GENERATIVE ADVERSARIAL APPROACH TO URBAN AREAS' NDVI ESTIMATION: A CASE STUDY OF ŁÓDŹ, POLAND

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

[1] ReasonField Lab, Warszawa, Polska
[2] University of Lodz, Faculty of Geographical Sciences, Institute of Urban Geography, Tourism Studies and Geoinformation, Łódź, Poland

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 utilising GANs in remote sensing (RS). The normalised difference vegetation index (NDVI) ground-truth labels were prepared by combining RGB and NIR orthophotos. The dataset was then utilised 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 256 px × 256 px patch using only information available in the panchromatic input. The network achieved a structural similarity index measure (SSIM) of 0.7569 ± 0.1083, a peak signal-to-noise ratio (PSNR) of 26.6459 ± 3.6577 and a root-mean-square error (RSME) of 0.0504 ± 0.0193 on the test set, which should be considered high. The perceptual evaluation was performed to verify the method's usability when working with a real-life scenario. The research confirms that the structure and texture of the panchromatic aerial RS image contain sufficient information for NDVI estimation for various objects of urban space. Even though these results can highlight areas rich in vegetation and distinguish them from the urban background, there is still room for improvement regarding the accuracy of the estimated values. The research aims to explore the possibility of utilising GAN to enhance panchromatic images (PAN) with information related to vegetation. This opens exciting opportunities for historical RS imagery processing and analysis.

KEY WORDS: generative adversarial networks, NDVI, green areas, orthophoto, artificial datasets

Corresponding author: Adam Bielecki, adam.bielecki@edu.uni.lodz.pl

Introduction

The normalised difference vegetation index (NDVI) is a metric that stores information pertaining to the degree of vegetation. Due to its characteristics, it is used in many research activities and remote-sensing (RS) measurements (Deering 1978, Jackson, Huete 1991). The main feature of the NDVI is that it directly depends on the quantity and quality of the vegetation in the selected area (Jarocińska, Zagajewski 2008). NDVI is widely recognised as a metric of vegetation vitality. It has a wide range of applications from vegetation cover segmentation (Tomaszewska et al. 2011), biomass estimation (Tucker 1979), crop maturity classification (Hatfield, Prueger 2010, Tuszynska et al. 2018), water stress detection (Hunt, Rock 1989, Jackson et al. 2004; Gu et al. 2007), nitrogen content indication (Bagheri et al. 2013), chlorophyll content estimation and yield

estimation (Turlej 2009, Sultana et al. 2014), to detection of disease and the effects of pest infestation (Chew et al. 2014) 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 (Nowak, Greenfield 2012, Kuang, Dou 2020, Zięba-Kulawik et al. 2021), including the increasingly important effect of cooling the city through shading (Li, Ratti 2018, McPherson et al. 2018).

The NDVI can be used in analysing 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 (Rouse et al. 1973):

$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad (1)$$

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 utilised in recent research projects related to urban areas conducted in Poland.

In RS, 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. (1973). The textural features concept is described as the spatial relationships between pixels of a given spectral band or several bands that enable discrimination between objects represented on the analysed image (Li et al. 2009). Texture parameters form the basis of the object-based image analysis (OBIA). As practice has shown, utilising textural features in different forms helps in accomplishing various tasks in the field of photogrammetry and RS. When combined with information acquired by calculating NDVI, textural features can serve as input for methods capable of detecting and segmenting vegetation in cities (Zhang 2001, Herold et al. 2003, Li et al. 2009, Marmol, Lenda 2010, Barley, Town 2014, Pyra, Adamczyk 2018).

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, i.e. greyscale 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

Table 1. Normalised difference vegetation index 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)
0.1	MODIS, GSD 250 m, 3 Jul. 2006	Warsaw	Tomaszewska et al. (2011)
0.1	Digital orthophoto, GSD 0.1 m, May 2014	Wroclaw	Kubalska and Preuss (2014)
0.2	IKONOS-2, GSD 1(4) m, 18 Aug. 2005	Lublin	Krukowski et al. (2016)
0.2	Landsat 8, GSD 30 m, 3 Jul. 2015	Łódź	Będkowski and Bielecki (2017)
0.1	Pléiades 1A, GSD 0.5 m, May 2012	Warsaw	Pyra and Adamczyk (2018)
0.1	CIR-orthophoto, GSD 0.25 m, 2015	Łódź	Pluto-Kossakowska et al. (2018)
0.2	IKONOS-2, GSD 1(4) m, 18 Aug. 2011	Lublin	Krukowski (2018)
0.1	CIR aerial orthophoto, GSD 0.25 m, 2015	Łódź	Worm et al. (2019)
0.6	Sentinel 2, GSD 10 m, summer 2018, 2019	Poland	Łachowski and Łęczek (2020)
0.2	IKONOS-2, June 2005, GSD = 0.8 m PAN (3.2 m MS) QuickBird-2, September 2006, GSD = 0.6 m PAN (2.4 m MS) WorldView-2, October 2014, GSD = 0.5 m PAN (2.0 m MS) Aerial orthophotomap (CIR), May 2017, GSD = 0.25 m	Poland	Zięba-Kulawik and Wężyk (2022)

the available bands. Recent research activities focus on utilising 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 ground truth output. Once learned, the conversion 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 convolution neural networks (CNN)-based regressor. Without the support of the discriminator mechanism offered by GAN networks, 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 the input image to the output image but also learns a loss function to train this mapping (Isola et al. 2017).

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. (2017) proposed using 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 (280,000) of pairs of patches (64 px × 64 px) 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 calculated using RGB and CIR-orthophoto images and NDVI synthesised. Suárez et al. (2019) proposed GAN for NDVI estimation based on grayscale images made from three-band RGB images. Importantly, RGB/CIR images that were not registered were not directly used in the training process. A grayscale image derived from RGB was used. 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 expressed in Eq. (1). As stated, the proposed model effectively translates the images from the visible spectrum to NIR and NDVI. Aslahishahri et al. (2021) used the 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 px × 256 px patches of RGB images and their corresponding NIR images acquired using a unmanned aerial vehicle (UAV). The trained network was able to synthesise 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 analyses of panchromatic orthoimagery by skipping the intermediate NIR generation step. The 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 some limited 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, such as using archived aerial imagery — that was acquired with the aid of technology prevalent in the pre-digital era — in RS.

Generative adversarial network

Generative adversarial networks (GANs) are neural networks characterised by a unique construction. GAN is composed of at least a single generator and accompanying discriminator. The role of the generator is to synthesise realistic results, while the discriminator must assess how close the generator output is to reality. By applying the discriminator judgement, 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 mini-max game, in which the generator tries to fool the discriminator (Salimans et al. 2016). 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 (Donahue, Simonyan 2019).

Exploiting the results of the duel between different GAN components opens interesting opportunities both for computer vision (CV) and 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 (Demir, Unal 2018, Dong et al. 2019). 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 (Adamiak et al. 2021). 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 cGAN (Mirza, Osindero 2014). 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 (Isola et al. 2018).

