WOJSKOWA AKADEMIA TECHNICZNA
im. Jarosława Dąbrowskiego

WYDZIAŁ INŻYNIERII LĄDOWEJ I GEODEZJI

ROZPRAWA DOKTORSKA

OPRACOWANIE METODYKI PODWYŻSZENIA ROZDZIELCZOŚCI PRZESTRZENNEJ ZOBRAZOWAŃ Z MAŁYCH SATELITÓW Z ZASTOSOWANIEM ALGORYTMÓW SZTUCZNEJ INTELIGENCJI

Autor: **por. mgr inż. Kinga KARWOWSKA**
Promotor pracy: **dr hab. inż. Damian Wierzbicki, prof. WAT**

Warszawa 2024

Rozprawa doktorska wykonana przez doktoranta: por. mgr inż. Kingę Karwowską

dziedzina nauki: nauki inżynieryjno-techniczne

dyscyplina naukowa: inżynieria lądowa, geodezja i transport

STRESZCZENIE W JĘZYKU POLSKIM

Wraz z rozwojem technologicznym oraz dynamicznym rozwojem sektora kosmicznego zwiększa się zapotrzebowanie na systemy umożliwiające obserwację Ziemi. Coraz bardziej popularne stają się małe satelity, które w przeciwieństwie do klasycznych satelitów obrazujących są prostsze w produkcji oraz znacznie tańsze. Jednak mają one dość znaczącą wadę – ze względu na ich konstrukcję mają one znacznie słabszą rozdzielczość przestrzenną. Ponadto posiadają niedoskonałe układy optyczne, co w konsekwencji przekłada się na ich możliwości fizyczne do pozyskania wysokorozdzielczych zobrazowań satelitarnych. Satelity te pozyskują zobrazowania panchromatyczne, wielospektralne lub w obu trybach (lecz różnica między rozdzielczością przestrzenną obrazu panchromatycznego i wielospektralnego jest niewielka). W związku z tym, zobrazowania pozyskane przez małe satelity nie spełniają wymagań do ich wyostrzenia za pomocą klasycznych metod poprawy rozdzielczości przestrzennej. Dlatego też konieczne jest opracowanie odpowiedniej metody poprawy rozdzielczości przestrzennej wykorzystującej jedynie obraz o niskiej rozdzielczości.

Wykonanie analizy obszaru badawczego pozwoliło mi na zidentyfikowanie problemów badawczych związanych z poprawą rozdzielczości przestrzennej zobrazowań satelitarnych pozyskanych przez małe satelity. Ich rozwiązanie zawarłam w cyklu czterech powiązanych tematycznie artykułach naukowych. W ramach rozprawy doktorskiej zaproponowałam nowe podejście do poprawy rozdzielczości przestrzennej zobrazowań satelitarnych. Przedstawiona metodyka umożliwia czterokrotną poprawę rozdzielczości przestrzennej oraz co najmniej dwukrotny wzrost możliwości interpretacyjnych zobrazowań satelitarnych lub sekwencji obrazów o dowolnym rozmiarze.

W ramach prowadzonych badań zaproponowałam nową metodykę przetwarzania całych zobrazowań satelitarnych w kontekście automatycznej interpretacji zobrazowań. Rozwiązanie to składa się z trzech etapów: (1) podział zobrazowania na mniejsze fragmenty (kafelki) o rozmiarze odpowiadającym parametrowi wejściowemu sieci neuronowej (SN), (2) poprawa rozdzielczości przestrzennej za pomocą SN, (3) ponowne połączenie zrekonstruowanych kafelków w jeden spójny obraz. Do scalania kafelków zastosowałam okna czasowe, wykazując, że użycie okien Hann, Hann-Poisson, Bartlett-Hann lub Triangular pozwala na uzyskanie obrazu o porównywalnej jakości.

Kluczowym elementem metodyki jest generatywna sieć przeciwstawna, do której szkolenia zastosowałam wielokolumnowy dyskryminator oraz funkcję straty Wasserseina. W ramach badań udowodniłam, że zastosowanie dyskryminatora wielokolumnowego pozwala na znaczne przyspieszenie procesu treningu sieci. Ponadto zaproponowałam trójetapowy proces oceny jakości obrazów oszacowanych (SR), w którego skład wchodzi: (1) ocena globalna, powszechnie stosowana przez badaczy, w której wskaźniki są często zawyżane przez wysokie odwzorowanie powierzchni jednorodnych, (2) ocena lokalna, umożliwiająca identyfikację obszarów o niższej lub wyższej jakości oszacowania SR, (3) analiza rozkładu częstotliwości energii sygnału (PSD), pozwalająca na określenie wzrostu możliwości interpretacyjnych danych obrazowych.

Podsumowaniem prowadzonych badań było opracowanie metodyki poprawy rozdzielczości przestrzennej sekwencji obrazów pozyskanych przez małe satelity. W toku tych badań zweryfikowałam dotychczasowe metody wyostrzania danych uzyskanych techniką obrazowania dynamicznego. Analizując wyniki eksperymentów, postanowiłam dodatkowo rozbudować generator zaprojektowanej sieci GAN, ponieważ generowane przez niego obrazy miały charakterystyczną teksturę spowodowaną niską szybkością uczenia modelu. W tym celu rozszerzyłam model generatora o sieć koder-dekoder w kształcie litery "U". Wprowadzona modyfikacja pozwoliła na dodatkową poprawę jakości interpretacyjnej obrazów oszacowanych przez generator.

Opracowana metodyka pozwala na czterokrotną poprawę rozdzielczości przestrzennej oraz co najmniej dwukrotny wzrost potencjału interpretacyjnego zobrazowań i sekwencji obrazów pozyskanych przez małe satelity. Właściwość ta jest kluczowa w obszarze bezpieczeństwa, zwłaszcza w dobie rozwijania się konfliktów i szybko zmieniającego się środowiska geopolitycznego.

STRESZCZENIE W JĘZYKU ANGIELSKIM

With technological advancements and the dynamic development of the space sector, the demand for Earth observation systems is increasing. Small satellites are becoming increasingly popular, as they are more simple to manufacture and significantly cheaper compared to traditional imaging satellites. However, they have a significant drawback – due to their design, they have a much lower spatial resolution. Additionally, they possess imperfect optical systems, which consequently limits their ability to capture high-resolution satellite imagery. These satellites acquire either panchromatic, multispectral or simultaneously both of these types of imagery (though the difference in spatial resolution between panchromatic and multispectral images is minimal). Therefore, the imagery obtained by small satellites does not meet the requirements for resolution enhancement using classical spatial resolution improvement methods. Thus, developing an appropriate method for improving spatial resolution using only low-resolution images is essential.

Analyzing the study area allowed me to identify research problems related to improving the spatial resolution of satellite imagery obtained by small satellites. I addressed these issues in a series of four thematically related scientific articles. As part of my doctoral dissertation, I proposed a new approach to enhancing the spatial resolution of satellite imagery. The presented methodology enables a fourfold improvement in spatial resolution and at least a twofold increase in the interpretative possibilities of satellite imagery or image sequences of any size.

As part of my research, I proposed a new methodology for processing entire satellite images in the context of automatic image interpretation. This solution consists of three stages: (1) dividing the image into smaller fragments (tiles) of a size corresponding to the input parameter of the neural network (NN), (2) improving the spatial resolution using the NN, and (3) recombining the reconstructed tiles into a single coherent image. For merging the tiles, I applied time windows, demonstrating that the use of Hann, Hann-Poisson, Bartlett-Hann, or Triangular windows results in an image of comparable quality.

A key element of the methodology is the generative adversarial network (GAN), for which I used a multi-column discriminator and the Wasserstein loss function during training. My research demonstrated that employing a multi-column discriminator significantly accelerates the training process of the network. Additionally, I proposed a three-stage process for evaluating the quality of super-resolution (SR) images, which includes:

(1) a global assessment, commonly used by researchers, where indicators are often inflated by a high representation of homogeneous surfaces, (2) a local assessment, enabling the identification of areas with lower or higher SR estimation quality, and (3) an analysis of the signal energy frequency distribution (PSD), which allows for determining the increase in the interpretative possibilitiesof the image data.

The culmination of my research was the development of a methodology for improving the spatial resolution of image sequences obtained by small satellites. During this research, I reviewed existing methods for sharpening data acquired through dynamic imaging techniques. Analyzing the experimental results, I decided to further enhance the generator of the designed GAN, as the generated images exhibited a characteristic texture caused by the model's low learning rate. To address this, I extended the generator model with a U-shaped encoder-decoder network. This modification led to an additional improvement in the interpretative quality of the images estimated by the generator.

The developed methodology allows for a fourfold improvement in spatial resolution and at least a twofold increase in the interpretative potential of images and image sequences obtained by small satellites. This capability is crucial in the field of security, especially in the context of escalating conflicts and the rapidly changing geopolitical environment.

Spis treści

WYKAZ UŻYTYCH SKRÓTÓW

BSP   – bezzałogowy statek powietrzny

CC   – *ang. Correlation Coefficient*, współczynnik korelacji

CGAN   – *ang. Conditional Generative Adversarial Network*, warunkowe generatywne sieci przeciwstawne

CMOS   – *ang. complementary metal–oxide–semiconductor*

CMPSIS   – *ang. Complementary Metal-Oxide-Semiconductor Image Sensors*

CMV   – *ang. Complementary Metal-Oxide-Semiconductor Video*

CNN   – *ang. Convolutional Neural Network*, neuronowa sieć konwolucyjna

DOTA   – *ang. Dataset for Object deTection in Aerial Images*, zbiór danych do wykrywania obiektów na zdjęciach lotniczych

EDSR   – *ang. enhanced deep super-resolution network*

ERGAS   – *ang. Relative Dimensionless Global Error in Synthesis*, bezwymiarowy globalny błąd względny

ESPCNN   – *ang. Efficient SubPixel Convolutional Neural Network*

ESRGAN   – *ang. Enhanced Super-Resolution Generative Adversarial Networks*

FSRCNN   – *ang. Fast Super-Resolution Convolutional Neural Network*

GAN   – *ang. Generative Adversarial Network*, generatywna sieć przeciwstawna

GRD   – *ang. Ground Resolved Distance*, terenowa zdolność rozdzielcza

GSD   – *ang. Ground Sampling Distance*, rozdzielczość przestrzenna

GSN   – głębokie sieci neuronowe

HR   – *ang. high resolution image*, obraz o wysokiej rozdzielczości

IHS   – *ang. intensity-hue-saturation*, intensywność-odcień-nasycenie

JCR   – *ang. Journal Citation Reports*

LeFF   – *ang. Locally-enhanced Feed-Forward Network*

LEO   – *ang. low Earth orbit*, niska orbita okołoziemska

LeWin   – *ang. locally enhanced window*

LN   – *ang. normalization layer*, warstwa normalizacji

LR   – *ang. low resolution image*, obraz o niskiej rozdzielczości

NN   – *ang. neural network*, sieć neuronowa

PCA   – *ang. principal component analysis*, analiza głównych składowych

PSD   – *ang. power spectral density*, gęstość widmowa mocy

PSNR   – *ang. peak signal-to-noise ratio*, szczytowy stosunek sygnału do szumu

QE   – *ang. quantum efficiency*, wydajność kwantowa

RASE   – *ang. Relative Average Spectral Error*, względny średni błąd spektralny

SAM   – *ang. Spectral Angle Mapper*, porównanie krzywych spektralnych

SAN   – *ang. Second-order Attention Network*

SID   – *ang. Spectral Information Divergence*, rozbieżność informacji spektralnych

SISR   – *ang. Single Image Super-Resolution*, super rozdzielczość pojedynczego obrazu

SR   – *ang. super resolution image*, obraz o super rozdzielczości (obraz wyostrzony za pomocą metod SISR)

SRCNN   – *ang. Super-Resolution Convolutional Neural Network*

SRGAN   – *ang. Super-Resolution Generative Adversarial Networks*

SRResNet   – *ang. Super Resolution Residual neural Network*

SSIM   – *ang. structural similarity index*, podobieństwo strukturalne

THR   – *fr. Très Haute Résolution,* bardzo wysoka rozdzielczość

UIQI   – *ang. Universal Image Quality Index*, uniwersalny wskaźnik jakości obrazu

W-MSA   – *ang. Window-based Multi-head Self-Attention*

WV2   – *ang. WorldView-2*

1. Wprowadzenie

Według bazy danych UCS Satellite [1] (na dzień 1 stycznia 2023 r.) od 1 stycznia 2000 r. na orbitę wyniesionych zostało ponad 1250 satelitów obserwacyjnych, a ponad 60% z nich umieszczono na orbicie w latach 2018-2022. Współczesne satelity wysoko-rozdzielcze rejestrują obrazy z rozdzielczością przestrzenną sięgającą w nadirze nawet 0.30 m, jednak coraz bardziej popularne stają się małe satelity, do których należą systemy nano, mikro oraz mini satelitarne [2], stanowiące ponad 70% satelitów umieszczonych na orbicie w latach 2018-2022. Systemy te, choć są łatwiejsze w produkcji niż duże satelity obrazujące oraz od nich tańsze, mają znaczącą (biorąc pod uwagę możliwości obserwacji Ziemi) wadę – ze względu na ich konstrukcję, mają one znacznie słabszą rozdzielczość przestrzenną. Niewielki rozmiar oraz waga (nanosatelity: 1-10kg, micro satelity: 10-100kg, mini satelity: 100-500kg [3]) determinują ograniczenia wyposażenia. Najważniej-szym elementem satelity obserwacyjnego jest układ optyczny, który w małych satelitach znacznie różni się od tych montowanych w dużych satelitach obserwacyjnych, takich jak WorldView-3 czy QuickBird. Małe satelity wyposażone są w matrycę o niewielkich roz-miarach np. CMOS [4] czy CMOSIS CMV [5], które charakteryzują się niską wydajno-ścią kwantową (ang. quantum efficiency – QE). Kolejnym ograniczeniem niewielkich satelitów jest niedoskonałość teleskopu, ponieważ do ich konstrukcji wykorzystywana jest znacznie mniejsza liczba soczewek, co prowadzi do obniżenia jakości pozyskanych zobrazowań m.in. poprzez rozmycie. Konsekwencją zastosowania matryc o niskiej wy-dajności i niedoskonałych teleskopów jest znacznie niższa rozdzielczość, mimo że sate-lity poruszają się po niskiej orbicie okołoziemskiej (ang. low Earth orbit - LEO).

Rozwój technologiczny sektora kosmicznego tworzy nowe możliwości obserwa-cji Ziemi. Pozyskane dane pozwalają na wykonywanie analiz znajdujących zastosowanie w wielu dziedzinach nauk m.in w monitorowaniu zmian czy analizach geoprzestrzen-nych. Informacje te mają zastosowanie np. w administracji publicznej, zarządzaniu kry-zysowym lub wojsku. To, jak pozyskane dane będą wykorzystane, determinuje ich roz-dzielczość. Dzięki wysokiej rozdzielczości czasowej możliwe jest wykonanie np. detekcji zmian. Rozdzielczość spektralna wyznacza zakres możliwych analiz w zakresie telede-tekcji, w tym analizy pokrycia terenu. Kolejnym ważnym parametrem jest rozdzielczość przestrzenna (ang. Ground Sampling Distance – GSD), która determinuje możliwości de-tekcji obiektów na zobrazowaniu. Ten parametr jest kluczowy w obszarze

bezpieczeństwa. Obecnie obserwowana jest i złożona sytuacja geopolityczna, a jej charakter wynika z wielu równocześnie występujących czynników. Na całym świecie obserwujemy napięcia, które mają potencjalnie dalekosiężne konsekwencje dla stabilności międzynarodowej. W dobie rozwijania się konfliktów i szybko zmieniającego się środowiska geopolitycznego, wysokorozdzielcze dane obrazowe są głównym źródłem informacji nie tylko na temat działań w obszarach objętych konfliktem, ale także budują świadomość sytuacyjną w zakresie działań państw leżących w obszarze operacyjnego zainteresowania.

Istnieje wiele metod poprawy rozdzielczości przestrzennej obrazów wielospektralnych [6]. Jedną z nich jest pansharpening, który łączy ze sobą wysoką rozdzielczość przestrzenną obrazu panchromatycznego z wysoką rozdzielczością spektralną[1] obrazu wielospektralnego [7], [8]. Dzięki tej operacji powstaje obraz wielospektralny o wysokiej rozdzielczości przestrzennej. Warunkiem koniecznym do zastosowania metody pansharpeningu jest dysponowanie obrazem o wysokiej rozdzielczości. Jednak nie wszystkie satelity pozyskują zobrazowania satelitarne w zakresie panchromatycznym i wielospektralnym. Jednym z rozwiązań tego problemu jest zastosowanie metod cyfrowego przetwarzania obrazu, do których należy algorytm liniowej rekonstrukcji zmiennych pikseli (Super-Resolution Variable-Pixel Linear Reconstruction) [9], czy próbkowanie quincunx (THR - Très Haute Résolution) wykorzystywane w przetwarzaniu zobrazowań pozyskiwanych przez SPOT5 [10], [11]. Metody wykorzystujące cyfrowe przetwarzanie obrazów dość rzadko stosowane są do poprawy rozdzielczości spektralnej zobrazowań satelitarnych ze względu na duże wymagania pozyskania scen satelitarnych, czy też ze względu na konieczność pozyskania wielu scen przedstawiających ten sam obszar przez więcej niż jedną satelitę w podobnym czasie. Innym rozwiązaniem jest zastosowanie algorytmów wykorzystujących głębokie sieci neuronowe. Grupa tych rozwiązań do poprawy rozdzielczości przestrzennej wykorzystuje jedynie wytrenowany model sieci neuronowej oraz obraz o niskiej rozdzielczości. W przypadku klasycznych rozwiązań (tj. interpolacja, pansharpening) na podstawie danych wejściowych (obraz o niskiej rozdzielczości- LR) przygotowywany jest algorytm odpowiedzialny za wyostrzenie obrazu. Natomiast, dla rozwiązań bazujących na sieciach neuronowych, przygotowanie algorytmu jest operacją bardziej złożony. Proces ten, na etapie uczenia, wymaga danych wejściowych (obrazów

---

[1] Jako zobrazowania o wysokiej rozdzielczości spektralnej rozumiem sceny satelitarne o co najmniej ośmiu kanałach.

LR), tożsamych obrazów docelowych (obrazów o wysokiej rozdzielczości – HR) oraz architektury sieci neuronowej. Następnie, w procesie uczenia na podstawie danych wejściowych i docelowych, odbywa się przygotowanie modelu, polegające na oszacowaniu wag łączących poszczególne warstwy. Można zatem stwierdzić, że algorytm wyostrzania powstaje w oparciu o dostarczone dane.

W 2014 r. zespół kierowany przez Goodfellowa dokonał przełomu w dziedzinie technologii generowania obrazów o realistycznym wyglądzie. W swojej publikacji [12] zaproponował generatywną sieć przeciwstawną (ang. generative adversarial networks – GAN). Sieć GAN składa się z dwóch sieci splotowych: generatora i dyskryminatora, jednakże dyskryminator jest wykorzystywany jedynie podczas etapu uczenia. Zadaniem generatora jest tworzenie realistycznych obrazów o wyższej rozdzielczości (ang. super resolution image, SR) na podstawie obrazów niskiej rozdzielczości (LR). Z kolei dyskryminator ocenia, czy dostarczony obraz należy do zbioru obrazów wysokiej rozdzielczości (HR) oraz czy został stworzony przez generator (rys. 1.1).

Rys. 1.1. *Proces uczenia sieci GAN na potrzeby poprawy rozdzielczości przestrzennej obrazów*
[źródło: opracowanie własne]

W procesie uczenia sieci GAN wagi dyskryminatora są zamrożone, co oznacza, że nie są one aktualizowane. Dzięki temu podczas treningu, jedynie wagi generatora podlegają zmianom. Gdyby wagi dyskryminatora nie były zamrożone, klasyfikator ten oceniałby obrazy według różnych standardów w każdej iteracji. W rezultacie, dyskryminator mógłby nauczyć się klasyfikować obrazy SR jako rzeczywiste z wysokim prawdopodobieństwem, co jest niepożądanym efektem. Mimo że sieci GAN stanowią przełom w generowaniu obrazów o realistycznym wyglądzie, ich trening jest bardzo wymagający. Sieci GAN, w przeciwieństwie do klasycznych sieci splotowych, nie mają ustalonego

minimum optymalizacji. W tradycyjnych rozwiązaniach optymalizator aktualizuje wagi na podstawie funkcji straty, a wprowadzone zmiany mają jedynie niewielki wpływ na uzyskane wyniki. Dzięki temu model trudno jest wyprowadzić z równowagi. Natomiast w sieciach GAN zmiana wartości funkcji straty powoduje znaczne zmiany w generowanych obrazach SR. W efekcie, sieci GAN wymagają precyzyjnego doboru nie tylko architektury, ale również parametrów uczenia.

Obrazy pozyskane przez małe satelity, do poprawy rozdzielczości przestrzennej wymagają zastosowania metod z grupy super rozdzielczości pojedynczego obrazu (SISR). Jak przedstawiono powyżej, jednym z możliwych rozwiązań są sieci GAN. Choć ich trening stanowi duże wyzwanie, ich zdolność do generowania realistycznych obrazów jest silną motywacją do podjęcia tego działania.

2. Tytuł rozprawy, cel, hipoteza, cykl publikacyjny

Rozprawę doktorską pt. *Opracowanie metodyki podwyższenia rozdzielczości przestrzennej zobrazowań z małych satelitów z zastosowaniem algorytmów sztucznej inteligencji* stanowi cykl czterech powiązanych tematycznie artykułów naukowych. Zawierają one autorskie propozycje rozwiązań w zakresie poprawy rozdzielczości całych zobrazowań satelitarnych i sekwencji obrazów pozyskanych przez małe satelity. Proponowane rozwiązania zostały porównane z popularnymi i powszechnie stosowanymi rozwiązaniami SISR wykorzystującymi głębokie sieci neuronowe. Do oceny jakości zrekonstruowanych zobrazowań satelitarnych wykorzystałam opracowany w ramach badań zbiór metod oceny jakościowej zawierający popularne wskaźniki stosowane w ocenie jakości zobrazowań optycznych i widzenia komputerowego oraz analizę obrazu w dziedzinie częstotliwości.

Małe satelity z uwagi na swoją konstrukcję posiadają niedoskonałe układy optyczne, co w konsekwencji przekłada się na ich możliwości fizyczne do pozyskania wysokorozdzielczych zobrazowań satelitarnych. Satelity te pozyskują zobrazowania panchromatyczne, wielospektralne lub w obu trybach. Należy jednak zwróć uwagę, że w przypadku obrazów pozyskiwanych przez małe satelity, różnica między rozdzielczością przestrzenną obrazu panchromatycznego i wielospektralnego jest niewielka (np. dla satelitów SkySat-3 - SkySat-15 to jedynie 0.16m). Dodatkowy problem pojawia się w przypadku potrzeby zwiększenia rozdzielczości przestrzennej obrazu, który nie ma swojego odpowiednika o wyższej rozdzielczości. W konsekwencji, **zobrazowania pozyskane przez małe satelity nie spełniają wymagań do ich wyostrzenia za pomocą klasycznych metod poprawy rozdzielczości. W związku z tym, konieczne jest zastosowanie metod bazujących jedynie na obrazach o niskiej rozdzielczości przestrzennej. Zagadnienie to jest pierwszym problemem badawczym**, którego rozwiązanie znajduje się w niniejszej pracy. W związku z powyższym, w prowadzonych badaniach za **cel główny (CG)** przyjęłam **opracowanie metodyki wykorzystującej algorytmy sztucznej inteligencji na potrzeby podwyższenia rozdzielczości przestrzennej oraz poprawy potencjału interpretacyjnego zobrazowań z małych satelitów.**

Drugi problem badawczy dotyczy niedostosowania istniejących metod SISR do poprawy rozdzielczości przestrzennej całych zobrazowań satelitarnych. Wykonując przegląd obszaru badawczego zauważyłam, że badacze prezentują swoje wyniki

w postaci małych obrazów, o rozmiarze równym wejściu do pierwszej warstwy sieci neuronowej[2]. Rozmiar ten jest znacznie mniejszy od rozmiaru zobrazowań satelitarnych pozyskanych przez małe satelity. W związku z tym, konieczne jest opracowanie metodyki pozwalającej na poprawę rozdzielczości przestrzennej całych zobrazowań satelitarnych przez dowolną sieć neuronową.

Trzeci problem badawczy dotyczy uodpornienia generatywnej sieci przeciwstawnej na zjawisko znikających lub eksplodujących gradientów. Jedną z cech, która odróżnia sieci GAN od klasycznych sieci splotowych jest to, że nie posiadają one ustalonego minimum optymalizacji. Sieci GAN, to system, który nie szuka minimum, a równowagi między dwoma modelami. W przypadku klasycznych sieci splotowych, algorytm spadku gradientowego śledzi funkcję straty, a wprowadzane zmiany wag oddziałują tylko nieznacznie na wyniki. Natomiast, w przypadku sieci GAN, każda aktualizacja wag powoduje znaczną zmianę generowanych wyników (zjawisko to jest bardzo dobrze widoczne w pierwszych krokach uczenia sieci GAN), przez co **trening sieci GAN charakteryzuje się niestabilnym procesem uczenia**. Z tego powodu sieć GAN wymaga dokładnego dopasowania architektury modelu oraz parametrów uczenia.

Czwarty problem badawczy dotyczy właściwej oceny obrazów, których rozdzielczość przestrzenna została podwyższona za pomocą metod SISR. Wykonując przegląd obszaru badawczego zauważyłam, że wskaźniki wykorzystywane do oceny jakościowej obliczane są dla całych obrazów. W związku z tym, wyniki te mogą być zawyżone w przypadku występowania powierzchni jednorodnych na obrazie. Z tego powodu wiarygodność wykonanej analizy maleje. Ponadto wartości wskaźników określających podobieństwo między obrazami nie pozwala na określenie wielkości poprawy możliwości interpretacyjnych oszacowanego obrazu[3].

W oparciu o cel główny **(CG)**, który dotyczy opracowania metodyki wykorzystującej algorytmy sztucznej inteligencji na potrzeby podwyższenia rozdzielczości

---

[2] Rozmiar wejścia do sieci – *ang. input size*, parametr określający rozmiar danych wejściowych, które sieć może przetworzyć. W przypadku obrazów, definiuje jego wysokość, szerokość oraz liczbę kanałów, jakie sieć przyjmuje na swojej pierwszej warstwie.

[3] Obraz oszacowany – obraz, którego rozdzielczość przestrzenna została poprawiona za pomocą splotowych sieci neuronowych (CNN), które przewidują, jak mógłby wyglądać obraz wejściowy przy wyższej rozdzielczości przestrzennej. Nazywany również obrazem o super rozdzielczości (SR).

przestrzennej oraz poprawy potencjału interpretacyjnego zobrazowań z małych satelitów, przyjęłam następujące cele szczegółowe:

CS.1. przygotowanie charakterystyki obszaru badawczego,

CS.2. opracowanie metodyki przetwarzania całych zobrazowań satelitarnych w zagadnieniach automatycznej interpretacji zobrazowań,

CS.3. opracowanie strategii budowy bazy danych do uczenia sieci GAN odpowiedzialnych za poprawę rozdzielczości przestrzennej,

CS.4. wybór funkcji straty, uwzględniając jej odporność na zjawisko znikających gradientów,

CS.5. opracowanie zbioru metod oceny jakości działania algorytmów poprawy rozdzielczości przestrzennej obrazów cyfrowych,

CS.6. opracowanie metodyki poprawy rozdzielczości sekwencji obrazów pozyskanych przez małe satelity.

W związku z powyższym podczas realizacji prac badawczych przyjęłam następujące założenia, które były podstawą do opracowana metodyki podwyższenia rozdzielczości przestrzennej zobrazowań z małych satelitów:

- sieci GAN pozwalają na generowanie realistycznych obrazów,
- okna czasowe[4] mogą zostać wykorzystane do tworzenia macierzy wag, na podstawie której łączone będą kafelki obrazu,
- strata Wassersteina[5] uwzględniająca ograniczenie Lipchitza[6] poprawia stabilność procesu optymalizacji.

Uwzględniając powyższe założenia oraz pytania badawcze, postawiłam następującą hipotezę główną: **Zastosowanie odpowiednio dobranych i wytrenowanych generatywnych sieci przeciwstawnych (GAN) pozwoli na opracowanie metodyki podwyższania rozdzielczości przestrzennej zobrazowań z małych satelitów.** Ponadto w prowadzonych badaniach przyjęłam pięć hipotez pomocniczych:

---

[4] Okno czasowe – *ang. window function*, funkcja opisująca sposób pobierania próbek z sygnału.

[5] Funkcja straty Wassersteina – *ang. Wesserstein loss*, to miara używana w trenowaniu sieci generatywno-przeciwstawnych (GAN), która ocenia różnicę między rozkładami rzeczywistych i generowanych danych, dążąc do minimalizacji odległości Wassersteina między tymi rozkładami.

[6] Ograniczenie Lipchitza – ang. *Lipschitz constraint*, definiuje, aby funkcja krytyka (dyskryminatora) była 1-Lipschitz ciągła, co oznacza, że jej gradient nie przekracza wartości 1. To ograniczenie zapewnia stabilność trenowania i poprawność obliczania odległości Wassersteina.

H1.   Klasyczne metody poprawy rozdzielczości przestrzennej zobrazowań satelitarnych zazwyczaj nie znajdują zastosowania w przypadku zobrazowań pozyskanych przez małe satelity.

H2.   Podział zobrazowań satelitarnych na mniejsze fragmenty (kafelki[7]), przetworzenie ich przez sieć konwolucyjną (dla zagadnień automatycznej interpretacji zobrazowań), a następnie ponowne połączenie fragmentów obrazu w jedno zobrazowanie za pomocą okien czasowych umożliwia przetwarzanie zobrazowań o dowolnym rozmiarze.

H3.   Przygotowanie bazy danych obrazów składającej się z fragmentów zobrazowań satelitarnych o różnej rozdzielczości przestrzennej ogranicza występowanie zjawiska znikających gradientów, a w konsekwencji ma korzystny wpływ na stabilność treningu sieci generatywnych.

H4.   Zastosowanie dyskryminatora wielokolumnowego[8] umożliwia lepszą ocenę obrazów na etapie treningu sieci generatywnej. Ponadto wykorzystanie straty Wessersteina i pozwala na lepsze odróżnienie obrazów o wysokiej rozdzielczości od obrazów rekonstruowanych, co dodatkowo poprawia jakość treningu generatora, a w konsekwencji jakość obrazów rekonstruowanych (SR).

H5.   Rozbudowanie architektury generatora o architekturę koder-dekoder[9] wykorzystującą połącznia pomijające (ang. skip connection[10]), pozwala na lepszą rekonstrukcję tekstury obrazu, a w konsekwencji dodatkowo poprawia możliwości interpretacyjne obrazów SR.

Badania stanowiące podstawę niniejszej rozprawy doktorskiej zostały przedstawione w powiązanych tematycznie artykułach naukowych, opublikowanych w czasopismach wyróżnionych w Journal Citation Reports (JCR). Suma punktów MEiN

---

[7] Kafelek – *ang. tile,* pojedynczy obraz będący wynikiem podziału dużego obrazu (np. zobrazowania satelitarnego) na mniejsze kawałki o zdefiniowanym rozmiarze i ustalonym pokryciu.

[8] Dyskryminator wielokolumnowy – *ang. multicolumn discriminator*, dyskryminator złożony z dwóch klasyfikatorów pracujących równolegle.

[9] Koder-dekoder – *ang. encoder-decoder*, znana również pod nazwą autoenkoder. Jest to sieć neuronowa składająca się z sieci kodera (odpowiedzialnego za kompresję wielowymiarowych danych wejściowych do wektora reprezentacji) oraz sieci dekodera (odpowiedzialnego za dekompresję wektora reprezentacji do wymiaru oryginalnego).

[10] Połączenia pomijające – *ang. skip connection*, mechanizm w sieciach neuronowych, który polega na bezpośrednim przekazywaniu sygnału z jednej warstwy do kolejnej, z pominięciem jednej lub więcej warstw pośrednich.

za publikacje uwzględnione w cyklu wynosi 480, a sumaryczny Impact Factor czasopism to 19.3 (z uwzględnieniem % udziału: punkty – 384, IF – 15.44).

Tabela 2.1.

Wykaz artykułów naukowych, z których składa się cykl [źródło: opracowanie własne]

Ozn.	Artykuł z cyklu	Punkty wg. MEiN, IF
P1	Karwowska K. (80%), Wierzbicki D. (20%), 2022, Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3292-3312, DOI: 10.1109/JSTARS.2022.3167646	140 pkt IF = 3.8
P2	Karwowska K. (80%), Wierzbicki D. (20%), 2022, Improving Spatial Resolution of Satellite Imagery Using Generative Adversarial Networks and Window Functions, *Remote Sensing*, 14(24):6285, DOI: 10.3390/rs14246285	100 pkt IF = 5.0
P3	Karwowska K. (80%), Wierzbicki D. (20%), 2023, MCWESRGAN: Improving Enhanced Super-Resolution Generative Adversarial Network for Satellite Images, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 9459-9479, DOI: 10.1109/JSTARS.2023.3322642	140 pkt IF = 5.5
P4	Karwowska K. (80%), Wierzbicki D. (20%), 2024, Modified ESRGAN with Uformer for Video Satellite Imagery Super-Resolution, *Remote Sensing*, *16(11):*1926. doi: 10.3390/rs16111926	100 pkt IF = 5.0

Powyższe publikacje stanowią odpowiedzi na rozpatrywane w cyklu hipotezy. Odnoszą się one do czterech rozpatrywanych w cyklu problemów badawczych:

- **problem badawczy 1**: klasyczne metody poprawy rozdzielczości przestrzennej nie pozwalają na poprawę możliwości interpretacyjnych zobrazowań pozyskanych przez małe satelity – **P1, P2, P3, P4**;

- **problem badawczy 2**: istniejące metody SISR nie są dostosowane do poprawy rozdzielczości przestrzennej obrazów o dużym rozmiarze – **P2**;

- **problem badawczy 3**: istniejące metody poprawy rozdzielczości przestrzennej wykorzystujące sieci GAN, charakteryzują się niestabilnym procesem uczenia – **P4**;
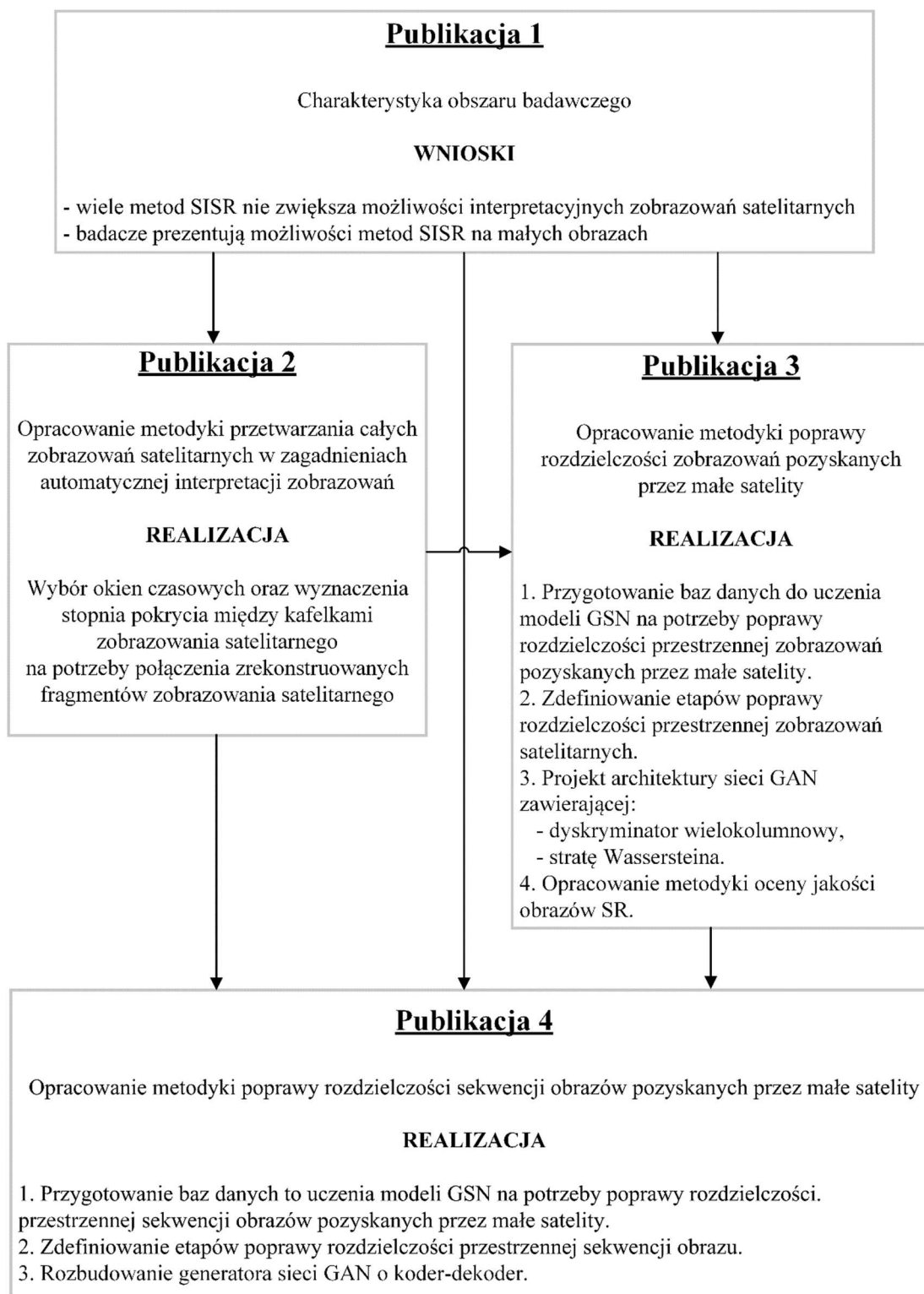
- **problem badawczy 4**: brak właściwych metod oceny zobrazowań satelitarnych, których rozdzielczość przestrzenna została poprawiona za pomocą algorytmów wykorzystujących głębokie sieci neuronowe – **P3**;

3. Metodyka badawcza i schemat postępowania

W rozprawie doktorskiej zaprezentowałam autorską metodykę podwyższania rozdzielczości przestrzennej zobrazowań pozyskanych przez małe satelity z zastosowaniem algorytmów sztucznej inteligencji. Metodyka ta obejmuje:

- wybór okien czasowych oraz wyznaczenia stopnia pokrycia między kafelkami zobrazowania satelitarnego na potrzeby połączenia zrekonstruowanych fragmentów zobrazowania satelitarnego,

- przygotowania baz danych do uczenia modeli sieci neuronowych na potrzeby poprawy rozdzielczości przestrzennej zobrazowań i sekwencji wideo pozyskanych przez małe satelity,

- zdefiniowanie etapów poprawy rozdzielczości przestrzennych zobrazowań oraz sekwencji obrazów pozyskanych przez małe satelity,

- projekt architektury generatywnej sieci przeciwstawnej umożliwiającej poprawę rozdzielczości przestrzennej zobrazowań satelitarnych, zawierającej:
  - dyskryminator wielokolumnowy,
  - stratę Wessersteina,
  - generator rozbudowany o koder-dekoder (w kształcie „U") wykorzystujący połączenia pomijające,

- opracowanie metodyki oceny jakości obrazów zrekonstruowanych przez głębokie sieci neuronowe, jednocześnie uwzględniając ocenę poprawy możliwości interpretacyjnych.

Schemat przyjętej metodyki, wraz z odniesieniem do publikacji przedstawiono na poniższym rysunku 3.1.

Publikacja 1

Charakterystyka obszaru badawczego

WNIOSKI

- wiele metod SISR nie zwiększa możliwości interpretacyjnych zobrazowań satelitarnych
- badacze prezentują możliwości metod SISR na małych obrazach

Publikacja 2

Opracowanie metodyki przetwarzania całych zobrazowań satelitarnych w zagadnieniach automatycznej interpretacji zobrazowań

REALIZACJA

Wybór okien czasowych oraz wyznaczenia stopnia pokrycia między kafelkami zobrazowania satelitarnego na potrzeby połączenia zrekonstruowanych fragmentów zobrazowania satelitarnego

Publikacja 3

Opracowanie metodyki poprawy rozdzielczości zobrazowań pozyskanych przez małe satelity

REALIZACJA

1. Przygotowanie baz danych do uczenia modeli GSN na potrzeby poprawy rozdzielczości przestrzennej zobrazowań pozyskanych przez małe satelity.
2. Zdefiniowanie etapów poprawy rozdzielczości przestrzennej zobrazowań satelitarnych.
3. Projekt architektury sieci GAN zawierającej:
   - dyskryminator wielokolumnowy,
   - stratę Wassersteina.
4. Opracowanie metodyki oceny jakości obrazów SR.

Publikacja 4

Opracowanie metodyki poprawy rozdzielczości sekwencji obrazów pozyskanych przez małe satelity

REALIZACJA

1. Przygotowanie baz danych to uczenia modeli GSN na potrzeby poprawy rozdzielczości. przestrzennej sekwencji obrazów pozyskanych przez małe satelity.
2. Zdefiniowanie etapów poprawy rozdzielczości przestrzennej sekwencji obrazu.
3. Rozbudowanie generatora sieci GAN o koder-dekoder.

Rys. 3.1. *Schemat graficzny przyjętej metodyki badawczej [źródło: opracowanie własne]*

W wyniku realizacji rozprawy doktorskiej wykonana została szczegółowa analiza obszaru badawczego (publikacja **P1**). Na podstawie wniosków wynikających z przeprowadzonej analizy, opracowana została metodyka podwyższenia rozdzielczości

przestrzennej zobrazowań z małych satelitów z zastosowaniem algorytmów sztucznej inteligencji (publikacja **P2** i **P3**). Metodyka ta została także zweryfikowana w zadaniu poprawy rozdzielczości przestrzennej sekwencji obrazów pozyskanych przez małe satelity (publikacja **P4**). W związku z tym, że wykorzystanie tych metod odbywa się na etapie opracowania danych (po pozyskaniu i przesłaniu na stację naziemną), rozwiązania te mogą znaleźć zastosowaniE w przypadku istniejących, jak i projektowanych małych satelitów. Zobrazowania (lub sekwencje obrazów) przetworzone przez proponowane w ramach pracy metody charakteryzują się znacznym wzrostem możliwości interpretacyjnych. Dzięki tej właściwości, zobrazowania o podwyższonej rozdzielczości przestrzennej mogą być wykorzystane w Wojsku Polskim (rozpoznanie obrazowe), jak i administracji publicznej (zarządzanie kryzysowe - analizy zmian).

Prace badawcze zrealizowane zostały etapami, poprzez opracowanie celów szczegółowych. Zestawienie problemów badawczych i celów szczegółowych wraz z metodami i sposobami ich rozwiązania przedstawione jest w tabeli 3.1.

Tabela 3.1.

Zestawienie celów szczegółowych wraz z metodami i sposobami ich rozwiązania [źródło: opracowanie własne]

Problem badawczy	Cele prac badawczych	Odniesienie do publikacji	Odniesienie do hipotezy
Klasyczne metody poprawy rozdzielczości przestrzennej nie pozwalają na poprawę możliwości interpretacyjnych zobrazowań pozyskanych przez małe satelity.	Charakterystyka obszaru badawczego.	**P1**	**H1, H4, H5**
	Opracowanie metodyki poprawy rozdzielczości zobrazowań pozyskanych przez małe satelity.	**P2, P3**	
	Opracowanie metodyki poprawy rozdzielczości sekwencji obrazów pozyskanych przez małe satelity.	**P2, P4**	
Istniejące metody SISR nie są dostosowane do poprawy rozdzielczości przestrzennej obrazów o dużym rozmiarze.	Opracowanie metodyki przetwarzania całych zobrazowań satelitarnych w zagadnieniach automatycznej interpretacji zobrazowań.	**P2**	**H2**

Istniejące metody poprawy rozdzielczości przestrzennej wykorzystujące sieci GAN, charakteryzują się niestabilnym procesem uczenia.	Opracowanie strategii budowy bazy danych do uczenia sieci GAN odpowiedzialnych za poprawę rozdzielczości przestrzennej.	**P3, P4**	**H3**
	Wybór straty uczenie, uwzględniając jej odporność na zjawisko znikających gradientów.		
Niewystarczające metody oceny zobrazowań satelitarnych, których rozdzielczość przestrzenna została poprawiona za pomocą algorytmów wykorzystujących głębokie sieci neuronowe	Opracowanie zbioru metod oceny jakości działania algorytmów poprawy rozdzielczości przestrzennej obrazów cyfrowych.	**P3**	**H5**

Opisy poszczególnych problemów badawczych znajdują się w kolejnym rozdziale niniejszej pracy. W kolejnych podrozdziałach znajdują się opisy badań, których podsumowaniem były artykuły należące do cyklu publikacyjnego. Podrozdział 4.1 zawiera szczegółowy przegląd metod poprawy rozdzielczości przestrzennej obrazów cyfrowych. Podrozdział 4.2 przedstawia opis metodyki mozaikowania scen satelitarnych za pomocą okien czasowych. W podrozdziale 4.3 zaprezentowany został model GAN pozwalającej na poprawę rozdzielczości przestrzennej zobrazowań satelitarnych oraz przedstawiony został zbiór metod, dzięki któremu możliwa jest rzetelna ocena obrazów SR. Natomiast w podrozdziale 4.4 zaprezentowano metodykę poprawy rozdzielczości przestrzennej sekwencji obrazów pozyskanych przez małe satelity.

4. Metody szczegółowe i badawcze

4.1. Charakterystyka obszaru badawczego [publikacja P1]

W ostatnich latach opublikowano wiele prac naukowych poświęconych poprawie rozdzielczości przestrzennej zobrazowań satelitarnych. Opracowane metody można podzielić na podstawie kilku kryteriów. Pierwszym z nich jest podział według liczby obrazów potrzebnych do oszacowania obrazu o wysokiej rozdzielczości. Tu można wyróżnić metody wykorzystujące informacje pochodzące z wielu obrazów oraz te, którym wystarczy tylko jeden obraz. Kolejny podział uwzględnia sposób działania algorytmów. Są to metody bazujące na interpolacji, pansharpeningu, innych algorytmach cyfrowego przetwarzania obrazu oraz wykorzystujące algorytmy uczenia głębokiego.

Do przygotowania charakterystyki obszaru badawczego w ramach publikacji P1 wykonano przegląd metod poprawy rozdzielczości przestrzennej, od powszechnie stosowanych rozwiązań wykorzystujących interpolację, po zyskujące popularność algorytmy głębokiego uczenia.

W przygotowanej publikacji starano się znaleźć odpowiedzi na następujące pytania badawcze:

PB1.1. Jakie są główne problemy podwyższania rozdzielczości zobrazowań satelitarnych?

PB1.2. Jak wyglądał rozwój algorytmów poprawy rozdzielczości?

PB1.3. Jakie są różnice pomiędzy metodami podwyższania rozdzielczości przestrzennej zobrazowań satelitarnych?

PB1.4. Jak tworzone są bazy danych obrazów na potrzeby podwyższenia rozdzielczości zobrazowań z małych satelitów za pomocą algorytmów głębokiego uczenia?

PB1.5. Jak wykonywana jest ocena skuteczności działania algorytmów odpowiedzialnych za poprawę rozdzielczości przestrzennej?

Wśród najpopularniejszych metod pozwalających na nadanie wartości nowo powstałym elementom macierzy obrazu są algorytmy interpolacji obrazu. Interpolacja jest wykorzystywana do rozwiązania problemu poprawy rozdzielczości dzięki swojej prostocie i szybkości obliczeń. Jest ona stosowana w metodach SISR- wykorzystujących do szacowania obrazu wyższej rozdzielczości (SR) tylko jeden obraz o niskiej rozdzielczości. Istnieje wiele metod interpolacji pozwalających na stworzenie nowego obrazu SR

o większej rozdzielczości. Do podstawowych metod można zaliczyć m.in. interpolację najbliższego sąsiada, dwuliniową oraz dwusześcienną. Metody te pozwalają na przypisanie nowo powstałym pikselom wartości jasności, jednak wpływają na zwiększenie rozmiaru obrazu, a nie ilość informacji, które mogą zostać z niego odczytane. Metody interpolacji są wykorzystywane w wielu algorytmach poprawy rozdzielczości przestrzennej obrazu, które poza etapem interpolacji, usuwają rozmycie oraz minimalizują występowanie szumów, co pozwala nie tylko na zmianę rozmiaru macierzy reprezentującej obraz wyjściowy, ale też poprawę jego jakości np. metoda Partial Volume Diffusion [13], bazujące na krawędziach [14], [15]. Jak już wspomniano, istnieją również metody, które do poprawy rozdzielczości wykorzystują więcej niż jeden obraz. Przykładem takiego rozwiązania jest połączenie interpolacji z metodami gradientu. Polega ono na połączeniu obrazów o niskiej rozdzielczości z obrazem o wysokiej rozdzielczości. W sposobie tym, iteracje metody gradientu są wykonywane w siatce współrzędnych widmowych i przestrzennych o wysokiej rozdzielczości [16]. Inne rozwiązanie zaproponował zespół X. Qifang [17], w którym do poprawy rozdzielczości przestrzennej wykorzystano serię obrazów o niskiej rozdzielczości oraz informacji z poruszających się pikseli reprezentujących ten sam obiekt.

W środowisku teledetekcyjnym bardzo popularny jest pansharpening, który łączy wysoką rozdzielczość przestrzenną zobrazowań panchromatycznych z wysoką rozdzielczością spektralną zobrazowań wielospektralnych. Rezultatem zastosowania pansharpeningu jest obraz wielospektralny o wysokiej rozdzielczości. Do głównych rozwiązań zaliczamy metodę IHS (ang. intensity-hue-saturation, intensywność-odcień-nasycenie) [7], analiza głównych składowych (ang. principal component analysis – PCA) [18], transformacje Broveya [19], [20], fuzję Ehlersa [21], wymnażanie obrazów [22] i metodę Gram-Schmidta [23].

W związku z tym, że naukowcy wciąż próbują ulepszać istniejące rozwiązania oraz tworzą kolejne metody, istnieje szereg nowych badań, z których wybrane zostały opisane w publikacji **P1** (wady i zalety w tabeli II i III, a ocena jakościowa w tabeli IV publikacji **P1**). Obrazy będące wynikiem poprawy rozdzielczości za pomocą pansharpeningu charakteryzują się znacznym wzrostem potencjału interpretacyjnego (w porównaniu z obrazami o niskiej rozdzielczości). Jednak mają one dwie zasadnicze wady: (1) wiele metod powoduje zniekształcenia widmowe, (2) wymagają dwóch zobrazowań – panchromatycznego (o wysokiej rozdzielczości) i wielospektralnego wykonanych w zbliżonym czasie.

Inną grupą rozwiązań są metody wykorzystujące wiele obrazów o niskiej rozdzielczości na potrzeby oszacowania obrazu o wysokiej rozdzielczości. Jedną z nich jest metoda próbkowania obrazu Très Haute Résolution (THR) stosowana do poprawy rozdzielczości przestrzennej scen panchromatycznych pozyskanych przez SPOT5. Jest to metoda pozyskiwania i przetwarzania obrazów wykorzystująca dwie macierze detektorów CCD, które są przesunięte w płaszczyźnie ogniskowej [24].

Kolejnym rozwiązaniem oszacowania wartości pikseli obrazu o wysokiej rozdzielczości na podstawie obrazów o niskiej rozdzielczości (LR) jest algorytm liniowej rekonstrukcji zmiennych pikseli (Super-Resolution Variable-Pixel Linear Reconstruction – SRVPLR). Opiera się on na algorytmie VPLR (nazywanym algorytmem Drizzle) [9], który służy do łączenia obrazów astronomicznych. W algorytmie tym piksele w obrazach o niskiej rozdzielczości są odwzorowane w nowym obrazie o wysokiej rozdzielczości, uwzględniając przesunięcia, obrót oraz zniekształcenie sensora obrazującego. Jednocześnie wielkość przenoszonego piksela jest zmniejszana, aby ograniczyć liczbę artefaktów na obrazie.

Oprócz klasycznych metod poprawy rozdzielczości (tj. interpolacja, pansharpening), istnieje grupa rozwiązań, której fundamentem są algorytmy głębokiego uczenia. Do poprawy rozdzielczości przestrzenniej za pomocą tych metod nie są potrzebne dodatkowe obrazy, jak w przypadku pansharpeningu. Receptura na oszacowanie obrazu SR pochodzi z wyuczonego modelu sieci neuronowej. Jakość modelu sieci neuronowej zależy od wielu czynników, a do najważniejszych można zaliczyć: architekturę sieci, metodę uczenia, sposób aktualizacji wag oraz zbiór danych treningowych, który symuluje środowisko w którym w przyszłości model będzie wykorzystywany. Najbardziej popularną metodą tworzenia baz danych przeznaczonych do poprawy rozdzielczości przestrzennej jest zmniejszenie rozdzielczości obrazu HR do rozdzielczości obrazu LR, a następnie wykonanie szkolenia w oparciu o obrazy LR, które są porównywane z obrazami prawdziwymi HR. W związku z tym, do szkolenia sieci może zostać wykorzystany każdy obraz cyfrowy bądź dowolna baza danych obrazów cyfrowych. Większość prowadzonych badań na temat poprawy rozdzielczości oparta jest na klasycznych obrazach cyfrowych (Set5-4x [25], Set14-4x [26], BSD100-4x [27], URBAN100-4x [28], FFHQ 256x256-4x [29], FFHQ 512x512-4x [29], FFHQ 1024x1024-4x [29]). Charakteryzują się one wysoką rozdzielczością, a co za tym idzie, dużą szczegółowością. W przypadku poprawy rozdzielczości obrazów pionowych pozyskanych z pułapu lotniczego, a szczególnie satelitarnego, te same obiekty charakteryzowane są przez inne cechy,

a rozdzielczość obrazu jest znacznie niższa. W związku z tym, aby wykorzystać zaproponowane przez naukowców rozwiązania poprawy rozdzielczości obrazów cyfrowych (pozyskanych głównie z pułapu ziemskiego), konieczne jest doszkolenie modeli, których wagi zostały oszacowane na podstawie klasycznych obrazów cyfrowych.

Podczas przeglądu literatury zwrócono uwagę na sposób oceny oszacowanych przez sieci neuronowe obrazów SR. Do najpopularniejszych wskaźników oceny jakościowej należy szczytowy stosunek sygnału do szumu (ang. peak signal-to-noise ratio – PSNR) [30] i podobieństwo strukturalne (ang. structural similarity index – SSIM) [30]. Pierwszy z nich zawiera informacje na temat maksymalnej mocy sygnału do mocy szumu zakłócającego ten sygnał, zaś druga uwzględnia zniekształcenia luminancji, kontrastu oraz struktury. Mniej popularnymi, lecz równie często wykorzystywanymi wskaźnikami oceny są: względny średni błąd spektralny (ang. Relative Average Spectral Error – RASE) [31], bezwymiarowy globalny błąd względny (ang. Relative Dimensionless Global Error in Synthesis – ERGAS) [32], wskaźnik korelacji (ang. Correlation Coefficient – CC) [33], porównanie krzywych spektralnych (ang. Spectral Angle Mapper – SAM) [34], rozbieżność informacji spektralnych (ang. Spectral Information Divergence - SID) [35], uniwersalny wskaźnik jakości obrazu (ang. Universal Image Quality Index - UIQI) [36]. Sposób ich działania został opisany w sekcji IIE publikacji **P1**.

Według bazy Scopus, od 2015 roku powstało ponad 2540 prac badawczych dotyczących algorytmów SISR bazujących na sieciach neuronowych. Liczba ta świadczy o dużej popularności tych rozwiązań, choć należy zwrócić uwagę, że większość z nich dotyczy wyostrzania obrazów klasycznych, jednak mogą one zostać wykorzystane w problemie poprawy rozdzielczości zobrazowań satelitarnych. Rozwiązania SISR wykorzystujące sieci splotowe można podzielić na dwie grupy: metody oparte na klasycznych warstwach splotowych oraz sieci GAN. Jednymi z najpopularniejszych rozwiązań należących do pierwszej grupy są SRCNN [37], FSRCNN [38], ESPCNN [39], EDSR [40]. Drugą grupą rozwiązań są metody wykorzystujące generatywne sieci przeciwstawne. Oryginalny model GAN składa się generatora oraz dyskryminatora. Zadaniem generatora jest przyjęcie na wejściu obrazu o niskiej rozdzielczości i dekodowaniu go w celu wygenerowania syntetycznego obrazu SR. Natomiast zadaniem dyskryminatora jest odróżnienie danych rzeczywistych od wygenerowanych przez generator. Sieć generatora trenowana jest tak, aby mogła oszukać dyskryminator. Sieci GAN mają ogromne możliwości rekonstrukcji obrazów. Niemniej jednak są one bardzo trudne do uczenia, z uwagi na swój niestabilny trening oraz tendencje do wystąpienia znikających lub eksplodujących

gradientów. W sekcji IID publikacji **P1** przedstawiono szczegółowy opis wybranych metod poprawy rozdzielczości przestrzennej wykorzystujących głębokie sieci neuronowe. Ponadto w tabeli VI wykonano porównanie jakości obrazów SR oszacowanych przez popularne metody wykorzystujące sieci neuronowe (np. SRResNet [41], EDSR [40], SRCNN [37], FSRCNN [38], SAN [42]) oraz rozwiązania oparte na GAN (np. SRGAN [41], ESRGAN [43]). Do porównania wykorzystano różne bazy danych testowych. W przypadku większości baz danych (Set14-4x [26], BSD100-4x [27], URBAN100-4x [28]) najlepszą rekonstrukcję obrazów LR wykonał model SAN [42], ale dla bazy danych Set5-4x najlepszy był model ESRGAN [43].

W ramach przeprowadzonych badań wykonano przegląd metod poprawy rozdzielczości przestrzennej obrazów cyfrowych za pomocą klasycznych metod wyostrzania tj. interpolacji i pansharpening oraz coraz bardziej popularnych rozwiązań wykorzystujących głębokie sieci neuronowe. Głównym ograniczeniem metod interpolacyjnych jest nieznaczny wzrost możliwości interpretacyjnych pomimo zmniejszenia rozmiaru piksela. Niemniej jednak rozwiązania te, nie wymagają dużej mocy obliczeniowej jednostek roboczych, a także należą do grupy metod SISR. Dysponując sceną panchromatyczną o wysokiej rozdzielczości i wielospektralną o niskiej rozdzielczości można połączyć atuty tych danych za pomocą metod pansharpeningu, jednak metoda ta nie pozwala na poprawę rozdzielczości zobrazowania, na którym znajdują się obiekty będące w ruchu. W tym przypadku, na obrazie SR widoczne są artefakty, będące konsekwencją przesunięcia między położeniem poruszającego się obiektu na wysokorozdzielczym obrazie panchromatycznym, a jego położeniem na obrazie wielospektralnym. Między innymi z tego powodu algorytmy te nie są wykorzystywane do poprawy rozdzielczości sekwencji obrazów. Dodatkowo, ze względu na metodykę działania tych rozwiązań, mogą one być zastosowane tylko do poprawy rozdzielczości obrazów wielospektralnych mających swój odpowiednik o wysokiej rozdzielczości.

W porównaniu z innymi metodami architektury oparte o GAN pozwalają uzyskać lepsze wyniki poprawy rozdzielczości w stosunku do metod klasycznych. Ze względu na wielki potencjał i szerokie zastosowanie GAN, naukowcy rozwijają ich zastosowanie w teledetekcji, jednak cały czas aktualne są problemy dotyczące zastosowania GAN do wyostrzenia całych scen satelitarnych, a nie tylko ich fragmentów. Innymi problemami w zastosowaniu sieci neuronowych (zwłaszcza GAN) jest niestabilność uczenia oraz problem znikającego gradientu. Ponadto na podstawie dokonanego przeglądu literatury można stwierdzić, że sieci GAN mają trudność z brakiem zbieżności oraz z doborem

hiperparametrów. Wiele metod np. SRCNN (zawartych w tabela VI publikacji P1) nie można zastosować do poprawy rozdzielczości przestrzennej zobrazowań satelitarnych ze względu na dość słabe odwzorowanie tekstury, co może prowadzić od powstania błędów podczas interpretacji zobrazowań. Najlepsze architektury zostały zaproponowane przez Wang [44] oraz Courtrai [45]. W metodach tych uwzględniono bloki rezydualne[11], co jest szczególnie istotne w przypadku zastosowania GAN, ponieważ bloki te stabilizują i znacznie przyspieszają trening modelu.

W ramach badań, których podsumowaniem jest publikacja P1, wykonano przegląd klasycznych oraz współczesnych metod poprawy rozdzielczości przestrzennej. Metody głębokiego uczenia wykazują ogromny potencjał jako SISR. Niektóre architektury bazujące na sieciach splotowych, pomimo swojego zaawansowania oraz rozbudowanej architektury, nie sprawdzają się do SISR zobrazowań satelitarnych, ponieważ pomimo poprawy rozdzielczości nie wzrastają możliwości interpretacyjne. Ponadto w publikacji P1 zwrócono uwagę na to, że wiele prac badawczych skupiało się do tej pory na poprawie rozdzielczości małych obrazów - niewiele jest rozwiązań dotyczących zastosowania SISR w teledetekcji, a szczególnie metod, które uwzględniają możliwość poprawy rozdzielczości całych scen satelitarnych o wymiarze znacznie większym niż wymiar obrazów prezentowanych w publikacjach. W publikacji **P1** przedstawiono przyszłe kierunki rozwoju w zakresie wyostrzania całych scen oraz optymalizacji parametrów i udoskonalenia architektur sieci neuronowych, a zwłaszcza GAN.

Przedstawiony przegląd literatury oraz wykonana analiza pozwoliły na uzyskanie odpowiedzi na wszystkie postawione na tym etapie pytania badawcze PB1.1 – PB1.5 oraz umożliwiły udział w osiągnięciu celu szczegółowego CS1, co bezpośrednio przyczyniło się do realizacji celu głównego CG.

---

[11] Blok rezydualny – *ang. residual block*, blok warstw sieci, w którym zastosowane zostało połączenie pomijające. W blokach tych, sygnał przekazywany jest do kolejnej warstwy boku oraz do warstwy znajdującej się nieco wyżej.

4.2. Metoda wykorzystania okna czasowego do poprawy rozdzielczości przestrzennej zobrazowań satelitarnych [publikacja P2]

Jak już wspomniano, większość naukowców skupiała się do tej pory na poprawie rozdzielczości małych obrazów. Nie poruszano zagadnień dotyczących zastosowania metod SISR w teledetekcji, a szczególnie w poprawie rozdzielczości całych scen satelitarnych, których rozmiar jest znacznie większy niż tych, które przykładowo prezentowano w publikacjach [46], [47], [48]. Ponadto analizując parametry charakterystyczne istniejących modeli SISR wykorzystywanych w widzeniu komputerowym, wymiar parametru wejścia do pierwszej warstwy jest znacznie mniejszy od wymiarów obrazów, które pozyskiwane są przez satelity. Co więcej, nawet niewielkie satelity wyposażone w matryce o niewielkich rozmiarach, np. CMOS [4] czy CMOSIS CMV [5], charakteryzujące się niską wydajnością kwantową pozyskują zobrazowania o rozmiarze znacznie większym od projektowanych wejść do sieci neuronowych.

Wiedząc, że pomimo tak dużych mocy obliczeniowych procesorów graficznych oraz możliwości wykorzystania wirtualnych maszyn, bardzo trudne jest przetworzenie tak dużych zobrazowań satelitarnych. W ramach badań starałam się znaleźć odpowiedź na poniższe pytania:

PB2.1. Czy możliwa jest poprawa rozdzielczości zobrazowań satelitarnych za pomocą algorytmów sztucznej inteligencji?

PB2.2. Czy istnieją metody łączenia obrazów po zastosowaniu algorytmów poprawy rozdzielczości za pomocą metod głębokiego uczenia?

PB2.3. Jaką należy przyjąć metodykę podczas łączenia obrazów SR oszacowanych przez generatywne sieci przeciwstawne?

PB2.4. Czy metoda ta może być również wykorzystywana do łączenia obrazów będących wynikiem działania algorytmów odpowiedzialnych za segmentację?

Jedną z możliwych odpowiedzi na pierwsze pytanie badawcze jest podzielenie procesu poprawy rozdzielczości przestrzennej zobrazowania satelitarnego na trzy etapy: (1) podział zobrazowania na mniejsze fragmenty (kafelki), o wymiarze równym parametrowi rozmiaru wejścia do sieci neuronowej (NN) (ang. input size), (2) poprawie rozdzielczości przestrzennej za pomocą NN, (3) ponownego połączenia zrekonstruowanych kafelków w jeden obraz. W ramach badań, których wynikiem jest publikacja **P2** zaproponowałam wykorzystanie okien czasowych w celu wykonania mozaikowania (łączenia) fragmentów zobrazowania. W proponowanej metodzie obrazy wynikowe łączone są ze

sobą uwzględniając wagi pochodzące z dwuwymiarowych okien czasowych. Zastosowanie okien czasowych polega na przygotowaniu odpowiedniej macierzy wag (symetrycznej względem środka obrazu), a następnie pomnożeniu jej przez obraz (SR), zrekonstruowany przez NN. Wykorzystanie tego rozwiązania umożliwiają właściwości okien czasowych: (1) niezerowa w skończonym przedziale czasu, (2) osiąga maksimum w środku przedziału, (3) jest symetryczna względem środka przedziału. Do analizy wybranych okien czasowych sprawdzono cztery parametry: wartość minimalną, maksymalną, średnią sumę wag oraz sumę wag dla obszaru wspólnego obrazów. Badania nad możliwością zastosowania okien czasowych na potrzeby łączenia obrazów przeprowadzono dla obrazów o wymiarze 384 × 384 pikseli. Wymiar ten nie został wybrany przypadkowo. Sieci SRGAN i ESRGAN (popularne modele SISR) jako wejście najczęściej pobierają obraz o niskiej rozdzielczości (LR) o wymiarze 96 × 96 pikseli, a zwracają obraz SR o wymiarach 384 × 384. Prace nad badaną tematyką zostały podzielone na dwa etapy. W pierwszym etapie zostały przeprowadzone badania wstępne, na podstawie których wybrano jedynie te okna czasowe, dla których: (1) w1: suma wag w punkcie należy do przedziału [0.95, 1.05], (2) w2: suma wag dla obszaru pokrycia należy do przedziału [190, 192], przy założeniu, że pokrycie łączonych obrazów (obrazów składowych) wynosi 50%. W badaniach sprawdzono możliwość wykorzystania okien czasowych tj. Welch, Sine, Hann, Bartlett-Hann, Triangular, Hann-Poisson, Gaussian, Lanchos, Blackmana, Blackman-Nuttall, Blackman-Harris, Flat top, Poisson, Hamming (w załączniku A publikacji **P2** znajdują się funkcje wykorzystanych okien czasowych).
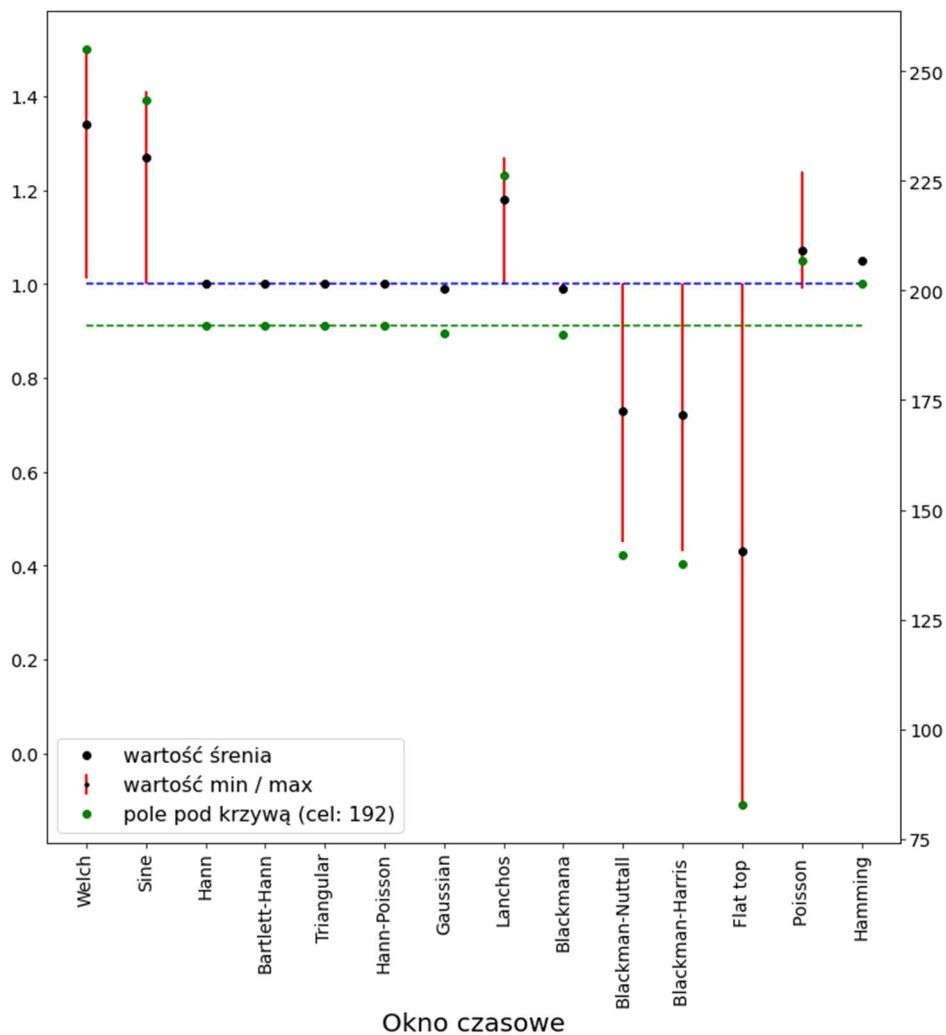
Rys. 4.1. *Zestawienie wartości średnich, minimalnych i maksymalnych wag badanych okien czasowych. Założono pokrycie obrazów równe 50%. Wartości liczbowe wyznaczono dla obszaru pokrycia obrazu. W celu lepszej interpretacji wyników, niebieską przerywaną linią zaznaczona została prosta pomocnicza y1 = 1, zaś zieloną przerywaną linią prosta pomocnicza y2 = 192 (opracowano na podstawie [49])*

Na rysunku 4.1 przedstawiono wizualizację otrzymanych wyników. Przedstawiony wykres jednoznacznie pokazuje, że zdefiniowane warunki w1 i w2 spełniają następujące okna czasowe: Hann, Bartlett-Hann, Triangular, Hann Poisson oraz Backmana. W celu weryfikacji otrzymanych wyników przeprowadzono ocenę jakościową obrazów. Do analizy wykorzystane zostały najczęściej wykorzystywane w środowisku teledetekcji i widzenia komputerowego wskaźniki oceny jakościowej. Do przeprowadzenia analizy wykorzystano jeden z obrazów pochodzących z sekwencji obrazów pozyskanych za pomocą minisatelity Jinlin-1. Przygotowany obraz podzieliłam na obrazy składowe o wymiarach 384 × 384 pikseli z pokryciem między nimi równym 192 piksele (50%). Następnie, przygotowane obrazy połączono ze sobą z wagami obliczonymi za pomocą okien

czasowych: Hann, Bartlett-Hann, Triangular, Hann-Poisson, Gaussian, Lanchos, Black-man. W analizie pominęłam okna czasowe, dla których średnia suma wag w punkcie nie należy do przedziału [0.9,1.1].

