doi:10.20944/preprints202202.0054.v2

Article

Generative Adversarial Approach to Urban Areas NDVI Estimation Using Panchromatic Orthoimagery: A Case Study of Łódź, Poland

Maciej Adamiak [1], Krzysztof Będkowski [2] and Adam Bielecki [2], *

[1] ReasonField Lab, ul. Na Uboczu 8/87, 02-791, Warszawa, Polska; maciej.adamiak@reasonfieldlab.com(M.A.)
[2] University of Lodz, Faculty of Geographical Sciences, Institute of Urban Geography, Tourism Studies and Geoinformation, ul. Kopcińskiego 31, 90-139 Łódź, Poland; krzysztof.bedkowski@geo.uni.lodz.pl(K.B.)
* Correspondence: adam.bielecki@edu.uni.lodz.pl

Abstract: Generative adversarial networks (GAN) opened new possibilities for image processing and analysis. Inpainting, dataset augmentation using artificial samples or increasing spatial resolution of aerial imagery are only a few notable examples of utilizing GANs in remote sensing. This is due to a unique construction and training process expressed as a duel between GAN components. The main objective of the research is to apply GAN to generate an artificial Normalized Difference Vegetation Index (NDVI) using panchromatic images. The NDVI ground-truth labels were prepared by combining RGB and NIR orthophoto. The dataset was then utilized as input for a conditional generative adversarial network (cGAN) to perform an image-to-image translation. The main goal of the neural network was to generate an artificial NDVI image for each processed 256px × 256px patch using only information available in the panchromatic input. The network achieved 0.7569 ± 0.1083 Structural Similarity Index Measure (SSIM), 26.6459 ± 3.6577 Peak Signal-to-Noise Ratio (PNSR) and 0.0504 ± 0.0193 Root-Mean-Square Error (RSME) on the test set. The perceptual evaluation was performed to verify the usability of the method when working with a real-life scenario. The research confirms that the structure and texture of the panchromatic aerial remote sensing image contains sufficient information for NDVI estimation for various objects of urban space. Even though these results can be used to highlight areas rich in vegetation and distinguish them from urban background, there is still room for improvement in terms of accuracy of estimated values. The purpose of the research is to explore the possibility of utilizing GAN to enhance panchromatic images (PAN) with information related to vegetation. This opens interesting possibilities in terms of historical remote sensing imagery processing and analysis. The panchromatic orthoimagery dataset was derived from RGB orthoimagery.

Keywords: generative adversarial networks; NDVI; green areas; orthophoto; artificial datasets

1. Introduction

Normalized Difference Vegetation Index (NDVI) is a metric that stores information regarding the degree of vegetation. Due to its characteristics, it is used in many research activities and remote sensing measurements [1,2]. The main feature of the NDVI is that it directly depends on the quantity and quality of the vegetation in the selected area [3]. NDVI is widely recognized as a metric of vegetation vitality. It has a wide range of applications from vegetation cover segmentation [4], biomass estimation [5], crop maturity classification [6,7], water stress detection [8–10], nitrogen content indication [11], chlorophyll content estimation, yield estimation [12,13], to detection of disease and the effects of pest infestation [14] and more. NDVI can help in determining the size and changes of vegetation resources, mainly woody, and further in estimating changes of ecosystem services provided by trees [15–17], including the increasingly important effect of cooling the city through shading [18,19].

doi:10.20944/preprints202202.0054.v2

The NDVI can be used in analyzing both aerial and satellite imagery in different scales, from single plants to global vegetation zones. The index is calculated by taking the spectral reflectance of near-infrared and red spectral bands of the image and applying the following formula:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{1}$$

where NIR and RED are the spectral reflectance of near-infrared and red bands respectively. NDVI is a number between -1 and 1, where higher values indicate denser vegetation canopy. To distinguish areas covered by vegetation from other areas, a threshold value is needed. There is no universal approach for setting the threshold. Therefore, it must be carefully adapted to the type of imaging system, natural conditions, type of vegetation and seasonality. Table 1 presents thresholds utilized in recent research projects related to urban areas conducted in Poland.

Table 1. NDVI threshold values used in urban studies in Poland.

NDVI Threshold for vegetation	Image data used	Research area	References
0.3	Landsat TM, GSD 30 m, 3 Jul. 2006	Warsaw	Tomaszewska et al. 2011 [4]
0.1	MODIS, GSD 250 m, 3 Jul. 2006	Warsaw	Tomaszewska et al. 2011 [4]
0.1	Digital orthophoto, GSD 0.1 m, May 2014	Wroclaw	Kubalska & Preuss 2014 [20]
0.2	IKONOS 2, GSD 1(4) m, 18 Aug 2005	Lublin	Krukowski et al. 2016 [21]
0.2	Landsat 8, GSD 30 m, 3 Jul. 2015	Łódź	Będkowski & Bielecki 2017 [22]
0.1	Pléiades 1A, GSD 0.5 m, May 2012	Warsaw	Pyra & Adamczyk 2018 [23]
0.1	CIR-orthophoto, GSD 0.25 m, 2015	Łódź	Pluto-Kossakowska et al. 2018 [24]
0.2	IKONOS 2, GSD 1(4) m, 18 Aug 2011	Lublin	Krukowski 2018 [25]
0.1	CIR aerial orthophoto, GSD 0.25 m, 2015	Łódź	Worm et al. 2019 [26]
0.6	Sentinel 2, GSD 10 m, summer 2018, 2019	Poland	Łachowski & Łęczek 2020 [27]

In remote sensing, the multitude of uses of aerial or satellite imagery in classification tasks is directly related to image textural features, originally introduced by Haralick et al. [28]. The textural features concept is described as the spatial relationships between pixels of a given spectral band or several bands that enables discrimination between objects represented on the analyzed image [29]. Texture parameters form the basis of the Object-Based Image Analysis (OBIA). As practice has shown, utilizing textural features in different forms helps in accomplishing various tasks in the field of photogrammetry and remote sensing. When combined with information acquired by calculating NDVI, textural features can serve as input for methods capable of detecting and segmenting vegetation in cities [23,29–33].

It is not always possible to compute NDVI due to the lack of the NIR band. Although it is uncommon for contemporary imagery not to contain NIR there are multiple examples of aerial imagery originating from archive data sources that store only RGB or panchromatic images. Obtaining information regarding vegetation, in this case, is difficult and associated with the need to estimate the NIR band.

Fortunately, there are successful attempts to create artificial spectral bands based solely on the available bands. Recent research activities focus on utilizing generative neural network models by treating the estimation of a spectral band as an image-to-image translation task. This supervised machine learning task is to discover a transformation between real input spectral bands and real ground truth output. Once learned, the transformation can be used to process real input and produce an artificial but realistic output. The output is being produced conditionally, i.e., it is influenced by the input composition regarding the learned transformation. A good candidate among multiple available neural network architectures to tackle the image-to-image translation task is the Conditional Generative Adversarial Network (cGAN). The use of cGAN for spectral bands estimation is justified in cases in which the input and output bands have little or no overlap. Therefore, acquiring the transformation is non-trivial. This is also a significant advantage of cGAN over the direct use of a CNN-based regressor. Without the support of the discriminator mechanism offered by GAN networks, the CNNs may have problems obtaining the desired result without crafting a fine-tailored loss function. In other words, cGAN not only learns the transformation from input image to output image, but also learns a loss function to train this mapping.

Contemporary research provides the basis for obtaining the answer regarding the possibility of NDVI estimation from RGB and panchromatic images. For example, Suárez et al. [34] proposed the usage of a cGAN approach in NDVI estimation. The proposed method allows the creation of an NDVI image based only on a single spectral band, which is near-infrared, NIR. Three network architectures were tested with gradually increasing complexity. The generative model was trained from a near-infrared image plus Gaussian noise. The network was trained on a large number (280000) of pairs of patches (64 × 64 pixels) of downloaded NIR images and corresponding NDVI images. In the validation phase, the result was compared with the proper NDVI calculated using the classical formula from the RGB and NIR bands. The research showed a high agreement of both NDVI ground truth and NDVI synthesized. In an article, Suárez et al. [35] proposed GAN for NDVI estimation based on grayscale images made from 3-bands RGB images. What is important, not registered RGB/NIR images were used in the training process. The procedure led to an efficient estimation of the synthetic NIR band, which was then used to calculate the NDVI according to the classical formula (1). As stated, the proposed model effectively translates the images from the visible spectrum to NIR and NDVI. Aslahishahri et al. [36] used the conditional GAN (cGAN) and confirmed the possibility of generating the NIR band, which is useful for examining the condition of crops (canola, lentil, dry bean, and wheat) from RGB images. The training sets included 256 × 256 pixels patches of RGB images and their corresponding NIR images acquired using a UAV (unmanned aerial vehicle). The trained network was able to synthesize the NIR band from which the NDVI was then derived.

This research aims to check the feasibility of using the generative adversarial approach to direct NDVI estimation using information exclusively from structural and textural analysis of panchromatic orthoimagery by skipping the intermediate NIR generation step. Authors assume that NDVI is strongly associated with certain structures and textures of a single-band panchromatic image and can therefore be determined with precision appropriate for practical purposes. The novelty of this approach when applied to aerial and satellite imagery is that it allows enhancing datasets that did not contain information regarding vegetation with a surrogate channel. This opens the possibility to reuse previously collected data in new research scenarios.

2. Materials and Methods

2.1. Research area

Łódź is a city located in central Poland, on the Łódź Upland, at an altitude of 162 to 278 m above sea level (Figure 1). It is the third-largest city in the country in terms of population, with approximately 668000 inhabitants.

In 1823, the city entered the phase of dynamic population and territorial development, which was related to the dynamic growth of the textile industry. In later years, large enterprises were established, employing huge numbers of workers, and the production of woolen, linen and cotton products was started. At that time, a characteristic strip-like spatial form of the urban layout of Łódź was formed [37]. The years 1860-1890 were the period of the most intensive industrial development. The city became the largest center of the textile industry in the Kingdom of Poland. Industrial buildings were mixed with residential ones. Colonies of houses for workers intertwined with factories and with factory owners' residences, which affects the image of the city to this day.

