Article

Continuous Satellite Image Generation from Standard Layer Maps Using Conditional Generative Adversarial Networks

Arminas Šidlauskas [ID], Andrius Kriščiūnas *[ID] and Dalia Čalnerytė [ID]

Department of Applied Informatics, Faculty of Informatics, Kaunas University of Technology, Studentu St. 50, LT-51368 Kaunas, Lithuania; arminas.sidlauskas@ktu.edu (A.Š.); dalia.calneryte@ktu.lt (D.Č.)
* Correspondence: andrius.krisciunas@ktu.lt

Abstract: Satellite image generation has a wide range of applications. For example, parts of images must be restored in areas obscured by clouds or cloud shadows or areas that must be anonymized. The need to cover a large area with the generated images faces the challenge that separately generated images must maintain the structural and color continuity between the adjacent generated images as well as the actual ones. This study presents a modified architecture of the generative adversarial network (GAN) pix2pix that ensures the integrity of the generated remote sensing images. The pix2pix model comprises a U-Net generator and a PatchGAN discriminator. The generator was modified by expanding the input set with images representing the known parts of ground truth and the respective mask. Data used for the generative model consist of Sentinel-2 (S2) RGB satellite imagery as the target data and OpenStreetMap mapping data as the input. Since forested areas and fields dominate in images, a Kneedle clusterization method was applied to create datasets that better represent the other classes, such as buildings and roads. The original and updated models were trained on different datasets and their results were evaluated using gradient magnitude (GM), Fréchet inception distance (FID), structural similarity index measure (SSIM), and multiscale structural similarity index measure (MS-SSIM) metrics. The models with the updated architecture show improvement in gradient magnitude, SSIM, and MS-SSIM values for all datasets. The average GMs of the junction region and the full image are similar (do not exceed 7%) for the images generated using the modified architecture whereas it is more than 13% higher in the junction area for the images generated using the original architecture. The importance of class balancing is demonstrated by the fact that, for both architectures, models trained on the dataset with a higher ratio of classes representing buildings and roads compared to the models trained on the dataset without clusterization have more than 10% lower FID (162.673 to 190.036 for pix2pix and 173.408 to 195.621 for the modified architecture) and more than 5% higher SSIM (0.3532 to 0.3284 for pix2pix and 0.3575 to 0.3345 for the modified architecture) and MS-SSIM (0.3532 to 0.3284 for pix2pix and 0.3575 to 0.3345 for the modified architecture) values.

Keywords: remote sensing; conditional generative adversarial network; satellite image; image generation

1. Introduction

Satellites come in diverse configurations and employ various sensors for Earth observation, including radar, optical, LiDAR, and other technologies. Remotely collected data are used in applications at various areas such as landscape monitoring [1], agriculture [2,3], disaster management [4,5], and others. Two main problems can be distinguished when generation of satellite products has to be applied instead of using the raw data. The first one is data anonymization which is more common for photos with faces [6,7] or medical images [8]. The same problem also arises for satellite images in military-related problems where it is required to disguise sensitive areas [9,10]. The second problem is related to data quality. While each sensor type offers distinct advantages, they also have drawbacks that can affect their effectiveness. For example, radar sensors are proficient at penetrating

cloud cover and operating in adverse weather conditions, ensuring consistent data acquisition capabilities. Optical sensors are known for providing high-resolution data across different wavelengths but are unable to penetrate clouds. This limitation is evident in optical satellites like Landsat 8 (L8) or Sentinel-2 (S2), which frequently encounter issues with cloud cover, impeding their ability to consistently capture clear images of the Earth's surface [11,12]. Cloud-caused issues include geometric, atmospheric, and topographic errors in feature extraction [13].

Generally, models for image generation have greatly improved when generative adversarial networks (GANs) have been proposed and their performance was demonstrated in a category-to-image problem [14], where generative and discriminative models are trained simultaneously. The discriminative model estimates the probability that the sample is from training data or a generated one, while the generator aims to maximize the probability for the discriminator to fail. Several architectures based on GANs were introduced for image-to-image translation problems [15–18]. In [15], the architecture of pix2pix has been presented which allows one to synthesize photos from label maps, colorize images, and apply other image generation tasks. The U-Net-based architecture [19] has been used for generators, while convolutional PatchGAN classifier has been applied as a discriminator. It operates patches instead of the full image to define whether it is generated or a real one. Later, the architecture was improved, such as an applied multiscale generator and discriminator in pix2pixHD [17] and spatially adaptive normalization was added in SPADE [20]. It must be noted that although the paired reference and ground truth data are usually used for training data [15,17,20], alternative solutions exist for image generation such as CycleGAN proposed in [18]. The training of the model is carried out between samples of two domains without an explicit relation between two items in the datasets, i.e., photos and paintings. Of course, while this solution has its advantages to easily prepare the dataset without exact mapping, it results in less precise translation compared to directly mapped ones if mapped data are available.

In regard to remote sensing a lot of research on GAN applications has been performed during recent years, mostly in cloud removal problems for optical imagery. Of course, non-GAN solutions for cloud removal problems exist, as in [21,22]. In [21], the method involves utilizing a time series of optical sensor images, each capturing clouds in different locations at various times and sections containing clouds are replaced with cloud-free areas sourced from other images. In [22], a synthetic aperture radar (SAR) data-guided residual multiscale feature fusion network (CERMF-Net) was proposed for cloud elimination. However, GAN-based applications also are heavily applied in this field. In [23], it was attempted to remove thin and transparent clouds from satellite images using pix2pix architecture with dilated residual inception blocks and an additional SSIM loss function was incorporated. The method initially translates an SAR image to optical one and later combines it with a cloudy optical image to generate a non-cloudy one. In [24], cycle-consistent GAN has been proposed for thin cloud removal only from optical images without explicit paired datasets. In [25], missing information of satellite images is filled in using surrounded image pixel values and static high-resolution visual priors. In [26], a snow elimination problem has been addressed and a snow-cover removal model based on Cycle GAN has been developed. In [27], a two-stage cloud removal algorithm is presented, where, in the first stage, cloud segmentation and removal from the image are applied, and, in the second stage, removed areas are inpainted using U-Net-based architecture for the generator, providing a real image and inpainting area mask. To deal with large-scale satellite images, where large areas may have been covered by clouds, an iterative recovery scheme by progressively filling in the missing areas has been employed to restore grids with missing area of less than 50%.

In this research, an approach to generate optical satellite images based on administrative data and ensuring continuity between adjacent images is presented. Similar to [15,18] where map to image generation was applied on data scraped from Google Maps, in our research we focus on area of interest generation using Sentinel-2 RGB images mapped with administrative data from OpenStreetMap (OSM). The area of interest may be selected in

both cases when it is affected by clouds or their shadows or the area needs to be anonymized. For example, by extracting generic information about the area from a mapping system, including features such as water bodies, roads, forests, and others, a broad overview of the area can be compiled. These outlines can then serve as input for a trained generative model, capable of producing synthetic cloud-free optical sensor data. Of course, in any case when generated data are used instead of raw data, it may affect the results [28] and that must be considered in the decision-making process. Analogically, in anonymization cases, the area of interest may be simply replaced by a synthetic image based on a made up administrative map of composed features. The main contribution in this article addresses the data imbalance and continuous Sentinel-2 map generation problems. The imbalance problem is solved by clustering the dataset using the k-means algorithm and for each cluster a generative model is created. Then, based on the conditional mask class distribution, the most suitable model is selected. In order to apply the continuous map generation, the pix2pix image generation model is adopted for the structure-guided image completion problem. Comprehensive analysis of the results is performed for the scenarios when the dataset comprises random points in the selected area and a continuous map is generated.