Materials and methods

Research area

Łódź is a city located in central Poland, on the Łódź Upland, at an altitude of 162–278 m above sea level (Fig. 1). It is the third-largest city in the
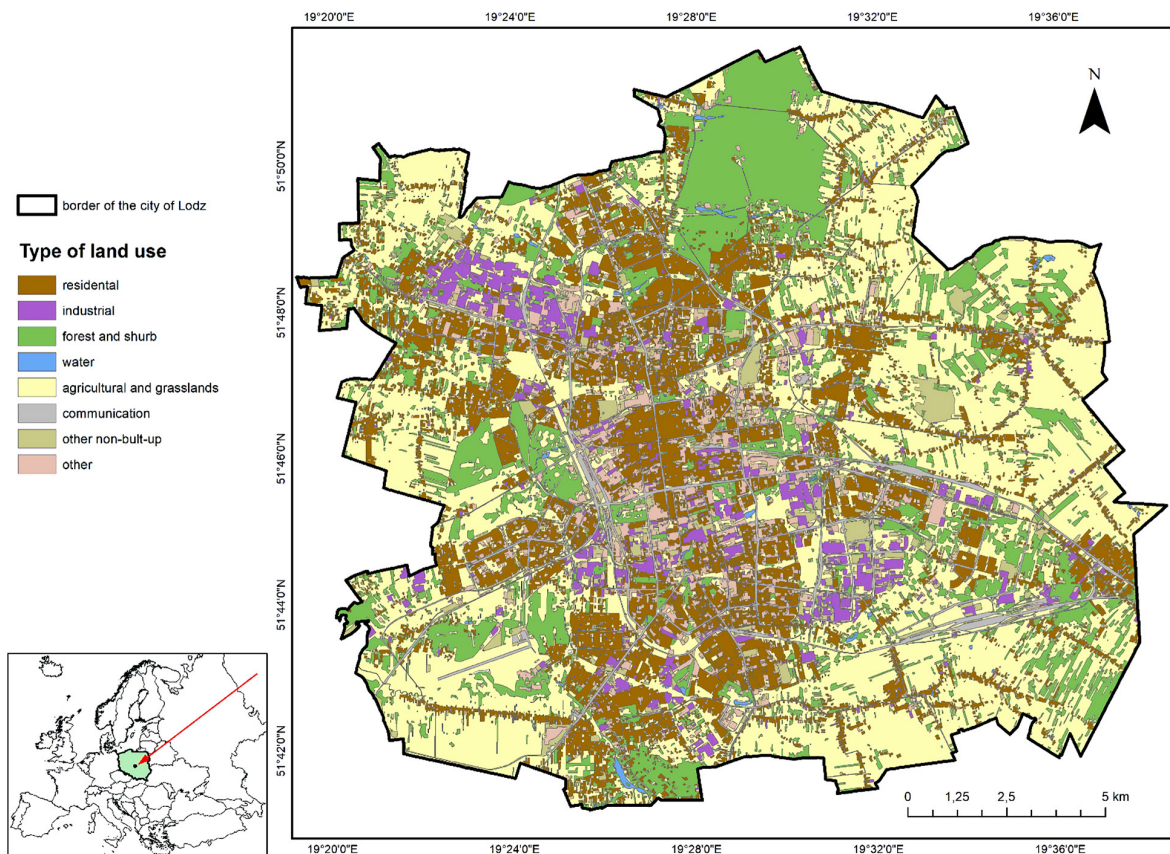

Fig. 1. Research area, based on the data from the Head Office of Geodesy and Cartography (Land and Building Records).

country in terms of population, with approximately 668,000 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 woollen, linen and cotton products was started. At that time, a characteristic strip-like spatial form of the urban layout of Łódź was formed (Barwiński 2009). The years 1860–1890 were the period of the most intensive industrial development. The city became the largest centre 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.

The city is situated between watersheds of the Vistula and Oder rivers. Although 18 rivers and streams flow within the city, none 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 and 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, amounting to 8.9% of the city area (Statistics Poland 2020).

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). Among the socio-economic changes that have played a significant role in the determination of the present-day (after 1989) appearance of the city, the most important include the consequences of the collapse of many textile industry plants,

impoverishment of the population and the degradation of a large part of the city's housing, industrial and communication infrastructure. The revitalisation process is underway, and new housing estates and industrial buildings appear. The former 19th-century industrial complexes receive new residential, office and commercial functions. Nevertheless, even in areas close to the city centre, many degraded areas are still subject to spontaneous vegetation succession, i.e. land abandonment.

Methods

The investigation was begun by collecting the sheets of orthophotomap in four bands (red, green, blue and near-infrared) covering 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 (grayscale) 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 on the test orthophotomap. A graphic overview is shown in Figure 2.
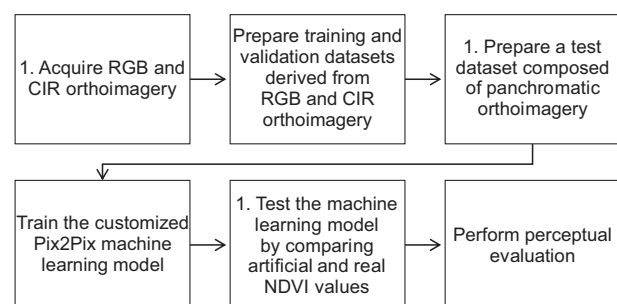

Fig. 2. The main steps of the investigation.

Dataset

All data have been downloaded from Geoportal (2021). The dataset included the orthophotomaps in two different compositions: RGB (bands: red, green and blue) and CIR (near-infrared, red and

Table 2. Structure of land use in Łódź [km^2].

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

Source: Statistics Poland (2020).

green). The Vexel Imaging UltraCam Eagle M3 (serial: 431S61782X117253-f100) camera model was used. The raids were made on 11 April 2021. The data download process has been carried out by using a plug-in for the Q-GIS software, GUGiK's data downloader (EnviroSolutions 2021). 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 have a ground sampling distance (GSD) (pixel size) of 5 cm and in total, they occupy 140 GB of disk space.

To compute a four-band orthophotomap (RGB + NIR), ESRI (Environmental Systems Research Institute) ArcGIS Desktop 10.5.1 Composite Bands were 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 resulted in 642,582 patches in total: 493,290 with 512 px, 120,582 with 1024 px and 28,710 with 2048 px edge lengths. Secondly, each patch was divided into two separate items.

The patch designed to be utilised as input during neural network training contained grayscale images derived as a simple sum of RGB bands stretched linearly into 256 grey levels, and can be defined by the following formula:

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

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

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 [256, 256, 1] target tensors. After the post-processing step had been completed, five patches representing different spatial features were randomly picked from each processed orthophoto sheet to form a 1305-item validation subset. Importantly, 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 needed to be rendered into a form that would demonstrate persistence across time, and for this purpose, NumPy archives in the NPZ (numpy zip) format ('NumPy documentation') were used. Dataset preparation was performed using Python programming language and machine learning and CV libraries such as scikit-learn (scikit-learn: machine learning in Python – scikit-learn 1.0.2 documentation b.d.), scikit-image (van der Walt et al. 2014), OpenCV (OpenCV b.d.) and TensorFlow Datasets (TensorFlow Datasets b.d.) (Fig. 3).

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 (Jung 2022).

The test dataset was prepared using RGB orthoimagery acquired in 2012. Test orthophoto patches did not overlap with training and
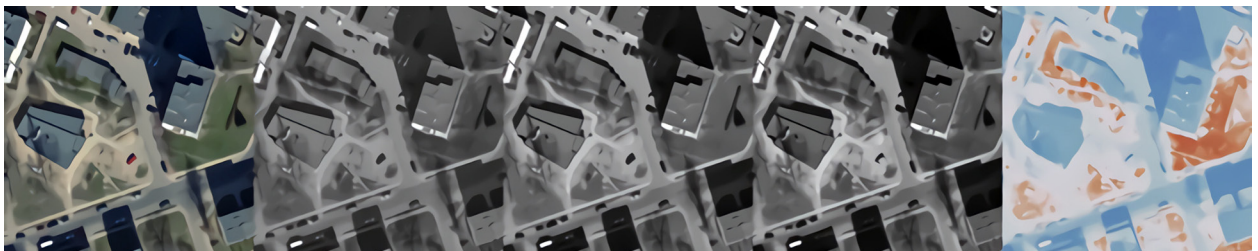

Fig. 3. Visualization Visualisation of a dataset item, based on Geoportal (2021). From left to right: RGB composition, R, G and B bands that form the input tensor and normalised difference vegetation index, which serves as a target tensor.

validation sets. Test items were similarly processed as those in the training set, excluding augmentation.

Evaluation metrics

Due to the high variety of content presented in RS imagery and the undoubted complexity of the natural image synthetisation process, GAN's artificial image quality assessment is considered an extremely complicated task. Relying only on the quality of colour reproduction would be inadequate, and thus there is also a need to pay attention to the structure and feature preservation, as well as evaluate the level of realism and diversity of acquired results. Additionally, when working with patches computed from a larger image, a confirmation is needed as to whether a reassembly yields a uniformly correct output. There are many reliable metrics available. In themselves, none of these metrics, considered alone, provides enough information to conduct a comprehensive assessment. Therefore, a combination of them is needed. The generative adversarial NDVI synthetisation process has been monitored using the following metrics: structural similarity index measure (SSIM), root mean square error (RMSE), peak signal-to-noise ratio (PSNR) and human eYe perceptual evaluation (HYPE). Implementation of SSIM, RMSE and PSNR algorithms is available in the Python library image-similarity-measures (Müller et al. 2020). Metrics were calculated by directly comparing values produced by the generator network with a sigmoid activation function output (NDVI_artificial) with ground-truth images (NDVI_real).

RMSE

RMSE is a standard statistical metric in such areas as meteorology, air quality and climate research, but its most popular application is in the field of geoscience because it gives more weight to errors with larger absolute values than errors with smaller absolute values (Chai, Draxler 2014).

The RMSE 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 in the range 0–1. The indicator formula is as follows:

$$RMSE(y, \hat{y}) = \sqrt{\text{MSE}(y, \hat{y})} \qquad (3)$$

where   represents predicted values,   observed values and n the number of observations.