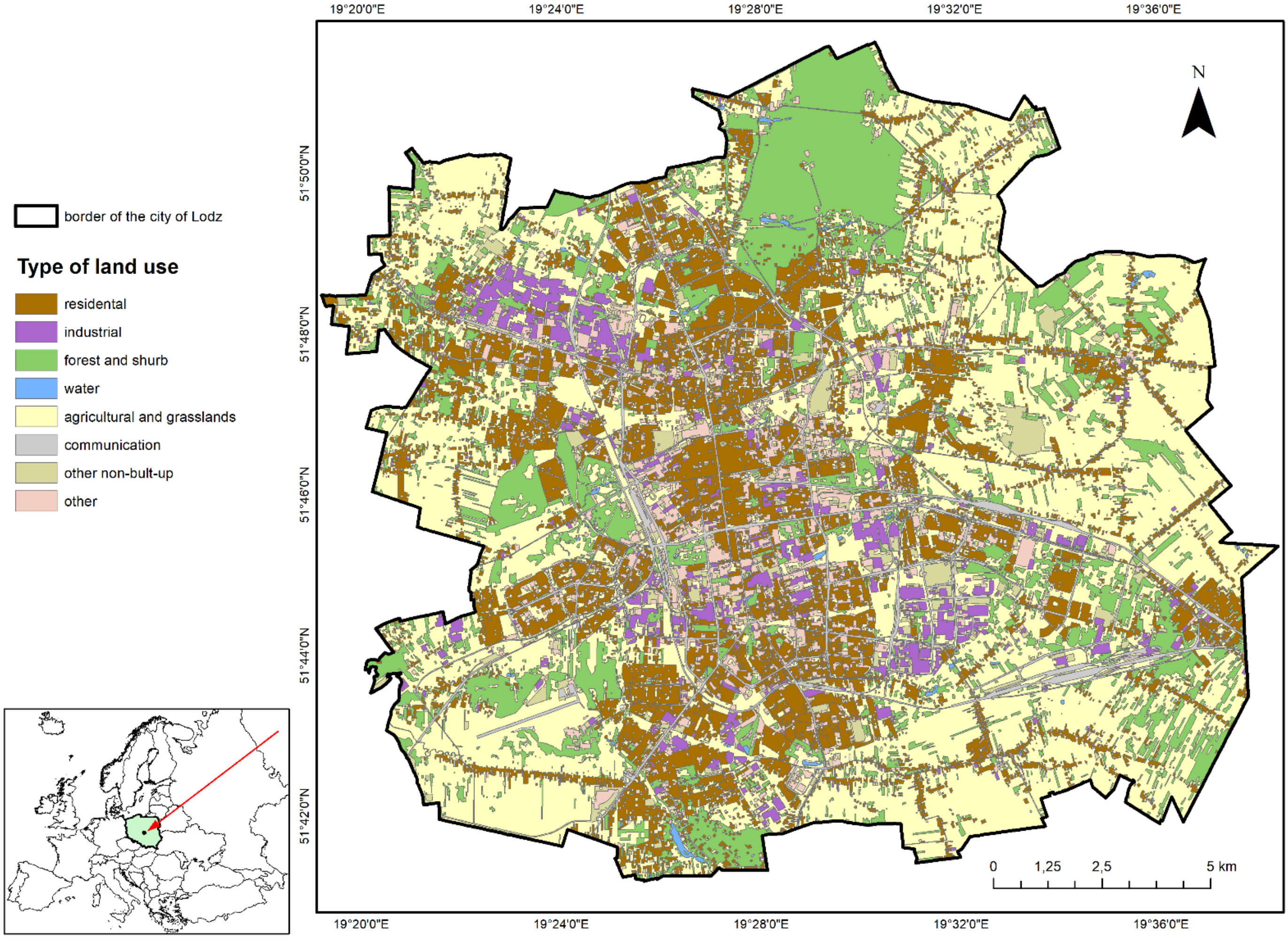

Figure 1. Research area, source: own elaboration based on Head Office of Geodesy and Cartography [38]

The city is situated between watersheds of the Vistula and Oder rivers. Although 18 rivers and streams flow within the city, none of them significantly affect the city’s physiognomy. There are no lakes or other larger bodies of water. Urban green areas - parks, lawns, estate green belts, a zoological garden, a botanical garden, nature reserves, landscape parks - cover 16.62 km^2 and constitute 5.67% of the city's area. Within the borders of Łódź there is the largest area of urban forests in the country - 8.9% of the city area [39].

The remote-sensing image of the city shows compact housing complexes, mainly from the years 1870-1914, constituting the largest complex of Art Nouveau architecture in the country. Workers' housing estates, villa complexes, large housing estates with high buildings in the form of blocks, industrial and commercial facilities, communication areas - roads and railroads mix to form the image of modern Łódź (Table 2). The most important socio-economic changes for the present appearance of the city after 1989 resulted in the collapse of many textiles industry plants, impoverishment of the population and the degradation of a large part of the city's housing, industrial and communication infrastructure. Currently, the revitalization process is underway, new housing estates and industrial buildings are appearing. The former 19th century industrial complexes receive new functions: residential, office and commercial. Nevertheless, even in areas close to the city center, there are still many degraded areas subject to spontaneous vegetation succession i.e., land abandonment.

Table 2. Structure of land use in Łódź [km^2]; source: Statistics Poland [39].

Total	Agricultural land	Forest, woody, and bushy land	Residential areas	Industrial areas	Transport areas	Groundwater	Other
293.25	113.76	24.67	47.13	13.91	42.37	1.33	1.15

2.1. Overview

The process of the investigation began by collecting the respective sheets of orthophotomap in 4 bands (Red, Green, Blue and Close Infrared) which covered the whole area of Łódź. Then the datasets from orthoimagery were prepared for the training and validation process. The next step was to compose a dataset of panchromatic orthoimagery, which was an input for the Pix2Pix machine learning model. This model has been tested by comparing the artificial and real NDVI values. The last step was to perform the perceptual evaluation. The graphic overview is shown in Figure 2.

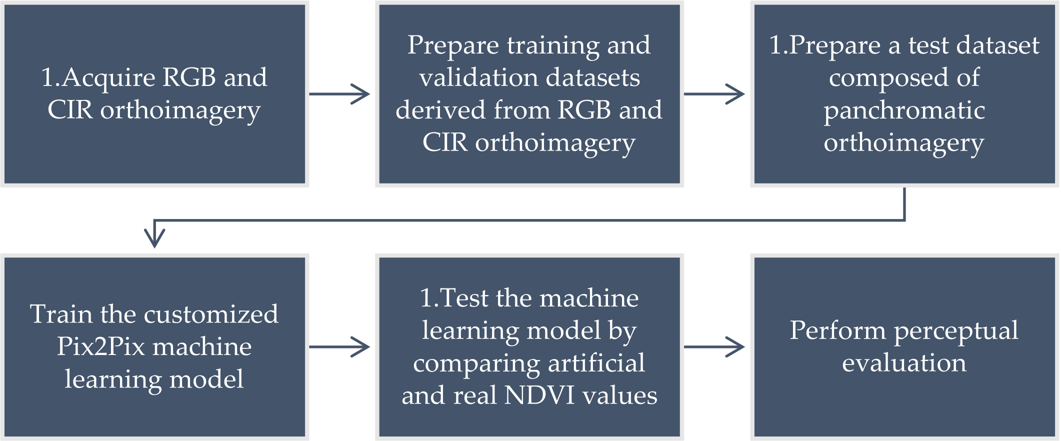

Figure 2. The main steps of the investigation.

2.3. Dataset

All data has been downloaded from Geoportal [40]. The dataset included the orthophotomaps in two different compositions: RGB (bands: RED, GREEN and BLUE) and CIR (NEAR INFRARED, RED and GREEN). The data download process has been carried out by using a plug-in for the Q-GIS software, GUGiK’s data downloader [41]. With this tool, it was possible to download all sheets, at least part of which was within the administrative boundaries of the city of Łódź. As a result, 261 sheets in TIFF format have been downloaded with both RGB and CIR compositions. The data has a Ground Sampling Distance (pixel size) of 5 cm and in total, they took up 140 GB of disk space.

To compute a 4-bands orthophotomap (RGB+NIR) ESRI ArcGIS Desktop 10.5.1 Composite Bands was used. The data processing pipeline was implemented in Python. Its main task was to merge RGB image bands and the first band from CIR.

Collected orthophoto imagery was post-processed to form neural network training and validation datasets. Firstly, each RGB+NIR sheet acquired from the initial dataset was split into multiple non-overlapping rectangular patches with different edge sizes. This procedure yielded 642582 patches in total: 493290 with 512 px, 120582 with 1024 px and 28710 with 2048 px edge lengths. Secondly, each patch was divided into two separate items.

The first item designed to be utilized as input during neural network training contained grayscale images derived as a simple sum of RGB bands stretched linearly into 256 grey levels by using the following formula:

$$PAN = \frac{((R + G + B) - min)}{max - min} \tag{2}$$

where R, G, B are respective spectral bands of orthophoto, max, min is the maximum and minimum pixel sum value R+G+B.

At 0.05 m GSD, the obtained images allow a good insight into the city space: the details of the construction of roofs, the presence of chimneys, ventilators, characteristic tile arrangements, roads and parking lots, fences, passenger cars and trucks are visible as well. There are also visible details of the structure of vegetation - individual trees, their groups or rows - treetops and in the case of leafless trees - the arrangement of branches and sharp shadows cast by them on the ground. The effects of plowing are visible in the cultivated fields not covered with vegetation. The fields covered with vegetation also often show streaky texture and traces of agricultural machinery passing resulting from the specificity of the cultivation itself and plant care treatments. Due to the small size of pixels, it can be assumed that there is no problem with the spectral mixture, which should be considered in the case of images with low resolution [42] and each of the pixels represents reflectance values typical of urban land cover forms, or actually - materials present in this space.

Network targets included the NDVI value calculated using the RED and NIR bands. All dataset items were scaled between 0 and 1 and resized to fit the network requirements forming a [256, 256, 1] input and a [256, 256, 1] target tensors. After the post-processing step has been completed 5 patches representing different spatial features were randomly picked from each processed orthophoto sheet to form a 1305 item validation subset. What is important is that these items did not overlap with the training dataset in terms of areas represented in the patch across all handled patch resolutions. For convenience and to speed up read times, the computed datasets were persisted as NumPy archives in the NPZ format. Dataset preparation was performed using Python programming language and machine learning and computer vision libraries such as scikit-learn [43], scikit-image [44], OpenCV [45] and TensorFlow Datasets [46].