2. Materials and Methods

The methodology consists of two main steps, that is, preparation of the dataset and quality analysis of images generated by GANs. For dataset preparation, two rasters are used: a mosaic of sentinel images and a standard map layer. The first raster is used for generative input and the second one is used as the model input. Then, both rasters are used to create datasets for the study area which is used to train and test GAN models under different scenarios for the addressed problems, namely, class imbalance in Sentinel-2 images (i.e., total area of houses or roads is much smaller than forest or water) and evaluation of the continuous mosaic composition of generated images using different GAN architectures.

2.1. Data Preparation

Lithuania was chosen as the study area. Although the proposed approach can be applied to any country, Lithuania, being the authors' country of residence, was chosen due to the planned future projects related to image generation and analysis for industrial and urban planning with respect to additional socio-economic parameters. Lithuania takes up around 65,300 km^2. The terrain is mostly flat with the highest point being about 293 m and contains many lakes and rivers. Forested areas cover around 33% of the country [29]. Because of rather consistent land coverage, enough data for training and analysis of the results are ensured and enable identification of possible challenges in image generation such as class imbalance, lack of continuity in generated images if large areas must be covered, and others.

2.1.1. Sentinel Data

In this research, Sentinel-2 was chosen because the need to generate parts of images originated from the need to recreate land images obstructed by clouds or cloud shadows. It is preferred in practical applications because of its higher spatial resolution (10 m) compared to similar spectral imaging satellites such as Landsat 8 (30 m).

A European Space Agency project, the Sentinel satellite missions are designed to observe the Earth's surface. These missions carry various different sensors, including radar and multispectral imaging equipment for ocean, land, and atmospheric monitoring. The Sentinel-2 mission uses a multispectral high-resolution sensor for land monitoring, able to provide a wide range of information based on its 13 different spectral bands, including visible, near infrared, and short-wave infrared bands.

Since Sentinel-2 satellite data are often used in research, this makes it a great subject for generative model research. This mission provides 13 different spectral bands, however, only 3, visible light bands, will be used in this research. The Sentinel-2 mission provides data with varied spatial resolution ranging from 10 to 60 m. The visible light bands are all

provided in 10 m spatial resolution so no further processing is required. In order to prepare the study area raster with selected spatial bands, multiple Sentinel-2 rasters have to be combined since a single Sentinel-2 raster only takes up part of the study area. Additional preprocessing of cloud filtering is also performed before the entire study area raster can be combined. Cloud filtering involves taking multiple rasters taken of the same area in a short period of time, in this case 1 month, then taking information from rasters which do not contain any clouds in that area. Data taken from Sentinel-2 satellites for the study area were captured in June of 2023. A study area raster after processing is presented in Figure 1.
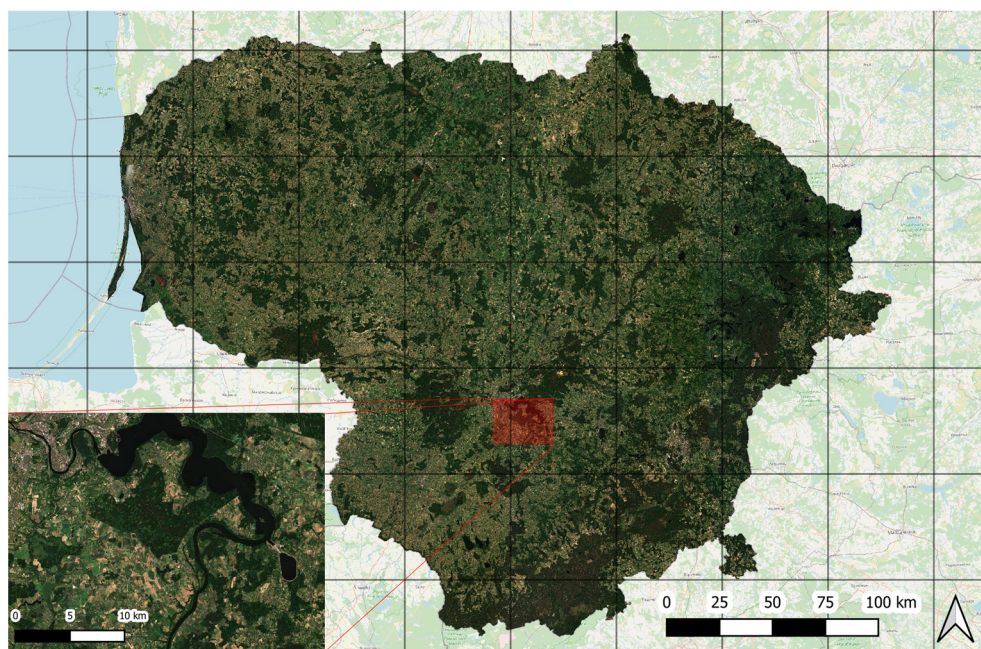

Figure 1. Sentinel-2 study area raster after preprocessing, LKS-94 coordinate format.

2.1.2. Administrative Data

OpenStreetMap (OSM) is an open-source mapping database that allows users to freely use its data. OSM provides various data including polygons of geographical features such as forests, buildings, lakes, rivers, and others. These data can be used as an input to generate Sentinel-2 images. For example, a model can be trained to take in OSM geographical features and then generate a filled in Sentinel-2 image. Data snippets are downloaded from the 2023 version and used as features. Since these data are provided in polygon form a rasterization process is performed to create an equivalent raster to the Sentinel-2 study area raster. However, before that, certain classes are merged together into single polygon classes. All different road types, such as motorway, secondary, primary, and others, are merged together into a single road class. Different water bodies such as rivers and lakes are also merged into a singular water class. Finally, after preprocessing the polygons, they are merged and rasterized, creating another study area raster as shown in Figure 2. The distribution of different classes in the study area is provided in Table 1.

Table 1. Distribution of study area by type.