Tabela 4.1.

*Ocena jakości łączenia obrazów za pomocą okien czasowych. Dla metod oznaczonych „ *", przed oceną zastosowano dopasowanie histogramów (opracowano na podstawie [49])*

wskaźnik funkcja okna	MSE [30]	RMSE [30]	PSNR [30]	UQI [36]	SCC [50]	SAM [34]	SSIM [30]	RASE [31]	VIFP [51]	NRMSE [30]
Hann a_0=0.5	0.42	0.64	51.89	1.00	1.00	0.01	1.00	0.35	1.00	0.01
Bartlett-Hann	3.66	1.91	42.49	1.00	0.91	0.01	1.00	101.21	0.98	0.02
Triangular	3.84	1.95	42.29	1.00	0.90	0.01	1.00	104.21	0.98	0.02
Hann-Poisson	3.42	1.85	42.78	0.99	0.92	0.01	1.00	96.91	0.98	0.02
Gaussian	92.79	9.63	28.46	0.99	0.89	0.08	0.99	401.44	0.89	0.08
Gaussian*	75.74	8.70	29.34	1.00	0.88	0.07	0.99	354.06	0.87	0.07
Lanchos	1288.29	35.89	17.03	0.90	0.85	0.30	0.88	1658.52	0.72	0.32
Lanchos*	863.22	29.38	18.77	0.95	0.81	0.24	0.88	1294.12	0.59	0.24
Blackmana	16.68	4.08	35.91	1.00	0.90	0.02	1.00	207.04	0.97	0.03
Blackmana*	3.23	1.80	43.03	1.00	0.89	0.01	1.00	87.88	0.97	0.01

W tabeli 4.1. przedstawiono ocenę jakości łączenia obrazów za pomocą okien czasowych. Analizując otrzymane wyniki wyraźnie widać, że okna czasowe, których suma wag dla pojedynczego piksela jest równa 1, prezentują najlepsze wyniki metryk oceny. Jednakże należy zwrócić uwagę, że zastosowanie okien czasowych nieznacznie pogarsza jakość obrazu, o czym świadczy niewielka wartość błędu MSE oraz RASE. Jednocześnie szczytowy stosunek sygnału do szumu (PSNR) przyjmuje wysokie wartości np. dla okna Hanna PSNR = 51.89 dB, co dowodzi o bardzo wysokim podobieństwie między obrazem wynikowym, a referencyjnym.

Do przeprowadzenia badań zasadniczych wybrane zostały jedynie cztery okna czasowe (Hann, Hann-Poisson, Bartlett-Hann, Triangular), które spełniały zdefiniowane na początku badań warunki w1 i w2. Okna te posłużyły do wyznaczenia stopnia pokrycia pomiędzy kafelkami, dzięki któremu zrekonstruowane zobrazowania satelitarne będą charakteryzowały się najlepszą jakością.

Na potrzeby przeprowadzenia badań przygotowana została sieć ESRGAN, której zadaniem była poprawa rozdzielczości kanałów 2, 3 i 5 zobrazowań wielospektralnych pozyskanych za pomocą satelity World View 2 (WV2). Do treningu modelu wykorzysta-łam własną bazę danych składającą się z 29 500 obrazów o niskiej rozdzielczości (LR)

(o wymiarach 96 × 96 pikseli) oraz ich odpowiedników o wysokiej rozdzielczości (HR) (o wymiarach 384 × 384 pikseli). Do przygotowania bazy danych obrazów o niskiej rozdzielczości wykorzystałam zobrazowania wielospektralne pozyskane przez WV2 przedstawiające obszar Radomia, Górnośląsko-Zagłębiowskiej Metropolii oraz Pomorza. W celu przygotowania zbioru obrazów HR połączyłam wysoką rozdzielczości przestrzenną obrazu panchromatycznego z wysoką rozdzielczością spektralną obrazu wielospektralnego wykonując pansharpenig metodą Grama-Schmidta. Głównym powodem wyboru tej metody są najmniejsze (w porównaniu z innymi metodami pansharpeningu) zniekształcenia barw.

Jako parametry początkowe szkolenia sieci ESRGAN zainicjowano parametry zalecane przez autorów rozwiązania [43]. Autorzy rozwiązania zalecają, aby zmniejszać szybkość uczenia (1×10^{-4}) o połowę przy 50 tys., 100 tys., 200 tys. i 300 tys. iteracjach. Niestety w przypadku tej bazy danych po wykonaniu ok. 35 000 iteracji można zauważyć zjawisko znikających gradientów, które bardzo łatwo zidentyfikować poprzez nagły wzrost straty generatora L_G (rys. 4.2).
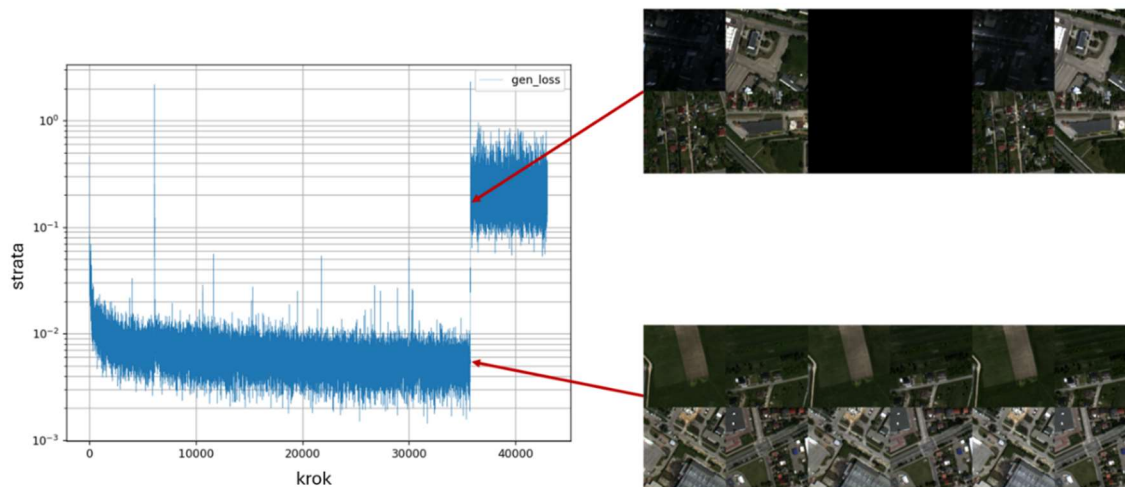
Rys. 4.2. *Zjawisko znikających gradientów przedstawione na wykresie L_G (opracowano na podstawie [49])*

W związku z powyższym zjawiskiem, konieczne było szybsze zmniejszenie parametru szybkości uczenia. W badaniach tych zastosowałam zmiany tego parametru według tabeli 3 znajdującej się w publikacji **P2**. Wartości te zostały dobrane doświadczalnie. Operacja ta zapobiega przetrenowaniu sieci, ale jednocześnie znacząco wydłuża proces szkolenia.

Na etapie badań zasadniczych nad możliwością połączenia obrazów SR za pomocą okien czasowych skupiono się również na określeniu najlepszego stopnia pokrycia między łączonymi obrazami. W tym celu wybrano pięć fragmentów scen satelitarnych o wymiarze nie mniejszym niż 900 × 900 pikseli (LR) (wybrane fragmenty nie brały udziału w treningu modelu ESRGAN). Obrazy te przedstawiają obszary miejskie, obrzeża miast, lasy, tereny rolne oraz fragment farmy wiatrowej. Do oceny, jako obrazy referencyjne o wysokiej rozdzielczości (HR) wykorzystano te same obrazy, których rozdzielczość przestrzenną poprawiono za pomocą pansharpeningu - metodą Grama-Schmidta.

Na potrzeby badań, obrazy LR zostały podzielone na mniejsze obrazy o rozdzielczości 96 × 96 pikseli. Przygotowano dziesięć zestawów obrazów składowych (dla każdego z fragmentów zobrazowań), do których przygotowania zastosowano różne stopnie pokrycia (od 50 do 5% z krokiem 5%). Przed przystąpieniem do badań, poprawiono rozdzielczość przestrzenną każdego obrazu częściowego za pomocą wytrenowanej sieci ESRGAN. Tak przygotowane obrazy połączono ze sobą za pomocą (wybranych w ramach badań wstępnych) okien czasowych. Schemat zastosowanego postępowania przedstawiono na rysunku 4.3.
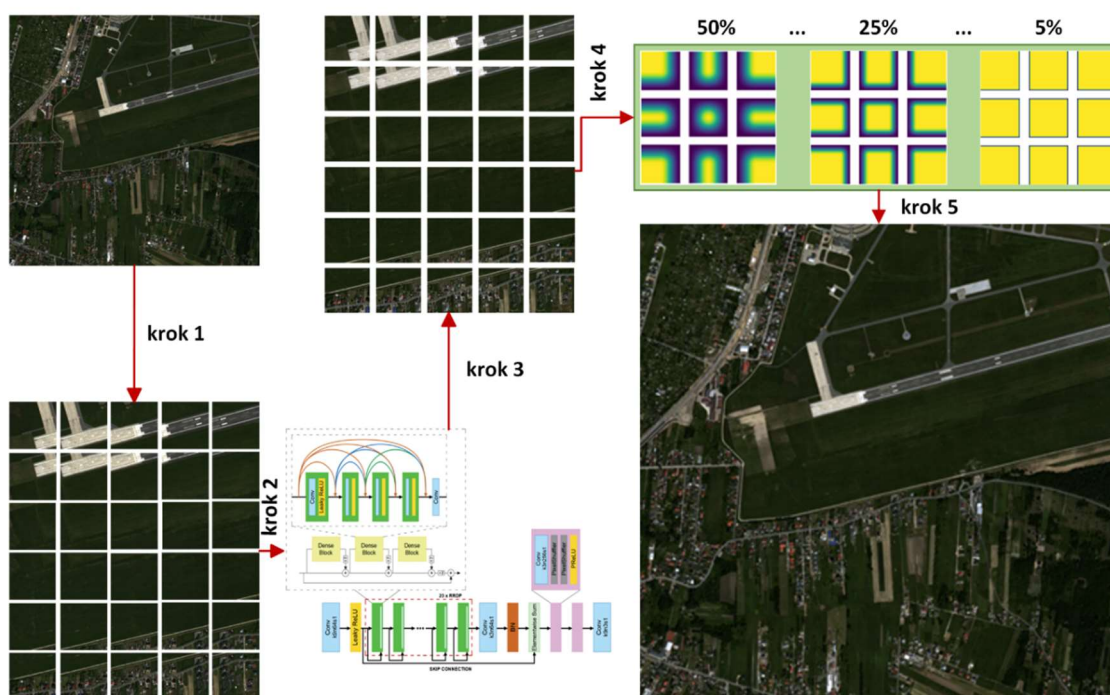
Rys. 4.3. *Schemat metodyki poprawy rozdzielczości całych scen satelitarnych (opracowano na podstawie [49])*

W wyniku wyżej opisanej operacji powstało 200 nowych fragmentów zobrazowań (5 fragmentów zobrazowań × 10 stopni pokrycia × 4 okna czasowe = 200 obrazów) o poprawionej za pomocą sieci ESRGAN rozdzielczości przestrzennej. Każdy z oszacowanych obrazów poddany został ocenie za pomocą wskaźników przedstawionych i opisanych w punkcie 3.2. publikacji **P2**. W celu lepszej wizualizacji wyników obliczono średnią wartość badanych wskaźników oceny dla każdego obrazu, z uwzględnieniem podziału na metodę łączenia obrazów. Otrzymane wyniki przedstawiono w załączniku B publikacji **P2**. Na ich podstawie można zauważyć, że jakość obrazu wynikowego rośnie wraz z zmniejszeniem się stopnia pokrycia między obrazami, a wartość maksymalną osiąga dla pokrycia wynoszącego ok. 10%. Dla wartości pokrycia mniejszej niż ta wartość jakość połączonego obrazu zaczyna spadać. Dodatkowo obrazy połączone za pomocą okna czasowego Triangular charakteryzują się najlepszą jakością dla pokrycia między obrazami większego od ok. 25%. W przypadku gdy pokrycie między obrazami jest mniejsze od tej wartości, najlepsze wyniki oceny mają obrazy połączone za pomocą okna Hann-Poisson (ale nie jest większe niż 15%).

W celu dodatkowej weryfikacji przedstawionej metody przetwarzania całych scen zobrazowań satelitarnych wykonałam badania, w których wykorzystałam okna czasowe do łączenia obrazów będących wynikiem operacji automatycznej interpretacji obrazu np. w wyniku zastosowania sieci U-Net, czy warunkowych generatywnych sieci przeciwstawnych (ang. conditional generative adversarial network – CGAN).
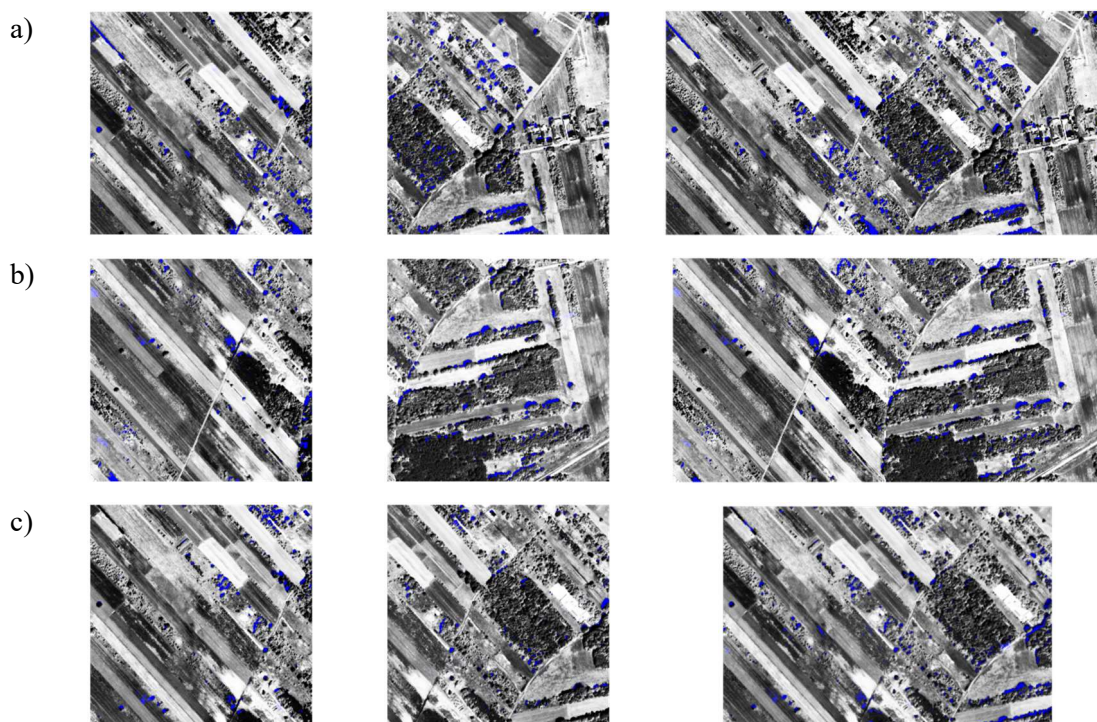
Rys. 4.4. *Przykład wykorzystania okien czasowych do łączenia masek cieni, które zostały wykryte za pomocą sieci U-Net, przedstawiono obrazy z pokryciem: a) 5%, b) 10%, c) 50% [49]*

Na rysunku 4.4 przedstawiono dwa obrazy, których porycie wynosi odpowiednio 5%, 10%, 50% oraz obraz wynikowy połączenia tych obrazów za pomocą okna Hann-Poisson. W przypadku zastosowania pokrycia między obrazami równego 5% i 10% przeprowadzona analiza wizualna nie zidentyfikowała błędów wynikających z łączenia obrazów. Błędy występujące na powyższym rysunku wynikają z błędów modelu odpowiadającego za segmentację.

W ramach badań przedstawiłam i zweryfikowałam metodę przetwarzania zobrazowań satelitarnych o dowolnych wymiarach za pomocą algorytmów głębokiego uczenia. Zaproponowane rozwiązanie umożliwia przetworzenie zobrazowania satelitarnego za pomocą modelu sieci neuronowej o dowolnym rozmiarze wejścia. Zastosowanie pokrycia między obrazami wynoszącego ok. 10% i połączenie kafelków za pomocą okna Triangular lub Bartlett-Hann pozwoliło na rekonstrukcję zobrazowań satelitarnych z najwyższą jakością. Jednocześnie biorąc pod uwagę wyniki przedstawione w załączniku B publikacji **P2**, według których różnice jakości obrazów połączonych za pomocą okien czasowych Hann, Hann-Poisson, Bartlett-Hann i Triangular są niewielkie, każde z wymienionych okien czasowych może być wykorzystane do wykonania mozaikowania.

Przeprowadzone badania pokazują, że obrazy oszacowane przez wytrenowany model ESRGAN zawierają wiele błędów, co szczególnie widoczne jest podczas analizy

wizualnej np. na dachach domów. Dodatkowo potwierdzają to wartości metryk prezentowe w tabeli 2 publikacji **P2** oraz na przykładach obrazów SR znajdujących się w załączniku na rysunku 4 w tej samej publikacji.

Mając na uwadze powyższe wyniki można wywnioskować że, możliwa jest poprawa rozdzielczości przestrzennej całych scen satelitarnych. Jednakże konieczna jest modyfikacja istniejących modeli głębokiego uczenia, bądź opracowanie nowych rozwiązań. Przeprowadzone analizy pozwoliły mi na uzyskanie odpowiedzi do wszystkich postawionych w tym etapie pytań badawczych PB2.1 – PB2.4 oraz umożliwiły udział w osiągnięciu celów szczegółowych CS2 i CS6, co bezpośrednio przyczyniło się do realizacji celów głównych CG.

4.3. Projekt modelu generatywnego do poprawy rozdzielczości przestrzennej zobrazowań satelitarnych [publikacja P3]

Jak przedstawiono w publikacji **P1** i **P2**, problem poprawy rozdzielczości zobrazowań satelitarnych, które nie spełniają wymagań do zastosowania pansharpeningu jest wciąż aktualny. Jedocześnie obrazy oszacowane przez popularne metody SISR wykorzystujące głębokie sieci neuronowe posiadają wiele artefaktów, nawet w przypadku krawędzi obiektów (załącznik 2 publikacji **P2**). W związku z powyższym, przeprowadziłam badania mające na celu wykonanie modyfikacji architektury ESRGAN [43]. Zadaniem wprowadzonych zmian była poprawa jakości generowanych obrazów SR oraz przyspieszenie procesu treningu generatora. W trakcie prowadzonych badań starałam się znaleźć odpowiedź na następujące pytania badawcze:

PB3.1. Czy zastosowanie wielogałęziowego dyskryminatora pozwala na lepszą ocenę obrazów SR generowanych przez generator?

PB3.2. Czy wykorzystanie starty Wassersteina do treningu sieci GAN umożliwia lepsze rozróżnienie obrazów HR i SR przez dyskryminator?

PB3.3. Jakie można podjąć działania, w celu ustabilizowania treningu sieci GAN (ograniczenia występowania zjawiska znikających (lub eksplodujących) gradientów)?

PB3.4. Jaką przyjąć metodykę do oceny obrazów SR? Jak przeprowadzić ocenę potencjału interpretacyjnego obrazów SR?

W związku z tym, że prace badawcze skupiają się nad opracowaniem nowej architektury sieci GAN przygotowane zostały trzy bazy danych, na podstawie których

wykonany został trening i weryfikacja projektowanego modelu sieci GAN. Bazy te składają się z tożsamych par obrazów o niskiej rozdzielczości (o wymiarach 96 × 96 pikseli) i wysokiej rozdzielczości (o wymiarach 384 × 384 pikseli).

Pierwsza z przygotowanych baz danych składa się z zdjęć lotniczych. Baza danych o wysokiej rozdzielczości powstała w wyniku podziału ortofotomapy o wielkości piksela wynoszącej 0.25m. Obrazy o niskiej rozdzielczości przygotowano poprzez przepróbkowanie obrazów HR do rozmiaru 96 × 96 pikseli. Do przygotowania drugiej bazy danych wykorzystano bazę danych Dataset for Object deTection in Aerial Images (DOTA) [52]. Do przygotowania zbioru obrazów HR wybrano 18 000 losowych fragmentów obrazów o rozmiarze 384 × 384 pikseli. Obrazy LR, podobnie jak w przypadku pierwszej bazy danych, przygotowano poprzez przepróbkowanie obrazów HR. Jako trzecią bazę danych wykorzystano bazę danych utworzoną na potrzeby badań nad metodą wykorzystania okna czasowego do poprawy rozdzielczości przestrzennej zobrazowań satelitarnych.

Wprowadzone modyfikacje modelu ESRGAN obejmują głównie model dyskryminatora, czyli część modelu odpowiadającą za ocenę obrazów. Pierwszą z zastosowanych zmian jest wprowadzenie dyskryminatora wielokolumnowego. Model ten składa się z dwóch gałęzi, gdzie każdą z nich jest klasyfikatorem wykorzystującym warstwy splotowe. Pierwszy z klasyfikatorów posiada układ warstw taki sam, jak dyskryminator sieci SRGAN [41]. W przypadku drugiego krytyka dodano jedną warstwę splotową, wykorzystano metodę aktywacji na ReLU oraz zmieniono rozmiar filtrów w warstwie konwolucyjnej (ang. kernel size) (w warstwach parzystych parametr ten wynosi 5 pikseli) (rys. 4.5).
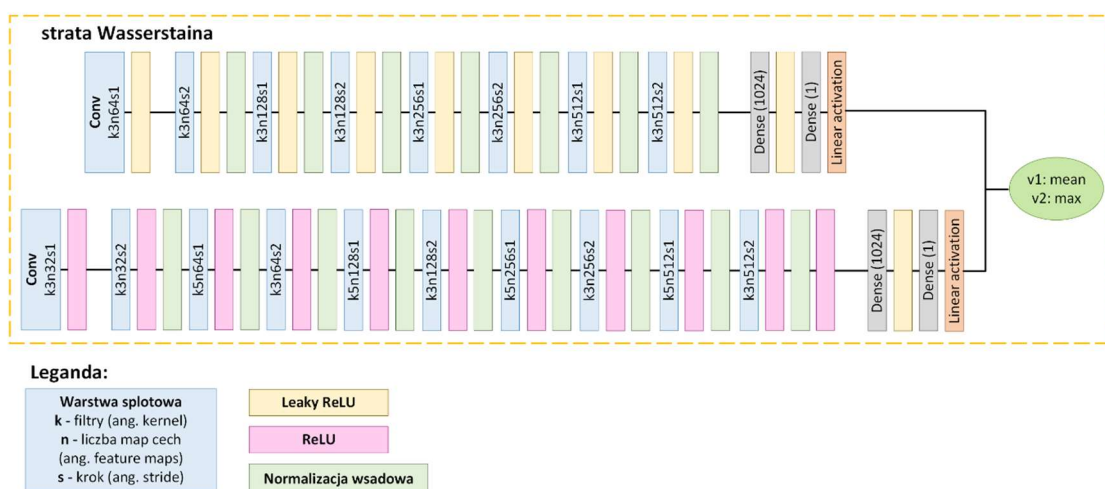
Rys. 4.5. *Model dyskryminatora (opracowano na podstawie* [53]*)*

Zastosowanie takiego dyskryminatora wielokolumnowego pozwoliło na lepszą ocenę jakości pracy generatora, a w konsekwencji, przyspieszyło trening generatora odpowiedzialnego za poprawę rozdzielczości przestrzennej zobrazowań satelitarnych. Kolejną wprowadzoną modyfikacją była zmiana metody oceny modeli generatora i dyskryminatora. W oryginalnym rozwiązaniu ESRGAN uczenie modelu odbywa się według strategii „gry o sumie zero", generator szkolony jest tak, aby oszukać dyskryminator (wz. 4.1):

$$\min_G \max_D \mathbb{E}_{x \sim P_r}\big[\log\big(D(x)\big)\big] + \mathbb{E}_{\tilde{x} \sim P_g}\big[\log\big(1 - D(\tilde{x})\big)\big] \tag{4.1}$$

gdzie:

P_r – rozkład docelowy;

P_g – rozkład bazowy.

Natomiast straty generatora (wz. 4.2) i dyskryminatora (wz. 4.3) modelu ESRGAN obliczane są za pomocą następujących wzorów:

$$L_D^{Ra} = -E_{I^{HR}}\big[\log\big(D_{Ra}(I^{HR}, I^{SR})\big)\big] - E_{I^{SR}}\big[\log\big(1 - D_{Ra}(I^{SR}, I^{HR})\big)\big] \tag{4.2}$$

$$L_G^{Ra} = -E_{I^{HR}}\big[\log\big(1 - D_{Ra}(I^{HR}, I^{SR})\big)\big] - E_{I^{SR}}\big[\log\big(D_{Ra}(I^{SR}, I^{HR})\big)\big] \tag{4.3}$$

gdzie:

I^{HR} – obraz HR;

I^{SR} – obraz SR;

D_{Ra} – dyskryminator relatywistyczny (średni).

W projektowanym modelu GAN podejście to zostało zastąpione stratą Wassersteina [54]. Zastosowanie tej funkcji straty pozwala na oszacowanie rozkładu danych obserwacyjnych wykorzystanych do uczenia. Zadaniem procesu uczenia jest minimalizacja odległości między rozkładem danych obserwowanym w zbiorze treningowym (rozkładem bazowym), a rozkładem danych oszacowanych przez generator (rozkładem docelowym). W celu określenia odległości między rozkładami prawdopodobieństwa w przestrzeni metrycznej wykorzystywana jest odległość Wassersteina. Odległość Wassersteina dla dwóch miar probabilistycznych μ i υ obliczana jest według poniższego wzoru (wz. 4.4).

$$W(\mu,\vartheta) = \max_{\alpha,\beta} \left[\mathbb{E}_x[\alpha(X)] + \mathbb{E}_{x'}[\beta(X')] - \gamma \sum_{x,x'\in\{0,1\}}^{d} \exp\left(\frac{1}{\gamma}\big(\alpha(x) + \beta(x') - D(x,x')\big) - 1\right) \right] \qquad (4.4)$$

gdzie:

D: $X \times X \rightarrow R_+$, $X = \{0,1\}^d$;

W - odległość Wassersteina;

α , β – funkcje zbioru X;

γ - rozkład prawdopodobieństwa.

Wykorzystanie odległości Wessersteina umożliwia określenie minimalnej ilości pracy, jaka jest potrzebna do przekształcenia rozkładu bazowego w rozkład docelowy. W projektowanej sieci GAN do wyznaczenia odległości Wessersteina wykorzystana została metryka Kantorovicha-Rubinsteina [55]. Metryka ta wykorzystuje warunek Lipschitza (szybkość zmiany funkcji, dla której obywa się optymalizacja jest ograniczona przez stałą Lipschitza). Pozwala on na utrzymanie równowagi między wielkością na wejściu do modelu, a wielkością zmian na jego wyjściu. W wyniku wprowadzonych modyfikacji, uproszczony jest proces optymalizacji generatora, co ma korzystny wpływ na jakość generowanych obrazów SR. Wprowadzone zmiany dotyczące metody oceny modelu, zmieniają strategię uczenia sieci (wz. 4.1). W konsekwencji przyjmuje ona postać (wz. 4.5):

$$\min_{G} \max_{D\in\mathcal{D}} \mathbb{E}_{x\sim P_r}[D(x)] - \mathbb{E}_{\tilde{x}\sim P_g}[D(\tilde{x})] \qquad (4.5)$$

gdzie:

$\mathcal{D}$ – zbiór funkcji 1-Lipschitz;

P_r – rozkład docelowy;

P_g – rozkład bazowy.

W związku z wprowadzonymi zmianami, dyskryminator zwraca dowolną liczbę z przedziału $(-\infty, +\infty)$. Z tego powodu nazywany on jest krytykiem. Zastosowanie straty Wessersteina dodatkowo stabilizuje proces uczenia modelu, na wypadek zmian hiperparametrów uczenia [54]. Jednakże wadą implementacji warunku Lipschitza jest skłonność modelu do występowania zjawiska znikających (lub eksplodujących) gradientów. Zjawisko to jest wynikiem tego, że krytyk może zwracać dowolną liczbę z przedziału $(-\infty, +\infty)$, czyli może osiągnąć bardzo duże wartości, które nie są pożądane w sieciach

neuronowych. W związku z tym konieczne jest wprowadzenie dodatkowych ograniczeń dla krytyka. W projektowanej sieci GAN uczenie modelu przeprowadzone zostało zgodnie z algorytmem opublikowanym przez zespół Ishaan Gulrajani [56]. Autorzy rozwiązania proponują zastosowanie ograniczeń dla funkcji Lipschitza. Zakładają, że 1-Lipschtiz jest różniczkowana, wtedy i tylko wtedy gdy norma gradientu nie jest większa od 1. W związku z tym, wprowadzili ograniczenie normy gradientu stosując ograniczenie z karą dla losowych próbek należących do rozkładu bazowego ($\hat{x} \sim P_{\hat{x}}$). Ostatecznie, strata projektowanego modelu GAN przyjęła postać (wz. 4.6):

$$L = \mathbb{E}_{\tilde{x} \sim P_g}[D(\tilde{x})] - \mathbb{E}_{x \sim P_r}[D(x)] + \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \qquad (4.6)$$

W wyniku wyżej opisanych modyfikacji opracowałam nowy model GAN - Multicolumn Wasserstain ESRGAN (MCWESRGAN) - odpowiedzialny za poprawę rozdzielczości przestrzennej zobrazowań satelitarnych. W celu określenia wpływu wprowadzonych zmian na jakość oszacowanych obrazów SR oraz czas potrzebny na uczenie modelu generatora, opracowanych i wytrenowanych zostało pięć modeli GAN, dla każdej z przygotowanych zbiorów danych:

- M1 – sieć ESRGAN,
- M2 – generator ESRGAN, dyskryminator multikolumnowy (do optymalizacji wykorzystana została większa z wartości strat obliczonych dla gałęzi dyskryminatorów),
- M3 – M2 (ale do optymalizacji wykorzystana została średnia wartość strat obliczonych dla gałęzi dyskryminatorów),
- M4 – M2 (ale do optymalizacji wykorzystana została średnia wartość strat Wassersteina),
- M5 – model M4 uwzględniający zjawisko znikających gradientów (model MCWESRGAN).

W tabeli 4.2 przedstawiono wyniki wskaźników oceny jakościowej przeprowadzonej dla każdej z przygotowanych baz danych. Ocenie poddano zbiór danych testowych, który nie brał udziału w treningu modeli. Analizując otrzymane wyniki widoczny jest znaczny spadek czasu potrzebnego na uczenie poszczególnych modeli SISR. Zaimplementowanie wielokolumnowego dyskryminatora pozwala na ok. 7-krotne skrócenie czasu treningu modelu, a wykorzystanie straty Wassersteina skraca ten czas ponad 20

razy. W przypadku modelu M4, w którym wykorzystana została strata Wassersteina, dość szybko (już po 4 godzinach dla bazy danych WorldView-2) pojawiło się zjawisko znikających gradientów. W tym przypadku, nawet zmiana szybkości uczenia nie ograniczyła tego zjawiska. Jednakże zastosowanie ograniczenia Lipschitza nie tylko umożliwiło wydłużenie procesu uczenia, ale także znacznie poprawiło jakość oszacowanych obrazów SR zwłaszcza w przypadku fragmentów zobrazowań satelitarnych (pomiędzy M4, a M5 średnia wartość SSIM wzrosła o 0.07).

Tabela 4.2.
Średnie wartości metryk oceny jakościowej danych testowych dla bazy WorldView-2, zdjęć lotniczych i DOTA (* - czas treningu z wykorzystaniem karty graficznej Nvidia TITAN RTX 24 GB) (opracowano na podstawie [53]*)*

model / wskaźnik	M1	M2	M3	M4	M5 **MCWESRGAN**
Czas uczenia [godz.]*	96/101/83	12/14/14	12/14/14	4/6/9	6/8/11
SSIM	0.83/0.15/0.60	0.82/0.54/0.68	0.84/0.56/0.69	0.85/0.61/0.69	0.92/0.62/0.70
PSNR [dB]	28.74/ 11.52/ 22.76	29.68/ 21.74/ 24.60	28.48/ 22.23/ 24.70	28.97/ 28.43/ 24.78	32.10/ 22.45/ 25.05
SAM	0.14/0.19/0.19	0.12/0.19/0.15	0.13/0.18/0.15	0.12/0.17/0.15	0.09/0.17/0.14
SCC	0.18/0.06/0.13	0.26/0.07/0.29	0.24/0.11/0.29	0.26/0.13/0.30	0.37/0.13/0.32
UQI	0.95/0.70/98	0.96/0.93/0.98	0.95/0.94/0.98	0.96/0.97/0.99	0.98/0.97/0.99

* - czas treningu z wykorzystaniem karty graficznej Nvidia TITAN RTX 24 GB

W przypadku obrazów pozyskanych z pułapu satelitarnego, na których mogą występować obszary o różnej charakterystyce (obszary zurbanizowane, leśne, łąki, pola) ich oszacowanie na obrazach SR może różnić się pod względem jakości. Wobec tego, wyżej przedstawiona analiza przedstawia wartości uśrednione dla całego obrazu, a wartości te mogą być zawyżone przez bardzo wysoką jakość oszacowania powierzchni jednorodnych np. zbiorników wodnych, łąk, terenów pokrytych trawą. W związku z tym, na potrzeby lepszej oceny obrazów SR wykonałam lokalną ocenę jakościową obrazów. W metodzie tej obraz referencyjny (HR) i oszacowany (SR) dzielony jest na mniejsze obszary o wymiarach określonych przez parametr p. Następnie, dla wyznaczonych obszarów wykonywana jest ocena jakościowa (rys. 4.6). Dzięki tej metodzie możliwe jest określenie obszarów, które zostały „gorzej" oszacowane przez algorytm poprawy rozdzielczości przestrzennej. Na podstawie tych informacji, można określić jakie należy wykonać

modyfikacje (np. jak zmodyfikować bazę danych treningowych), aby zwiększyć jakość oszacowanych obrazów SR.
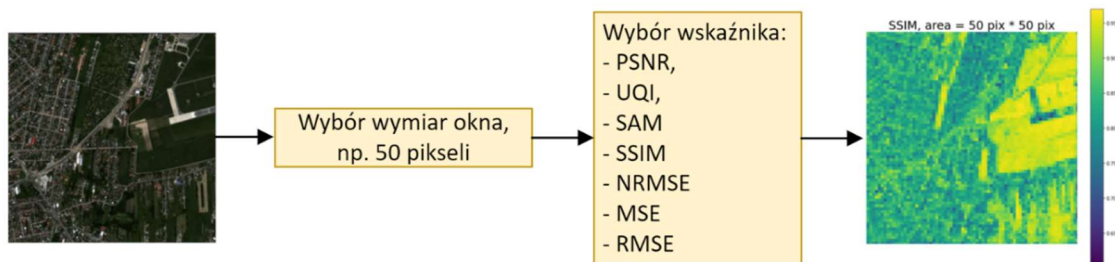
Rys. 4.6. *Schemat działania algorytmu oceny lokalnej obrazów*
(opracowano na podstawie [53])

Przedstawione powyżej metody oceny globalnej i lokalnej pozwalają określić jakość oszacowania obrazów SR. Jednakże nie umożliwiają one oceny wzrostu potencjału interpretacyjnego między obrazami LR i SR. Dlatego konieczne jest zastosowanie dodatkowej metody, która pozwoli określić, czy zwiększenie liczby pikseli przekłada się na wzrost zdolności rozróżniania obiektów na obrazie SR. W tym celu przeprowadziłam analizę widma gęstości mocy obrazu w dziedzinie częstotliwości (PSD), która, mimo że nie jest powszechnie stosowana w teledetekcji, dostarcza cennych informacji na temat ocenianego obrazu. Umożliwia ona określenie nie tylko zmiany rozdzielczości przestrzennej (GSD), ale również terenowej zdolności rozdzielczej (GRD) (rys. 4.7), gdzie parametr GRD określa najmniejszą wielkość, jaka może zostać rozróżniona na zobrazowaniu. Wartość tego parametru zależy od kontrastu miedzy obiektem, a tłem. Dla obiektów charakteryzujących się wysokim kontrastem z tłem GRD przyjmuje dwukrotną wartość parametru GSD, natomiast w przypadku wystąpienia niskiego kontrastu $GRD = 2\sqrt{2}GSD$.
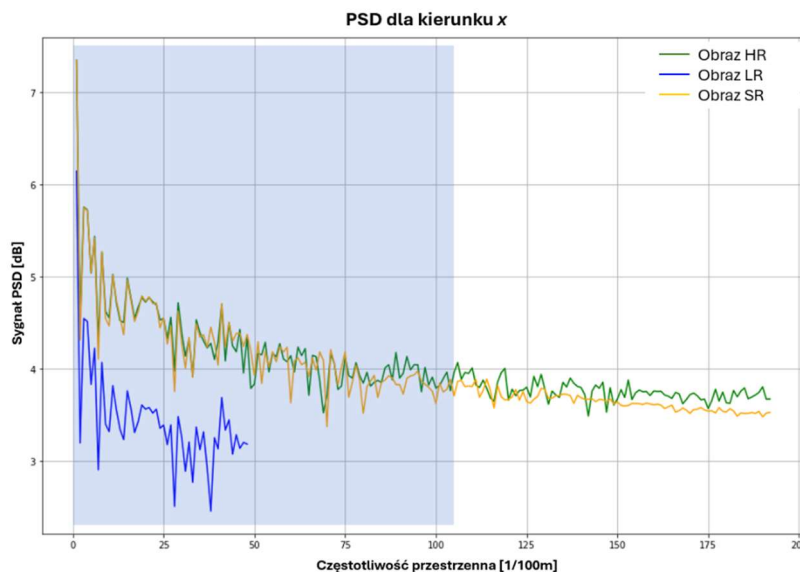
Rys. 4.7. *Przykład wykresu widma gęstości mocy, do częstości przestrzennej dla obrazu HR o GSD = 0.5m, na kierunku x (opracowano na podstawie* [53]*)*

Przedstawione wyniki pokazują, że obraz SR (żółta krzywa na wykresie), posiada taką samą wartość parametru GSD, jak obraz HR (rozdzielczość przestrzenna wzrosła czterokrotnie, w porównaniu do obrazu LR). Dodatkowo, analizując stosunek widma gęstości mocy do częstości przestrzennej obrazu HR i SR, jednoznacznie widać, że poprawa rozdzielczości obrazu miała miejsce, tylko w obszarze zaznaczonym niebieskim prostokątem, czyli do gęstości przestrzennej równej ok. 105 [1/100m] (rys. 4.7). W przedziale tym, ma miejsce najlepsze odwzorowanie widmowej gęstości mocy. A co za tym idzie, wartość parametru GRD wynosi ok. 1.8-2.1 GSD.

Analizując otrzymane wyniki (głównie na podstawie oceny lokalnej obrazów), można stwierdzić, że obiekty pochodzenia antropogenicznego są lepiej odwzorowywane na obrazach SR niż obiekty pochodzenia naturalnego (np. korony drzew). Wyjątkiem są polany i łąki, których jakość oszacowania rośnie, wraz ze spadkiem rozdzielczości przestrzennej obrazów LR.

Połączenie powyżej przedstawionych metod oceny obrazu pozwala na jednoznaczne określenie jakości oszacowanych obrazów SR. Ich ocena powinna się składać z trzech etapów: (1) oceny globalnej, (2) oceny lokalnej w celu identyfikacji obszarów zawierających błędy, (3) analizy zmienności widma gęstości mocy.

W celu dalszej oceny zaprojektowanego modelu MCWESRGAN, wykonano porównanie oszacowanych obrazów SR z obrazami oszacowanymi przez inne modele SISR oparte na głębokich sieciach neuronowych. Na potrzeby porównania, wykonano ocenę

globalną obrazów należących do zbiorów testowych przygotowanych baz danych. Wyniki analizy przedstawione są w poniższej tabeli 4.3.

Tabela 4.3.

Porównanie MCWESRGAN z innymi modelami SISR dla baz danych (BD): I. DOTA, II. zdjęcia lotnicze, III. WorldView-2. Kolorem zielonym zaznaczono najlepsze wyniki oceny jakościowej, kolorem żółtym te, które były drugie, zaś kolorem czerwonym najgorsze

wskaź-nik	BD	EDSR [40]	ESPCN [39]	FSRCNN [38]	RDN [58]	SRCNN [37]	SRDen-seNet [59]	SRGAN [41]	ESRGAN [43]	MCWE-SRGAN
SSIM	I	0.54	0.69	0.67	0.67	0.61	0.68	0.60	0.60	0.70
	II	0.45	0.60	0.56	0.60	0.56	0.61	0.54	0.15	0.62
	III	0.79	0.91	0.90	0.91	0.82	0.89	0.83	0.83	0.92
PSNR	I	20.62	24.11	24.58	23.61	22.87	24.49	21.48	22.76	25.05
	II	18.64	20.95	21.97	21.41	21.05	22.01	20.44	11.52	22.45
	III	24.76	27.15	31.27	32.01	25.4	30.12	29.18	28.74	32.10
SAM	I	0.24	0.15	0.15	0.16	0.18	0.15	0.17	0.19	0.14
	II	0.26	0.19	0.14	0.18	0.20	0.19	0.21	0.19	0.17
	III	0.18	0.14	0.10	0.11	0.14	0.12	0.15	0.14	0.09
SCC	I	0.12	0.29	0.17	0.25	0.14	0.23	0.13	0.13	0.32
	II	0.06	0.11	0.13	0.12	0.08	0.11	0.08	0.06	0.13
	III	0.22	0.31	0.20	0.34	0.16	0.31	0.15	0.14	0.37
UQI	I	0.96	0.99	0.98	0.98	0.97	0.98	0.95	0.98	0.99
	II	0.93	0.96	0.95	0.96	0.96	0.96	0.95	0.70	0.97
	III	0.91	0.98	0.97	0.98	0.93	0.97	0.91	0.95	0.98

Na podstawie analizy wizualnej (załącznik 4 publikacji **P3**) oraz wyników jakościowej oceny obrazu (tabela 4.3) jednoznacznie widać, że opracowany model MCWE-SRGAN umożliwia najdokładniejszą rekonstrukcję obrazu, szczególnie dla elementów i powierzchni liniowych. Model MCWESRGAN osiągnął najlepsze wyniki w każdej z ocenianych metryk - SSIM, PSNR, SAM, SCC i UQI. Dodatkowo można zauważyć, że obrazy SR z fragmentów zobrazowań satelitarnych dają lepsze wyniki w porównaniu do obrazów SR z bazy DOTA. Zjawisko to jest wynikiem znacznie niższej rozdzielczości przestrzennej zobrazowań satelitarnych w porównaniu z obrazami w bazie DOTA, co skutkuje mniejszą ilością szczegółów jakie muszą zostać zrekonstruowane. Ponadto na uwagę zasługuje model SRDenseNet. Obrazy SR oszacowane tym modelem osiągają równie dobre wyniki w ocenie jakościowej. Jednakże na podstawie analizy wizualnej można zauważyć, że obrazy SR sprawiają wrażenie namalowanych pędzlem, co odróżnia je od obrazów HR. Najgorsze wyniki uzyskała sieć EDSR (wartość SSIM poniżej 0.80

dla bazy zobrazowań satelitarnych, podczas gdy dla MCWESRGAN SSIM = 0.92). Prowadząc analizę wizualną (przykłady obrazów można znaleźć w załączniku 4 publikacji **P3**), widoczna jest degradacja potencjału interpretacyjnego.

Przeprowadzone badania pokazują, że zastosowanie wielokolumnowego dyskryminatora pozwala na znaczne przyspieszenie procesu treningu sieci. Dodatkowo proces ten można przyspieszyć poprzez wykorzystanie funkcji kosztu, zdefiniowanej jako starta Wassersteina (ang. Wasserstein loss), która uwzględnia odległość Wassersteina. Pomimo tego, że strata Wassersteina zwiększa podatność modelu na przetrenowanie (zwłaszcza w przypadku obrazów pozyskanych z pułapu lotniczego i kosmicznego), wytrenowane modele charakteryzują się większą jakością w zadaniu poprawy rozdzielczości przestrzennej. Analiza uzyskanych wyników (w głównej mierze oparta na lokalnej ocenie obrazów) pozwala stwierdzić, że obiekty antropogeniczne są lepiej reprezentowane na zdjęciach SR niż obiekty pochodzenia naturalnego (np. korony drzew). Wyjątek stanowią polany i łąki, których jakość oceny wzrasta wraz ze spadkiem rozdzielczości przestrzennej obrazów LR.

Na podstawie otrzymanych wyników, można stwierdzić, że jakość obrazów SR spada wraz ze wzrostem rozdzielczości przestrzennej obrazów LR (tabela III i IV publikacji **P3**). W przypadku obrazów pozyskanych z pułapu lotniczego (gdzie wartość parametru GSD wynosi poniżej 0.3 m) wartości badanych wskaźników odbiegają od pozostałych baz danych. Zjawisko to jest wynikiem licznych nieprawidłowości, będących skutkiem błędnego oszacowania szczegółów obiektów znajdujących się na zobrazowaniu. Dodatkowo w toku przeprowadzonych badań zauważono, że w przypadku stosowania jednorodnych baz danych (wykorzystane obrazy pochodzą z tego samego zobrazowania satelitarnego bądź nalotu fotogrametrycznego), gdzie warunki są zbliżone, prawdopodobieństwo wystąpienia zjawiska znikających gradientów wzrasta. W przypadku zastosowania baz niejednorodnych (np. DOTA, gdzie znajdują się zobrazowania satelitarne, lotnicze i pozyskane przez BSP) zjawiska znikających gradientów nie występują z uwagi na „pobudzanie" optymalizatorów przez obrazy o innej rozdzielczości przestrzennej.

Przeprowadzone analizy pozwoliły na uzyskanie odpowiedzi do wszystkich postawionych w tym etapie pytań badawczych PB3.1 – PB3.4 oraz umożliwiły udział w osiągnięciu celów szczegółowych CS3 i CS4, a także osiągnięcie celu szczegółowego C5, co bezpośrednio przyczyniło się do realizacji celu głównego CG.

4.4. Metoda poprawy rozdzielczości przestrzennej sekwencji obrazów pozyskanych przez małe satelity [publikacja P4]

Rozwój małych satelitów nie dotyczy jedynie rejestrowania scen satelitarnych, ale także sekwencji obrazów. Wiele z nich, wykorzystuje technikę obrazowania dynamicznego, która znajduje wiele zastosowań w zakresie teledetekcji. Rozwiązanie to, pomimo możliwości detekcji celów dynamicznych posiada również ograniczenia. Ze względu na obrazowanie w trybie wideo oraz stosowanie kompresji stratnej, ma miejsce degradacja rozdzielczości przestrzennej pozyskiwanych sekwencji. Jednym z możliwych rozwiązań tego problemu jest zastosowanie najnowszych osiągnięć uczenia głębokiego opartych o techniki single image super-resolution (SISR).

Problem jakości interpretacji kontekstu obrazu jest bardzo istotny w przypadku danych wideo, gdzie występują rozległe, skomplikowane wzorce, które zmieniają swoje położenie na kolejnych obrazach sekwencji. Zmiana położenia obiektów na kolejnych obrazach sekwencji jest wynikiem ruchu obiektów na obrazie, jak i ruchu satelity pozyskującej dane. Ponadto jakość odwzorowania obiektów infrastruktury budowlanej (w tym krytycznej) oraz drogowej jest kluczowe w zadaniach detekcji mniejszych obiektów np. pojazdów, samolotów.

Model MCWESRGAN opracowany i przetestowany w ramach publikacji **P3** posiada jedną dość istotną (pod względem interpretacyjnym) wadę. Oszacowane obrazy SR posiadają dość charakterystyczną (trudno zauważalną) teksturę. Powodem jej powstania jest niska wartość parametru odpowiedzialnego za szybkość uczenia, jednakże jest ona konieczna w celu ograniczenia (lub opóźnienia) wystąpienia zjawiska znikających gradientów. Z tego powodu, konieczne jest opracowanie metody umożliwiającej usunięcie błędnej tekstury.

W związku z powyższym, w ramach badań wykonano adaptację zaprojektowanego w poprzednich badaniach modelu sieci MCWESRGAN. W trakcie prowadzonych badań starałam się znaleźć odpowiedź na następujące pytania:

PB4.1. Jaką przyjąć metodykę do poprawy rozdzielczości przestrzennej sekwencji obrazu?

PB4.2. Czy możliwe jest przetworzenie sekwencji o dowolnej rozdzielczości czasowej?

PB4.3. Jak ograniczyć występowanie charakterystycznej tekstury na obrazach SR?

PB4.4. Jak ocenić jakość sekwencji obrazów po zastosowaniu algorytmów SISR?

W ramach prowadzonych badań (szczegółowo opisanych w **P4**) zaproponowano nową metodykę poprawy rozdzielczości sekwencji obrazów. Rozwiązanie to pozwala na poprawę rozdzielczości przestrzennej danych wideo, które mogą składać się z dowolnej liczby klatek o rozmiarze, który nie jest ściśle określony. Składa się ona z pięciu etapów: (1) wczytanie pojedynczej klatki wideo, (2) podział klatki na mniejsze kafelki (wymiar jest zdeterminowany parametrem wejścia do sieci GAN, która zostanie wykorzystana do poprawy rozdzielczości przestrzennej) z pokryciem 10% (wartość ta została określona podczas badań opisanych w publikacji **P2**), (3) poprawa rozdzielczości przestrzennej za pomocą algorytmu SISR, (4) poprawa jakości interpretacyjnej obrazów SR1, w wyniku której powstaną obrazy SR2, (5) ponowne połączenie obrazów oszacowanych (SR2) w jedną klatkę. Schemat wyżej przedstawionej metodyki został pokazany na rysunku 4.7.
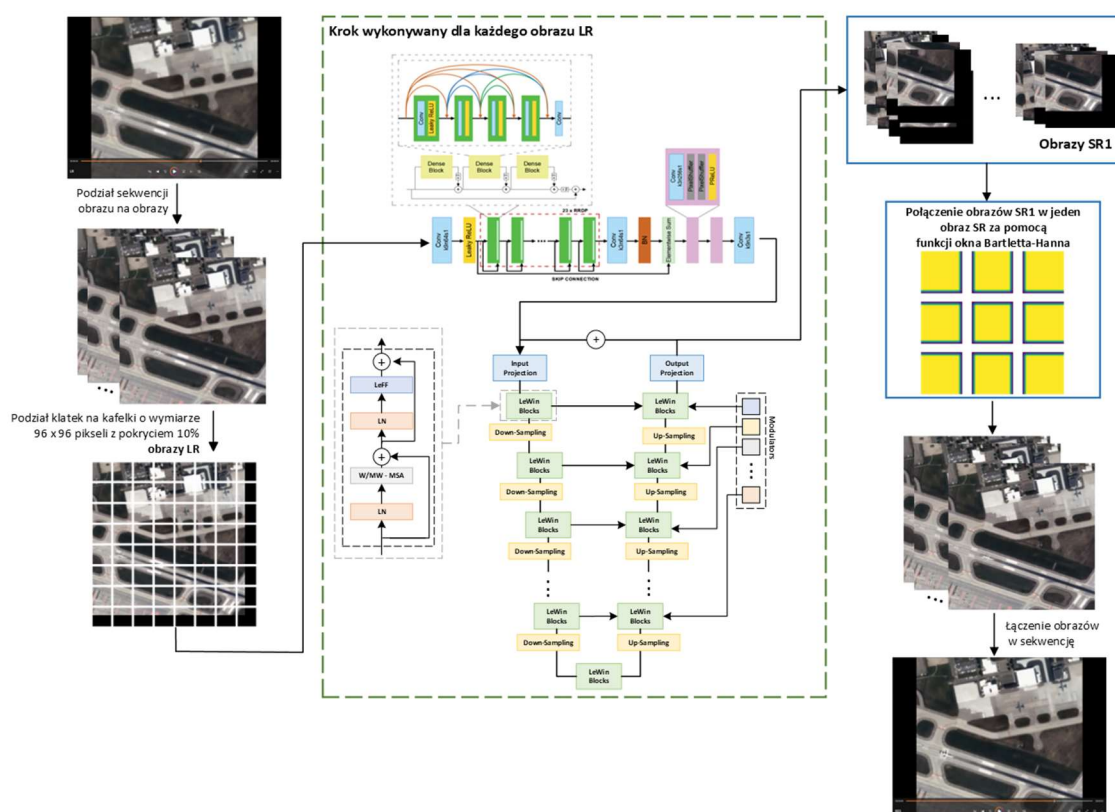
Rys. 4.7. *Schemat działania algorytmu (opracowano na podstawie* [60]*)*

Dotychczas przeprowadzone badania pozwalają na wykonanie wszystkich wyżej opisanych etapów, poza poprawą jakości interpretacyjnej obrazów SR1 (zakładając, że etap 1 i 5 to podstawowe operacje na danych wideo i nie wymagają szerszych badań). W związku z tym, przed przystąpieniem do testowania metodyki poprawy rozdzielczości

sekwencji obrazów, przystąpiono do badań nad poprawą jakości obrazów zrekonstruowanych przez model MCWESRGAN. W pierwszej kolejności zweryfikowano klasyczne metody przetwarzania obrazów cyfrowych, oraz nieco bardziej zaawansowane metody usuwania rozmycia - algorytm Lucy-Richardsona [61] oraz dekonwolucję Wienera [62]. Jednakże metody te nie pozwoliły na poprawę jakości obrazów SR. W związku z tym, konieczne było zastosowanie innego rozwiązania. Wykonując przegląd prac badawczych dotyczących usunięcia szumów oraz rozmycia na obrazach, na uwagę zasługuje model Uformer [63] zaproponowany przez zespół Zhendong Wang.

Model Uformer ma charakter klasycznej sieci koder-dekoder w kształcie litery „U". Dodatkowo pomiędzy koderem i dekoderem wykorzystane zostały połączenia pomijalne, dzięki którym możliwe jest dodanie wykrytych przez algorytm cech do obrazu wejściowego. Zadaniem modelu jest pozyskanie informacji o kontekście lokalnym pikseli „zanieczyszczonych", na potrzeby przywrócenia ich „czystej" wersji. W tym celu wykorzystywane są bloki transformatora (ang. transformer block) wyposażone w mechanizm samouwagi, dzięki któremu możliwe jest przechwytywanie długo-zasięgowych zależności. Ponadto mapy cech, które trafiają do bloków LeWin (ang. locally enhanced window, zaznaczone zielonym kolorem na rysunku 4.8) napotykają warstwę normalizacji (LN), a następnie Window-based Multi-head Self-Attention (W-MSA) korzystające z mechanizmu samouwagi wewnątrz niezachodzących na siebie okienek lokalnych. W kolejnym kroku, za pomocą połączeń pomijalnych, tensor ten dodawany jest do wejściowych map cech i przekazywany do kolejnej warstwy normalizacyjnej. Znormalizowany tensor przekazywany jest do Locally-enhanced Feed-Forward Network (LeFF), a tensor wynikowy ponownie (za pomocą skip connection) dodawany jest do map jakie przekazywane były do bloków LeWin. Dodatkowo dekoder sieci wyposażony jest w modulatory, których zadaniem jest poprawa jakości cech. W wyniku tej operacji możliwe jest odzyskanie większej liczby szczegółów (dzięki uwzględnieniu charakterystyk wzorców zaburzających).

W związku z tym, że prezentowane przez autorów wyniki oceny jakościowej były bardzo obiecujące, do dalszych badań przyjęto generator modelu MCWESRGAN rozbudowany o model Uformer. Na poniższym rysunku (rys. 4.8) przedstawiono finalną postać generatora odpowiedzialnego za poprawę rozdzielczości przestrzennej sekwencji obrazów.
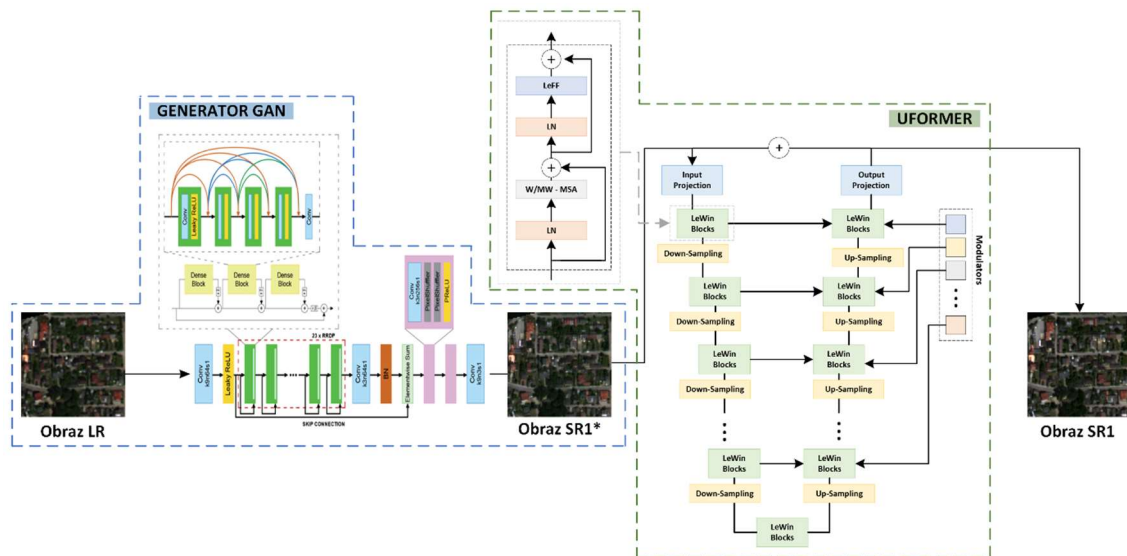
Rys. 4.8. *Schemat poprawy rozdzielczości pojedynczej klatki wideo (opracowano na podstawie [60])*

Podczas badań nad siecią MCWESRGAN (publikacja **P3**) zauważono, że uczenie modeli z wykorzystaniem baz danych złożonych z obrazów o różnej rozdzielczości przestrzennej jest mniej podane na zjawisko znikających gradientów. W związku z tym, przygotowana została nowa baza danych zawierająca zdjęcia pozyskane przez BSP, sensory optyczne zamieszczone na samolotach oraz satelity obserwacyjne (w tym, małe satelity). W wyniku przeprowadzonych prac, powstała baza danych złożona z obrazów o rozdzielczości przestrzennej od 0.15 m do 20 m składająca się z 100 000 obrazów. Tak różnorodna baza danych została utworzona w wyniku połączenia bazy danych DOTA, fragmentów zdjęć lotniczych, zobrazowań pozyskanych przez satelitę WorldView-2, bazę danych Jilin-1 [64] oraz inne dane pozyskane przez nanosatelity SkySat-1 i Jilin-1.

Do treningu modeli odpowiedzialnych za poprawę rozdzielczości konieczne jest przygotowanie danych o wysokiej rozdzielczości (HR) oraz o niskiej rozdzielczości (LR). Do przygotowania naszej bazy danych, w większości przypadków obrazy HR przygotowano poprzez podział obrazów na kafelki o wymiarach 384 × 384 pikseli, zaś ich odpowiedniki LR powstały w wyniki przepróbkowania obrazów HR do rozmiaru 96 × 96 pikseli. Wyjątkiem są sceny satelitarne pozyskane przez satelitę WV2. W przypadku tych danych, postąpiono tak samo jak w poprzednich badaniach (publikacja **P2** i **P3**).

W celu lepszej oceny oszacowanego obrazu zastosowano lokalną ocenę obrazu dla klatek wynikowych (po połączeniu kafelków za pomocą okien czasowych). Z uwagi na rozmiar ocenianych klatek wynikowych równych 640 × 640 pikseli, założono, że ocena lokalna wykonana zostanie na fragmentach obrazu o powierzchni 20 × 20

pikseli. Na poniższych rysunkach (rys. 4.9, 4.10) przedstawiono wyniki oceny lokalnej za pomocą metryk jakościowych SSIM i PSNR. W przypadku metryki SSIM większość badanych obszarów zaznaczonych została na kolor żółty, czyli wartość metryki wyniosła więcej niż 0.97. Istnieją jedynie pojedyncze obszary o niskiej jakości oszacowania. Natomiast ocena lokalna za pomocą PSNR jest znacznie różnorodna (różnica wartości PSNR pomiędzy sąsiednimi obszarami jest większa, niż w przypadku SSIM), co umożliwia bardziej dokładną analizę obrazu. W oparciu o ocenę lokalną z wykorzystaniem PSNR można wywnioskować, że obiekty poruszające się (np. samolot na pasie), bądź obiekty niewielkie o nieregularnych kształtach oszacowane zostały z wysoką jakością, choć niewiele niższą niż płaszczyzny postojowe, pas startowy i drogi kołowania.

Rys. 4.9. *Ocena lokalna- metryka SSIM (dla pola oceny wynoszącego: 20 × 20 pikseli)* [60]

Rys. 4.10. *Ocena lokalna- metryka PSNR (dla pola oceny wynoszącego: 20 × 20 pikseli)* [60]

Kolejną metodą wykorzystaną na potrzeby określenia jakości obrazów SR jest wykonanie PSD, która opisuje rozkład częstotliwości energii sygnału. Jak już wspomniano wcześniej, ocena ta pozwala na określenie terenowej zdolności rozdzielczej.
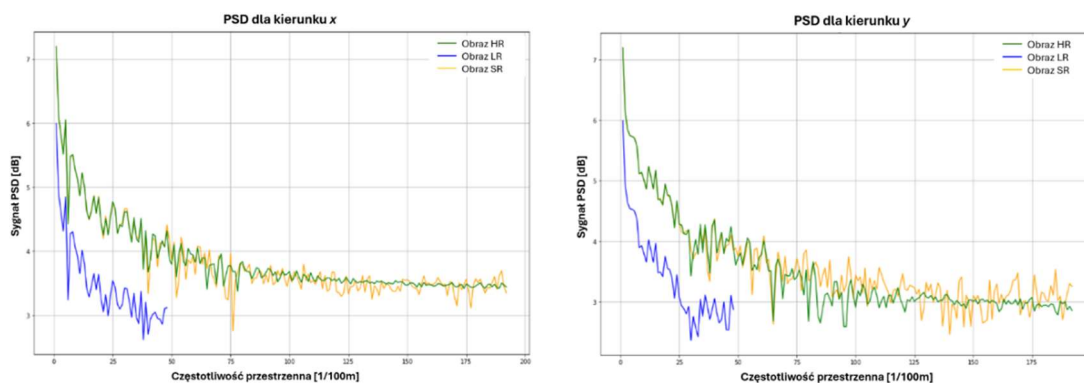
Rys. 4.11. *Wykres PSD na kierunku x i y dla przykładowego obrazu (opracowano na podstawie* [60]*)*

Na rysunku 4.11 przedstawiono analizę PSD dla kierunku x i y. Wykonując analizę PSD można zauważyć wysokie podobieństwo rozkładu częstotliwości energii sygnału pomiędzy obrazem HR (żółta krzywa), a obrazem SR (zielona krzywa) zwłaszcza na kierunku wertykalnym obrazu (kierunek y). Na kierunku horyzontalnym poprawa rozdzielczości widoczna jest dla gęstości przestrzennej wynoszącej ok. 100 [1/100m], natomiast na kierunku wertykalnym ok. 120 [1/100m]. W związku z tym wartość parametru GRD na kierunku x wynosi GRDx = 2.1 × GSD, a na kierunku y GRDy = 2.5 × GSD. Otrzymane wartości potwierdzają, że wartość parametru GRD dla obiektów o wysokim kontraście przyjmuje wartość nie mniejszą niż dwukrotność GSD.

Proponowana metodyka poprawy rozdzielczości sekwencji obrazów składa się z trzech zasadniczych etapów: przygotowania obrazów, poprawy rozdzielczości przestrzennej oraz połączenia oszacowanych obrazów w sekwencję obrazu. Konstrukcja rozwiązania pozwala na wykorzystanie innych algorytmów SISR wykorzystujących sieci neuronowe. W związku z tym, w ramach badań wykonano porównanie jakości poprawy rozdzielczości przestrzennej obrazów za pomocą różnych metod wykorzystujących sieci splotowe. Dodatkowo każdy z badanych modeli SISR rozbudowany został o model Uformer, co pozwoliło na lepszą ocenę wpływu modelu Uformer na jakość wynikowych obrazów SR. Wyniki globalnej oceny jakości zestawu testowego przedstawiono w tabeli 4.4. Ponadto w tabeli 4.5 przedstawiono przykłady obrazów SR oszacowanych przez różne modele SISR. Więcej próbek znajduje się w załączniku A publikacji **P4.** W załączniku tym, znajdują się również obrazy, które zostały oszacowane przez inne modele SISR rozbudowane o warstwy sieci Uformer. Globalna ocena jakościowa tych obrazów znajduje się w załączniku B publikacji **P4.**

Tabela 4.4.

Porównanie MCWESRGAN z innymi modelami SISR. Kolorem zielonym zaznaczono najlepsze wyniki oceny jakościowej, kolorem żółtym te, który były drugie, zaś kolorem czerwonym najgorsze. "*" zaznaczono modele SISR rozbudowane o warstwy sieci Uformer (opracowano na podstawie [60]*)*

Metryka / Model SISR	SSIM	PSNR	SAM	SCC	UQI
LR	0.92	31.37	0.07	0.15	0.991
EDSR	0.87	25.36	0.11	0.20	0.889
EDSR*	0.90	25.62	0.11	0.26	0.955
ESPCN	0.95	27.63	0.08	0.26	0.939
ESPCN*	0.94	27.16	0.08	0.25	0.970
FSRCNN	0.90	27.23	0.08	0.18	0.775
FSRCNN*	0.93	27.23	0.08	0.24	0.904
RDN	0.94	27.59	0.08	0.26	0.958
RDN*	0.94	27.45	0.08	0.25	0.967
SRCNN	0.93	32.40	0.06	0.19	0.964
SRCNN*	0.94	33.06	0.06	0.25	0.987
SRDenseNet	0.94	27.56	0.08	0.24	0.970
SRDenseNet*	0.94	27.46	0.08	0.25	0.973
SRGAN	0.62	19.92	0.235	0.06	0.758
SRGAN*	0.67	20.91	0.209	0.17	0.760
ESRGAN	0.78	29.35	0.086	0.14	0.788
ESRGAN*	0.83	30.04	0.080	0.21	0.793
MCWESRGAN	0.96	36.21	0.03	0.26	0.993
MCWESRGAN*	0.98	38.32	0.02	0.27	0.996

Tabela 4.5.