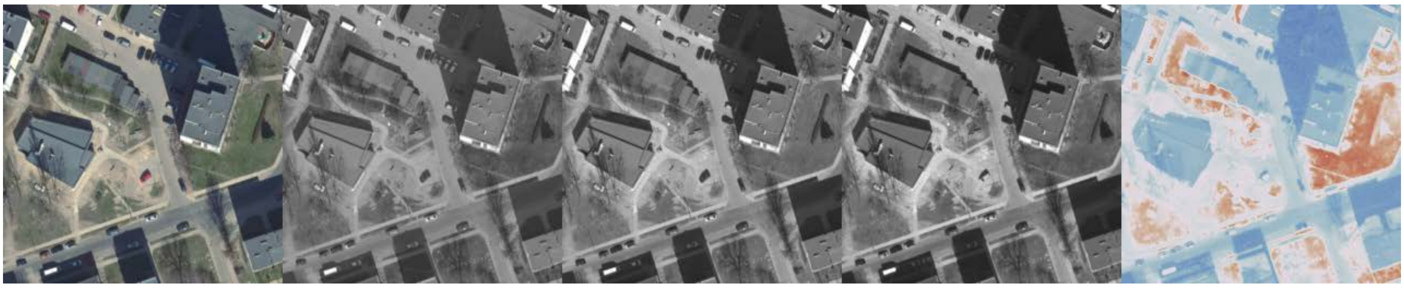

Figure 3. Visualization of a dataset item. From left to right: RGB composition, R, G, B bands that form the input tensor and NDVI, which serves as a target tensor. Source: own elaboration based on Geoportal [40].

During network training, each of the input-target pairs was extensively augmented by applying horizontal and vertical flipping, pixel intensity manipulation and performing random perspective transform. Augmentation was performed using imgaug – a library for image augmentation in machine learning experiments [47].

The test dataset was prepared using RGB orthoimagery acquired in 2012. Test orthophoto patches did not overlap with training and validation sets. Test items were similarly processed as those in the training set, excluding augmentation.

2.4. Evaluation Metrics

Due to the high variety of content presented in remote sensing imagery and the undoubted complexity of the natural image synthetization process, GAN's artificial image quality assessment is considered an extremely complicated task. One can not only rely on the quality of color reproduction but also pay attention to the structure, feature preservation and evaluate the level of realism and diversity of acquired results. What is more, when working with patches computed from a larger image, a confirmation is needed whether a reassembly yields a uniformly correct output. There are many reliable metrics available. Neither metric alone provides enough information to conduct a comprehensive assessment. Therefore, a combination of them is needed. The generative adversarial NDVI synthetization process has been monitored by the following metrics: Structural Similarity Index Measure (SSIM), Root Mean Square Error (RMSE), Peak Signal-to-Noise Ratio (PSNR) and Perceptual Evaluation (HYPE). Implementation of SSIM, RMSE and PSNR algorithms is available in the Python library image-similarity-measures [48]. Metrics were calculated against values directly produced by the generator network with sigmoid activation function output.

2.4.1. Root Mean Square Error (RMSE)

The Root Mean Square Error is an index used for determining the error of a model in forecasting quantitative data. In practice, it is the root of the mean square of the distance between estimated and obtained values. The indicator formula is as follows:

$$MSE(y,\hat{y}) = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{3}$$

$$RMSE(y,\hat{y}) = \sqrt{MSE(y,\hat{y})} \tag{4}$$

where $\hat{y}_i$ - predicted values, y_i - observed values, n - number of observations. RMSE is a standard statistical metric in such areas as meteorology, air quality and climate research, but the most popular is in the field of geoscience because it gives more weight to errors with larger absolute values than errors with smaller absolute values [49].

2.4.2. Structural Similarity Index Measure

Natural image signals are highly structured. Their pixels exhibit strong dependencies, especially when they are spatially proximate, and these dependencies carry important information about the structure of the objects in the visual scene [50]. Structural Similarity Index Measure (SSIM) provides a way to directly compare two images in terms of contrast, structure, and luminance. It is frequently used to assess image compression and reconstruction quality. SSIM value ranges between -1 and 1, where -1 indicates dissimilarity and 1 similarity. Its use in orthoimagery analysis is justified because by imitating human perception, it can recognize patterns and thus measure the quality of spatial features reproduction.

$$SSIM(y,\hat{y}) = \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_{\hat{y}}^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_{\hat{y}}^2 + \sigma_{\hat{y}}^2 + C_2)} \tag{5}$$

where μ - pixel value mean, σ -pixel values standard deviation and C_k-constant.

2.4.3. Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) is calculated based on mean square error (MSE). Its value approaches infinity as the MSE approaches zero. Therefore, higher image quality yields higher PSNR values. On the other hand, low values indicated high numerical differences between images. PSNR performs badly in discriminating structural content in images since various types of degradations applied to the same image can yield the same value of the MSE [51]. At the same time, PSNR can assess the results of image transformations which tend to apply different levels of noise to their output.

$$PSNR(y,\hat{y}) = 10\, log_{10}\left(\frac{255^2}{MSE(y,\hat{y})}\right) \tag{6}$$

2.4.4. Perceptual Evaluation

The Human eYe Perceptual Evaluation (HYPE) method measures the human deception rate. The $HYPE_{\infty}$ score rates a total error on a task supported by 50 fake and 50 real images. It enables capturing errors on both fake and real images and effects of hyper-realistic generation when fake images look even more realistic than real images [52]. $HYPE_{\infty}$ is calculated as a proportion of images that were judged incorrectly and aggregated the judgments over the *n* evaluators on *k* images to produce the final score. $HYPE_{\infty}$ values range from 0 to 1, where 0 indicates low synthetization capabilities, 0.5 is a sign of the good properties of creating artificial samples and 1 means that results are hyper-realistic and easy to distinguish from the real output. Two experts in the field of remote sensing were subject to the study to perceptually evaluate synthetic NDVI layers. Due to the complexity of the task experts were granted infinite evaluation time. Experts working independently visually inspected 155 and 101 randomly selected image sets, respectively, consisting of a panchromatic image, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff.}$ In the first stage, their task was to find out where the $NDVI_{artificial}$ images differed significantly from the $NDVI_{true}$ image, i.e., they contained overstated or underestimated NDVI values. In the second stage, attention was drawn to areas where the NDVI values were similar.

$$\mathrm{NDVI_{diff}} = \mathrm{NDVI_{true}} - \mathrm{NDVI_{artificial}} \tag{7}$$

2.5. Generative Adversarial Network

Generative adversarial networks (GANs) are neural networks characterized by a unique construction. GAN is composed of at least a single generator and accompanying discriminator. The role of the generator is to synthesize realistic results, while the discriminator must assess how close the generator output is to reality. By applying the discriminator judgment, the generator can iteratively improve its performance. Therefore, the main goal of these two networks is to compete in the form of a two-player minimax game, in which the generator tries to fool the discriminator [53]. More complex GANs are also known. They are built from multiple different subnetworks and frequently possess more than one instance of a single subnet type. A notable example is bidirectional GAN (BigBiGAN) which consists of a generator, three distinct discriminators and an encoder that can represent the image as a lower-dimensional vector [54].

Exploiting the results of the duel between different GAN components opens interesting opportunities both for computer vision (CV) and remote sensing (RS) activities. Due to their versatility, GANs are applicable to a wide variety of settings, such as unsupervised semantic segmentation, conditional artificial image preparation, inpainting or style transfer [55,56]. It is a common practice to extract a trained and fine-tuned GANs subnet to apply it to solve a certain research problem. Frequently, the generator is the target of such a procedure but there are also cases of using only the encoder [57]. The model architecture complexity is also a reason for multiple problems that one must face when training a GAN: vanishing gradient, mode collapse and convergence failure are the most common.

In the study, a simple yet powerful GAN architecture has been chosen as a baseline. Pix2Pix is a generator-discriminator network capable of conditional output creation. This makes it a variant of a conditional generative adversarial network (cGAN) [58]. Pix2Pix is a general-purpose solution to image-to-image translation problems, i.e., creating a mapping between input image x and output image y [59]. For the purpose of the study and after performing extensive experiments with different parameters, loss functions and hyperparameters tuning, the baseline TensorFlow architecture [60] has been utilized with following modifications:

- Applying depth-wise separable convolution layers [61] instead of default convolution layers to reduce resource requirements of the model.
- Multiplying the number of filters in each convolution layer by a factor of 2 to increase network capacity. Originally the first convolutional downsampling block starts with 64 filters. In the utilized approach 128 initial filters were used.
- Modifying the generator loss function by enhancing the existing implementation composed of a sum of binary cross-entropy (BCE) and mean absolute error (MAE) with an additional structure similarity index measure (SSIM).

Figures 4 and 5 present the fully modified Pix2Pix architecture. The overall network objective G* is an extension of the objective present in the original Pix2Pix paper [59]. It is as follows:

$$\mathcal{L}_{\mathrm{cGAN}}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,}\big[\log\big(1 - D(x,G(x))\big)\big] \tag{8}$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y}[||y\ - G(x)||_1] \tag{9}$$

$$\mathcal{L}_{SSIM}(G) = \mathbb{E}_{x,y}\big[1 - SSIM\big(y,G(x)\big)\big] \tag{10}$$

$$G^* = \arg\min_{D}\max_{G}\mathcal{L}_{\mathrm{cGAN}}(G,D) + \lambda\mathcal{L}_{L1}(G) + \omega\mathcal{L}_{SSIM}(G) \tag{11}$$

where x denotes the input PAN image, y denotes $\mathrm{NDVI}_{\mathrm{true}}$ and G(x) is the calculated $\mathrm{NDVI}_{\mathrm{artificial.}}$

The model training workflow is as follows. The input grayscale image (256px × 256px × 1 band) is fed to the generator. The generator produces a synthesized NDVI composition image (256px × 256px × 1 band). Both real and artificial NDVI composition images are transferred to the discriminator, which tries to assess in each case whether it is dealing with real or artificial images. After the assessment, the discriminator loss is calculated to measure the quality of the verdict. In the next phase, the generator loss is being calculated to consider both the quality of the synthesized output (MAE + SSIM) and the discriminator verdict (BCE). The training step ends when both generator and discriminator gradients are applied.

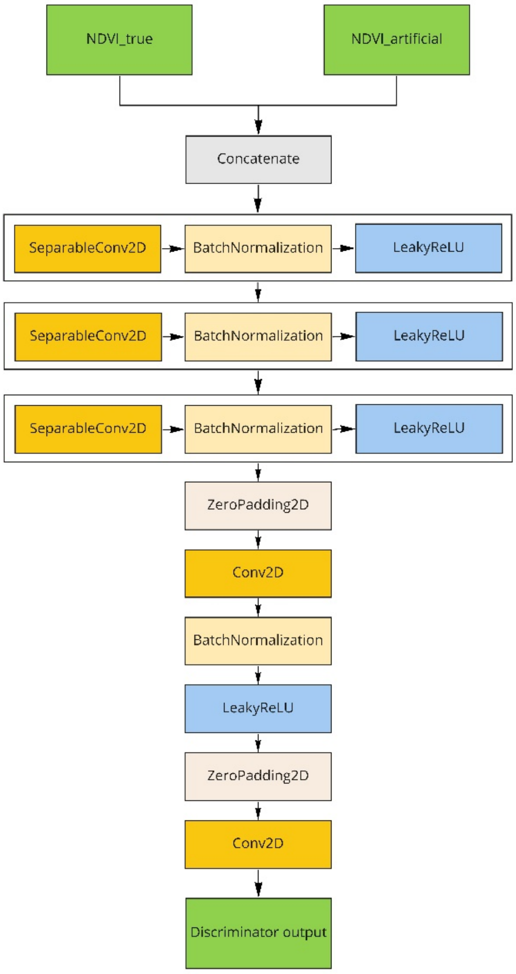

Figure 4. Modified Pix2Pix discriminator model visualization, source: own elaboration.