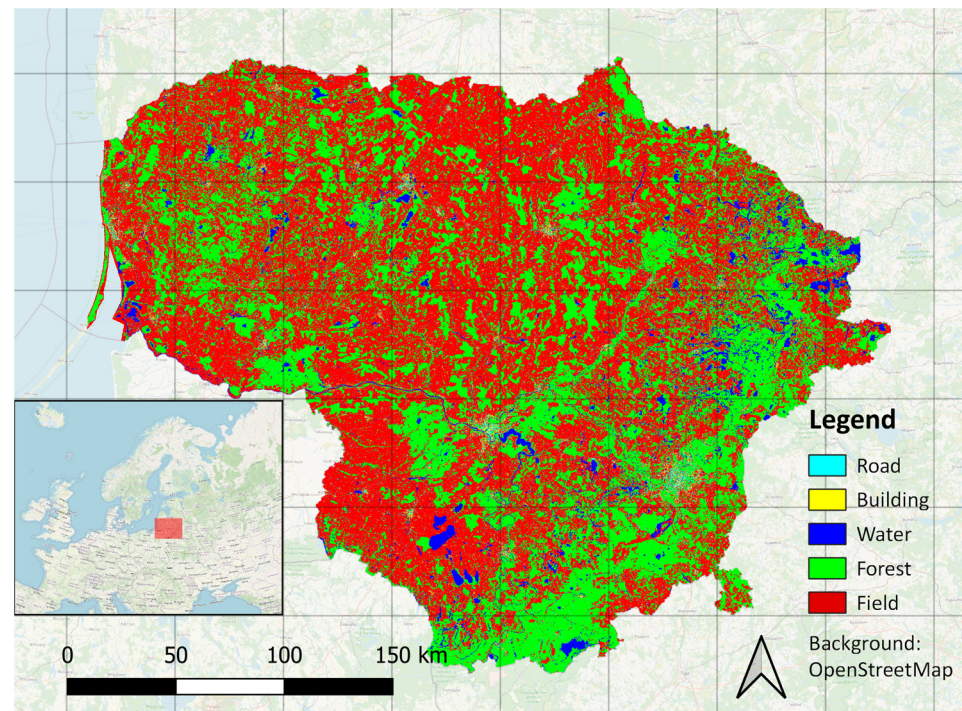| No	Type	Area Occupied
1	Road	2063.48 km^2 (3.16%)
2	Building	822.78 km^2 (1.26%)
3	Water	2461.81 km^2 (3.77%)
4	Forest	23,769.2 km^2 (36.40%)
5	Field	36,182.73 km^2 (55.41%)

Figure 2. Study area raster of land type.

2.1.3. Combining Dataset

Dataset preparation involves multiple processing steps. Firstly, the numbers of samples for training n_{train} and validation n_{val} are defined. Then, for every sample a random longitude and latitude coordinate is generated and used as a reference point to crop out a single sample of input and output images. Of course, randomization of coordinates does not strictly separate data and some overlap between generated images may occur. However, in this research we assume that the overlap is not significant. A single sample crop 256×256 px in size used for the generative model is shown in Figure 3a. After cropping out the dataset, another mirrored dataset is created with 3 inputs as shown in Figure 3b.
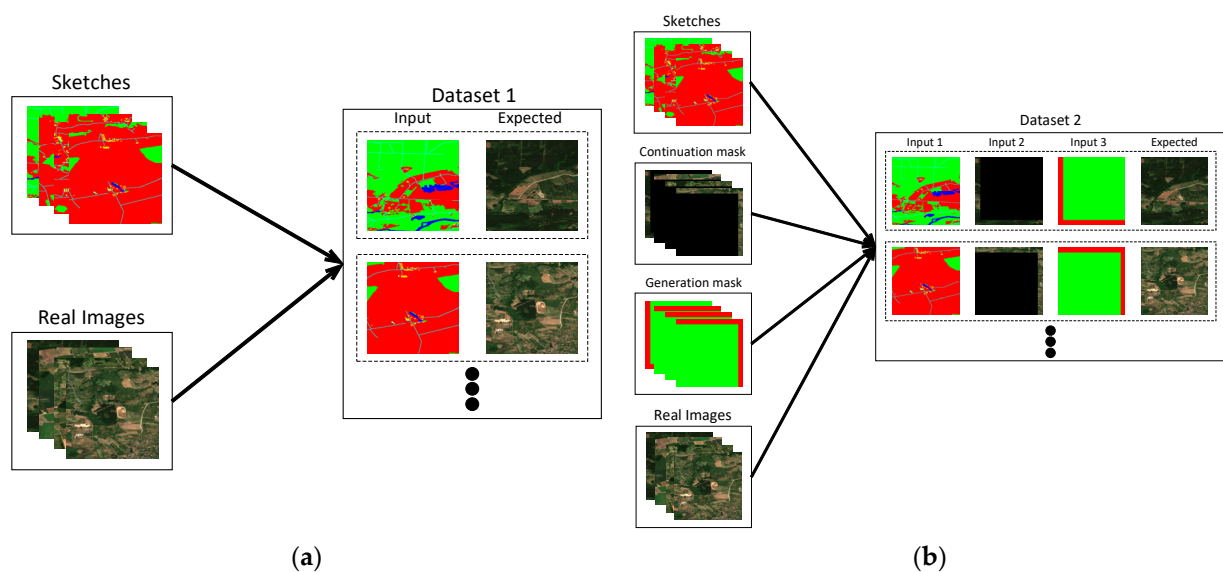

(**a**)　　　(**b**)

Figure 3. Composition of 2 datasets. (**a**) Dataset 1 composed of standard OSM input and Sentinel-2 output. (**b**) Dataset 2 composed of OSM input, continuation mask that is taken from a real image, and generation mask that defines the generation boundaries with red being non-generated area and green as generated area.

These inputs include the original OSM class input, partial ground truth input, and input defined with generative bounds, which consists of narrow strips at one or two image borders. In our research, we used $f = 20$ pixels for strip width. Procedure 1 to construct coverage of the strips on image borders is as follows:

Procedure 1:

1. Add one strip with uniform probability at one of the borders on the top, bottom, left, or right;
2. Add the strip on any other border with probability of 25%.

The first step of the procedure ensures that at least one border is covered, while the second step provides the samples with various covering scenarios, such as covering two borders at the corner of the image or covering ahead lines. Such border covering represents the problem of continuous raster generation for a large area, when an iterative image mosaic is assembled with narrow overlap between small images, which in a general case may be in any direction.

The testing dataset consists of another n_{test} samples of crop size of 512×512 pixels. Larger images for testing are used in order to simulate a situation when evaluation is performed on a continuously generated raster. The cropped image of 256×256 pixels with offsets of 128 pixels from the top and right represents the evaluation zone as shown in Figure 4. If the mosaic generation starts from the top-left corner as numbered in Figure 4, even though overlap between generated elements exists and the dimensions of the final mosaic of four images are k pixels smaller in both directions, the evaluation zone represents only the generated area and the junction of four generated images is in the middle of it.
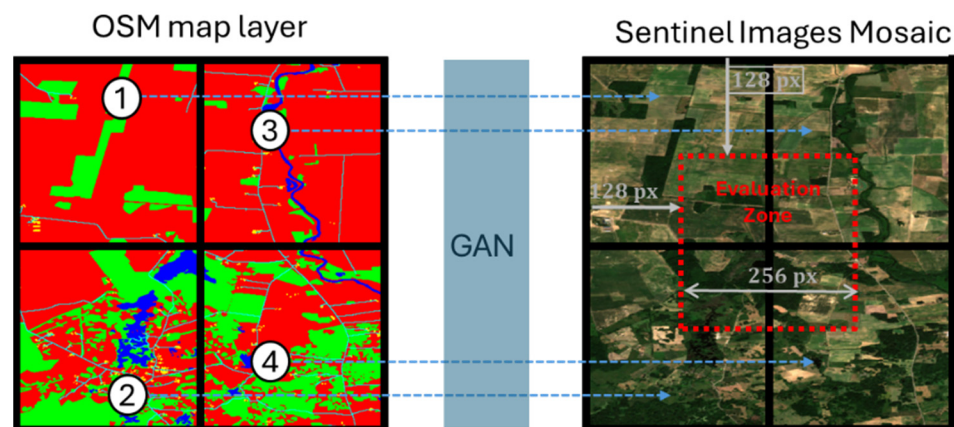

Figure 4. Example of a four-image generation for the capture of evaluation zone.