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 (Wang et al. 2004). 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 0 and 1, where 0 indicates dissimilarity and 1 similarity. Its use in orthoimagery analysis is justified because by imitating human perception, it can recognise patterns and thus measure the quality of spatial features' reproduction.

$$SSIM(y, \hat{y}) = \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)} \qquad (4)$$

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

PSNR

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. It is worth mentioning that PSNR performs poorly in discriminating structural content in images since various types of degradations applied to the same image can yield the same value of the MSE (Horé, Ziou 2010). At the same time, PSNR can assess the results of image transformations that tend to apply different noise levels to their output.

$$PSNR(y, \hat{y}) = 10 \, log_{10}\left(\frac{255^2}{MSE(y, \hat{y})}\right) \qquad (5)$$

Perceptual evaluation

The 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 artificial and real images and effects of the hyper-realistic generation when fake images look even more realistic than real images (Zhou et al. 2019). HYPE$_\infty$ is calculated as a proportion of images judged incorrectly and aggregated the judgements over the n evaluators on k images to produce the final score. HYPE$_\infty$ values range from 0 to 1, where 0 indicates low synthetisation 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 RS were recruited to participate in the study to perceptually evaluate synthetic NDVI layers. Due to the complexity of the task, experts were granted unlimited evaluation time. Experts working independently visually inspected 155 and 101 randomly selected image sets, variously 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.

$$NDVI_{diff} = NDVI_{true} - NDVI_{artificial} \qquad (6)$$

For the purpose of the study and after performing extensive experiments with different parameters, loss functions and hyperparameters tuning, the baseline TensorFlow architecture (TensorFlow 2022) has been utilised with the following modifications:

Depth-wise separable convolution layers (Chollet 2017) were applied instead of default convolution layers to reduce the resource requirements of the model.

The number of filters in each convolution layer was multiplied by a factor of 2 to increase network capacity. Originally, the first convolutional downsampling block started with 64 filters. In the utilised approach, 128 initial filters were used.

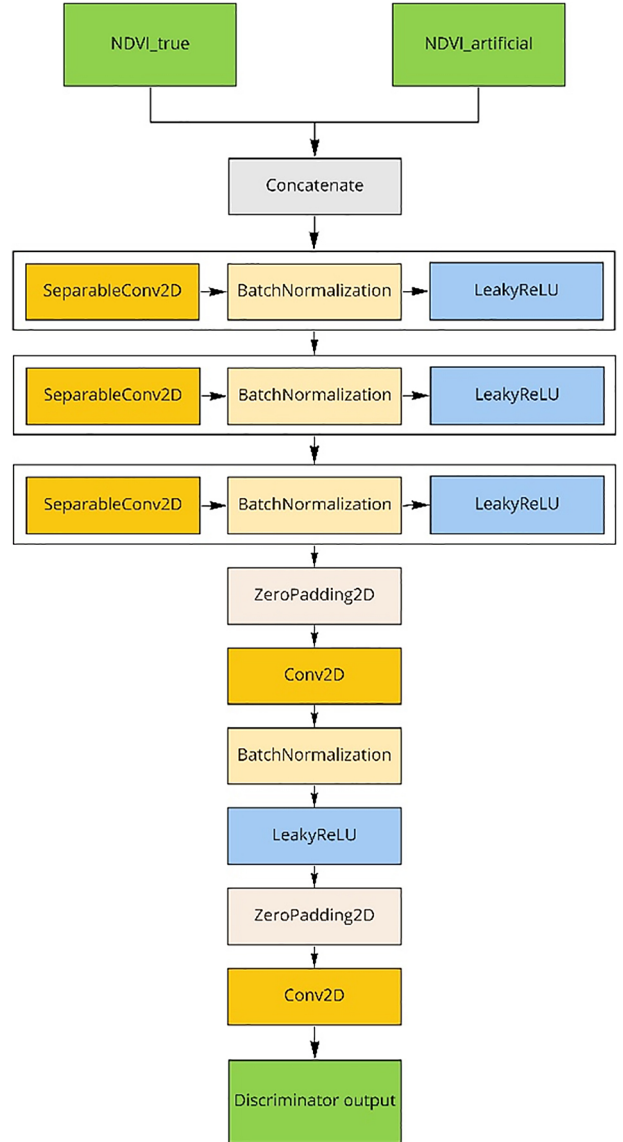Modifying the generator loss function by enhancing the existing implementation comprised

Fig. 4. Modified Pix2Pix discriminator model visualisation.

applying 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 (Isola et al. 2018). It is as follows:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x,y)] + \\ \mathbb{E}_{x,}[\log(1 - D(x, G(x)))] \qquad (7)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y}[\|y - G(x)\|_1] \qquad (8)$$

$$\mathcal{L}_{SSIM}(G) = \mathbb{E}_{x,y}[1 - SSIM(y, G(x))] \qquad (9)$$

$$G^* = \arg \min_{D} \max_{G} \mathcal{L}_{cGAN}(G, D) +$$
$$\lambda \mathcal{L}_{L1}(G) + \omega \mathcal{L}_{SSIM}(G) \qquad (10)$$

where D represents the distriminator, G the generator, x the input PAN image, y the NDVI$_{true}$ and G(x) the calculated NDVI$_{artificial}$.

The model training workflow is as follows. The input grayscale image (256 px × 256 px × 1 band) is fed to the generator. The generator produces a synthesised NDVI composition image (256 px × 256 px × 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 calculated to consider both the quality of the synthesised output (MAE + SSIM) and the discriminator verdict (BCE). The training step ends when both generator and discriminator gradients are applied.

Results

In the interpretation process, the main forms of land cover were identified with which the
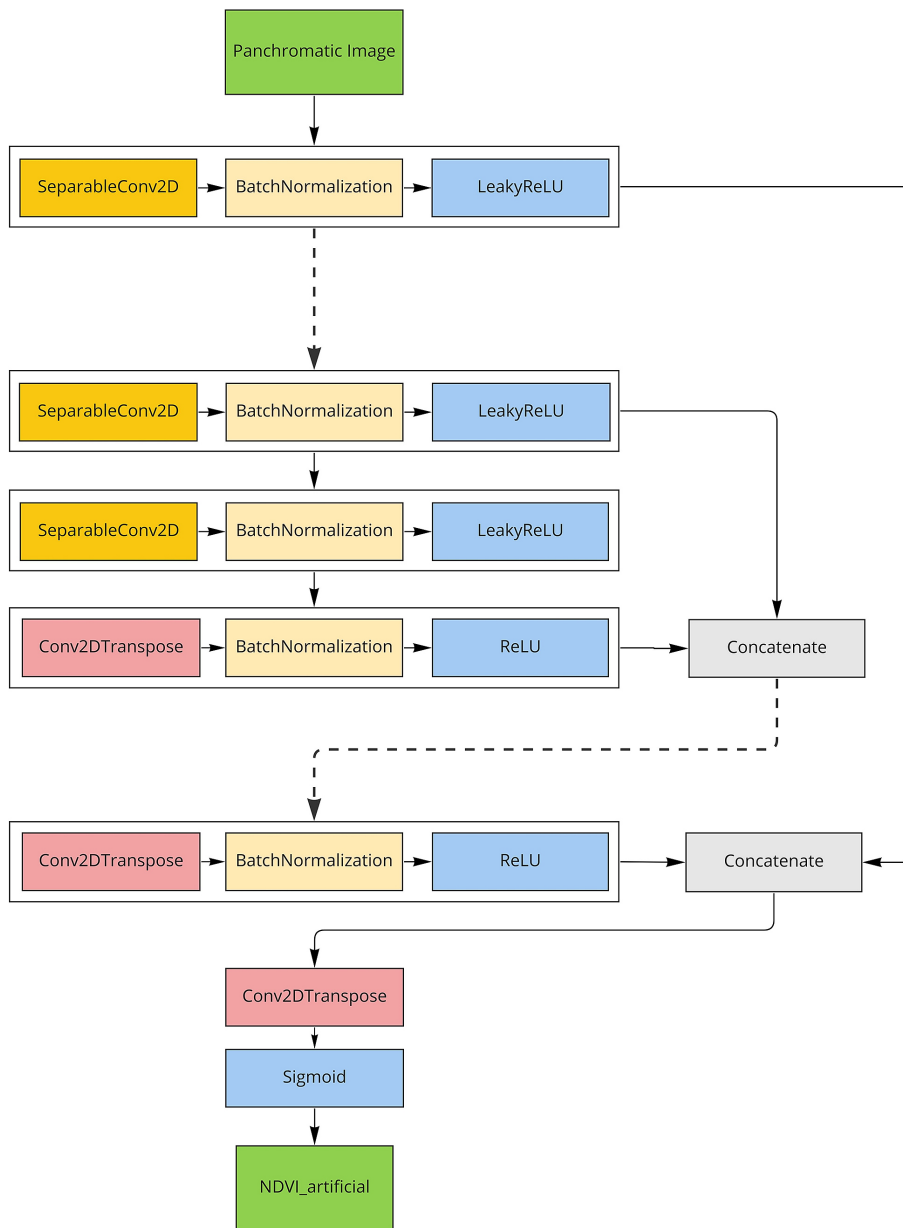

Fig 5. Modified Pix2Pix generator sub-model visualisation.

visible regularities of the algorithms' behaviour can be associated, i.e. using which it can be ascertained 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 utilised colour map to produce the NDVI$_{true}$, NDVI$_{artificial}$ and NDVI$_{diff}$ images.