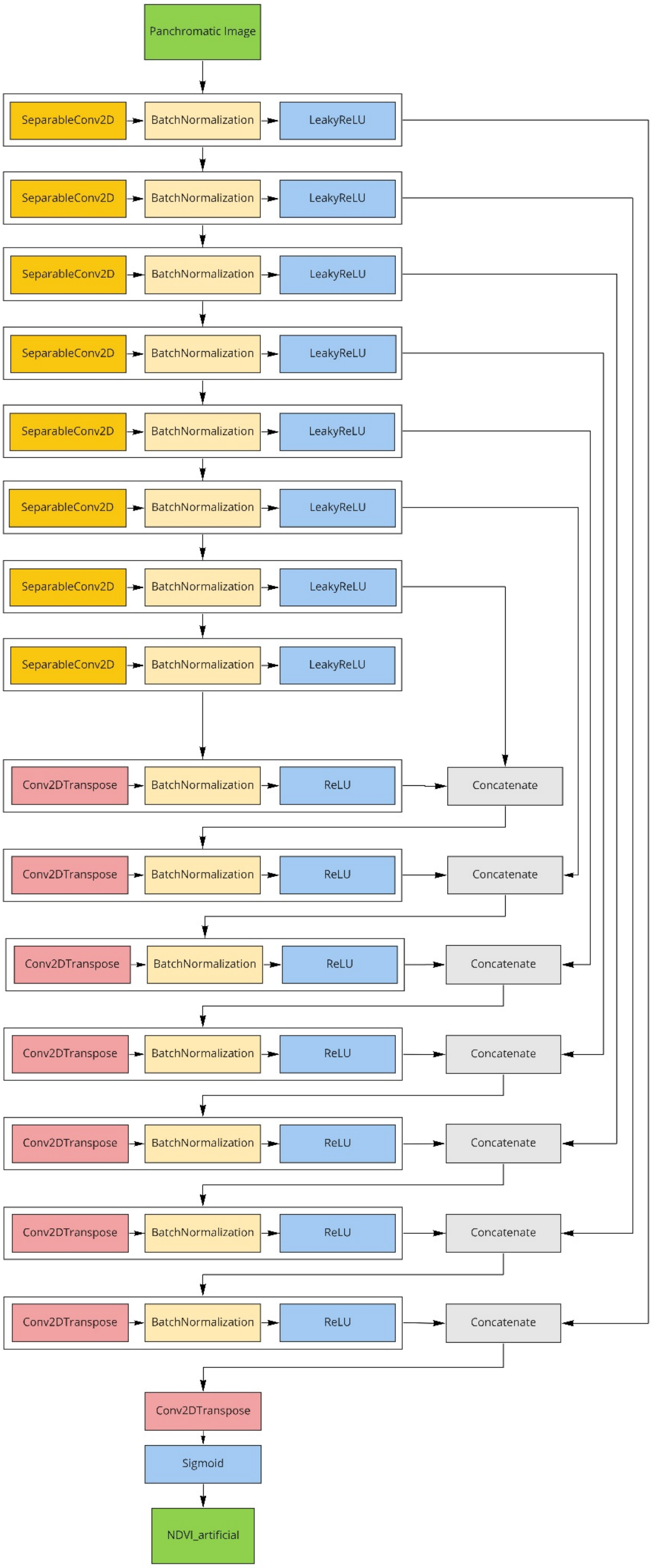

Figure 5. Modified Pix2Pix generator submodel visualization, source: own elaboration.

3. Results

In the interpretation process, the main forms of land cover were identified with which the visible regularities of the algorithms' behavior can be associated, i.e., whether the estimated $NDVI_{artificial}$ value was higher, close, or lower than the $NDVI_{true}$ value calculated from the infrared and red bands. Each presented figure shows, in order, panchromatic images (PAN), ground truth ($NDVI_{true}$), model inference result ($NDVI_{artificial}$) and differences between NDVI images ($NDVI_{diff}$). Figure 6 presents the utilized color map to produce the $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$ images.

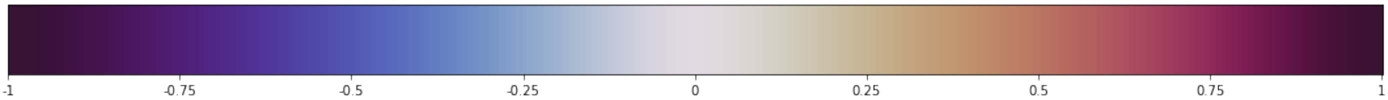

Figure 6. Color palette used during visual inspection of figures 7-15, source: own elaboration.

In the first analyzed area, a single row of low trees is visible (Figure. 7). On $NDVI_{artificial}$ image the line of trees has lost its rectilinear shape. $NDVI_{true}$ and $NDVI_{artificial}$ are comperable over a large area of meadows, where the texture is not very differentiated. The evaluation of the sample yielded 0.8764 SSIM and 0.0329 RMSE.

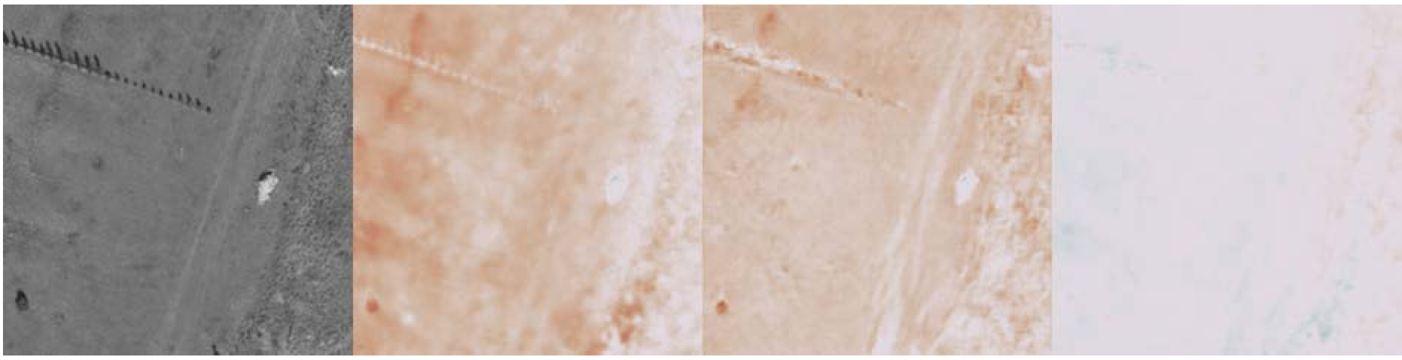

Figure 7. Inference result – abandoned land: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; source: own elaboration based on Geoportal [40].

The area presented in Figure 8 is entirely wooded. $NDVI_{true}$ has exposed only a few trees and one group, the rest of the trees and terrain are blue-toned. Some crowns or their fragments (in the tree group) are missing on $NDVI_{artificial}$. Two other trees have been found. The rest of the picture (background under the trees) is shown in the same way in both pictures ($NDVI_{artificial}$ equals $NDVI_{true}$). The evaluation of the sample yielded 0.7667 SSIM and 0.0519 RMSE.

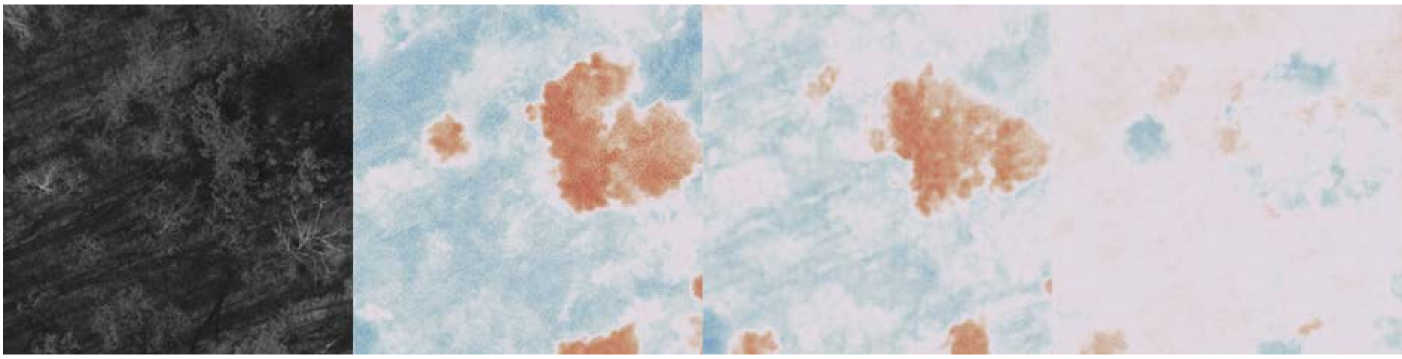

Figure 8. Inference result – wooded area: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; source: own elaboration based on Geoportal [40].

Figure 9 presents a parking lot with passenger cars. Most of the cars are properly handed over, one is underestimated (its shadow – falls on the surface of the yard - overestimated), the other car (white bodywork area on NDVItrue) is underestimated. The green stripe (a bit deviated) is better exposed on the $NDVI_{artificial}$ image. The garage roof was slightly overestimated. The shadow of the lantern is reproduced similarly in both NDVI images. The evaluation of the sample yielded 0.8553 SSIM and 0.0469 RMSE.