Przykłady fragmentów obrazu SR oszacowanych za pomocą różnych modeli sieci neuronowych
[60]

LR	HR	EDSR	ESPCN
FSRCNN	RDN	SRCNN	SRDenseNet
SRGAN	ESRGAN	MCWESRGAN	MCWESRGAN z modelem Uformer
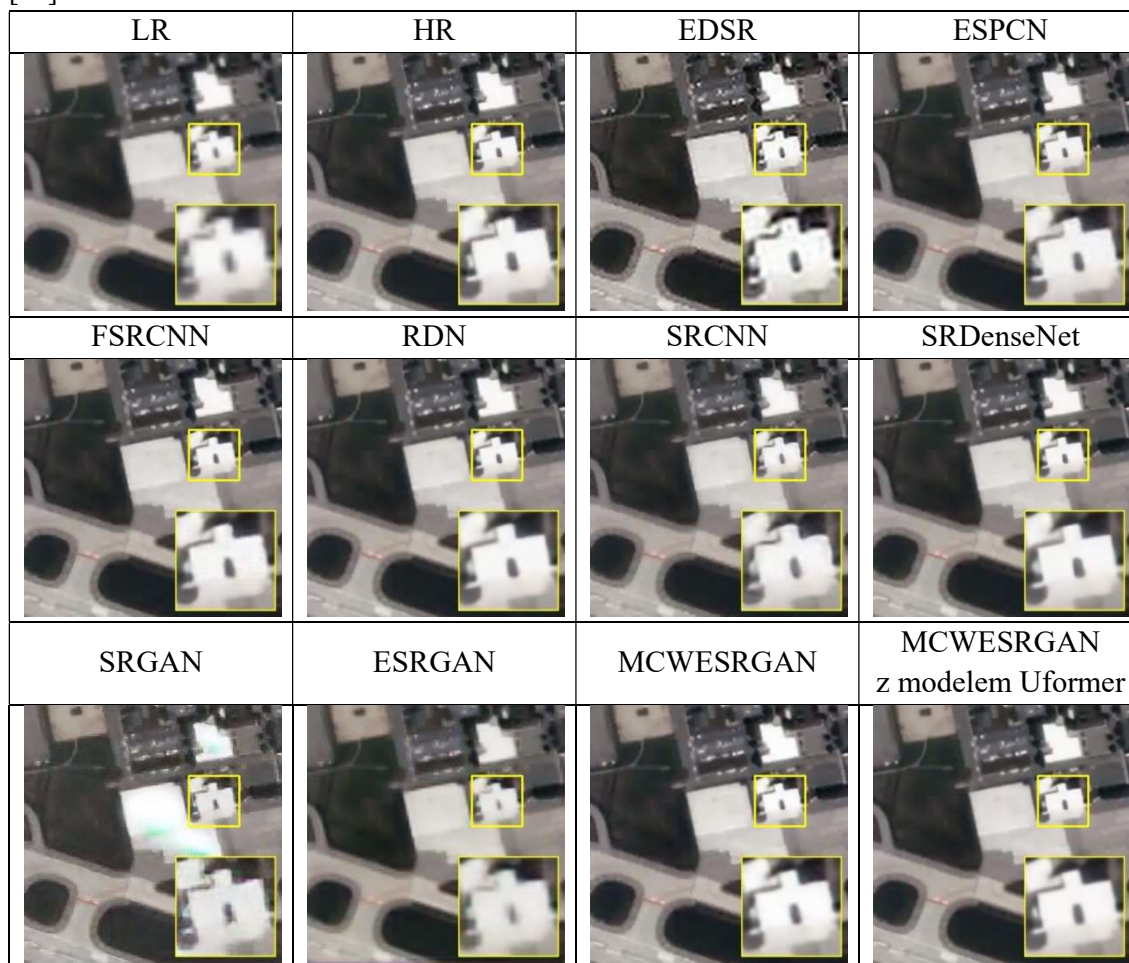
Analizując zestawienie obrazów SR wygenerowanych przez różne rozwiązania SISR wykorzystujące sieci splotowe (CNN) możemy zauważyć duże różnice, zwłaszcza w odwzorowaniu krawędzi obiektów antropogenicznych (stworzonych przez człowieka). Przykład takiego obiektu został zaznaczony żółtą ramką (tabela 4.5), a jego powiększenie znajduje się w prawym dolnym rogu rysunku. Wiele z badanych modeli np. EDSR, FSRCNN, SRCNN, SRGAN, ESRGAN popełnia błędy podczas odwzorowania obrazów HR. W przypadku obrazów oszacowanych przez te metody, krawędzie obiektów (nawet tych posiadających duży kontrast z tłem) są rozmyte, a narożniki zaokrąglone. Ocenę wizualną potwierdziła analiza jakościowa (tabela 4.5). Zaprojektowany model MCWE-SRGAN rozszerzony o warstwy Uformer pozwala na poprawę rozdzielczości przestrzennej z najwyższą jakością (SSIM = 0.98, PSNR = 38.32 dB). W przypadku większości innych modeli SISR, które zostały rozbudowane o warstwy Uformer widoczna jest poprawa jakości obrazów SR. W przypadku modeli ESPCN, RDN i SRDenseNet

zaobserwowano nieznaczny spadek jakości obrazów SR (np. w przypadku sieci RDN spadek PSNR wynosił 0.14 dB).

Zaprojektowana metoda poprawy rozdzielczości sekwencji obrazów składa się z kilku etapów działających niezależnie. Dzięki działaniu etapowemu proponowane rozwiązanie nie wymaga wysokich mocy obliczeniowych. Moc obliczeniowa stacji roboczej determinuje jedynie czas przetwarzania sekwencji obrazu. W tabeli 4.6 przedstawiono porównanie czasów pracy modułu poprawy rozdzielczości przestrzennej danych wideo składających się z 100 klatek o wymiarach 120 × 120 pikseli. Zestawienie to pokazuje, że proponowany algorytm może zostać zaimplementowany nawet na urządzeniach niewyposażonych w procesor graficzny. Chociaż wykorzystanie karty graficznej znacznie skraca proces przetworzenia sekwencji obrazów.

Tabela 4.6.

*Porównanie czasu potrzebnego na przetworzenie pojedynczego filmu na różnych jednostkach obliczeniowych (wymagany czas na przetworzenie: * jednego kafelka, ** jednego obrazu sekwencji) (opracowano na podstawie* [60]*)*

stacja robocza / etap	Czas			
	Intel Xeon Silver 4216 z GPU TITAN RTX 24 GB	Intel Xeon Silver 4216	Intel Corei5-1235	Intel Corei3-7100U
Przygotowanie danych	0.004 (0.00004**)	0.04 (0.0004**)	0.2 (0.002**)	0.4 (0.004**)
Poprawa rozdzielczości	76.80 (0.19*)	988 (2.47*)	4316 (10.79*)	8861 (22.15*)
Połączenie w dane wideo	0.07 (0.0007**)	0.9 (0.009**)	4 (0.04**)	8 (0.08**)
Całkowity czas poprawy rozdzielczości przestrzennej jednego filmu	76.87 (~1 min 17s) (0.77**)	988.94 (~16 min 29s) (9.9**)	4320.2 (~72 min) (43.2**)	8869.4 (~147 min 49s) (88.7**)

Otrzymane wyniki potwierdzają zasadność rozbudowania modelu MCWE-SRGAN o warstwy modelu Uformer. Modyfikacja ta znacząco poprawia jakość obrazów SR, a w konsekwencji poprawę możliwości interpretacyjnych. Wadą tego rozwiązania jest wydłużenie pracy modułu odpowiedzialnego za poprawę rozdzielczości przestrzennej obrazu sekwencji. Poprawa rozdzielczości przestrzennej pojedynczego kafelka przez model MCWESRGAN (nierozbudowany o warstwy Uformer) trwa 0.03s (przy

korzystaniu z Intel Xeon Silver 4216 z TITAN RTX 24 GB), czyli proces ten trwa ponad 6 razy krócej, niż w przypadku modelu MCWESRGAN rozbudowanego o model Uformer.

W ramach niniejszych badań zaproponowana została nowa metodyka poprawy rozdzielczości przestrzennej sekwencji obrazów. Otrzymane wyniki pokazują czterokrotną poprawę rozdzielczości przestrzennej obrazów o niskiej rozdzielczości, jednocześnie możliwości interpretacyjne wzrosły ponad dwukrotnie. Ponadto rozbudowanie generatora odpowiedzialnego za poprawę rozdzielczości przestrzennej pozwoliło na poprawę podobieństwa strukturalnego (SSIM) pomiędzy obrazami HR i SR o ok. 0.02, zaś szczytowy stosunek sygnału do szumu wzrósł o ponad 2dB. Zaproponowana metodyka może znaleźć zastosowanie w problemach poprawy rozdzielczości przestrzennej obrazów i wideo pozyskanych przez małe satelity, w których konieczne jest zastosowanie metod single image super resolution (SISR). Ponadto poprawa rozdzielczości przestrzennej za pomocą wyżej opisanej metodyki nie wymaga dużych mocy obliczeniowych (te wymagane są jedynie na etapie treningu modelu), przez co może być obsługiwana przez jednostki robocze niewyposażone w procesory graficzne. Przeprowadzone analizy pozwoliły mi na uzyskanie odpowiedzi do wszystkich postawionych w tym etapie pytań badawczych PB4.1 – PB4.4 oraz umożliwiły udział w osiągnięciu celów szczegółowych CS3, CS4 i CS6, co bezpośrednio przyczyniło się do realizacji celu głównego CG.

5. Posumowanie i wnioski

Niniejsza rozprawę doktorską pt. *Opracowanie metodyki podwyższenia rozdzielczości przestrzennej zobrazowań z małych satelitów z zastosowaniem algorytmów sztucznej inteligencji* stanowi spójny tematycznie cykl czterech publikacji, którego celem było zaproponowanie nowego podejścia do poprawy rozdzielczości przestrzennej zobrazowań satelitarnych. Przedstawiona metodyka umożliwia czterokrotną poprawę rozdzielczości przestrzennej oraz co najmniej dwukrotny wzrost możliwości interpretacyjnych zobrazowań satelitarnych lub sekwencji obrazów o dowolnym rozmiarze.

W niniejszej rozprawie doktorskiej zweryfikowałam postawioną hipotezę główną: **zastosowanie odpowiednio dobranych i wytrenowanych generatywnych sieci przeciwstawnych (GAN) pozwoli na opracowanie metodyki podwyższania rozdzielczości przestrzennej zobrazowań z małych satelitów.** W ramach prowadzonych badań opracowałam metodykę poprawy rozdzielczości przestrzennej zobrazowań i sekwencji obrazów pozyskanych przez małe satelity. Rdzeniem metodyki jest generatywna sieć przeciwstawna, do której szkolenia zastosowałam wielokolumnowy dyskryminator oraz funkcję straty Wasserseina. Jakość obrazów zrekonstruowanych według opisanej metodyki przewyższa popularne rozwiązania SISR, co zostało wykazane na podstawie analizy jakościowej w publikacjach **P3** i **P4**.

Pierwszym etapem badań był szczegółowy przegląd metod poprawy rozdzielczości przestrzennej zobrazowań satelitarnych, ze szczególnym uwzględnieniem metod bazujących wyłącznie na obrazach o niskiej rozdzielczości (publikacja P1). Na podstawie tego przeglądu zauważyłam, że: **(1) wiele metod SISR nie prowadzi do zwiększenia możliwości interpretacyjnych obrazów SR, (2) badacze często prezentują efektywność metod SISR na obrazach znacznie mniejszych od zobrazowań satelitarnych, (3) wskaźniki wykorzystywane do oceny obrazów SR nie zawsze pozwalają na jednoznaczną ocenę wzrostu możliwości interpretacyjnych**. Wnioski te umożliwiły mi udowodnienie pierwszej hipotezy pomocniczej (**H1**) oraz osiągnięcie pierwszego celu szczegółowego (**CS1**).

Celem kolejnego etapu badań było opracowanie metodyki przetwarzania całych zobrazowań satelitarnych w kontekście automatycznej interpretacji zobrazowań (publikacja **P2**). Proponowana metoda składa się z trzech etapów: (1) podział zobrazowania na mniejsze fragmenty (kafelki) o rozmiarze odpowiadającym parametrowi wejściowemu

sieci neuronowej (SN), (2) poprawa rozdzielczości przestrzennej za pomocą SN, (3) ponowne połączenie zrekonstruowanych kafelków w jeden spójny obraz. Do scalania kafelków zastosowałam okna czasowe, wykazując, że **użycie okien Hann, Hann-Poisson, Bartlett-Hann lub Triangular pozwala na uzyskanie obrazu o porównywalnej jakości. Kluczowym czynnikiem wpływającym na jakość końcowego obrazu jest stopień pokrycia między kafelkami**. Na podstawie przeprowadzonych eksperymentów ustaliłam, że **najlepsze rezultaty osiąga się przy pokryciu wynoszącym około 10%**. Zaletą tej metody jest fakt, że **poprawa rozdzielczości przestrzennej zobrazowań o dowolnym rozmiarze nie wymaga większych mocy obliczeniowych, niż wyostrzenie pojedynczego kafelka**. Opracowana metodyka pozwoliła mi na udowodnienie drugiej hipotezy pomocniczej (**H2**) oraz osiągnięcie drugiego celu szczegółowego (**CS2**).

Kluczowym etapem moich badań było opracowanie metodyki poprawy rozdzielczości przestrzennej obrazów satelitarnych pozyskanych przez małe satelity, która wykorzystuje sieci GAN (publikacja **P3**). W ramach badań udowodniłam, że **zastosowanie dyskryminatora wielokolumnowego pozwala na znaczne przyspieszenie procesu treningu sieci.** Co więcej**, zaimplementowanie straty Wassersteina dodatkowo przyspiesza ten proces**. Chociaż strata Wassersteina zwiększa podatność modelu na przetrenowanie, zwłaszcza w przypadku obrazów pozyskanych z pułapu kosmicznego, wytrenowane modele cechują się wyższą jakością. Ponadto zaproponowałam **trójetapowy proces oceny jakości obrazów oszacowanych (SR)**: (1) ocena globalna, powszechnie stosowana przez badaczy, w której wskaźniki są często zawyżane przez wysokie odwzorowanie powierzchni jednorodnych, (2) ocena lokalna, umożliwiająca identyfikację obszarów o niższej lub wyższej jakości oszacowania SR, (3) analiza rozkładu częstotliwości energii sygnału (PSD), pozwalająca na określenie wzrostu możliwości interpretacyjnych danych obrazowych. Podczas realizacji badań zauważyłam, że zastosowanie baz niejednorodnych, złożonych z obrazów o różnej rozdzielczości przestrzennej, sprawia, że model jest bardziej odporny na pojawienie się zjawiska znikających gradientów. Opracowana metodyka pozwoliła mi na udowodnienie dwóch hipotez pomocniczych: **H3**, **H4** oraz osiągnięcie trzech celów szczegółowych **CS3**, **CS4** oraz **CS5**.

Podsumowaniem moich prac badawczych było opracowanie metodyki poprawy rozdzielczości przestrzennej sekwencji obrazów pozyskanych przez małe satelity. W toku tych badań zweryfikowałam dotychczasowe metody wyostrzania danych uzyskanych techniką obrazowania dynamicznego. Analizując wyniki eksperymentów, postanowiłam dodatkowo rozbudować generator zaprojektowanej sieci GAN, ponieważ generowane

przez niego obrazy miały charakterystyczną teksturę spowodowaną niską szybkością uczenia modelu. W tym celu **rozszerzyłam model generatora o sieć koder-dekoder w kształcie litery "U"**. **Wprowadzona modyfikacja pozwoliła na poprawę podobieństwa strukturalnego między obrazami HR i SR o około 0.02, a szczytowy stosunek sygnału do szumu wzrósł o ponad 2dB**. Rozbudowany model generatora może być wykorzystywany do poprawy rozdzielczości przestrzennej zarówno zobrazowań satelitarnych, jak i sekwencji obrazów. Opracowana metodyka pozwoliła mi na udowodnienie dwóch hipotez pomocniczych: **H3**, **H5** oraz osiągnięcie trzech celów szczegółowych **CS3** oraz **CS6**.

Wszystkie opracowane w ramach badań metody przyczyniły się do osiągnięcia celu głównego (**CG**) jakim było **opracowanie metodyki wykorzystującej algorytmy sztucznej inteligencji na potrzeby podwyższenia rozdzielczości przestrzennej oraz poprawy potencjału interpretacyjnego zobrazowań z małych satelitów.** Jednocześnie, przeprowadzone prace badawcze pozwoliły na potwierdzenie hipotezy głównej: **Zastosowanie odpowiednio dobranych i wytrenowanych generatywnych sieci przeciwstawnych (GAN) pozwoli na opracowanie metodyki podwyższania rozdzielczości przestrzennej zobrazowań z małych satelitów.**

Opracowana metodyka stanowi rozwiązanie problemów, które mogą się pojawić, gdy Polska stanie się posiadaczem własnych operacyjnych systemów obserwacyjnych, w skład których będą wchodzić małe satelity. Projektując tę metodykę, zwróciłam uwagę na niskie wymagania obliczeniowe, co pozwala na jej wykorzystanie na powszechnie używanych stacjach roboczych.

Uważam, że zaprezentowane rozwiązania i wyniki prac badawczych stanowią „novum" w literaturze przedmiotu badań, co zostało potwierdzone publikacjami w czasopismach wyróżnionych przez Journal Citation Reports.

LITERATURA

[1]   „Satellite Database | Union of Concerned Scientists". Dostęp: 17 wrzesień 2021. [Online]. Dostępne na: https://www.ucsusa.org/resources/satellite-database

[2]   G. W. Nagel, E. M. L. de M. Novo, i M. Kampel, „Nanosatellites applied to optical Earth observation: a review", *Rev. Ambient. Água*, t. 15, cze. 2020, doi: 10.4136/ambi-agua.2513.

[3]   E. Buchen i D. DePasquale, „SpaceWorks' 2014 Nano/Microsatellite Market Assessment", *SpaceWorks Enterprises*, nr 12, 2014.

[4]   H. Kuuste *i in.*, „Imaging system for nanosatellite proximity operations", *Proc. Estonian Acad. Sci.*, t. 63, nr 2S, s. 250, 2014, doi: 10.3176/proc.2014.2S.06.

[5]   J. Blommaert *i in.*, *CHIEM: A new compact camera for hyperspectral imaging.* 2017.

[6]   K. Karwowska i D. Wierzbicki, „Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, t. 15, s. 3292–3312, 2022, doi: 10.1109/JSTARS.2022.3167646.

[7]   K. Edwards i P. Davis, „The use of intensity-hue-saturation transformation for producing color shaded relief images", *Photogrammetric Engineering and Remote Sensing*, t. 60, nr 11, s. 1369–1374, 1994.

[8]   L. Wald, T. Ranchin, i M. Mangolini, „Fusion of Satellite Images of Different Spatial Resolutions: Assessing the Quality of Resulting Images", *Photogrammetric Engineering and Remote Sensing*, t. 63, nr 6, s. 691–699, 1997.

[9]   A. Fruchter i R. N. Hook, „Novel image reconstruction method applied to deep Hubble space telescope images", w *Applications of Digital Image Processing XX*, International Society for Optics and Photonics, paź. 1997, s. 120–125. doi: 10.1117/12.292751.

[10] P. Dherete i B. Rouge, „Image de-blurring and application to SPOT5 THR satellite imaging", w *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*, lip. 2003, s. 318–320 t.1. doi: 10.1109/IGARSS.2003.1293762.

[11] C. Latry i B. Rouge, „Super resolution: quincunx sampling and fusion processing", w *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*, lip. 2003, s. 315–317 t.1. doi: 10.1109/IGARSS.2003.1293761.

[12] I. Goodfellow *i in.*, „Generative Adversarial Nets", w *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2014. Dostęp: 8 grudzień

2021. [Online]. Dostępne na: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

[13] O. Salvado, C. M. Hillenbrand, i D. L. Wilson, „Partial Volume Reduction by Interpolation with Reverse Diffusion", *Int J Biomed Imaging*, t. 2006, 2006, doi: 10.1155/IJBI/2006/92092.

[14] G. Chaudhary, K. Das, i M. Gopi, „Curvature Minimizing Depth Interpolation for Intuitive and Interactive Space Curve Sketching", s. 2.

[15] B. S. Morse i D. Schwartzwald, „Isophote-based interpolation", w *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269)*, paź. 1998, s. 227–231 t.3. doi: 10.1109/ICIP.1998.999013.

[16] A. Belov i A. Denisova, „Spatial interpolation methods for spectral-spatial remote sensing image super-resolution algorithm based on gradient descent approach", *Journal of Physics: Conference Series*, t. 1368, s. 032006, lis. 2019, doi: 10.1088/1742-6596/1368/3/032006.

[17] X. Qifang, Y. Guoqing, i L. Pin, „Super-resolution Reconstruction of Satellite Video Images Based on Interpolation Method", *Procedia Computer Science*, t. 107, s. 454–459, sty. 2017, doi: 10.1016/j.procs.2017.03.089.

[18] J. Chavez, S. C. Sides, i J. A. Anderson, „Comparison of three different methods to merge multiresolution and multispectral data: LANDSAT TM and SPOT panchromatic", *AAPG Bulletin (American Association of Petroleum Geologists); (USA)*, t. 74:6, Art. nr CONF-900702-, cze. 1990, Dostęp: 17 czerwiec 2024. [Online]. Dostępne na: https://www.osti.gov/biblio/6165108

[19] A. R. Gillespie, A. B. Kahle, i R. E. Walker, „Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques", *Remote Sensing of Environment*, t. 22, nr 3, s. 343–365, sie. 1987, doi: 10.1016/0034-4257(87)90088-5.

[20] W. A. Hallada i S. Cox, „Image sharpening for mixed spatial and spectral resolution satellite systems", zaprezentowano na International Symposium on Remote Sensing of Environment, Ann Arbor, MI, sty. 1983. Dostęp: 17 czerwiec 2024. [Online]. Dostępne na: https://ntrs.nasa.gov/citations/19850028100

[21] M. Ehlers, S. Klonus, P. J. Åstrand, i P. Rosso, „Multi-sensor image fusion for pansharpening in remote sensing", *International Journal of Image and Data Fusion*, mar. 2010, doi: 10.1080/19479830903561985.

[22] R. E. Crippen, „A simple spatial filtering routine for the cosmetic removal of scanline noise from Landsat TM P-tape imagery", *Photogrammetric Engineering and Remote Sensing*, t. 55, mar. 1989, Dostęp: 17 czerwiec 2024. [Online]. Dostępne na: https://ntrs.nasa.gov/citations/19890044964

[23] C. A. Laben i B. V. Brower, „Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening", US6011875A, 4 styczeń 2000 Dostęp: 17

czerwiec 2024. [Online]. Dostępne na: https://patents.google.com/patent/US6011875A/en

[24] C. E. Shannon, „A mathematical theory of communication", *The Bell System Technical Journal*, t. 27, nr 3, s. 379–423, lip. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

[25] M. Bevilacqua, A. Roumy, C. Guillemot, i M. A. Morel, „Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding", w *Proceedings of the British Machine Vision Conference 2012*, Surrey: British Machine Vision Association, 2012, s. 135.1-135.10. doi: 10.5244/C.26.135.

[26] R. Zeyde, M. Elad, i M. Protter, „On Single Image Scale-Up Using Sparse-Representations", w *Curves and Surfaces*, t. 6920, J.-D. Boissonnat, P. Chenin, A. Cohen, C. Gout, T. Lyche, M.-L. Mazure, i L. Schumaker, Red., w Lecture Notes in Computer Science, vol. 6920. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, s. 711–730. doi: 10.1007/978-3-642-27413-8_47.

[27] D. Martin, C. Fowlkes, D. Tal, i J. Malik, „A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics", w *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, lip. 2001, s. 416–423 t.2. doi: 10.1109/ICCV.2001.937655.

[28] J.-B. Huang, A. Singh, i N. Ahuja, „Single image super-resolution from transformed self-exemplars", w *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, cze. 2015, s. 5197–5206. doi: 10.1109/CVPR.2015.7299156.

[29] T. Karras, L. Samuli, i A. Timo, „A style-based generator architecture for generative adversarial networks", zaprezentowano na CVF conference on computer vision and pattern recognition, Proceedings of the IEEE, 2019.

[30] B. Keelan, *Handbook of Image Quality: Characterization and Prediction*. Boca Raton: CRC Press, 2002. doi: 10.1201/9780203910825.

[31] Y. Chen i G. Zhang, „A Pan-Sharpening Method Based on Evolutionary Optimization and IHS Transformation", *Mathematical Problems in Engineering*, t. 2017, nr 1, s. 8269078, 2017, doi: 10.1155/2017/8269078.

[32] L. Wald, „Quality of high resolution synthesised images: Is there a simple criterion ?", w *Third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images"*, T. Ranchin i L. Wald, Red., Sophia Antipolis, France: SEE/URISCA, sty. 2000, s. 99–103. Dostęp: 17 wrzesień 2021. [Online]. Dostępne na: https://hal.archives-ouvertes.fr/hal-00395027

[33] Q. Du, O. Gungor, i J. Shan, *Performance evaluation for pan-sharpening techniques*. 2005, s. 4266. doi: 10.1109/IGARSS.2005.1525860.

[34] A. Goetz, W. Boardman, i R. Yunas, „Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm", *Proceedings of the 3rd Annu*, nr 147–149, 1992.

[35] C.-I. Chang, „Spectral information divergence for hyperspectral image analysis", w *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*, cze. 1999, s. 509–511 t.1. doi: 10.1109/IGARSS.1999.773549.

[36] Z. Wang i A. C. Bovik, „A universal image quality index", *IEEE Signal Processing Letters*, t. 9, nr 3, s. 81–84, mar. 2002, doi: 10.1109/97.995823.

[37] C. Dong, C. C. Loy, K. He, i X. Tang, „Image Super-Resolution Using Deep Convolutional Networks", *arXiv:1501.00092 [cs]*, lip. 2015, Dostęp: 15 lipiec 2021. [Online]. Dostępne na: http://arxiv.org/abs/1501.00092

[38] C. Dong, C. C. Loy, i X. Tang, „Accelerating the Super-Resolution Convolutional Neural Network", *arXiv:1608.00367 [cs]*, sie. 2016, Dostęp: 15 lipiec 2021. [Online]. Dostępne na: http://arxiv.org/abs/1608.00367

[39] W. Shi *i in.*, „Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", *arXiv:1609.05158 [cs, stat]*, wrz. 2016, Dostęp: 19 lipiec 2021. [Online]. Dostępne na: http://arxiv.org/abs/1609.05158

[40] B. Lim, S. Son, H. Kim, S. Nah, i K. M. Lee, „Enhanced Deep Residual Networks for Single Image Super-Resolution", *arXiv:1707.02921 [cs]*, lip. 2017, Dostęp: 15 lipiec 2021. [Online]. Dostępne na: http://arxiv.org/abs/1707.02921

[41] C. Ledig *i in.*, „Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", *arXiv:1609.04802 [cs, stat]*, maj 2017, Dostęp: 14 czerwiec 2021. [Online]. Dostępne na: http://arxiv.org/abs/1609.04802

[42] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, i L. Zhang, „Second-Order Attention Network for Single Image Super-Resolution", w *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, cze. 2019, s. 11057–11066. doi: 10.1109/CVPR.2019.01132.

[43] X. Wang *i in.*, „ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks", *arXiv:1809.00219 [cs]*, wrz. 2018, Dostęp: 15 czerwiec 2021. [Online]. Dostępne na: http://arxiv.org/abs/1809.00219

[44] X. Wang, J. Ma, i J. Jiang, „Hyperspectral Image Super-Resolution via Recurrent Feedback Embedding and Spatial-Spectral Consistency Regularization", *IEEE Transactions on Geoscience and Remote Sensing*, s. 1–13, 2021, doi: 10.1109/TGRS.2021.3064450.

[45] L. Courtrai, M.-T. Pham, i S. Lefèvre, „Small Object Detection in Remote Sensing Images Based on Super-Resolution with Auxiliary Generative Adversarial

Networks", *Remote Sensing*, t. 12, nr 19, Art. nr 19, sty. 2020, doi: 10.3390/rs12193152.

[46] R. Dong, L. Zhang, i H. Fu, „RRSGAN: Reference-Based Super-Resolution for Remote Sensing Image", *EEE Transactions on Geoscience and Remote Sensing*, nr 60, s. 1–17, sty. 2021, doi: doi: 10.1109/TGRS.2020.3046045.

[47] A. Bordone Molini, D. Valsesia, G. Fracastoro, i E. Magli, „DeepSUM: Deep Neural Network for Super-Resolution of Unregistered Multitemporal Images", *IEEE Transactions on Geoscience and Remote Sensing*, t. 58, nr 5, s. 3644–3656, 2020, doi: 10.1109/TGRS.2019.2959248.

[48] Z. Wang *i in.*, „FeNet: Feature Enhancement Network for Lightweight Remote-Sensing Image Super-Resolution", *IEEE Transactions on Geoscience and Remote Sensing*, t. 60, s. 1–12, 2022, doi: 10.1109/TGRS.2022.3168787.

[49] K. Karwowska i D. Wierzbicki, „Improving Spatial Resolution of Satellite Imagery Using Generative Adversarial Networks and Window Functions", *Remote Sensing*, t. 14, nr 24, Art. nr 24, sty. 2022, doi: 10.3390/rs14246285.

[50] Y. Chen, „A New Methodology of Spatial Cross-Correlation Analysis", *PLOS ONE*, t. 10, nr 5, s. e0126158, maj 2015, doi: 10.1371/journal.pone.0126158.

[51] H. R. Sheikh i A. C. Bovik, „Image information and visual quality", *IEEE Transactions on Image Processing*, t. 15, nr 2, s. 430–444, luty 2006, doi: 10.1109/TIP.2005.859378.

[52] G.-S. Xia *i in.*, „DOTA: A Large-Scale Dataset for Object Detection in Aerial Images", zaprezentowano na Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, s. 3974–3983. Dostęp: 13 grudzień 2023. [Online]. Dostępne na: https://openaccess.thecvf.com/content_cvpr_2018/html/Xia_DOTA_A_Large-Scale_CVPR_2018_paper.html

[53] K. Karwowska i D. Wierzbicki, „MCWESRGAN: Improving Enhanced Super-Resolution Generative Adversarial Network for Satellite Images", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, t. 16, s. 9459–9479, 2023, doi: 10.1109/JSTARS.2023.3322642.

[54] M. Arjovsky, S. Chintala, i L. Bottou, „Wasserstein GAN". arXiv, 6 grudzień 2017. doi: 10.48550/arXiv.1701.07875.

[55] D. Wang i Q. Liu, „Learning to Draw Samples: With Application to Amortized MLE for Generative Adversarial Learning". arXiv, 25 listopad 2016. doi: 10.48550/arXiv.1611.01722.

[56] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, i A. Courville, „Improved Training of Wasserstein GANs", *arXiv:1704.00028 [cs, stat]*, grudz. 2017, Dostęp: 16 czerwiec 2021. [Online]. Dostępne na: http://arxiv.org/abs/1704.00028

[57] Z. Kurczyński, *Lotnicze i satelitarne obrazowanie Ziemi*, 1. wyd. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej, 2006.

[58] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, i Y. Fu, „Residual Dense Network for Image Super-Resolution". arXiv, 26 marzec 2018. doi: 10.48550/arXiv.1802.08797.

[59] T. Tong, G. Li, X. Liu, i Q. Gao, „Image Super-Resolution Using Dense Skip Connections", w *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, paź. 2017, s. 4809–4817. doi: 10.1109/ICCV.2017.514.

[60] K. Karwowska i D. Wierzbicki, „Modified ESRGAN with Uformer for Video Satellite Imagery Super-Resolution", *Remote Sensing*, t. 16, nr 11, Art. nr 11, sty. 2024, doi: 10.3390/rs16111926.

[61] W. H. Richardson, „Bayesian-Based Iterative Method of Image Restoration*", *J. Opt. Soc. Am., JOSA*, t. 62, nr 1, s. 55–59, sty. 1972, doi: 10.1364/JOSA.62.000055.

[62] F. Orieux, J.-F. Giovannelli, i T. Rodet, „Bayesian estimation of regularization and PSF parameters for Wiener-Hunt deconvolution", *J. Opt. Soc. Am. A*, t. 27, nr 7, s. 1593, lip. 2010, doi: 10.1364/JOSAA.27.001593.

[63] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, i H. Li, „Uformer: A General U-Shaped Transformer for Image Restoration". arXiv, 25 listopad 2021. doi: 10.48550/arXiv.2106.03106.

[64] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, i L. Zhang, „Satellite Video Super-Resolution via Multiscale Deformable Convolution Alignment and Temporal Grouping Projection", *IEEE Transactions on Geoscience and Remote Sensing*, t. 60, s. 1–19, 2022, doi: 10.1109/TGRS.2021.3107352.

ZAŁĄCZNIKI

Artykuły stanowiące cykl publikacji wraz z oświadczeniami współautorów o procentowym udziale w poszczególnej publikacji:

1. Karwowska K. (80%) and Wierzbicki D. (20%), 2022, Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3292-3312, DOI: 10.1109/JSTARS.2022.3167646

2. Karwowska K. (80%) and Wierzbicki D. (20%), 2022, Improving Spatial Resolution of Satellite Imagery Using Generative Adversarial Networks and Window Functions, *Remote Sensing*, 14(24):6285, DOI: 10.3390/rs14246285

3. Karwowska K. (80%) and Wierzbicki D. (20%), 2023, MCWESRGAN: Improving Enhanced Super-Resolution Generative Adversarial Network for Satellite Images, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 9459-9479, DOI: 10.1109/JSTARS.2023.3322642

4. Karwowska K. (80%) and Wierzbicki D. (20%), 2024, Modified ESRGAN with Uformer for Video Satellite Imagery Super-Resolution, *Remote Sensing*, *16(11):*1926. doi: 10.3390/rs16111926

Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review

Kinga Karwowska and Damian Wierzbicki

Abstract—In recent years, we have witnessed significant development in the space sector, in particular regarding Earth imaging. Small satellites, whose size and construction make their production much cheaper, are becoming increasingly popular. As a result, a larger number of satellites may be placed in space, and thus, they may perform more frequent observations of selected spots on Earth. Unfortunately, the construction of these satellites also affects their observation capacity as they have a weaker spatial resolution. Scientists have been dealing with the problem of improving the spatial resolution of satellite imaging for many years. Numerous methods were developed that allow for the best possible representation of high-resolution images based on low-resolution images. However, the application of traditional solutions to improve the resolution of digital images requires an additional high-resolution image. As far as images obtained by small satellites (e.g., nano, micro, or mini) are concerned, the difference between the spatial resolution of panchromatic and multispectral images is small (e.g., for SkySat-3 – SkySat-15 satellites, it is only 0.16 m). The need to increase the spatial resolution of an image that does not have a corresponding higher resolution image (e.g., a panchromatic image or a sequence of images) causes additional problems. This article presents a review of the methods to improve the spatial resolution of small-satellite imaging. The authors analyze the interpolation, pansharpening, and digital image processing methods. Additionally, the article focuses on presenting solutions based on deep learning that enables the enhancement of the spatial resolution of images obtained from small satellites. The methodology of creating databases used for network training is described. Finally, the authors present the main limitations of the analyzed solutions and future development trends that will enable to improve the spatial resolution with the use of a single image.

Index Terms—Convolutional neural networks, deep learning, neural networks, single image super-resolution (SISR), super-resolution.

I. INTRODUCTION

IN RECENT years, we have witnessed dynamic progress in technology, which, in turn, has contributed to a significant development of the space sector. According to the data provided by the Union of Concerned Scientists, as of the 1st of January 2021, over 3000 operational satellites were orbiting the Earth. Every sixth of them performs operational tasks [1]. Over the past five years, over 380 new imaging satellites have been placed in orbit (of which approx. 300 weighed less than 500 kg). This number is more than $4.5\times$ higher than in 2011–2015. High-resolution (HR) satellite imaging may be applied in numerous fields of research, such as land-use and landcover mapping [2], [3], urban mapping [4], detection and tracking of objects [5], [6], maritime monitoring [7], and automatic building classification [8], [9].

Contemporary satellites of very HR can record images with a spatial resolution that may reach, in nadir, even 0.30 m. However, nano-, micro-, and minisatellite systems are becoming increasingly popular [10]. Although they are both easier to manufacture and cheaper than large satellites, they have one major disadvantage (as far as the possibility of Earth observation is concerned): their resolution is significantly lower due to their construction. Their small size and weight (nanosatellites: 1–10 kg, microsatellites: 10–100 kg, and minisatellites: 100–500 kg) result in equipment limitations. The main element of an imaging satellite is its optic system, which, in the case of nano- or microsatellites, differs significantly from those installed in large observation satellites, such as WorldView-3 or QuickBird. Nano- and microsatellites are equipped with small matrices, e.g., complementary metal–oxide–semiconductor [11] or CMOSIS CMV [12], which are characterized by low quantum efficiency. Another limitation of small satellites is the inadequacy of the telescope caused by the much smaller number of applied lenses, which leads to a deteriorated quality of the obtained images, e.g., through blurring. The resolution obtained with the use of low-efficiency matrices and imperfect telescopes is significantly lower, even if low earth orbit satellites are employed. An example may be nSight-1, a nanosatellite designed and manufactured by the privately owned South African company SCS Space. The nanosatellite was placed in a 400 km orbit in 2017. It is equipped with the SCS Gecko imager that consists of the sensor unit (SU), control unit, optics, and mechanical support structure. The 2-Mega Pixel SU and Bayer filter allow it to take RGB images of a bandwidth of 64 km and a spatial resolution of 32 m [13] (see Fig. 1).

Scientists have been attempting to improve the spatial resolution of satellite imaging for many years. Numerous methods were developed that allow for the best possible representation of HR images based on low-resolution (LR) images. These methods may be divided into groups based on the number of images necessary to estimate an HR image. This allows us to distinguish between methods that use information obtained from several images and those that require only one image. Another classification is based on the operation of algorithms. These

Fig. 1.   Sample photo taken by nSight-1, GSD= 32 m [13].

TABLE I
SUMMARY OF THE NUMBER OF PUBLICATIONS ON THE ISSUE OF IMPROVING
THE RESOLUTION OF DIGITAL IMAGES (INCLUDING SATELLITE IMAGERY) WITH
THE USE OF SELECTED INTERPOLATION, PANSHARPENING, AND DEEP
LEARNING ALGORITHMS

METHOD		Scopus	Web of Science
INTERPOLATION	Nearest Neighbour	1395	1113
	Bilinear	2710	1911
	Cubic	7995	5052
PANSHARPENING	Intensity-hue-saturation (IHS)	135	141
	Principal component analysis (PCA)	128	138
	Brovey	73	53
	Ehlers	28	33
	Gramm-Schmidt	49	30
	Multiplicative	53	60
CNN	SRGAN	185	106
	SRResNet	21	13
	EDSR	63	107
	ESRGAN	46	35
	FSRCNN	43	20
	SRCNN	195	129
	SAN	204	330
	ESPCN	26	13
	EnhanceNet	8	7

methods are based on interpolation, pansharpening, and deep learning algorithms. The problem of the breadth of increasing digital image resolution has often been discussed in the literature. The most popular and oldest ones are the image interpolation methods (see Table I). Additionally, attention should be paid to the methods based on deep learning algorithms, which were described in a greater number of publications than the most popular and most frequently used (in the case of satellite imagery) pansharpening methods.

The least demanding group of solutions are interpolation methods. They only require one image and employ simple mathematical operations to calculate a new pixel value. Unfortunately, the simplicity of calculations significantly affects the interpretation possibilities of the generated images. Despite the increased spatial resolution, the improvement in interpolation possibilities is low. In the 1970s, space imagery became more common, which led to the necessity to improve satellite imagery resolution. Scientists saw the possibility of combining HR panchromatic (PAN) images with spectral information of multispectral images (MS) with a lower resolution very quickly. This method was called pansharpening. Over the years, many solutions based on this dependence have been developed. The great possibilities of these methods are best evidenced by the fact that they are widely used today. On the other hand, their main limitation is the need to have both HR images and the corresponding LR images.

However, as far as images obtained by small satellites (e.g., nano, micro, or mini) are concerned, the difference between the spatial resolution of PAN and multispectral images is small (e.g., for SkySat-3 – SkySat-15 satellites, it is only 0.16 m). The need to increase the spatial resolution of an image when a corresponding higher resolution image (e.g., a PAN image or a sequence of images) is unavailable causes additional problems. Research on improving spatial resolution based on a single image has been conducted for many years, initially without any noteworthy effects. Finally, a breakthrough was driven by the increased computing power of workstations that enabled deep neural networks to process digital images. They allowed for automatic classification of images [14]–[16], detection and tracking of objects [17]–[20], detection of changes [21]–[23], and segmentation [24]–[28]. They were also used to create algorithms to increase the resolution, e.g., SRCNN [29]. Taking the values of the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) metrics into consideration, it may be concluded that the use of convolutional neural networks may help improve the resolution of digital images significantly. Another milestone in processing digital images was generative adversarial nets (GAN) [30]—as described in the presentation by Goodfellow. The characteristic feature of this solution is network training. During this training, the generator's results (convolutional network) are assessed by another convolutional network (discriminator). This solution provides many possibilities, including generating new (false images) [30] or performing image translation [32]–[34]. Having these results in mind, scientists used GAN to develop new methods to improve the resolution of digital images based on a single LR image. At the same time, considering the construction of satellite images, which may be treated as classic digital images, all solutions using GAN may improve the spatial resolution of satellite images because it is the structure of the image that is important, not what is shown in it. In 2016, the team led by Ledig et al. [35] in the publication photorealistic single image super-resolution (SISR) using a generative adversarial network presented the super-resolution generative adversarial networks (SRGAN)

model—the super-resolution using a generative adversarial network that improves the spatial resolution of images, which received great acclaim. Several other authors introduced various modifications to this solution, including enhanced super-resolution generative adversarial networks (ESRGAN) [36]. However, they do not fully solve the problem of improving the resolution of satellite imaging.

This article reviews the methods to improve spatial resolution, starting from commonly used solutions based on interpolation. Then, the pansharpening methods are discussed, which are some of the most widely used solutions. The subsequent sections discuss digital image processing methods that are used, among others, in the SPOT and Pleiades satellites. Finally, the authors focus on solutions based on deep learning, which, as research has demonstrated, offers a variety of new possibilities. They will improve the spatial resolution of images obtained from nano- and microsatellite systems.

The article provides the answers to the following research questions.

1) What are the main problems related to improving the resolution of satellite imaging?
2) How are resolution improvement algorithms developed?
3) What is the difference between resolution improvement methods of small-satellite imaging?
4) What methodology should be applied to build training and validation data to improve the resolution of imaging from mini and nanosatellites using deep learning algorithms?
5) What methods should be utilized to assess the correctness of the operation of the algorithms that are responsible for the enhancement of spatial resolution?

To evaluate individual methods, quantitative analyses were performed in terms of spectral and spatial quality using traditional quality assessment metrics and visual analysis.

The rest of this article is organized as follows. Section II contains a review of the methods to improve the spatial resolution of satellite imaging and a description of the metrics used to assess the correctness of the operation of the analyzed algorithms. Section III includes results and discussion. Finally, Section IV concludes this article.

II. RELATED WORKS

The issue of improving the resolution of images has been known for many years, and the dynamic development of the space sector revealed the urgency of this problem in the context of improving the resolution of satellite imaging. This section presents a review of the methods to improve spatial resolution, starting from the most traditional ones, i.e., those based on interpolation, to the dynamically developing SISR methods that employ deep learning.

A. Interpolation

The most popular methods that enable to assign values to newly created elements of the image matrix are image interpolation algorithms. Thanks to its simplicity and computation speed, interpolation is often used to solve the problem of improving image resolution. It is applied in the SISR methods that use only a single LR image to render a super-resolution image (SR). There

TABLE II
COMPARISON OF SAMPLE INTERPOLATION METHODS THAT ARE USED TO
ENHANCE THE SPATIAL RESOLUTION OF SATELLITE IMAGES

Method	Advantages	Disadvantages
Nearest Neighbor, Bilinear, cubic interpolation	- They do not require large computing power - Smooth, visible edges, other methods	- They do not improve the interpretation ability - Lack of clearly visible edges – Nearest Neighbor interpolation
Super-Resolution Reconstruction of Satellite Video Images Based on Interpolation Method [45]	- Visual enhancement of Satellite Video Images - Smooth, visible edges	- No significant improvement in interpretation ability
Integration of the bilinear and the bicubic interpolation methods [46]	- Significant visual enhancement of the image, smoothened edges - They do not require large computing power - Possibility of modifing the weight at which interpolation results are combined	- No significant improvement in interpretation ability
Splines interpolation [47]	- The method does not require large computing power. It is faster than traditional interpolation methods - Significant visual enhancement of the image, smoothened edges	- No significant improvement in interpretation ability

are numerous methods of interpolation that allow for creating a new super-resolution image with a higher resolution. Table II presents examples of some popular methods along with their advantages and disadvantages.

The objective of linear interpolation methods is to find u function that will meet the condition specified in the following and will then be used for sampling the input image $I_{n,m}^{LR}$ of the dimensions $n \times m$, being integers [37]

$$I_{n,m}^{LR} = u(m,n) \wedge m, n \in \mathbb{Z} \qquad (1)$$

where $I_{n,m}^{LR}$ is the input image with $n \times m$ size and u denotes the image resampling function.

The main methods include, among others, the nearest neighbor, bilinear, and bicubic interpolation. The most used one is the nearest neighbor method (also called pixel duplication), where the pixel of the resulting image takes the value of the pixel of the original image that is situated nearest to the analyzed point. The u function is presented in (2), and the kernel of the nearest neighbor interpolation is shown as follows:

$$u(x, y) = I_{[x],[y]}^{LR} \qquad (2)$$

$$K(x, y) = K_1(x) K_1(y), K_1(t)$$
$$= \begin{cases} 1 \text{ for } -0,5 \le t < 0,5 \\ 0 \text{ for others} \end{cases} \qquad (3)$$

where $I_{n,m}^{\text{LR}}$ is the input image with $n \times m$ size, u denotes the image resampling function, $K(x,y)$ is the kernel, $K_1(x)K_1(y)$ denotes the tensor product kernel $K(x,y)$, and t is the processed pixel.

A slightly more elaborate method is the bilinear interpolation, where the pixel of the resulting image takes the average value of four neighboring pixels of the original image. Formula (4) presents u function of the bilinear method, where $\langle x \rangle$ means the fraction part equal to $x - [x]$, while $[\cdot]$ implies the function of rounding real numbers down to integers. Finally, (5) presents the interpolation kernel, where $(\cdot)+$ means the positive part

$$
\begin{aligned}
u\ (x,\ y) =\ & (1 - \langle x \rangle)(1 - \langle y \rangle)\, I_{[x],[y]}^{\text{LR}} \\
& + x - (1 - \langle y \rangle)\, I_{[x]+1,[y]}^{\text{LR}} \\
& + y\,(1 - \langle x \rangle)\, I_{[x],[y]+1}^{\text{LR}} + \langle x \rangle \langle y \rangle\, I_{[x]+1,[y]+1}^{\text{LR}}
\end{aligned}
\tag{4}
$$

$$
K\,(x,y) = K_1\,(x)\,K_1\,(y)\,, \quad K_1\,(t) = (1 - |t|)^+ \tag{5}
$$

where $u(x,y)$ denotes the image resampling function, $\langle x \rangle$ denotes the fractional part equal to $x - [x]$, $[x]$ and $[y]$ denote the real numbers that were rounded down to integers function, $K(x,y)$ denotes the kernel, $K_1(x)K_1(y)$ denotes the tensor product kernel $K(x,y)$, t denotes the processed pixel, and $(.)^+$ denotes the positive part of a real or extended real-valued function.

Bilinear Interpolation creates affine functions, and it is the simplest method that delivers a continuous function. Thanks to its efficiency and simplicity, it is commonly used in numerous algorithms.

Another method is a bicubic interpolation, where the value is calculated based on the values of all pixels that are adjacent to the analyzed pixel. The interpolation kernel formula (6) contains an additional parameter α, which generates an additional degree of freedom of the system

$$
K_1\,(t) = \begin{cases} (\alpha + 2)\,|t|^3 - (\alpha + 3)\,|t|^2 + 1 \text{ if } \leq 1 \\ \alpha|t|^3 - 5\alpha|t|^2 + 8\alpha|t| - 4\alpha \text{ if } 1 < |t| < 2 \\ 0 \text{ dla pozostałych wartości} \end{cases} \tag{6}
$$

where α denotes the parameter creating an additional degree of freedom of the system, $K(x,y)$ is the kernel, and t denotes the processed pixel.

The methods presented above enable to assign a brightness value to the newly emerging pixels, but this increases only the size of the image instead of the amount of information that may be read. Interpolation methods are used by numerous algorithms that improve the spatial resolution of images. Apart from the interpolation stage, these algorithms also remove blurring and minimize noise, which allows changing the size of the matrix that represents the output image and improves its quality.

One of such methods is partial volume diffusion, developed by Salvado *et al.* [38]. It enables image interpolation while at the same time recovering the blurred edges. Another solution is the curvature-based technique that is based on bilinear interpolation [39], [40] or the new edge detection interpolation algorithm that uses the concept of self-similarity and provides satisfactory super-resolution quality. Still another solution is the projection onto convex sets (POCS) [41] that is used for interpolation in

the fusion process. In this method, pixel values of the SR image are determined as the points of intersection of closed convex sets. The most crucial property of POCS consists in the fact that the values of the newly generated image may be included in the model to enable finding the searched point. Another popular method is iteration backprojection (IBP). In this approach, the SR image is estimated by representing the difference between simulated LR images and observed LR images by blurring the image. The literature on remote sensing features the IBP algorithm, modified by Li *et al.* [42], where different blur kernels are applied for each sensor. Additional elements that allow enhancing the resolution of the estimated image with the use of the discussed methods are blurring and noise reduction. Gilman *et al.* [43] found that optimal kernels are similar for different images, so it is possible to use one kernel for multiple images, and the obtained results are satisfactory. The authors developed two image models that enable capturing the important properties of the image and using them to develop the optimum kernels. The introduction of optimal interpolators allowed combining linear and cubic interpolation. Certain other methods use more than one image to enhance resolution. An example of such a solution is the combination of interpolation and gradient methods. It consists in combining LR images with the target HR image. In this solution, the gradient method iterations are performed in a HR grid of spectral and spatial coordinates [44].

Authors of numerous studies have attempted to use these relations to enhance the spatial resolution of satellite imaging. An example is an approach that uses a series of LR images to generate a single HR image by using the information from moving pixels representing the same object [45]. Another solution consists in integrating the bilinear and the bicubic interpolation methods, where the results of the application of interpolation are combined, being assigned the relevant weights [46]. On the other hand, Malpica [47] presents splines interpolation. This method enhances the radiometric resolution of satellite images using less computational power than classic interpolation methods, such as the nearest neighbor, linear, and cubic interpolation, which makes this method faster, and its results are comparable to those of linear interpolation (see Fig. 2).

B. Pansharpening

The optic systems installed on Earth-imaging satellites enable them to take photographs of various spatial and spectral resolutions. PAN images are characterized by very HR that may reach even up to 0.31 m (civilian satellites). In contrast, multispectral images provide plenty of valuable information, thanks to imaging in multiple spectral bands, but their spatial resolution is approximately four times lower. Fortunately, a method exists that enables to combine HR PAN images with the HR of a multispectral (or hyperspectral) image [48]–[51]. It is referred to as pansharpening. This operation results in generating an HR multispectral image. The main pansharpening methods include intensity-hue-saturation (IHS) [52], [53], principal component analysis (PCA), Brovey color standardization, Ehlers fusion, image multiplication, and Gram–Schmidt orthogonalization, but there are also numerous algorithms [54]–[56] (see Table III).
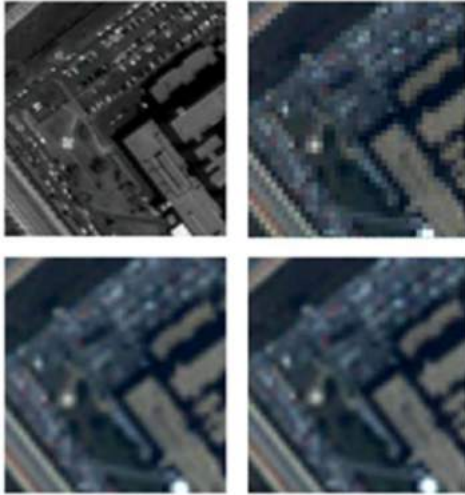
Fig. 2.   Detail of Ikonos image scene from Alcala University Campus. From top left, in clockwise direction, Pan image (GSD = 1 m), nearest-neighbour, splines, linear interpolation [47].

The IHS method uses three selected LR spectral bands to generate a color composition. It consists of four stages: The first one is the transformation of the selected MS bands (M_i) to the IHS space [(7) and (8)], where N is the number of selected bands [57]–[59]. Then the color composite image is sampled to the resolution of the PAN image. At this stage, each image band is also normalized to the [0, 1] range, and the histogram of the HR image is adjusted. The correctness of this operation is verified by observation of the mean and standard deviation values of the PAN image and the multispectral image (9). In the third phase, the intensity component is replaced with the PAN image (10). The final fourth stage is returning to the RGB color scheme [60]

$$I = \sum_{i=1}^{N} \alpha_i M_i \qquad (7)$$

$$\alpha_i = \frac{1}{N} \qquad (8)$$

$$P = \frac{\sigma I}{\sigma P} (P - \mu P) + \mu I \qquad (9)$$

$$F_i = M_i + (P - I) \qquad (10)$$

where N represents the number of channels, M_i represents the selected multispectral Image channels, F_i represents pan-sharpened image, P_H represents the histogram-matched high-spatial-resolution PAN data, P represents PAN data, I represents up-sampled intensity component, σ represents deviation, and μ represents mean.

However, some authors noticed that the resulting image contains several spectral distortions despite excellent efficiency and high spatial resolution [61]–[63]. These findings inspired numerous modifications to the algorithm. One of them is the adaptive IHS method. Rehmani *et al.* [64] noted that spectral distortions were smaller and the intensity component was more similar to the PAN image. Another solution to the problem of

TABLE III
COMPARISON OF SAMPLE INTERPOLATION METHODS THAT ARE USED TO
ENHANCE THE SPATIAL RESOLUTION OF SATELLITE IMAGES

Method	Advantages	Disadvantages
Intensity-hue-saturation	- Enhanced spatial resolution - Improved interpretation ability - Excellent performance	- The resolution is enhanced only in three selected LR bands - Presence of spectral distortions
Principal Component Analysis	- Any number of spectral ranges are used - Enhanced spatial resolution - Improved interpretation ability - Excellent performance	- Presence of spectral distortions
Brovey color standardization	- Enhanced spatial resolution - Improved interpretation ability - Excellent performance - Limited spectral distortions in comparison to the other methods	- The resolution is enhanced only in three selected LR bands - Presence of spectral distortions
Ehlers fusion	- Enhanced spatial resolution - Improved interpretation ability	- The resolution is enhanced only in three selected LR bands - Presence of spectral distortions (transformation to IHS color space)
Image multiplication	- Slight color distortions	
Gram-Schmidt orthogonalisation	- Enhanced spatial resolution - Improved interpretation ability - Slight colour distortions - Excellent performance	

spectral distortions was the edge-adaptive IHS. In this method, first, the edges are detected in the PAN image, and then the IHS method is applied to the location of the edges. The other pixels of the newly generated image (F) (those that do not belong to edges) are represented by the multispectral image (formula 11), where $h(x)$ is the edge detection function that takes the value of 1 if an edge occurs, and 0 in other cases

$$F_i = M_i + h(x)(P - I) \qquad (11)$$

where F_i denotes the pan-sharpened image, P_H denotes the histogram-matched high-spatial-resolution PAN data, M_i denotes the selected multispectral image channels, and $h(x)$ denotes the edge detection function.

Another method that enables the reduction of spectral distortions is the use of discrete ripple transform and compressed sensing [65]. In this solution, the transform is implemented for the intensity component and the PAN image to obtain large-scale subimages. The resulting HR subimages are subjected to local variance, and compressed sensing is applied to LR subimages to reconstruct the intensity component. This enables the integration of local information from the intensity component and the PAN image. The final stage consists in applying inverse ripplet transform and reverse IHS transform.

A similar principle is employed in the Brovey color standardization method that uses any three MS bands. In this method, the selected bands are normalized by dividing these values by the sum of the triplet of bands. The next stage consists in multiplying the obtained results by the PAN image containing spatial information, which may be presented in the form of the following equation, where i denotes the three selected bands [66]–[68]:

$$\mathrm{DN}\ (i) = \left[\frac{\mathrm{DN}\ (i)}{\sum \mathrm{DN}\ (i)} \right] \cdot \mathrm{DN}_{\mathrm{pan}} \qquad (12)$$

where $\mathrm{DN}(i)$ represents the digital number, I–3 are the selected channels, and $\mathrm{DN}_{\mathrm{pan}}$ is the digital number of PAN images.

This method may also be modified in certain ways. One of them is the combination of the Brovey method with high-pass filters, such as the Laplacian filter [68], [69]. This operation enhances the contract between the edges that are present in the image and the background, which significantly improves the visibility of details.

Another solution is the PCA pansharpening method. It uses any number of spectral ranges, transforming generally correlated bands into a new set of noncorrelated image data. The transformation matrix (V) (13) consists of the eigenvectors of the matrix of covariance or the matrix of correlation of the dataset for the nonstandardized PCA and normalized PCA, respectively. The new HR image is generated with the use of the three first spectral ranges of the PCA, and the PC1 component (which mainly contains spatial information) is replaced with the PAN image (PAN^s) after the histogram is adjusted to PC1. The final stage consists in performing a reverse transformation of the PCA, which creates a new HR multispectral image (F) [70]

$$\begin{bmatrix} F^1 \\ F^2 \\ \dots \\ F^n \end{bmatrix} = \begin{bmatrix} v^{11} & v^{12} & \dots & v^{1n} \\ v^{21} & v^{22} & \dots & v^{2n} \\ \dots & \dots & \dots & \dots \\ v^{n1} & v^{n2} & \dots & v^{nn} \end{bmatrix} \begin{bmatrix} \mathrm{PAN}^s \\ \mathrm{PC}^2 \\ \dots \\ \mathrm{PC}^n \end{bmatrix} = V \begin{bmatrix} \mathrm{PAN}^s \\ \mathrm{PC}^2 \\ \dots \\ \mathrm{PC}^n \end{bmatrix} \qquad (13)$$

where PAN denotes the panchromatic image, PC^n denotes the principal component, F^n denotes n component of the F image, where $F = [F^1, F^2, \dots, F^n]^T$, and V denotes the transformation matrix.

The PCA pansharpening method may be modified in several ways [71]–[73]. One of them is the additive wavelet principal component (AWPC). In this method, the PAN image's histogram is adjusted to the histogram of the first principal component PC1, and then the "à trous" algorithm is applied to distinguish wavelet planes. The last stage before applying the reverse transformation
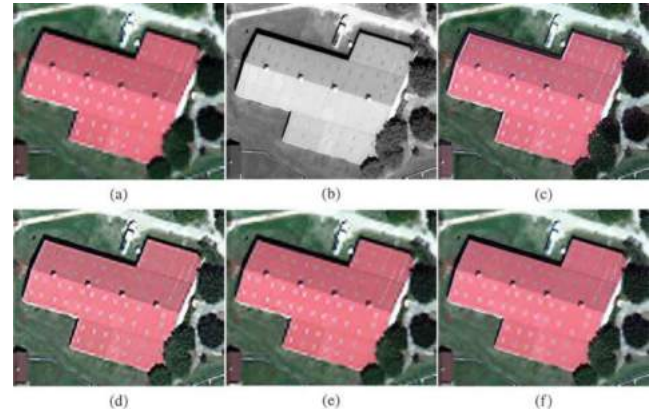
Fig. 3. Subsets from the study area in true color combination of (a) LMS image ($\mathrm{GSD}_{\mathrm{MS}} = 1.24$ m), (b) PAN image ($\mathrm{GSD}_{\mathrm{PAN}} = 0.31$ m), (c) PCA fused image, (d) AWPC fused image, (e) IPCA fused image (one iteration), and (f) IPCA fused image (four iterations) [75].

of the PCA consists in adding wavelet planes to the principal component [74]. This method was improved by Ghadjati *et al.* In their publication; these authors presented an approach where the spatial information content obtained from the PAN image is automatically adjusted until the highest quality is obtained. This dependence is presented in the following equation, where the HPF is the high-pass filtration of the PAN image (see Fig. 3) [75]:

$$\begin{bmatrix} F_0^1 \\ F_0^2 \\ \dots \\ F_0^n \end{bmatrix} = V \begin{bmatrix} \mathrm{PC}_0^1 + \mathrm{HPF}\ [\mathrm{PAN}] \\ \mathrm{PC}_0^2 \\ \dots \\ \mathrm{PC}_0^n \end{bmatrix} \qquad (14)$$

where F_0^n represents the n component of the F image, where $F = [F_0^1, F_0^2, \dots, F_0^n]^T$, V is the transformation matrix, PC^n is the principal component, PAN is the panchromatic image, and HPF is the high-pass filtering of a PAN image.

Another method that uses the selected bands and the transformation to the IHS color space is the Ehler method. In this solution, the PAN image and the intensity component (I) are transformed to the frequency level with the fast Fourier transform. After that, details of the image are separated from the PAN image with the use of high-pass filtering, and component I is subjected to low-pass filtering in order to blur the image. Images prepared in this way are adjusted to the original histogram based on intensity and transformed into an HR RGB image with the use of the reverse Fourier transform. This method utilizes several iterations for the combinations of all available bands [76], [77]. The application of this method was discussed in numerous publications on the enhancement of the spatial resolution of multispectral satellite images [78], [79], hyperspectral images (HSI) [80], and radar images [81].

The method with the smallest degree of color distortion, as a result of the presence of the intensity component, is the multiplicative method. The new pixel value is the product of the multiplication of the DN value of the PAN image pixel and the pixel of each of the multispectral image bands [82]. This solution is frequently used [78], [83]–[86].

TABLE IV
QUANTITATIVE EVALUATION WITH REFERENCE IMAGES: QUALITY METRICS RESULTS FOR DIFFERENT PANSHARPENING METHODS USING DIFFERENT SATELLITE IMAGES (MEAN VALUES)

Test Images	Metrics Method	CC	ERGAS	RASE	RMSE	SAM	SID	SCC	UIQI	SSIM
Deimos-2 images	GIHS [62]	0.0285	5.6958	22.5444	30.1408	1.7161	0.0311	0.9916	0.9280	0.8886
	AIHS [92]	0.0043	5.7104	22.4819	29.7967	2.1875	0.0443	0.9903	0.9315	0.8961
	AWLP [92]	0.0213	4.3249	17.4812	23.4077	2.33E-7	0.0013	0.9509	0.9627	0.9581
	PCA [93]	0.2683	7.7880	31.1688	41.2881	4.0265	0.0515	0.9471	0.8141	0.8007
	PRACS [94]	0.0204	4.2847	17.4401	23.1224	1.2166	0.0677	0.9688	0.9643	0.9483
	FRSQ [95]	0.0317	4.7057	18.9366	25.3195	1.1819	0.0620	0.8451	0.9560	0.9128
	Wady et al. Method [61]	0.0132	3.4207	13.5333	18.0741	0.7302	0.0082	0.9832	0.9755	0.9763
WV-2 images	GIHS	0.0089	4.5698	17.2795	50.7240	1.4450	0.0076	0.9910	0.9555	0.8103
	AIHS	0.0010	4.3303	16.7635	49.1225	1.2774	0.0133	0.9896	0.9635	0.8301
	AWLP	0.0086	4.5449	18.2989	53.7066	2.43E-7	0.0004	0.9481	0.9579	0.8338
	PCA	0.1783	8.6275	32.9548	97.1180	4.6856	0.0329	0.9168	0.7952	0.7236
	PRACS	0.0056	4.4537	18.1805	53.2766	1.0780	0.0032	0.9820	0.9638	0.8391
	FRSQ	0.0119	5.5075	21.1254	61.8930	1.4337	0.0692	0.8298	0.9495	0.7992
	Wady et al. Method [61]	0.0055	3.7033	14.0013	41.0648	0.8565	0.0110	0.9874	0.9717	0.8579
WV-3 images	GIHS	0.0406	5.7943	18.8313	60.6717	2.2267	0.0512	0.9818	0.9309	0.5300
	AIHS	0.0111	6.0724	19.7170	63.4380	2.0189	0.0478	0.9849	0.9246	0.5421
	AWLP	0.0514	4.4063	17.0148	54.9083	0.0306	0.0721	0.9373	0.9534	0.7377
	PCA	0.3438	7.4928	26.6934	85.4653	3.4903	0.0487	0.9416	0.8712	0.4411
	PRACS	0.0171	3.7532	12.8878	41.4724	1.3918	0.0016	0.9293	0.9721	0.7387
	FRSQ	0.0151	5.1888	15.1875	48.9939	2.3728	0.1263	0.8228	0.9531	0.6688
	Wady et al. Method [61]	0.0089	3.6405	11.8389	38.0540	1.2723	0.0397	0.9676	0.9728	0.7017

The best values of the analyzed metrics are highlighted in green and the worst ones in red [61].

Gram–Schmidt orthogonalization is another method that replaces the intensity component with the PAN band. In this solution, the algorithm takes nonorthogonal vectors and then rotates them so that they become orthogonal. For images, each band is treated like a multidimensional vector [77], [87]. This method is very popular and brings very good results, which was proven by a number of publications that emphasized the high quality of the obtained images [63], [88]–[91].

Table IV presents an example of testing the correctness of the method proposed by the authors, and Fig. 4 shows a visual comparison of the application of several methods to an exemplary image. Based on the metrics, it can be concluded that the method proposed by the Wada team [61] gives the best results, which is confirmed by visual assessment [see Fig. 4(i)]. For this method, the relative dimensionless global error in synthesis (ERGAS), relative average spectral error (RASE), and root-mean-squared error (RMSE) metrics take the lowest values among the compared methods, while universal image quality index (UIQI) and SSIM take the highest values (close to +1). The AWLP method [92] presents slightly weaker values of the assessment metrics, although they do not differ significantly from the winners of this comparison in the visual assessment. The values of the examined metrics confirm the small differences: ERGAS, RASE, and RMSE are much lower than other compared methods, whereas

the value of spectral angle mapper (SAM) takes a value close to zero, which means the high quality of mapping of the spectral components. Due to the highest spectral distortions, the worst results were achieved by the PCA method, which is confirmed by the tested metrics, e.g., very high RMSE value (for WV-2 images RMSE = 97.1180) or low SSIM value (for WV-3 images SSIM = 0.44) [93]. In Table IV, the best values of the analyzed metrics are highlighted in green, while the worst ones are in red.

C. Other Digital Image Processing Methods to Improve Image Resolution

Other solutions are methods that utilize numerous LR images in order to estimate an HR image. One of them is the algorithm that enhances the resolution of the PAN SPOT5 scene, a quincunx sampling mode named Très Haute Résolution (THR), proposed by the CNES company. This innovative model of acquiring and processing images allows obtaining PAN images with a resolution of approx. 2.5 m is based on two images of the resolution of 5 m that are taken simultaneously. This is enabled by the fact that the images are obtained with the use of two CCD detector matrices that are shifted in the focal plane so that they meet the Shannon requirement at first order [96] (see Fig. 5). This means that the acquisition system is well optimized, thanks to
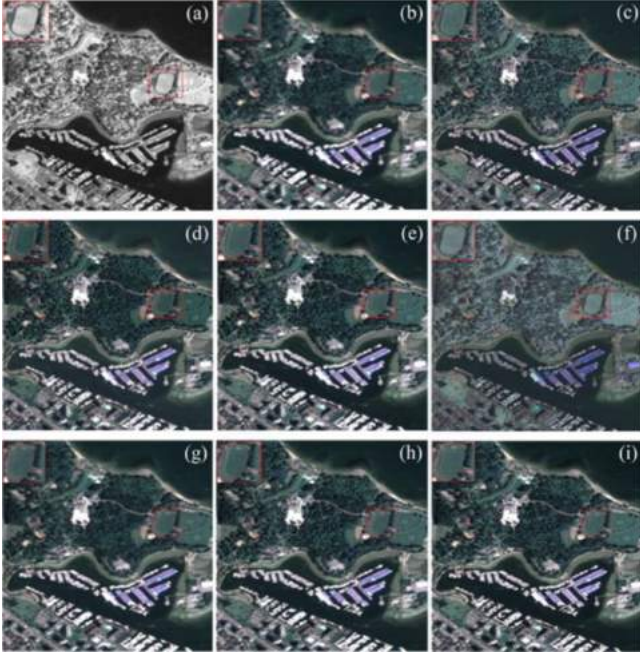
Fig. 4. Fusion results of the second Deimos-2 image. (a) PAN image (GSD_{PAN} = 1 m). (b) Up-sampled MS image (GSD_{MS} = 4 m). (c) GIHS. (d) AIHS. (e) AWLP. (f) PCA. (g) PRACS. (h) FPSQ. (i) Wady *et al*. Method [61].
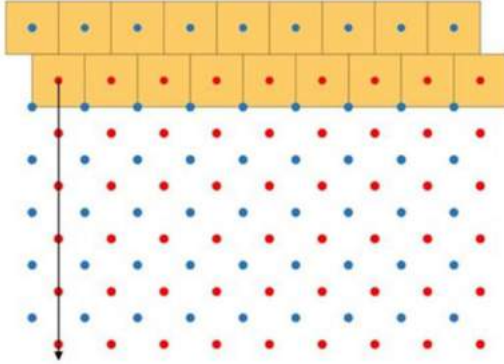
Fig. 5. SPOT5 THR quincunx sampling grid. A double CCD linear array generates two classical square sampling grids, shifted by half a sampling interval in both row and column directions. Interleaving the two shifted grids yields a quincunx sampling [98].

the correct match of the sampling and the modulation transfer function (MTF).