In the first analysed area, a single row of low trees is visible (Fig. 7). On the NDVI$_{artificial}$ image, the line of trees has lost its rectilinear shape. NDVI$_{true}$ and NDVI$_{artificial}$ are comparable 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, which should be considered good results.

The area presented in Figure 8 is entirely wooded. NDVI$_{true}$ has exposed only a few trees and one group, whereas the remainder 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, which should be considered mediocre results.

Figure 9 presents a parking lot with passenger cars. Most of the cars are properly handed over, and one is underestimated (its shadow – falls on the surface of the yard – overestimated), whereas 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.

Big buildings with vast shadows are shown in Figure 10. In the shadows, a warehouse of pallets and building materials can be observed; however,
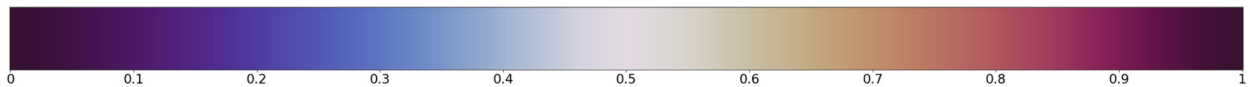

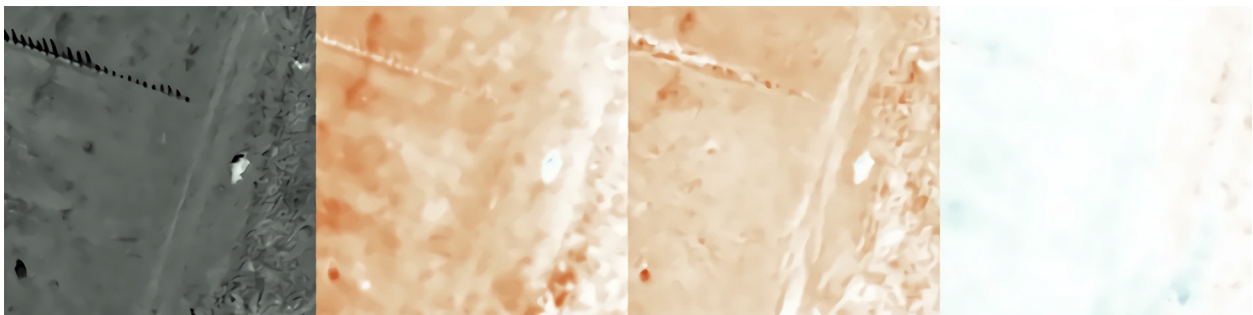Fig. 6. Colour palette used during visual inspection of Figs 7–15.

Fig. 7. Inference result – abandoned land; based on Geoportal (2021). From left to right: PAN, NDVI$_{true}$, NDVI$_{artificial}$ and NDVI$_{diff}$.
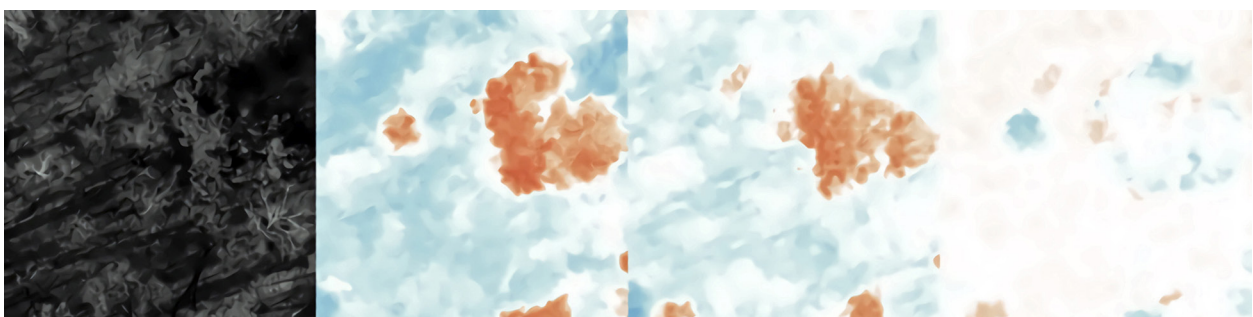

Fig. 8. Inference result – wooded area: PAN, NDVI$_{true}$, NDVI$_{artificial}$ and NDVI$_{diff}$; based on Geoportal (2021).

in places of greater chaos, such textures on the $NDVI_{artificial}$ image have been incorrectly marked as greenery. An increase of the $NDVI_{artificial}$ (correctly) was identified on lawns in the lit zone. Roofs and roads manifested with no major differences between $NDVI_{artificial}$ and $NDVI_{true}$. The evaluation of the sample yielded 0.7859 SSIM and 0.0689 RMSE.

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 (~90%) stand out as NDVI compliant. Other roofs may have inflated $NDVI_{artificial}$ in places where slightly blurred edges are observed (e.g. in the left corner of the image).

Underruns occur in the shadows cast by roofs. In the centre 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 geometrised plant shapes are recognised as inanimate objects, while roofs with blurred edges as lawns. The evaluation of the sample yielded 0.5651 SSIM and 0.0736 RMSE.

The next figure (Fig. 12) shows a large-scale commercial facility (40%), parking lot (25%) and grass. The biggest mistake obtained as output from the NDVIartificial revaluation is that a long row of skylights on the roof of the building has been incorrectly recognised, possibly as a line of trees. On the same roof, next to it, there are
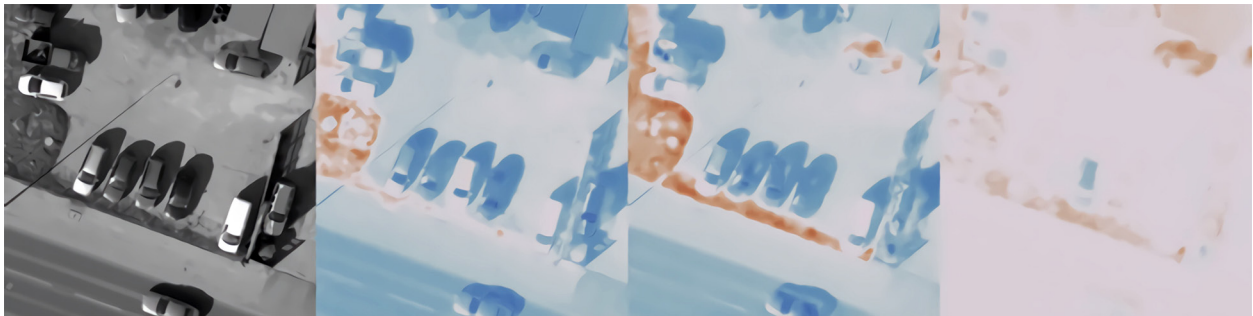

Fig. 9. Inference result – small parking lot: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; based on Geoportal (2021).

Fig. 10. Inference result – buildings: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; based on Geoportal (2021).
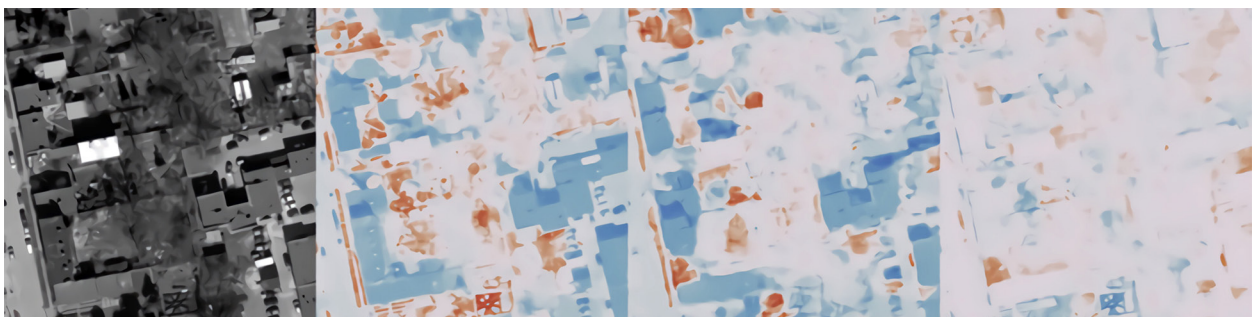

Fig. 11. Inference result – residential area: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; based on Geoportal (2021).

irregularly placed ventilators, and 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.