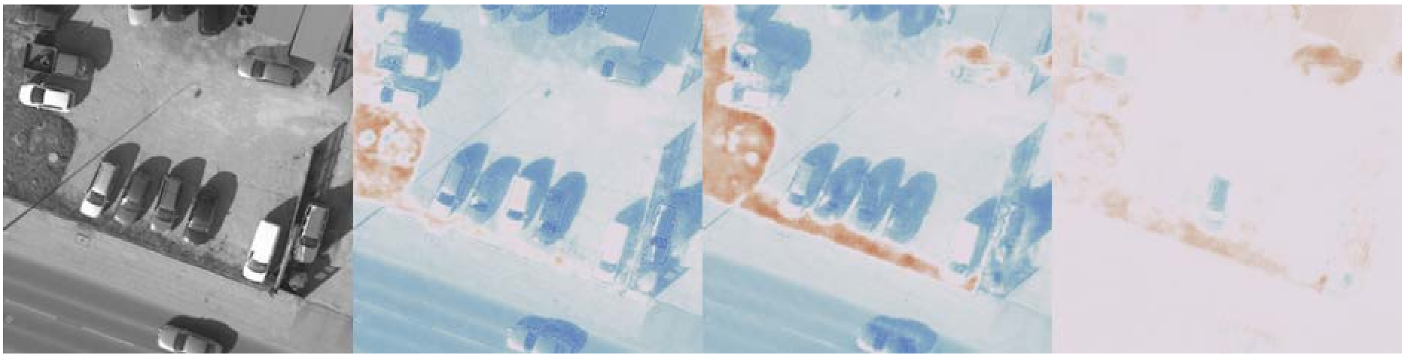

Figure 9. Inference result – small parking lot: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; source: own elaboration based on Geoportal [40].

Big buildings with vast shadows are shown in Figure 10. In the shadows, a warehouse of pallets, building materials – in places of greater chaos, textures marked on $NDVI_{artificial}$ image incorrectly as greenery. An increase of the $NDVI_{artificial}$ (correctly) was identified on lawns in the lit zone. Roofs and roads with no major differences between $NDVI_{artificial}$ and $NDVI_{true}$. The evaluation of the sample yielded 0.7859 SSIM and 0.0689 RMSE.

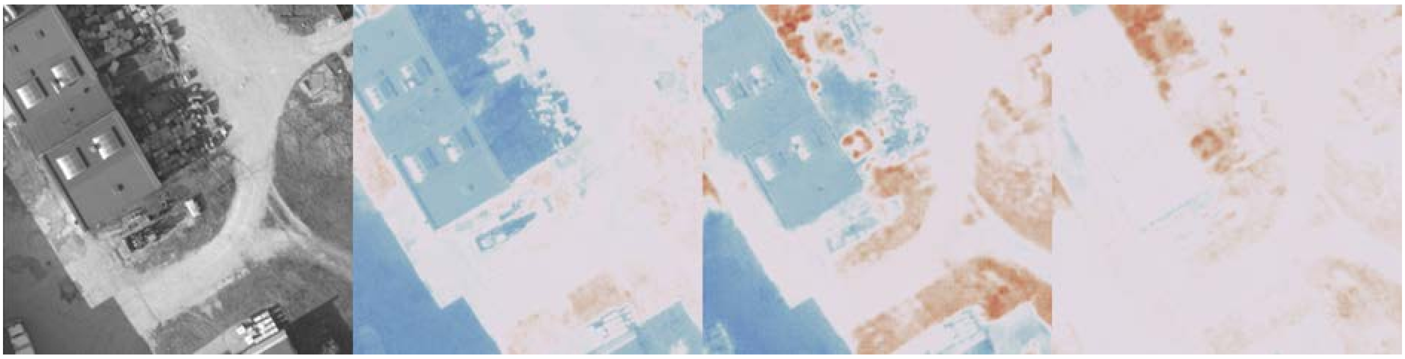

Figure 10. Inference result – buildings: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; source: own elaboration based on Geoportal [40].

Figure 11 represents an area of low and intensive development—an example of areas where there are large differences between $NDVI_{true}$ and $NDVI_{artificial}$. Many roof areas (approx. 90%) stand out as NDVI compliant. Other roofs may have $NDVI_{artificial}$ inflated when their edges are slightly blurred (e.g., in the left corner of the image). Underruns occur in the shadows cast by roofs. In the center of the bottom edge of the painting, the floral pattern, as if arranged in a scout cross, has a low $NDVI_{artificial}$, which can be interpreted in such a way that geometrized plant shapes are recognized as inanimate objects, while roofs with blurred edges - as lawns. The evaluation of the sample yielded 0.5651 SSIM and 0.0736 RMSE.

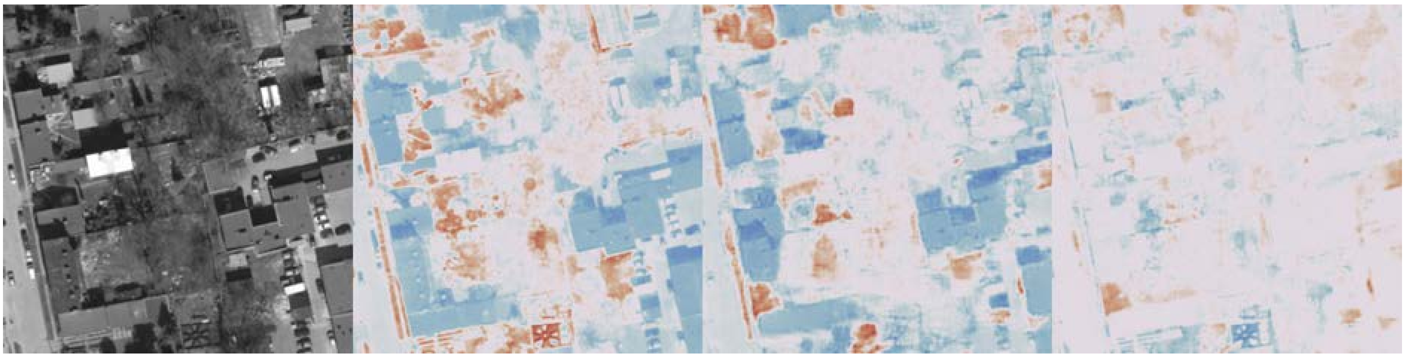

Figure 11. Inference result – residential area: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; source: own elaboration based on Geoportal [40].

The next figure (Figure 12) shows a large-scale commercial facility (40%), parking lot (25%) and grass. The biggest mistake – $NDVI_{artificial}$ revaluation – on the roof of the building in place of a long row of skylights, possibly recognized as a line of trees. On the same roof, next to it, there are irregularly placed ventilators, yet the entire area received the correct NDVI. Correct NDVI was identified on large sections of roads and large sections of roofs. $NDVI_{artificial}$ of the grassy area on the right side of the image is visibly underestimated. The evaluation of the sample yielded 0.6698 SSIM and 0.0740 RMSE.

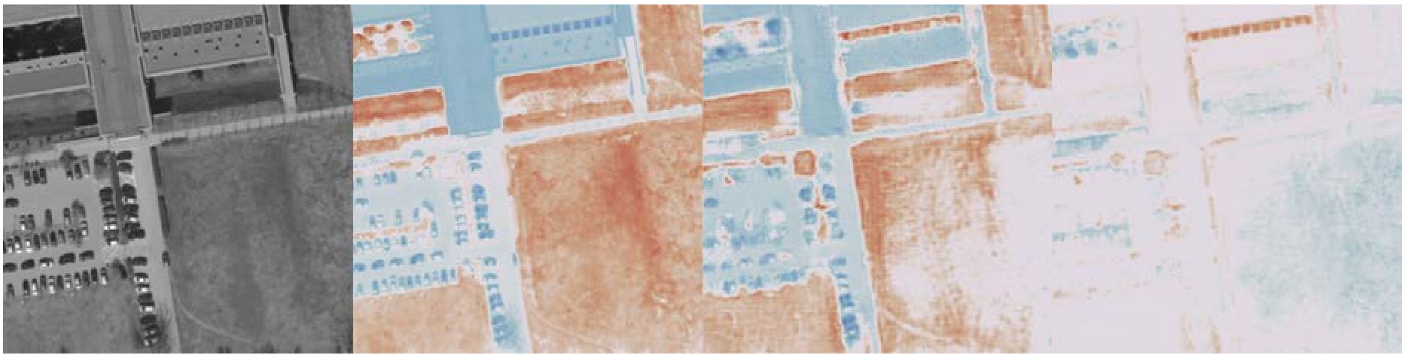

Figure 12. Inference result – commercial facility: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; source: own elaboration based on Geoportal [40].

The whole figure (Figure 13) presents an empty parking lot of a shopping center with large areas for vehicles and small, long, and narrow islands of greenery. Most or some of these islets have undervalued $NDVI_{artificial}$. One of the contours of the green area was blurred because, at some point it was tonally similar to the paving covering the adjacent part of the parking lot. The evaluation of the test sample yielded 0.8318 SSIM and 0.0496 RMSE.

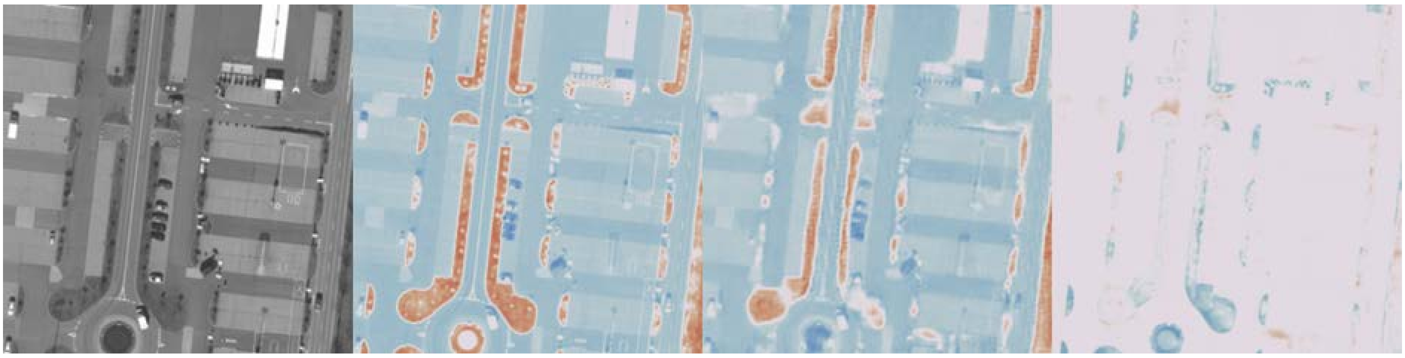

Figure 13. Inference result – commercial facility parking lot: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; source: own elaboration based on Geoportal [40].

Figure 14 shows a very large farmland area with an overestimated $NDVI_{artificial}$, where the textural features were very similar to those of the fields covered with vegetation. Other discrepancies in the fields arise along the lines resulting from the technology of establishing and cultivating the crop. The evaluation of the test sample yielded 0.7450 SSIM and 0.1317 RMSE.

Figure 14. Inference result –farmland: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; source: own elaboration based on Geoportal [40].

It was found that the algorithm heavily depends on textural information to function properly. It was clearly visible, for example, that lawns, if they had an amorphous structure (lack of texture) and were surrounded by straight, clear edges, generally had values lower than in the $NDVI_{true}$ image, i.e., they were recognized as surfaces of roofs or paved parking lots or roads. In addition, sunlit lawns, and thus brighter in the panchromatic image, erroneously obtained higher results than lawns in the shade and consequently darker ones. For arable fields, the algorithm overestimated the values for darker parts of crops.