2.1.4. Data Balance

As can be seen in Table 1, the data imbalance in the masks is very high, especially comparing such classes as forest and fields with roads and buildings (36.4%, 55.41% to 3.16%, 1.26%). Although a GAN can be used to solve imbalanced data problems for other machine learning problems, when, i.e., image data are augmented with GANs instead of using standard data transformation procedures such as rotation or scaling [30], results of GAN-generated images also depend on training data [31]. Several other approaches may be used to deal with imbalanced data such as resampling techniques or using adopted loss functions [32,33]. In our research, the imbalance problem is solved by clustering the dataset using the k-means algorithm with each image represented by a feature vector that contains the number of pixels of each class. Then, a separate GAN model is created for each cluster. In the image generation step, the most suitable model is selected with respect to the distribution of classes in the conditional mask. The motivation for this approach is based on the fact that there is practically no limitation on creating dataset by selecting larger n_{train}, n_{val}, and n_{test} values and generating input data automatically. Datasets can be selected respectively to have enough data in each cluster even if the data distribution between

clusters is not uniform. This may be performed by multiplying initial data amounts by coefficient *r* with procedure 2 as follows:

Procedure 2:

1. Define n_{train}, n_{val}, n_{test}, and *r*;
2. Create dataset with $r \cdot n_{train}$, $r \cdot n_{val}$, $r \cdot n_{test}$ samples;
3. Cluster training dataset based on the image feature vector that contains the class distribution, select *k* optimal clusters for data, and assign the samples of all datasets to clusters based on the centroids obtained for the training dataset;
4. If any cluster has fewer samples than the initially defined n_{train}, n_{val}, n_{test} then increase the *r* value and repeat steps 2–3;
5. Select n_{train}, n_{val}, n_{test} samples from each cluster for further analysis.

Training several models is a time-consuming task. However, at the inference stage only the model selection based on context information is additionally performed. Also, this solution can be implemented in GAN-based problems unlike the other machine learning problems like semantic segmentation. In this case, the mask of the generated image is known in advance, while in semantic segmentation the image generally cannot be assigned to a more domain-specific model as no information about its probable class distribution is known.

2.2. Generative Models and Evaluation of Results

Generative models are machine learning models that are trained to take an input, capture certain details and patterns about it, and then generate new similar data according to those patterns. In this research a generative adversarial network (GAN) has been selected.

2.2.1. Generative Models

Depending on the context, different methods to generate images may be formulated, i.e., conditional image generation [20] tackles new image generation based on the same inputs or conditions. Other methods, such as structure-guided image completion, try to complete images by filling in missing parts by using the structural information of existing parts [34,35]. This research aims to implement a structure-guided image completion task for Sentinel-2 image generation based on conditional GAN pix2pix architecture [15]. The architecture is based on the pix2pix model and comprises two main components, that is, the generator *G* and the discriminator *D*.

The generator comprises the U-Net architecture shown in Figure 5a, and the discriminator uses PatchGAN architecture. Figure 5b shows the modification of the U-Net generator to enable implementation of structure-guided image completion. In addition to the conditional mask x_1, extra inputs into the standard skip connection part of the generator, namely, an image with a small part of target image x_2 and respective mask x_3, are used in the modification. Using a part of a target image in x_2 maintains continuity in the structure of the generated image. The mask x_3 enables locking in the existing part of the image according to the x_3 mask and then forcing the model to use the original image parts at the end of generation. The modified architecture consists of seven input layers and therefore the resulting GAN is denoted as pix2pix I7 later in the text.

The network aims to minimize the following PatchGAN loss function:

$$L(D, G) = E_y[log D(y)] + E_x[log(1 - D(G(\{x_n\})))] \tag{1}$$

where *D* is the discriminator network, *G* is the generator network, $\{x_n\}$ represents all possible inputs, that is, $\{x_1\}$ for the pix2pix and $\{x_1, x_2, x_3\}$ for the modified architecture pix2pix I7, E_y is the expectation operator over the distribution of real images *y*, E_x is the expectation operator over the distribution of input images *x* that are fed into the generator.
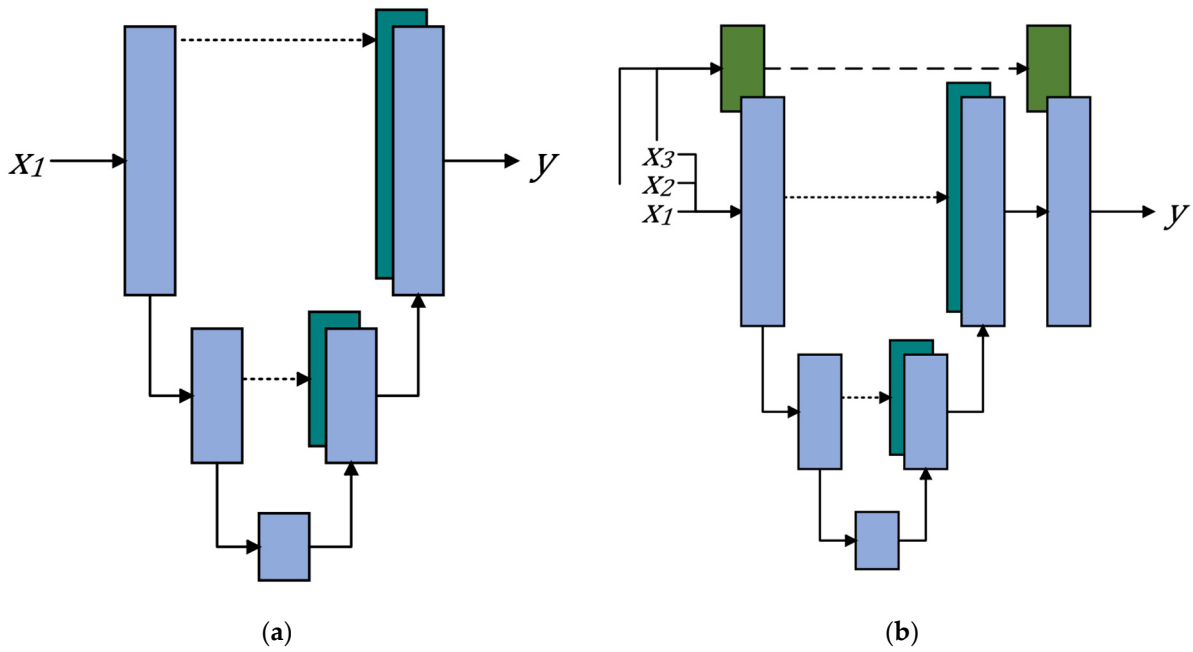
(**a**) (**b**)

Figure 5. Architectures of a generator: (**a**) denotes the standard U-Net architecture with skip connections; (**b**) denotes modified U-Net architecture, which takes in 3 different inputs, uses the additional inputs in the standard skip connection training in order to provide continuous generation, and additionally uses two extra inputs for locking in which areas the generator is supposed to generate and which are already complete.