The THR process consists of three stages: interpolation, deconvolution, and noise reduction. In the first stage, the obtained LR images are interleaved, creating an image of a radiometric resolution of 2.5 m. However, only half of the pixels originate from the images obtained by the sensor ($\frac{2 \cdot x \cdot y}{2 \cdot x \cdot 2 \cdot y} = \frac{1}{2}$, where x and y denote the number of rows and columns of the image, respectively). As a result, the values of the newly created pixels are interpolated based on the values of the adjacent pixels. A characteristic blur is visible in the generated image, which is caused by low MTF values for high frequencies. Due to that, the second stage involves removing the blur by applying a filter that

represents the instrument's inverse transfer function. This operation results not only in removing the blur but also in enhancing high-frequency noise, which is then removed at the last stage of the algorithm's operation. This phase utilizes the fixed chosen noise restoration algorithm [97], whose task is to remove noise in a controlled way in the most sensitive places. This is enabled by a nonlinear reduction of the wavelet packet coefficients, employing joint-adaptive space and frequency wavelet packet decomposition. This means that the noise removal process consists of thresholding noisy wavelet coefficients in the image at different image restoration levels [98], [99]. The application of this algorithm enables to enhance the spatial resolution of the image, which, in turn, improves the level of detail, and, as a consequence, the interpretation ability of the image, which was ultimately reflected in numerous studies on monitoring the environment [100]–[102], detecting changes [103]–[105], and mapping [106]–[108].

Another solution to estimate the value of pixels of HR images based on LR ones is the super-resolution variable-pixel linear reconstruction (SRVPLR) algorithm. It is based on the VPLR algorithm (also referred to as the Drizzle algorithm) [109], which combines astronomical dithered undersampled images. In this algorithm, pixels in LR images are represented in the newly created HR image, taking into account the shift, rotation, and distortion of the imaging sensor. At the same time, the transferred pixel's size is decreased to limit the number of artifacts in the image. Such decreased pixel i_{xy} of the weight w_{xy} is then added to the pixel of image I_{xy} with the weight W_{xy} and a fraction, overlaying pixel $0 < a_{xy} < 1$. As a result, the new values of the image I'_{xy} and weight W'_{xy} equal

$$W'_{xy} = a_{xy}w_{xy} + W_{xy} \tag{15}$$

$$I'_{xy} = \frac{a_{xy}i_{xy}w_{xy} + I_{xy}W_{xy}}{W'_{xy}} \tag{16}$$

where x, y denote pixel position, i_{xy} denotes the DN value of the reduced pixel, i_{xy} denotes the reduced pixel weight value, W'_{xy} is the final pixel weight value, I_{xy} is the DN value of pixel before modification, I_{xy} is the DN value of pixel after modification, a_{xy} denotes the fractional pixel overlap parameter, where $a_{xy} \in (0, 1)$, and DN(i) is the digital number.

However, satellite images, as opposed to astronomical images that present stars, asteroids, nebulae, or galaxies, contain multiple elements that differ in terms of shape, structure, or texture. Due to that, the SRVPLR algorithms adjust the histogram of LR images, taking into account the selected values of pixels of the original LR image as reference values (i.e., their value should remain unchanged after changing the resolution) and using tools to mask the areas that are not subject to the resolution enhancement process.

Other authors have also used the multiframe super-resolution technique to enhance the resolution of electro-optic images [110] and radar images. Ito [111] proposed an algorithm to enhance the resolution of SAR images by performing local arithmetic operations [112] that take into account the orbit deviations of a satellite and temporal changes. The presented solution enables us to estimate a HR image based on n-frames of LR images.

TABLE V
COMPARISON OF SAMPLE DEEP LEARNING METHODS

	Method	Advantages	Disadvantages
CNN	Classic solutions	- improved interpretation ability of SR images - only a LR image is required to enhance resolution	- access to large databases is required, as they are the basis for network training - correct functioning only for specific image parameters - high computational capacity of the workstation is required, particularly at the network training stage
	Networks using residual connections	- improved interpretation ability of SR images - only a LR image is required to enhance resolution - solution for the problem of gradient fading and representation bottlenecks - accelerated network training process - improved interpretation ability of SR images	- access to large databases is required, as they are the basis for network training - correct functioning only for specific image parameters - high computational capacity of the workstation is required, particularly at the network training stage
	Multi-column generator	- the algorithm may accept images of various spatial resolutions Additionally, for solutions that utilize residual connections: - solution for the problem of gradient fading and representation bottlenecks - accelerated network training process - improved interpretation ability of SR images	- access to large databases is required, as they are the basis for network training - high computational capacity of the workstation is required, particularly at the network training stage
GAN	Classic solutions	- significant improvement in the interpretation ability of SR images compared to SR images obtained with use of algorithms that employ classic deep convolutional networks	- unstable training - modifications of network training parameters have a strong influence on the generated HR images - at the training stage, LR images are required, along with corresponding images of high (target) resolution - high computational capacity of the workstation is required, particularly at the network training stage
	Classic solutions	- significant improvement in the interpretation ability of SR images compared to SR images obtained with use of algorithms that employ classic deep convolutional networks - solution for the problem of gradient fading and representation bottlenecks - accelerated network training process	- unstable training - modifications of network training parameters have a strong influence on the generated HR images - at the training stage, LR images are required, along with corresponding images of high (target) resolution - high computational capacity of the workstation is required, particularly at the network training stage

As far as HSI are concerned, tensor-based methods are becoming increasingly popular. An example of the application of such a method is the coupled tensor factorization approach [113], which utilizes the multidimensional structure of the HSI and a multispectral image (MSI) based on the coupled tensor factorization framework. This solution enables solving the problem of the loss of structural information of the HSI and MSI data "cubes," and allows the identifiability of the super-resolution image (SRI) under realistic assumptions. Another solution combines the LR HSI with an HR multispectral image to obtain an HR HSI [114]. This method utilizes nonlocal similar patches to form a nonlocal patch tensor, which is modeled with the use of the method based on the tensor-tensor product (t-product). Another approach to solving this problem exploits ideas from the field of tensor completion to directly impose a low-rank property spatially and spectrally while avoiding the computationally complex patch clustering and dictionary learning common to competing for fusion techniques [115].

D. Deep Learning

Apart from the interpolation methods that utilize HR PAN images or digital image processing, there is also a group of methods that employ deep learning algorithms. These solutions recreate a HR image based on the knowledge about the links between LR images and the corresponding HR ones, which were obtained during model learning. Over the last decade, we have been witnessing a significant improvement in the efficiency of deep learning algorithms that enhance spatial resolution. Even more, they now enable a much more accurate recreation of HR images based on LR ones. Another advantage of this group of solutions over the pansharpening methods, which are a very popular means of enhancing satellite images' spatial resolution, is the fact that deep learning solutions do not require another corresponding HR image. As a result, the interpretation ability of digital images is improved, provided that the neural network has been prepared correctly.

The methods presented in the literature may be divided into two groups: solutions that utilize convolutional neural networks [116]–[121] and methods based on generative adversarial networks [122]–[128]. However, these solutions are used in a slightly different way than those presented above. It may be divided into two stages: training of the network on which the algorithm is based and using the algorithm (see Table V). At the same time, one should bear in mind that the method's capability depends on the training stage. Apart from the adequate architecture and network parameters, network training requires a database. The most common method of creating databases dedicated to improving spatial resolution is reducing the resolution of an HR image to the resolution of the LR image and then conducting training based on LR images that are compared to the actual HR images. As a result, any digital image or any database of digital images (e.g., ImageNet [129], LSUN [130], MC COCO [131]) may be used to train the network. Still, a database dedicated to the issue of improving resolution exists. It contains LR images and the corresponding HR images [132]. However, to use an existing database for network training, it should be selected so that the image parameters are similar to those the algorithm will work on in the future.

TABLE VI
COMPARISON OF THE PSNR [DB]/SSIM METRICS OF THE MOST POPULAR MODELS OF ENHANCING THE RESOLUTION OF DIGITAL IMAGES BASED ON DATABASES:
SET5-4x, SET14-4x, BSD100- 4x, URBAN100- 4x, FFHQ 256x256- 4x, FFHQ 512x512- 4x, AND FFHQ 1024x1024- 4x

Model	Set5-4x [139]	Set14-4x [140]	BSD100- 4x [141]	URBAN100- 4x [142]	FFHQ 256x256- 4x [143]	FFHQ 512x512- 4x [143]	FFHQ 1024x1024- 4x [143]
SRGAN [35]	29.40/0.8472	26.02/0.7397	25.16/0.6688	-	17.57/0.415	27.494/0.735	21.49/0.515
SRResNet [35]	32.05/0.9019	28.49/0.8184	27.58/0.7620		-	-	-
Bicubic [35]	28.43/0.8211	25.99/0.7486	25.94/0.6935	-	-	-	-
EDSR [144]	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/0.8033	22.47/0.706	30.188/0.824	28.34/0.827
ESRGAN [36]	32.73/0.9011	28.99/0.7917	27.85/0.7455	27.03/0.8153	15.43/0.267	27.134/0.741	19.84/0.353
FSRCNN [138]	30.70/0.8657	27.59/0.7535	26.98/0.7398 (BSD200)	-	-	-	-
SRCNN [29]	30.49/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221	23.12/0.688	-	27.40/0.801
SAN [145]	32.70/0.9013	29.05/0.7921	27.86/0.7457	27.23/0.8169			
HAN+ [146]	32.75/0.9016	28.99/0.7907	27.85/0.7454	27.02/0.8131	-	-	-
WSRGAN [147], [148]	28.24/0.8779	25.08/0.7396	-	-	-	-	-
ESPCN [149]	28.16/0.8001	24.76/0.6962	-	-	-	-	-
EnhanceNet [149]–[151]	31.74/0.8869	28.42/0.7774	27.50/0.7326	25.66/0.7703	23.64/0.701	-	29.42/0.832
SRFBN [152]	32.47/0.8983	28.81/0.7868	27.72/0.7409	26.60/0.8015	21.96/0.693	29.58/0.8270	27.90/0.822

For several years now, participants in many image processing and machine learning competitions have attempted to solve the problem of improving the resolution of digital images. One of them is the NTIRE Challenge organized as part of the Conference on Computer Vision and Pattern Recognition. The NTIRE Challenge deals with numerous issues related to digital image processing, including the improvement of digital image resolution [133]–[136] and video [137].

Most of the research on improving resolution is based on classic digital images (Set5-4x, Set14-4x, BSD100- 4x, URBAN100- 4x, FFHQ 256x256- 4x, FFHQ 512x512- 4x, FFHQ 1024x1024- 4x). These images are characterized by HR and, thus, a great level of detail. In the case of improving the resolution of vertical images acquired from the aerial level, especially for satellite images, the same objects are characterized by different features, and the image resolution is much lower. Therefore, to use the solutions proposed by scientists to improve the resolution of digital images (obtained mainly from the Earth's ceiling), it is necessary to train models whose weights have been estimated based on the images mentioned above (or conduct training from the beginning—which is more time-consuming and requires Big Data).

The application of deep learning neural networks to enhance the spatial resolution of satellite images was inspired by the results of works on improving the resolution of digital images. The authors used the PSNR and the SSIM to assess the correct operation of the applied algorithms (see Table VI). The first indicator carries information about the ratio between the maximum signal strength and the power of noise interfering with the signal, while the other one takes into account distortions of luminance, contrast, and structure. One of the most popular solutions is

the SRCNN, which consists of three parts: patch extraction and representation, nonlinear mapping, and reconstruction. Its modification that enables slightly improve the PSNR coefficient (by 0.21 dB for the Set5-4x dataset) is the FSRCNN, where the deconvolution layer was introduced at the end of the network, so the mapping is learned directly from the original LR image (without interpolation) to the HR one, and the mapping layer was modified by changing the dimensions of the element before and after mapping. Additionally, the size of filters was reduced while at the same time increasing the number of mapping layers (see Fig. 6).

Another solution is the ESPCNN, which is a convolutional neural network that allows playing a video recording in real-time, at the resolution of 1080 p, with the use of the K2 graphic processor. The property maps are separated in the LR space in the proposed model architecture. Additionally, the subpixel convolution layer was introduced to enable improved reconstruction of the HR image. An example of a solution that employs generative adversarial networks is WSRGAN, which was created based on the analysis of MSE and VGG losses of the SRGAN model. The authors modified the range of weight values from 0 to 1 and the sampling interval to 0.1, which improved the PSNR and SSIM values of the generated images. However, numerous solutions present higher values of the coefficients. An example is the EDSR model, where the redundant modules of convolutional neural networks (Conv + ReLU + Conv) were removed, and the model's size was increased (32 residual blocks for EDSR and 80 residual blocks for MDSR). Another solution is a second-order attention network (SAN) that focuses on improved learning of correlations between properties, which, as a consequence, results in a better representation of the image. The authors
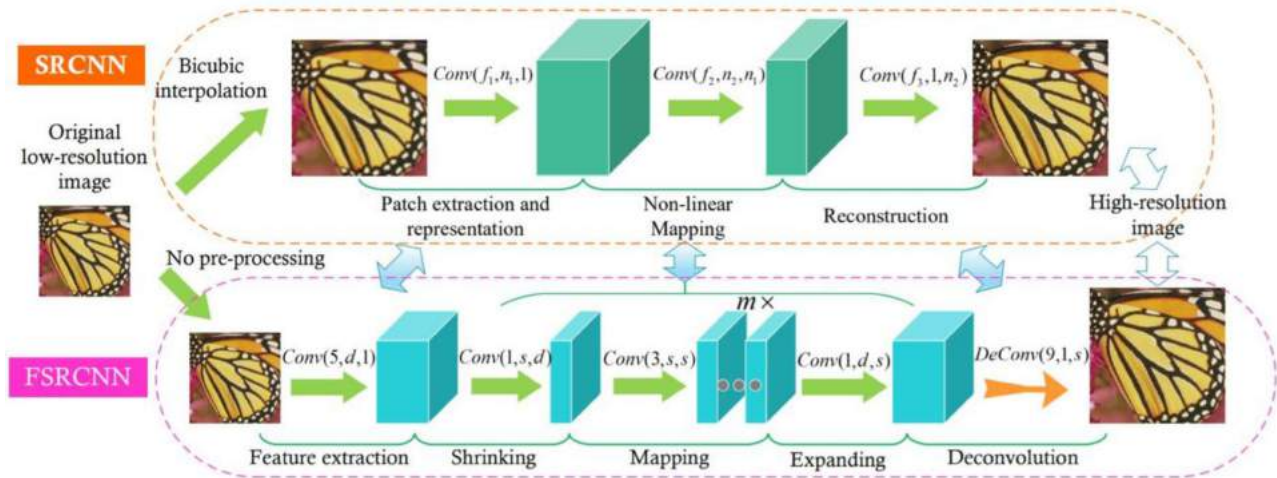
Fig. 6.   Comparison of the structures of SRCNN and FSRCNN [138].

have developed an innovative second-order channel attention module to adaptively rescale the channelwise features by using second-order feature statistics for more discriminative representations. The super-resolution feedback network (SRFBN) that utilizes feedback also brings very good results. On the other hand, the authors of the EnhanceNet model focused on the correct representation of the textures of generated images. They proposed the application of automated texture synthesis in combination with perceptual loss. The application of feed-forward fully convolutional neural networks in the discriminator model resulted in an improved quality of the generated SR images. The model that presents the best PSNR and SSIM results for the Set5-4x dataset is the holistic attention network (HAN+) which consists of a layer attention module and a channel-spatial attention module that enables holistic modeling of the correlations between layers, channels, and positions. The comparison of the PSNR and SSIM metrics of the most commonly used resolution enhancement methods revealed that the ESRGAN method [36] brought the best results for the Set-4x database, whereas the best solution for the Set14-4x, BSD100- 4x, and URBAN100-4x was SAN, developed by Dai *et al.* [145]. On the other hand, the worst results were obtained with the use of the ESPCNN algorithm [149]. What is interesting, this solution was characterized by poorer metrics than bicubic interpolation for the Set5-4x and Set14-4x databases (no data were available for the other analyzed databases).

1) Solutions Based on Classic Deep Neural Networks: An example of the application of deep neural networks is the VDSR network [153] which enables the enhancement of the spatial resolution of multispectral images. This solution utilizes blocks that consist of the convolutional layer and an activation unit (ReLU). At the last stage of operation of this model, the input image is added to the obtained results. This combination results in an SR image [154]. Several solutions that also use residual connections have been introduced in recent years.

An example is an algorithm used to enhance the resolution of space photographs presented by Feng *et al.* [155], whose task is to remove noise from the image and then enhance its

resolution. The proposed network architecture consists of nine residual blocks containing two convolutional layers separated by the activation unit.

Some of the other solutions presented in the literature are methods that utilize multicolumn networks. In these solutions, each branch is dedicated to one of the resolutions of the input image [156], [157]. As a result of the use of many branches, input data may take different dimensions, which allows adjusting the size of the image to the dimensions of the object whose resolution is to be enhanced. The authors of these publications assume that observing the given object in various spatial resolutions allows us to see more. This translates into practical applications, as is shown in Fig. 7. New SR images are characterized by a much better representation of the shape of the contours of objects in the image, which is additionally confirmed by the PSNR coefficient.

The methods presented in the previous sections allow for enhancing the resolution of images that consist of several bands. However, deep learning solutions also enable enhancing the resolution of HSI. This task is much more difficult than improving the resolution of multispectral images. The reason is the problem with obtaining HR images that are necessary at the training stage of convolutional neural networks.

However, numerous teams have attempted to solve this issue [158]–[163]. One of them was Li *et al.* [164]. The authors presented a new network consisting of two streams: spatial and spectral. The 1-D spectral convolution encodes minor changes in the spectrum, while the 2-D convolution, combined with the attention mechanism, encodes spatial information. Additionally, hierarchical lateral connections are used to enable combining spectral and spatial information. Moreover, attention blocks have been applied. They are implemented between blocks in the spatial network and enable considering the relations between nonlocal regions.

2) GAN-Based Solutions: The second group of solutions that utilize deep learning methods is those based on generative adversarial networks. The original GAN model consists of the generator G and discriminator D. The task of the G generator is to accept a random vector on the input and decode it in
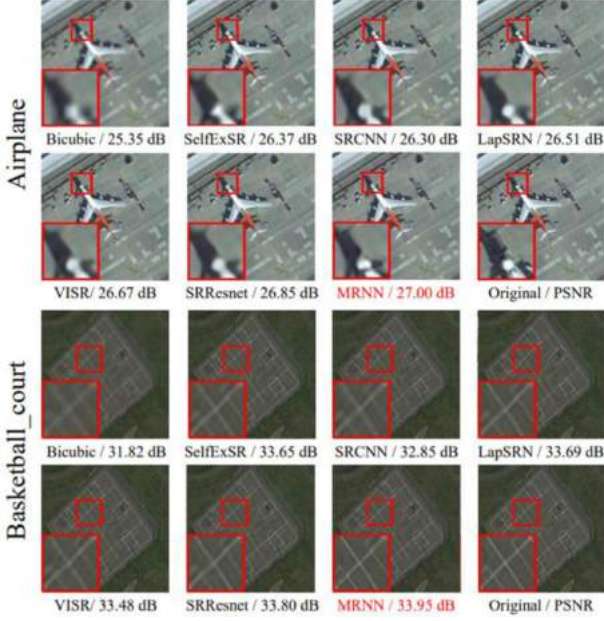
Fig. 7. Comparison of the results of the operation of the multicolumn network MRNN with Bicubic, SelfExSR, SRCNN, LapSRN, VISR, SRResnet, and the original image [157].

order to generate a synthetic image, while the discriminator is responsible for distinguishing between actual and decoded data. The generator network is trained so as to enable it to "cheat" discriminator D. Thus, with the progress in training, it will generate more and more realistic images, which enables, among others, to enhance their resolution. The original GAN model may be expressed as [165]

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log D(x)]$$
$$+ E_{z \sim p_{\text{data}}(z)} [\log (1 - D(G(z)))] \quad (17)$$

where x denotes the input image $x \sim p_{\text{data}}(x)$, z represents the random noise from a probability distribution of the *a priori* distribution $z \sim pz(z)$, and E is the empirical approximation of the expected probability. The above formula means that the process is iterative in a way that allows it to maximize the probability that the discriminator will correctly distinguish between real and synthetic images. At the same time, the generator learns how to minimize the probability of detection.

Based on this dependence, Ledig *et al.* [35] proposed the SRGAN algorithm (see Fig. 8). In this solution, LR images are generated based on HR ones. This is achieved by blurring the image with the use of the Gaussian filter, followed by decimation of the image at the sampling coefficient r. The training of the generator network is conducted in the form of feedback, while the weights and bias of the l-layer deep neural network may be calculated as

$$\theta_G = \{W_{1:L}; b_{1:L}\}$$
$$= \operatorname{argmin} \frac{1}{N} \sum_{n=1}^{N} l^{\text{SR}} \left(G_{\theta_G} \left(I_n^{\text{LR}} \right), I_n^{\text{HR}} \right) \quad (18)$$

where $\theta_G = \{W_{1:L}; b_{1:L}\}$ denotes the weights and biases of an L layer deep network and is obtained by optimizing an SR specific loss function l^{SR}, N denotes the number of HR images in the database I^{HR} and the corresponding LR images I^{LR}, and G_{θ_G} is the generator model (feed-forward CNN).

The network training takes place with the use of the method of maximizing the minimum gain, which is based on a theoretical game scenario, where images estimated by the generator compete with original HR images in the discriminator model. The training of generative adversarial networks utilizes a zero-sum game, where the reward of the generating network is determined (19). This solution motivates the discriminator to classify the received images (samples) correctly as true or false. At the same time, the generator is trying to cheat the adversary by saying that the provided samples are true

$$\min_{\theta_G} \max_{\theta_D} E_{I^{\text{HR}} \sim p_{\text{train}}(I^{\text{HR}})} \left[\log D_{\theta_D} \left(I^{\text{HR}} \right) \right]$$
$$+ E_{I^{\text{LR}} \sim p_G(I^{\text{LR}})} \left[\log \left(1 - D_{\theta_D} \left(G_{\theta_G} \left(I^{\text{HR}} \right) \right) \right) \right] \quad (19)$$

where E is the empirical approximation of the expected probability, D_{θ_D} is the discriminator network, I^{HR} is the HR image, G_{θ_G} is the generator network, and I^{LR} is the LR image.

The authors of SRGAN noted that for images with high PSNR, calculating the content loss based on the MSE of pixels results in the emergence of smooth textures (which is caused by the absence of high frequencies). As a result, VGG loss was applied based on the ReLU activation layers of the pretrained VGG19 network

$$l_{\text{VGG}/ \, i,j}^{\text{SR}} = \frac{1}{W_{i, \, j} H_{i,j}}$$
$$\times \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I^{\text{HR}})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{\text{LR}}))_{x,y} \right)^2. \quad (20)$$

In this formula, $\phi_{i,j}$ denotes a map of properties of the dimensions $W_{i,j}$ and $H_{i,j}$, obtained by the jth convolution (after activation) before the ith max-pooling layer of the VGG19 network. Then, the VGG loss is defined as the Euclidean distance between the representations of properties of the recreated image $G_{\theta_G}(I^{\text{LR}})$ and the reference image I^{HR}.

The SRGAN model inspired numerous modifications [166], [167], including HSI [168]. However, the most popular modification of this model is ESRGAN [36]. The batch normalization (BN) layers were removed from its generator (BN), and the basic block was replaced with a residual-in-residual dense block (RRDB), being a combination of a multilevel residual block and dense connections.

The removal of the BN layers resulted in stable training and improved network capacity (the time required for training became significantly shorter), which resulted from reduced computational complexity. The authors of ESRGAN also modified the discriminator by replacing it with a relativistic discriminator. As opposed to the standard discriminator used in SRGAN (21), which estimates the probability of whether the assessed image belongs to the set of HR images, the relativistic discriminator attempts to predict the probability that the real image I^{HR} is
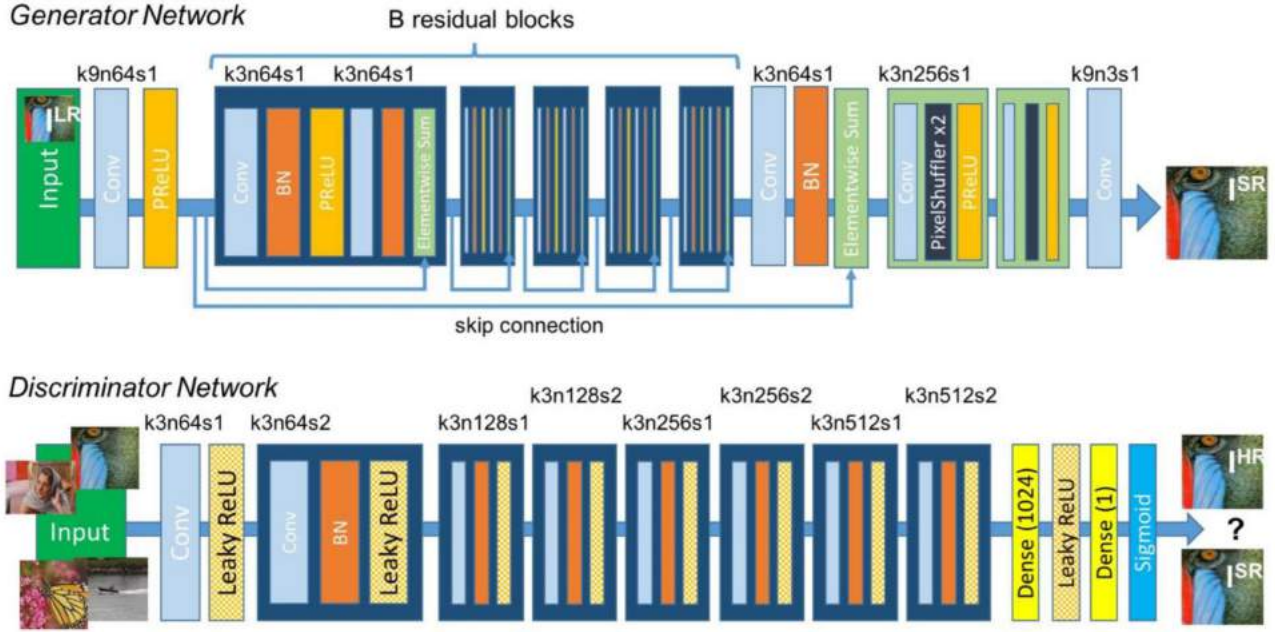
Fig. 8.   Architecture of the SRGAN network [35].

relatively more realistic than the fake image I^{SR} (22)

$$D\left(x\right) = \sigma\left(C\left(x\right)\right) \tag{21}$$

$$D_{Ra}\left(I^{\text{HR}},\, I^{\text{SR}}\right) = \sigma\left(C\left(I^{\text{HR}}\right) - E_{x_f}\left[C\left(I^{\text{SR}}\right)\right]\right) \tag{22}$$

where σ denotes the sigmoid function, $C(x)$ represents the output data of the generator before the application of the final activation function, and $E[\cdot]$ is the average of all fake data in the minibatch.

Another modification of the SRGAN model is the fact that perceptual loss is applied before the activation layers (not after them). This allows for increasing the number of features taken for calculating $l^{\text{SR}}_{\text{VGG}/\,i,j}$, which improves network efficiency. Additionally, it enables significantly better reconstruction of the brightness of the SR images. The ESRGAN network is trained based on network interpolation, which removes the noise on the estimated SR images. It consists in training the network G_{PSNR} focused on PSNR, and then the network is adjusted to obtain the G_{GAN} network. As it was in the previous case, this solution is also used to enhance the resolution of satellite images [169]–[171]. Two datasets, a satellite video sequence, were used to verify the possibilities and differences between the methods mentioned above. The first set of data contains video sequences collected by SkySat-1 on March 25, 2014 from Las Vegas in the USA, on April 9, 2014 from Burj Khalifa in Dubai, and on February 1, 2019 from Burj al-Arab in Dubai. The resolution was approximately 1.5 m, and the frame size was 1280×720 pixels at 30 frames per second [172].

The second dataset included video sequences from the Jilin-1 mission video collected on October 6, 2017 from Beirut in Lebanon and on September 10, 2018 from Florence in Italy. For the Beirut (Lebanon) set, the resolution was approximately 1.3 (or 1.12 m), and the frame size was 1920×1080 pixels at 25 frames per second with a video duration of 32 s. For

the Florence (Italy) set, the resolution was approximately 0.9 m, and the frame size was 3840×2160 pixels at ten frames per second with a video duration of 31 s [Chang Guang Satellite Technology, Company, Ltd.]. First, the authors used the pretrained generator G of ESRGAN, originally referenced on ImageNet and distributed as RRDB_ESRGAN_x4.pth. Then the networks (SRGAN and ESRGAN) were trained based on the above-described datasets, where the LR image was obtained by downsampling the HR image. Fig. 11 shows examples of improvement of spatial resolution of video frame fragments obtained by Jilin-1 Smart Video Satellite using SRGAN and ESRGAN.

Based on the visual analysis, it may be noticed that the introduced modification, i.e., the ESRGAN network, reproduces the structure and texture of the image much better. Moreover, by sharpening small objects, the image interpretation possibilities increase.

The application of residual connections allowed for a significant improvement of results while at the same time shortening the time required for network training [156], [157], [164], [169], [173]–[176] (see Figs. 9 and 10). This solution is used in numerous algorithms. One of them is the solution proposed by Courtrai *et al.* [177], who developed a method to improve the spatial resolution of small objects in both aerial and satellite images (see Fig. 11). The generator model utilizes residual connections and integration with the cycle model. Additionally, the authors used the Wasserstein GAN version [178] with the addition of a gradient penalty, which is the last component of the following loss function of the discriminator (or critic). Such a structure of the model resulted in significant improvement.

The possibilities and reliability of images recreated using deep learning methods are proven by the fact that they are utilized for the improvement of the resolution of medical imaging [179]–[181]. An example is a solution proposed by Zamzmi
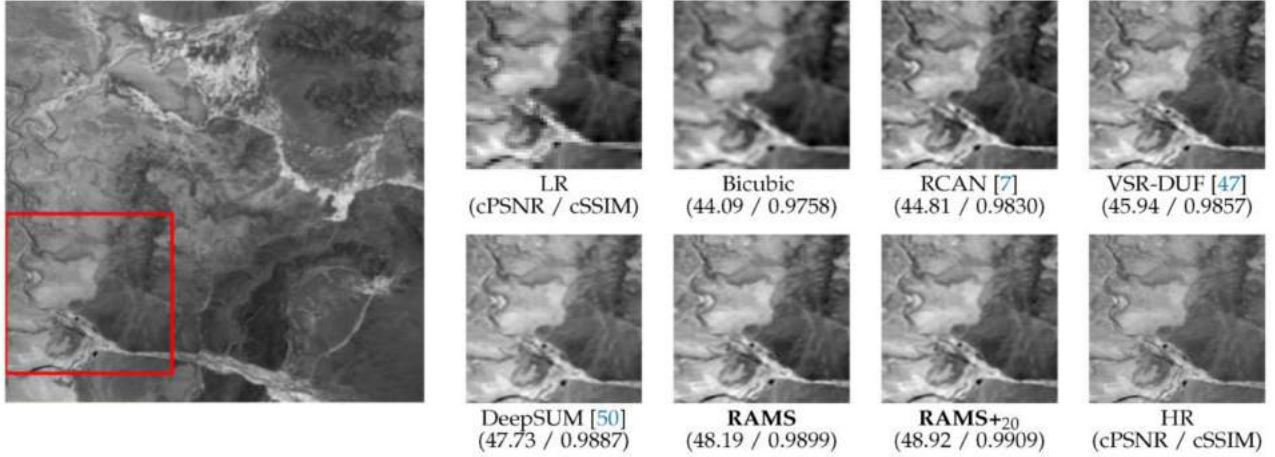
Fig. 9. Visual effects of the application of residual connections on the improved interpretation ability of PAN satellite images (PROBA-V, GSD= 1 km) [174].
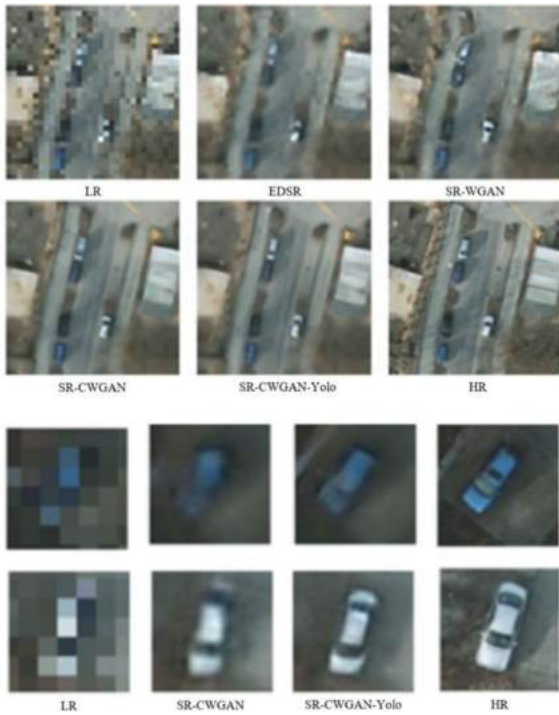
Fig. 10. Possibilities of enhancing the resolution of small objects. (a) Super-resolved images compared to LR (GSD = 1 m) and HR images (GSD = 12.5 cm). (b) Zoom on two vehicles [177].

et al. [182] that enables to enhance the resolution of X-ray images. The first stage of this algorithm consists in performing bicubic interpolation. Then the image is sent to the BN layer. The subsequent element of the architecture is a system of blocks consisting of three layers each (convolutional, zero padding, and activation (ReLU) layers). These blocks differ in terms of the number of convolutional layer filters. The image after bicubic interpolation is added to the resulting image at the last stage. The network analyses conducted by the authors demonstrated that it is characterized by much fewer training parameters than the VDSR [153] so that the training duration is significantly shortened with a simultaneous improvement of results.

As the above examples demonstrate, methods of enhancing the resolution of digital images with the use of deep neural networks are widely applied in issues related to the enhancement of satellite imaging resolution and for all digital images from space photography to medical imaging.

E. Evaluation Metrics

In order to assess the correctness of the operation of the methods, several metrics are used. One of them is the very popular RMSE, defined as

$$\text{RMSE} = \sqrt{\frac{1}{m \cdot n} \sum_{i=1}^{m} \sum_{j=1}^{n} |\text{MS}(i, j) - \text{PS}(i, j)|^2} \quad (23)$$

where MS and PS are original multispectral images and fused multispectral images, and $m \times n$ represents the size of the image.

RMSE error is the basis of RASE [183], which is used to evaluate the average performance of image fusion methods for each spectral band and is calculated using

$$\text{RASE} = \frac{100}{\mu} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \text{RMSE}^2(B_i)} \quad (24)$$

where μ is the mean, N is the number of bands, and B_i represents an ith band of the input MS image [184]. Also, ERGAS [185] determines the number of spectral distortions of the image

$$\text{ERGAS} = 100 \cdot \frac{h}{l} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\text{RMSE}(i)}{\mu(i)} \right)^2} \quad (25)$$

where $\frac{h}{l}$ is the ratio of pixel sizes of the input PAN and MS images, $\mu(i)$ is the mean of the ith band, while N is the number of bands [184].

Another solution for assessing the correctness of the method is the correlation coefficient (CC) [186], which determines the correlation measure between the PAN and multispectral (MS) images determined according to (26) shown at the bottom of this page. Additionally, modification of the spatial correlation
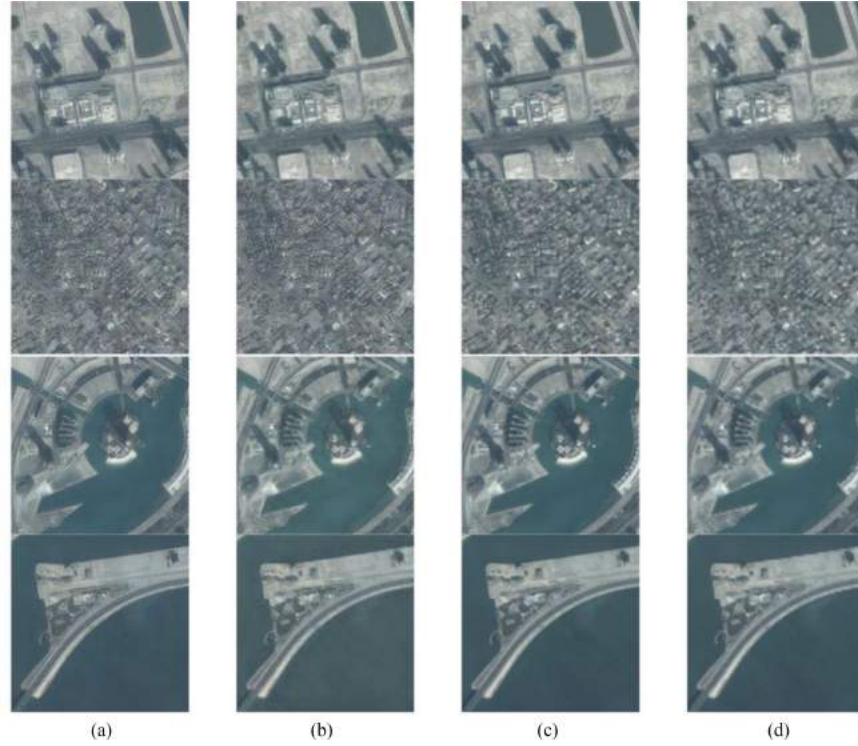
Fig. 11.   Comparison of the results of spatial resolution enhancement with the use of SRGAN and ESRGAN on fragments of a video frame captured by microsatellite (95 kg) Jilin-1 Smart Video Satellite. (a) ESRGAN. (b) SRGAN. (c) HR (GSD = 1.3 m). (d) LR (GSD = 5.2 m).

coefficient (SCC) metric [187] can be used, (26) shown at the bottom of this page where PAN represents the panchromatic image, MS the multispectral image, $\overline{MS}$ the mean value of the MS images, $\overline{PAN}$ the mean value of the PAN images, and n,m the image dimension [184].

Another example of a metric is the SAM [188], which defines the average change of all angles in the spectral component

$$\text{SAM}\,(v, w) = \cos^{-1}\left(\frac{\sum_{i=1}^{L} v_i w_i}{\sqrt{\sum_{i=1}^{L} v_i^2}\sqrt{\sum_{i=1}^{L} w_i^2}}\right) \quad (27)$$

where L represents the number of layers and v, w the test spectrum and reference spectrum (each has L components).

The spectral information divergence (SID) [189] method computes spectral similarity based on the divergence between the probability distributions of the two spectra as

$$\text{SID}\,(x, y) = \sum_{i=1}^{L} p_i\, \log\left(\frac{p_i}{q_i}\right) + \sum_{i=1}^{L} q_i \log\left(\frac{q_i}{p_i}\right) \quad (28)$$

$$q_i = \frac{r_i}{\sum_{i=1}^{N} r_i} \quad (29)$$

$$p_i = \frac{t_i}{\sum_{i=1}^{N} t_i} \quad (30)$$

where r represents the reference spectrum, t the test spectrum, and L the band number of multispectral images.

Another group of solutions for assessing the correctness of the resolution improvement methods are metrics that take into account three types of distortions: luminance, contrast, and image structure after pansharpening (y) in relation to the reference image (x): UIQI [190] (31) and SSIM (32). Their numerical values are determined based on the following formulas:

$$\text{UIQI}\,(x, y) = \frac{4\mu_x\mu_y\mu_{xy}}{\left(\mu_x^2 + \mu_y^2\right)\left(\sigma_x^2 + \sigma_y^2\right)} \quad (31)$$

$$\text{SSIM}\,(x, y) = \frac{(2\mu_x\mu_y + C_1)\,(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \quad (32)$$

where μ_x, μ_y, σ_x, σ_y, σ_{xy} are the local means, standard deviations, and cross covariance for images x and y, x is the reference image, and y is the pansharpening image.

Another metric to evaluate the processed image is PSNR. It defines the ratio of the maximum signal power (the maximum image value) to the power disturbing this signal, i.e., mean square

$$\text{CC}\,(\text{MS}, \text{PAN}) = = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\text{MS}_{ij} - \overline{\text{MS}}\right)\left(\text{PAN}_{ij} - \overline{\text{PAN}}\right)}{\sqrt{\left(\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\text{MS}_{ij} - \overline{\text{MS}}\right)^2\right)\left(\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\text{PAN}_{ij} - \overline{\text{PAN}}\right)^2\right)}} \quad (26)$$

error (33). PSNR values are expressed in decibels

$$PSNR = 10 \cdot \log_{10} \frac{[\max (HR(n,m))]^2}{MSE} \qquad (33)$$

where MSE represents the mean square error and max(HR(*n,m*)) represents the maximum reference image value with *n x m* size.

The evaluation metrics presented above are commonly used in remote sensing tasks. The application of those metrics enables to compare the correctness of the operation of various algorithms, which is given in Tables IV and V. Tasks related to the enhancement of the spatial resolution of digital images involve attempts to achieve the lowest possible values of the CC, ERGAS, RASE, RMSE, SAM, and SID errors and the highest possible values of the PSNR, SCC, UIQI, and SSIM metrics. As far as the last three methods of assessment are concerned, the values of errors approach 1 if the SR images are represented with high quality.

III. DISCUSSION

A. Review of the Methods

This article presents a review of the methods to enhance the spatial resolution of satellite images with the use of classic image sharpening methods, i.e., interpolation and pansharpening, and new solutions that employ deep learning: CNN and GAN. Until now, no such detailed review of the solutions for sharpening satellite images from the point of view of remote sensing has been performed.

Tables II and III present the advantages and disadvantages of classic solutions. The main limitation of the interpolation methods is the lack of a significant increase in interpretative abilities despite reducing the pixel size or the visual improvement of the image quality. At the same time, these solutions do not require the high computing power of working units and use only one image. On the other hand, the use of a PAN image to improve the resolution of multispectral images (MS) with lower resolution allows for a significant improvement in the interpretability of MS images, which provides a significant advantage over solutions that use interpolation. Unfortunately, this method does not prove effective in enhancing the resolution of images that depict moving objects. In such cases, SR images contain artifacts, which emerge as a consequence of the shift between the position of the moving object in the HR PAN image and its position in the multispectral image. This is one of the reasons why these algorithms cannot be applied to enhance the resolution of image sequences. Additionally, considering the methodology of operation of these solutions, they can only be used after improving the resolution of multispectral images that have their HR counterpart.

Table VI presents a summary and comparison of various neural network methods to enhance the resolution of digital images based on the same datasets. The comparison was based on the PSNR and SSIM quality indicators. The highest values of these indicators were obtained for ESRGAN [36] and SR-ResNet [35]. This method is characterized by significantly more stable training, higher efficiency, and improved reconstruction of details in the image. The obtained results confirm that this architecture may be applied to enhance the resolution of images from nanosatellite systems even up to four times. Besides, the ESRGAN method offers an excellent ability to represent detail, which is confirmed by a vast number of studies in the field of computer vision [36], [191].

In comparison to other methods, architectures based on GANs enable to obtain better results of resolution enhancement than classic methods. Due to the great potential and wide scope of applications of GAN, scientists are developing their applications in remote sensing analysis. However, the problems related to the application of GAN to sharpen whole satellite scenes, not only their fragments, still remain valid. Some other issues connected to the use of deep learning methods include the instability of data training or the vanishing gradient issue. Moreover, the conducted literature review allows us to state that GAN networks have a problem with nonconvergence and with the sensitivity to the selection of hyperparameters. Numerous methods, such as SRCNN (presented in Table VI), cannot be employed to enhance the spatial resolution of satellite images due to relatively poor representation of the texture, which may lead to misinterpretation of the images. The best architecture models were proposed by Wang [161] and Courtrai [177]. These methods use residual blocks, which is particularly important in the event of the application of generative adversarial networks, as they stabilize and significantly accelerate network training. Other important aspects that have been discussed in this literature review are the PSNR, SSIM, and other quality indicators that allow for an objective assessment of the usefulness of the given method for the sharpening of satellite images.

B. Future Possibilities of SISR Architecture

Currently used resolution enhancement methods that employ deep neural networks, including GAN, require very large datasets and high computational capacity of the equipment that is used to train the networks. This is caused by a very high number of layers, which, in turn, results in a large number of parameters that have to be adjusted. They are still based on large volumes of training data so that the trained models are universal. However, the model may become overtrained if the database is too small. In such an event, even if the model achieves good results during network training, it does not function correctly for other data. Although creating databases is not a problem for tasks that enhance the resolution of LR images generated based on HR images, it may become a major limitation in the enhancement of images from nanosatellite systems, where no corresponding HR images are available. Due to that, the SISR methods require further improvement, particularly with sharpening nanosatellite images. The success of deep neural networks is attributed precisely to the vast databases that enable training on powerful representations. Unfortunately, scientists are still unable to recreate the learning process, so deep neural networks are considered a "black box." Another challenge in the enhancement of the resolution of nanosatellite images is the need to define unambiguous evaluation metrics. Currently, the used MSE, SSIM, or PSNR metrics may often take excellent values, although the interpretation ability of the SR image does not improve, e.g., due to blurring. In addition, in publications, the results of image resolution improvement algorithms using

DL are presented in small images. Often, LR images consist of less than 10 000 pixels (e.g., Set5-4x, Set14-4x, BSD100-4x, URBAN100- 4x, FFHQ 256x256- 4x, FFHQ 512x512- 4x, FFHQ 1024x1024- 4x). This results, to a large extent, from the computational requirements of the algorithms that use GAN.

IV. Conclusion

This literature review presents state of the art. SISR has the potential to be applied to small satellite images. Apart from presenting the advantages and limitations of classic methods, the authors conducted an extensive review of contemporary super-resolution methods based on deep learning, particularly on generative adversarial networks. Deep learning methods show great potential in super-resolution satellite imagery. Despite their advanced sophisticated structure, some types of architecture based on convolutional networks cannot be successfully applied in SISR satellite imaging because they do not improve the interpretation abilities despite the enhanced resolution.

The authors have also demonstrated that numerous research projects so far have focused on enhancing the resolution of small images and the issues of applying SiSR methods in remote sensing, particularly in satellite imagery, to whole satellite scenes, whose dimensions are bigger than the dimension of images presented in publications. Furthermore, the authors presented future development directions with respect to sharpening whole scenes, optimizing parameters, and improving the architecture based on GAN to make them resistant to training instability.

References

[1] Satellite Database | Union of Concerned Scientists, Accessed: Sep. 25, 2022. [Online]. Available: https://www.ucsusa.org/resources/satellite-database

[2] D. Amarsaikhan, H. Blotevogel, J. van Genderen, M. Ganzorig, R. Gantuya, and B. Nergui, "Fusing high-resolution SAR and optical imagery for improved urban land cover study and classification," *Int. J. Image Data Fusion*, vol. 1, pp. 83–97, Mar. 2010, doi: 10.1080/19479830903562041.

[3] H. Luo, L. Li, H. Zhu, X. Kuai, Z. Zhang, and Y. Liu, "Land cover extraction from high resolution ZY-3 satellite imagery using ontology-based method," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 3, Mar. 2016, Art. no. 31, doi: 10.3390/ijgi5030031.

[4] P. Han *et al.*, "Monitoring trends in light pollution in China based on nighttime satellite imagery," *Remote Sens.*, vol. 6, no. 6, pp. 5541–5558, Jun. 2014, doi: 10.3390/rs6065541.

[5] W. Guo, W. Yang, H. Zhang, and G. Hua, "Geospatial object detection in high resolution satellite images based on multi-scale convolutional neural network," *Remote Sens.*, vol. 10, no. 1, Jan. 2018, Art. no. 131, doi: 10.3390/rs10010131.

[6] C. Witharana and H. J. Lynch, "An object-based image analysis approach for detecting penguin guano in very high spatial resolution satellite images," *Remote Sens.*, vol. 8, no. 5, May 2016, Art. no. 375, doi: 10.3390/rs8050375.

[7] H. Greidanus, "Satellite imaging for maritime surveillance of the european seas," in *Remote Sensing of the European Seas*, Berlin, Germany: Springer, 2008, pp. 343–358, doi: 10.1007/978-1-4020-6772-3_26.

[8] N. Gavankar and S. Ghosh, "Automatic building footprint extraction from high-resolution satellite image using mathematical morphology," *Eur. J. Remote Sens.*, vol. 51, pp. 182–193, Jan. 2018, doi: 10.1080/22797254.2017.1416676.

[9] J. Lee, H. Jang, J. Yang, and K. Yu, "Machine learning classification of buildings for map generalization," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 10, Oct. 2017, Art. no. 309, doi: 10.3390/ijgi6100309.

[10] G. W. Nagel, E. M. L. de M Novo, and M. Kampel, "Nanosatellites applied to optical earth observation: A review," *Rev. Ambiente Água*, vol. 15, pp. 1–19, Jun. 2020, doi: 10.4136/ambi-agua.2513.

[11] H. Kuuste *et al.*, "Imaging system for nanosatellite proximity operations," *Proc. Estonian Acad. Sci.*, vol. 63, no. 2S, 2014, Art. no. 250, doi: 10.3176/proc.2014.2S.06.

[12] J. Blommaert *et al.*, "CHIEM: A new compact camera for hyperspectral imaging," in *Proc. Small Satell. Conf.*, 2017, Paper 103.

[13] P. Mhangara, W. Mapurisa, and N. Mudau, "Image interpretability of nSight-1 nanosatellite imagery for remote sensing applications," *Aerospace*, vol. 7, Feb. 2020, Art. no. 19, doi: 10.3390/aerospace7020019.

[14] C. Yao, X. Luo, Y. Zhao, W. Zeng, and X. Chen, "A review on image classification of remote sensing using deep learning," in *Proc. 3rd IEEE Int. Conf. Comput. Commun.*, Dec. 2017, pp. 1947–1955, doi: 10.1109/CompComm.2017.8322878.

[15] J. Zhang, H. Zhao, and J. Li, "TRS: Transformers for remote sensing scene classification," *Remote Sens.*, vol. 13, no. 20, Jan. 2021, Art. no. 4143, doi: 10.3390/rs13204143.

[16] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of image classification algorithms based on convolutional neural networks," *Remote Sens.*, vol. 13, no. 22, Jan. 2021, Art. no. 4712, doi: 10.3390/rs13224712.

[17] G. Tochon, J. Chanussot, M. D. Mura, and A. L. Bertozzi, "Object tracking by hierarchical decomposition of hyperspectral video sequences: Application to chemical gas plume tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4567–4585, Aug. 2017, doi: 10.1109/TGRS.2017.2694159.

[18] D. Yu and S. Ji, "A new spatial-oriented object detection framework for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2021, doi: 10.1109/TGRS.2021.3127232.

[19] Z. Dong, M. Wang, Y. Wang, Y. Zhu, and Z. Zhang, "Object detection in high resolution remote sensing imagery based on convolutional neural networks with suitable object scale features," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2104–2114, Mar. 2020, doi: 10.1109/TGRS.2019.2953119.

[20] X. Li, J. Deng, and Y. Fang, "Few-shot object detection on remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 14–14, 2022, doi: 10.1109/TGRS.2021.3051383.

[21] Y. Li, C. Peng, Y. Chen, L. Jiao, L. Zhou, and R. Shang, "A deep learning method for change detection in synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5751–5763, Aug. 2019, doi: 10.1109/TGRS.2019.2901945.

[22] W. Zhao, L. Mou, J. Chen, Y. Bo, and W. J. Emery, "Incorporating metric learning and adversarial network for seasonal invariant change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2720–2731, Apr. 2020, doi: 10.1109/TGRS.2019.2953879.

[23] J. Zhang *et al.*, "Water body detection in high-resolution SAR images with cascaded fully-convolutional network and variable focal loss," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 316–332, Jan. 2021, doi: 10.1109/TGRS.2020.2999405.

[24] H. Liu, L. He, and J. Li, "Remote sensing image classification based on convolutional neural networks with two-fold sparse regularization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2017, pp. 992–995, doi: 10.1109/IGARSS.2017.8127121.

[25] Z. Li, X. Cui, L. Wang, H. Zhang, X. Zhu, and Y. Zhang, "Spectral and spatial global context attention for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 4, Jan. 2021, Art. no. 771, doi: 10.3390/rs13040771.

[26] J. M. Ramirez and H. Arguello, "Multiresolution compressive feature fusion for spectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9900–9911, Dec. 2019, doi: 10.1109/TGRS.2019.2930093.

[27] Z. Xue, J. Li, L. Cheng, and P. Du, "Spectral–spatial classification of hyperspectral data via morphological component analysis-based image separation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 70–84, Jan. 2015, doi: 10.1109/TGRS.2014.2318332.

[28] J.-Y. Rau, J.-P. Jhan, and Y.-C. Hsu, "Analysis of oblique aerial images for land cover and point cloud classification in an urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1304–1319, Mar. 2015, doi: 10.1109/TGRS.2014.2337658.

[29] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," Jul. 2015, Accessed: Jul. 15, 2021. [Online]. Available: http://arxiv.org/abs/1501.00092

[30] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, Paper 27, Accessed: Dec. 8, 2021. [Online]. Available: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

[31] B. Zhao, S. Zhang, C. Xu, Y. Sun, and C. Deng, "Deep fake geography? When geospatial data encounter artificial intelligence," *Cartogr. Geographic Inf. Sci.*, vol. 48, no. 4, pp. 338–352, Jul. 2021, doi: 10.1080/15230406.2021.1910075.

[32] Z. Tan, M. Gao, X. Li, and L. Jiang, "A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Jan. 2022, doi: 10.1109/TGRS.2021.3050551.

[33] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 14–34, Sep. 2021, doi: 10.1016/j.isprsjprs.2021.07.007.

[34] Y. Liu *et al.*, "CscGAN: Conditional scale-consistent generation network for multi-level remote sensing image to map translation," *Remote Sens.*, vol. 13, no. 10, Jan. 2021, Art. no. 1936, doi: 10.3390/rs13101936.

[35] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," May 2017, Accessed: Jun. 14, 2021. [Online]. Available: http://arxiv.org/abs/1609.04802

[36] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," Sep. 2018, Accessed: Jun. 15, 2021. [Online]. Available: http://arxiv.org/abs/1809.00219

[37] P. Getreuer, "Linear methods for image interpolation," *Image Process. Line*, vol. 1, pp. 238–259, Sep. 2011, doi: 10.5201/ipol.2011.g_lmii.

[38] O. Salvado, C. M. Hillenbrand, and D. L. Wilson, "Partial volume reduction by interpolation with reverse diffusion," *Int. J. Biomed. Imag.*, vol. 2006, 2006, Art. no. 92092, doi: 10.1155/IJBI/2006/92092.

[39] G. Chaudhary, K. Das, and M. Gopi, "Curvature minimizing depth interpolation for intuitive and interactive space curve sketching," [Online]. Available: https://www.researchgate.net/profile/MGopi/publication/228964270_Curvature_Minimizing_Depth_Interpolation_for_Intuitive_and_Interactive_Space_Curve_Sketching

[40] B. S. Morse and D. Schwartzwald, "Isophote-based interpolation," in *Proc. Int. Conf. Image Process.*, Oct. 1998, pp. 227–231, doi: 10.1109/ICIP.1998.999013.

[41] P. Oskoui-Fard and H. Stark, "Tomographic image reconstruction using the theory of convex projections," *IEEE Trans. Med. Imag.*, vol. 7, no. 1, pp. 45–58, Mar. 1988, doi: 10.1109/42.3928.

[42] F. Li, D. Fraser, and X. Jia, "Improved IBP for super-resolving remote sensing images," *Ann. Gis Geographic Inf. Sci.*, vol. 12, pp. 106–111, Dec. 2006, doi: 10.1080/10824000609480624.

[43] A. Gilman, D. G. Bailey, and S. R. Marsland, "Interpolation models for image super-resolution," in *Proc. 4th IEEE Int. Symp. Electron. Des. Test Appl.*, Jan. 2008, pp. 55–60, doi: 10.1109/DELTA.2008.104.

[44] A. Belov and A. Denisova, "Spatial interpolation methods for spectral-spatial remote sensing image super-resolution algorithm based on gradient descent approach," *J. Phys., Conf. Ser.*, vol. 1368, Nov. 2019, Art. no. 032006, doi: 10.1088/1742-6596/1368/3/032006.

[45] X. Qifang, Y. Guoqing, and L. Pin, "Super-resolution reconstruction of satellite video images based on interpolation method," *Procedia Comput. Sci.*, vol. 107, pp. 454–459, Jan. 2017, doi: 10.1016/j.procs.2017.03.089.

[46] W. Shi, Y. Tian, and K. Liu, "An integrated method for satellite image interpolation," *Int. J. Remote Sens.*, vol. 28, no. 6, pp. 1355–1371, Mar. 2007, doi: 10.1080/01431160600851876.

[47] J. A. Malpica, "Splines interpolation in high resolution satellite imagery," in *Advances in Visual Computing*. Berlin, Germany: Springer, 2005, pp. 562–570, doi: 10.1007/11595755_68.

[48] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017, doi: 10.1109/MGRS.2016.2637824.

[49] J. Marcello, E. Ibarrola-Ulzurrun, C. Gonzalo-Martín, J. Chanussot, and G. Vivone, "Assessment of hyperspectral sharpening methods for the monitoring of natural areas using multiplatform remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8208–8222, Oct. 2019, doi: 10.1109/TGRS.2019.2918932.

[50] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, *Hyperspectral Image Fusion*. Berlin, Germany: Springer, 2021.

[51] A. Mookambiga and V. Gomathi, "Comprehensive review on fusion techniques for spatial information enhancement in hyperspectral imagery," *Multidimensional Syst. Signal Process.*, vol. 27, no. 4, pp. 863–889, Oct. 2016, doi: 10.1007/s11045-016-0415-2.

[52] W. J. Carper, T. M. Lillesand, and R. W. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and ~ultispectraIImage data," *Photocramm. Eng. Remote Sens.*, vol. 56, Apr. 1990, Art. no. 459467.

[53] A. D. Philip, "The use of intensity-hue-saturation transformation for producing color shaded relief images," *Photogramm. Eng. Remote Sens.*, vol. 60, pp. 1369–1374, 1994.

[54] K. Kpalma, M. Chikr El-Mezouar, and N. Taleb, "Recent trends in satellite image pan-sharpening techniques," Vrniacka Banja, Serbia, Jun. 2014, Accessed: Jun. 17, 2021. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01075703

[55] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015, doi: 10.1109/TGRS.2014.2361734.

[56] Shruti and S. Budhiraja, "Multiscale image fusion for pan-sharpening of multispectral images using saliency detection," in *Proc. 9th Int. Conf. Contemporary Comput.*, Aug. 2016, pp. 1–6. doi: 10.1109/IC3.2016.7880253.

[57] Y. Zhang, "(PDF) understanding image fusion," 2008, Accessed: May 31, 2021. [Online]. Available: https://www.researchgate.net/publication/342318144_Understanding_Image_Fusion

[58] M. Choi, "A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1672–1682, Jul. 2006, doi: 10.1109/TGRS.2006.869923.

[59] M.-J. Choi, H.-C. Kim, N. Cho, and H. Kim, "An improved intensity-hue-saturation method for IKONOS image fusion," *Int. J. Remote Sens.*, vol. 13, pp. 1–5, Jan. 2006.

[60] R. Welch and M. Ehlers, "Merging multiresolution SPOT HRV and landsat TM data," *Photogramm.Eng. Remote Sens.*, vol. 53, pp. 301–303, Mar. 1987.

[61] S. M. A. Wady, Y. Bentoutou, A. Bengermikh, A. Bounoua, and N. Taleb, "A new IHS and wavelet based pansharpening algorithm for high spatial resolution satellite imagery," *Adv. Space Res.*, vol. 66, no. 7, pp. 1507–1521, Oct. 2020, doi: 10.1016/j.asr.2020.06.001.

[62] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001, doi: 10.1016/S1566-2535(01)00036-7.

[63] J. Pushparaj and A. V. Hegde, "Evaluation of pan-sharpening methods for spatial and spectral quality," *Appl. Geomat.*, vol. 9, no. 1, pp. 1–12, Mar. 2017, doi: 10.1007/s12518-016-0179-2.

[64] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, " An adaptive IHS pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 746–750, Oct. 2010, doi: 10.1109/LGRS.2010.2046715.

[65] C. Yang, Q. Zhan, H. Liu, and R. Ma, "An IHS-based pan-sharpening method for spectral fidelity improvement using ripplet transform and compressed sensing," *Sensors*, vol. 18, no. 11, Oct. 2018, Art. no. 3624, doi: 10.3390/s18113624.

[66] W. Hallada and S. Cox, "Image sharpening for mixed spatial and spectral resolution satellite systems," in *Proc. Int. Symp. Remote Sens. Environ.*, 1983, Accessed: May 31, 2021. [Online]. Available: /paper/Image-sharpening-for-mixed-spatial-and-spectral-Hallada-Cox/42fdced145103292dd3ea3c5441145abf6a835e5

[67] H. R. Shahdoosti, "MS and PAN image fusion by combining brovey and wavelet methods," p. 8, doi: 10.48550/arXiv.1701.01996.

[68] H. R. Shahdoosti, "Improved adaptive brovey as a new method for image fusion," Jul. 2018, Accessed: May 24, 2021. [Online]. Available: http://arxiv.org/abs/1807.09610

[69] S. S. Khan, Q. Ran, M. Khan, and Z. Ji, "Pan-sharpening framework based on laplacian sharpening with brovey," in *Proc. IEEE Int. Conf. Signal Inf. Data Process.*, Dec. 2019, pp. 1–5, doi: 10.1109/ICSIDP47821.2019.9173129.

[70] J. Chavez Pat, S. Sides, and J. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, pp. 265–303, Mar. 1991.

[71] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008, doi: 10.1109/TGRS.2008.916211.

[72] Z. Zhou, N. Ma, Y. Li, P. Yang, P. Zhang, and Y. Li, "Variational PCA fusion for pan-sharpening very high resolution imagery," *Sci. China Inf. Sci.*, vol. 57, no. 11, pp. 1–10, Nov. 2014, doi: 10.1007/s11432-014-5108-6.

[73] V. Shah, N. H. Younan, and R. King, "An adaptive PCA-based approach to pan-sharpening," in *Proc. SPIE, Int. Soc. Opt. Eng.*, Oct. 2007, Paper 678402, doi: 10.1117/12.736674.

[74] A. M. Gonzalez, X. Otazu, O. Fors, R. García, and J. Nunez, "Fusion of different spatial and spectral resolution images: Development, application and comparison of new methods based on wavelets," in *Proc. RAQRS*, Jan. 2002, pp. 228–237.

[75] M. Ghadjati, A. Moussaoui, and A. Boukharouba, "A novel iterative PCA–based pansharpening method," *Remote Sens. Lett.*, vol. 10, no. 3, pp. 264–273, Mar. 2019, doi: 10.1080/2150704X.2018.1547443.

[76] M. Ehlers, "Multisensor image fusion techniques in remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 46, no. 1, pp. 19–30, Feb. 1991, doi: 10.1016/0924-2716(91)90003-E.

[77] C. A. Laben and V. B. Brower, "Process for enhancing the spatial resolution of multispectral imagery using Pan-sharpening," U.S. Patent 6 011 875 A, Apr. 29, 1998. [Online]. Available: https://lens.org/135-660-046-023-136

[78] H. Sunuprapto, P. Danoedoro, and S. Ritohardoyo, "Evaluation of Pan-sharpening method: Applied to artisanal gold mining monitoring in gunung pani forest area," *Procedia Environ. Sci.*, vol. 33, pp. 230–238, 2016, doi: 10.1016/j.proenv.2016.03.074.

[79] S. Klonus and M. Ehlers, "Image fusion using the Ehlers spectral characteristics preservation algorithm," *GIScience Remote Sens.*, vol. 44, no. 2, pp. 93–116, Jun. 2007, doi: 10.2747/1548-1603.44.2.93.

[80] S. Xu and M. Ehlers, "Hyperspectral image sharpening based on ehlers fusion," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W7, pp. 941–947, Sep. 2017, doi: 10.5194/isprs-archives-XLII-2-W7-941-2017.

[81] S. Klonus and M. Ehlers, "Pansharpening with Terrasar-X and optical data," *3rd TerraSAR-X Sci. Team Meeting*, pp. 25–26, 2008.

[82] R. E. Crippen, "A simple spatial filtering routine for the cosmetic removal of scan-line noise from landsat TM P-Tape imagery," *Photogramm. Eng. Remote Sens.*, vol. 55, pp. 327–331, 1989.

[83] A. Garzelli, B. Aiazzi, L. Alparone, S. Lolli, and G. Vivone, "Multispectral pansharpening with radiative transfer-based detail-injection modeling for preserving changes in vegetation cover," *Remote Sens*, vol. 10, no. 8, Aug. 2018, Art. no. 1308, doi: 10.3390/rs10081308.

[84] S. Kahraman and A. Ertürk, "A comprehensive review of pansharpening algorithms for Göktürk-2 satellite images," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. IV-4/W4, pp. 263–270, Nov. 2017, doi: 10.5194/isprs-annals-IV-4-W4-263-2017.

[85] M. S. Karoui, K. Djerriri, and I. Boukerch, "Pansharpening multispectral remote sensing data by multiplicative joint nonnegative matrix factorization," *Int. J. Remote Sens.*, vol. 37, pp. 805–818, Feb. 2016, doi: 10.1080/01431161.2015.1137650.

[86] G. Palubinskas, "Model-based image adjustment for a successful pansharpening," Mar. 2021, Accessed: Jun. 08, 2021. [Online]. Available: http://arxiv.org/abs/2103.03062

[87] P. F. Fougere, "A defense of the Gram-Schmidt orthogonalization procedure applied to spherical harmonic analysis," *J. Geophys. Res. 1896–1977*, vol. 71, no. 21, pp. 5171–5174, 1966, doi: 10.1029/JZ071i021p05171.

[88] T. Maurer, "How to pan-sharpen images using the Gram-Schmidt pan-sharpen method—A recipe," *ISPRS—Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XL-1/W1, pp. 239–244, May 2013, doi: 10.5194/isprsarchives-XL-1-W1-239-2013.

[89] Z. Huang, Q. Chen, Q. Chen, and X. Liu, "Variational pansharpening for hyperspectral imagery constrained by spectral shape and Gram–Schmidt transformation," *Sensors*, vol. 18, no. 12, Dec. 2018, Art. no. 4330, doi: 10.3390/s18124330.

[90] H. Li, Y. Zhang, Y. Gao, and S. Yue, "Using guided filtering to improve gram-schmidt based pansharpening method for geoeye-1 satellite images," in *Proc. 4th Int. Conf. Inf. Syst. Comput. Technol.*, 2016, pp. 33–37, doi: 10.2991/isct-16.2016.6.

[91] G. Sarp, "Spectral and spatial quality analysis of pan-sharpening algorithms: A case study in Istanbul," *Eur. J. Remote Sens.*, vol. 47, no. 1, pp. 19–28, Jan. 2014, doi: 10.5721/EuJRS20144702.

[92] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005, doi: 10.1109/TGRS.2005.856106.

[93] M. Gonzalez-Audicana, J. L. Saleta, R. G. Catalan, and R. Garcia, "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1291–1299, Jun. 2004, doi: 10.1109/TGRS.2004.825593.

[94] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan. 2011, doi: 10.1109/TGRS.2010.2051674.

[95] H. R. Shahdoosti and H. Ghassemian, "Fusion of MS and PAN images preserving spectral quality," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 611–615, Mar. 2015, doi: 10.1109/LGRS.2014.2353135.

[96] C. E. Shannon, "A mathematical theory of communication," *Bell Labs Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

[97] P. Dherete and B. Rouge, "Image de-blurring and application to SPOT5 THR satellite imaging," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2003, vol. 1, pp. 318–320, doi: 10.1109/IGARSS.2003.1293762.

[98] C. Latry and B. Rouge, "Super resolution: Quincunx sampling and fusion processing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2003, vol. 1, pp. 315–317, doi: 10.1109/IGARSS.2003.1293761.

[99] C. Latry and B. Rougé, "In-flight commissioning of SPOT5 THR Quincunx sampling mode," in *Proc. Sensors, Syst. Next-Gener. Satellites VI*, 2003, Paper 199, doi: 10.1117/12.462632.

[100] A. Navarro *et al.*, "Crop monitoring based on SPOT-5 take-5 and Sentinel-1A data for the estimation of crop water requirements," *Remote Sens.*, vol. 8, no. 6, Jun. 2016, Art. no. 525, doi: 10.3390/rs8060525.