For water-occupied areas, the artificially generated NDVI values were generally lower than $NDVI_{true}$. On the other hand, the opposite was true in the case of the immediate vicinity of a body of water or watercourse, which, due to higher humidity, had a darker shade in the panchromatic image (similar to the darker parts of crops).

The results closest to the actual $NDVI_{true}$ values were obtained for sheets with homogeneous coverage. If the terrain coverage is similar over a large part of the image, then $NDVI_{artificial}$ does not differ in practice from the $NDVI_{true}$ calculated from the red and infrared channels. Such objects are large areas of roofs, roads, and crops. On the other hand, the areas with the most differences between $NDVI_{artificial}$ and $NDVI_{true}$ were housing estates built with single-family houses, with their backup facilities such as a small driveway or a yard with garages, and gardens with diverse vegetation, consisting of irregular fragments of lawns, single shrubs and trees, or their groups. The whole picture is complemented by numerous fences (usually tight, sometimes in the form of walls), often planted with dense lines of not very tall trees or shrubs (it is difficult to distinguish individual specimens in the pictures), as well as various types of surfaces of housing estate roads or yards. It was noticed that the greatest differences between $NDVI_{true}$ and $NDVI_{artificial}$ in this area arose in situations where the structure and texture of the image in a fragment were more complex, less ordered. In such cases, they were recognized as areas covered with vegetation.

An interesting observation is that $NDVI_{artificial}$ does not indicate overestimated values for images of people, visible e.g., in the background of pavement, even though this indicator, calculated from RED and IR channels, incorrectly inflates the value in the place occupied by people in a given area.

In the case of tree canopy, which generally has high $NDVI_{true}$ values, the algorithm found their limits fairly well. The branch of a tree index values themselves could have varied due to the level of insolation – branch fragments on the sunny side generally received lower values.

$NDVI_{artificial}$ images do not accurately reproduce the textural information present in the original RGB, panchromatic and $NDVI_{true}$, such as the drawing of roof tiles, parking lot pavement tiles, details of the structure of branches of a tree, or they showed the deformation of linear green systems – tree rows, perfectly straight in the field, have been depicted as skewed in the synthesized image. An important result of the interpretation process is the statement that the artificially generated image is less accurate, especially in the case of buildings, because the edges of these objects are blurred, less sharp. As a result, one can get the impression that the image obtained has a lower spatial resolution.

Table 3 presents evaluation metrics calculated for all orthophoto patches stored in the test dataset. It shows the average, standard deviation, minimum and maximum values of the metrics for monitoring the generative adversarial NDVI synthetization process.

Table 3. Test set evaluation metrics.

	SSIM	**PSNR**	**RSME**
AVG	0.7569	26.6459	0.0504
STD	0.1083	3.6577	0.0193
MIN	0.3589	16.3343	0.0026
MAX	0.9987	51.7674	0.1525

4. Discussion

The performed experiment showed that the structure and texture of the panchromatic aerial remote sensing image contain information that allows us to estimate the NDVI value for various objects of urban space, although we do not know the real values of the brightness of bands based on which NDVI is calculated, i.e., RED and NIR. Figure 15 presents NDVI calculated solely from a panchromatic version of the aerial image utilizing the sliding window inference approach.

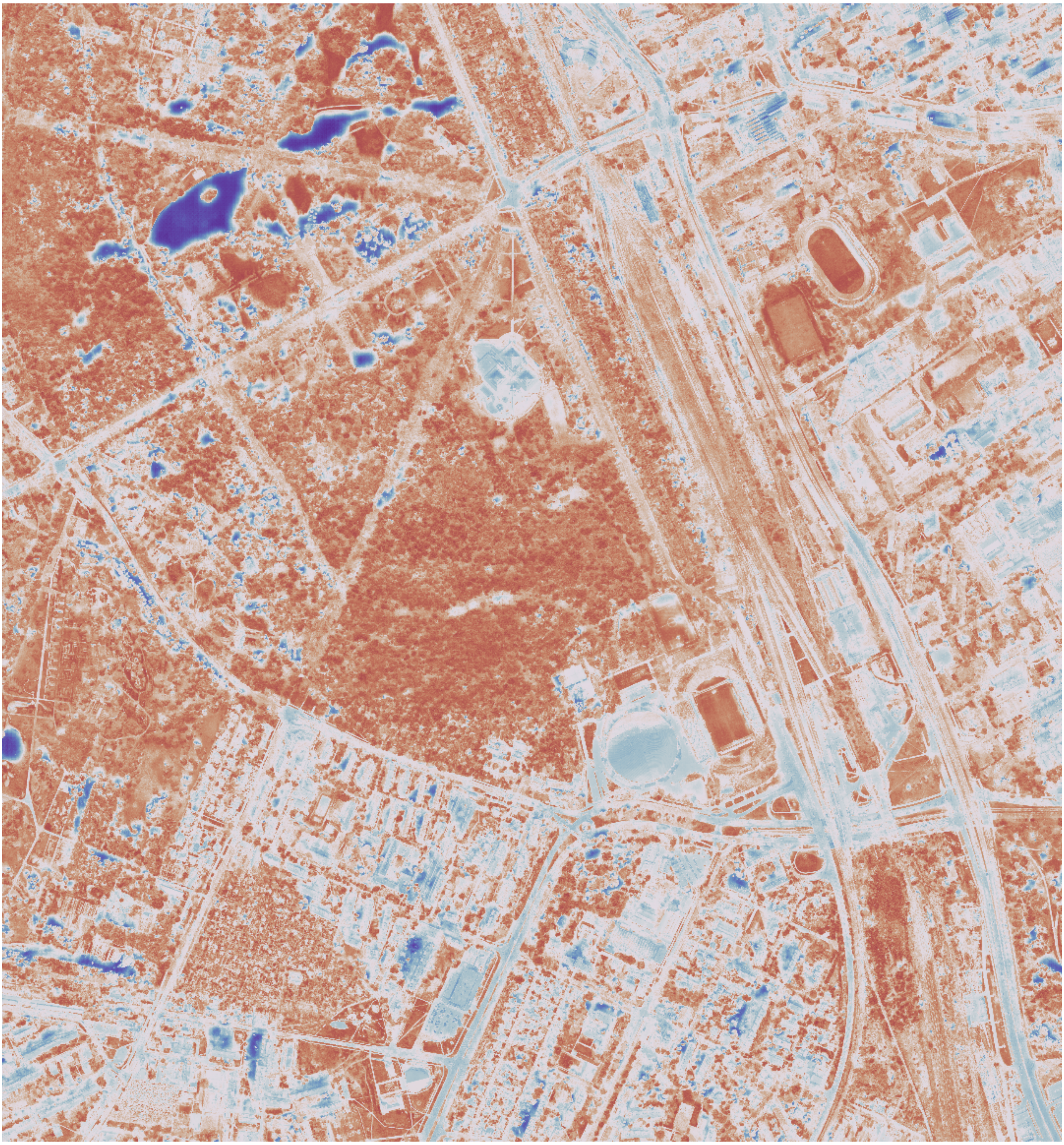

Figure 15. Sliding window inference ($NDVI_{artificial}$) of an orthophoto used to compute the test dataset. Scene presents Łódź, Poland. Red values indicate high NDVI values (closer to 1). Blue ones represent small values (closer to -1); source: own elaboration based on Geoportal [40].

In the light of the assessment of the obtained $NDVI_{artificial}$ images, let us consider the factors that facilitated or, on the contrary, made it difficult to achieve the intended goal. First, note that the aerial photos based on which the grayscale orthomosaic used in the experiment were made from the early spring period, so the trees and some shrubs were not covered with leaves, or the leaves were not yet fully developed. Therefore, in the real picture ($NDVI_{true}$), there should not be especially high NDVI values in places of shrubs and trees because trunks and branches are similar to objects of inanimate nature. However, even under such conditions, the algorithm was able to appropriately assign values to the $\mathrm{NDVI}_{\mathrm{artificial}}$ image similar to those it learned to recognize on the $\mathrm{NDVI}_{\mathrm{true}}$ ground truth.

At this point, the question arises whether the result would not be even better if we used pairs of pictures for learning: panchromatic (from spring, as these are mainly performed for cities in our geographical zone) and $\mathrm{NDVI}_{\mathrm{true}}$ generated from the summer image when the vegetation is fully developed? Perhaps we could get $\mathrm{NDVI}_{\mathrm{artificial}}$ values that describe vegetation better? From this perspective, spring images (originally taken as panchromatic or linearly synthesized from RGB channels) could be used to create images showing the predicted NDVI distribution in the full growing season. The practical significance of such an image (let's call it the approximated summer NDVI) may result from allowing us to quickly identify places where changes in the spatial structure of vegetation were made during the leafless period (change detection). It would be enough to take panchromatic or even RGB + NIR photos in spring, apply a trained algorithm (here you would need photos from summer once) and you could create $\mathrm{NDVI}_{\mathrm{artificial}}$ every spring.

One more aspect needs to be considered, namely the classification of objects, especially vegetation, which is typical for remote sensing. Summer photos are considered to be the best for this purpose as the vegetation is in full growth and it is possible to capture the spectral differences between particular species or their genera. You can increase the accuracy of the classification by entering information about the structure and texture of the image. It should be noted, however, that we can only have information about the structure and texture of the top layer of tree crowns. The use of additional information from spring images will allow you to penetrate deep into the tree crowns, show the details of their internal structure, which should increase the identification potential. In addition, it is worth remembering that panchromatic images, those recorded by photogrammetric cameras as panchromatic, and not synthesized from RGB + NIR channels, have a good spatial resolution, better than each of these channels considered separately (at least 1: 2 - linear, or 1: 4 - surface).

When photographing from an aerial view, however, we must consider the problem of a specific representation of reality in the perspective projection. Each object with a third dimension (building, tree) has shifted images of the upper parts (roofs, trees - crowns) in relation to the lower parts (building bases, tree trunks), generally from the center of the photo to the outside. In photos taken from different points in space, this deviation for the same objects will be visible in different directions. The shift effect gets greater as the subject is further from the center of the image, and it decreases as the shooting height increases. This makes it difficult to take pictures, e.g., in spring and summer and at any other time that would have similar geometric properties (i.e., deviations of spatial objects in the same directions) due to the need to overlap them because it would be difficult to take the shots from exactly the same points in space. Part of the solution to the problem is to significantly increase the height of aerial photography or even the use of high-resolution satellite images. In the context of creating an $\mathrm{NDVI}_{\mathrm{artificial}}$ image, this observation should be read in such a way that if we had greyscale images of a given area, taken on the same date but from different points in space, we could expect differences in the estimation of $\mathrm{NDVI}_{\mathrm{artificial}}$ in many places. These differences would result from the fact that the images of the crowns of the same trees would be visible in the background of other objects and in a different context.