2.2.2. Evaluation of Results

Unlike many segmentation models, GANs lack well-defined metrics for evaluation and several metrics for quantitative evaluation may be used [36]. In our research on the qualitative evaluation of the generated images we use the Fréchet inception distance (FID) metric [37] and structural similarity index measure (SSIM) [38]. The FID metric evaluates the quality of generated images by comparing the distribution of features extracted from the generated images to the distribution of features extracted from real images using a pretrained Inception v3 network. A lower FID value indicates that the generated images are more similar to ground truth images in terms of their feature distribution. The FIDis calculated as follows:

$$\text{FID} = \|m - m_w\|_2^2 + Tr\left(C + C_w - 2(CC_w)^{1/2}\right) \tag{2}$$

where m, m_w refer to the feature-wise mean of real and generated images; C, C_w are the covariance matrices of real and generated images.

The SSIM, on the other hand, measures the structural similarity between two images by comparing their luminance, contrast, and structural information. Here, an SSIM value of 1 indicates perfect similarity, 0 indicates no similarity, and negative values indicate dissimilarity. For two images x, y, the general SSIM formula is as follows:

$$\text{SSIM}(x, y) = (l(x,y))^{\alpha}(c(x,y))^{\beta}(s(x,y))^{\gamma} \tag{3}$$

where x, y are the input images; $l(x,y)$ is the luminescence comparison; $c(x,y)$ is the contrast comparison; $s(x,y)$ is the structural comparison; α, β, γ are parameters used to adjust the relative importance of each component, usually all set to 1 to have equal importance. The formula of luminance l is as follows:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{4}$$

where μ_x, μ_y are mean values of each compared image; C_1 is a small constant used to stabilize the equation (avoid division by 0). The formula of contrast c is as follows:

$$c(x,\ y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{5}$$

where σ_x, σ_y are the variance values for each specific input; C_2 is a small constant used to stabilize the equation. Lastly, the formula of structure s is as follows:

$$s(x,\ y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{6}$$

where σ_x, σ_y are the variance values for each specific input; σ_{xy} are the covariance values for both of the inputs; C_3 is a small constant used to stabilize the equation. If $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$, then the entire equation can be simplified to:

$$\text{SSIM}(x,\ y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{7}$$

The multiscale structural similarity index measure (MS-SSIM) is an updated approach using the same SSIM metric that improves the measurements by adding variations of image resolution and viewing conditions [39]. This method iteratively applies a low-pass filter and then downsamples the image. Given two input images, MS-SSIM is computed as follows:

$$\text{MS--SSIM}(x,\ y) = [l_M(x,\ y)]^{\alpha_M} \cdot \prod_{j=1}^{M} \left[c_j(x,y)\right]^{\beta_j} \left[s_j(x,y)\right]^{\gamma_j} \tag{8}$$

where the exponents α_M, β_j, γ_j are again used to adjust the importance of each component; M denotes the highest scale (usually $M = 5$) and luminance $l_M(x,\ y)$ is calculated only at that scale; j denotes finer scales that are calculated for both contrast $c_j(x,y)$ and structure $s_j(x,y)$; the specific component calculations remain the same as described above.

In order to evaluate how good the continuous image generation is, we compare the images generated by standard and modified architectures using the gradient magnitude (GM) metric which measures how strong the change in image intensity is [40]. Although the value of this metric depends on the content of the image (more details lead to a higher GM value), visible transitions between mosaic regions should increase the GM, making this comparison method suitable for our analysis, where a lower value of GM represents better results. GM can be calculated as follows:

$$\text{GM} = \sqrt{G_x^2 + G_y^2} \tag{9}$$

where G_x is the gradient of the image intensity in the horizontal x axis; G_y is the gradient of the image intensity in the vertical y axis. Both of the intensities can be calculated using a Sobel filter.

Additionally, in all cases the visual inspection of the generated images is performed.

3. Results

3.1. Experimental Setup

In order to train the generative model, we use $n_{train} = 2500$ samples. For validation and testing datasets, we select $n_{val} = n_{test} = 500$ samples each. In all cases the models are trained for 10,000 epochs with a learning rate of 0.01. In order to compare the impact of the data balancing, the following cases are analyzed:

1. $n_{train} \times k$ images are selected randomly from whole dataset (DS1), where k represents the number of clusters;
2. n_{train} elements from each of the data clusters are used (DS2);
3. k distinct models are trained for each cluster, where the ith ($i = \overline{1, k}$) model is trained with n_{train} samples from the ith cluster (DS3 − DS(2 + i)).

3.2. Dataset Preparation

In order to have at least the minimum number of samples after clusterization in each cluster, procedure 2 was applied. The coefficient $r = 50$ was selected and initially 100,000, 25,000, and 25,000 samples were generated for training, validation, and testing datasets. After the clusterization stage, the number of clusters $k = 3$ was selected using the Elbow method and the optimal number of clusters was numerically defined with the Kneedle algorithm, when k values have been analyzed in the range of $[1 \dots 10]$ as shown in Figure 6a. The images in the training data after clusterization have been distributed as follows: cluster 1: 38,020; cluster 2: 59,368; cluster 3: 2612. The class distributions in clusters are provided in Figure 6b.
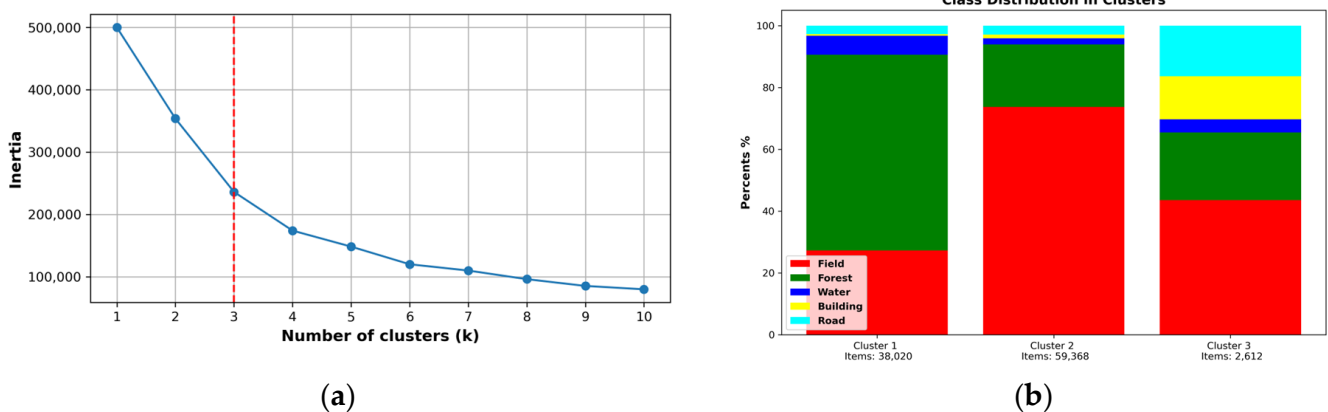

(a)

(b)

Figure 6. (**a**) Selection of optimal number of clusters (marked by the red dashed line) after applying Kneedle method, (**b**) Class distribution in different clusters, in addition to their sample sizes.