[101] D. C. Wehlage, J. A. Gamon, D. Thayer, and D. V. Hildebrand, "Interannual variability in dry mixed-grass prairie yield: A comparison of MODIS, SPOT, and field measurements," *Remote Sens.*, vol. 8, no. 10, Oct. 2016, Art. no. 872, doi: 10.3390/rs8100872.

[102] M. A. Castillo-Santiago, M. Ricker, and B. H. J. de Jong, "Estimation of tropical forest structure from SPOT-5 satellite images," *Int. J. Remote Sens.*, vol. 31, no. 10, pp. 2767–2782, May 2010, doi: 10.1080/01431160903095460.

[103] M. Hao, W. Shi, H. Zhang, Q. Wang, and K. Deng, "A scale-driven change detection method incorporating uncertainty analysis for remote sensing images," *Remote Sens.*, vol. 8, no. 9, Sep. 2016, Art. no. 745, doi: 10.3390/rs8090745.

[104] N.-B. Chang, M. Han, W. Yao, L.-C. Chen, and S. Xu, "Change detection of land use and land cover in an urban region with SPOT-5 images and partial Lanczos extreme learning machine," *J. Appl. Remote Sens.*, vol. 4, no. 1, Nov. 2010, Art. no. 043551, doi: 10.1117/1.3518096.

[105] D. Lu, M. Batistella, and E. Moran, "Integration of landsat TM and SPOT HRG images for vegetation change detection in the Brazilian Amazon," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 4, pp. 421–430, Apr. 2008, doi: 10.14358/PERS.74.4.421.

[106] E. Sertel and S. S. Akay, "High resolution mapping of urban areas using SPOT-5 images and ancillary data," *Int. J. Environ. Geoinform.*, vol. 2, no. 2, pp. 63–76, Aug. 2015, doi: 10.30897/ijegeo.303545.

[107] V. Pasqualini *et al.*, "Use of SPOT 5 for mapping seagrasses: An application to Posidonia oceanica," *Remote Sens. Environ.*, vol. 94, no. 1, pp. 39–45, Jan. 2005, doi: 10.1016/j.rse.2004.09.010.

[108] A. M. Borghuis, K. Chang, and H. Y. Lee, "Comparison between automated and manual mapping of typhoon-triggered landslides from SPOT-5 imagery," *Int. J. Remote Sens.*, vol. 28, no. 8, pp. 1843–1856, Apr. 2007, doi: 10.1080/01431160600935638.

[109] A. Fruchter and R. N. Hook, "Novel image reconstruction method applied to deep hubble space telescope images," in *Proc. Appl. Digit. Image Process. XX*, Oct. 1997, vol. 3164, pp. 120–125, doi: 10.1117/12.292751.

[110] M. T. Merino and J. Nunez, "Super-resolution of remotely sensed images with variable-pixel linear reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1446–1457, May 2007, doi: 10.1109/TGRS.2007.893271.

[111] Y. Ito, "Resolution enhancement of SAR image using a multiframe super resolution technique," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2009, vol. 4, pp. 446–449, doi: 10.1109/IGARSS.2009.5417409.

[112] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, May 1991, doi: 10.1016/1049-9652(91)90045-L.

[113] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, Dec. 2018, doi: 10.1109/TSP.2018.2876362.

[114] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 3034–3047, Jun. 2019, doi: 10.1109/TIP.2019.2893530.

[115] N. Liu, L. Li, W. Li, R. Tao, J. E. Fowler, and J. Chanussot, "Hyperspectral restoration and fusion with multispectral imagery via low-rank tensor-approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7817–7830, Sep. 2021, doi: 10.1109/TGRS.2020.3049014.

[116] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pan-sharpening method with deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1037–1041, May 2015, doi: 10.1109/LGRS.2014.2376034.

[117] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, Jul. 2016, Art. no. 594, doi: 10.3390/rs8070594.

[118] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018, doi: 10.1109/TGRS.2018.2817393.

[119] J. Wang, Z. Shao, X. Huang, T. Lu, and R. Zhang, "A dual-path fusion network for pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jun. 2022, Art. no. 5403214, doi: 10.1109/TGRS.2021.3090585.

[120] Z. Shao, L. Wang, Z. Wang, and J. Deng, "Remote sensing image super-resolution using sparse representation and coupled sparse autoencoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2663–2674, Aug. 2019, doi: 10.1109/JSTARS.2019.2925456.

[121] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018, doi: 10.1109/JSTARS.2018.2805921.

[122] C. Shang *et al.*, "Spatiotemporal reflectance fusion using a generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2021, Art. no. 5400915, doi: 10.1109/TGRS.2021.3065418.

[123] H. Zhang, Y. Song, C. Han, and L. Zhang, "Remote sensing image spatiotemporal fusion using a generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4273–4286, May 2021, doi: 10.1109/TGRS.2020.3010530.

[124] D. Feng, D. An, L. Chen, and X. Huang, "Holographic SAR tomography 3-D reconstruction based on iterative adaptive approach and generalized likelihood ratio test," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 305–315, Jan. 2021, doi: 10.1109/TGRS.2020.2994201.

[125] J. Leinonen, D. Nerini, and A. Berne, "Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 1–13, Sep. 2020, doi: 10.1109/TGRS.2020.3032790.

[126] F. Han, H. Zhang, S. Chatterjee, Q. Guo, and S. Wan, "A modified generative adversarial nets integrated with stochastic approach for realizing super-resolution reservoir simulation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1325–1336, Feb. 2020, doi: 10.1109/TGRS.2019.2945946.

[127] J. Wang, Z. Shao, X. Huang, T. Lu, R. Zhang, and J. Ma, "Enhanced image prior for unsupervised remoting sensing super-resolution," *Neural Netw.*, vol. 143, no. C, pp. 400–412, Nov. 2021, doi: 10.1016/j.neunet.2021.06.005.

[128] Z. Shao, J. Cai, P. Fu, L. Hu, and T. Liu, "Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product," *Remote Sens. Environ.*, vol. 235, Dec. 2019, Art. no. 111425, doi: 10.1016/j.rse.2019.111425.

[129] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[130] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, "LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop," Jun. 2016, Accessed: Dec. 28, 2021. [Online]. Available: http://arxiv.org/abs/1506.03365

[131] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context," Feb. 2015, Accessed: Dec. 28, 2021. [Online]. Available: http://arxiv.org/abs/1405.0312

[132] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Honolulu, HI, USA, Jul. 2017, pp. 1122–1131, doi: 10.1109/CVPRW.2017.150.

[133] H. Ren, A. Kheradmand, M. El-Khamy, S. Wang, D. Bai, and J. Lee, "Real-world super-resolution using generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Seattle, WA, USA, Jun. 2020, pp. 1760–1768, doi: 10.1109/CVPRW50498.2020.00226.

[134] Z.-S. Liu, W.-C. Siu, L.-W. Wang, C.-T. Li, M.-P. Cani, and Y.-L. Chan, "Unsupervised real image super-resolution via generative variational autoencoder," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Seattle, WA, USA, Jun. 2020, pp. 1788–1797, doi: 10.1109/CVPRW50498.2020.00229.

[135] R. M. Umer, G. L. Foresti, and C. Micheloni, "Deep generative adversarial residual convolutional networks for real-world super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Seattle, WA, USA, Jun. 2020, pp. 1769–1777, doi: 10.1109/CVPRW50498.2020.00227.

[136] A. Lugmayr *et al.*, "NTIRE 2021 learning the super-resolution space challenge," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2021, pp. 596–612, doi: 10.1109/CVPRW53098.2021.00072.

[137] S. Son *et al.*, "NTIRE 2021 challenge on video super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2021, pp. 166–181, doi: 10.1109/CVPRW53098.2021.00026.

[138] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," Aug. 2016, Accessed: Jul. 15, 2021. [Online]. Available: http://arxiv.org/abs/1608.00367

[139] M. Bevilacqua, A. Roumy, C. Guillemot, and M. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 135.1–135.10, doi: 10.5244/C.26.135.

[140] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surfaces*, 2010, vol. 6920, pp. 711–730, doi: 10.1007/978-3-642-27413-8_47.

[141] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, Jul. 2001, vol. 2, pp. 416–423, doi: 10.1109/ICCV.2001.937655.

[142] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5197–5206, doi: 10.1109/CVPR.2015.7299156.

[143] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," Mar. 2019, Accessed: Sep. 17, 2021. [Online]. Available: http://arxiv.org/abs/1812.04948

[144] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," Jul. 2017, Accessed: Jul. 15, 2021. [Online]. Available: http://arxiv.org/abs/1707.02921

[145] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 11057–11066, doi: 10.1109/CVPR.2019.01132.

[146] B. Niu *et al.*, "Single image super-resolution via a holistic attention network," Aug. 2020, Accessed: Jul. 15, 2021. [Online]. Available: http://arxiv.org/abs/2008.08767

[147] H. Cao and S. Mi, "Weighted SRGAN and reconstruction loss analysis for accurate image super resolution," *J. Phys., Conf. Ser.*, vol. 1903, no. 1, Apr. 2021, Art. no. 012050, doi: 10.1088/1742-6596/1903/1/012050.

[148] J.-T. Jiang, H. L. Huang, and L. Hu, "Single image super-resolution: Depthwise separable convolution super-resolution generative adversarial network," *Appl. Sci.*, vol. 10, Jan. 2020, Art. no. 375, doi: 10.3390/app10010375.

[149] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," Sep. 2016, Accessed: Jul. 19, 2021. [Online]. Available: http://arxiv.org/abs/1609.05158

[150] M. S. M. Sajjadi, B. Schölkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," Jul. 2017, Accessed: Jul. 15, 2021. [Online]. Available: http://arxiv.org/abs/1612.07919

[151] R. Kalarot, T. Li, and F. Porikli, "Component attention guided face super-resolution network: CAGFace," Oct. 2019, Accessed: Jul. 15, 2021. [Online]. Available: http://arxiv.org/abs/1910.08761

[152] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," Jun. 2019, Accessed: Jul. 15, 2021. [Online]. Available: http://arxiv.org/abs/1903.09814

[153] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1646–1654, doi: 10.1109/CVPR.2016.182.

[154] L. Wagner, L. Liebel, and M. Körner, "Deep residual learning for single-image super-resolution of multi-spectral satellite imagery," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. IV-2/W7, pp. 189–196, Sep. 2019, doi: 10.5194/isprs-annals-IV-2-W7-189-2019.

[155] X. Feng, X. Su, J. Shen, and H. Jin, "Single space object image denoising and super-resolution reconstructing using deep convolutional networks," *Remote Sens.*, vol. 11, no. 16, Jan. 2019, Art. no. 1910, doi: 10.3390/rs11161910.

[156] N. Latte and P. Lejeune, "PlanetScope radiometric normalization and sentinel-2 super-resolution (2.5m): A straightforward spectral-spatial fusion of multi-satellite multi-sensor images using residual convolutional neural networks," *Remote Sens.*, vol. 12, no. 15, Jan. 2020, Art. no. 2366, doi: 10.3390/rs12152366.

[157] T. Lu, J. Wang, Y. Zhang, Z. Wang, and J. Jiang, "Satellite image super-resolution via multi-scale residual deep neural network," *Remote Sens.*, vol. 11, no. 13, Jan. 2019, Art. no. 1588, doi: 10.3390/rs11131588.

[158] W. Dong, Y. Yang, J. Qu, W. Xie, and Y. Li, "Fusion of hyperspectral and panchromatic images using generative adversarial network and image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2021, Art. no. 5508413, doi: 10.1109/TGRS.2021.3078711.

[159] T. Li and Y. Gu, "Progressive spatial-spectral joint network for hyperspectral image reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2021, Art. no. 5507414, doi: 10.1109/TGRS.2021.3079969.

[160] J. Ying, H.-L. Shen, and S.-Y. Cao, "Unaligned hyperspectral image fusion via registration and interpolation modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2021, Art. no. 5511114, doi: 10.1109/TGRS.2021.3081136.

[161] X. Wang, J. Ma, and J. Jiang, "Hyperspectral image super-resolution via recurrent feedback embedding and spatial-spectral consistency regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2021, Art. no. 5503113, doi: 10.1109/TGRS.2021.3064450.

[162] D. Liu, J. Li, and Q. Yuan, "A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7711–7725, Sep. 2021, doi: 10.1109/TGRS.2021.3049875.

[163] R. Dian, S. Li, L. Fang, and Q. Wei, "Multispectral and hyperspectral image fusion with spatial-spectral sparse representation," *Inf. Fusion*, vol. 49, pp. 262–270, Sep. 2019, doi: 10.1016/j.inffus.2018.11.012.

[164] J. Li, R. Cui, B. Li, R. Song, Y. Li, and Q. Du, "Hyperspectral image super-resolution with 1D–2D attentional convolutional neural network," *Remote Sens.*, vol. 11, no. 23, Jan. 2019, Art. no. 2859, doi: 10.3390/rs11232859.

[165] I. J. Goodfellow *et al.*, "Generative adversarial networks," Jun. 2014, Accessed: Sep. 13, 2021. [Online]. Available: http://arxiv.org/abs/1406.2661

[166] L. Liebel and M. Körner, "Single-image super resolution for multispectral remote sensing data using convolutional neural networks," *ISPRS—Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 41B3, pp. 883–890, Jun. 2016, doi: 10.5194/isprs-archives-XLI-B3-883-2016.

[167] Y. Xiong *et al.*, "Improved SRGAN for remote sensing image super-resolution across locations and sensors," *Remote Sens.*, vol. 12, no. 8, Jan. 2020, Art. no. 1263, doi: 10.3390/rs12081263.

[168] X. Dou, C. Li, Q. Shi, and M. Liu, "Super-resolution for hyperspectral remote sensing images based on the 3D attention-SRGAN network," *Remote Sens.,* vol. 12, no. 7, Jan. 2020, Art. no. 1204, doi: 10.3390/rs12071204.

[169] L. S. Romero, J. Marcello, and V. Vilaplana, "Super-resolution of Sentinel-2 imagery using generative adversarial networks," *Remote Sens.*, vol. 12, no. 15, Jan. 2020, Art. no. 2424, doi: 10.3390/rs12152424.

[170] M. Pashaei, M. J. Starek, H. Kamangir, and J. Berryhill, "Deep learning-based single image super-resolution: An investigation for dense scene reconstruction with UAS photogrammetry," *Remote Sens.*, vol. 12, no. 11, Jan. 2020, Art. no. 1757, doi: 10.3390/rs12111757.

[171] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, and D. Chao, "Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network," *Remote Sens.*, vol. 12, no. 9, Jan. 2020, Art. no. 1432, doi: 10.3390/rs12091432.

[172] Planet | Homepage, *Planet*. Accessed: Dec. 10, 2021. [Online]. Available: https://www.planet.com/

[173] M. Galar, R. Sesma, C. Ayala, L. Albizua, and C. Aranda, "Super-resolution of Sentinel-2 images using convolutional neural networks and real ground truth data," *Remote Sens.*, vol. 12, no. 18, Jan. 2020, Art. no. 2941, doi: 10.3390/rs12182941.

[174] F. Salvetti, V. Mazzia, A. Khaliq, and M. Chiaberge, "Multi-image super resolution of remotely sensed images using residual attention deep neural networks," *Remote Sens.*, vol. 12, no. 14, Jan. 2020, Art. no. 2207, doi: 10.3390/rs12142207.

[175] M. Qin *et al.*, "Achieving higher resolution lake area from remote sensing images through an unsupervised deep learning super-resolution method," *Remote Sens.*, vol. 12, no. 12, Jan. 2020, Art. no. 1937, doi: 10.3390/rs12121937.

[176] H. Ahn and C. Yim, "Convolutional neural networks using skip connections with layer groups for super-resolution image reconstruction based on deep learning," *Appl. Sci.*, vol. 10, no. 6, Jan. 2020, Art. no. 1959, doi: 10.3390/app10061959.

[177] L. Courtrai, M.-T. Pham, and S. Lefèvre, "Small object detection in remote sensing images based on super-resolution with auxiliary generative adversarial networks," *Remote Sens.*, vol. 12, no. 19, Jan. 2020, Art. no. 3152, doi: 10.3390/rs12193152.

[178] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein GANs," Dec. 2017, Accessed: Jun. 16, 2021. [Online]. Available: http://arxiv.org/abs/1704.00028

[179] Y. Gao, H. Li, J. Dong, and G. Feng, "A deep convolutional network for medical image super-resolution," in *Proc. Chin. Automat. Congr.*, Oct. 2017, pp. 5310–5315, doi: 10.1109/CAC.2017.8243724.

[180] D. Mahapatra and B. Bozorgtabar, "Progressive generative adversarial networks for medical image super resolution," Feb. 2019, Accessed: May 28, 2021. [Online]. Available: http://arxiv.org/abs/1902.02144

[181] T. Gulati, S. Sengupta, and V. Lakshminarayanan, "Application of an enhanced deep super-resolution network in retinal image analysis," in *Proc. Ophthalmic Technol. XXX*, Feb. 2020, vol. 11218, Paper 112181K, doi: 10.1117/12.2543791.

[182] G. Zamzmi, S. Rajaraman, and S. Antani, "Accelerating super-resolution and visual task analysis in medical images," *Appl. Sci.*, vol. 10, no. 12, Jan. 2020, Art. no. 4282, doi: 10.3390/app10124282.

[183] "A pan-sharpening method based on evolutionary optimization and IHS transformation," Accessed: Dec. 10, 2021. [Online]. Available: https://www.hindawi.com/journals/mpe/2017/8269078/

[184] S. Panchal and R. Thakker, "Implementation and comparative quantitative assessment of different multispectral image pansharpening approaches," *Signal Image Process., Int. J.*, vol. 6, no. 5, pp. 35–48, Jun. 2015, doi: 10.5121/sipij.2015.6503.

[185] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion ?," in *Proc. 3rd Int. Conf. Fusion Earth Data, Merging Point Meas. Raster Maps Remotely Sensed Images*, Sophia Antipolis, France, Jan. 2000, pp. 99–103, Accessed: Sep. 17, 2021. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00395027

[186] Q. Du, O. Gungor, and J. Shan, "Performance evaluation for pan-sharpening techniques," in *Proc. Geosci. Remote Sens. Symp.*, 2005, Art. no. 4266, doi: 10.1109/IGARSS.2005.1525860.

[187] Y. Chen, "A new methodology of spatial cross-correlation analysis," *PLoS One*, vol. 10, no. 5, May 2015, Art. no. e0126158, doi: 10.1371/journal.pone.0126158.

[188] A. Goetz, W. Boardman, and R. Yunas, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," *Proc. 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.

[189] C.-I. Chang, "Spectral information divergence for hyperspectral image analysis," in *Proc. Int. Geosci. Remote Sens. Symp.*, Jun. 1999, vol. 1, pp. 509–511, doi: 10.1109/IGARSS.1999.773549.

[190] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002, doi: 10.1109/97.995823.

[191] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," Aug. 2021, Accessed: Mar. 16, 2022. [Online]. Available: http://arxiv.org/abs/2107.10833

Kinga Karwowska received the M.Sc. degree in geoinformatics in 2020 from the Military University of Technology, Warsaw, Poland, where she is currently working toward the Ph.D. degree with the Doctoral School.

Damian Wierzbicki received the Ph.D. degree in photogrammetry and remote sensing from the Military University of Technology, Warsaw, Poland, in 2015.

He is currently an Associate Professor with the Department of Imagery Intelligence, Faculty of Civil Engineering and Geodesy, Military University of Technology, where he teaches photogrammetry and remote sensing and image processing. His research interests include UAV navigation and image processing, deep learning in remote sensing, the development of new algorithms for object detection, and the classification of image sequences from UAVs.

MDPI

Article

Improving Spatial Resolution of Satellite Imagery Using Generative Adversarial Networks and Window Functions

Kinga Karwowska * and Damian Wierzbicki

Department of Imagery Intelligence, Faculty of Civil Engineering and Geodesy, Military University of Technology, 00-908 Warsaw, Poland
* Correspondence: kinga.karwowska@wat.edu.pl

Abstract: Dynamic technological progress has contributed to the development of systems imaging of the Earth's surface as well as data mining methods. One such example is super-resolution (SR) techniques that allow for the improvement of the spatial resolution of satellite imagery on the basis of a low-resolution image (LR) and an algorithm using deep neural networks. The limitation of these solutions is the input size parameter, which defines the image size that is adopted by a given neural network. Unfortunately, the value of this parameter is often much smaller than the size of the images obtained by Earth Observation satellites. In this article, we presented a new methodology for improving the resolution of an entire satellite image, using a window function. In addition, we conducted research to improve the resolution of satellite images acquired with the World View 2 satellite using the ESRGAN network, we determined the number of buffer pixels that will make it possible to obtain the best image quality. The best reconstruction of the entire satellite imagery using generative neural networks was obtained using a Triangular window (for 10% coverage). The Hann-Poisson window worked best when more overlap between images was used.

Keywords: remote sensing; satellites; neural network application; image processing; image resolution

---

1. Introduction

In recent years, we have been witnessing a growing interest in imagery obtained from space altitudes. According to the Union of Concerned Scientists (UCS) Satellite Database, (as of 1 January 2022) 1031 Earth Observation satellites have been launched since 1994, and 68% of them were launched in the last 5 years. At least 55% of Earth Observation satellites enable the acquisition of the following types of images: optical (e.g., Pleiades Neo, SkySat, Gaofen, Worldview), optical stereo (e.g., Gaofen 14), multispectral (e.g., Dove 4p-1-5, 7, 10-11, CSO-1,2), hyperspectral (e.g., ÑuSat-4, 5, Spark- 1,2, OHS), infrared (e.g., Tianjin Daxue 1, HOPSAT-TD, TTU100), radar (e.g., Yaogan, COSMO-Skymed, SAR-Lupe, Capella, ICEYE), as well as video materials (Jilin-1, UVS). As Earth Observation systems are evolving, small satellites are becoming increasingly popular, including nanosatellites (weight 1–10 kg, e.g., Dove, NAPA- 1, 2, and Jin Zijing 2 constellations), microsatellites (weight 10–100 kg, e.g., Jilin, BlackSky Global, ICEYE, OHS, and GRUS constellations), and minisatellites (weight 100–500 kg, e.g., Capella, SkySat, and Kanopus-V-IK constellations). These small satellites, with weights that do not exceed 500 kg, enable the acquisition of images of a spatial resolution lower than 1 m. Their potential is also emphasized by their number: according to data provided in the UCS Satellite Database, they account for 71% of all imagery acquiring satellites.

Satellite imagery is commonly applied in numerous areas. It is frequently used in environmental protection, spatial planning, monitoring changes, or in military applications. However, in order to improve the interpretational capacity of the acquired satellite imagery, it is necessary to perform certain operations on the images. The traditional and most popular of which include digital image processing methods, such as segmentation or

detection of changes. However, new methods that employ machine learning and especially deep learning have been becoming increasingly popular in recent years. Multiple solutions that enable the detection and recognition of objects have been developed [1–4], as well as those that allow for the segmentation of scenes [5–7] or improvement of the resolution of satellite images [8,9] and using linear regression [10,11].

Unlike classification, the task of linear regression is to predict numeric variables, not discrete (categorical) variables. In the literature, there are solutions in which classification tasks are solved by regression. This is possible by using the appropriate loss function—mean squared error (MSE), root mean squared error (L2), mean absolute error (L1) or Huber loss. This method is used in tasks related to ice concentration estimation [12–14], vegetation index estimation [15], motion parameters estimation of moving targets [16] and ship orientation angle estimation [17].

Methods that use deep neural networks make it possible to process large datasets quickly and to extract information that would be impossible to extract with the use of digital image processing methods. In order to use the designed architectures for other purposes than those for which they were dedicated, it is necessary to prepare a database of training data in such a way that the data will meet the input parameters of the network (height, width, and number of channels). Unfortunately, this operation leads to changing the values of pixels in the image, which may, in turn, result in the loss of information from the image. A review of the size of the parameters that define the entry to the network reveals that they are significantly smaller than the dimensions of images that are acquired by optoelectronic sensors installed on UAVs, Reconnaissance aircrafts or EO satellites. Moreover, even nanosatellite systems that are equipped with small arrays, e.g., CMOS [18] or CMOSIS CMV [19], characterized by a low quantum yield have a much higher resolution than that of the designed input to the neural networks. In order to apply the solutions that use deep neural networks, it is thus necessary to divide the imagery into smaller images of specific dimensions, and the training process should be performed according to the diagram below (Figure 1). The result of this procedure is a set of resulting images, yet not the whole image. This gives rise to the question: how to combine the obtained results?
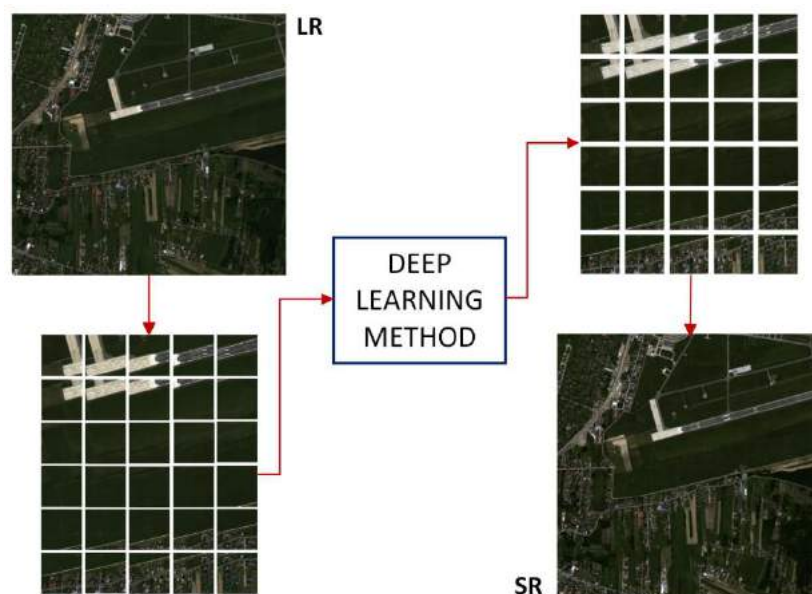
Figure 1. Diagram illustrating the process of improving the resolution of satellite imagery.

Moreover, as far as image to image [20] algorithms are concerned, as in the example of improving the spatial resolution of satellite imagery with the use of generative adversarial networks, the problem is additionally complicated as each of the pixels in the image is re-evaluated. Additionally, in order to use AI algorithms to improve the resolution of images that depict urban areas and contain a large number of details, it is necessary to minimize

the phenomenon where the same object that is present in several images is represented in different ways.

It is well known that in spite of the very high computational power of graphic processors and the possibility to use virtual machines, it is very difficult to process such large satellite images. As a result, the following research questions arise:

1. Are there any methods to combine images after the application of algorithms to improve spatial resolution with the use of deep learning methods?
2. What methodology should be adopted to combine images evaluated by generative adversarial networks?
3. What is the number of buffer pixels that will result in the best quality of the resulting image?
4. Can this method also be used to combine images that are the outcome of segmentation algorithms?

The paper consists of the following sections: Section 2 discusses the methods of improving the resolution of satellite imagery. The proposed methodology is presented in Section 3. Section 4 contains the Discussion, while the final conclusions are presented in Section 5.

2. Related Works

Considering the review of the solutions used to improve the resolution of satellite images [21,22], the methods may be divided into two groups. The first group contains methods that enable the processing of whole images. These include interpolation, solutions using signal processing (e.g., MUltiple SIgnal Classification (MUSIC) algorithm [23]) and pansharpening methods. The second group are the methods that allow for the improvement of the resolution only for images of specific dimensions. Examples of this are all solutions that employ convolutional neural networks (CNN). They enable the learning of local patterns of the image, which are the basis for image classification. Classification is applied both to groups of pixels (object classification) and single pixels (image segmentation). One of the elements that characterize each convolutional neural network is the dimension of input to the first hidden layer. This parameter defines the dimensions of square images that will be processed by the network. This value is determined mainly by the computational power of workstations, as the number of searched parameters increases with the growing size of the image. Table 1 presents examples of the input dimensions to the first hidden layer in sample networks that solve the problems of the classification, detection, segmentation, and translation of digital images. Based on the presented information, one may notice that the size of the processed images usually does not exceed one million pixels.

Table 1. Sample dimensions of the input layers used in processing satellite data.

Method	Input Size
classification	224×224 [24–26], 299×299 [27]
object detection	400×400 [28], 668×668 [29], 1024×1024 [30]
segmentation	32×32 [31], 128×128 [32], 512×512 [33], 513×513 [34]
image-to-image translation	256×256 [35,36], 500×500 [37,38], 64×64 [39], 96×96 [40,41], 128×128 [41], 192×192 [41]

Apart from that, it is noticeable that the results presented by other authors have the form of small resultant samples [42–44]. Unfortunately, researchers do not address the issue of improving the resolution of whole satellite images or developing the methodology to improve entire satellite images.

The problem of combining images most often appears in the literature, in the context of creating panoramas or stitching several photos [45,46]. An example is the work of the Mingyuan Lin team. They drew attention to the problem of combining images by disparity-guided multi-plane alignment. In this solution, the researchers used the algorithm guided

by the disparity map, which allowed for limiting the occurrence of parallax artifacts [47]. However, the use of this solution for a large number of images (which we deal with when improving spatial resolution with solutions using neural networks) would require the use of large image coverage, which would significantly extend the work of the algorithm. Meanwhile, the problem of processing large images using deep neural networks has been noticed by scientists involved in the processing of biological and medical images, where samples containing several gigapixels are used. They found that the use of the Hann window to combine images significantly reduces the number of unwanted artifacts (e.g., edge effects) [48]. Unfortunately, in the case of satellite imagery, where the number of details in the photos is very large, the problem of combining photos is more complicated.

3. Experiments and Results

3.1. The Proposed Method

The aim of the experiment is to combine super resolution (SR) images estimated with the use of a GAN network based on low resolution (LR) images. In the proposed methodology, the fragments of the scene are combined with the use of two-dimensional window functions. The application of the window function consists of preparing the adequate matrix of weights (which is symmetrical in relation to the center of the image) and then multiplying it by the super resolution (SR) image, estimated by the generative adversarial network. The specific stages of the process are presented in Figure 2.
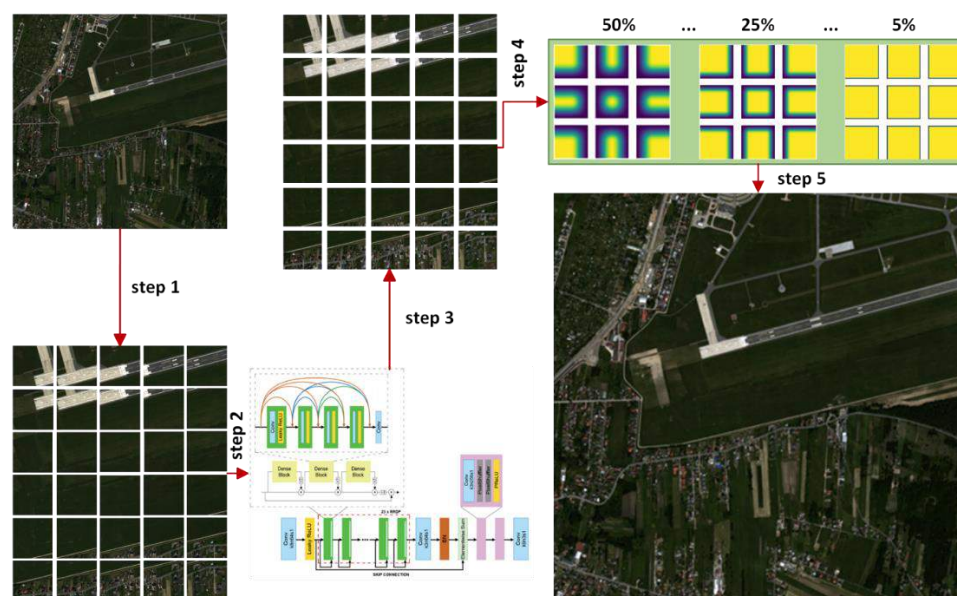
Figure 2. Diagram of the methodology to improve the resolution of whole satellite scenes. Step 1: divided of image, step 2: upload images to the network, step 3: use a method of improving spatial resolution using neural networks, step 4: select a window function to combine images, step 5: combine images.

This solution may be applied thanks to the properties of the window function: (1) non-zero on a finite interval, (2) reaches a maximum at the center of the interval, (3) is symmetrical relative to the center of the interval. Additionally, for combining images, another condition should be checked: (4) the sum of weights for each of the pixels equals 1. If the sum of weights is lower (or higher) than 1, the image after combining has a characteristic grid that consists of pixels of lower (or higher) DN values. For the purpose of the analysis of selected windows, four parameters were tested: minimum and maximum value, the average sum of weights and the sum of weights for the shared area of the images. The tests of the possibility to apply windows for the purpose of combining images were conducted for images with the dimensions 384×384 pixels. The combination of two images

of the size 384 × 384 pixels was simulated. This dimension was not selected at random. The SRGAN and ESRGAN networks usually collect low resolution (LR) images with the dimensions of 96 × 96 pixels as input, and return an SR image with the dimensions of 384 × 384 pixels. The works on the analyzed issue were divided into two stages. At the first stage, preliminary tests were conducted, which were the basis for selecting only those windows for which: (1) the sum of weights at a point belonged to the range 0.95 to 1.05], and (2) the sum of weights for the overlap area belonged to the range 190 to 192, based on the assumption that the overlap of the combined images (component images) was 50%. Based on the set of metrics proposed in the subsequent sections of this paper, the quality of the image after combining (the resultant image) was assessed. Additionally, it was assumed that the analyzed image would not be subjected to any digital image processing operations. Thus, the image would be divided into component images and then combined, with the use of windows, to create one resulting image. At the main stage tests, the influence of the size of overlap of component images on the quality of the resulting image was analyzed. Only four windows that brought the best results in the preliminary research phase were used.

3.2. Equations

The quality of the SR images estimated by the ESRGAN network was determined with the use of some of the most popular metrics used in the fields of remote sensing and computer vision.

3.2.1. Peak Signal-to-Noise Ratio

Peak signal-to-noise ratio (PSNR) [49] (Equation (1)) is the ratio of the maximum signal power (maximum value of the image) to the power that interferes with his signal, i.e., the mean square error. *PSNR* values are expressed in decibels.

$$PSNR = 10 \cdot log_{10} \frac{[max(HR(n,m))]^2}{MSE},\tag{1}$$

where *MSE*—mean square error, *max(HR(n,m))*—maximum value of the image.

For images that have been recreated in high quality, i.e., when the *MSE* approaches zero, the value of *PSNR* tends towards infinity. This means that the higher the *PSNR* value, the better the images had been combined.

3.2.2. Universal Quality Measure

Universal Quality Measure (UQI) [50,51] is another metric that compares the reference image (HR) and the image after processing (*HR'*). The value of the UQI metric is determined based on the values of the image pixels, but also their average and variance. UQI is calculated from equation (Equation (2)):

$$Q = \frac{\sigma'_{HRHR}}{\sigma_{HR}\sigma'_{HR}} \cdot \frac{2 \cdot HR \cdot HR'}{\left(\overline{(HR)}\right)^2 \left(\overline{(HR')}\right)^2} \cdot \frac{2\sigma_{HR}\sigma'_{HR}}{\sigma^2_{HR} + \sigma^2_{HR'}},\tag{2}$$

where *HR*—reference image, *HR'*—image after processing.

3.2.3. Spatial Correlation Coefficient

The Spatial Correlation Coefficient (SCC) [49] is a method of assessing image processing based on CC. In this method, maps of the properties of high frequency images which emerge after the application of edge detection filters are assessed.

3.2.4. Spectral Angle Mapper

Spectral Angle Mapper (*SAM*) [49,52] is a metric that defines the average change of all angles in the spectral component (Equation (3)).

$$SAM(v, w) = \cos^{-1} \left(\frac{\sum_{i=1}^{N} HR'_i HR_i}{\sqrt{\sum_{i=1}^{N} HR'^2_i} \sqrt{\sum_{i=1}^{N} HR^2_i}} \right), \tag{3}$$

where *N*—number of channels, *HR'*, *HR*, respectively: test spectrum and reference spectrum (each has n components).

3.2.5. Spectral Angle Mapper

Structural similarity index measure (SSIM) [49] (Equation (4)) is a measure of the structural similarity in the image domain, additionally taking into account the changes in brightness and contrast. The measure of brightness changeability is defined by the difference in the value of average brightness in the image, while the change in contrast is defined by the standard deviation. SSIM takes the values from the range <−1, 1>, where if SSIM = 1, the reference image is the same as the processed image.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}, \tag{4}$$

where μ_x—average brightness in the *X* window, μ_y—average brightness in the *Y* window $\sigma_x{}^2$—variance in the *X* window, $\sigma_y{}^2$—variance in the *Y* window, σ_{xy}—covariance of pixels in the *X* and *Y* windows, C_1 and C_2—permanent coefficients.

3.2.6. VIFP

The VIFP metrics is also known under the name VIF. It is the quantification of two mutual information quantities: the mutual information between the input and the output of the HVS channel when no distortion channel is present (this is referred to as the reference image information) and the mutual information between the input of the distortion channel and the output of the HVS channel for the test image [53].

3.2.7. Normalized Root Mean-Squared Error

Another modification of the mean square error MSE is the Normalized root mean-squared error (*NRMSE*) [49] (Equation (5)). Literature does not provide a standard normalization method. Depending on the chosen method, this error is calculated as the quotient of the square root of the MSE (i.e., the *RMSE*) and the mean (in subject literature, the mean value is sometimes replaced with standard deviation, difference between maximum and minimum or the interquartile range). A sample equation is represented as.

$$SNRMSE = \frac{RMSE}{HR}, \tag{5}$$

where *RMSE* is the Root mean square error, *HR*—reference image.

3.2.8. Mean Square Error

The Mean square error (*MSE*) [49] between images is based on calculating the square error between the estimator (the *HR* image) and the estimated value (the *HR'* image) (6). It is the main measure to assess the quality of the generated image. As a result, lower values of the *MSE* correspond to a better recreation of the image. This error is calculated with the equation:

$$MSE = \frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} \left[HR(n, m) - HR'(n, m)\right]^2, \tag{6}$$

where N, M—image resolution, n,m—coordinates of the analyzed pixel, HR—high-resolution image, HR'—image combined with the use of the window function.

3.2.9. Root Mean Square Error

The Root mean square error (*RMSE*) [49] is another measure used to assess the quality of the generated image. *RMSE* is the square root of the mean square error (Equation (7)):

$$RMSE = \sqrt{\frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} [HR(n,m) - HR'(n,m)]^2}, \tag{7}$$

where N, M image resolution, n,m—coordinates of the analyzed pixel, HR—high-resolution image, HR'—image combined with the use of the window function.

3.3. Preliminary Tests

The tests verified the possibility to use the chosen windows (i.e., Welch, Sine, Hann, Bartlett-Hann, Triangular, Hann-Poisson, Gaussian, Lanchos, Blackman, Blackman-Nuttall, Blackman–Harris window, Flat top window, Poisson, and Hamming [53]). For the simulated image combination, the sum of weights at a point, the sum of all weights, and the average value in the image overlap area were calculated. Appendix A presents the obtained results and the window formula that was used for the calculations. Considering the properties of windows (2) reaches a maximum at the center of the interval, (3) is symmetrical relative to the center of the interval], for the purposes of tests it was assumed that the windows are one-dimensional.

Considering the obtained results, which are additionally visualized in Figure 3, one may notice that the properties defined above are met only by the following windows: Hann, Bartlett-Hann, Triangular, and Hann Poisson. Additionally, it is worth noting the Backman window, which does not meet only the condition that refers to the sum of weights of the windows. As far as the Backman window is concerned, the sum of all weights equals 190.08, which is 1.92 lower than the target. Additionally, at the further stages of research, the Gaussian, Lanchos, and Blackman window functions will be tested. They do not meet the condition introduced by the authors at all, but, at the same time, the sums of weights do not diverge significantly from the target.

Additionally, Appendix A presents the functions that describe the manner of collecting samples of the image, i.e., calculating the weights that will be used while combining the images. Some of the presented window function formulas (Hann-Poisson) use (constant) parameters that were determined by the authors of the relevant solutions. However, in order to meet the condition of the sum of weights of image pixels, new values of these parameters were defined. As a result of this operation, the sum of weights of the pixels for those solutions is close to 1.

Table 2 and Figure 4 present the results of the final stage of preliminary research. The tests were conducted on one of the images from the sequence of images acquired by the Jinlin minisatellite. The prepared image was divided into component images of the dimensions 384×384 pixels with an overlap of 192 pixels (50%). Then, the prepared images were combined with the weights calculated with the use of the Hann, Bartlett-Hann, Triangular, Hann-Poisson, Gaussian, Lanchos, and Blackman window functions. The quality of the obtained resultant image was analyzed based on the evaluation metrics described in point A. The obtained results clearly reveal that those windows, where the sum of weights for a single pixel equals 1, present the best results of the evaluation metrics. Apart from that, combining images based on weights determined with the use of window functions leads to a slight deterioration in image quality, which is proven by the low value of MSE and RASE errors. At the same time, the Peak signal-to-noise ratio (PSNR) takes high values, e.g., for the Hann window PSNR = 51.89 dB, which means very high similarity between the resulting image and the reference image. As for the windows, for which the sum of weights in a point in the overlap area is different from 1, horizontal and vertical

stripes are visible in the resulting images. If the value of the sum of weights is lower than one, the DN values for those pixels are lower than the target value, which, in consequence, leads to the emergence of a dark grid. On the other hand, if the sum of weights is higher than 1, the DN value of the pixel is higher than the target value, which results in the emergence of lighter stripes (Figure 5).
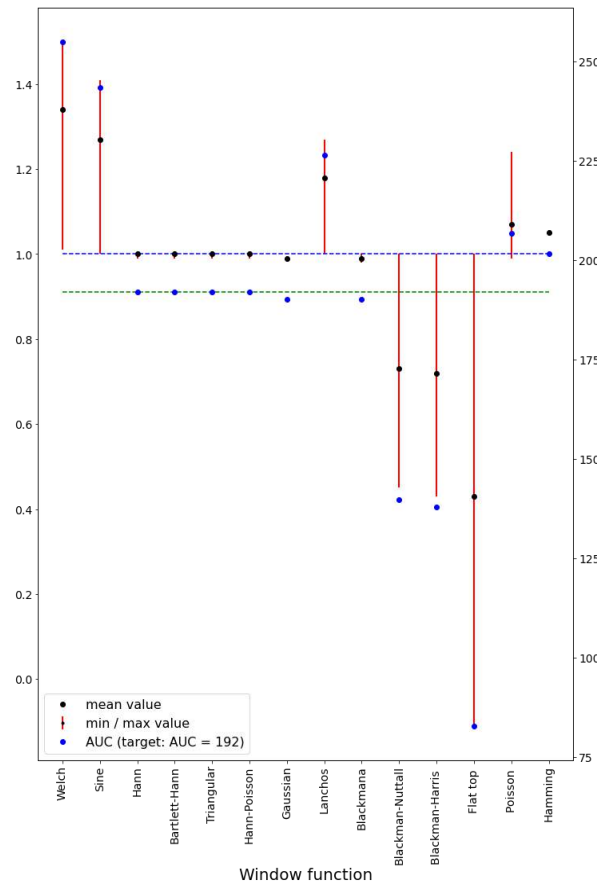
Figure 3. Presentation of the values of the average, minimum, and maximum values of the weights of the analyzed windows. It was assumed that the images overlapped in 50%. Numerical values were determined for the overlap area.

Table 2. Assessment of the quality of combining images with the use of windows. For methods marked with "*", histogram adjustment was applied before assessment.

Window Function\Metrics	MSE	RMSE	PSNR	UQI	SCC	SAM	SSIM	RASE	VIFP	NRMSE
Overlap	0.00	0.00	-	1.00	1.00	0.00	1.00	0.00	1.00	0.00
Hann $a_0 = 0.5$	0.42	0.64	51.89	1.00	1.00	0.01	1.00	0.35	1.00	0.01
Bartlett-Hann	3.66	1.91	42.49	1.00	0.91	0.01	1.00	101.21	0.98	0.02
Triangular	3.84	1.95	42.29	1.00	0.90	0.01	1.00	104.21	0.98	0.02
Hann-Poisson	3.42	1.85	42.78	0.99	0.92	0.01	1.00	96.91	0.98	0.02
Gaussian	92.79	9.63	28.46	0.99	0.89	0.08	0.99	401.44	0.89	0.08
Gaussian *	75.74	8.70	29.34	1.00	0.88	0.07	0.99	354.06	0.87	0.07
Lanchos	1288.29	35.89	17.03	0.90	0.85	0.30	0.88	1658.52	0.72	0.32
Lanchos *	863.22	29.38	18.77	0.95	0.81	0.24	0.88	1294.12	0.59	0.24
Blackman	16.68	4.08	35.91	1.00	0.90	0.02	1.00	207.04	0.97	0.03
Blackman *	3.23	1.80	43.03	1.00	0.89	0.01	1.00	87.88	0.97	0.01
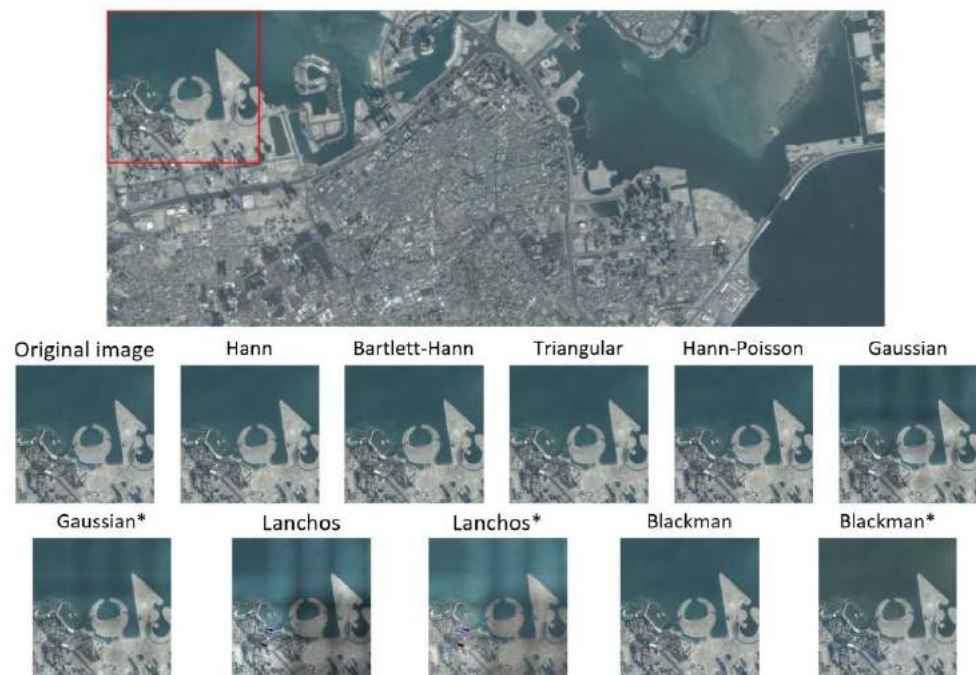
Figure 4. Sample image generated as a result of combining images with the use of selected windows. For methods marked with "*", histogram adjustment was applied before combining images after adjusting the histogram to the reference image.
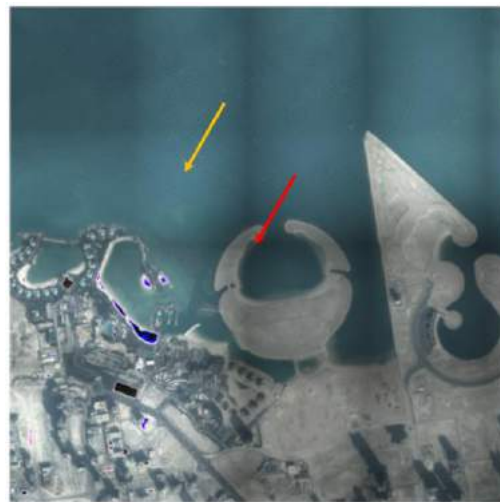
Figure 5. Fragment of the resulting image that was generated with the use of the Lanchos window. The lighter stripe of pixels that shows that the sum of weights is higher than 1 and is marked with the yellow arrow, while the red arrow indicates the stripe of darker pixels, where the sum of weights is lower than 1.

Considering the results presented above, the main tests will be conducted for the Hann, Bartlett-Hann, Triangular, and Hann Poisson windows.

3.4. Results

The aim of the main tests was to verify the windows determined in the preliminary phase of research and to determine the best level of overlap between component images. For the purposes of tests, a GAN model was prepared that makes it possible to increase the spatial resolution of the input images four times. To achieve it, the ESRGAN network was trained with the use of an own database consisting of low-resolution (LR) images and

corresponding high-resolution (HR) images. The quality of the combination of component images was assessed for images, whose spatial resolution was improved with the use of the trained network.

3.4.1. Database

For the purposes of these tests, a database was prepared that consisted of low-resolution (LR) images and corresponding high-resolution (HR) images. The task of the ESRGAN network was to improve the resolution of channels 2, 3, and 4 of multispectral images obtained by the World View 2 (WV2) satellite. This satellite captures panchromatic images with a spatial resolution of 0.5 m and multispectral (8-band) images with a resolution of 2 m. The database of low-resolution images was created with the use of multi-spectral images presenting the areas located in south-eastern, southern, and northern parts of Poland. The corresponding high-resolution images were prepared based on colored satellite images after traditional pansharpening.

The spatial resolution of the images was improved with the use of the Gram–Schmidt method [54–56]. The main reason for choosing this method was the fact that the color distortions are the lowest (in comparison to other methods). Figure 6 presents sample pairs of LR and HR images.

Figure 6. Sample pairs of (**A**) LR and (**B**) HR images that were used to train the ESRGAN network.

As a result of the operations described above, a database containing 29,500 LR images with the dimensions of 96×96 pixels and corresponding HR images with the dimensions of 384×384 pixels was created.

3.4.2. The ESRGAN Network

The Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) [41] are the most popular modification of the SRGAN networks. Their task is to estimate a

high-resolution image based on a low-resolution one. The ESRGAN model uses the low-resolution image and a deep convolutional network that contains residual blocks to estimate high-resolution images. It consists of two models: the generator and the discriminator. The task of the generator network is to improve the resolution of the input image, while the discriminator model evaluates the generated image and is used only during network training.

As it was mentioned before, the ESRGAN model is a modification of the SRGAN network. The batch normalization (BN) layers have been removed from its generator, and the basic block was replaced with a Residual-in-Residual Dense Block (RRDB), being a combination of a multi-level residual network and dense connections (Figure 7). The removal of the BN layers resulted in stable training and improved network capacity (the time required for training became much shorter), which is a result of the reduced computational complexity. In cases when the statistical data of the training and testing processes differ significantly, the BN layers tend to generate artefacts in the SR image. This phenomenon comes from the difference between the datasets that are used to calculate the average and the variance. During network training, they are calculated based on a certain batch of images, while at the stage of testing information from the whole dataset is used. Another modification of the generator network can be found in the implementation of RRDB blocks that have a residual structure. This solution makes it possible to increase the network capacity, which, in turn, improves its performance.
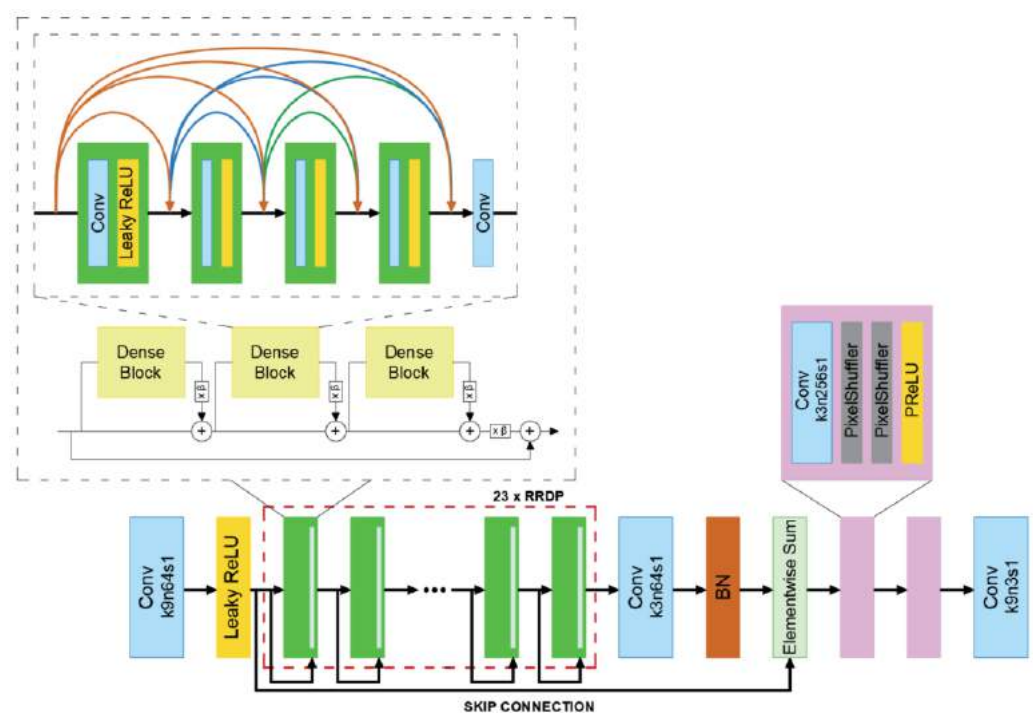
Figure 7. Model of the ESRGAN network generator (base on [41]).

The authors of ESRGAN also modified the discriminator, by replacing it with a relativistic discriminator. As opposed to the standard discriminator used in SRGAN, which estimates the probability that the evaluated image belongs to the set of *HR* images, the relativistic discriminator attempts to predict the probability that the true image *IHR* is relatively more realistic than the false image *ISR* (Equations (8)–(9)).

$$RMSE = \sqrt{\frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} [HR(n,m) - HR'(n,m)]^2}, \tag{8}$$

$$D\_Ra\left(I^{HR}, I^{SR}\right) = \sigma\left(C\left(I^{HR}\right) - E_{x_f}\left[C\left(I^{SR}\right)\right]\right) \tag{9}$$

where σ—sigmoid function, $C(x)$—output data of the generator, before applying the last activation function, $E[\cdot]$—the average of all false data in the mini-batch.

The losses of the generator (Equation (10)) and discriminator (Equation (11)) may be formulated as follows:

$$L_D^{Ra} = -E_{x_r}\left[log\left(D_{Ra}\left(I^{HR}, I^{SR}\right)\right)\right] - E_{x_r}\left[log\left(1 - D_{Ra}\left(I^{SR}, I^{HR}\right)\right)\right], \tag{10}$$

$$L_G^{Ra} = -E_{x_r}\left[\log\left(1 - D_{Ra}\left(I^{HR}, I^{SR}\right)\right)\right] - E_{x_r}\left[\log\left(D_{Ra}\left(I^{SR}, I^{HR}\right)\right)\right] \tag{11}$$

where l^{SR}—Perceptual loss function.

Another modification of the SRGAN model is the application of the perceptual loss before activation layers (instead of after them). This allows for an increase in the number of properties used to calculate $l_{VGGi,j}^{SR}$, which makes it possible to improve network performance. Additionally, it enables much better reconstruction of the brightness of *SR* images. As a result, the total loss of the generator may be presented in form of Equation (12).

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta E_{x_1}\|G(x_i) - y\|_1, \tag{12}$$

where L_{percep}—perceptual loss, λ, η—coefficients compensating various losses, $E_{x_1}\|G(x_i) - y\|_1$—(also denoted as L_1), distance between the *SR* and *HR* images.

The ESRGAN network is trained based on network interpolation, whose task is to remove the noise from the estimated SR images. It consists of the training of the *G_PSNR* network that is oriented on the *PSNR* and then the *G_GAN*. The network is obtained as a result of adjustments. The generator model is obtained as the interpolation of other models according to the equation below (Equation (13)).

$$\theta_G^{INTERP} = (1 - \alpha)\theta_G^{PSNR} + \alpha\theta_G^{GAN}, \tag{13}$$

where: α—interpolation parameter, $\alpha = [0, 1]$, θ_G^{INTERP}, $\theta_G^{PSNR}, \theta_G^{GAN}$—parameters of the G_{INTERP}, G_{PSNR}, G_{GAN} networks.

This modification enables the generating of results for any value of the α coefficient, reducing the presence of artefacts in the image. Secondly, it is possible to modify SR images without the need to re-train the model.

3.4.3. Network Training

The network was trained on an Nvidia TITAN RTX 24 GB graphics card, Intel Xeon Silver 4216 processor, and an Ubuntu 18.04 operating system. The initial parameters for the ESRGAN network training were those recommended by the authors of the solution: learning rate is initialized as 2×10^{-4}, decayed by a factor of 2 every 2×10^5 of mini-batch updates. The generator is trained using the loss function in (12) with $\lambda = 5{\cdot}10^{-3}$ and $\eta = 10^{-2}$. For optimization, we use Adam with $\beta1 = 0.9$, $\beta2 = 0.999$. The learning rate is set to 1×10^{-4} and halved at [50 k, 100 k, 200 k, 300 k] iterations [41]. The aim of introducing a change to the learning rate during network training is to improve the model's resistance to overtraining.

The application of the above approach resulted in the observed phenomenon of disappearing gradients after performing approximately 35,000 iterations, which are very easy to identify through a rapid increase of the LG loss (Figure 8).

As a result, it is necessary to reduce the learning values earlier in the network training process. Based on the conducted experiments, a modification of the learning rate parameter (Table 3) was proposed, in order to improve the resistance of the neural network to the unstable gradient syndrome. As a consequence of this operation, the training process becomes significantly longer.
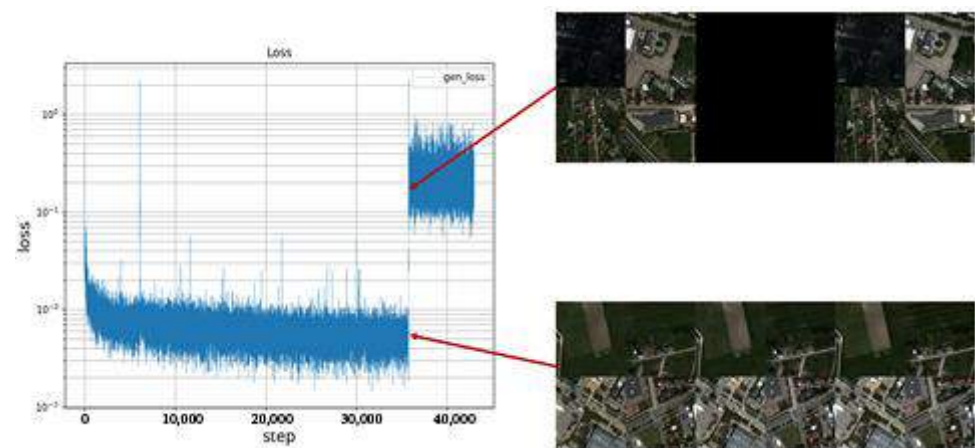
Figure 8. Sample pairs of LR and HR images that were used to train the ESRGAN network.

Table 3. Learning rate values used in network training.

Iterations	Learning Rate
35,000	2×10^{-4}
80,000	1×10^{-4}
80,000	5×10^{-5}
100,000	2×10^{-5}

3.4.4. Combining Images

At the main stage of the research on the possibility to combine SR images with the use of window functions, the authors also focused on determining the best degree of overlap between the combined images. For this purpose, five fragments of multi-spectral satellite scenes with dimensions not smaller than 900×900 pixels (LR) were selected. These images show urban areas and outskirts of cities, forests, agricultural areas, and a fragment of a wind farm. The reference high-resolution (HR) images selected for evaluation were the same images, whose spatial resolution was improved with the use of pansharpening with the Gram–Schmidt method (Figure 9).

Figure 9. Sample test image—fragment of a multi-spectral image captured by the World View-2 satellite, depicting the suburbs of the town of Radom: (**A**) low-resolution (LR) image (dimensions: 917×921 pixels), (**B**) high-resolution (reference) image obtained as a result of pansharpening with the Gram–Schmidt method (dimensions: 3667×3684 pixels).

For test purposes, the LR images were divided into smaller images with a resolution of 96×96 pixels. Ten sets of component images were prepared (for each of the scene

fragments), using various degrees of overlap (from 50% to 5% at 5% intervals). Before starting the tests, the spatial resolution of every component image was improved with the use of the trained ESRGAN network. Images prepared in this way were then combined with the use of the window functions selected in the preliminary research phase (Figure 2).

The operation described above resulted in creating 200 new image fragments (5 scene fragments × 10 degrees of overlap × 4 window functions = 200 images) of a spatial resolution improved with the use of the ESRGAN network. Each of the estimated images was then evaluated based on the metrics presented and described in Section III.A. The average value of the analyzed metrics for each image was calculated taking into account the classification based on the image combining method. The obtained results are presented in Appendix B. They reveal that the quality of the resulting image improves with the decreasing degree of overlap between images, and the maximum value is achieved for the overlap of 10%. For lower values of overlap, the quality of the combined image deteriorated. Additionally, one may notice that the images combined with the use of the triangular window achieve the best quality results for the overlap between images that is higher than 25%.

For an overlap lower than this threshold, the best evaluation results were achieved by images combined with the use of the Hann-Poisson window, although this method works best when the overlap between images does not exceed 15%. For further verification, the values of the SSIM and PSNR metrics were analyzed for the image presented in Figure 10.
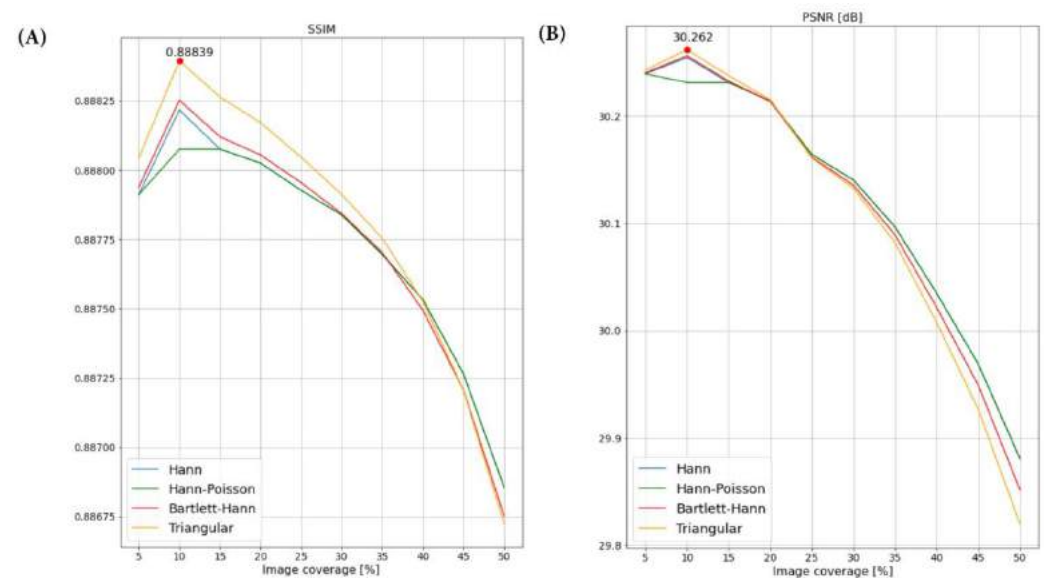
Figure 10. The values of the (**A**) SSIM and (**B**) PSNR metrics for the image presented in Figure 9.

The obtained results confirm that the window functions allow for the combination of image fragments whose spatial resolution was improved with the use of the generative adversarial network. Although one may notice that combining images with the use of the Triangular window resulted in the best values of evaluation metrics, the differences between the other analyzed window functions were small, e.g., for SSIM it was approximately 0.0002.

As for the comparison of quality of different resulting images, a significant shift of the SSIM and PSNR metrics along the y axis was noted for the analyzed images (Figure 11). The reason is, however, the quality of evaluating the SR images by GAN, not the method of combining images, which is confirmed by the shape of the curves being the interpolation of the results of the analyzed metrics. Moreover, the analysis of the obtained results clearly demonstrates that the application of a large overlap between the combined results has a negative influence on the values of the evaluation metrics. At the same time, using an overlap of approximately 10% of the image enables the best estimation of the super resolution (SR) image.
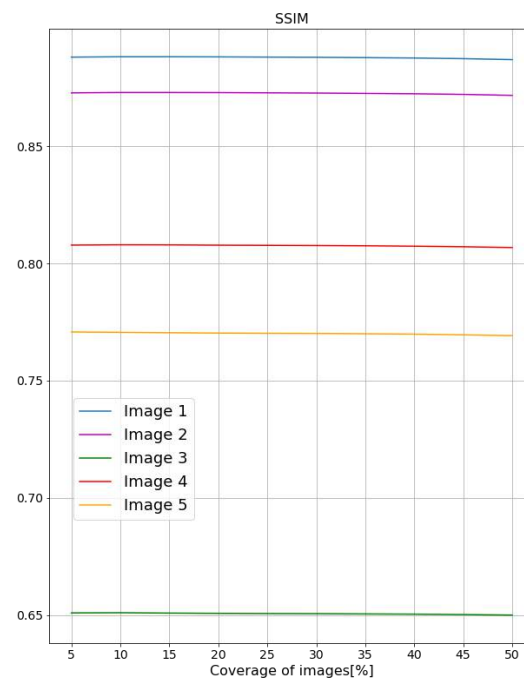
Figure 11. SSIM values for the analyzed images.

4. Discussion

This article presents a new methodology for improving the spatial resolution of whole satellite scenes with the use of deep learning methods.

In this solution, the input low-resolution image is divided into smaller fragments with dimensions equal to the dimensions of the input data of the neural network. Based on the tests presented above, the recommended overlap between images should be approximately 10%. Then, the spatial resolution of all LR images is improved with the use of any deep learning method (the authors used the ESRGAN network). This stage is followed by using window functions to combine the SR images of a higher resolution created using the methods described above. If the overlap between images equals 10%, it is recommended to use the Triangular window. For an overlap exceeding 20%, the authors recommend using the Hann-Poisson window instead. At the same time, considering the results of the main tests, the degree of overlap between images has a stronger influence on the quality of the resulting image than the window function used to combine images. Therefore, each of the window functions that were verified in the main phase of research, i.e., Hann, Hann-Poisson, Bartlett-Hann, and Triangular may be applied to combine the estimated SR images.

At the same time, window functions may be successfully implemented to combine other images that result from image translation operations, for example the activity of the U-Net network or conditional generative adversarial networks (CGAN). The illustration below shows an example of shadow detection in panchromatic images with the use of CGAN. Figure 12 shows two images with 5%, 10%, and 50% degrees of overlap, respectively, and the result of their combination with the use of the Hann-Poisson window. The resulting images were generated based on the assumption that, if the probability of defining the resulting pixel as a shadow is lower than 70%, then it will not be assigned to the "shadow" class. Such a solution allows for the elimination of errors that may appear at the borders of the image, where only parts of the object are often visible.
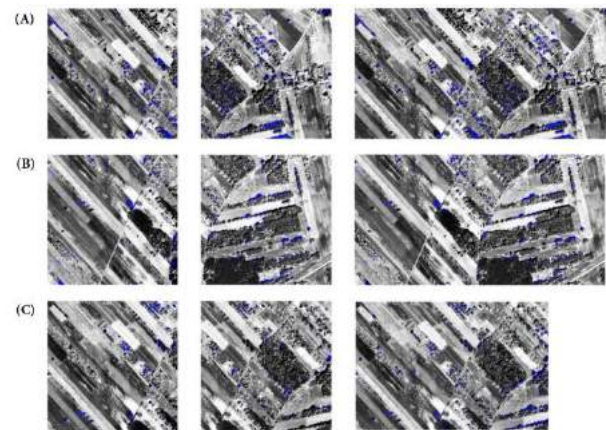
Figure 12. An example of the application of window function to combine shadow masks that were detected with the use of the UNet network. The images shown have an overlap of: (**A**) 5%, (**B**) 10%, (**C**) 50%. For panchromatic images (where histogram equalization was applied) window functions were not used.

Unfortunately, the literature review performed by the authors revealed that the results of the application of deep learning algorithms are presented for small images. One of the numerous examples is the solution presented by Xiaoyu Dong [57], which makes it possible to improve the spatial resolution of images with dimensions of 48×48 pixels. A similar problem may be encountered in the segmentation of images. Binge Cui et al. presented a method that enables Sea-Land Segmentation of images with dimensions of 256×256 pixels [58].

On the other hand, the methodology presented in this paper could make it possible to apply solutions proposed by scientists to satellite images of any dimensions.

5. Conclusions

The conducted research revealed that it is possible to improve the spatial resolution of whole satellite image scenes with the use of deep learning algorithms. However, as those algorithms require large computational power, processing whole satellite images is very difficult, and sometimes even impossible. This problem may be solved by the methodology presented above, which enables the processing of digital images of any dimension. Moreover, this solution may be applied to combine images generated as a result of image translation operations, including segmentation. At the same time, using an overlap between images of approximately 10% allows for a significant shortening of the duration of the spatial resolution improvement process, which results from the reduced number of necessary operations (an example is also presented in Figure 11). Additionally, this approach also enables the application of trained neural network models regardless of their input size parameters.