At this point, there is probably a high potential for NDVI estimation based on panchromatic images forming stereoscopic pairs and thus containing 3D information. Having full control over all geometrical relations between two stereoscopic images and the terrain, i.e., having the so-called parameters of internal and external orientation, we enter the area of known issues, typical in photogrammetry and remote sensing. Then finding corresponding fragments of two images is not difficult. Appropriate algorithms have long been used to correlate images. Digital image matching (DIM) techniques are now widely used to automatically generate Digital Surface Models and Digital Terrain Models from satellite or aerial images of different resolutions [62,63] and further to detect changes, e.g., as a result of the development of plant cover or movements [64]. The $NDVI_{artifcial}$ estimation method proposed by us can be interwoven into ML-based digital image correlation (DIC) procedures to find the same objects, e.g., in oblique photos, multi-view oblique images [65], or pairs of images even with large differences in viewpoints [66]. We expect that having information about the third dimension of objects can be used to develop expansions of known "two-dimensional" textural features into their "three-dimensional" versions [28], which will facilitate the recognition of objects, determining their shapes, and estimating their $NDVI_{artificial}$ on a level similar to $NDVI_{true}$. The prospects for improving the obtained results, including some guidelines for further research, maybe the inclusion in the research of orthoimages with the fully removed radial displacement effect, i.e., true orthophoto. For their production, normalized Digital Surface Models (nDSM) are used, compiled based on LiDAR laser scanning data. Currently, it is difficult to find a more effective and accurate source of information about the topography and objects located on it. The nDSM geometry is correct - the projection is orthogonal, there are no displacements. However, there is some loss of photographic image quality due to pixel resampling. However, it is positive that LiDAR data (point clouds) have features that make it possible to distinguish vegetation from other objects, such as buildings.

It was noticed that the algorithm used in the research could "repair" the image, in a way "correct" the NDVI value. This effect is visible in places of shadows cast by tall buildings and is manifested by an increase in the index value. This is partly in line with the results obtained by Myeong et al. [67], who used high-resolution digital infrared images and divided the urban space into trees and shrubs, lawns, soil without cover, water, and impervious surfaces. As they stated, the shaded areas were a challenge (similar problems are also reported by Pyra & Adamczyk [23]), but also large spectral similarities between the distinguished land cover classes. Multiple attempts to select and apply textures, masks and majority filters allowed to increase the accuracy of the classification. As we believe, the condition for obtaining good results with our algorithm is that the structure and texture should be visible in the shadows, digitally expressed, although often invisible to the human eye, which can be associated with vegetation.

The obtained results were also influenced by a certain property of the textural features, namely the sensitivity to changes in some image parameters. As reported by Książek [68], changes in contrast, brightness, sharpness and image rotation may have an impact on several groups of textural features of roofs selected by her for research (asbestos roofs were dealt with). The obtained results were also analyzed in terms of parameters insensitive to image transformations, but no texture features were found that would be resistant to all factors causing their change. In our experience, this observation is of such importance that when estimating NDVI, one should probably consider the generation of an extended set of textural features from images in a similar manner to the one mentioned above, modified.

Estimating the NDVI based on archival panchromatic images makes it possible to study the condition and changes in urban vegetation cover in the long term. The research may concern with changes in the amount of greenery resulting from its expansion as a result of the economic activity of municipal authorities and natural growth, but it also helps to reveal places of unnecessarily excessive or even illegal tree removal.

It is hard to expect that the generated $NDVI_{artificial}$ images will fulfill all the functions that NDVI applies to in modern scientific research and practice. We expect that the differences between the synthesized and real NDVI are significant enough to allow only a simple classification: vegetation - not vegetation, as well as estimating the size of biomass, or, e.g., determining the size of the shade effect, in the sense in which they discuss this issue [18,19]. However, it may be problematic to try to classify vegetation even into large groups (e.g., coniferous - deciduous), and certainly in such subtle tasks as assessing the health condition and water stress of vegetation, crop maturity, nutrient content, e.g., nitrogen, chlorophyll content, estimated yields, and detection of disease or insect infestation.

The lack of high precision does not reduce the applicability of this method when combined with other machine learning methods. Achieved results enable utilizing the results both as a data augmentation technique and an inpainting mechanism. The first approach could involve preparing a dataset of $NDVI_{artificial}$ images enriching an existing dataset, and supporting the training process with new realistic but artificial samples. Data augmentation is nowadays an indispensable but difficult step in scientific research using neural networks. A similar situation occurs in the case of inpainting which can be utilized to fill in missing data in remote sensing imagery. Not only can the method be used to generate missing NDVI values from panchromatic imagery, but it also enables calculating the NIR band when dealing with RGB images.

5. Conclusions

This research discussed and confirmed the potential of utilizing conditional generative adversarial networks (cGAN) to perform an image-to-image translation capable of estimating NDVI from a single panchromatic orthoimagery. The customized Pix2Pix model was trained on patches acquired from an orthophoto presenting an urban area. The network achieved 0.7569 ± 0.1083 SSIM, 26.6459 ± 3.6577 PNSR and 0.0504 ± 0.0193 RSME on the test set. Furthermore, the model passed a thorough inspection during perceptual evaluation, during which the main areas of error were identified. The overall results are satisfactory. The estimated NDVI values can be utilized to distinguish areas of different vegetation volumes. Although the predicted NDVI values in pixel scope are not precise enough to utilize the model as a replacement for traditionally computed NDVI, it can be used in a variety of use cases, from archival RGB and panchromatic imagery processing to serving as a data augmentation or an inpainting technique.

Author Contributions: Conceptualization, M.A., K.B., A.B.; Methodology, M.A.; Software, M.A.; Validation, M.A., K.B., A.B.; Formal analysis, M.A.; Investigation, M.A., K.B., A.B.; Resources, M.A., A.B.; Data curation, M.A., A.B.; Writing—original draft preparation, M.A., K.B., and A.B..; Writing—review and editing, M.A., K.B., A.B.; Visualization, M.A., A.B.; Supervision, K.B.; Project administration, M.A.

All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were analyzed in this study: https://www.geoportal.gov.pl/.

Conflicts of Interest: The authors declare no conflict of interest

References

1. Deering, D.W. Rangeland Reflectance Characteristics Measured by Aircraft and Spacecraft Sensors. Thesis, Texas A&M University. Libraries, 1978.
2. Jackson, R.D.; Huete, A.R. Interpreting Vegetation Indices. *Preventive Veterinary Medicine* **1991**, *11*, 185–200, doi:10.1016/S0167-5877(05)80004-2.
3. Jarocińska, A.; Zagajewski, B. Correlations of ground- and airborne-level acquired vegetation indices of the Bystrzanka catchment. *Teledetekcja Środowiska* **2008**, *40*, 100–124.
4. Tomaszewska, M.; Lewiński, S.; Woźniak, E. Use of MODIS Satellite Images to Study the Percentage of Vegetation Cover. *Teledetekcja Środowiska* **2011**, *46*, 15–22.

5. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment* **1979**, *8*, 127–150, doi:10.1016/0034-4257(79)90013-0.
6. Hatfield, J.L.; Prueger, J.H. Value of Using Different Vegetative Indices to Quantify Agricultural Crop Characteristics at Different Growth Stages under Varying Management Practices. *Remote Sensing* **2010**, *2*, 562–578, doi:10.3390/rs2020562.
7. Tuszynska, J.; Gatkowska, M.; Wrobel, K.; Jagiello, K. A Pilot Study on Determining Approximate Date of Crop Harvest on the Basis of Sentinel-2 Satellite Imagery. *Geoinformation Issues* **2018**, *Vol. 10*.
8. Hunt, E.R.; Rock, B.N. Detection of Changes in Leaf Water Content Using Near- and Middle-Infrared Reflectances. *Remote Sensing of Environment* **1989**, *30*, 43–54, doi:10.1016/0034-4257(89)90046-1.
9. Jackson, T.J.; Chen, D.; Cosh, M.; Li, F.; Anderson, M.; Walthall, C.; Doriaswamy, P.; Hunt, E.R. Vegetation Water Content Mapping Using Landsat Data Derived Normalized Difference Water Index for Corn and Soybeans. *Remote Sensing of Environment* **2004**, *92*, 475–482, doi:10.1016/j.rse.2003.10.021.
10. Gu, Y.; Brown, J.F.; Verdin, J.P.; Wardlow, B. A Five-Year Analysis of MODIS NDVI and NDWI for Grassland Drought Assessment over the Central Great Plains of the United States. *Geophysical Research Letters* **2007**, *34*, doi:10.1029/2006GL029127.
11. Bagheri, N.; Ahmadi, H.; Alavipanah, S.K.; Omid, M. Multispectral Remote Sensing for Site-Specific Nitrogen Fertilizer Management. *Pesq. agropec. bras.* **2013**, *48*, 1394–1401, doi:10.1590/S0100-204X2013001000011.
12. Turlej, K. Comparison of NDVI index based on NOAA AVHRR, SPOT-VEGETATION and TERRA MODIS satellite data. *Teledetekcja Środowiska* **2009**, 83–88.
13. Sultana, S.R.; Ali, A.; Ahmad, A.; Mubeen, M.; Zia-Ul-Haq, M.; Ahmad, S.; Ercisli, S.; Jaafar, H.Z.E. Normalized Difference Vegetation Index as a Tool for Wheat Yield Estimation: A Case Study from Faisalabad, Pakistan. *The Scientific World Journal* **2014**, *2014*, e725326, doi:10.1155/2014/725326.
14. Chew, W.C.; Hashim, M.; Lau, A.M.S.; Battay, A.E.; Kang, C.S. Early Detection of Plant Disease Using Close Range Sensing System for Input into Digital Earth Environment. *IOP Conf. Ser.: Earth Environ. Sci.* **2014**, *18*, 012143, doi:10.1088/1755-1315/18/1/012143.
15. Nowak, D.J.; Greenfield, E.J. Tree and Impervious Cover Change in U.S. Cities. *Urban Forestry & Urban Greening* **2012**, *11*, 21–30, doi:10.1016/j.ufug.2011.11.005.
16. Kuang, W.; Dou, Y. Investigating the Patterns and Dynamics of Urban Green Space in China's 70 Major Cities Using Satellite Remote Sensing. *Remote Sensing* **2020**, *12*, 1929, doi:10.3390/rs12121929.
17. Zięba-Kulawik, K.; Hawryło, P.; Wężyk, P.; Matczak, P.; Przewoźna, P.; Inglot, A.; Mączka, K. Improving Methods to Calculate the Loss of Ecosystem Services Provided by Urban Trees Using LiDAR and Aerial Orthophotos. *Urban Forestry & Urban Greening* **2021**, *63*, 127195, doi:10.1016/j.ufug.2021.127195.
18. Li, X.; Ratti, C. Mapping the Spatial Distribution of Shade Provision of Street Trees in Boston Using Google Street View Panoramas. *Urban Forestry & Urban Greening* **2018**, *31*, 109–119, doi:10.1016/j.ufug.2018.02.013.
19. McPherson, E.G.; Xiao, Q.; Doorn, N.S. van; Johnson, N.; Albers, S.; Peper, P.J. Shade Factors for 149 Taxa of In-Leaf Urban Trees in the USA. *Urban Forestry & Urban Greening. 31: 204-211* **2018**, *31*, 204–211, doi:10.1016/j.ufug.2018.03.001.
20. Kubalska, J.; Preuss, R. Use of the Photogrammetric Data for Vegetation Inventory on Urban Areas. *Archiwum Fotogrametrii, Kartografii i Teledetekcji* **2014**, 75–86, doi:10.14681/AFKIT.2014.006.
21. Krukowski, M.; Cebrykow, P.; Płusa, J. Classification of Green Areas in Lublin Based on Satellite Images Ikonos 2. *Barometr Regionalny* **2016**, *14*, 35–44.