All three clusters show varying distributions. The first cluster is heavily biased towards forested areas, the second towards fields, and the third towards more urban areas. However, unlike the first two clusters, the 3rd cluster does not have a heavy overload of a single class, as multiple classes are represented in a balanced way rather than one of the main two classes (forest and field) being overloaded. This occurs because the urban areas take up a smaller part of satellite images. Additionally, it can be noted that with more urban areas, the road class also becomes more represented. From these clusters, five dataset cases (DS1–5) are created. DS1 is created from random images without taking any clusters into account with 7500 samples for training and 1500 samples for validation and testing. DS2 is created from a combination of all three clusters, by taking 2500 samples from each for training and 500 samples from each for validation and testing. The DS3–DS5 datasets take 2500 samples for training and 500 for validation and testing from each of the clusters. Figure 7 provides examples of each cluster. The first cluster is overloaded with "forest" data, the second cluster provides samples with the "field" class, and the third cluster provides examples of more urban areas where there are different classes, but most importantly "building" and "road" classes are better represented.

Figure 7. Examples of clusters: (**a**–**c**)—data from cluster 1; (**d**–**f**)—data from cluster 2; (**g**–**i**)—data from cluster 3.

3.3. Evaluation of Combined Images

After running the tests on all the models trained with their specific datasets, GM was calculated as shown in Table 2. The table provides results of the average GM for the respective testing sets. Additionally, the average GM is calculated in the junction area of the respective bandwidths, that is, the bands according to the horizontal and vertical axis in the middle of the combined image. The metric is provided for both types of models, that is, the standard architecture of pix2pix and the modified one of pix2pix I7. The GM value itself describes the local rate of change in the scalar field. If this metric is applied in the junction where the models are intended to generate images with a continuous structure, localized change can be measured. The results of the updated model pix2pix I7 suggest that it has a smaller rate of change in the junctions of generated images, implying that it creates smoother transitions between multiple generated images. This can be clearly seen

in cases where the average GM is calculated in the junction areas. For the pix2pix model, the average GM increases as the bandwidth decreases. For the narrowest analyzed band (10 px), the average value was more than 13% higher than the average of the full image (more that 20% for DS1 and DS3 datasets). In comparison, the average GM in the junction regions of the images generated by pix2pix did not differ significantly from the values obtained for the full image, that is, the difference did not exceed 5% with the exception of DS3 (forest-dominated cluster) in which the average GM was less than 7% higher than the average GM of the full image. Moreover, the trend that the GM increases as the analyzed area moves closer to the junction does not hold for the images generated using the pix2pix I7 model.

Table 2. Results of gradient magnitude on the specific datasets if average value is calculated for the full image or the vertical and horizontal bands in the transition area.

	Average (Full Image)		Average (Bandwidth 20 px)		Average (Bandwidth 10 px)	
	Pix2Pix	Pix2Pix I7	Pix2Pix	Pix2Pix I7	Pix2Pix	Pix2Pix I7
DS1	54.05	53.71	60.60	54.72	64.90	53.38
DS2	59.48	55.14	63.41	55.53	67.56	55.07
DS3	42.11	38.50	47.42	40.64	51.04	41.10
DS4	59.32	52.29	63.92	50.97	68.90	50.40
DS5	82.94	79.36	88.69	78.85	93.98	80.92

Additional visual examples of GM calculation are provided in Figure 8. These examples illustrate a visual distinction between the continuous images generated by the different models. Figure 8a presents the original image and the generated alternatives are provided in Figure 8b,c. Figure 8g–j show the GM filter on the generated images and respective junction areas for both models. It can be seen that the pix2pix I7 model provides a smoother transition and avoids creating sharp lines in the transition between multiple generated images.
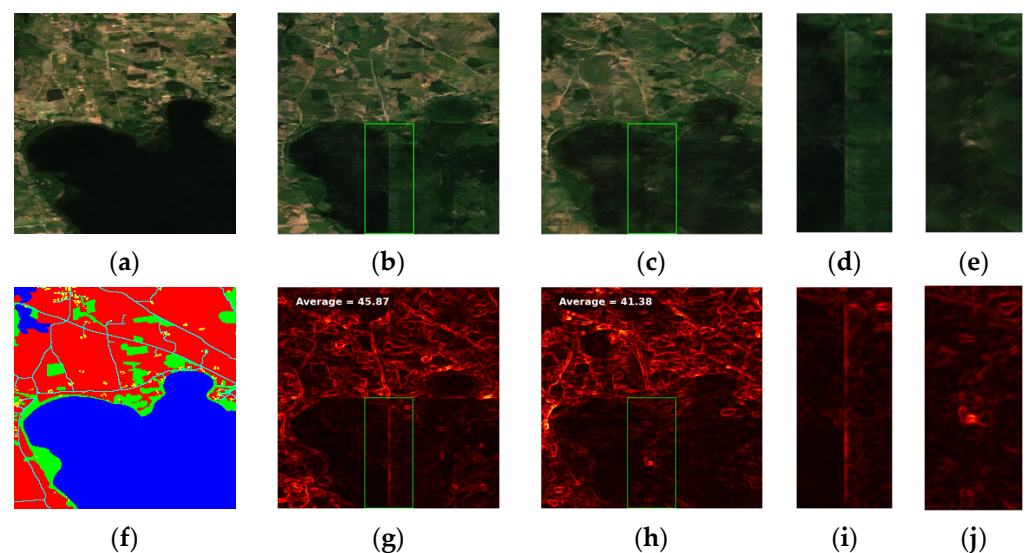

Figure 8. Visualization of gradient magnitude evaluation: (**a**) Example of ground truth image; (**b**) Example of an image generated with pix2pix model; (**c**) Example of an image generated with pix2pix I7 model; (**d**) Zoomed-in junction area from (**b**) image bounded in green; (**e**) Zoomed-in junction area from (**c**) image bounded in green; (**f**) Example of input sketch image with mapping data; (**g**) Example of an image generated with pix2pix model illustrated with gradient magnitude filter; (**h**) Example of an image generated with pix2pix I7 model with gradient magnitude filter; (**i**) Zoomed-in junction area from (**g**) image bounded in green; (**j**) Zoomed-in junction area from (**j**) image bounded in green.

3.4. Generated Image Quality Evalution

The evaluation of the proposed approach is provided in Table 3. As a baseline, the standard architecture of pix2pix and proposed pix2pix I7 architecture have been trained on DS1–5 datasets and each trained model was compared with separate testing data from clusters 1–3 (DS3–5 testing data). FID, SSIM, and MS-SSIM metrics were used for performance evaluation of numerical models.