The whole figure (Fig. 13) presents an empty parking lot of a shopping centre 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 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.

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,
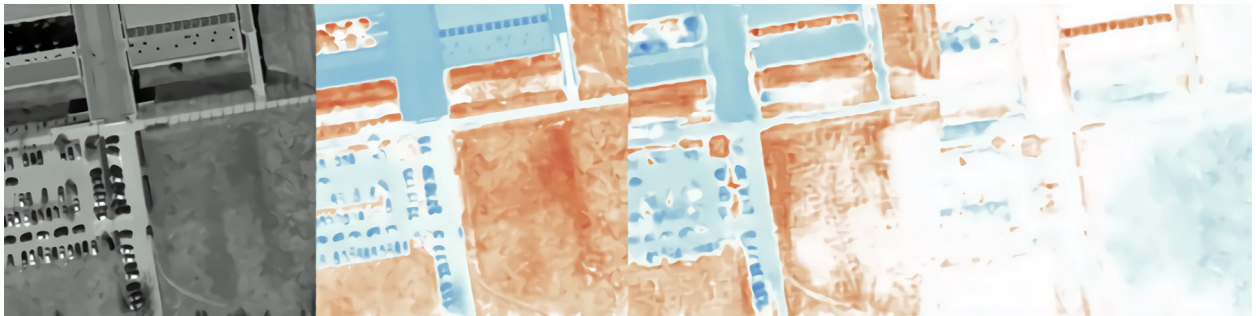

Fig. 12. Inference result – commercial facility: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; based on Geoportal (2021).
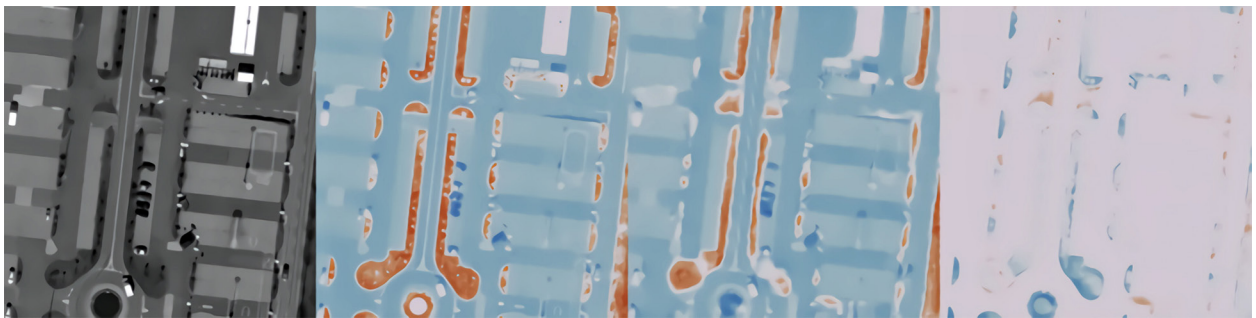

Fig. 13. Inference result – commercial facility parking lot: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; based on Geoportal (2021).
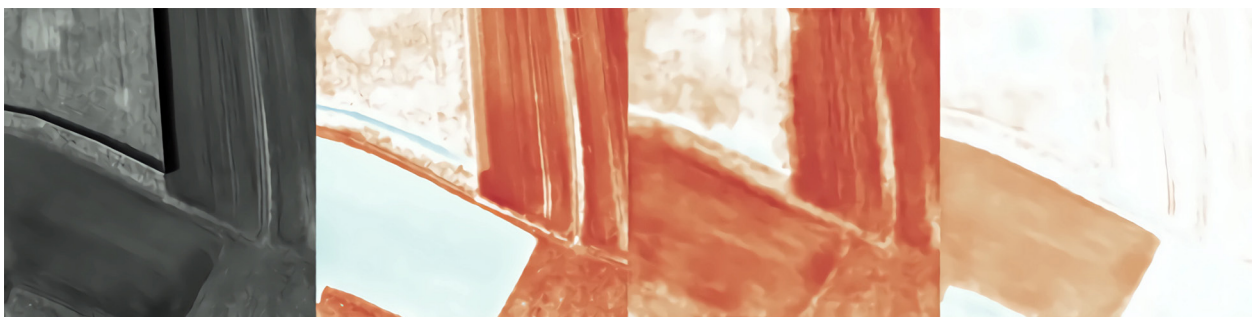

Fig. 14. Inference result –farmland: PAN, $NDVI_{true}$, $NDVI_{artificial}$ and $NDVI_{diff}$; based on Geoportal (2021).

clear edges, generally had values lower than in the NDVI$_{true}$ image, i.e. they were recognised as surfaces of roofs or paved parking lots or roads. In addition, lawns that were sunlit, and accordingly registered brighter in the panchromatic image, erroneously obtained higher results than lawns that were in the shade and consequently darker. 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 or less ordered. In such cases, they were recognised as areas covered with vegetation.

An interesting observation is that NDVI$_{artificial}$ does not indicate overestimated values for ima-

ges 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. Owing to variances in the level of insolation, corresponding variations could also be present in the index values assigned to tree-branches, which is evidenced by the observation that branch fragments on the sunny side generally received lower values.

NDVIartificial images do not accurately reproduce the textural information present in the original RGB, panchromatic and NDVItrue renderings. For instance, drawings of roof tiles, parking lot pavement tiles and the structure of tree branches do not capture the detailed patterns of surface-markings naturally seen in these objects. Additionally, deformation was observed in the rendering of linear green systems, as evidenced by the fact that tree rows, perfectly straight in the field, have been depicted as skewed in the synthesised 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 or 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 synthetisation process.

Discussion

The performed experiment showed that the structure and texture of the panchromatic aerial RS image contain information that allows us to estimate the NDVI value for various objects of urban space. However, 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 utilising the sliding window inference approach.

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

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 ploughing 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,
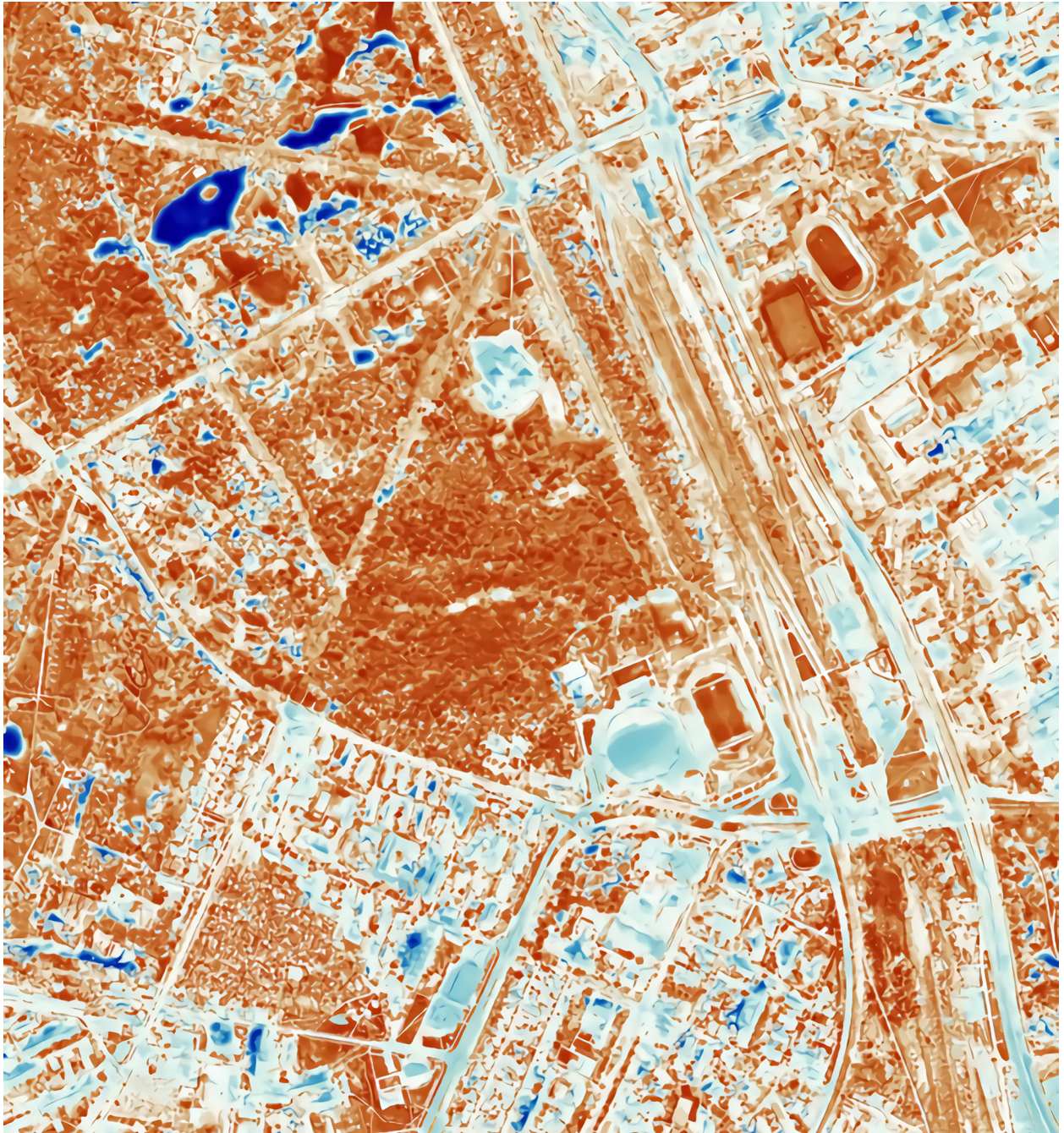