The conducted experiments also demonstrate that the ESRGAN network is not completely successful in improving the spatial resolution of satellite imagery, and the estimated images contain multiple errors, which is particularly noticeable during visual analysis, e.g., of roofs of houses. This is also confirmed by the values of metrics presented in Appendix B.

Considering the obtained results, one may conclude that it is possible to improve the spatial resolution of whole satellite scenes. However, this requires the modification of existing deep learning models or the development of completely new solutions.

Appendix A

Table A1. Windows formulas that were used for the calculations.

WINDOW	Formula	Min	Max	Mean	Surface for Image 384 pix	Figure (the Modification, When in Non-Overlapping Areas, i.e., on the External Edges, the Value of the Window Function Equals 1) Image Overlap of 50% Was Assumed.
Welch	$w[n] = 1 - \left(\dfrac{n - \frac{N}{2}}{\frac{N}{2}}\right)^2$	1.01	1.50	1.34	254.99	
Sine	$w[n] = \sin\left(\frac{\pi n}{N}\right)$	1.00	1.41	1.27	243.46	
Hann	$w[n] = a_0 * \left[1 - \cos\left(\frac{2\pi n}{N}\right)\right]$ $a_0 = 0.5$	0.9 (9)	1.00	1.00	192	
Bartlett-Hann	$w[n] = a_0 - a_1\left|\frac{n}{N} - \frac{1}{2}\right| - a_2 \cos\left(\frac{2\pi n}{N}\right)$ $a_0 = 0.62;\ a_1 = 0.48;\ a_2 = 0.38$	0.9 (9)	1.00	1.00	192	

Table A1. *Cont.*

WINDOW	Formula	Min	Max	Mean	Surface for Image 384 pix	Figure (the Modification, When in Non-Overlapping Areas, i.e., on the External Edges, the Value of the Window Function Equals 1) Image Overlap of 50% Was Assumed.
Triangular	$w[n] = 1 - \left\| \frac{n - \frac{N}{2}}{\frac{L}{2}} \right\|,\ 0 \leq n \leq N$	0.9 (9)	1.00	1.00	192	
Hann-Poisson	$w[n] = \frac{1}{2}\left(1 - \cos\left(\frac{2\pi n}{N}\right)\right) e^{\frac{-\alpha\|N - 2n\|}{N}}$	0.9 (9)	1.00	1.00	192	
Gaussian	$w[n] = \exp\left(-\frac{1}{2}\left(\frac{n - \frac{N}{2}}{\frac{\sigma N}{2}}\right)^2\right)$ $0 \leq n \leq N,\ \sigma \leq 0.5,$ Selected: $\sigma = 0.4$	0.92	1.05	0.99	190.11	
Lanchos	$w[n] = sinc\left(\frac{2n}{N} - 1\right)$	1.0	1.27	1.18	226.36	
Blackmana	$w[n] = a_0 - a_1 \cos\left(\frac{2\pi n}{N}\right) + a_2 \cos\left(\frac{4\pi n}{N}\right)$ $a_0 = \frac{1-\alpha}{2};\ a_1 = \frac{1}{2};\ a_2 = \frac{\alpha}{2}$ Selected: $\alpha = 0.01$	0.98	1.00	0.99	190.08	

Table A1. *Cont.*

WINDOW	Formula	Min	Max	Mean	Surface for Image 384 pix	Figure (the Modification, When in Non-Overlapping Areas, i.e., on the External Edges, the Value of the Window Function Equals 1) Image Overlap of 50% Was Assumed.
Blackman-Nuttall	$w[n] =$ $a_0 - a_1 \cos\left(\frac{2\pi n}{N}\right) +$ $+a_2 \cos\left(\frac{4\pi n}{N}\right) -$ $a_3 \cos\left(\frac{6\pi n}{N}\right)$ $a_0 = 0.3635819;$ $a_1 = 0.4891775;$ $a_2 = 0.1365995;$ $a_3 = 0.0106411;$	0.45	1.00	0.73	139.62	
Blackman–Harris window	$w[n] =$ $a_0 - a_1 \cos\left(\frac{2\pi n}{N}\right) +$ $+a_2 \cos\left(\frac{4\pi n}{N}\right) -$ $a_3 \cos\left(\frac{6\pi n}{N}\right)$ $a_0 = 0.35875;$ $a_1 = 0.48829;$ $a_2 = 0.14128;$ $a_3 = 0.01168;$	0.43	1.00	0.72	137.76	
Flat top window	$w[n] =$ $a_0 - a_1 \cos\left(\frac{2\pi n}{N}\right) +$ $+a_2 \cos\left(\frac{4\pi n}{N}\right) +$ $-a_3 \cos\left(\frac{6\pi n}{N}\right) +$ $a_4 \cos\left(\frac{8\pi n}{N}\right)$ $a_0 = 0.21557895;$ $a_1 = 0.41663158;$ $a_2 = 0.277263158;$ $a_3 = 0.083578947;$ $a_4 = 0.006947368;$	−0.11	1.0	0.43	82.78	
Exponential or Poisson window	$w[n] = e^{-\left|n - \frac{N}{2}\right| \cdot \frac{1}{\tau}}$ $\tau = \frac{N}{2} \cdot \frac{8.69}{D}$ $\text{Selected}: D = 12,166$	0.99	1.24	1.07	206.65	
Hamming	$w[n] =$ $\alpha - (1 - \alpha) \cos\left(\frac{2\pi n}{N}\right)$ $\text{Recommended}: \alpha = 0.53836$ $\text{Selected}: \alpha = 0.525$	1.05	1.05	1.05	201.60	

Appendix B

Table A2. Results.
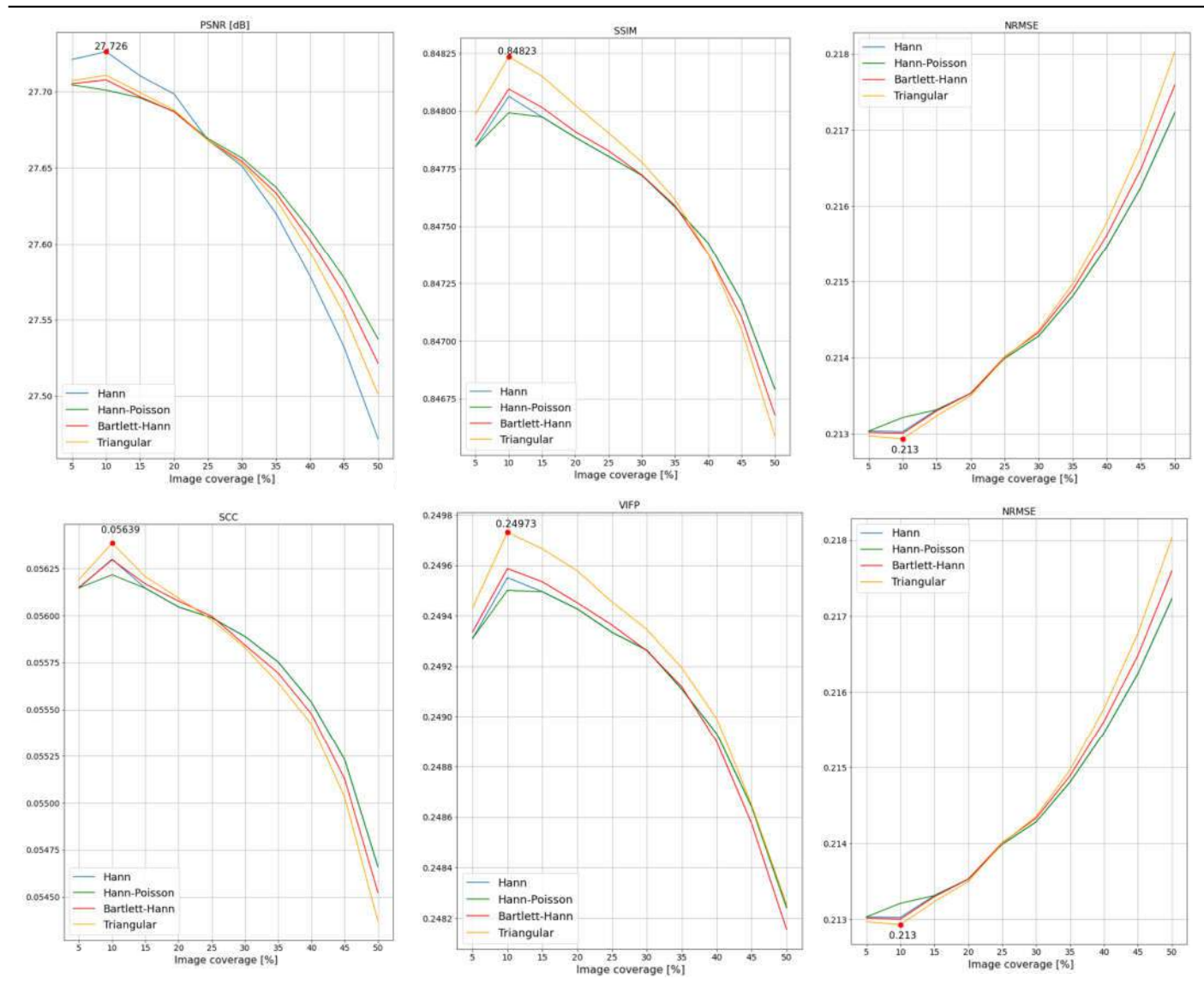
References

1. Liu, Y.; Li, Q.; Yuan, Y.; Du, Q.; Wang, Q. ABNet: Adaptive Balanced Network for Multiscale Object Detection in Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [CrossRef]
2. Cao, Z.; Fang, W.; Song, Y.; He, L.; Song, C.; Xu, Z. DNN-Based Peak Sequence Classification CFAR Detection Algorithm for High-Resolution FMCW Radar. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [CrossRef]
3. Cui, Y.; Hou, B.; Wu, Q.; Ren, B.; Wang, S.; Jiao, L. Remote Sensing Object Tracking With Deep Reinforcement Learning Under Occlusion. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [CrossRef]
4. Oveis, A.H.; Giusti, E.; Ghio, S.; Martorella, M. A Survey on the Applications of Convolutional Neural Networks for Synthetic Aperture Radar: Recent Advances. *IEEE Aerosp. Electron. Syst. Mag.* **2022**, *37*, 18–42. [CrossRef]
5. Singh, A.; Kalke, H.; Loewen, M.; Ray, N. River Ice Segmentation With Deep Learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7570–7579. [CrossRef]
6. Saha, S.; Mou, L.; Qiu, C.; Zhu, X.X.; Bovolo, F.; Bruzzone, L. Unsupervised Deep Joint Segmentation of Multitemporal High-Resolution Images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8780–8792. [CrossRef]
7. Zhang, B.; Chen, T.; Wang, B. Curriculum-Style Local-to-Global Adaptation for Cross-Domain Remote Sensing Image Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]

8.   Mei, S.; Jiang, R.; Li, X.; Du, Q. Spatial and Spectral Joint Super-Resolution Using Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4590–4603. [CrossRef]

9.   Song, H.; Huang, B.; Liu, Q.; Zhang, K. Improving the Spatial Resolution of Landsat TM/ETM+ Through Fusion With SPOT5 Images via Learning-Based Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1195–1204. [CrossRef]

10.   Lima, P.; Steger, S.; Glade, T.; Murillo-García, F.G. Literature Review and Bibliometric Analysis on Data-Driven Assessment of Landslide Susceptibility. *J. Mt. Sci.* **2022**, *19*, 1670–1698. [CrossRef]

11.   Xia, D.; Tang, H.; Sun, S.; Tang, C.; Zhang, B. Landslide Susceptibility Mapping Based on the Germinal Center Optimization Algorithm and Support Vector Classification. *Remote Sens.* **2022**, *14*, 2707. [CrossRef]

12.   Wang, L.; Scott, K.A.; Xu, L.; Clausi, D.A. Sea Ice Concentration Estimation During Melt From Dual-Pol SAR Scenes Using Deep Convolutional Neural Networks: A Case Study. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4524–4533. [CrossRef]

13.   Wang, L.; Scott, K.A.; Clausi, D.A. Sea Ice Concentration Estimation during Freeze-Up from SAR Imagery Using a Convolutional Neural Network. *Remote Sens.* **2017**, *9*, 408. [CrossRef]

14.   Cooke, C.L.V.; Scott, K.A. Estimating Sea Ice Concentration From SAR: Training Convolutional Neural Networks With Passive Microwave Data. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4735–4747. [CrossRef]

15.   Scarpa, G.; Gargiulo, M.; Mazza, A.; Gaetano, R. A CNN-Based Fusion Method for Feature Extraction from Sentinel Data. *Remote Sens.* **2018**, *10*, 236. [CrossRef]

16.   Oveis, A.H.; Giusti, E.; Ghio, S.; Martorella, M. CNN for Radial Velocity and Range Components Estimation of Ground Moving Targets in SAR. In Proceedings of the 2021 IEEE Radar Conference (RadarConf21), Atlanta, GA, USA, 7–14 May 2021; pp. 1–6.

17.   Wang, J.; Lu, C.; Jiang, W. Simultaneous Ship Detection and Orientation Estimation in SAR Images Based on Attention Module and Angle Regression. *Sensors* **2018**, *18*, 2851. [CrossRef]

18.   Kuuste, H.; Eenmäe, T.; Allik, V.; Agu, A.; Vendt, R.; Ansko, I.; Laizans, K.; Sünter, I.; Lätt, S.; Noorma, M. Imaging System for Nanosatellite Proximity Operations. *Proc. Est. Acad. Sci.* **2014**, *63*, 250. [CrossRef]

19.   Blommaert, J.; Delauré, B.; Livens, S.; Nuyts, D.; Moreau, V.; Callut, E.; Habay, G.; Vanhoof, K.; Caubo, M.; Vandenbussche, J.; et al. CHIEM: A New Compact Camera for Hyperspectral Imaging. 2017. Available online: https://www.researchgate.net/publication/321214165_CHIEM_A_new_compact_camera_for_hyperspectral_imaging (accessed on 18 October 2022).

20.   Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.

21.   Karwowska, K.; Wierzbicki, D. Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3292–3312. [CrossRef]

22.   Lu, T.; Wang, J.; Zhang, Y.; Wang, Z.; Jiang, J. Satellite Image Super-Resolution via Multi-Scale Residual Deep Neural Network. *Remote Sens.* **2019**, *11*, 1588. [CrossRef]

23.   Schmidt, R. Multiple Emitter Location and Signal Parameter Estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [CrossRef]

24.   Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.

25.   Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.

26.   He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.

27.   Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv* **2015**, arXiv:1512.00567.

28.   Han, J.; Zhang, D.; Cheng, G.; Guo, L.; Ren, J. Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3325–3337. [CrossRef]

29.   Wang, Z.; Zhang, Y.; Yu, Y.; Zhang, L.; Min, J.; Lai, G. Prior-Information Auxiliary Module: An Injector to a Deep Learning Bridge Detection Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6270–6278. [CrossRef]

30.   Yu, D.; Ji, S. A New Spatial-Oriented Object Detection Framework for Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [CrossRef]

31.   Kemker, R.; Luu, R.; Kanan, C. Low-Shot Learning for the Semantic Segmentation of Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6214–6223. [CrossRef]

32.   Vinayaraj, P.; Sugimoto, R.; Nakamura, R.; Yamaguchi, Y. Transfer Learning With CNNs for Segmentation of PALSAR-2 Power Decomposition Components. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6352–6361. [CrossRef]

33.   Šćepanović, S.; Antropov, O.; Laurila, P.; Rauste, Y.; Ignatenko, V.; Praks, J. Wide-Area Land Cover Mapping With Sentinel-1 Imagery Using Deep Learning Semantic Segmentation Models. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10357–10374. [CrossRef]

34.   Feng, Y.; Sun, X.; Diao, W.; Li, J.; Gao, X.; Fu, K. Continual Learning with Structured Inheritance for Semantic Segmentation in Aerial Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [CrossRef]

35.   Zuo, Z.; Li, Y. A SAR-to-Optical Image Translation Method Based on PIX2PIX. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Kuala Lumpur, Malaysia, 17–22 July 2021; pp. 3026–3029.

36.   Chen, X.; Chen, S.; Xu, T.; Yin, B.; Peng, J.; Mei, X.; Li, H. SMAPGAN: Generative Adversarial Network-Based Semisupervised Styled Map Tile Generation Method. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4388–4406. [CrossRef]

37. Kaiser, P.; Wegner, J.D.; Lucchi, A.; Jaggi, M.; Hofmann, T.; Schindler, K. Learning Aerial Image Segmentation from Online Maps. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6054–6068. [CrossRef]

38. Fu, Y.; Liang, S.; Chen, D.; Chen, Z. Translation of Aerial Image Into Digital Map via Discriminative Segmentation and Creative Generation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [CrossRef]

39. Vandal, T.J.; McDuff, D.; Wang, W.; Duffy, K.; Michaelis, A.; Nemani, R.R. Spectral Synthesis for Geostationary Satellite-to-Satellite Translation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [CrossRef]

40. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *arXiv* **2017**, arXiv:1609.04802.

41. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Loy, C.C.; Qiao, Y.; Tang, X. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. *arXiv* **2018**, arXiv:1809.00219.

42. Choi, J.-S.; Kim, Y.; Kim, M. S3: A Spectral-Spatial Structure Loss for Pan-Sharpening Networks. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 829–833. [CrossRef]

43. Ji, H.; Gao, Z.; Mei, T.; Ramesh, B. Vehicle Detection in Remote Sensing Images Leveraging on Simultaneous Super-Resolution. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 676–680. [CrossRef]

44. Tang, W.; Deng, C.; Han, Y.; Huang, Y.; Zhao, B. SRARNet: A Unified Framework for Joint Superresolution and Aircraft Recognition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 327–336. [CrossRef]

45. Shen, C.; Ji, X.; Miao, C. Real-Time Image Stitching with Convolutional Neural Networks. In Proceedings of the 2019 IEEE International Conference on Real-time Computing and Robotics (RCAR), Irkutsk, Russia, 4–9 August 2019; pp. 192–197.

46. He, X.; He, L.; Li, X. Image Stitching via Convolutional Neural Network. In Proceedings of the 2021 7th International Conference on Computer and Communications (ICCC), Chengdu, China, 9–12 December 2021; pp. 709–713.

47. Lin, M.; Liu, T.; Li, Y.; Miao, X.; He, C. Image Stitching by Disparity-Guided Multi-Plane Alignment. *Signal Process.* **2022**, *197*, 108534. [CrossRef]

48. Pielawski, N.; Wählby, C. Introducing Hann Windows for Reducing Edge-Effects in Patch-Based Image Segmentation. *PLoS ONE* **2020**, *15*, e0229839. [CrossRef] [PubMed]

49. Keelan, B. *Handbook of Image Quality: Characterization and Prediction*; CRC Press: Boca Raton, FL, USA, 2002; ISBN 978-0-429-22280-1.

50. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

51. Wang, Z.; Bovik, A.C. A Universal Image Quality Index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]

52. Goetz, A.; Boardman, W.; Yunas, R. Discrimination among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm. In *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop*; AVIRIS Workshop: Pasadena, CA, USA, 1992.

53. Sheikh, H.R.; Bovik, A.C. Image Information and Visual Quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444. [CrossRef]

54. Prabhu, K.M.M. *Window Functions and Their Applications in Signal Processing*; CRC Press: Boca Raton, FL, USA, 2017; ISBN 978-1-315-21638-6.

55. Li, H.; Zhang, Y.; Gao, Y.; Yue, S. Using Guided Filtering to Improve Gram-Schmidt Based Pansharpening Method for GeoEye-1 Satellite Images. In Proceedings of the 4th International Conference on Information Systems and Computing Technology, Shanghai, China, 22–23 December 2016; pp. 33–37.

56. Sekrecka, A.; Kedzierski, M. Integration of Satellite Data with High Resolution Ratio: Improvement of Spectral Quality with Preserving Spatial Details. *Sensors* **2018**, *18*, 4418. [CrossRef] [PubMed]

57. Dong, X.; Sun, X.; Jia, X.; Xi, Z.; Gao, L.; Zhang, B. Remote Sensing Image Super-Resolution Using Novel Dense-Sampling Networks. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1618–1633. [CrossRef]

58. Cui, B.; Jing, W.; Huang, L.; Li, Z.; Lu, Y. SANet: A Sea–Land Segmentation Network Via Adaptive Multiscale Feature Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 116–126. [CrossRef]

MCWESRGAN: Improving Enhanced Super-Resolution Generative Adversarial Network for Satellite Images

Kinga Karwowska [ID] and Damian Wierzbicki [ID]

Abstract—With the dynamic technological development, we are witnessing a major progress in solutions that allow for the observation of Earth's surface. Small satellites have a significant drawback. Due to their limitations, the installed optic systems are not perfect. As a result, the quality of the obtained images is lower, including lower resolution, although the satellites move on the Low Earth Orbit. In the case of images lacking a high-resolution counterpart, the spatial resolution of the imagery can be improved using single-image super-resolution algorithms. In this article, we present an SISR solution based on a new network called MCWESRGAN, which is a modification of the popular ESRGAN network. We propose a novel strategy that introduces a multi-column discriminator model. The generator model is trained using Wasserstein loss. The introduced modifications enable a tenfold reduction in the training time of the network. The proposed algorithm is verified using images obtained from space, aerial imagery, and the Dataset for Object deTection in Aerial Images (DOTA) database. A set of evaluation methods for super-resolution (SR) images is proposed to verify the results. These evaluation methods indicate areas that are poorly estimated by the algorithm. Furthermore, as part of the conducted experiments, an absolute assessment method for interpretational potential based on the power spectral density of the image (PSD) is proposed, allowing for determining the magnitude of interpretational improvement after applying resolution enhancement algorithms. The conducted research demonstrates that we achieve better qualitative and quantitative results than classical ESRGAN methods and other state-of-the-art (SOTA) approaches.

Index Terms—Convolutional neural networks, deep learning, enhanced super-resolution generative adversarial network (ESRGAN), neural networks, power spectral density (PSD), single-image super-resolution (SISR), super resolution (SR).

I. INTRODUCTION

IN RECENT years, we have been witnessing a dynamic development in space technology, and the possibilities to observe the Earth have been improving every year. The obtained data allow us to conduct analyses that may be applied in numerous fields of science, including monitoring changes [1], [2], [3], [4], [5], geospatial analysis [6], [7], automated object detection [8], [9], [10], [11], [12], [13], or monitoring havens [14], [15], [16].

The possibility to use the given imagery is determined by its resolution. High temporal resolution enables, for example, conducting the analyses of change detection. Spectral analysis determines the scope of potential remote sensing analyses, including the analyses of land coverage [17], [18], [19], [20].

Another quite important parameter is a spatial resolution that informs about ground sampling distance (GSD), i.e., the distance between two subsequent centers of pixels measured on the surface of the Earth. There are numerous methods used to improve the spatial resolution of an image [21], [22], [23], [24]. One of them is pansharpening, which combines the high resolution of a panchromatic image with the high resolution of a multi-spectral image [25], [26], [27], [28], [29]. This operation results in generating a high-resolution multispectral image. However, the pansharpening method may be used only provided that there is an existing high-resolution (HR) image. Unfortunately, not all satellites acquire imagery in the panchromatic and multispectral range. In addition, the problem also tends to appear as a result of the attempts to improve the resolution of imagery acquired by nanosatellites. Nano- and minisatellites are much smaller, and the matrices and telescopes that they carry are imperfect. As a result, the acquired images have much lower spatial resolution. Apart from that, the difference between the spatial resolution of the panchromatic and multispectral images is small, e.g., for the satellites from the SkySat-3 - SkySat-15 constellation it is only 0.16 m. This problem intensifies when it is necessary to increase the spatial resolution of an image that does not have a higher resolution corresponding image (e.g., a panchromatic image or a sequence of images). One of the solutions is the application of digital image processing methods. One of them is the Super-Resolution Variable-Pixel Linear Reconstruction algorithm [30], or the quincunx sampling mode named Très Haute Résolution that is applied for processing the imagery acquired by SPOT5 [31], [32].

Solutions that apply digital image processing are rather seldom used to improve the spectral resolution of satellite imagery due to the high requirements of acquiring satellite scenes or due to the need to acquire multiple scenes that present the same area by more than one satellite at a similar time. Another solution to

this problem is the application of algorithms that employ deep neural networks. This group of solutions uses only a trained network model and low-resolution (LR) image to improve the spatial resolution.

In order to solve the problems with improving the satellite imagery based on a LR image, we propose a modification to the enhanced super-resolution generative adversarial network (ES-RGAN) architecture [33]—multicolumn Wasserstein enhanced super-resolution generative adversarial networks (MCWESR-GAN). It will enable to train of the neural network over 10 times faster and, at the same time, to improve the quality of the generated images.

To address these above issues, we made the following improvements.

1) We employed a multicolumn discriminator, which allowed for a better evaluation of SR images generated by the generator, resulting in up to a tenfold acceleration of the model training process.

2) We used Wasserstein loss, which allowed for better discrimination between HR and SR images by the discriminator, consequently enabling improved control over the generator. This modification significantly reduces the issue of vanishing gradients and accelerates GAN training.

3) We paid attention to the quality of training databases. In the case of training single-image super-resolution (SISR) models dedicated to satellite images, it is recommended to create diverse databases, e.g., by using images obtained at different times of the day by other satellites. By taking care of this aspect, we can eliminate the occurrence of vanishing or exploding gradients.

4) We introduced a novel method for evaluating generated SR images based on the power spectral density (PSD) of the image. This analysis allows us to determine the ground resolved distance (GRD) parameter and, consequently, the interpretation possibilities of SR images.

5) We conducted research and assessed the feasibility of the proposed method in this article. The experiments were conducted on three sets of databases. To objectively evaluate the generated SR images, we employed quality metrics commonly used in the remote sensing and computer vision domains. In addition, we proposed two additional evaluation methods. The first method allows for a local assessment of the images, while the second method enables an evaluation of the interpretational potential based on the PSD of the image.

This article consists of the following parts: Section II presents an overview of methods to improve the spatial resolution of satellite imagery using deep neural networks. A description of the proposed method is provided in Section III. Section IV describes the research methodology and presents the results of the work. Section V contains a discussion of the results. Finally, Section VI concludes this article.

II. RELATED WORKS

Research on the methods to improve the spatial resolution of satellite imagery with the use of a single image (Single Image Super Resolution—SISR) has been conducted for years [34], [35], [36], [37]. However, the methods proposed by scientists do not allow for solving the problem of improving the spatial resolution of those satellite images that do not have a corresponding HR image. The main reason for this is the insufficient computational capacity of the workstations and the restrictions on free access to satellite data. However, the technological progress that took place in recent years enables us to solve these problems. Modern workstations enable the processing of very large imagery databases in a short time. In addition, the application of graphic processors enables the training and implementation of algorithms that employ deep neural networks.

Solutions based on convolutional networks may be applied in the tasks of detection [38], [39], [40], [41], classification [42], [43], segmentation [44], [45], [46], [47], [48], or improving the spatial resolution of the acquired images [49], [50], [51], [52].

This article will focus on the last of these applications, i.e., improving the spatial resolution of digital images. In this group of methods, we may distinguish solutions that are based on classical convolutional networks and solutions that employ generative adversarial networks (GANs). Although both these methods are based on a similar idea (the solutions recreate a HR image based on a neural network model), they differ significantly in terms of the architecture of the artificial neural network and the method of training. The first group of solutions includes models that are based on convolutional layers. It includes models that employ residual blocks or multicolumn networks, where each of the branches is responsible for improving the spatial resolution, and at the final stage of the operation of the model the results obtained on every branch are combined.

An example of a SISR algorithm is the single image super-resolution diffusion probabilistic model (SRDiff), which is designed to address the issues of over-smoothing and mode collapse [53]. SRDiff is a diffusion-based model for SISR, optimized using a variance-corrected likelihood. Another solution is the efficient super-resolution transformer model, which achieves significant computational cost reduction by combining the lightweight convolutional neural network (CNN) backbone and a lightweight transformer backbone [54].

Another solution employs the GAN described by Goodfellow [55]. The models of these networks find applications in many areas: image synthesis, texture synthesis, object detection, vision, natural language processing (NLP), music, and SR [56]. In the case of GAN models responsible for SR, apart from the model that is responsible for improving the resolution of the LR input image is the neural network that deals with the assessment of the estimated images. The application of two networks (the generator and the discriminator) significantly accelerates the process of training the generator (which is responsible for improving the resolution of LR images). Moreover, the final super resolution (SR) images are characterized by much higher quality than

those acquired with the use of other SISR methods. However, these resolutions also have some disadvantages. Due to the fact that two neural network models are employed, GAN networks require quite large computational power. In addition, during the training of the network, one may encounter the problem of unstable gradient (in particular if low-diversity databases are used, e.g., fragments of satellite images). In spite of these difficulties, an increasing number of algorithms are proposed to improve the spatial resolution with the use of GAN.

An example is the cycle-in-cycle GANs, where the image resolution enhancement occurs in three stages. In the first stage, noise is removed from the LR image. Next, using a generator model, the image's resolution is increased. Finally, in the third stage, the two modules are fine-tuned in an end-to-end manner to obtain the HR output [57]. Another solution, the super-resolution generative adversarial networks with Ranker, is aimed at optimizing the generator in the direction of perceptual metrics by training a Ranker and introducing a novel rank-content loss to optimize the perceptual quality [58].

However, one of the most popular solutions is the SRGAN algorithm [59]. In this solution, the SR image is created as a result of the application of a series of operations. The first of them is blurring the LR image with the use of a Gaussian filter. Then the image is decimated with the use of the r-sampling coefficient. The neural network model is trained based on the strategy of maximizing the minimum profit, which is based on a theoretical game scenario, where images estimated by the generator compete with the original HR images in the discriminator model.

The SRGAN model has contributed to many modifications that enable to improve multispectral and hyperspectral images. Still, the most popular modification of the SRGAN model is the ESRGAN model [33]. The authors introduced several changes. The most important ones include the modification of the generator. The scientists removed the batch normalization (BN) layers from the model. In addition, the basic block was replaced with the residual-in-residual dense block (RRDB), which is a combination of a multilevel residual network and dense connections. The introduced modifications improved the stability of training, while the removal of the BN layers allowed reducing the number of trained parameters, which, in turn, allowed accelerating network training. Another important change was the modification of the discriminator.

The authors replaced the standard discriminator (used in the SRGAN) with a relativistic discriminator.

The task of this discriminator is to estimate the probability that image I^{HR} is relatively more realistic than the false image I^{SR}. The introduced changes have significantly improved the quality of SR images, e.g., for the database Set5-4x [60] the value of the structural similarity index (SSIM) increased from 0.847 to 0.901 [33].

The authors of numerous solutions that employ GAN networks used residual links. This solution allowed for a significant reduction of the phenomenon of vanishing gradients and accelerated the network training process. An example of the application of residual links is the algorithm proposed by Courtrai et al. [61]
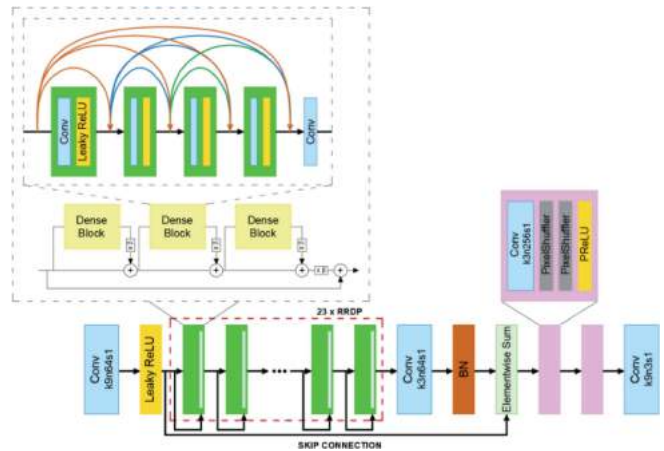
Fig. 1. Model of the ESRGAN generator (based on [33]).

that enables to improve the spatial resolution of small objects that are present in aerial and satellite photos.

The algorithms to improve resolution are developed mainly based on databases of digital images that are not images acquired at the aerial or satellite altitudes, e.g., Set5-4x [60], Set14-4x [62], BSD100- 4x [63], URBAN100- 4x [64], FFHQ 256x256-4x [65], FFHQ 512x512- 4x [65], and FFHQ 1024x1024- 4x [65].

Nevertheless, the developed methods are implemented to improve the spatial resolution of satellite imagery.

III. PROPOSED METHOD

In order to improve the quality of SR images, the proposed solution involved a modification of the enhanced super-resolution generative adversarial networks (ESRGANs) architecture published in 2018 by Xintao Wang et al. [33]. The introduced modifications mainly refer to the part of the model that is responsible for the assessment of SR images.

The SR images are generated based on the original generator of the ESRGAN [33]. The generator of the ESRGAN is a modification of the SRGAN generator, where BN layers have been removed. In addition, basic blocks have been replaced with RRDBs, which combine a multilevel residual network with dense connections (see Fig. 1). By removing the BN layers, the authors of the solution observed stable training and improved network efficiency (the training process is significantly faster), which resulted from reduced computational complexity. The second modification of the generator network is the implementation of RRDB blocks. These blocks have a residual structure, where residual learning occurs on the main branch of the network (although they can be utilized at different levels). This solution increases network throughput, leading to improved performance [33].

The first of the modifications consisted of introducing a multicolumn discriminator. The proposed solution contains two branches, where each of them creates a classifier that uses the convolutional layers. The first of these classifiers has the same structure of layers as the discriminator of the SRGAN
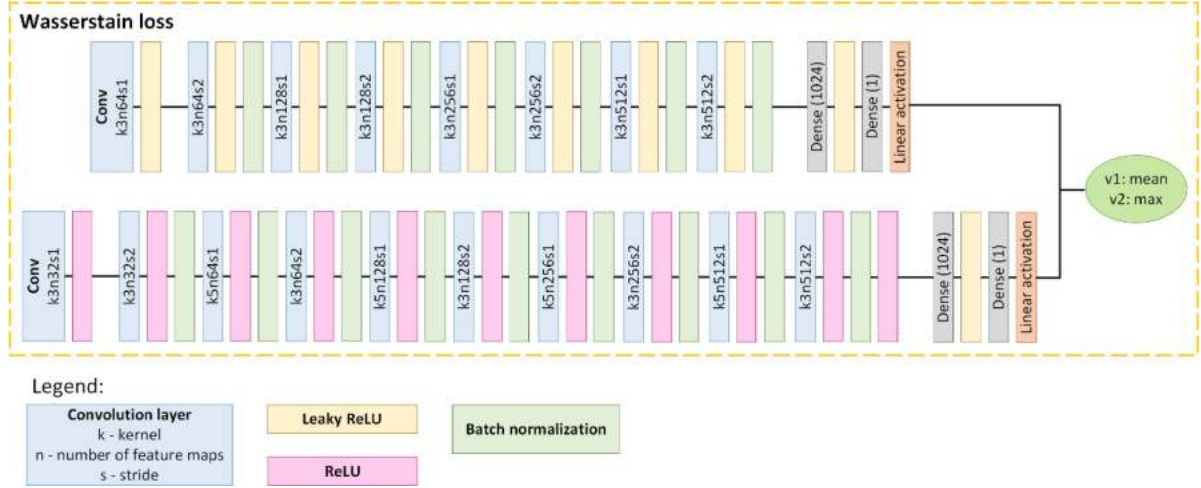
Fig. 2.   Discriminator model.

network. For the second critic, one convolutional layer was added, the method of activation on ReLU was used (in the part of the network that is responsible for the extraction of attributes) and the kernel size was changed (in even-number layers this parameter is 5 pixels) (see Fig. 2).

The application of such a multicolumn discriminator will allow for a better assessment of the operation of the generator, and thus, it will significantly accelerate the training of the network of generators that are responsible for improving the spatial resolution of satellite imagery. Another modification was changing the method of assessing the operation of the generator and discriminator. In the original ESRGAN solution, the generator losses (1) and discriminator losses (2) were calculated from the following equations:

$$L_D^{Ra} = -E_{x_r}\left[\log\left(D_{Ra}\left(I^{HR}, I^{SR}\right)\right)\right]$$
$$- E_{x_r}\left[\log\left(1 - D_{Ra}\left(I^{SR}, I^{HR}\right)\right)\right] \quad (1)$$
$$L_G^{Ra} = -E_{x_r}\left[\log\left(1 - D_{Ra}\left(I^{HR}, I^{SR}\right)\right)\right]$$
$$- E_{x_r}\left[\log\left(D_{Ra}\left(I^{SR}, I^{HR}\right)\right)\right] \quad (2)$$

where $l^{SR}-$ preceptual loss function, $I^{HR}-$ HR image, $I^{SR}-$ SR image, and $D_{Ra}-$ relativistic discriminator.

The proposed solution also applied the Wasserstein Loss. This solution was introduced by Martin Arjovsky et al. in 2017, in the article entitled *Wasserstein GAN* [66]. The authors of the solution have changed the approach to teaching the generator model, which will enable to approximate the distribution of observation data that are used for network training. They modify the discriminator that defines the probability of the image belonging to the SR/HR classes by a critic that determines whether the given image is "true or false". The aim of the training process is to minimize the distance between the distribution of data that is observed in the training dataset and the distribution observed in the generated examples. The distance is determined with the use of the Wasserstein distance method, which defines the distance between probability distributions in metric space. The

Wasserstein distance between two probability measures μ and υ is calculated using the following formula (3) [67]

$$W\left(\mu, \vartheta\right) = \max_{\alpha,\beta}\left[\mathbb{E}_x\left[\alpha\left(X\right)\right] + \mathbb{E}_{x'}\left[\beta\left(X'\right)\right]\right.$$
$$- \gamma \sum_{x,x'\in\{01\}}^{d} \exp\left(\frac{1}{\gamma}\left(\alpha\left(x\right) + \beta\left(x'\right)\right.\right.$$
$$\left.\left.- D\left(x, x'\right)\right) - 1\right)\right] \quad (3)$$

where D: $X \times X \to R_+$, $X = \{0,1\}^d$, W is Wasserstein distance, α, β are functions in the set X, and γ is a joint probability distribution.

The distance carries information about the minimum amount of work that is needed to transform the base distribution into the target distribution, i.e., to improve the spatial resolution of LR data. Additional advantages of the Wasserstein distance method are its properties: it is continuous and differential. Moreover, this solution is more stable during the changes in the architecture of the model or the modification of hyperparameters. An additional advantage of the introduced modification is the fact that it accelerates the model training process. The aim of training the GAN network is to strive to achieve alignment between the generator and the critic by reducing generator's loss. This is enabled by the application of the Kantorovich-Rubinstein duality (4) [68]

$$W\left(\boldsymbol{P}_r, \boldsymbol{P}_\theta\right) = \sup_{||f||_L \leq 1} \mathbb{E}_{x\sim P_r}\left[f\left(x\right)\right] - \mathbb{E}_{z\sim P_\theta}\left[f\left(x\right)\right] \quad (4)$$

where $\boldsymbol{P}_r-$ real data distribution, $\boldsymbol{P}_\theta-$ is the distribution of the parametrized density P_θ, $f: X \to \mathbb{R}$.

In the case where we are dealing with 1-Lipschitz functions, where $f: X \to \mathbb{R}$ and we consider K-Lipschitz for some constant K, then $||f||_L \leq 1$ will take the form $||f||_L \leq K$. Thanks to this relationship, we get the result $K \cdot W(P_r, P_\theta)$. Therefore, when we have a case where $\{f_w\}_{w\in W}$, which are all K-Lipschitz

Hyperparameter	value
Activation	linear activation
Learning rate	0.00005
Clipping parameter	0.01
Batch size	64
Number of iterations of critic per generator iteration	5
Image label	HR: -1, LR: 1
Optimization	RMSProp

for some K, We can write this operation as (5) [66]

$$\max_{w \in W} \mathbb{E}_{x \sim P_r} \left[f_w \left(x \right) \right] - \mathbb{E}_{z \sim p(z)} \left[f_w(g_\theta \left(z \right)) \right]. \quad (5)$$

If the above supremum (3) is reached for some $w \in W$ this process will yield a calculation of $W(\boldsymbol{P}_r, \boldsymbol{P}_\theta)$ up to a multiplicative constant [66].

Considering the specificity of the operation of the GAN network, it takes the form presented in the following equation:

$$\min_G \max_{D \in \mathcal{D}} \mathbb{E}_{x \sim P_r} \left[D \left(x \right) \right] - \mathbb{E}_{\tilde{x} \sim P_g} \left[D \left(\tilde{x} \right) \right] \quad (6)$$

where $\mathcal{D}$ - set of 1-Lipschitz functions, P_r - data distribution, and P_g - the model distribution implicitly defined by $\tilde{x} = G(z), z \sim p(z)$.

In order to implement the Wasserstein Loss, the authors recommend using a special configuration of hyperparameters that is presented in Table I.

IV. EXPERIMENTS AND RESULTS

A. Database

The proposed method was evaluated with the use of three databases that contained identical pairs of LR images (96 × 96 pixels) and HR ones (384 × 384 pixels). The first prepared database comprised aerial photos. The HR database was created as a result of dividing an orthophotomap of the pixel size of 0.25 m. LR images were prepared by sampling HR images to the size of 96 × 96 pixels. The second database was prepared with the use of the Dataset for Object deTection in Aerial images (DOTA) database [69]. The DOTA database is a collection of data created for applications in the field of object detection in aerial or satellite images. The DOTA dataset contains images with varying spatial resolutions, encompassing various scenarios and types of objects, such as vehicles, buildings, sports fields, and roads. In the conducted research, 18 000 random fragments of images of the size of 384 × 384 pixels were selected. LR images, as it was in the case of the first database, were prepared by sampling the HR images. The third database was prepared with the use of imagery acquired by World View 2 (WV2).In this case, the LR database was created by dividing the multispectral imagery of the spatial resolution of 2 m (for the purposes of the research, three channels were prepared: 2, 3, and 4) into
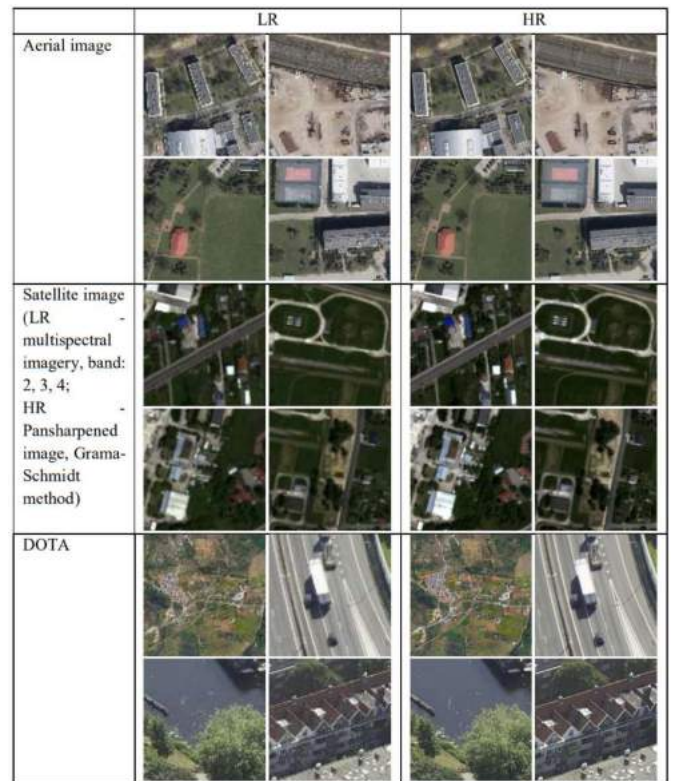
Fig. 3. Sample LR and HR images from the prepared databases.

smaller images of the dimensions of 96 × 96 pixels. To prepare the HR database, it was necessary to apply pansharpening (the Gram-Schmidt method), which enabled to increase the spatial resolution of the multispectral image four times. The resulting image was divided into smaller images (384 × 384 pixels). The prepared pairs of LR and HR images had the same range of imaging. Sample pairs of LR images from specific databases are presented in Fig. 3.

For training purposes, the prepared LR and HR image databases were divided into three datasets: training data (used to train the network and accounting for 70% of the whole database), validation data (used to evaluate the model during network training and accounting for 20% of the whole database) and test data (used for the final assessment of the model).

B. Analyzed Models

In order to verify the proposed solution, five models were trained for each of the prepared datasets. The aim of this approach was to select the solution that would best solve the problem of improving the resolution of satellite imagery. The first of these models was the original ESRGAN. In the second model, the multicolumn discriminator described above was applied. The model assumed that the maximum value of the assessment of the discriminators was the final loss value. In the third solution, the manner of calculating the final value of model losses was modified: the adopted final value was the average value of the losses of all discriminators.

In order to accelerate the network training process even more, in the fourth model, apart from introducing the multicolumn discriminator, the method of network training was modified: the Wasserstein loss (7) was applied. Moreover, considering the results obtained during the training of models 2 and 3, the weights of the generator were updated based on the average value of loss of each of the branches of the discriminator

$$L = \mathbb{E}_{\tilde{x} \sim P_g} \left[D\left(\tilde{x}\right) \right] - \mathbb{E}_{x \sim P_r} \left[D\left(x\right) \right]. \tag{7}$$

However, it was observed during training that this approach was prone to the occurrence of the vanishing gradients phenomenon. As a result, in the final (fifth) model, the method of network training was modified according to the recommendations by Ishaan Gulrajani et al. [70].

This operation resulted in creating the MCWESRGAN model.

The authors of this solution propose a different method of enforcing the Lipschitz constraint. Based on the assumption that a differentiable function is 1-Lipschtiz if and only if it has gradients with norm at most 1 everywhere, so they consider directly constraining the gradient norm by constraint with a penalty on the gradient norm for random samples $\hat{x} - P_{\hat{x}}$ (8)

$$L = \mathbb{E}_{\tilde{x} \sim P_g} \left[D\left(\tilde{x}\right) \right] - \mathbb{E}_{x \sim P_r} \left[D\left(x\right) \right]$$
$$+ \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}} \left[\left(||\nabla_{\hat{x}} D(\hat{x})||_2 - 1 \right)^2 \right]. \tag{8}$$

Considering the above modifications, the last model was trained in compliance with the algorithm published by Ishaan Gulrajani et al. [70].

The conducted research involved the preparation of five models of the GAN network (mainly by introducing modifications to the structure of the discriminator and modifying the network training method). In order to select the best solution, all the models discussed above were trained based on each of the prepared databases.

C. Methods of Assessing the Results

The quality of the operation of the proposed network model was determined with the use of the most popular metrics used in the fields of remote sensing and computer vision. A list of the metrics used is provided in Table II.

Moreover, to improve the assessment of the generated images, an application was created that enables the local assessment of the obtained results. In this solution, the image is divided into smaller areas (determined by the p parameter). Then, the aforementioned metrics of qualitative assessment are defined. This method allows for the determination of areas that were "worse" assessed by the algorithm to improve spatial resolution. This information may become the basis to draw conclusions concerning the modifications that should be introduced, for example, during network training or on the manner of modifying the training database so as to improve the quality of the estimated SR images (see Fig. 4).

Another method proposed by the authors to assess this type of images is the analysis of the PSD of image, which describes the distribution of the frequency of signal power. This method allows us not only to assess the change in the GSD parameter but also in the GRD parameter (see Fig. 5). The GRD parameter defines the smallest size that may be distinguished in the image. Its value depends on the contract between the object and the background. For objects that are characterized by high contrast with the background the value of the GRD is two times the value of the GSD parameter, while if the contrast is low, then $GRD = 2\sqrt{2}GSD$ [71].

The combination of the image assessment methods discussed above enables an unambiguous determination of the quality of the estimated images. Qualitative metrics of image assessment allow for a global assessment of SR images. Applying them to local areas also enables to identify areas that are poorly represented, e.g., containing errors or artifacts. On the other hand, conducting an analysis of PSD for various spatial frequencies enables us to determine the level of improvement of the interpretation capacity of SR images.

TABLE II
LIST OF THE MOST IMPORTANT ASSESSMENT METHODS

PSNR [65]	$PSNR = 10 \cdot \log_{10} \dfrac{\left[\max(HR(n,m)) \right]^2}{MSE}$
UQI [66]	$Q = \dfrac{\sigma_{HR HR\prime}}{\sigma_{HR} \sigma_{HR\prime}} \cdot \dfrac{2 \cdot \overline{HR} \cdot \overline{HR\prime}}{(\overline{HR})^2 (\overline{HR\prime})^2} \cdot \dfrac{2\sigma_{HR} \sigma_{HR\prime}}{\sigma_{HR}^2 + \sigma_{HR\prime}^2}$
SAM [67]	$SAM(v,w) = \cos^{-1}\left(\dfrac{\sum_{i=1}^{N} HR\prime_i HR_i}{\sqrt{\sum_{i=1}^{N} HR\prime_i^2} \sqrt{\sum_{i=1}^{N} HR_i^2}} \right)$
SSIM [65]	$SSIM(x,y) = \dfrac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$
Normalized root mean-squared error (NRMSE) [65]	$NRMSE = \dfrac{RMSE}{HR}$
Root mean square error (RMSE) [65]	$RMSE = \sqrt{\dfrac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} [HR(n,m) - HR\prime(n,m)]^2}$
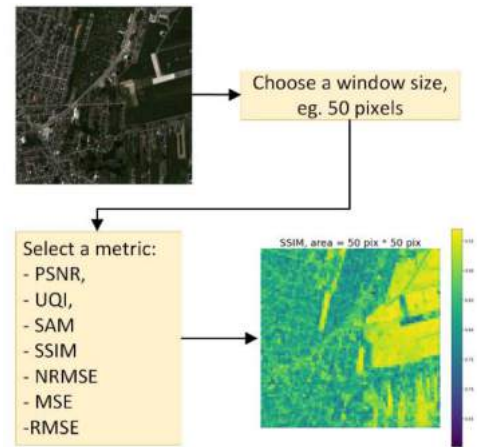
Fig. 4. Block diagram of the functioning of the image assessment algorithm.

TABLE III
PRESENTATION OF RESULTS FOR SATELLITE IMAGES

	ESRGAN	ESRGAN + MC discriminator (mean dis_loss)	ESRGAN + MC discriminator (max dis_loss)	ESRGAN + MC discriminator (mean dis_loss) + Wasserstein	ESRGAN + MC discriminator (mean dis_loss) + Wasserstein VG **MCWESRGAN**
training time *	96 h	12 h	12 h	4 h	6 h
SSIM	0.83	0.84	0.82	0.85	0.92
PSNR	28.74	28.48	29.68	28.97	32.10
SAM	0.14	0.13	0.12	0.12	0.09
SCC	0.18	0.24	0.26	0.26	0.37
UQI	0.95	0.95	0.96	0.96	0.98

The following abbreviations were used: MC − multicolumn, VG − vanish gradient.
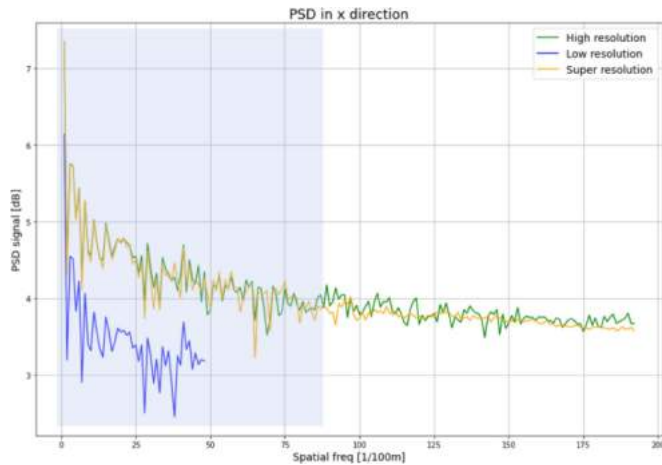
Fig. 5. Sample diagram of PSD, for the spatial frequency of the HR image of the GSD = 0.5 m, along the *x*-axis.

D. Results

All the analyzed artificial neural networks were trained based on the Nvidia TITAN RTX 24 GB graphics card, Intel Xeon Silver 4216 processor and the Ubuntu 18.04 operating system. The input parameters adopted for the ESRGAN training were the authors' recommendations (learning rate is initialized as $2 \cdot 10^{-4}$, decayed by a factor of 2 every $2 \cdot 10^5$ of mini-batch updates, Adam optimization with $\beta1 = 0.9$, $\beta2 = 0.999$). However, due to the vanishing gradient phenomenon that occurred for the ESRGAN after 30 000 steps, the rate of network training was slowed down by reducing the value of the learning rate parameter [33].

Based on the above assumptions, three models (that differed in terms of the training dataset) were trained for each of the analyzed architectures of the conditional GAN network.

Table III presents the qualitative results of the evaluation metrics conducted on the database of images acquired by the WorldView-2 satellite. Analyzing the results, we observe a significant decrease in the training time required for each SISR model. The implementation of a multicolumn discriminator allows for approximately a 7-fold reduction in the training time of the model. At the same time, the use of Wasserstein loss shortens this time by over ten times. However, a drawback of the modified loss method is a relatively high likelihood of encountering the issue of vanishing gradients, especially in the case of satellite imagery. In the case of the MCWESRGAN, the "enforce the Lipschitz constraint" method was applied to mitigate the vanishing gradient issue, enabling longer network training and consequently improving the quality of the generated images, as confirmed in the qualitative analysis. The results of the quality assessment of SR images for the remaining databases are presented in Appendix 1.

To improve the evaluation of the proposed strategies for training GANs, qualitative metrics were calculate for the assessment of the image. These metrics are presented in Tables V and VI (Appendix 1). Apart from that, the table contains information about the number of hours that were spent on training the analyzed models (generators).

In order to improve the evaluation of SR images, the values of local SSIM and peak signal-to-noise ratio (PSNR) metrics were calculated. In this approach, the reference image (HR) and the estimated image (SR) are divided into smaller areas of the dimensions determined by the *p* parameter. Then, the values of the selected metrics of qualitative assessment were calculated for the defined areas (for the purposes of the presentation of results, only two metrics of qualitative assessment were selected, as, in the opinion of the authors, they best reflect the differences between the quality of estimation of the areas in SR images). Appendix 2 presents the examples of the conducted local analysis of SR images.

In order to determine the quality of improvement of the spatial resolution of SR images, the ratio of power density spectrum to spatial frequency was analyzed. This method enables us to determine the level of increase in the interpretation capacity of SR images. The analyses conducted for sample images on the *x*- and *y*-axes are presented in Appendix 3.

The presented results demonstrate that the SR image (the yellow curve in the diagram) has the same value of the GSD

TABLE IV
QUANTITATIVE EVALUATION OF MCWESRGAN

		EDSR [50]	ESPCN [69]	FSRCNN [70]	RDN [71]	SRCNN [49]	SRDenseNet [72]	SRGAN [54]	ESRGAN [33]	MCWESRGAN (our)
	I	0.54	0.69	0.67	0.67	0.61	0.68	0.60	0.60	0.70
SSIM	II	0.45	0.60	0.56	0.60	0.56	0.61	0.54	0.15	0.62
	III	0.79	0.91	0.90	0.91	0.82	0.89	0.83	0.83	0.92
	I	20.62	24.11	24.58	23.61	22.87	24.49	21.48	22.76	25.05
PSNR	II	18.64	20.95	21.97	21.41	21.05	22.01	20.44	11.52	22.45
	III	24.76	27.15	31.27	32.01	25.4	30.12	29.18	28.74	32.10
	I	0.24	0.15	0.15	0.16	0.18	0.15	0.17	0.19	0.14
SAM	II	0.26	0.19	0.14	0.18	0.20	0.19	0.21	0.19	0.17
	III	0.18	0.14	0.10	0.11	0.14	0.12	0.15	0.14	0.09
	I	0.12	0.29	0.17	0.25	0.14	0.23	0.13	0.13	0.32
SCC	II	0.06	0.11	0.13	0.12	0.08	0.11	0.08	0.06	0.13
	III	0.22	0.31	0.20	0.34	0.16	0.31	0.15	0.14	0.37
	I	0.96	0.99	0.98	0.98	0.97	0.98	0.95	0.98	0.99
UQI	II	0.93	0.96	0.95	0.96	0.96	0.96	0.95	0.70	0.97
	III	0.91	0.98	0.97	0.98	0.93	0.97	0.91	0.95	0.98

The best results are marked in green, the second in yellow and the worst in red the following abbreviations were used: I: dota database, II: aerial photos, III: satellite images.

as the HR image (the spatial resolution increased four times in comparison to the LR image). In addition, the analysis of the ratio of power density spectrum to spatial frequency of the HR and SR images revealed that image resolution was improved only in the area marked by the blue rectangle, i.e., to a spatial density of approx. 105 [1/100 m] (see Fig. 5). The power density spectrum is best reflected in this area. As a result, the value of the GRD parameter is approx. 1.8–2.1 GSD.

To further verify the proposed model, the SR images generated by MCWESRGAN were compared with the SR images estimated by other SISR models based on deep neural networks. Based on visual analysis (see Appendix 4) and the results of the qualitative image assessment (see Table IV), it can be observed that the proposed solution allows for the most accurate image reconstruction, especially for linear elements and surfaces. Examining the obtained results, our model exhibits the best evaluation metric values (see Table IV), indicated in green. Metrics with the worst values are highlighted in red (mainly attributed to the EDSR model). From the conducted analysis, it is clear that our proposed MCWESRGAN achieves the best results in each of the evaluated metrics (SSIM, PSNR, spectral angle mapper [SAM], SCC, and universal quality measure [UQI]).

In addition, it can be noticed that SR images from satellite imagery fragments exhibit better results compared to SR images from the DOTA database. That may be due to the significantly lower spatial resolution of satellite imagery compared to the images in the DOTA database, resulting in fewer details present in the images. Furthermore, attention should be drawn to the SRDenseNet model. SR images estimated by this model also achieve equally good results during the qualitative assessment. However, through visual analysis, it can be observed that the SR images appear to be painted with a brush, distinguishing them from the HR images. The worst results were obtained by the EDSR network (SSIM value below 0.80 for the satellite imagery database, whereas for our model, SSIM = 0.92). Conducting visual analysis (examples of images can be found in
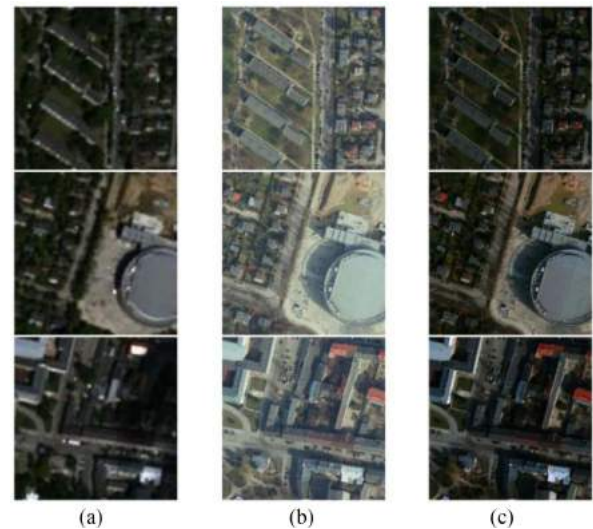
Fig. 6. Examples of image pairs. (a) LR. (b) HR. (c) HR after histogram adjustment.

Appendix 3), even degradation of interpretational potential can be observed.

The obtained results demonstrated that the application of a multicolumn network in the generator model significantly accelerates the process of network training. For the analyzed databases, the time required to train the generator was shortened more than 10 times. The implementation of Wasserstein loss additionally accelerated the training of the model. In addition, the quality of the estimated SR images was better than that of those estimated by the ESRGAN.

The analysis of the obtained results (mainly based on local assessment of images) allows us to claim that anthropogenic objects are better represented in SR images than objects of natural origin (e.g., tree crowns). The exceptions are clearings and meadows, whose assessment quality increases with the decrease in spatial resolution of LR images.

Fig. 7. Examples of the improvement of resolution of satellite images based on the example of aerial imagery. (a) LR. (b) SR. (c) HR.

V. DISCUSSION

This article presents a new architecture of a conditional GAN that enables to improve the resolution of images acquired at the aerial and space altitudes. The core of the proposed solution is the ESRGAN, where the structure of the discriminator model was modified (the multicolumn network was used and the arrangement of layers was modified). The training strategy of the GAN network was also improved. The introduction of Wasserstein Loss enabled us to significantly increase the rate of network training: more than 10 times faster with a simultaneous improvement of the quality of estimated SR images (Appendix 1).

The conducted research has demonstrated that the use of a multicolumn discriminator and Wasserstein loss during the generator training process:

1) significantly accelerates the network training process,
2) mitigates the issue of vanishing gradients, leading to a more stable model training,
3) results in estimated SR images of higher quality, as confirmed through both local and global image evaluations.

However, it is worth noting that this proposed solution does not address the problem of the characteristic texture artifacts that appear in SR images (this phenomenon is also visible in images estimated by the ESRGAN model).

During the research works, an experiment was additionally conducted, in which the model was trained based on satellite data (LR image database) and aerial imagery (HR image database). During the creation of the database, the authors attempted to select the data so that the time of data acquisition was similar. Examples of LR and HR images are presented in Fig. 6. To reduce the differences in the quality of the obtained data, resulting from various methods of image acquisition, and different structures of the optical systems, the histogram of the HR image was matched to the LR image. This operation allowed us to significantly reduce the differences between LR and HR areas.

Analyzing the obtained results, it is worth noting the quality of the work of the generators (which were prepared based on multispectral images and aerial imagery). The values of qualitative assessment metrics are much poorer in comparison to the models that were trained on other databases. Fig. 7 presents an example of the operation of a generator.

When the database was created, the authors attempted to select the data so that the time and conditions of image acquisition were similar. However, with the use of open data sources, it was possible to acquire aerial images that were taken only in the same year as the WV2 satellite imagery. Unfortunately, the satellite imagery and aerial photos were acquired:

1) in different seasons of the year, so the plant cover visible in the images is at various stages of vegetation,
2) at different times of day, so the difference in the façade angle and azimuth of the Sun influenced the position of shadows that are visible in the photos,
3) at different values of sensor tilt angle, so various walls of the facades of tall buildings are visible.

The aim of the research was to improve the spatial resolution of satellite imagery of the GSD of approximately 2 m. Due to that, the differences within pairs of images (LR and HR) had a significant influence on the quality of network training. During the research, none of the tested networks managed to cope with his problem. At the same time, in a situation when it is necessary to improve the resolution of satellite images of lower resolution (when the GSD value exceeds 10 m), the differences that result from different ways of acquiring the data will have a lower influence [72], [73], [74].

VI. CONCLUSION

The conducted research demonstrated that the application of a multicolumn discriminator allows us for a significant acceleration of the network training process. The rate of this process may be additionally accelerated by implementing the Wasserstein Loss that uses Wasserstein distance. Although Wasserstein Loss increases the model's proneness to overtraining (especially for images acquired at aerial and space altitudes). Moreover, we would like to attract the readers' attention to the manner of assessment of the images estimated by GAN. During the global qualitative assessment of SR images, one may encounter cases where estimation errors are visible, in spite of the high results of the visual assessment of the image. In these cases, the high value of the assessment metrics may result in a very good representation e.g., of homogeneous surfaces (e.g., surfaces covered with

TABLE V
PRESENTATION OF RESULTS FOR AERIAL PHOTOS

	ESRGAN	ESRGAN + MC discriminator (mean dis_loss)	ESRGAN + MC discriminator (max dis_loss)	ESRGAN + MC discriminator (mean dis_loss) + Wasserstein	ESRGAN + MC discriminator (mean dis_loss) + Wasserstein VG **MCWESRGAN**
training time *	101 h	14 h	14 h	6 h	8 h
SSIM	0.15	0.54	0.56	0.61	0.62
PSNR	11.52	21.74	22.23	28.43	22.45
SAM	0.19	0.19	0.18	0.17	0.17
SCC	0.06	0.07	0.11	0.13	0.13
UQI	0.70	0.93	0.94	0.97	0.97

TABLE VI
PRESENTATION OF RESULTS FOR THE DOTA DATABASE (NO VANISHING GRADIENT PHENOMENON)

	ESRGAN	ESRGAN + MC discriminator (mean dis_loss)	ESRGAN + MC discriminator (max dis_loss)	ESRGAN + MC discriminator (mean dis_loss) + Wasserstein	ESRGAN + MC discriminator (mean dis_loss) + Wasserstein VG **MCWESRGAN**
training time *	83 h	14 h	14 h	9 h	11 h
SSIM	0.6	0.68	0.69	0.69	0.70
PSNR	22.76	24.60	24.70	24.78	25.05
SAM	0.19	0.15	0.15	0.15	0.14
SCC	0.13	0.29	0.29	0.30	0.32
UQI	0.98	0.98	0.98	0.99	0.99

shadows, or roads), which "overstate" the final score. The application of local assessment enables the identification of areas of lower or higher SR assessment quality. Such information may be valuable for preparing an additional database of images, which may provide a basis for supplementary training of an existing model. Furthermore, we recommend assessing interpretational potential based on the PSD of the image. This method allows for a numerical determination of the improvement in recognition capabilities in the estimated SR images. The proposed approach enables a precise evaluation of SR images, facilitating a better assessment of the generator and detecting outliers that may lead to overestimation or underestimation of global metrics in the qualitative image assessment.

The obtained results revealed that the quality of SR images deteriorated with the increase in spatial resolution of LR images (see Tables III–V). For images acquired at aerial altitudes (where the value of the GSD parameter is lower than 30 cm), deformations that result from incorrect estimation of the details of objects that are in the photograph have a significant influence on the result of the assessment of SR images.

In addition, it was noted that if homogenous databases are used (i.e., if the images used originate from the same set of satellite imagery or photogrammetric flight), where the conditions are similar, the probability of occurrence of the vanishing gradient phenomenon increases. If heterogeneous databases are used (e.g., DOTA that contains aerial and satellite images as well as images acquired by UAV) the phenomenon does not occur.

Considering the obtained results, future research will verify the proposed method for improving the spatial resolution of image sequences acquired by nano- and minisatellites.

APPENDIX 1
RESULTS OF THE ASSESSMENT OF THE TESTED NETWORK MODELS WITH THE USE OF QUALITATIVE METRICS

See Tables V and VI.