Table 3. Evaluation results (FID, SSIM, and MS-SSIM) of pix2pix and pix2pix I7 models with testing data selected from clusters 1, 2, and 3.

	Cluster 1 (DS3)			Cluster 2 (DS4)			Cluster 3 (DS5)		
	FID	SSIM	MS-SSIM	FID	SSIM	MS-SSIM	FID	SSIM	MS-SSIM
Pix2Pix (DS1)	**141.202**	0.4299	0.5083	184.036	0.3132	0.3736	190.036	0.3284	0.4455
Pix2Pix (DS2)	141.685	0.4342	0.5193	**182.238**	0.3152	0.3817	197.586	0.3362	0.4587
Pix2Pix (DS3)	144.690	0.4461	0.5288	X	X	X	X	X	X
Pix2Pix (DS4)	X	X	X	182.477	0.3110	0.3714	X	X	X
Pix2Pix (DS5)	X	X	X	X	X	**X**	**162.673**	0.3532	0.5159
Pix2Pix I7 (DS1)	144.667	0.4456	0.5360	195.249	0.3267	0.3961	195.621	0.3345	0.4597
Pix2Pix I7 (DS2)	145.593	0.4595	0.5465	198.271	**0.3400**	**0.4056**	200.117	0.3453	0.4675
Pix2Pix I7 (DS3)	143.407	**0.4634**	**0.5477**	X	X	X	X	X	X
Pix2Pix I7 (DS4)	X	X	X	194.465	0.3352	0.3984	X	X	X
Pix2Pix I7 (DS5)	X	X	X	X	X	X	173.408	**0.3575**	**0.5165**

The FID evaluation shows that the original pix2pix model in most cases achieves better results than the updated pix2pix I7 model. For example, the FID metrics are 141.202 vs. 144.667 when models trained on DS1 have been compared on cluster 1 testing data. However, comparing the SSIM metric in all cases pix2pix I7 outperforms the standard architecture. For example, in the same scenario when models trained on DS1 data have been compared on cluster 1 testing data, the best results were obtained for pix2pix I7 with an SSIM value of 0.4634 while the standard pix2pix architecture trained on the same data obtained an SSIM value of 0.4461. When calculating the SSIM, the structural similarities of ground truth and generated images are compared for each image independently. Although the locked parts of the image are not included in the area for which the SSIM is calculated, the pix2pix I7 model tends to maintain better structural similarity between the adjacent images. The MS-SSIM corresponds to the SSIM results by demonstrating the better multiscale structural similarity.

It should be noted that the highest impact of the training models on the clustered data was noticed for cluster 3 (DS5) in which the ratio of classes "road" and "building" is significantly higher than in the others. The FID values for the pix2pix and pix2pix I7 models trained on the cluster 3 data (DS5) decreased by 14% and 11% compared to the respective models trained on the full dataset. Similarly, SSIM values increased by 7.5% and 6.86%, respectively. For other clusters, the impact of the training on clustered data was not significant, that is, the FID and SSIM values differed less than 5% compared with the models trained on the full dataset. This is explained by the fact that classes "road" and "building" highly dominate in clusters 1 and 2, similarly to the full dataset, and therefore there is no significant difference in the clustered and initial data.

Moreover, the highest FID and lowest SSIM values were obtained for the DS3 cluster with dominating class "forest" which demonstrated good application possibility for generating images with the forest-dominated conditional mask. Models for cluster 2 (DS4) and cluster 3 (DS5) demonstrated similar FID and SSIM values except for the models trained on DS5 data for the DS5 testing dataset.

Visual examples of data from each cluster are provided in Figure 9. The results of models of pix2pix trained on DS1 and pix2pix I7 trained on specific cluster data are compared. Examples show the standard test cases from each cluster, where the "forest" class dominates in the first cluster, the "field" class dominates in the second cluster, and in the last one, urban areas with classes "building" and "road" are better represented compared

with other clusters. Predictions of both models are shown as examples to illustrate what the updated model is attempting to achieve. The original pix2pix model contains hard lines in areas where multiple image predictions join, however, the updated pix2pix I7 model has smoother and more continuous transitions. However, certain drawbacks are still visible, for example, the results of the updated model trained on DS4 data still contain transitioning marks, even if they are not hard lines. This shows that although the updated model has the capabilities to create continuous transitions between generated images, further research might be needed in order to achieve smoother results.
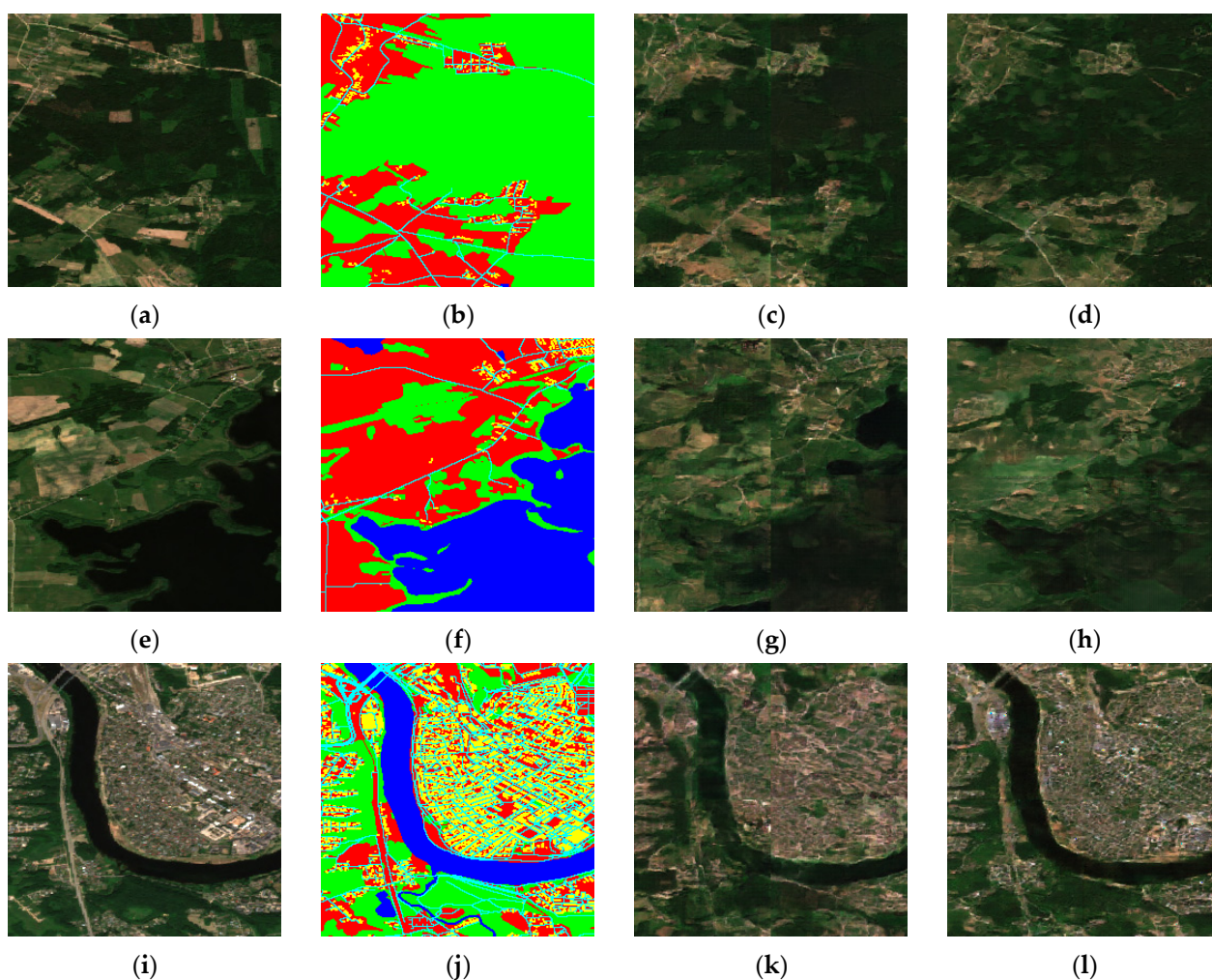