Fig. 15. Sliding window inference (NDVI$_{artificial}$) of an orthophoto used to compute the test dataset); based on Geoportal (2021). The scene (51.76174 E, 19.42149 N) presents Łódź, Poland. Red values indicate high NDVI values (closer to 1). Blue ones represent small values (closer to −1).

which should be considered in the case of images with low resolution (Small 2001), and each of the pixels represents reflectance values typical of urban land cover forms, or actually, materials present in this space.

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 are from the early spring period, and thus 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 could appropriately assign values to the NDVI$_{artificial}$ image, similar to those it learned to recognise on the NDVI$_{true}$ ground truth.

At this point, the question arises as to 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 NDVI$_{true}$ generated from the summer image when the vegetation is fully developed. Could we perhaps get NDVI$_{artificial}$ values that describe vegetation better? From this perspective, spring images (originally taken as panchromatic or linearly synthesised from RGB channels) could be used to create images showing the predicted NDVI distribution in the full growing season. The evident practical significance of such an image (which we refer to as the approximated summer NDVI) may be attributable to the fact that the present authors encountered circumstances that permitted them 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 and apply a trained algorithm (here, photos from summer on one occasion would be required), and creation of NDVI$_{artificial}$ would be feasible every spring.

One more aspect needs to be considered, namely the classification of objects, mainly vegetation, which is typical for RS. 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. The accuracy of the classification can be increased 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 data from spring images allows deep penetration of visibility into the tree crowns, thus showing 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 synthesised from RGB + NIR channels, have a good spatial resolution, better than each of these channels considered separately.

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 centre 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 centre 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 result in the pictures being characterised by 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 NDVI$_{artificial}$ 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 NDVI$_{artificial}$ 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 RS. Then, finding corresponding fragments of two images would not be 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 (Koza 2006, Davis, Wang 2011) and further to detect changes, e.g. as a result of the development of plant cover or movements (Dematteis, Giordan 2021). 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 (Verykokou, Ioannidis 2019), or pairs of images even with significant differences in viewpoints (Yao et al. 2021). 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 (Haralick et al. 1973), 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, may be the inclusion in the study of orthoimages with the radial displacement effect entirely removed, i.e. true orthophoto. For their production, normalised 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 for information about the topographic and spatial positions of objects whose location is discerned through the use of LiDAR laser scanning technology. The nDSM geometry is correct – the projection is orthogonal, and there are no displacements. However, there is some loss of photographic image quality due to pixel resampling. Even so, it is certain 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 and 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. (2003), 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 and Adamczyk [2018]), but considerable spectral similarities were also observed between the distinguished land cover classes. Multiple attempts to select and apply textures, masks and majority filters allowed the accuracy of the classification to be increased. We believe that it is possible to conclude that our algorithm has output good results only if the condition is satisfied that, when digitally expressed, the structure and texture of the represented objects should be visible even when they are enveloped in shadows, and even when they are often invisible to the human eye; these circumstances usually pertain to dense, layered vegetation.

The obtained results were also influenced by a specific property of the textural features, namely the sensitivity to changes in some image parameters. As reported by Książek (2018), changes in contrast, brightness, sharpness and image rotation may impact several groups of textural features of roofs selected by her for research (asbestos roofs were dealt with). The obtained results were also analysed 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 manner similar to the one mentioned above, modified.

Estimating the NDVI based on archival panchromatic images makes it possible to study the conditions characterising, and changes in, urban vegetation cover in the long term. The research may primarily concern changes in the amount of greenery resulting from the expansion of vegetation cover due to the economic activity of municipal authorities and natural growth. Still, it also helps reveal places of excessive or illegal

tree removal. Figure 16 presents an example of utilising the model architecture to predict NDVI values of an aerial image acquired in 1966.

It is hard to expect that the generated $NDVI_{artificial}$ images will fulfil all the functions that NDVI applies to in modern scientific research and practice. We expect that the differences between the synthesised 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 researchers discuss this issue (Li, Ratti 2018, McPherson et al. 2018). However, it may be problematic to try to classify vegetation even into large groups (e.g. coniferous – deciduous), and subtle tasks—such as assessing the health condition and water stress of vegetation, crop maturity, nutrient content (e.g. nitrogen,
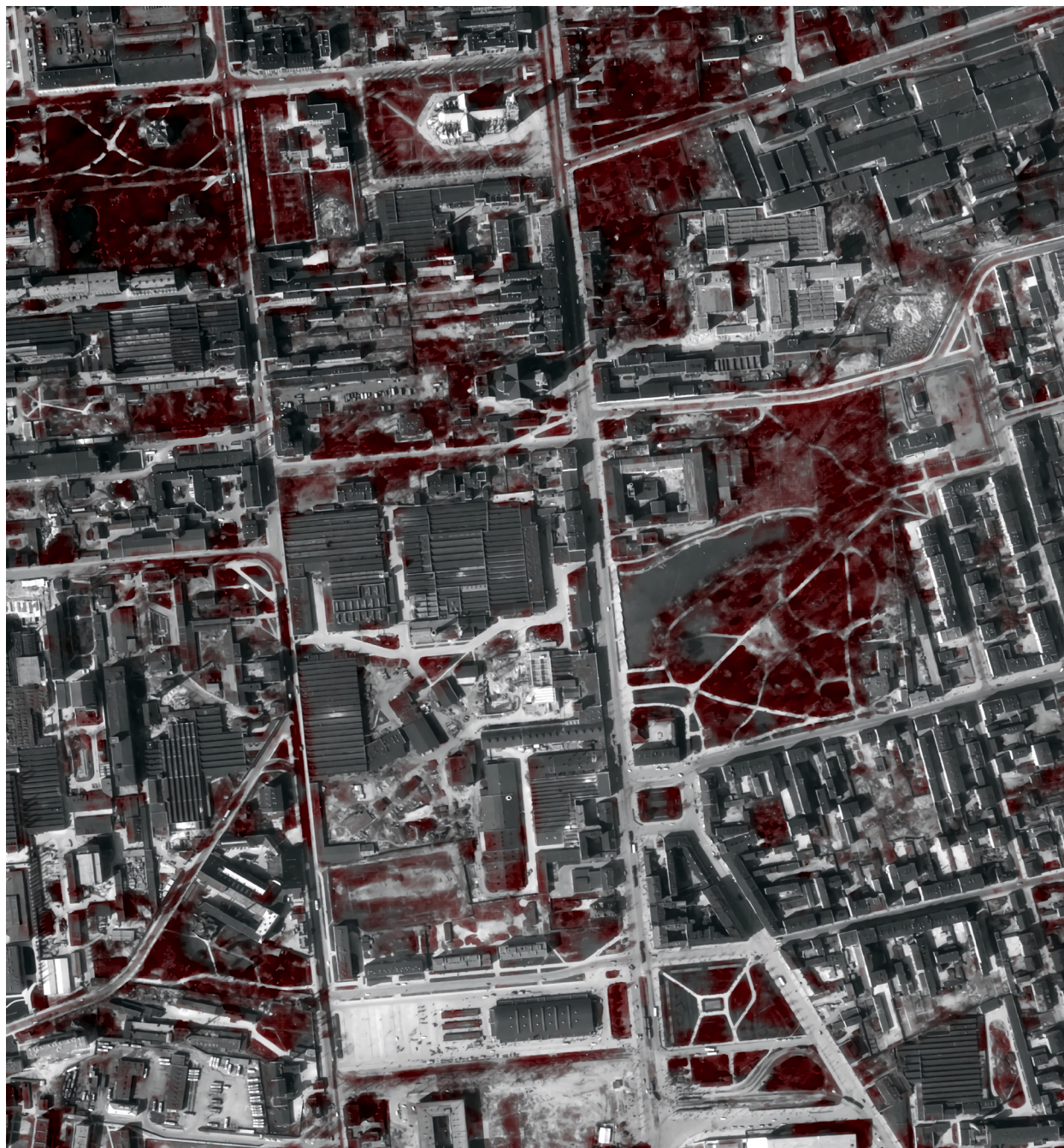