22. Będkowski, K.; Bielecki, A. Assessment of the Availability of Greenery in the Place of Residence in Cities Using NDVI and the Lorenz's Concentration Curve. *Teledetekcja Środowiska* **2017**, *57*, 5–14.
23. Pyra, M.; Adamczyk, J. Object-Oriented Classification in the Inventory of Green Infrastructure Objects on the Example of the Ursynów District in Warsaw. *Teledetekcja Środowiska* **2018**, *T. 59*.
24. Pluto-Kossakowska, J.; Władyka, M.; Tulkowska, W. Assessment of remote sensing image data to identify objects in green and blue infrastructure. *Teledetekcja Środowiska* **2018**, *T. 59*.
25. Krukowski, M. Modelowanie Kartograficzne w Ocenie Jakości Życia w Mieście – Aspekt Zieleni Miejskiej w Lublinie. *Annales Universitatis Mariae Curie-Sklodowska, sectio B – Geographia, Geologia, Mineralogia et Petrographia* **2018**, *73*, 7–27, doi:10.17951/b.2018.73.0.7-27.
26. Worm, A.; Będkowski, K.; Bielecki, A. The use of surface and volume indicators from high resolution remote sensing data to assess the vegetation filling of urban quarters in Łódź city centre, Poland. *Teledetekcja Środowiska* **2019**, *T. 60*.
27. Łachowski, W.; Łęczek, A. Tereny zielone w dużych miastach Polski. Analiza z wykorzystaniem Sentinel 2. *Urban Development Issues* **2021**, *66*, 77–90, doi:10.51733/udi.2020.68.07.
28. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics* **1973**, *SMC-3*, 610–621, doi:10.1109/TSMC.1973.4309314.
29. Li, P.; Cheng, T.; Guo, J. Multivariate Image Texture by Multivariate Variogram for Multispectral Image Classification. *Photogrammetric Engineering & Remote Sensing* **2009**, *75*, 147–157, doi:10.14358/PERS.75.2.147.
30. Zhang, Y. Texture-Integrated Classification of Urban Treed Areas in High-Resolution Color-Infrared Imagery. *Photogrammetric Engineering & Remote Sensing* **2001**, *67*, 1359–1365.
31. Herold, M.; Liu, X.; Clarke, K. Spatial Metrics and Image Texture for Mapping Urban Land Use. *Photogrammetric Engineering and Remote Sensing* **2003**, *69*, 991–1001, doi:10.14358/PERS.69.9.991.
32. Marmol, U.; Lenda, G. Texture Filters in the Process of Automatic Object Classification. *Archiwum Fotogrametrii, Kartografii i Teledetekcji* **2010**, *21*.
33. Barley, A.; Town, C. Combinations of Feature Descriptors for Texture Image Classification. *Journal of Data Analysis and Information Processing* **2014**, *2*, 67–76, doi:10.4236/jdaip.2014.23009.
34. Suarez, P.L.; Sappa, A.D.; Vintimilla, B.X. Learning Image Vegetation Index through a Conditional Generative Adversarial Network. In Proceedings of the 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM); October 2017; pp. 1–6.
35. Suárez, P.L.; Sappa, A.D.; Vintimilla, B.X.; Hammoud, R.I. Image Vegetation Index Through a Cycle Generative Adversarial Network. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); June 2019; pp. 1014–1021.
36. Aslahishahri, M.; Stanley, K.G.; Duddu, H.; Shirtliffe, S.; Vail, S.; Bett, K.; Pozniak, C.; Stavness, I. From RGB to NIR: Predicting of near Infrared Reflectance from Visible Spectrum Aerial Images of Crops. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW); October 2021; pp. 1312–1322.
37. Barwiński, M. Spatial Development and Functional Changes in Łódź - Geographic, Economic and Political Conditions. *Geografia w szkole* **2009**, 38–50.
38. Head Office of Geodesy and Cartography Integrated copies of databases of topographic objects Available online: http://www.gugik.gov.pl/pzgik/zamow-dane/baza-danych-obiektow-topograficznych-bdot-10k (accessed on 11 November 2020).
39. Statistics Poland Statistics of Łódź 2020 Available online: https://lodz.stat.gov.pl/en/publications/statistical-yearbook/statistics-of-lodz-2020,1,16.html (accessed on 27 January 2022).

40. Geoportal Available online: http://geoportal.gov.pl (accessed on 11 November 2020).
41. EnviroSolutions Sp. z o.o. - Michał Włoga Pobieracz Danych GUGiK Available online: https://plugins.qgis.org/plugins/pobieracz_danych_gugik/ (accessed on 27 January 2022).
42. Small, C. Estimation of Urban Vegetation Abundance by Spectral Mixture Analysis. *International Journal of Remote Sensing* **2001**, *22*, 1305–1334, doi:10.1080/01431160151144369.
43. Scikit-Learn: Machine Learning in Python — Scikit-Learn 1.0.2 Documentation Available online: https://scikit-learn.org/stable/ (accessed on 27 January 2022).
44. van der Walt, S.; Schönberger, J.L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J.D.; Yager, N.; Gouillart, E.; Yu, T. Scikit-Image: Image Processing in Python. *PeerJ* **2014**, *2*, e453, doi:10/gftp3s.
45. OpenCV Available online: https://opencv.org/ (accessed on 27 January 2022).
46. TensorFlow Datasets Available online: https://www.tensorflow.org/datasets (accessed on 27 January 2022).
47. Jung, A. *Imgaug*; 2022;
48. Müller, M.U.; Ekhtiari, N.; Almeida, R.M.; Rieke, C. Super-Resolution of Multispectral Satellite Images Using Convolutional Neural Networks. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2020**, *V-1–2020*, 33–40, doi:10.5194/isprs-annals-V-1-2020-33-2020.
49. Chai, T.; Draxler, R.R. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments against Avoiding RMSE in the Literature. *Geoscientific Model Development* **2014**, *7*, 1247–1250, doi:10.5194/gmd-7-1247-2014.
50. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing* **2004**, *13*, 600–612, doi:10.1109/TIP.2003.819861.
51. Horé, A.; Ziou, D. Image Quality Metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition; August 2010; pp. 2366–2369.
52. Zhou, S.; Gordon, M.L.; Krishna, R.; Narcomey, A.; Fei-Fei, L.; Bernstein, M.S. HYPE: A Benchmark for Human EYe Perceptual Evaluation of Generative Models. *arXiv:1904.01121 [cs]* **2019**.
53. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved Techniques for Training GANs. *arXiv:1606.03498 [cs]* **2016**.
54. Donahue, J.; Simonyan, K. Large Scale Adversarial Representation Learning. *arXiv:1907.02544 [cs, stat]* **2019**.
55. Demir, U.; Unal, G. Patch-Based Image Inpainting with Generative Adversarial Networks. *arXiv:1803.07422 [cs]* **2018**.
56. Dong, J.; Yin, R.; Sun, X.; Li, Q.; Yang, Y.; Qin, X. Inpainting of Remote Sensing SST Images With Deep Convolutional Generative Adversarial Network. *IEEE Geoscience and Remote Sensing Letters* **2019**, *16*, 173–177, doi:10.1109/LGRS.2018.2870880.
57. Adamiak, M.; Będkowski, K.; Majchrowska, A. Aerial Imagery Feature Engineering Using Bidirectional Generative Adversarial Networks: A Case Study of the Pilica River Region, Poland. *Remote Sensing* **2021**, *13*, 306, doi:10.3390/rs13020306.
58. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv:1411.1784 [cs, stat]* **2014**.
59. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv:1611.07004 [cs]* **2018**.
60. TensorFlow *TensorFlow Documentation*; tensorflow, 2022;
61. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. *arXiv:1610.02357 [cs]* **2017**.
62. Davis, C.H.; Wang, X. High-Resolution DEMS for Urban Applications from NAPP Photography. *Photogrammetric Engineering and Remote Sensing* **2011**, *67*.

63. Koza, P. Orientation of Ikonos stereo images and automatic acquisition of height models. *Archiwum Fotogrametrii, Kartografii i Teledetekcji* **2006**, *Vol. 16*.
64. Dematteis, N.; Giordan, D. Comparison of Digital Image Correlation Methods and the Impact of Noise in Geoscience Applications. *Remote Sensing* **2021**, *13*, 327, doi:10.3390/rs13020327.
65. Verykokou, S.; Ioannidis, C. *A Global Photogrammetry-Based Structure from Motion Framework: Application in Oblique Aerial Images*; 2019;
66. Yao, G.; Yilmaz, A.; Zhang, L.; Meng, F.; Ai, H.; Jin, F. Matching Large Baseline Oblique Stereo Images Using an End-to-End Convolutional Neural Network. *Remote Sensing* **2021**, *13*, 274, doi:10.3390/rs13020274.
67. Myeong, S.; Nowak, D.J.; Hopkins, P.F.; Brock, R.H. Urban Cover Mapping Using Digital, High-Resolution Aerial Imagery. *Urban Ecosystems. 5: 243-256.* **2003**, *5*.
68. Książek, J. Study of Selected Textural Features Properties on Asbestos Roof Images. *Geomatics and Environmental Engineering* **2018**, *Vol. 12*, doi:10.7494/geom.2018.12.4.45.