APPENDIX 2
RESULTS OF LOCAL ASSESSMENT OF SR IMAGES

LR

HR

SR

SSIM (p = 8pix)

SR + SSIM

PSNR

SR + PSNR

SAM

SR + SAM
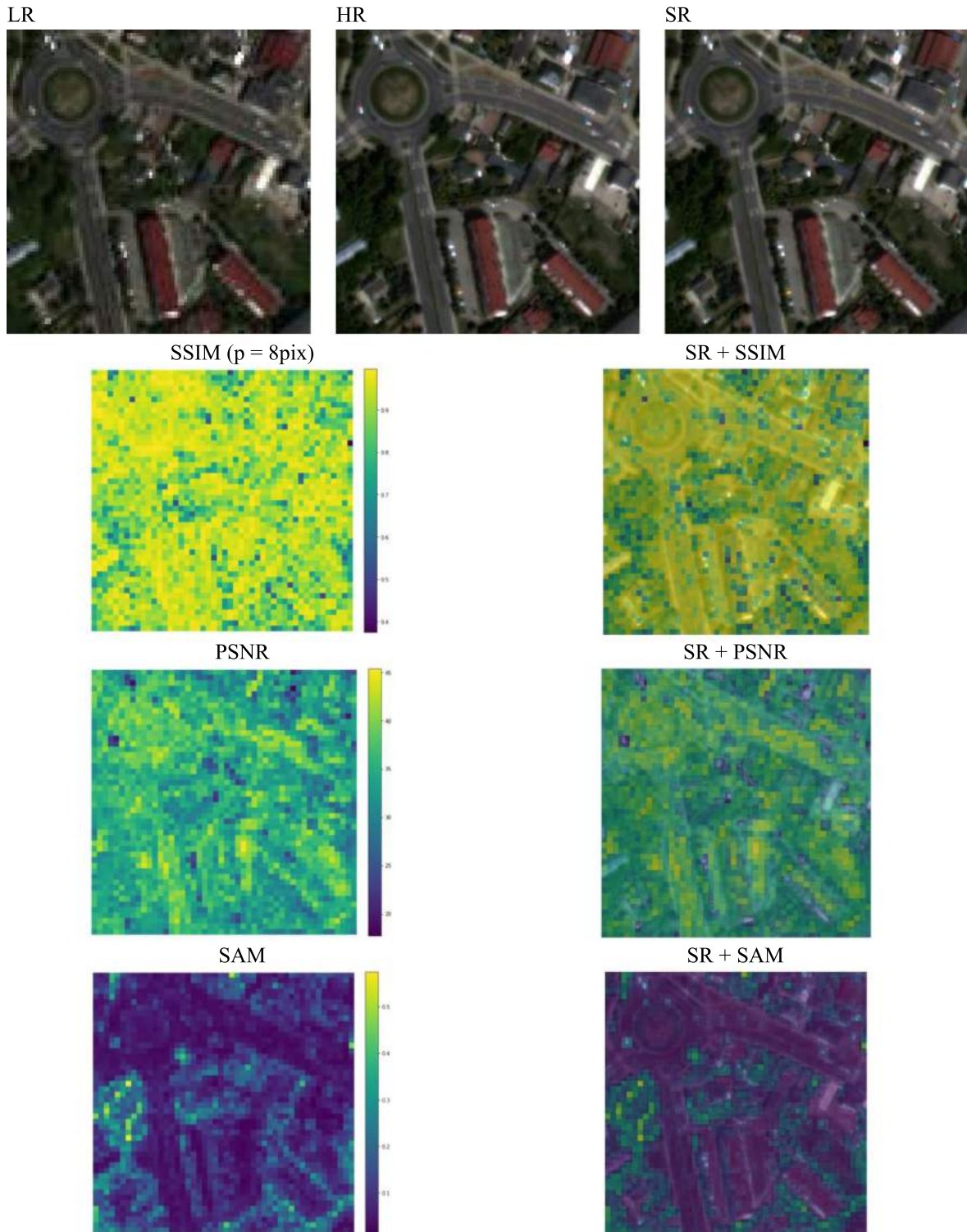
APPENDIX 3
RESULTS OF THE RATIO OF PSD TO THE SPATIAL FREQUENCY OF SR IMAGES

a) ESRGAN
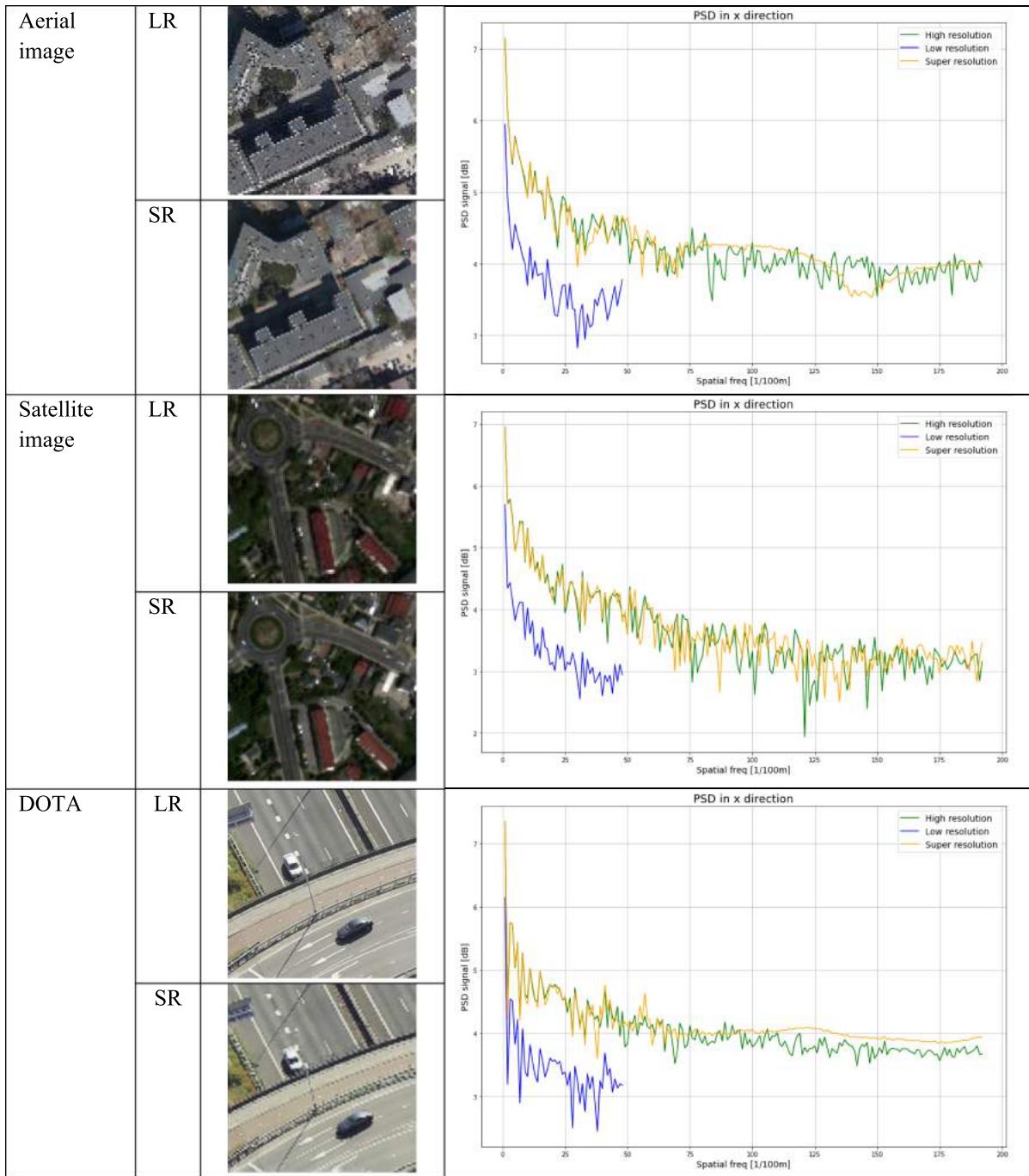
b) Multicolumn ESRGAN (max)
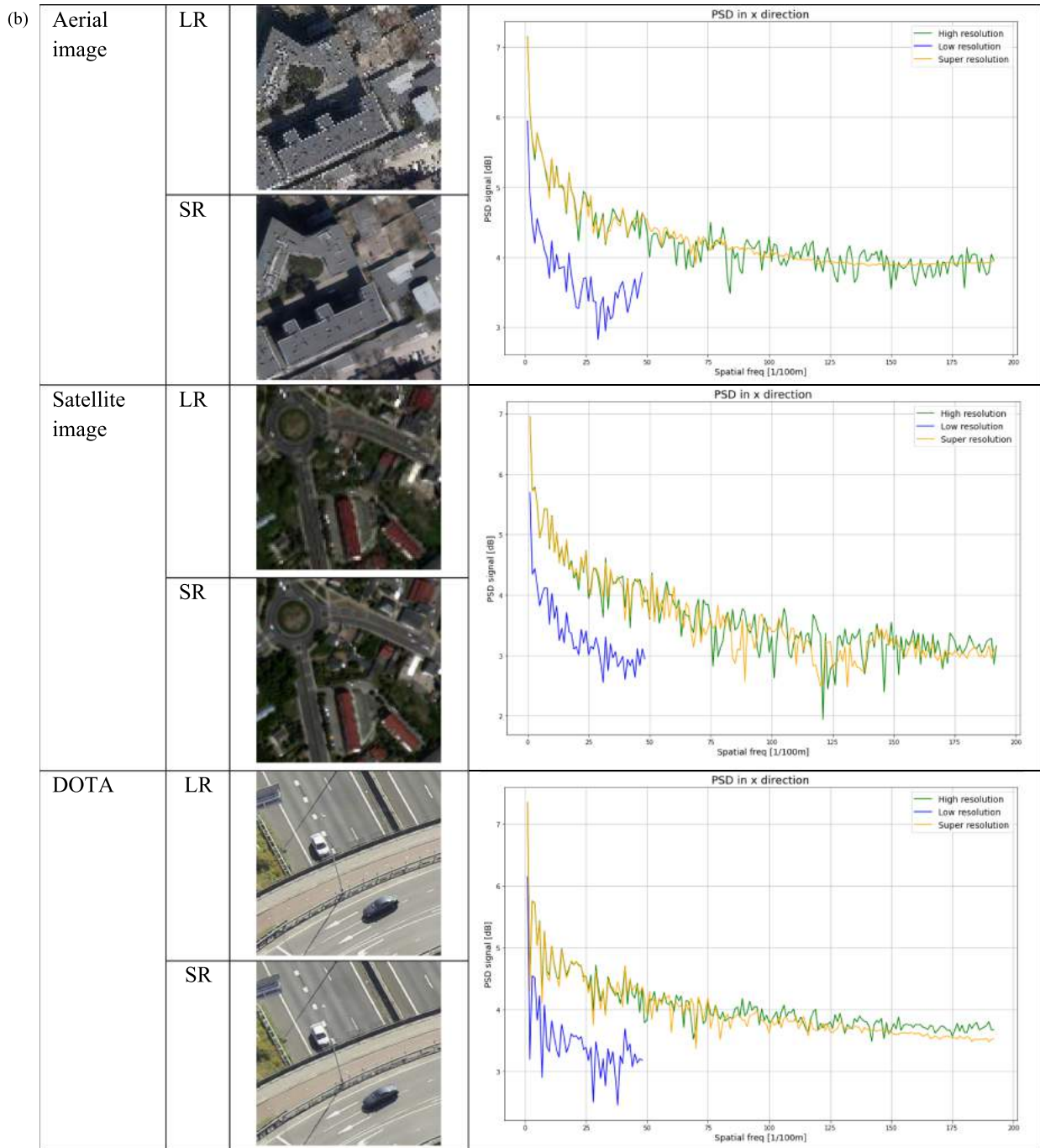
c) Multicolumn ESRGAN (mean)
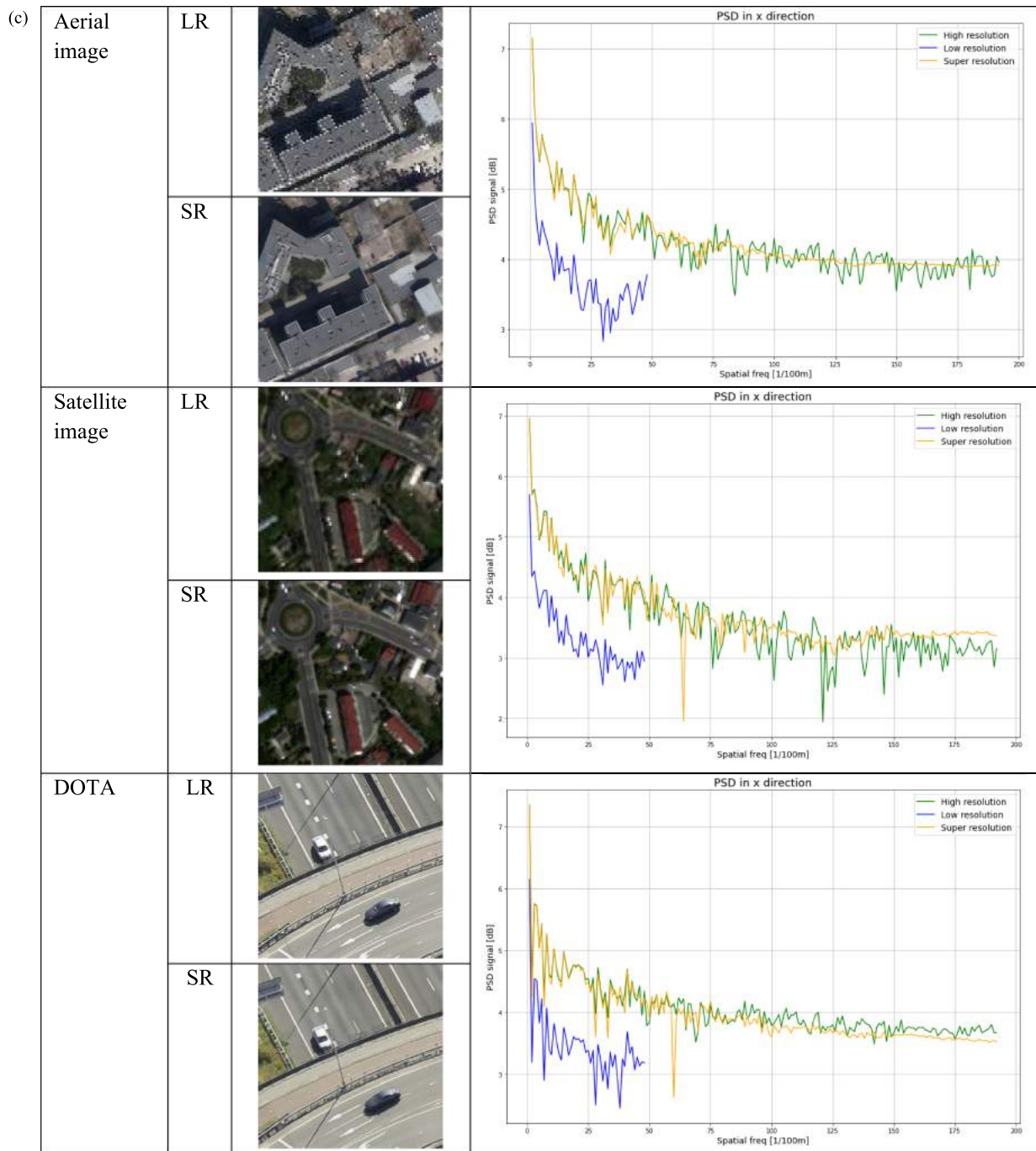
d) Multicolumn Wasserstein
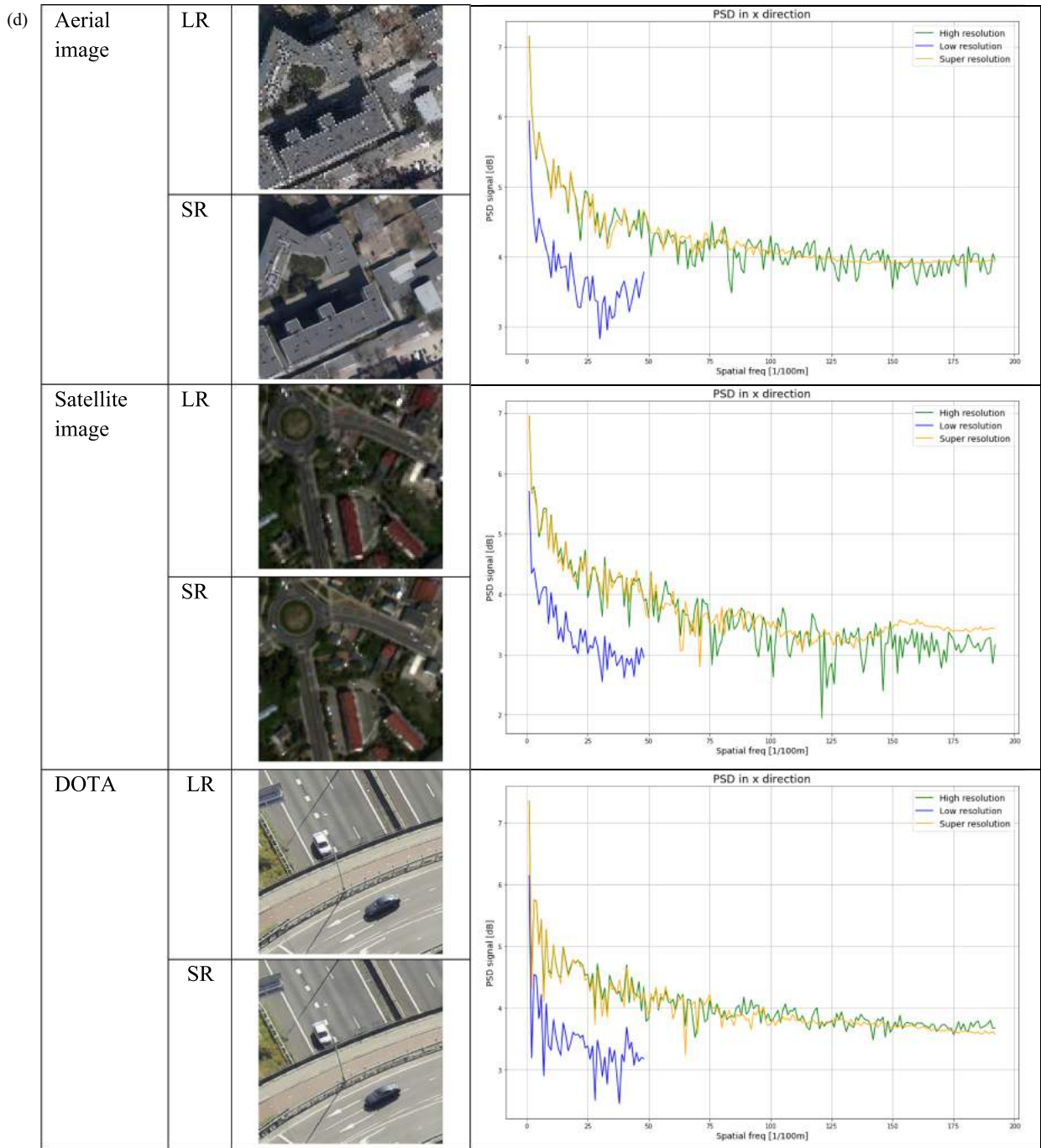
e) Multicolumn Wasserstein VGESRGAN (mean), Vanish gradient
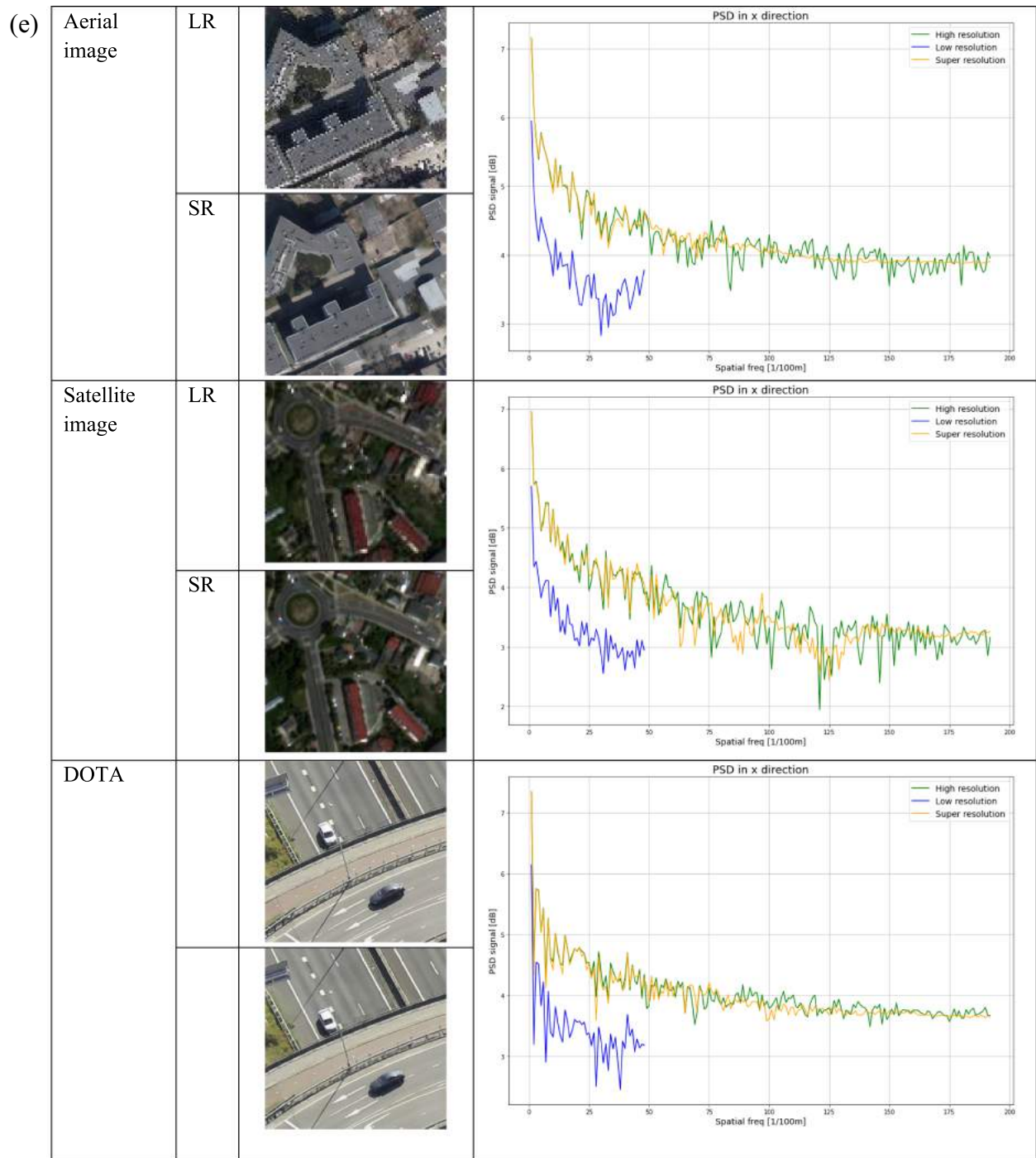
APPENDIX 4

	DOTA DATABASE	SATELLITE IMAGES	AERIAL PHOTOS
LR			
EDSR			
ESPCN			
FSRCNN			

HR
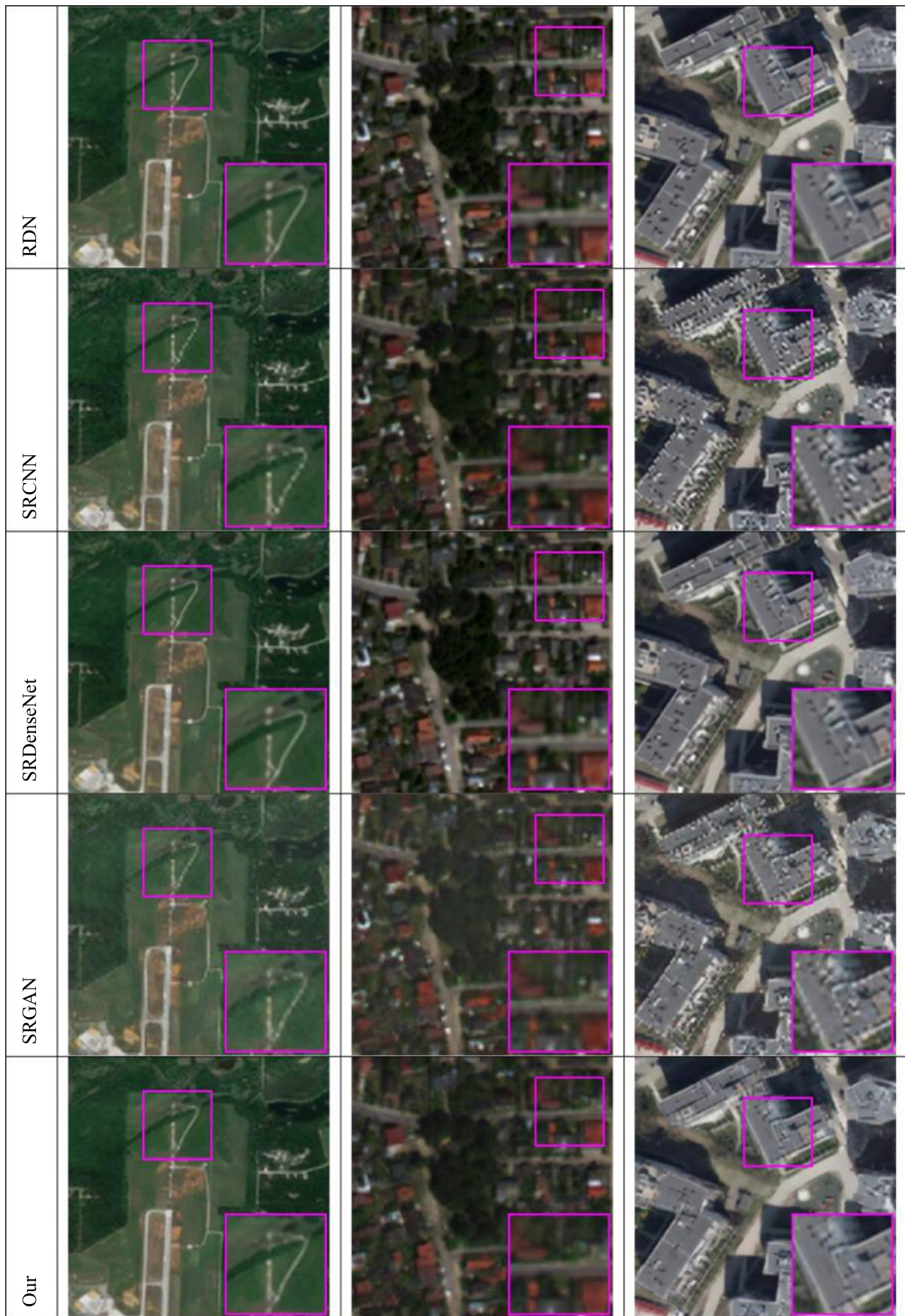
REFERENCES

[1] A. Jenerowicz, R. Kaczynski, K. Siok, and K. Palkiewicz, "Change detection of urban area based on multisensor imagery," in *Remote Sensing Technologies and Applications in Urban Environments IV*, N. Chrysoulakis, T. Erbertseder, Y. Zhang, and F. Baier, Eds. Strasbourg, France: SPIE, Oct. 2019, pp. 27, doi: 10.1117/12.2533379.

[2] J. Lu, J. Li, G. Chen, L. Zhao, B. Xiong, and G. Kuang, "Improving pixel-based change detection accuracy using an object-based approach in multitemporal SAR flood images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3486–3496, Jul. 2015, doi: 10.1109/JSTARS.2015.2416635.

[3] B. P. Salmon, W. Kleynhans, F. van den Bergh, J. C. Olivier, T. L. Grobler, and K. J. Wessels, "Land cover change detection using the internal covariance matrix of the extended Kalman filter over multiple spectral bands," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1079–1085, Jun. 2013, doi: 10.1109/JSTARS.2013.2241023.

[4] Y. Peng, B. Cui, H. Yin, Y. Zhang, and P. Du, "Automatic SAR change detection based on visual saliency and multi-hierarchical fuzzy clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7755–7769, 2022, doi: 10.1109/JSTARS.2022.3199017.

[5] W. Li, X. Xiao, P. Xiao, H. Wang, and F. Xu, "Change detection in multitemporal SAR images based on slow feature analysis combined with improving image fusion strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3008–3023, 2022, doi: 10.1109/JSTARS.2022.3166234.

[6] P. Han et al., "Monitoring trends in light pollution in China based on nighttime satellite imagery," *Remote Sens.*, vol. 6, no. 6, Jun. 2014, Art. no. 6, doi: 10.3390/rs6065541.

[7] E. Bielecka, A. Jenerowicz, K. Pokonieczny, and S. Borkowska, "Land cover changes and flows in the Polish Baltic Coastal Zone: A qualitative and quantitative approach," *Remote Sens.*, vol. 12, no. 13, Jan. 2020, Art. no. 13, doi: 10.3390/rs12132088.

[8] Z. Dong, M. Wang, Y. Wang, Y. Zhu, and Z. Zhang, "Object detection in high resolution remote sensing imagery based on convolutional neural networks with suitable Object scale features," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2104–2114, Mar. 2020, doi: 10.1109/TGRS.2019.2953119.

[9] S. Qiu, G. Wen, and Y. Fan, "Occluded object detection in high-resolution remote sensing images using partial configuration object model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1909–1925, May 2017, doi: 10.1109/JSTARS.2017.2655098.

[10] Q. Yin et al., "Detecting and tracking small and dense moving objects in satellite videos: A benchmark," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5612518, doi: 10.1109/TGRS.2021.3130436.

[11] M. J. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 374–387, Feb. 2005, doi: 10.1109/TGRS.2004.841481.

[12] T. Zhang, Y. Zhuang, G. Wang, S. Dong, H. Chen, and L. Li, "Multiscale semantic fusion-guided fractal convolutional object detection network for optical remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5608720, doi: 10.1109/TGRS.2021.3108476.

[13] Q. Lin, J. Zhao, B. Du, G. Fu, and Z. Yuan, "MEDNet: Multiexpert detection network with unsupervised clustering of training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4703114, doi: 10.1109/TGRS.2021.3093556.

[14] K. Thiagarajan, M. Manapakkam Anandan, A. Stateczny, P. B. Divakarachari, and H. Kivudujogappa Lingappa, "Satellite image classification using a hierarchical ensemble learning and correlation coefficient-based gravitational search algorithm," *Remote Sens.*, vol. 13, 2021, Art. no. 4351.

[15] Z. Liu, J. Xu, J. Li, A. Plaza, S. Zhang, and L. Wang, "Moving ship optimal association for maritime surveillance: Fusing AIS and Sentinel-2 data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5635218, doi: 10.1109/TGRS.2022.3227938.

[16] H. Greidanus, "Satellite imaging for maritime surveillance of the European Seas," *Remote Sens. Eur. Seas*, 2008, pp. 343–358, doi: 10.1007/978-1-4020-6772-3_26.

[17] R. Jalal et al., "Toward efficient land cover mapping: An overview of the National land representation system and land cover Map 2015 of Bangladesh," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3852–3861, Oct. 2019, doi: 10.1109/JSTARS.2019.2903642.

[18] A. Jamil and B. Bayram, "Tree species extraction and land use/cover classification from high-resolution digital orthophoto maps," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 89–94, Jan. 2018, doi: 10.1109/JSTARS.2017.2756864.

[19] H. Shih, D. A. Stow, J. R. Weeks, and L. L. Coulter, "Determining the type and starting time of land cover and land use change in southern Ghana based on discrete analysis of dense landsat image time series," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 5, pp. 2064–2073, May 2016, doi: 10.1109/JSTARS.2015.2504371.

[20] Y. Zhang, H. Balzter, B. Liu, and Y. Chen, "Analyzing the impacts of urbanization and seasonal variation on land surface temperature based on subpixel fractional covers using landsat images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1344–1356, Apr. 2017, doi: 10.1109/JSTARS.2016.2608390.

[21] H. Lestiana and Sukristiyanti, "Spatial resolution enhancement of satellite image data using fusion approach," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 118, no. 1, Feb. 2018, Art. no. 012047, doi: 10.1088/1755-1315/118/1/012047.

[22] K. Karwowska and D. Wierzbicki, "Using super-resolution algorithms for small satellite imagery: A systematic review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3292–3312, 2022, doi: 10.1109/JSTARS.2022.3167646.

[23] X. Wang et al., "A review of image super-resolution approaches based on deep learning and applications in Remote sensing," *Remote Sens.*, vol. 14, no. 21, Jan. 2022, Art. no. 21, doi: 10.3390/rs14215423.

[24] P. Wang, B. Bayram, and E. Sertel, "A comprehensive review on deep learning based remote sensing image super-resolution methods," *Earth-Sci. Rev.*, vol. 232, Sep. 2022, Art. no. 104110, doi: 10.1016/j.earscirev.2022.104110.

[25] A. D. Philip, "The use of intensity-hue-saturation transformation for producing color shaded relief images Photogramm," *Eng. Remote Sens.*, no. 60, pp. 1369–1374, 1994.

[26] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and Multispectral Data Fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017, doi: 10.1109/MGRS.2016.2637824.

[27] A. Mookambiga and V. Gomathi, "Comprehensive review on fusion techniques for spatial information enhancement in hyperspectral imagery," *Multidim. Syst. Signal Process.*, vol. 27, no. 4, pp. 863–889, Oct. 2016, doi: 10.1007/s11045-016-0415-2.

[28] Shruti and S. Budhiraja, "Multiscale image fusion for pansharpening of multispectral images using saliency detection," in *Proc. 9th Int. Conf. Contemporary Comput.*, 2016, pp. 1–6, doi: 10.1109/IC3.2016.7880253.

[29] G. Vivone et al., "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015, doi: 10.1109/TGRS.2014.2361734.

[30] A. Fruchter and R. N. Hook, "Novel image reconstruction method applied to deep Hubble space telescope images," *Appl. Digit. Image Process. XX, Int. Soc. Opt. Photon.*, pp. 120–125, Oct. 1997, doi: 10.1117/12.292751.

[31] P. Dherete and B. Rouge, "Image de-blurring and application to SPOT5 THR satellite imaging," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. Proc. (IEEE Cat. No.03CH37477)*, 2003, pp. 318–320, doi: 10.1109/IGARSS.2003.1293762.

[32] C. Latry and B. Rouge, "Super resolution: Quincunx sampling and fusion processing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. Proc. (IEEE Cat. No.03CH37477)*, 2003, pp. 315–317, doi: 10.1109/IGARSS.2003.1293761.

[33] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," Sep. 2018. arXiv:1809.00219 [cs], Jun. 15, 2021. [Online]. Available: http://arxiv.org/abs/1809.00219

[34] H. Ren, A. Kheradmand, M. El-Khamy, S. Wang, D. Bai, and J. Lee, "Real-world super-resolution using generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 1760–1768, doi: 10.1109/CVPRW50498.2020.00226.

[35] Z.-S. Liu, W.-C. Siu, L.-W. Wang, C.-T. Li, M.-P. Cani, and Y.-L. Chan, "Unsupervised real image super-resolution via generative variational AutoEncoder," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 1788–1797, doi: 10.1109/CVPRW50498.2020.00229.

[36] R. M. Umer, G. L. Foresti, and C. Micheloni, "Deep generative adversarial residual convolutional networks for real-world super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 1769–1777, doi: 10.1109/CVPRW50498.2020.00227.

[37] A. Lugmayr et al., "NTIRE 2021 learning the super-resolution space challenge," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2021, pp. 596–612, doi: 10.1109/CVPRW53098.2021.00072.

[38] M. Lalak and D. Wierzbicki, "Automated detection of Atypical aviation obstacles from UAV images using a YOLO algorithm," *Sensors*, vol. 22, no. 17, Jan. 2022, Art. no. 17, doi: 10.3390/s22176611.

[39] Z. Zakria, J. Deng, R. Kumar, M. S. Khokhar, J. Cai, and J. Kumar, "Multiscale and direction target detecting in remote sensing images via modified YOLO-v4," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1039–1048, 2022, doi: 10.1109/JSTARS.2022.3140776.

[40] C. Zhao, X. Fu, J. Dong, R. Qin, J. Chang, and P. Lang, "SAR ship detection based on end-to-end morphological feature pyramid network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4599–4611, 2022, doi: 10.1109/JSTARS.2022.3150910.

[41] Q. Xu, Y. Li, and Z. Shi, "LMO-YOLO: A ship detection model for low-resolution optical satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4117–4131, 2022, doi: 10.1109/JSTARS.2022.3176141.

[42] L. E. Falqueto, J. A. S. Sá, R. L. Paes, and A. Passaro, "Oil rig recognition using convolutional neural network on sentinel-1 SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1329–1333, Aug. 2019, doi: 10.1109/LGRS.2019.2894845.

[43] G. Sheng, W. Yang, L. Chen, and H. Sun, "Satellite image classification using sparse codes of multiple features," in *Proc. IEEE 10th Int. Conf. SIGNAL Process. Proc.*, 2010, pp. 952–955, doi: 10.1109/ICOSP.2010.5655718.

[44] A. H. Khan, Z. Zafar, M. Shahzad, K. Berns, and M. M. Fraz, "Crop type classification using multi-temporal Sentinel-2 satellite imagery: A deep semantic segmentation approach," in *Proc. Int. Conf. Robot. Autom. Ind.*, 2023, pp. 1–6, doi: 10.1109/ICRAI57502.2023.10089586.

[45] Y.-C. Sun, I. D. Gerg, and V. Monga, "Iterative, deep synthetic aperture sonar image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4206615, doi: 10.1109/TGRS.2022.3162420.

[46] I. D. Gerg and V. Monga, "Deep multi-look sequence processing for synthetic aperture sonar image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4200915, doi: 10.1109/TGRS.2023.3234229.

[47] I. Ulku, E. Akagündüz, and P. Ghamisi, "Deep semantic segmentation of trees using multispectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7589–7604, 2022, doi: 10.1109/JS-TARS.2022.3203145.

[48] G. Lenczner, A. Chan-Hon-Tong, B. L. Saux, N. Luminari, and G. L. Besnerais, "DIAL: Deep interactive and active learning for semantic segmentation in remote sensing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3376–3389, 2022, doi: 10.1109/JS-TARS.2022.3166551.

[49] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," Jul. 2015. arXiv:1501.00092 [cs], Accessed: Jul. 15, 2021. [Online]. Available: http://arxiv.org/abs/1501.00092

[50] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," Jul. 2017. arXiv:1707.02921 [cs], Accessed: Jul. 15, 2021. [Online]. Available: http://arxiv.org/abs/1707.02921

[51] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," Accessed: Apr. 14, 2023. arXiv, Mar. 26, 2016. [Online]. Available: http://arxiv.org/abs/1603.08155

[52] T. R. Shaham, T. Dekel, and T. Michaeli, "SinGAN: Learning a generative model from a single natural image," Sep. 04, 2019. arXiv. Accessed: Apr. 14, 2023. [Online]. Available: http://arxiv.org/abs/1905.01164

[53] H. Li et al., "SRDiff: Single image super-resolution with diffusion probabilistic models," May 18, 2021. arXiv. Accessed: Sep. 14, 2023. [Online]. Available: http://arxiv.org/abs/2104.14951

[54] Z. Lu, J. Li, H. Liu, C. Huang, L. Zhang, and T. Zeng, "Transformer for single image super-resolution," Apr. 22, 2022. arXiv. Accessed: Sep. 14, 2023. [Online]. Available: http://arxiv.org/abs/2108.11084

[55] I. Goodfellow et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, 2014. Accessed: Dec. 08, 2021. [Online]. Available: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

[56] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on Generative adversarial Networks: Algorithms, theory, and applications," Jan. 19, 2020. arXiv. Accessed: Sep. 14, 2023. [Online]. Available: http://arxiv.org/abs/2001.06937

[57] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks," Sep. 02, 2018. arXiv. Accessed: Sep. 14, 2023. [Online]. Available: http://arxiv.org/abs/1809.00437

[58] W. Zhang, Y. Liu, C. Dong, and Y. Qiao, "RankSRGAN: Generative adversarial networks with ranker for image super-resolution," Aug. 26, 2019. arXiv. Accessed: Sep. 14, 2023. [Online]. Available: http://arxiv.org/abs/1908.06382

[59] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," May 2017. arXiv:1609.04802 [cs, stat], Accessed: Jun. 14, 2021. [Online]. Available: http://arxiv.org/abs/1609.04802

[60] M. Bevilacqua, A. Roumy, C. Guillemot, and M. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 135.1–135.10, doi: 10.5244/C.26.135.

[61] L. Courtrai, M.-T. Pham, and S. Lefèvre, "Small object detection in remote sensing images based on super-resolution with auxiliary generative adversarial networks," *Remote Sens.*, vol. 12, no. 19, Jan. 2020, Art. no. 19, doi: 10.3390/rs12193152.

[62] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. 7th Int. Conf. Curves Surfaces*, vol. 6920, 2010, Paper 730, doi: 10.1007/978-3-642-27413-8_47.

[63] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, 2001, pp. 416–423, doi: 10.1109/ICCV.2001.937655.

[64] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5197–5206, doi: 10.1109/CVPR.2015.7299156.

[65] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," Mar. 2019. arXiv:1812.04948 [cs, stat], Accessed: Sep. 17, 2021. [Online]. Available: http://arxiv.org/abs/1812.04948

[66] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," Dec. 06, 2017. arXiv, doi: 10.48550/arXiv.1701.07875.

[67] G. Montavon, K.-R. Müller, and M. Cuturi, "Wasserstein training of restricted Boltzmann machines," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, Inc., 2016. Accessed: Jul. 11, 2023. [Online]. Available: https://papers.nips.cc/paper_files/paper/2016/hash/728f206c2a01bf572b5940d7d9a8fa4c-Abstract.html

[68] D. Wang and Q. Liu, "Learning to draw Samples: With application to amortized MLE for generative adversarial learning," Nov. 25, 2016. arXiv, doi: 10.48550/arXiv.1611.01722.

[69] G.-S. Xia et al., "DOTA: A large-scale dataset for object detection in aerial images," May 19, 2019. *arXiv*. Accessed: Apr. 14, 2023. [Online]. Available: http://arxiv.org/abs/1711.10398

[70] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," Dec. 2017. arXiv:1704.00028 [cs, stat], Accessed: Jun. 16, 2021. [Online]. Available: http://arxiv.org/abs/1704.00028

[71] Z. Kurczyński, *Lotnicze i Satelitarne Obrazowanie Ziemi*, 1st ed. Warszawa, Poland: Oficyna Wydawnicza Politechniki Warszawskiej, 2006.

[72] L. Salgueiro Romero, J. Marcello, and V. Vilaplana, "Super-resolution of Sentinel-2 imagery using generative adversarial networks," *Remote Sens.*, vol. 12, no. 15, Jan. 2020, Art. no. 15, doi: 10.3390/rs12152424.

[73] M. T. Razzak, G. Mateo-García, G. Lecuyer, L. Gómez-Chova, Y. Gal, and F. Kalaitzis, "Multi-spectral multi-image super-resolution of Sentinel-2 with radiometric consistency losses and its effect on building delineation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 195, pp. 1–13, Jan. 2023, doi: 10.1016/j.isprsjprs.2022.10.019.

[74] Y. Li and B. Li, "Super-resolution of sentinel-2 images at 10m resolution without reference images," Preprints, Apr. 21, 2021, doi: 10.20944/preprints202104.0556.v1.

Kinga Karwowska received the M.Sc. degree in geoinformatics from Military University of Technology, Warsaw, Poland, in 2020. She is currently working toward the Ph.D. degree in Doctoral School of the Military University of Technology, Warsaw, Poland.

Damian Wierzbicki received the Ph.D. degree in photogrammetry and remote sensing from Military University of Technology, Warsaw, Poland, in 2015.

He is currently an Associate Professor with the Department of Imagery Intelligence, Faculty of Civil Engineering and Geodesy, Military University of Technology where he teaches: "Photogrammetry and Remote Sensing" and "Image Processing." His research interests include UAV navigation and image processing, deep learning in remote sensing. His research interests also include development of new algorithms for object detection and classification in satellite optical and radar imagery.

Article

Modified ESRGAN with Uformer for Video Satellite Imagery Super-Resolution

Kinga Karwowska * and Damian Wierzbicki

Department of Imagery Intelligence, Faculty of Civil Engineering and Geodesy,
Military University of Technology, 00-908 Warsaw, Poland; damian.wierzbicki@wat.edu.pl
* Correspondence: kinga.karwowska@wat.edu.pl

Abstract: In recent years, a growing number of sensors that provide imagery with constantly increasing spatial resolution are being placed on the orbit. Contemporary Very-High-Resolution Satellites (VHRS) are capable of recording images with a spatial resolution of less than 0.30 m. However, until now, these scenes were acquired in a static way. The new technique of the dynamic acquisition of video satellite imagery has been available only for a few years. It has multiple applications related to remote sensing. However, in spite of the offered possibility to detect dynamic targets, its main limitation is the degradation of the spatial resolution of the image that results from imaging in video mode, along with a significant influence of lossy compression. This article presents a methodology that employs Generative Adversarial Networks (GAN). For this purpose, a modified ESRGAN architecture is used for the spatial resolution enhancement of video satellite images. In this solution, the GAN network generator was extended by the Uformer model, which is responsible for a significant improvement in the quality of the estimated SR images. This enhances the possibilities to recognize and detect objects significantly. The discussed solution was tested on the Jilin-1 dataset and it presents the best results for both the global and local assessment of the image (the mean values of the SSIM and PSNR parameters for the test data were, respectively, 0.98 and 38.32 dB). Additionally, the proposed solution, in spite of the fact that it employs artificial neural networks, does not require a high computational capacity, which means it can be implemented in workstations that are not equipped with graphic processors.

Keywords: singe-image super-resolution; generative adversarial networks; convolutional neural networks; video satellite; super-resolution

1. Introduction

In recent years, we have been witnessing a rapid development in Very-High-Resolution Satellite imaging (VHRS). It may be applied in numerous fields: landcover mapping [1,2], urban mapping [3], the detection and tracking of objects [4–7], maritime monitoring [8,9], automatic building classification, etc., [10,11].

Contemporary very-high-resolution satellites are capable of recording images with a spatial resolution of less than 0.30 m. However, until now, these scenes were acquired in a static way. The new technique of the dynamic acquisition of video satellite imagery has been available only for a few years. It has multiple applications related to remote sensing. However, in spite of the offered possibility to detect dynamic targets, its main limitation is the degradation of the spatial resolution of the image that results from imaging in video mode, along with a significant influence of lossy compression. Satellite observations in video mode offers new possibilities in comparison to traditional satellite systems for the observation of the Earth. This technology may be used to track targets or to capture a sequence of video images instead of a static scene [12,13]. As it has been mentioned before, video imaging is characterized by high temporal resolution; however, at the cost of lower

spatial resolution in comparison to VHRS systems. Methods that allow for the improvement of spatial resolution may be divided into two groups: the first one includes multiple-image super-resolution methods [14], while the other one is based on single-image super-resolution [15]. As far as video imaging is concerned, the multiple-image methods cannot always be applied as the content of the image changes too quickly to be able to use a combination of several sequences of images to improve the resolution. Another common barrier is the presence of clouds, shadows, and moving objects, as well as changes in plant vegetation.

However, classical methods of spatial resolution enhancement for image sequences, as well as images acquired from space, often exhibit:

- A lack of improvement in interpretational capabilities.
- The occurrence of spectral distortions.
- Color distortions.
- For pansharpening-based methods, it is necessary to have an identical high-resolution panchromatic image.
- The occurrence of artifacts in the case of video data where objects move quickly and the temporal resolution of the data is low.
- Correctness of operation only for specific types of data.

As a result, one of the solutions to the above problem is the application of Single-image super-resolution (SISR) algorithms that are based on Deep convolutional neural networks (CNNs). In the last decade, many Super-Resolution (SR) techniques were introduced that were developed with the use of deep learning methods. Some of the popular techniques include Generative Adversarial Networks (GAN) [16] and very deep convolutional networks (VDSR) that show promising results on generating realistic HR imagery from low-resolution (LR) input data [17]. Moreover, they show promising potential in terms of their high resolution (HR) and computational speed.

The applications of super-resolution algorithms may be found in multiple research studies. The authors of [18] applied SRCNN and VDSR solutions for Pleiades and SPOT imagery. On the other hand, Lanaras et al. [19] increased the resolution of channels in the Sentinel 2 imagery from 20 m and 60 m to 10 m with the use of high-resolution bands in order to transfer spatial details from lower-resolution channels. The developed architecture was named DSen2 and was based on the EDSR solution [20]. A similar solution was applied in the studies conducted by [21] with the difference that Sentinel-2 images were used to train the neural network in order to improve the spatial resolution for Landsat imagery.

A much greater challenge lies in improving the quality of the image sequences acquired by nanosatellites. In this regard, researchers must consider the following challenges:

- A large number of images (frames), which increases the computational demands on processing units or prolongs the time required for data processing.
- Maintaining the estimation style of images for all images belonging to the sequence.
- The occurrence of noise and disturbances resulting from imperfect optical systems.
- Low temporal resolution due to limited computational power of the computers on board nanosatellites.

The authors of the present article propose a new solution that allows for the improvement of the spatial resolution of video data. This method belongs to the group of Single-Image Super-Resolution (SISR) solutions and it uses the GAN network extended by an encoder–decoder network that is responsible for improving the quality of SR images. The method presented here introduces the following improvements:

- An individual approach to each video frame enables the processing of video data of any temporal resolution;
- The frames are divided into smaller fragments (tiles), which allows the improvement of the spatial resolution of video data of unlimited dimensions;

- The use of a 10% overlap between tiles enables the achievement of the maximum quality of SR images while, at the same time, minimizing the amount of time required to improve the resolution of the video data (a higher overlap leads to a larger number of images to be processed);
- It is recommended to train the network using databases that consist of images of various spatial resolutions, as this allows for the model training to be made even more resistant to the vanishing gradients phenomenon (the probability of its occurrence for satellite data is much higher than in the case of images obtained from the surface of the Earth);
- Developing the GAN network model by layers of the Uformer model (the encoder–decoder network) allows for a significant improvement of the quality of the estimated SR images;
- The application of time windows to connect the tiles estimated by the GAN network generator enables the combination of them into one video data frame;
- The global quality assessment does not reflect the actual quality of the estimated SR images, but the application of the local assessment and image evaluation in the frequency domain enables a better assessment of SR images.

This paper consists of the following sections: Section 2 presents a review of the methods to improve the spatial resolution of video sequences with the use of deep neural networks. The description of the proposed methodology is provided in Section 3. Section 4 contains a description of the research and the results. Section 5 is the discussion, while the final conclusions are presented in Section 6.

2. Related Works

The need to improve the spatial resolution of a sequence of images poses additional requirements: the method's operation has to be fast (due to a large number of frames), enable processing images of various dimensions, and the estimated frames must be recreated in the same style (as estimation errors will be visible in subsequent images in the sequence). Algorithms based on SISR may be divided into solutions that employ interpolation, reconstruction, and machine learning. The issue of super-resolution is a typical problem of connecting the relations between low-resolution (LR) and high-resolution (HR) images [22]. On the other hand, other sharpening methods, in particular concerning video sequences, were based on motion blurring and hybrid regularization [23]. Yet another method used discrete and stationary wavelet transforms to increase the spatial resolution and to bring out the details of a video sequence [24]. With the development of modern deep learning algorithms, more and more attention has been paid to developing new methods related to the improvement of the spatial resolution of the imagery. One of the first research works was the study by [25], which presented the results of enhancing the resolution of an image with the use of Deep Convolutional Networks. Other authors proposed using an hourglass-shaped convolutional neural network structure (FSRCNN) to accelerate SRCNN [26]. Yet another method that employs the Deep Back-Projection Network (DBPN) allowed for an even eight-fold increase in the resolution of images by connecting a series of iterative up- and down-sampling steps [27]. Furthermore, Yu Xiao et al. [28] addressed the issue of enhancing the spatial and temporal resolution of satellite recordings. They introduced a module of attribute interpolation that is equipped, among others, with multi-scale deformable convolution to predict unknown frames and a multi-scale spatial–temporal transformer is proposed to aggregate the contextual information in long-time series video frames effectively. Other authors [29] introduced the HiRN model that employs the hierarchical recurrent propagation and residual block-based backbone with a temporal wavelet attention (TWA) module. Another solution consists of the application of a generative adversarial network to perform multi-frame super-resolution (MFSR). An example is the MRSISR model [30], which introduces comments to the module generator and space-based net that is

responsible for improved information extraction. Jin et al. [31] noted that most of the existing VSR methods focused mainly on the local information between frames, which lack the ability to model long-distance correspondence. In order to solve this problem, they proposed a two-branch alignment network with an efficient fusion module. This solution uses deformable convolution (DCN) and transformer-like attention. Another solution that uses information between frames is the framework model proposed by Liu et al. [32]. These authors proposed a framework based on locally spatiotemporal neighbors and nonlocal similarity modeling. This allowed them to use local motion information without explicit motion estimations. Yet another VSR solution that employs artificial neural networks is the model proposed by He et al. [33]. They proposed a method that consists of three modules: the degradation estimation module (which estimates the blurring and the level of noise in LR images), the intermediate image generation module (that generates frames of intermediate super-resolution), and the multi-frame feature fusion module (feature fusion subnetwork is leveraged to fuse the features from multiple video frames).

Another literature review is [34]. The authors of this publication highlighted the difficulties and challenges of enhancing the resolution of the sequences of images acquired by satellites. They pointed to poor data continuity, global motion due to platform movement, large changes in illumination, large redundancy between video frames, an extreme foreground–background imbalance, a complex background environment, a huge scene size, and, finally, large differences in object scales. Apart from a detailed overview of the methods to improve the resolution of image sequences, the study also contains a wide presentation of Satellite VSR methods. The authors also described the public Dataset: VISO dataset [35], SatSOT dataset [36], and Air-MOT dataset [37]. Another database that has been published in recent years is the Jilin-189 proposed by Wuhan University [38]. Li et al. noted that in the last few years, there has been a dynamic increase in the number of publications concerning VSR.

A review of these publications reveals that multiple methods use SISR solutions to improve the spatial resolution of image sequences. The number of frameworks that present proposals of solutions to the SISR problem is several times higher that the number of solutions that focus on VSR. In the last few years, numerous overviews of the SISR methods were published. An example may be the study by Wang et al. [39]. In 2022, we also noted the problem of enhancing the resolution of imagery acquired by Small Satellites. The publication [40] contains a review of the methods to enhance the spatial resolution of satellite imagery, with particular focus on enhancing the resolution of images acquired by nanosatellites. In our review, we highlighted the methods that employ pansharpening and SISR methods that use interpolation, Digital Image Processing, and artificial neural networks. During our research, we noted that, in spite of their advanced, sophisticated structure, some types of architecture based on convolutional networks cannot be successfully applied in SISR satellite imaging because they do not improve interpretation abilities in spite of the enhanced resolution.

After a series of tests, we found that SR images estimated by SISR algorithms that employ artificial intelligence often had characteristic textures or blurred edges. An example of an algorithm that focuses on removing the noise from an image is the solution proposed by [41]. The authors of [42] modified the variance model based on the edge adaptive-guiding function. It uses an adaptive function that is responsible for enhancing the edges. It employs a standard gradient into the non-linear diffusion term. Another solution [43] is the non-linear filter, whose task is to remove constant-impulse noise (e.g., salt and pepper) from multiple-channel images. The authors of this method performed a fusion of a standard median filter and morphological operation. Among noise removal methods, one may distinguish solutions that employ non-blind techniques [44]. These authors proposed new images prior based on a parameterized scaled Gaussian model and a gamma distribution, with hyperparameters based on the statistical properties of tens of thousands of images. Another solution is the Uformer [45], where the novel, locally

enhanced window was introduced (LeWin), with a transformer block and a learnable multi-scale restoration modulator in the form of a multi-scale spatial bias to adjust features in multiple layers of the Uformer decoder. This method stands out for its ability to capture both local and global features of images, making it applicable in noise removal and image deblurring tasks. The resulting images are characterized by a high quality (PSNR/SSIM: 30.90 dB/0.953 for HIDE database [46], 36.19 dB/0.956 for RealBlur-R database [47], and 29.09 dB/0.886 for RealBlur-J database [47]), ref. [45], compared to other methods used for image denoising (for example: Nah [48], DeblurGAN [49,50], DeblurGAN-v2 [51], DBGAN [52], SPAIR [53,54], SRN [55], DMPHN [56], and MPRNet [57]).

3. Methodology

When we were conducting our research on MCWESRGAN (Multicolumn Wasserstein Enhanced Super-Resolution Generative Adversarial Networks) [58], we noted that SR images were characterized by a specific texture. Based on this finding, we decided to search for a method that would improve the quality of images by limiting the occurrence of deformations.

The problem of the quality of interpretation of image context is very important in the case of video data, which contain large, complex patterns that change their position in consecutive images of the sequence. The change in the position of objects in consecutive images in a sequence results from the movement of objects in the image and the movement of the satellite that acquires data. Furthermore, the quality of representation of building infrastructure (including critical infrastructure) and road infrastructure is essential in tasks that consist in detecting smaller objects, e.g., vehicles or airplanes.

During our research, we noted that both classical methods of processing digital imagery and methods of blur removal, i.e., the Lucy–Richardson Algorithm [59] and Wiener deconvolution [60], fail to solve the problem (which was presented in Section 4). Due to that, we decided to find a solution that would employ deep neural networks. Based on empirical experiments, we noted that combining the Uformer model with a MCWESRGAN network allows for the removal of mistakes generated by the SISR network. Thanks to the application of the locally enhanced window (LeWin) Transformer block mechanisms, the extended model interprets the global and local context of an image better. Consequently, the resulting SR images are characterized by higher quality and improved interpretation capacity. Moreover, the conducted literature review demonstrated that expanding SISR algorithms by adding models that improve the quality of estimated SR images was not used by other researchers. Due to that, the proposed solution is innovative and may be commonly applied to enhance the spatial resolution of digital imagery.

This paper presents a method to enhance the resolution of video sequences with the use of Generative adversarial network (GAN) of a small input size (96 × 96 pixels). The effectiveness of the proposed method of image processing was verified with the use of evaluation metrics presented in Section 3.2.

3.1. Proposed Methodology

The aim of the conducted research was to develop a method to enhance the resolution of sequences of satellite images while at the same time recreating as much of the informational content of images as possible. For the purposes of our research, we propose an approach that takes into account the Generative Adversarial Networks' architecture. These networks enable us to generate images of increased resolution by forcing the artificially generated images to be non-distinguishable from original high-resolution images in terms of statistical parameters [16]. The proposed methodology enables the improvement of the spatial resolution of video data that may consist of any number of frames whose size is not precisely defined. It consists of five main stages: (1) loading a single video frame, (2) dividing the frame into smaller tiles (the dimensions are determined by the input_size parameter of the GAN network) with an overlap of 10% (the

value of overlap between images was calculated in another research project of the authors [61]), (3) improving spatial resolution with the use of the SISR algorithm (e.g., a GAN network generator), (4) enhancing the interpretational quality of SR1 image with the use of artificial neural networks (Uformer model), and (5) recombining the estimated images (SR2) into a single frame. The diagram of the presented methodology is shown in Figure 1.
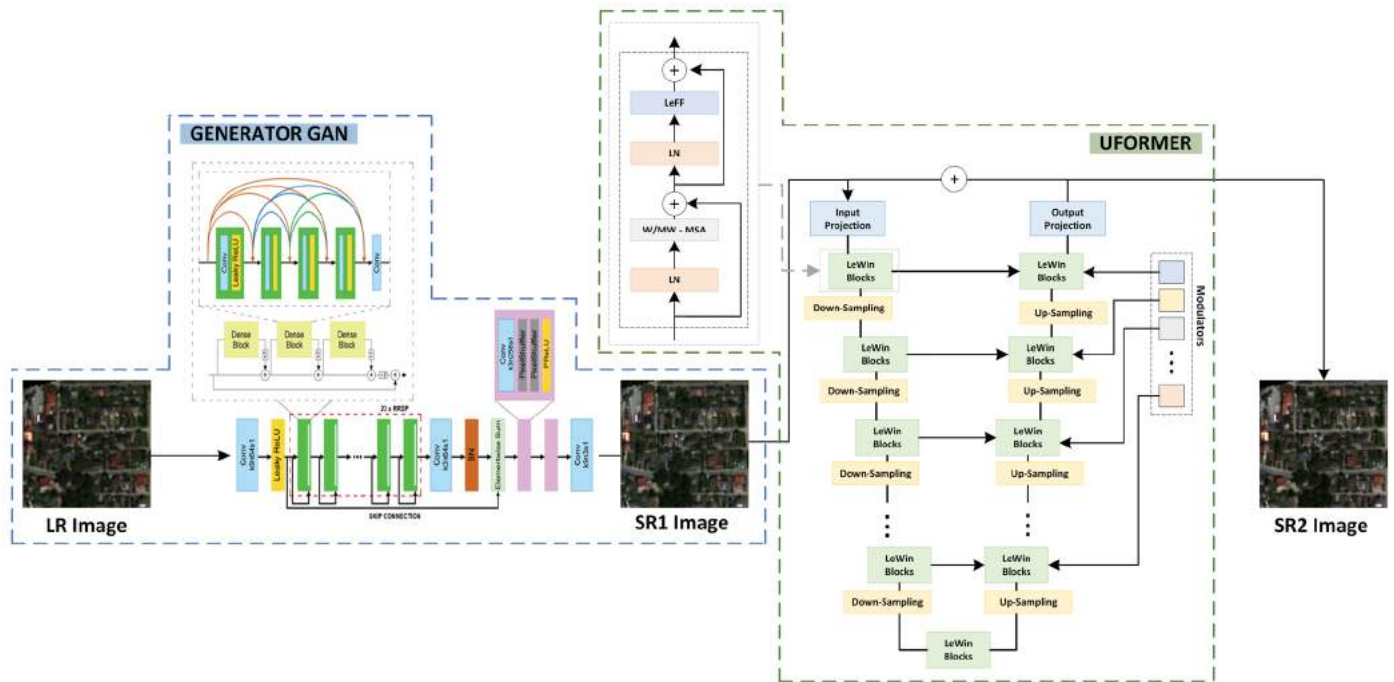
Figure 1. Diagram of enhancement of spatial resolution of a single video frame.

In order to conduct the tests to determine the quality of the proposed method, the network used as the one responsible for enhancing spatial resolution was the MCWESRGAN network created by the authors, which was described in detail in the publication [58]. The quality of SR images was improved with the use of the Uformer algorithm that employs the locally enhanced window Transformer block. The resulting images were combined into a single frame with the use of the Bartlett–Hann Window Functions.

3.1.1. MCWESRGAN

The resolution of each tile was enhanced with the use of the MCWESRGAN model designed by the authors. When preparing the MCWESRGAN network, we modified one of the most popular SISR networks, i.e., the ESRGAN [62]. In our work, we focused mainly on the discriminator and the generator learning technique. The discriminator of generative adversarial networks (GAN) is a convolutional neural network that is responsible for the evaluation of images during the training of the generator. In the MCWESRGAN network, we used a discriminator consisting of two branches: the first of these classifiers has the same structure of layers as the discriminator of the SRGAN network, the second one was expanded by adding another convolutional layer and, for every layer, ReLU activation was applied. Moreover, for even convolutional layers, the kernel size was increased to 5 (pixels) (Figure 2). The introduced modification allowed for a much better assessment of the generator work quality (two scores are received from each critic), which, in consequence, will enable the acceleration of the generator training process.
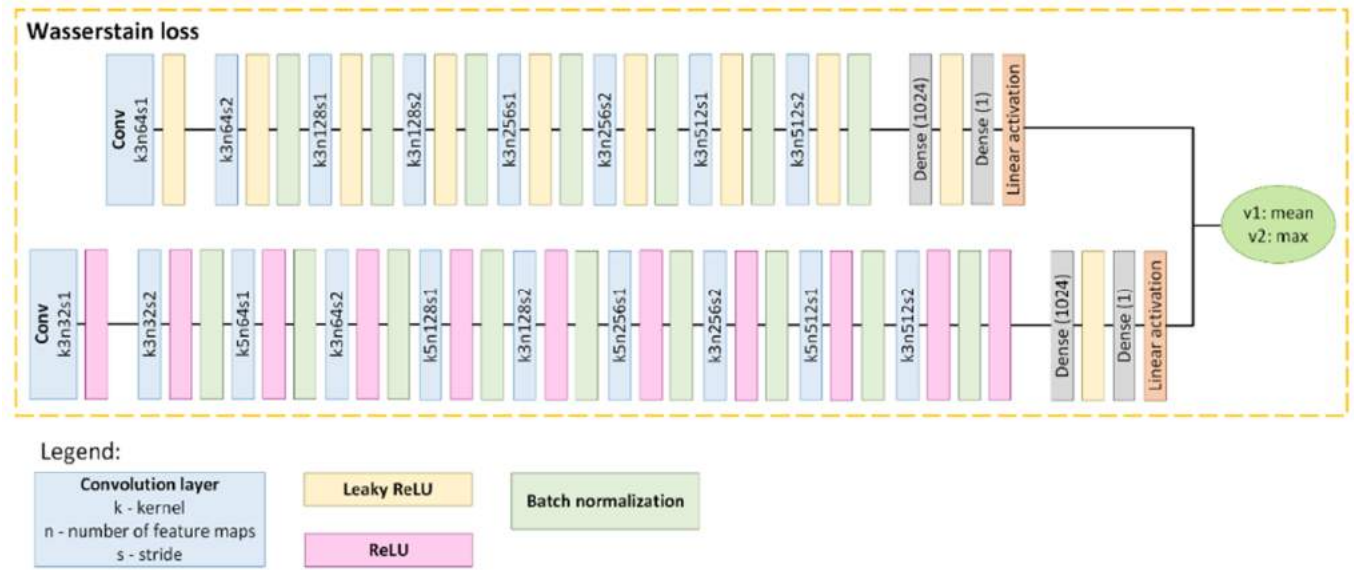
Figure 2. Discriminator model [58].

The MCWESRGAN network generator was trained with the use of Wasserstein Loss. The main task of this method is to minimize the distance between the HR and SR images. The distance is determined with the use of the Wasserstein distance method (Equation (1)).

$$W(\mu, \vartheta) = \max_{\alpha, \beta} \left[\mathbb{E}_x[\alpha(X)] + \mathbb{E}_{x'}[\beta(X')] - \gamma \sum_{x, x' \in \{0,1\}}^{d} \exp\left(\frac{1}{\gamma}\big(\alpha(x) + \beta(x') - D(x, x')\big) - 1\right) \right] \quad (1)$$

where μ, υ—probabilistic measures, D: $X \times X \rightarrow R_+$, $X = \{0,1\}^d$, W is Wasserstein distance, α, β are functions in the set X, and γ is joint probability distribution.

The aim of training the MCWESRGAN network with the use of Wasserstein Loss is to strive to achieve alignment between the generator. For this purpose, Kantorovich–Rubinstein duality is used (Equation (2)) [63,64].

$$W(P_r, P_\theta) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x \sim P_r}[f(x)] - \mathbb{E}_{z \sim P_\theta}[f(x)] \quad (2)$$

where P_r is real data distribution, P_θ is the distribution of the parametrized density P_θ, and $f: X \rightarrow \mathbb{R}$.

For GAN (composed of a generator and a discriminator), Wasserstein Loss takes Equation (3).

$$\min_G \max_{D \in \mathcal{D}} \mathbb{E}_{x \sim P_r}[D(x)] - \mathbb{E}_{\tilde{x} \sim P_g}[D(\tilde{x})] \quad (3)$$

where P_r is data distribution, P_g is the model distribution implicitly defined by $\tilde{x}$ = G(z), z ~ p(z), and $\mathcal{D}$ is set of 1-Lipschitz functions.

The introduced modifications allowed us to significantly (even by 10 times) shorten the time required to train the generator while improving the quality of super-resolution (SR) images, which has been proven in our paper.

3.1.2. Uformer

Training a GAN network is a very demanding process, during which one may often encounter the phenomenon of exploding or vanishing gradients. We discussed this phenomenon in detail in one of our other works [61]. This phenomenon may be reduced by decreasing the learning parameter, which allows to limit the problem of vanishing gradients and, in consequence, to achieve the optimum alignment of the model. However,

this solution has one disadvantage, which is the emergence of a specific, unnatural texture in SR images (visually, it resembles a combination of deblurring, denoising, and deraining). In order to avoid this phenomenon, the MCWESRGAN network generator was expanded by adding the Uformer model presented by Zhendong Wang et al. [45]. The authors of this model introduced the locally enhanced window (LeWin) transformer block and a learnable modulator. Moreover, the authors recommended using Charbonnier loss [65,66] to train this model.

The *Uformer* model is of the nature of a classical U-shaped encoder–decoder network (a diagram of the network is presented in Figure 1). Additionally, skip connections were used between the encoder and decoder, which enable adding the features detected by the algorithm to the input image ($I \in \mathbb{R}^{3 \times H \times W}$). The first stage of operation of the model is to extract low-level features $X_0 \in \mathbb{R}^{C \times H \times W}$ with a convolutional layer from a kernel of the size 3×3 and activation LeakyReLU. Then, the tensor undergoes further encoder stages, where each of the stages contains a stack of the LeWin Transformer blocks and one convolutional down-sampling layer (stride: 2). Additionally, LeWin uses the self-attention mechanism for capturing long-range dependencies. Moreover, thanks to the overlapping of windows on feature maps, it derives the calculation cost that results from the use of self-attention. Apart from that, the authors of this solution provide an example, where for feature maps, $X_0 \in \mathbb{R}^{C \times H \times W}$ and the l-th degree of the encoder, feature maps equal $X_l \in \mathbb{R}^{2^l C \times \frac{H}{2^l} \times \frac{W}{2^l}}$. The encoder of the model ends with a bottleneck created from the stack of LeWin Transformer blocks.

The task of the module is to obtain information about the local context of "contaminated" pixels in order to restore their "clean" version. For this purpose, LeWin transformer blocks are used. Moreover, the authors of this solution introduced a convolution operator into the Transformer to capture useful local context. In the feature maps, the LeWin Transformer block meets the normalization layer (LN) first. Then, it is transferred to the Window-based Multi-head Self-Attention (W-MSA) that uses self-attention inside the non-overlapping local windows. In the next step, the skip connection is used to add this result to input feature maps and the result is transferred to the next normalization layer. The normalized tensor is transmitted to the locally enhanced Feed-Forward Network (LeFF), and the resultant tensor is once again (with the use of skip connection) added to the feature maps that were transferred to the LeWin Transformer blocks. Furthermore, the application of the bottleneck stage with a stack of LeWin Transformer blocks allows the capture of additional dependencies of features, and if the size of the window equals the size of the Transformer block, global dependencies are captured.

The network decoder is equipped with modulators. The aim of this module is to calibrate the features, which enable recovering a larger number of details (by taking into account the characteristics of interfering patterns). The multi-scale restoration modulator is embedded in each block of the LeWin Transformer, taking a form of a tensor of the shape W × W × C, where W is the window size and C is the size of the channel of the current feature map. The modulator takes the form of a matrix that is added to all nonoverlapping windows before the self-attention module.

The branch of the decoder module is responsible for the reconstruction of features. It is built of the up-sampling layer and a stack of LeWin Transformer blocks. In this model, the convolution layers were replaced with transposed convolution (convolution: 2 × 2, stride: 2). In accordance with the principle of functioning of transposed convolution, it is responsible for increasing the feature maps twofold and reduces half of the feature channels. The authors of the solution used the LeWin Transformer blocks to learn to restore the image. In order to obtain the resulting image $R \in \mathbb{R}^{3 \times H \times W}$, the authors propose 3×3 convolution layer, which will allow for the feature maps to be flattened to 2D feature maps. The resultant image is a result of adding features to the input image $I' = I + R$.