Fig. 16. Sliding window inference ($NDVI_{artificial}$) of an archival 1966 greyscale aerial image; based on GUGiK (Head Office of Geodesy and Cartography b.d.). The scene presents Łódź, Poland. Red overlay indicates values where $0.5 < NDVI < 1$.

chlorophyll content) and estimated yields, as well as detection of disease or insect infestation—would almost certainly pose a major challenge.

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

Conclusions

This research discussed and confirmed the potential of utilising cGAN in the performance of an image-to-image translation capable of estimating NDVI from a single panchromatic orthoimagery. The customised 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 PSNR 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 utilised to distinguish areas of different vegetation volumes. Although the predicted NDVI values in pixel scope are not precise enough to utilise the model as a replacement for traditionally computed NDVI, it can be used in a variety of avenues of application, from archival RGB and panchromatic imagery processing to serving as a data augmentation or an inpainting technique.

Author's contribution

Conceptualisation: M.A., K.B., A.B.; Methodology: M.A.; Software: M.A.; Validation: M.A.,

Data availability statement

The datasets analysed in this study are publicly available at https://www.geoportal.gov.pl/

Conflicts of interest

The authors declare no conflict of interest.

References

Adamiak M., Będkowski K., Majchrowska A., 2021. Aerial imagery feature engineering using bidirectional generative adversarial networks: A case study of the Pilica River Region, Poland. *Remote Sensing* 13(2): 306. DOI 10.3390/rs13020306.

Aslahishahri M., Stanley K.G., Duddu H., Shirtliffe S., Vail S., Bett K., Pozniak C., Stavness I., 2021. From RGB to NIR: Predicting of near infrared reflectance from visible spectrum aerial images of crops. In: *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 1312–1322. DOI 10.1109/ICCVW54120.2021.00152.

Bagheri N., Ahmadi H., Alavi Panah S., Omid M., 2013. Multispectral remote sensing for site-specific nitrogen fertilizer management. *Pesquisa Agropecuária Brasileira* 48: 1394–1401. DOI 10.1590/S0100-204×2013001000011.

Barley A., Town C., 2014. Combinations of feature descriptors for texture image classification. *Journal of Data Analysis and Information Processing* 2(3): 67–76. DOI 10.4236/jdaip.2014.23009.

Barwiński M., 2009. Spatial development and functional changes in Łódź – Geographic, economic and political conditions. *Geografia w szkole* 6: 38–50.

Będkowski K., Bielecki A., 2017. Assessment of the availability of greenery in the place of residence in cities using NDVI and the Lorenz's concentration curve. *Teledetekcja Środowiska* 57: 5–14.

Chai T., Draxler R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development* 7(3): 1247–1250. DOI 10.5194/gmd-7-1247-2014.

Chew W.C., Hashim M., Lau A.M.S., Battay A.E., Kang C.S., 2014. Early detection of plant disease using close range sensing system for input into digital earth environment. *IOP Conference Series: Earth and Environmental Science* 18: 012143. DOI 10.1088/1755-1315/18/1/012143.

Chollet F., 2017. Xception: Deep learning with depthwise separable convolutions. *arXiv:1610.02357 [cs]*, April. Online: http://arxiv.org/abs/1610.02357.

Davis C.H., Wang X., 2011. High-resolution DEMS for urban applications from NAPP photography. *Photogrammetric Engineering and Remote Sensing* 67: 4–11.

Deering D., 1978. Rangeland reflectance characteristics measured by aircraft and spacecraft sensors. Thesis, Texas A&M University. Libraries. Online: https://oaktrust.library.tamu.edu/handle/1969.1/DISSERTATIONS-253780.

Dematteis N., Giordan D., 2021. Comparison of digital image correlation methods and the impact of noise in geoscience applications. *Remote Sensing* 13(2): 327. DOI 10.3390/rs13020327.

Demir U., Unal G., 2018. Patch-based image inpainting with generative adversarial networks. *arXiv:1803.07422 [cs]*. Online: http://arxiv.org/abs/1803.07422.

Donahue J., Simonyan K., 2019. Large scale adversarial representation learning. *arXiv:1907.02544 [cs, stat]*. Online: http://arxiv.org/abs/1907.02544.

Dong J., Yin R., Sun X., Li Q., Yang Y., Qin X., 2019. Inpainting of remote sensing SST images with deep convolutional generative adversarial network. *IEEE Geoscience and Remote Sensing Letters* 16(2): 173–177. DOI 10.1109/LGRS.2018.2870880.

EnviroSolutions Sp. z o.o. – Michał Włoga., 2021. Pobieracz danych GUGiK. Online: https://plugins.qgis.org/plugins/pobieracz_danych_gugik/.

Geoportal, 2021. Online: http://geoportal.gov.pl.

Gu Y., Brown J., Verdin J., Wardlow B., 2007. 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* 34(6). DOI 10.1029/2006GL029127.

Haralick R.M., Shanmugam K., Dinstein I., 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics* SMC 3(6): 610–621. DOI 10.1109/TSMC.1973.4309314.

Hatfield J., Prueger J., 2010. Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices. *Remote Sensing* 2(2): 562–578. DOI 10.3390/rs2020562.

Head Office of Geodesy and Cartography., b.d. Integrated copies of databases of topographic objects. Główny Urząd Geodezji i Kartografii. Główny Urząd Geodezji i Kartografii. Online: https://www.geoportal.gov.pl/dane/baza-danych-obiektow-topograficznych-bdot (accessed 11 November 2020).

Head Office of Geodesy and Cartography., b.d. Online: https://www.gov.pl/web/gugik-en (accessed 8 August 2022).

Herold M., Liu X., Clarke K., 2003. Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering and Remote Sensing* 69: 991–1001. DOI 10.14358/PERS.69.9.991.

Horé A., Ziou D., 2010. Image quality metrics: PSNR vs. SSIM. In: *2010 20th International Conference on Pattern Recognition*, 2366–2369. DOI 10.1109/ICPR.2010.579.

Hunt E.R., Rock B., 1989. Detection of changes in leaf water content using near – And middle-infrared reflectances. *Remote Sensing of Environment* 30(1): 43–54. DOI 10.1016/0034-4257(89)90046-1.

Isola P., Zhu J-Y., Zhou T., Efros A., 2017. Image-to-image translation with conditional adversarial networks. *arXiv:1611.07004 [cs]*, November 2021. Online: http://arxiv.org/abs/1611.07004.

Jackson R., Huete A., 1991. Interpreting vegetation indices. *Preventive Veterinary Medicine* 11(3): 185–200. DOI 10.1016/S0167-5877(05)80004-2.

Jackson T., Chen M., Cosh M., Li F., Anderson M., Walthall C., Doriaswamy P., Ray Hunt R., 2004. Vegetation water content mapping using landsat data derived normalized difference water index for corn and soybeans. *Remote Sensing of Environment*, 2002 Soil Moisture Experiment (SMEX02), 92(4): 475–482. DOI 10.1016/j.rse.2003.10.021.

Jarocińska A., Zagajewski B., 2008. Correlations of ground – And airborne-level acquired vegetation indices of the Bystrzanka catchment. *Teledetekcja Środowiska* 40: 100–124.

Jung A., 2022. Imgaug. Python. Online: https://github.com/aleju/imgaug.

Koza P., 2006. Orientation of Ikonos stereo images and automatic acquisition of height models. *Archiwum Fotogrametrii, Kartografii i Teledetekcji* 16. Online: http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-3514d2c7-31a9-49d8-ad2d-c35825c950f8.

Krukowski M., 2018. 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* 73: 7–27. DOI 10.17951/b.2018.73.0.7-27.