3.1.3. Bartlett–Hann Window Function

The Bartlett–Hann Window Function was used to recombine the SR images into a single frame. This Window Function was not selected at random. The choice was motivated by the research results presented by the authors in the publication [61], which provides a comparative analysis of various Window Functions for various degrees of overlapping between neighboring tiles. During the research, it was noted that the best quality of the resultant image was obtained when the tiles of the SR image were combined with the use of Hann, Hann–Poisson, Bartlett–Hann, or Triangular window functions.

For the purposes of this research project, the Bartlett–Hann Window Function was used to recombine the tiles into a single image. In this method, for each pixel in the image fragment (which is in the area of overlap between images), the weight is calculated (Equation (4)), with which the given pixel will be added to the resultant SR image.

$$w[n] = 0.62 - 0.48 \cdot \left| \frac{n}{N} - \frac{1}{2} \right| - 0.38 \cdot \cos\left(\frac{2\pi n}{N} \right) \tag{4}$$

where n is position of the pixel in the image and N is tile size.

3.2. Evaluation Metrics

The quality assessment of the resulting SR images consisted of three stages. The first one consisted of conducting a global quality assessment of all SR images before combining them into images of the sequence. This was performed with the use of the most popular metrics that are used in the fields of remote sensing and computer vision (Table 1).

Table 1. Presentation of the main evaluation metrics.

Peak signal-to-noise ratio (PSNR) [67]	$PSNR = 10 \cdot \log_{10} \dfrac{\left[\max\left(HR(n,m)\right)\right]^2}{MSE}$
Universal Quality Measure (UQI) [68]	$Q = \dfrac{\sigma_{HRHR\prime}}{\sigma_{HR}\sigma_{HR\prime}} \cdot \dfrac{2 \cdot \overline{HR} \cdot \overline{HR\prime}}{(\overline{HR})^2 (\overline{HR\prime})^2} \cdot \dfrac{2\sigma_{HR}\sigma_{HR\prime}}{\sigma_{HR}^2 + \sigma_{HR\prime}^2}$
Spectral Angle Mapper (SAM) [69]	$SAM(v,w) = \cos^{-1}\left(\dfrac{\sum_{i=1}^{N} HR\prime_i HR_i}{\sqrt{\sum_{i=1}^{N} HR\prime_i^2} \sqrt{\sum_{i=1}^{N} HR_i^2}} \right)$
Structural Similarity Index (SSIM) [67]	$SSIM(x,y) = \dfrac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}$
Root mean square error (RMSE) [67]	$RMSE = \sqrt{\dfrac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} [HR(n,m) - HR\prime(n,m)]^2}$
Spatial Correlation Coefficient (SCC) [67]	$CC = \dfrac{\sum_{mn}(LR - \overline{LR})\,(HR\prime - \overline{HR\prime})}{\sqrt{\sum_{mn}(LR - \overline{LR})^2 \sum_{mn}(HR\prime - \overline{HR\prime})^2}}$

When performing the global assessment of the quality of SR images, one should note that the results are the average evaluation of the whole image. Due to that, it is difficult to determine how specific areas of the image are represented in the SR images. As a result, the second stage of the SR image evaluation involves local assessment. In this method, the SR image is divided into smaller fragments of a defined surface area, and then evaluation metrics values are calculated for these fragments. This method enables one to determine which areas of the SR image are better represented, and which are poorer.

The final method of assessing the quality of SR images is the determination of the physical enhancement of the spatial resolution of SR images by means of power spectral density analysis (PSD) of the image. PSD explains the distribution of the frequency of signal power. This method allows you to determine the improvement in the ability to detect objects in the image using the Ground Resolved Distance (GRD) parameter.

4. Experiments and Results

4.1. Datasets

While training generative adversarial networks, one should bear in mind that this type of network is very prone to the phenomenon of vanishing gradients. During the tests, it was found that databases that consisted of images of varied spatial resolution were more stable during the training of the generator. Due to that, the final database that was the basis for training the prepared network contained images obtained by unmanned aerial vehicles (UAV), optic sensors mounted on airplanes, observation satellites, and nanosatellites. As a result, the created database contained images of a spatial resolution ranging from 0.15 m to 20 m and consisted of 100,000 images. Such a diversified database was created as a result of combining the Dataset for Object deTection in Aerial Images (DOTA) [70], fragments of aerial images (that were a result of dividing the orthophotomap), imagery acquired by the WorldView-2 (WV2) satellite, the Jilin-1 database [38], and other data acquired by the SkySat-1 and Jilin-1 nanosatellites. The Jilin database contains video data acquired by the Jilin-1 satellite. It consists of 201 clips, where each clip contains 100 images of the size of 640 × 640 pixels. Apart from that, other video sequences captured by the Jilin-1 mission were added to the database [71]:

- Beirut in Liban—collected on 06 October 2017, the resolution: 1.12 m, the frame size: 1920 × 1080 pixels, 25 frames per second (FPS), and video duration: 32 s.
- Florence in Italy—collected on 10th September 2018, the resolution: 0.9 m, frame size: 3840 × 2160 pixels, 10 FPS, and video duration: 31 s.
- Other materials added to the database were the video sequences collected by SkySat-1 from Las Vegas (25 March 2014), Burj Khalifa (9 April 2014), and Burj al-Arab (1 February 2019) [72].

In order to train the models that are responsible for enhancing the resolution, it is necessary to prepare high-resolution (HR) and low-resolution (LR) data. To create the database used here, most of the HR images were prepared by dividing images into tiles of the dimensions of 384 × 384 pixels, and the corresponding LR images were prepared as a result of sampling the HR images to the dimensions of 96 × 96 pixels. The exceptions were the satellite scenes acquired by the WV2 satellite. In this case, the LR database was created by dividing the multispectral imagery. To prepare the HR database, we used an image whose resolution was improved using pansharpening (the Gram–Schmidt method).

4.2. Methodology of Processing Video Data

The process of enhancing the spatial resolution of a sequence of images may be divided into three main stages: (1) preparing data, (2) enhancing spatial resolution, and (3) combining the images into sequences. The detailed flowchart is presented in Figure 3.

The process of preparing data for spatial resolution enhancement starts with loading a video file, which, in the next step is divided into single frames. Then each of the frames is divided into smaller tiles (LR images). Due to the fact that the MCWESRGAN network was used to improve the resolution, the frames were divided into fragments of the dimensions of 96 × 96 pixels (the size was determined by the input size parameter of the GAN network used). If the size of the tile is smaller than the input size parameter, the missing areas are filled with pixels of the digital number (DN) equal to 0 (black stripes that are shown in Figure 3). Data prepared in this wat are then transferred to the MCWESRGAN network with the added Uformer network. This module first enhances the spatial resolution of the resulting LR image, and then improves its quality with the use of the Uformer model. When the spatial resolution of all tiles belonging to a single frame has been improved, the third stage of the proposed methodology commences. In this module, the tiles are combined into a single frame with the use of the Bartlett–Hann Window Function, and when all frames have been processed, they are recombined into a sequence of images.
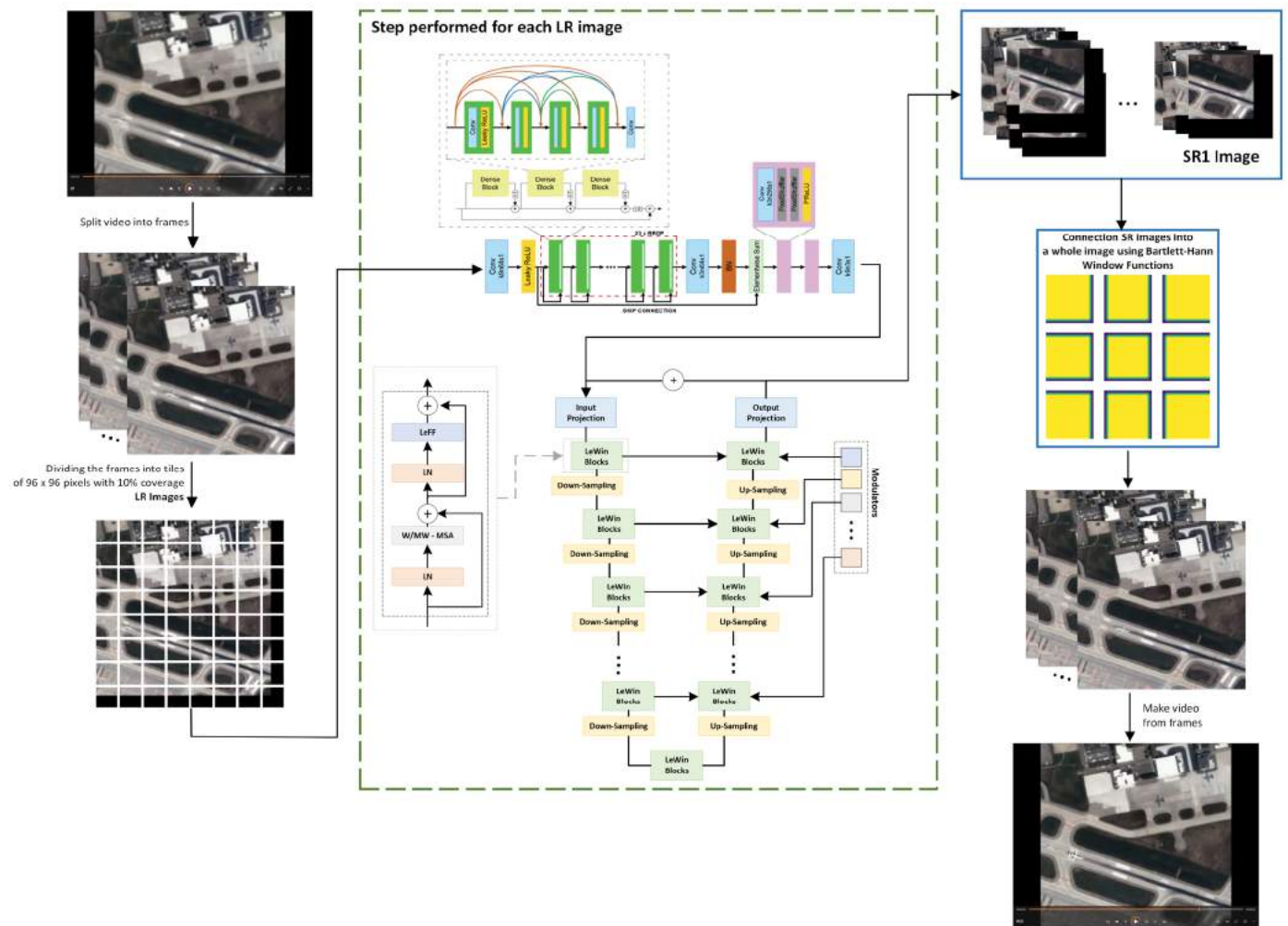
Figure 3. The flowchart of the algorithm.

4.3. Implementation

The SISR model was trained on a Nvidia TITAN RTX 24 GB graphics card, Intel Xeon Silver 4216 processor, and the Ubuntu 18.04 operating system. The network training parameters are presented in Table 2.

Table 2. Network training parameters.

Input size	96 × 96 pixels
Learning rate	$2 \cdot 10^{-4} - 2 \cdot 10^{-6}$
Optimization	Adam ($\beta\_1 = 0.9$, $\beta\_2 = 0.999$, $\varepsilon = 1 \cdot 10^{-8}$)
Batch size	4
Epochs	150

The training dataset was used to train the model that is responsible for enhancing the spatial resolution of LR images. This was the basis for estimating the weights between the layers of the model. The weights were updated based on the results of the assessment of the network model with the use of validation data. Adam optimization was used to find the target function. The application of this model enables the high-quality achievement of the model in a short time. Moreover, as a result of training the generator of the GAN network that is responsible for the SISR of data acquired from space altitude, it is more

resistant to the vanishing gradient phenomenon (in comparison to other optimization methods, e.g., SGD, RMSprop, and Adagrad).

4.4. Results

The proposed MCWESRGAN model, expanded by adding the Uformer model layers, was tested on a set of video frames from the Jinlin-1 dataset (the test dataset). To assess the results, the metrics presented in Table 1 were employed. The obtained results were compared to the SR images estimated by the MCWESRGAN model, from which the specific texture was removed with the use of the Lucy–Richardson Algorithm and Wiener deconvolution. The Lucy–Richardson Algorithm is an iterative method that enables the removal of blurring and noise from an image. The functioning of the algorithm may be presented in the form of Equations (5) and (6).

$$w\hat{u}_j^{(t+1)} = \hat{u}_j^t \sum_i \frac{d_i}{c_i} p_{ij} \tag{5}$$

$$c_i = \sum_j p_{ij} \hat{u}_j^t \tag{6}$$

where $\hat{u}_j^{(t+1)}$ is the image estimated in $(t+1)$ iterations, d_i is the pixel value, and p_{ij} is the spatial displacement between the value of the source pixel j and the value of the observation pixel i.

On the other hand, Wiener deconvolution enables the determination of the range of frequency of the non-disturbed signal. Thanks to this property, it is possible to remove blurring from the image. If the blurring is large, a disadvantage of the solution is the fact that the characteristic texture emerges in the image. In the frequency domain, the Weiner Filter was presented in Equation (7).

$$G(f) = \frac{H^*(f)S(f)}{|H(f)|^2 S(f) + N(f)} \tag{7}$$

where *G(f)*, *H(f)* are the Fourier transforms *g(t)* and *h(t)*, *S(f)* is the average spectral density of the power of the original signal *x(t)*, *N(f)* is the average spectra density of the noise power, and * is the feedback operation.

In order to remove blurring, it is necessary to possess information about the characteristics of the image deformations and the balance parameter. The aim of this parameter is to reduce the occurrence of noise artifacts in the resulting image, i.e., to maintain the balance between the reconstructed image and the image containing noise.

The evaluation of the estimated high-resolution images started with the assessment of the results, and the values of the described quality evaluation indicators for the whole test dataset were calculated.

In Table 3, we presented a qualitative analysis of the test dataset, where we compared the SR images generated by the MCWESRGAN generator enhanced with (1) the Lucy–Richardson Algorithm, (2) Wiener deconvolution, and (3) the Uformer network. Additionally, in Figure 4, we provided examples of images. The results unequivocally demonstrate the superiority of the MCWESRGAN network extended with the Uformer network over the generator enhanced with the Lucy–Richardson Algorithm and Wiener deconvolution. In the case of SR data generated using the Lucy–Richardson Algorithm, in addition to the deterioration in quality, artifacts are visible (mainly at the edges of the image) along with significant color changes. Conversely, for SR images generated using Wiener deconvolution, the distortion of colors is much greater (confirmed by the qualitative analysis).

Table 3. Quality analysis of the test database. The best results are marked in green and the poorest ones in red.

Metrics	LR	MCWESRGAN	MCWESRGAN with Lucy–Richardson Algorithm	MCWESRGAN with Wiener Deconvolution	MCWESRGAN with Uformer
SSIM	0.92	0.96	0.91	0.77	0.98
PSNR	31.37	36.21	23.83	15.18	38.32
SAM	0.07	0.03	0.08	0.10	0.02
SCC	0.15	0.26	0.08	0.08	0.27
UQI	0.991	0.993	0.966	0.818	0.996
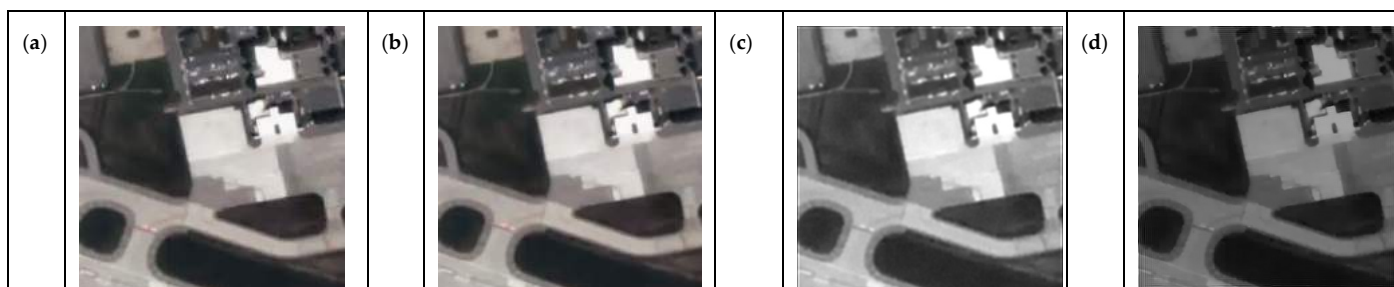
Figure 4. Examples of images from test data with quality results are shown in Table 3: (**a**) HR image, (**b**) MCWESRGAN with Uformer, (**c**) MCWESRGAN with Lucy–Richardson Algorithm, and (**d**) MCWESRGAN with Wiener deconvolution.

Additionally, Figures 5 and 6 show the differences in the values of the SSIM and PSNR metrics for a random video, where, however, the frame has the dimensions of 160×160 pixels (the Jinlin-1 database). Due to the size of the frame, it was divided into four tiles (pixels of DN = 0 were added in the second column and the second line of tiles so that the dimensions would equal the input size). The ID of the image is marked on the X axis of the diagrams. Analyzing the obtained results, one may notice that groups of images of a similar quality exist. Images with ID300-400 are characterized by the highest quality, but they have the added groups of pixels of DN = 0, which increase the metrics of the quality assessment. At the same time, one may notice that the SR images estimated by the MCWESRGAN network with the Uformer model are characterized by higher quality in all fragments of the image.
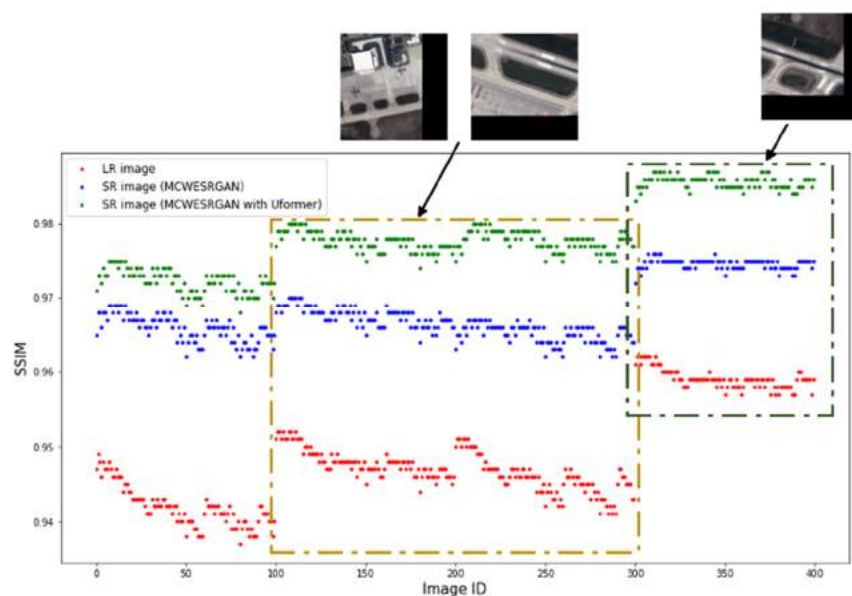
Figure 5. Structural similarity between the estimated images (tiles) (SR) and the reference HR images.
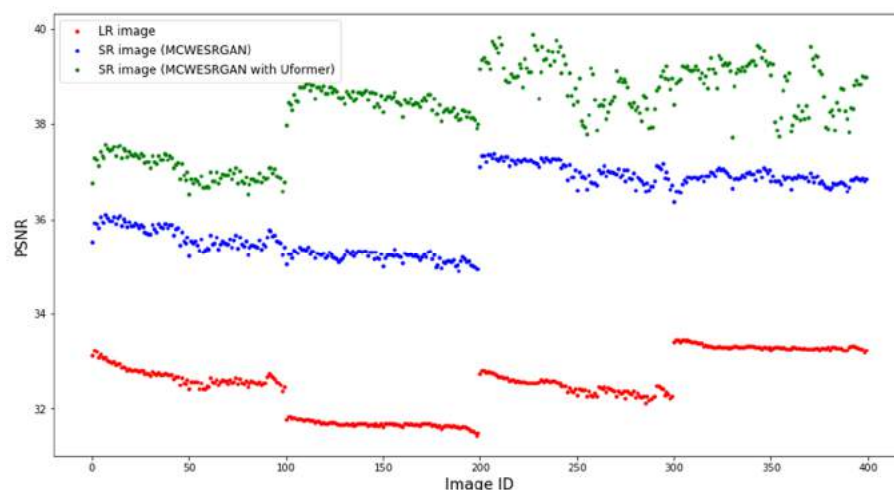
Figure 6. Peak signal-to-noise ratio (PSNR [dB]) between the estimated images (tiles) (SR) and the reference HR images.

In order to improve the evaluation of the estimated image, local image evaluation was used for the resulting frames (after the tiles had been combined with the Window function). Due to the fact that the size of the assessed resulting frames equalled 640 × 640 pixels, it was assumed that the local assessment would be conducted on image fragments of the size of 20 × 20 pixels. The illustrations below (Figures 7 and 8) show the results of the local assessment conducted with the use of the SSIM and PSNR metrics. In the case of the SSIM metric, the majority of the analyzed areas are marked in yellow, indicating a value of the metric exceeding 0.97. There are only isolated areas with a low estimation quality. However, a local assessment using a PSNR enables a more precise analysis of the image—the differences between the adjacent areas under examination are greater. Based on a local assessment using a PSNR, it can be inferred that moving objects (the airplane on the runway) or small objects of irregular shapes were estimated with high quality, although it was slightly lower than that of parking areas, the runway, and taxiways.
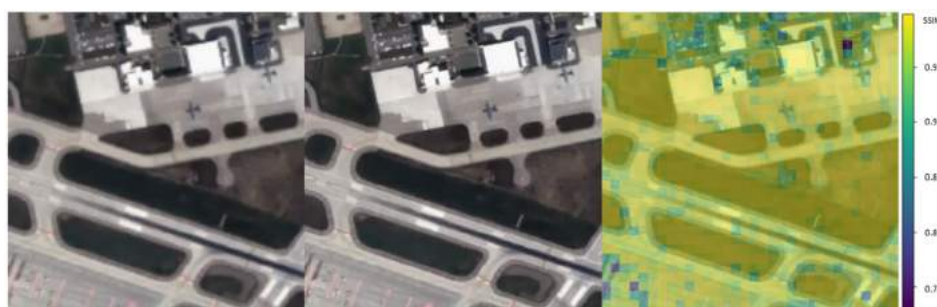
Figure 7. Local assessment—SSIM metrics (for the evaluated field of the size of 20 × 20 pixels).

Figure 8. Local assessment—PSNR metrics (for the evaluated field of the size of 20 × 20 pixels).

Another method to assess the quality of SR images is PSD analysis. The PSD describes the distribution of the frequency of the power signal. This allows for the determination of the smallest size that may be identified in the image.

Figure 9 presents the PSD analysis for the x and y directions. The PSD analysis reveals a high similarity between the signal power frequency distribution in the HR image (yellow curve) and the SR image (green curve), in particular, in the vertical direction of the image (the y direction). In the horizontal direction, resolution enhancement is noticeable for a spatial density of approx. 100 [1/100 m], and in the vertical direction of approx. 120 [1/100 m]. As a result, the value of the GRD parameter on the x direction is GRDx = 2.1 × GSD, and in the y direction, GRDy = 2.5 × GSD. The obtained values confirm that the value of the GRD parameter for high-contrast objects should take the value no lower than two times the GSD.
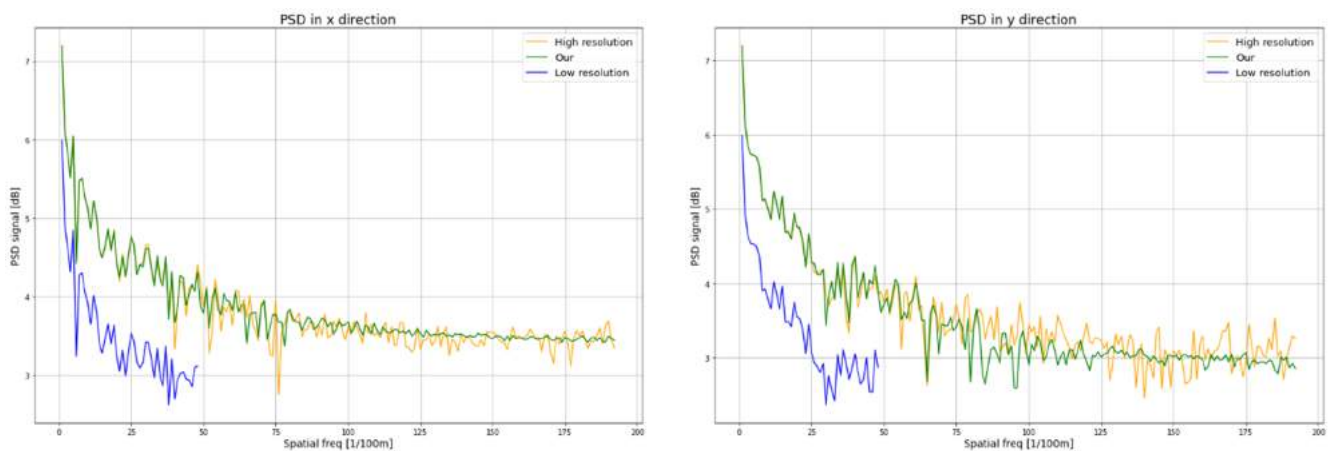
Figure 9. PSD diagram on the x and y directions for a sample image.

5. Discussion

The proposed methodology of enhancing the resolution of a sequence of images consists of three main stages: the preparation of the images, enhancing the spatial resolution, and combining the estimated images into a sequence. The structure of this solution enables one to use other SISR algorithms that use artificial neural networks. During the research, the authors compared the quality of the enhancement of the spatial resolution of images with the use of various methods using convolutional networks. Additionally, each of the analyzed SISR models was extended by adding the Uformer model, which allowed for a better evaluation of the influence of the Uformer model in the quality of the resultant SR images. The results of the global quality assessment of the test dataset are presented in Table 4. Apart from that, examples of SR images estimated by the analyzed SISR models are provided in Appendices A and B.

Table 4. Quality assessmentof convolutional network models dedicated to SISR. The best results are marked in green, the second in yellow, and the worst in red. "*" marks the SISR models with added Uformer network layers.

Metric	LR	EDSR	EDSR *	ESPCN	ESPCN *	FSRCNN	FSRCNN *	RDN	RDN *	SRCNN	SRCNN *	SRDenseNet	SRDenseNet *	SRGAN	SRGAN *	ESRGAN	ESRGAN *	MCWESRGAN	MCWESRGAN *
SSIM	0.92	0.87	0.90	0.95	0.94	0.90	0.93	0.94	0.94	0.93	0.94	0.94	0.94	0.62	0.67	0.78	0.83	0.96	0.98
PSNR	31.37	25.36	25.62	27.63	27.16	27.01	27.23	27.59	27.45	32.40	33.06	27.56	27.46	19.92	20.91	29.35	30.04	36.21	38.32

SAM	0.07	0.11	0.11	0.08	0.08	0.08	0.08	0.08	0.08	0.06	0.06	0.08	0.08	0.23	0.21	0.08	0.080	0.03	0.02
SCC	0.15	0.20	0.26	0.26	0.25	0.18	0.24	0.26	0.25	0.19	0.25	0.24	0.25	0.06	0.17	0.14	0.21	0.26	0.27
UQI	0.991	0.889	0.955	0.939	0.970	0.775	0.904	0.958	0.967	0.964	0.987	0.970	0.973	0.758	0.760	0.788	0.793	0.993	0.996

The analysis of the presentation of SR images generated by various SISR solutions that employ convolutional networks (CNN) revealed large differences, in particular in the representation of the edges of anthropogenic objects (i.e., objects created by humans). An example of such an object was marked with a yellow frame (Table 5), and the magnified image is in the bottom-right corner of the illustration. Additional examples are provided in Appendices A and B. Many of the analyzed models, e.g., EDSR, FSRCNN, SRCNN, SRGAN, and ESRGAN, make errors during the representation of HR images. In images that were estimated with those methods, the edges of objects (even those with high contrast between them and the background) are blurred and their corners are rounded. This visual assessment is confirmed by the qualitative analysis, which was performed with the use of the metrics presented in Section 3.2. The MCWESRGAN architecture designed by the authors and expanded by adding Uformer layers enables the highest quality of enhancing the spatial resolution (SSIM = 0.98, PSNR = 38.32 dB). For the purposes of the conducted research, the authors expanded the most popular models responsible for SISR by adding layers of the Uformer model. The aim of the research was to provide a better assessment of the influence of the Uformer model on the quality of the estimated SR images. In most cases, the obtained results confirmed the beneficial effects of developing the SISR models by adding Uformer network layers. However, for models ESPCN, RDN, and SRDenseNet, a slight deterioration in the quality of SR images was observed (e.g., for the RDN network, the decrease in the PSNR was 0.14 dB).

Table 5. Examples of SR image fragments estimated by different convolutional network models.
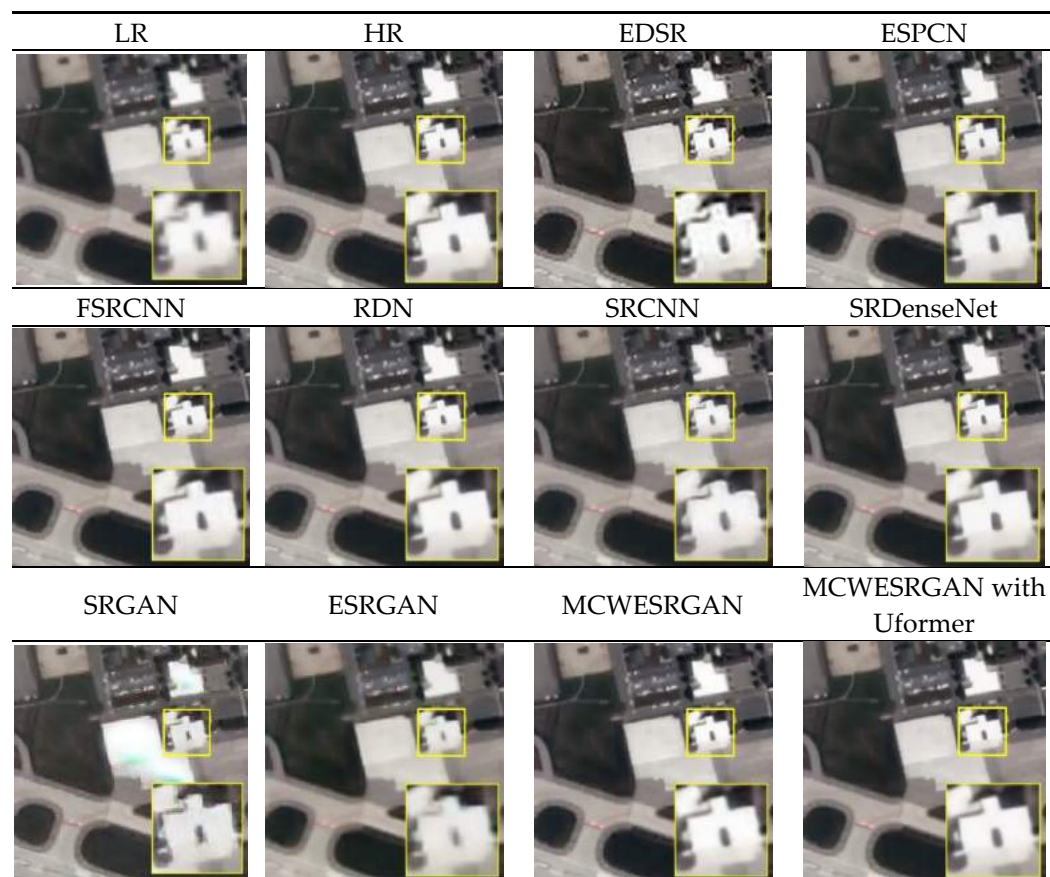
In order to improve the assessment of the SR images that are presented in Table 4, the SR images were transformed with the Fourier domain. Images estimated with MCWESRGAN, for which attempts were made to improve the quality of the estimated image by applying the Lucy–Richardson Algorithm and Wiener deconvolution, were added to the presentation. The aim of including these images was to compare them (where their quality is worse than that of SR images generated by MCWESRGAN) with images estimated with the use of other SISR models.

The transformation of the image with the Fast Fourier Transform (FFT) allows the movement from the spatial domain to the frequency domain while maintaining all attributes of the image. Decomposing an image into sines and cosines of varying amplitudes and phases enables one to distinguish mistakes in the image (Figure 10). Considering the results of the qualitative analyses that are presented in Tables 4 and 5, one may notice that the poorest quality of the estimated images is obtained if the quality of the SR images obtained with the use of the MCWESRGAN generator is "enhanced" using Wiener deconvolution or the Lucy–Richardson Algorithm. These conclusions are confirmed by the images presented in the frequency domain. These images contain low-frequency areas that are located at a large distance from the point of the coordinates (0,0) (the (0,0) point is situated in the center of the image and it corresponds to the intensity of the constant part of the original function). A similar phenomenon occurs in the SR images that were estimated by the SRGAN, FSRCNN, and EDSR models. An interesting case is the image in the frequency domain that was estimated by ESPCN. During the quality assessment, SR images estimated with the ESPCN model were characterized by high quality, but the analysis of sample SR images in the frequency domain revealed low frequencies that were located approximately 130 pixels from the centre of the image. This phenomenon may result from the fact that areas of the image are represented by pixels of a similar DN value and by the stepped representation of object edges (which is confirmed by the visual assessment of the estimated images).
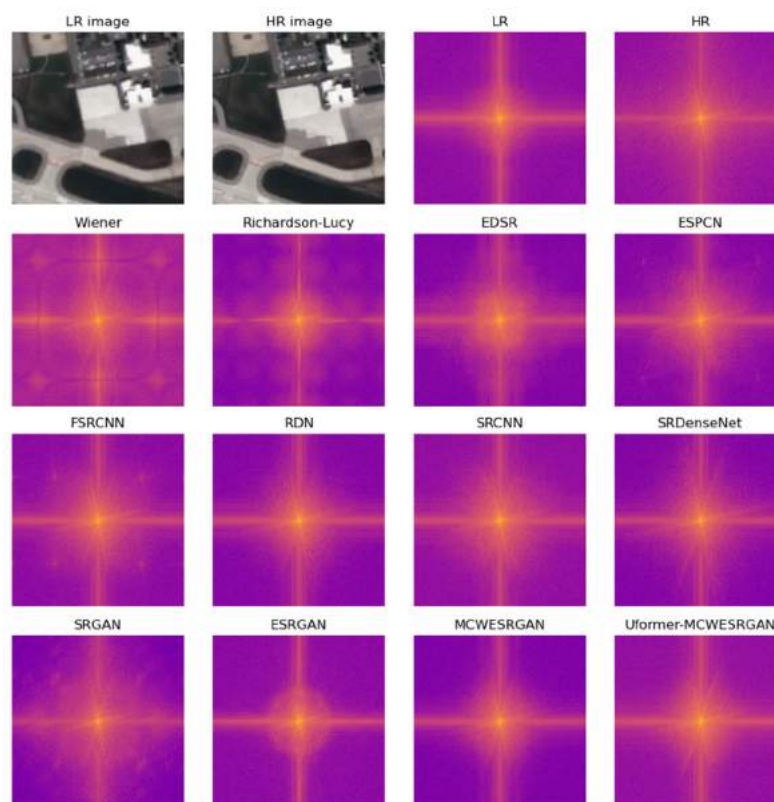
Figure 10. Images in the frequency domain.

The proposed method consists of several stages that occur consecutively. Thanks to its multiple-stage structure, the proposed solution does not require a large computational capacity. This feature allows the model to be operated on workstations that are not equipped with graphics processors (GPU). The computational capacity of the workstation determines only the duration of processing image sequences. Table 6 presents a comparison of the working times of the module that enhances spatial resolution of video data that consist of 100 frames of the dimensions of 120 × 120 pixels. The presentation demonstrates that the proposed algorithm may be implemented even on hardware that is not equipped with a graphics processor. However, the use of a graphics card significantly shortens the time of processing the sequence of images.

Table 6. Comparison of the time required to process a single video on various computational units.

STAGE	Time [s] (*—per Frame)			
	Intel Xeon Silver 4216 with TITAN RTX 24 GB	**Intel Xeon Silver 4216**	**Intel Corei5-1235**	**Intel Corei3-7100U**
Preparation of data	0.004 0.00004 *	0.04 0.0004 *	0.2 0.002 *	0.4 0.004 *
Resolution enhancement	76.80 (0.19 per tile)	988 (2.47 per tile)	4316 (10.79 per tile)	8861 (22.15 per tile)
Combining into video data	0.07 0.0007 *	0.9 0.009 *	4 0.04 *	8 0.08 *
Total time required to enhance the spatial resolution of one video	76.87 (~1 min 17 s) 0.77 s *	988.94 (~16 min 29 s) 9.9 s *	4320.2 (~72 min) 43.2 s *	8869.4 (~147 min 49 s) 88.7 s *

The obtained results confirm that it is reasonable to expand the MCWESRGAN model by adding layers of the Uformer network. This modification significantly improves the quality of SR images, and, in consequence, the interpretation capacity. However, the disadvantage of this solution is the prolonged duration of the work of the module that is responsible for enhancing the spatial resolution of the image in the sequence. It takes 0.03 s to enhance the spatial resolution of a single tile with the MCWESRGAN model (without the Uformer layers) (with the use of GPU), which means that the process is over in a time six times shorter than with the use of the MCWESRGAN model expanded by adding Uformer layers.

6. Conclusions

The authors of the present article have proposed a new methodology to enhance the spatial resolution of video data. This solution may, in particular, be applied to sequences of images acquired by nanosatellites. The obtained results demonstrate a fourfold increase in the spatial resolution of low-resolution images, while at the same time, the interpretation capacity has increased by more than two times. Moreover, the development of the generator that is responsible for improving the spatial resolution enabled the improvement of the structural similarity between HR and SR images by approx. 0.02, and the peek signal-to-noise ratio increased by over 2 dB (the average value of the SSIM parameter for the test data increased from 0.96 to 0.98, and the average PSNR value increased from 36.21 to 38.32 dB). The proposed methodology may be applied to issues related to the enhancement of the resolution of images and video sequences obtained by nanosatellites, which require the application of single-image super-resolution (SISR) methods. Apart from that, enhancing the spatial resolution with the use of the methodology described here does not require a large computational capacity (this is

required only at the stage of training the model), so it may be used on workstations that are not equipped with graphics processors.

The conducted research focused on proposing a new methodology to enhance the spatial resolution of images obtained from space altitudes. At the same time, it was noted that sequences of images captured by small satellites were characterized not only by a relatively low spatial resolution, but also by a low temporal resolution (approx. 10 frames per sec.) in comparison to aerial systems. This limitation was clearly noticeable during the observations of moving objects (such as an airplane during take-off). Due to that, further works on enhancing the quality of the sequences of images acquired by nanosatellites will focus on improving their temporal resolution.

Author Contributions: Conceptualization, K.K.; methodology, K.K. and D.W.; software, K.K.; validation, K.K.; formal analysis, K.K. and D.W.; investigation, K.K. and D.W.; resources, K.K.; data curation, K.K.; writing—original draft preparation, K.K.; writing—review and editing, D.W.; visualization, K.K.; supervision, D.W.; project administration, D.W.; funding acquisition, D.W. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data will be made available on request.
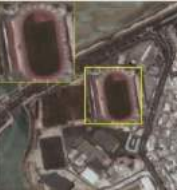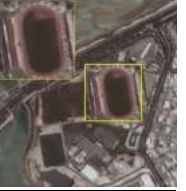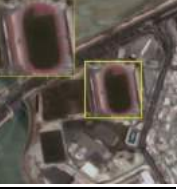
Appendix A

Table A1. Models extended with Uformer network layers are marked by adding "*" to the name.

Appendix B

Table A2. Quality assessment of convolutional network models dedicated to SISR for images with Appendix A. The best results are marked in green and the worst in red. "*" marks the SISR models with added Uformer network layers.

	Image 1																	
Metric	EDSR	EDSR *	ESPCN	ESPCN *	FSRCNN	FSRCNN *	RDN	RDN *	SRCNN	SRCNN *	SRDenseNet	SRDenseNet *	SRGAN	SRGAN *	ESRGAN	ESRGAN *	MCWESRGAN	MCWESRGAN *
SSIM	0.84	0.88	0.92	0.94	0.92	0.93	0.94	0.94	0.89	0.92	0.94	0.94	0.74	0.83	0.84	0.87	0.94	0.96
PSNR	26.36	27.2	31.26	27.16	30.88	32.37	32.99	32.79	29.0	30.65	32.88	32.64	22.56	24.57	26.96	28.15	33.02	34.08
SAM	0.09	0.08	0.05	0.08	0.05	0.04	0.04	0.04	0.06	0.05	0.04	0.04	0.13	0.11	0.08	0.07	0.04	0.03
SCC	0.34	0.43	0.42	0.25	0.32	0.41	0.44	0.43	0.32	0.43	0.40	0.42	0.11	0.28	0.26	0.38	0.45	0.46
UQI	0.99	0.99	1.0	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.99	0.99	1.00	1.00	1.00
	Image 2																	
SSIM	0.85	0.89	0.93	0.94	0.93	0.94	0.95	0.94	0.91	0.93	0.94	0.94	0.78	0.86	0.86	0.88	0.95	0.96
PSNR	25.26	25.77	29.78	30.01	29.93	31.08	31.34	31.41	27.85	29.15	30.91	31.12	22.8	24.57	26.02	27.36	31.68	32.46
SAM	0.10	0.09	0.06	0.06	0.06	0.05	0.05	0.05	0.07	0.06	0.05	0.05	0.13	0.10	0.09	0.07	0.05	0.04
SCC	0.25	0.31	0.30	0.31	0.22	0.29	0.30	0.30	0.23	0.29	0.27	0.29	0.09	0.21	0.19	0.25	0.31	0.32
UQI	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.99	0.99	0.99	1.00	1.00
	Image 3																	
SSIM	0.93	0.95	0.98	0.97	0.97	0.97	0.97	0.97	0.96	0.97	0.98	0.97	0.85	0.92	0.92	0.94	0.98	0.98
PSNR	32.26	33.9	38.35	37.57	35.93	37.2	37.01	37.01	34.34	36.37	38.35	37.36	23.86	25.4	30.88	31.79	37.67	37.41
SAM	0.06	0.05	0.03	0.03	0.04	0.04	0.04	0.04	0.05	0.04	0.03	0.03	0.15	0.12	0.07	0.07	0.03	0.03
SCC	0.17	0.22	0.20	0.20	0.13	0.19	0.18	0.18	0.14	0.21	0.17	0.19	0.05	0.16	0.11	0.16	0.19	0.20
UQI	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.97	0.99	0.99	1.00	1.00
	Image 4																	
SSIM	0.92	0.95	0.98	0.98	0.96	0.97	0.97	0.98	0.96	0.97	0.98	0.98	0.87	0.93	0.91	0.94	0.98	0.98
PSNR	31.68	33.13	38.58	39.62	35.79	38.58	38.02	39.46	34.31	37.77	39.41	39.44	26.54	28.98	32.0	34.52	39.29	40.17
SAM	0.06	0.05	0.02	0.02	0.03	0.02	0.03	0.02	0.04	0.03	0.02	0.02	0.1	0.07	0.05	0.04	0.02	0.02
SCC	0.22	0.31	0.28	0.28	0.18	0.26	0.28	0.29	0.22	0.29	0.26	0.28	0.09	0.23	0.14	0.24	0.30	0.30
UQI	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00
	Image 5																	
SSIM	0.89	0.92	0.95	0.95	0.95	0.96	0.96	0.96	0.93	0.95	0.96	0.96	0.82	0.89	0.88	0.90	0.96	0.98
PSNR	28.99	29.89	34.23	34.59	33.62	35.14	35.65	35.46	31.87	33.25	35.67	35.25	24.77	26.96	29.73	30.89	35.87	36.94
SAM	0.09	0.08	0.05	0.05	0.05	0.04	0.04	0.04	0.06	0.05	0.04	0.04	0.14	0.11	0.08	0.07	0.04	0.03
SCC	0.26	0.34	0.31	0.33	0.25	0.32	0.34	0.33	0.26	0.33	0.31	0.32	0.09	0.24	0.18	0.28	0.34	0.36
UQI	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00

References

1. Rujoiu-Mare, M.-R.; Mihai, B.-A. Mapping Land Cover Using Remote Sensing Data and GIS Techniques: A Case Study of Prahova Subcarpathians. *Procedia Environ. Sci.* **2016**, *32*, 244–255. https://doi.org/10.1016/j.proenv.2016.03.029.
2. Hejmanowska, B.; Kramarczyk, P.; Głowienka, E.; Mikrut, S. Reliable Crops Classification Using Limited Number of Sentinel-2 and Sentinel-1 Images. *Remote Sens.* **2021**, *13*, 3176. https://doi.org/10.3390/rs13163176.
3. Han, P.; Huang, J.; Li, R.; Wang, L.; Hu, Y.; Wang, J.; Huang, W. Monitoring Trends in Light Pollution in China Based on Nighttime Satellite Imagery. *Remote Sens.* **2014**, *6*, 5541–5558. https://doi.org/10.3390/rs6065541.
4. Hall, O.; Hay, G.J. A Multiscale Object-Specific Approach to Digital Change Detection. *Int. J. Appl. Earth Obs. Geoinf.* **2003**, *4*, 311–327. https://doi.org/10.1016/S0303-2434(03)00010-2.

5. Nasiri, V.; Hawryło, P.; Janiec, P.; Socha, J. Comparing Object-Based and Pixel-Based Machine Learning Models for Tree-Cutting Detection with PlanetScope Satellite Images: Exploring Model Generalization. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *125*, 103555. https://doi.org/10.1016/j.jag.2023.103555.

6. Li, W.; Zhou, J.; Li, X.; Cao, Y.; Jin, G. Few-Shot Object Detection on Aerial Imagery via Deep Metric Learning and Knowledge Inheritance. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *122*, 103s397. https://doi.org/10.1016/j.jag.2023.103397.

7. Stateczny, A.; Kazimierski, W.; Gronska-Sledz, D.; Motyl, W. The Empirical Application of Automotive 3D Radar Sensor for Target Detection for an Autonomous Surface Vehicle's Navigation. *Remote Sens.* **2019**, *11*, 1156. https://doi.org/10.3390/rs11101156.

8. de Moura, N.V.A.; de Carvalho, O.L.F.; Gomes, R.A.T.; Guimarães, R.F.; de Carvalho Júnior, O.A. Deep-Water Oil-Spill Monitoring and Recurrence Analysis in the Brazilian Territory Using Sentinel-1 Time Series and Deep Learning. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102695. https://doi.org/10.1016/j.jag.2022.102695.

9. Greidanus, H. Satellite Imaging for Maritime Surveillance of the European Seas. In *Remote Sensing of the European Seas*; Barale, V., Gade, M., Eds.; Springer: Dordrecht, The Netherlands, 2008; pp. 343–358, ISBN 978-1-4020-6772-3.

10. Gavankar, N.L.; Ghosh, S.K. Automatic Building Footprint Extraction from High-Resolution Satellite Image Using Mathematical Morphology. *Eur. J. Remote Sens.* **2018**, *51*, 182–193. https://doi.org/10.1080/22797254.2017.1416676.

11. Reda, K.; Kedzierski, M. Detection, Classification and Boundary Regularization of Buildings in Satellite Imagery Using Faster Edge Region Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 2240. https://doi.org/10.3390/rs12142240.

12. d'Angelo, P.; Máttyus, G.; Reinartz, P. Skybox Image and Video Product Evaluation. *Int. J. Image Data Fusion* **2016**, *7*, 3–18. https://doi.org/10.1080/19479832.2015.1109565.

13. Yang, T.; Wang, X.; Yao, B.; Li, J.; Zhang, Y.; He, Z.; Duan, W. Small Moving Vehicle Detection in a Satellite Video of an Urban Area. *Sensors* **2016**, *16*, 1528. https://doi.org/10.3390/s16091528.

14. Kawulok, M.; Tarasiewicz, T.; Nalepa, J.; Tyrna, D.; Kostrzewa, D. Deep Learning for Multiple-Image Super-Resolution of Sentinel-2 Data. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3885–3888.

15. Dong, H.; Supratak, A.; Mai, L.; Liu, F.; Oehmichen, A.; Yu, S.; Guo, Y. TensorLayer: A Versatile Library for Efficient Deep Learning Development. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA USA, 23–27 October 2017; pp. 1201–1204.

16. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv* **2014**, arXiv:1406.2661.

17. Kim, J.; Lee, J.K.; Lee, K.M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.

18. Fu, J.; Liu, Y.; Li, F. Single Frame Super Resolution with Convolutional Neural Network for Remote Sensing Imagery. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 8014–8017.

19. Lanaras, C.; Bioucas-Dias, J.; Galliani, S.; Baltsavias, E.; Schindler, K. Super-Resolution of Sentinel-2 Images: Learning a Globally Applicable Deep Neural Network. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 305–319. https://doi.org/10.1016/j.isprsjprs.2018.09.018.

20. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140.

21. Pouliot, D.; Latifovic, R.; Pasher, J.; Duffe, J. Landsat Super-Resolution Enhancement Using Convolution Neural Networks and Sentinel-2 for Training. *Remote Sens.* **2018**, *10*, 394. https://doi.org/10.3390/rs10030394.

22. Xiao, A.; Wang, Z.; Wang, L.; Ren, Y. Super-Resolution for "Jilin-1" Satellite Video Imagery via a Convolutional Network. *Sensors* **2018**, *18*, 1194. https://doi.org/10.3390/s18041194.

23. Yi Wang; Fevig, R.; Schultz, R.R. Super-Resolution Mosaicking of UAV Surveillance Video. In Proceedings of the 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 345–348.

24. Demirel, H.; Anbarjafari, G. IMAGE Resolution Enhancement by Using Discrete and Stationary Wavelet Decomposition. *IEEE Trans. Image Process.* **2011**, *20*, 1458–1460. https://doi.org/10.1109/TIP.2010.2087767.

25. Dong, W.; Zhang, L.; Shi, G.; Wu, X. Nonlocal Back-Projection for Adaptive Image Enlargement. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 349–352.

26. Dong, C.; Loy, C.C.; Tang, X. Accelerating the Super-Resolution Convolutional Neural Network. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 391–407.

27. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep Back-Projection Networks for Super-Resolution. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1664–1673.

28. Xiao, Y.; Yuan, Q.; He, J.; Zhang, Q.; Sun, J.; Su, X.; Wu, J.; Zhang, L. Space-Time Super-Resolution for Satellite Video: A Joint Framework Based on Multi-Scale Spatial-Temporal Transformer. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102731. https://doi.org/10.1016/j.jag.2022.102731.

29. Choi, Y.-J.; Kim, B.-G. HiRN: Hierarchical Recurrent Neural Network for Video Super-Resolution (VSR) Using Two-Stage Feature Evolution. *Appl. Soft Comput.* **2023**, *143*, 110422. https://doi.org/10.1016/j.asoc.2023.110422.

30. Wang, P.; Sertel, E. Multi-Frame Super-Resolution of Remote Sensing Images Using Attention-Based GAN Models. *Knowl. Based Syst.* **2023**, *266*, 110387. https://doi.org/10.1016/j.knosys.2023.110387.

31. Jin, X.; He, J.; Xiao, Y.; Yuan, Q. Learning a Local-Global Alignment Network for Satellite Video Super-Resolution. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. https://doi.org/10.1109/LGRS.2023.3250009.

32. Liu, H.; Gu, Y.; Wang, T.; Li, S. Satellite Video Super-Resolution Based on Adaptively Spatiotemporal Neighbors and Nonlocal Similarity Regularization. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8372–8383. https://doi.org/10.1109/TGRS.2020.2987400.

33. He, Z.; Li, X.; Qu, R. Video Satellite Imagery Super-Resolution via Model-Based Deep Neural Networks. *Remote Sens.* **2022**, *14*, 749. https://doi.org/10.3390/rs14030749.

34. Li, S.; Sun, X.; Gu, Y.; Lv, Y.; Zhao, M.; Zhou, Z.; Guo, W.; Sun, Y.; Wang, H.; Yang, J. Recent Advances in Intelligent Processing of Satellite Video: Challenges, Methods, and Applications. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 6776–6798. https://doi.org/10.1109/JSTARS.2023.3296451.

35. Yin, Q.; Hu, Q.; Liu, H.; Zhang, F.; Wang, Y.; Lin, Z.; An, W.; Guo, Y. Detecting and Tracking Small and Dense Moving Objects in Satellite Videos: A Benchmark. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. https://doi.org/10.1109/TGRS.2021.3130436.

36. Zhao, M.; Li, S.; Xuan, S.; Kou, L.; Gong, S.; Zhou, Z. SatSOT: A Benchmark Dataset for Satellite Video Single Object Tracking. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. https://doi.org/10.1109/TGRS.2022.3140809.

37. He, Q.; Sun, X.; Yan, Z.; Li, B.; Fu, K. Multi-Object Tracking in Satellite Videos With Graph-Based Multitask Modeling. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. https://doi.org/10.1109/TGRS.2022.3152250.

38. Xiao, Y.; Su, X.; Yuan, Q.; Liu, D.; Shen, H.; Zhang, L. Satellite Video Super-Resolution via Multiscale Deformable Convolution Alignment and Temporal Grouping Projection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–19. https://doi.org/10.1109/TGRS.2021.3107352.

39. Wang, X.; Sun, L.; Chehri, A.; Song, Y. A Review of GAN-Based Super-Resolution Reconstruction for Optical Remote Sensing Images. *Remote Sens.* **2023**, *15*, 5062. https://doi.org/10.3390/rs15205062.

40. Karwowska, K.; Wierzbicki, D. Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3292–3312. https://doi.org/10.1109/JSTARS.2022.3167646.

41. Wang, X.; Liu, Y.; Zhang, H.; Fang, L. A Total Variation Model Based on Edge Adaptive Guiding Function for Remote Sensing Image De-Noising. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 89–95. https://doi.org/10.1016/j.jag.2014.06.001.

42. Irum, I.; Sharif, M.; Raza, M.; Mohsin, S. A Nonlinear Hybrid Filter for Salt & Pepper Noise Removal from Color Images. *J. Appl. Res. Technology. JART* **2015**, *13*, 79–85. https://doi.org/10.1016/S1665-6423(15)30015-8.

43. Wang, B.; Ma, Y.; Zhang, J.; Zhang, H.; Zhu, H.; Leng, Z.; Zhang, X.; Cui, A. A Noise Removal Algorithm Based on Adaptive Elevation Difference Thresholding for ICESat-2 Photon-Counting Data. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *117*, 103207. https://doi.org/10.1016/j.jag.2023.103207.

44. Yang, H.; Su, X.; Wu, J.; Chen, S. Non-Blind Image Blur Removal Method Based on a Bayesian Hierarchical Model with Hyperparameter Priors. *Optik* **2020**, *204*, 164178. https://doi.org/10.1016/j.ijleo.2020.164178.

45. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. Uformer: A General U-Shaped Transformer for Image Restoration. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 17662–17672.

46. Shen, Z.; Wang, W.; Lu, X.; Shen, J.; Ling, H.; Xu, T.; Shao, L. Human-Aware Motion Deblurring. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 5571–5580.

47. Rim, J.; Lee, H.; Won, J.; Cho, S. Real-World Blur Dataset for Learning and Benchmarking Deblurring Algorithms. In *Computer Vision—ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 184–201.

48. Nah, S.; Kim, T.H.; Lee, K.M. Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 257–265.

49. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8183–8192.

50. Xu, L.; Zheng, S.; Jia, J. Unnatural L0 Sparse Representation for Natural Image Deblurring. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1107–1114.

51. Kupyn, O.; Martyniuk, T.; Wu, J.; Wang, Z. DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27–28 October 2019; pp. 8877–8886.

52. Zhang, K.; Luo, W.; Zhong, Y.; Ma, L.; Stenger, B.; Liu, W.; Li, H. Deblurring by Realistic Blurring. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 2734–2743.

53. Purohit, K.; Suin, M.; Rajagopalan, A.N.; Boddeti, V.N. Spatially-Adaptive Image Restoration Using Distortion-Guided Networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2309–2319.

54. Zhang, J.; Pan, J.; Ren, J.; Song, Y.; Bao, L.; Lau, R.W.H.; Yang, M.-H. Dynamic Scene Deblurring Using Spatially Variant Recurrent Neural Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2521–2529.

55. Tao, X.; Gao, H.; Shen, X.; Wang, J.; Jia, J. Scale-Recurrent Network for Deep Image Deblurring. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8174–8182.

56. Zhang, H.; Dai, Y.; Li, H.; Koniusz, P. Deep Stacked Hierarchical Multi-Patch Network for Image Deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5978–5986.

57. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H.; Shao, L. Multi-Stage Progressive Image Restoration. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 14816–14826.

58. Karwowska, K.; Wierzbicki, D. MCWESRGAN: Improving Enhanced Super-Resolution Generative Adversarial Network for Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 9459–9479. https://doi.org/10.1109/JSTARS.2023.3322642.

59. Richardson, W.H. Bayesian-Based Iterative Method of Image Restoration. *J. Opt. Soc. Am.* **1972**, *62*, 55–59. https://doi.org/10.1364/JOSA.62.000055.

60. Orieux, F.; Giovannelli, J.-F.; Rodet, T. Bayesian Estimation of Regularization and PSF Parameters for Wiener-Hunt Deconvolution. *J. Opt. Soc. Am. A* **2010**, *27*, 1593–1607. https://doi.org/10.1364/JOSAA.27.001593.

61. Karwowska, K.; Wierzbicki, D. Improving Spatial Resolution of Satellite Imagery Using Generative Adversarial Networks and Window Functions. *Remote Sens.* **2022**, *14*, 6285. https://doi.org/10.3390/rs14246285.

62. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In *Computer Vision—ECCV 2018 Workshops*; Leal-Taixé, L., Roth, S., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 63–79.

63. Kantorovich, L.V.; Rubinshtein, S.G. On a Space of Totally Additive Functions. *Vestn. St. Petersburg Univ. Math.* **1958**, *13*, 52–59.

64. Wang, D.; Liu, Q. Learning to Draw Samples: With Application to Amortized MLE for Generative Adversarial Learning. *arXiv* **2016**, arXiv:1611.01722.

65. Charbonnier, P.; Blanc-Feraud, L.; Aubert, G.; Barlaud, M. Two Deterministic Half-Quadratic Regularization Algorithms for Computed Imaging. In Proceedings of the 1st International Conference on Image Processing, Austin, TX, USA, 13–16 November 1994; Volume 2, pp. 168–172.

66. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H.; Shao, L. Learning Enriched Features for Real Image Restoration and Enhancement. *arXiv* **2020**, arXiv:1611.01722.

67. Keelan, B. *Handbook of Image Quality: Characterization and Prediction*; CRC Press: Boca Raton, FL, USA, 2002; ISBN 978-0-429-22280-1.

68. Wang, Z.; Bovik, A.C. A Universal Image Quality Index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. https://doi.org/10.1109/97.995823.

69. Goetz, A.; Boardman, W.; Yunas, R. Discrimination among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm. JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop; 1992. Available online: https://ntrs.nasa.gov/search.jsp?R=19940012238 (accessed on 24 November 2023).

70. Xia, G.-S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3974–3983.

71. Chang Guang Satellite Technology Co., Ltd. Available online: http://www.jl1.cn/EWeb/ (accessed on 24 November 2023).

72. Planet|Homepage. Available online: https://www.planet.com/ (accessed on 24 November 2023).

mgr inż. Kinga Karwowska

Katedra Rozpoznania Obrazowego

Wydział Inżynierii Lądowej i Geodezji

Wojskowa Akademia Techniczna im. J. Dąbrowskiego

Warszawa, Polska

Oświadczenie

Oświadczam, że w pracy badawczej, której wynikiem była publikacja:

Karwowska K., Wierzbicki D., 2022, Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, *vol. 15, pp. 3292-3312, DOI: 10.1109/JSTARS.2022.3167646*

mój wkład polegał na:

- współudziale w opracowaniu koncepcji przeglądu obszaru badawczego,
- współudziale w formułowaniu celów i założeń pracy badawczej,
- współudziale w zaprojektowaniu badań,
- przygotowaniu przeglądu literatury,
- przygotowaniu baz danych, które zostały wykorzystane do testów modeli sieci SRGAN i ESRGAN,
- przygotowaniu oprogramowania,
- przeprowadzeniu prac badawczych,
- przeprowadzeniu testów modeli sieci SRGAN i ESRGAN,
- wykonaniu dyskusji wyników i sformułowaniu wniosków końcowych,
- współudziale w przygotowaniu manuskryptu,
- współudziale w wykonaniu korekt do artykułu i sformułowaniu odpowiedzi do recenzentów publikacji.

Swój wkład oceniam na **80%** całości artykułu.

....

podpis

dr hab. inż. Damian Wierzbicki, prof. WAT

Katedra Rozpoznania Obrazowego

Wydział Inżynierii Lądowej i Geodezji

Wojskowa Akademia Techniczna im. J. Dąbrowskiego

Warszawa, Polska

Oświadczenie

Oświadczam, że w pracy badawczej, której wynikiem była publikacja:

Karwowska K., Wierzbicki D., 2022, Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, *vol. 15, pp. 3292-3312, DOI: 10.1109/JSTARS.2022.3167646*

mój wkład polegał na:

- współudziale w opracowaniu koncepcji przeglądu obszaru badawczego,
- współudziale w formułowaniu celów i założeń pracy badawczej,
- współudziale w zaprojektowaniu badań,
- współudziale w przygotowaniu manuskryptu,
- współudziale w wykonaniu korekt do artykułu i sformułowaniu odpowiedzi do recenzentów publikacji,
- nadzorze merytorycznym nad prowadzonymi pracami badawczymi.

Swój wkład oceniam na **20%** całości artykułu.

....
podpis

mgr inż. Kinga Karwowska

Katedra Rozpoznania Obrazowego

Wydział Inżynierii Lądowej i Geodezji

Wojskowa Akademia Techniczna im. J. Dąbrowskiego

Warszawa, Polska

Oświadczenie

Oświadczam, że w pracy badawczej, której wynikiem była publikacja:

Karwowska K., Wierzbicki D., 2022, Improving Spatial Resolution of Satellite Imagery Using Generative Adversarial Networks and Window Functions, Remote Sensing, 14(24):6285, DOI: 10.3390/rs14246285

mój wkład polegał na:

– współudziale w opracowaniu koncepcji przetwarzania całych zobrazowań satelitarnych w zagadnieniach automatycznej interpretacji zobrazowań,

– współudziale w formułowaniu celów i założeń pracy badawczej,

– zaprojektowaniu badań,

– przygotowaniu przeglądu literatury,

– przygotowaniu oprogramowania,

– przeprowadzeniu prac badawczych,

– przeprowadzeniu testów rozwiązania,

– opracowaniu graficznym ilustracji i wykresów,

– wykonaniu dyskusji wyników i sformułowaniu wniosków końcowych,

– współudziale w przygotowaniu manuskryptu,

– współudziale w wykonaniu korekt do artykułu i sformułowaniu odpowiedzi do recenzentów publikacji.

Swój wkład oceniam na **80%** całości artykułu.

....

podpis

Warszawa, 24.06.2024 r.

dr hab. inż. Damian Wierzbicki, prof. WAT

Katedra Rozpoznania Obrazowego

Wydział Inżynierii Lądowej i Geodezji

Wojskowa Akademia Techniczna im. J. Dąbrowskiego

Warszawa, Polska

Oświadczenie

Oświadczam, że w pracy badawczej, której wynikiem była publikacja:

Karwowska K., Wierzbicki D., 2022, Improving Spatial Resolution of Satellite Imagery Using Generative Adversarial Networks and Window Functions, Remote Sensing, 14(24):6285, DOI: 10.3390/rs14246285

mój wkład polegał na:

– współudziale w opracowaniu koncepcji badań,

– współudziale w formułowaniu celów i założeń pracy badawczej,

– współudziale w przygotowaniu manuskryptu,

– współudziale w wykonaniu korekt do artykułu i sformułowaniu odpowiedzi do recenzentów publikacji,

– nadzorze merytorycznym nad prowadzonymi pracami badawczymi.

Swój wkład oceniam na **20%** całości artykułu.

....
podpis

Warszawa, 24.06.2024 r.

mgr inż. Kinga Karwowska

Katedra Rozpoznania Obrazowego

Wydział Inżynierii Lądowej i Geodezji

Wojskowa Akademia Techniczna im. J. Dąbrowskiego

Warszawa, Polska

Oświadczenie

Oświadczam, że w pracy badawczej, której wynikiem była publikacja:

Karwowska K., Wierzbicki D., 2023, MCWESRGAN: Improving Enhanced Super-Resolution Generative Adversarial Network for Satellite Images, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 9459-9479, DOI: 10.1109/JSTARS.2023.3322642

mój wkład polegał na:

- współudziale w opracowaniu koncepcji,
- współudziale w formułowaniu celów i założeń pracy badawczej,
- zaprojektowaniu badań,
- przygotowaniu przeglądu literatury,
- przygotowaniu baz danych, na podstawie których opracowany i testowany był opracowany model sieci GAN,
- przygotowaniu oprogramowania,
- przeprowadzeniu prac badawczych,
- przeprowadzeniu testów rozwiązania,
- opracowaniu graficznym ilustracji i wykresów,
- wykonaniu dyskusji wyników i sformułowaniu wniosków końcowych,
- współudziale w przygotowaniu manuskryptu,
- współudziale w wykonaniu korekt do artykułu i sformułowaniu odpowiedzi do recenzentów publikacji.

Swój wkład oceniam na **80%** całości artykułu.

....
podpis

Warszawa, 24.06.2024 r.

dr hab. inż. Damian Wierzbicki, prof. WAT

Katedra Rozpoznania Obrazowego

Wydział Inżynierii Lądowej i Geodezji

Wojskowa Akademia Techniczna im. J. Dąbrowskiego

Warszawa, Polska

Oświadczenie

Oświadczam, że w pracy badawczej, której wynikiem była publikacja:

Karwowska K., Wierzbicki D., 2023, MCWESRGAN: Improving Enhanced Super-Resolution Generative Adversarial Network for Satellite Images, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 9459-9479, DOI: 10.1109/JSTARS.2023.3322642

mój wkład polegał na:

- współudziale w opracowaniu koncepcji,
- współudziale w formułowaniu celów i założeń pracy badawczej,
- współudziale w przygotowaniu manuskryptu,
- współudziale w wykonaniu korekt do artykułu i sformułowaniu odpowiedzi do recenzentów publikacji,
- nadzorze merytorycznym nad prowadzonymi pracami badawczymi.

Swój wkład oceniam na **20%** całości artykułu.

podpis

mgr inż. Kinga Karwowska

Katedra Rozpoznania Obrazowego

Wydział Inżynierii Lądowej i Geodezji

Wojskowa Akademia Techniczna im. J. Dąbrowskiego

Warszawa, Polska

Warszawa, 24.06.2024 r.

Oświadczenie

Oświadczam, że w pracy badawczej, której wynikiem była publikacja:

Karwowska K., Wierzbicki D., 2024, Modified ESRGAN with Uformer for Video Satellite Imagery Super-Resolution, Remote Sensing, 16(11):1926. doi: 10.3390/rs16111926

mój wkład polegał na:

- współudziale w opracowaniu koncepcji, formułowaniu celów i założeń pracy badawczej,
- zaprojektowaniu badań wykorzystania sieci MCWESRGAN na potrzeby poprawy rozdzielczości sekwencji obrazów pozyskanych przez małe satelity,
- przygotowaniu przeglądu literatury,
- przygotowaniu nowej bazy danych, która została wykorzystana do przygotowania modelu sieci MCWESRGAN rozbudowanego o model Uformer,
- przygotowaniu oprogramowania,
- przeprowadzeniu prac badawczych,
- przeprowadzeniu testów rozwiązania,
- opracowaniu graficznym ilustracji i wykresów,
- wykonaniu dyskusji wyników i sformułowaniu wniosków końcowych,
- współudziale w przygotowaniu manuskryptu,
- współudziale w wykonaniu korekt do artykułu i sformułowaniu odpowiedzi do recenzentów publikacji.

Swój wkład oceniam na **80%** całości artykułu.

....

podpis

dr hab. inż. Damian Wierzbicki, prof. WAT

Katedra Rozpoznania Obrazowego

Wydział Inżynierii Lądowej i Geodezji

Wojskowa Akademia Techniczna im. J. Dąbrowskiego

Warszawa, Polska

Oświadczenie

Oświadczam, że w pracy badawczej, której wynikiem była publikacja:

Karwowska K., Wierzbicki D., 2024, Modified ESRGAN with Uformer for Video Satellite Imagery Super-Resolution, Remote Sensing, 16(11):1926. doi: 10.3390/rs16111926

mój wkład polegał na:

- współudziale w opracowaniu koncepcji, formułowaniu celów i założeń pracy badawczej,
- współudziale w przygotowaniu manuskryptu,
- współudziale w wykonaniu korekt do artykułu i sformułowaniu odpowiedzi do recenzentów publikacji,
- nadzorze merytorycznym nad prowadzonymi pracami badawczymi.

Swój wkład oceniam na **20%** całości artykułu.

podpis