Krukowski M., Cebrykow P., Płusa J., 2016. Classification of green areas in Lublin based on satellite images Ikonos 2. *Barometr Regionalny* 14(2): 35–44.

Książek, J., 2018. Study of selected textural features properties on asbestos roof images. *Geomatics and Environmental Engineering* 12(4). DOI 10.7494/geom.2018.12.4.45.

Kuang, W., Dou Y., 2020. Investigating the patterns and dynamics of urban green space in China's 70 major cities using satellite remote sensing. *Remote Sensing* 12(12): 1929. DOI 10.3390/rs12121929.

Kubalska J., Preuss R., 2014. Use of the photogrammetric data for vegetation inventory on urban areas. *Archiwum Fotogrametrii, Kartografii i Teledetekcji* 26: 75–86. DOI 10.14681/AFKIT.2014.006.

Łachowski W., Łęczek A., 2020. Tereny zielone w dużych miastach Polski. Analiza z wykorzystaniem Sentinel 2. *Urban Development Issues* 66(1): 77–90. DOI 10.51733/udi.2020.68.07.

Li P., Cheng T., Guo J., 2009. Multivariate image texture by multivariate variogram for multispectral image classification. *Photogrammetric Engineering & Remote Sensing* 75(2): 147–157. DOI 10.14358/PERS.75.2.147.

Li X., Ratti C., 2018. Mapping the spatial distribution of shade provision of street trees in Boston using google street view Panoramas. *Urban Forestry & Urban Greening* 31: 109–119. DOI 10.1016/j.ufug.2018.02.013.

Marmol U., Lenda G., 2010. Texture filters in the process of automatic object classification. *Archiwum Fotogrametrii, Kartografii i Teledetekcji* 21: 235–243.

McPherson G., Xiao Q., van Doorn N., Johnson N., Albers S., Peper P., 2018. Shade factors for 149 taxa of in-leaf urban trees in the USA. *Urban Forestry & Urban Greening* 31: 204–211. DOI 10.1016/j.ufug.2018.03.001.

Mirza M., Osindero S., 2014. Conditional generative adversarial nets. *arXiv:1411.1784 [cs, stat]*. Online: http://arxiv.org/abs/1411.1784.

Müller M., Ekhtiari N., Almeida R., Rieke C., 2020. Super-resolution of multispectral satellite images using convolutional neural networks. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-1–2020 (August): 33–40. DOI 10.5194/isprs-annals-V-1-2020-33-2020.

Myeong S., Nowak D., Hopkins P., Brock R., 2003. Urban cover mapping using digital, high-resolution aerial imagery. *Urban Ecosystems* 5: 243–256. Online: http://www.fs.usda.gov/treesearch/pubs/18820.

Nowak D., Greenfield E., 2012. Tree and impervious cover change in U.S. cities. *Urban Forestry & Urban Greening* 11(1): 21–30. DOI 10.1016/j.ufug.2011.11.005.

NumPy documentation. Online: https://numpy.org/doc/stable/reference/generated/numpy.savez.html (accessed 27 January 2022).

OpenCV., b.d. Online: https://opencv.org/ (accessed 27 January 2022)

Pluto-Kossakowska J., Władyka M., Tulkowska W., 2018. Assessment of remote sensing image data to identify objects in green and blue infrastructure. *Teledetekcja Środowiska* T 59. Online: http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-9632f302-e255-497e-a9dd-368ea620f9b4.

Pyra M., Adamczyk J., 2018. Object-oriented classification in the inventory of green infrastructure objects on the example of the Ursynów District in Warsaw. *Teledetekcja Środowiska* T. 59. Online: http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-8bd759f8-2ab3-4b35-946d-b34b73f28b88.

Rouse J.W., Jr., Haas R.H., Schell J.A., Deering D.W., 1973. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Texas A&M Univ. College Station, TX, United States.

Salimans T., Goodfellow I., Zaremba W., Cheung V., Radford A., Chen X., 2016. Improved techniques for training GANs. *arXiv:1606.03498 [cs]*. Online: http://arxiv.org/abs/1606.03498.

Scikit-learn: Machine learning in Python – Scikit-learn 1.0.2 documentation., b.d. Online: https://scikit-learn.org/stable/ (accessed 27 January 2022)

Small, C., 2001. Estimation of urban vegetation abundance by spectral mixture analysis. *International Journal of Remote Sensing* 22(7): 1305–1334. DOI 10.1080/01431160151144369.

Statistics Poland., 2020. Statistics of Łódź 2020. Lodz.Stat.Gov.Pl. Online: https://lodz.stat.gov.pl/en/publications/statistical-yearbook/statistics-of-lodz-2020,1,16.html.

Suarez P., Sappa A., Vintimilla B., 2017. Learning image vegetation index through a conditional generative adversarial network. In: *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, 1–6. DOI 10.1109/ETCM.2017.8247538.

Suárez P., Sappa A., Vintimilla B., Hammoud R., 2019. Image vegetation index through a cycle generative adversarial network. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1014–1021. DOI 10.1109/CVPRW.2019.00133.

Sultana S., Ali A., Ahmad A., Mubeen M., Zia-Ul-Haq M., Ahmad S., Ercisli S., Jaafar H., 2014. Normalized difference vegetation index as a tool for wheat yield estimation: A case study from Faisalabad, Pakistan. *The Scientific World Journal* 2014: e725326. DOI 10.1155/2014/725326.

TensorFlow., (2018). 2022. TensorFlow documentation. Jupyter notebook. Tensorflow. Online: https://github.com/tensorflow/docs/blob/d58904052034c0870678709d-c1ee8eb35e2fd34c/site/en/tutorials/generative/pix2pix.ipynb.

TensorFlow Datasets., b.d. Online: https://www.tensorflow.org/datasets (accessed 27 January 2022)

Tomaszewska M., Lewiński S., Woźniak E., 2011. Use of MODIS satellite images to study the percentage of vegetation cover. *Teledetekcja Środowiska* 46: 15–22.

Tucker C., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8(2): 127–150. DOI 10.1016/0034-4257(79)90013-0.

Turlej K., 2009. Comparison of NDVI index based on NOAA AVHRR, SPOT-VEGETATION and TERRA MODIS satellite data. *Teledetekcja Środowiska* 42: 83–88.

Tuszynska J., Gatkowska M., Wrobel K., Jagiello K., 2018. A pilot study on determining approximate date of crop harvest on the basis of sentinel-2 satellite imagery. *Geoinformation Issues* 10(1): 65–77. Online: http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.baztech-46991614-3b5b-429e-892e-b1a2556684c5.

van der Walt S., Schönberger JL., Nunez-Iglesias J., Boulogne F., Warner JD., Yager N., Gouillart E., Yu T., 2014. Scikit-image: Image processing in Python. *PeerJ* 2: e453. DOI 10.7717/peerj.453.

Verykokou S., Ioannidis C., 2019. *A Global Photogrammetry-Based Structure from Motion Framework: Application in Oblique Aerial Images*. Conference paper: FIG Working Week 2019: Geospatial information for a smarter life and environmental resilience. Hanoi, Vietnam

Wang Z., Bovik A., Sheikh H., Simoncelli E., 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4): 600–612. DOI 10.1109/TIP.2003.819861.

Worm A., Będkowski K., Bielecki A., 2019. 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* T. 60. Online: http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-4a024b76-0072-48be-94a6-cee-a9e001322.

Yao G., Yilmaz A., Zhang L., Meng F., Ai H., Jin F., 2021. Matching large baseline oblique stereo images using an end-to-end convolutional neural network. *Remote Sensing* 13(2): 274. DOI 10.3390/rs13020274.

Zhang Y., 2001. Texture-integrated classification of urban treed areas in high-resolution color-infrared imagery. *Photogrammetric Engineering & Remote Sensing* 67(12): 1359–1365.

Zhou S., Gordon M., Krishna R., Narcomey A., Fei-Fei L., Bernstein M., 2019. HYPE: a benchmark for human eYe perceptual evaluation of generative models. *arXiv:1904.01121 [cs]*. Online: http://arxiv.org/abs/1904.01121.

Zięba-Kulawik K., Hawryło P., Wężyk P., Matczak P., Przewoźna P., Inglot A., Mączka K., 2021. Improving methods to calculate the loss of ecosystem services provided by urban trees using LiDAR and aerial orthophotos. *Urban Forestry & Urban Greening* 63(sierpień): 127195. DOI 10.1016/j.ufug.2021.127195.

Zięba-Kulawik K., Wężyk P., 2022. Monitoring 3D changes in urban forests using landscape metrics analyses based on multi-temporal remote sensing data. *Land* 11(6): 883. DOI 10.3390/land11060883.