Figure 9. Examples of images from specific datasets and the predictions of the respective models trained on them: (**a**) DS3 ground truth image; (**b**) DS3 input mask image; (**c**) DS3 pix2pix model generation on DS1; (**d**) DS3 pix2pix I7 model generation on DS3; (**e**) DS4 ground truth image; (**f**) DS4 input mask image; (**g**) DS4 pix2pix model generation on DS1; (**h**) DS4 pix2pix I7 model generation on DS4; (**i**) DS5 ground truth image; (**j**) DS5 input mask image; (**k**) DS5 pix2pix model generation on DS1; (**l**) DS5 pix2pix I7 model generation on DS5. In the mask images, green represents Forest, red represents Field, blue represents Water, cyan represents Road, yellow represents Building.

4. Discussion

The goal of this study was to propose an approach to generate satellite images covering a large area by ensuring their visual and structural integrity with the adjacent generated and the known actual satellite images. The study contributes to the analysis of the data imbalance problem as well as proposes the modified pix2pix I7 GAN architecture to use known parts of the image as input.

The quality of the visual and structural integrity of the combined images was evaluated with FID, SSIM, MS-SSIM, and GM metrics. Although the FID values did not demonstrate the benefits of the modified architecture, the SSIM and MS-SSIM values were higher for the images generated using the pix2pix I7 model which implies better structural integrity of the images. The analysis of average GM demonstrated that the values in the junction region are higher compared to the ones obtained for the full image for the images generated using the original architecture whereas the images generated using the modified architecture maintain approximately uniform average GM throughout the image. Obviously, the GM can be affected by structures with sharp contours, such as roads or buildings, thus, the local effect on the junction area was analyzed.

The data imbalance problem was tackled by clustering the training images with respect to the class distribution in the conditional mask. Although the "forest" and "field" classes dominate in all three clusters, the third cluster represents the "building" and "road" classes better than the other clusters. Therefore, the models trained on only this dataset show a significant improvement compared to the models trained on the dataset without clustering. The disadvantages of using this strategy to train different models for each cluster include the limitations of time and computational resources required to train and store several models. On the other hand, in the inference stage, the model selection is performed by evaluating similarity to the training data centroids and using the model representing the most similar one.

According to the FID, SSIM, and MS-SSIM evaluation metrics, the best images were generated for the dataset cluster with the dominating "forest" class (more than 50% of pixel values in the images of the cluster). This could be explained by the rather simple and repetitive forest pattern throughout the dataset and a limited color scheme. In contrast, for the cluster in which the "field" class dominates over 50% of pixel values, higher FID values and lower SSIM values are observed, meaning worse conformity with the original input. This may be caused by a more complex pattern and a wider color range caused by the diversity of the field types (grassland, crops, plowed, and others).

Although the suggested approach demonstrates good results for the used datasets, the used datasets are rather specific to the geographic and climatic conditions. The research does not take into account the land coverage diversity that occurs in other geographical areas (mountains, deserts, other forest and field types). Thus, fine-tuning to the specific geographical conditions is necessary in order to transfer the models to other geographical areas or take into account global diversity. Similarly, the datasets were limited to images of the summer period because of the high overcast and cloudiness percentage (~80%) during the winter period. This leads to a limitation of image generation to only the summer period.

Using parts of the satellite image as GAN input contributes to generating a smooth transition between the images. However, the size of the input image needed to achieve acceptable integrity between the input and generated parts is not clear and could be considered in future work.

As technologies develop and the quality of satellite images improves, the selection of GAN architecture and its parameters, such as loss function, is not a finite process. The study presents a framework for analysis of generated images. However, the comparison of results obtained using different GAN architectures is a matter of future research. In addition, the evaluation of the quality of generated images and especially their junctions still lacks a clear methodology.

5. Conclusions

In this study, open access remote sensing data and open-source mapping data were used to create a GAN model that could generate continuous remote sensing data. Reference model pix2pix was taken and then modified with the purpose of making a model that was more suited to smoothly generating multiple remote sensing images. The modification included adding two additional inputs to the standard layer that allow the new model to have a reference point of where to start generating from. Multiple iterations of both

the original pix2pix and updated pix2pix I7 models were trained and evaluated using the following metrics: gradient magnitude, Fréchet inception distance, and structural similarity index measure. Models were trained on randomized data and on data clustered by the class distribution in an image. Five datasets were created, one from random data as DS1 of 7500 samples and four others from clustered data, DS2 of size 7500 with 2500 samples from each cluster and DS3–DS5 with 2500 samples from a specific cluster. Gradient magnitude results indicate that the updated model manages to make smoother transitions between multiple images when measuring the transition between multiple generated images. When measured against randomly composed DS1 data, the original pix2pix model scores 54.05 and the updated pix2pix I7 model manages to achieve a lower value of 53.71. Results are also better for the updated model when evaluating models trained on combined clustered data: DS2 59.48 against 55.14. Evaluation of FID and SSIM was performed for DS3–DS5 testing data, each dataset consisted of 500 testing samples. Evaluating the SSIM also yielded better results for the pix2pix I7 model. Models trained on DS3 and DS5 data achieved the highest results when tested on their own cluster datasets and for DS4 testing data the best evaluated model was pix2pix I7 trained on combined clustered data from DS2. On the other hand, FID results showed the opposite. All three testing datasets achieved the best results with the original pix2pix model. This could indicate that although the model obtains smoother transitions between multiple generated images as shown by gradient magnitude and SSIM results, the feature quality of the generated image becomes slightly worse with the pix2pix I7 model. The results show promising possibilities of modifying an existing pix2pix model for the generation of continuous images.

Author Contributions: Conceptualization, Arminas Šidlauskas, Andrius Kriščiūnas and Dalia Čalnerytė; methodology, Arminas Šidlauskas, Andrius Kriščiūnas and Dalia Čalnerytė; software, Arminas Šidlauskas, Andrius Kriščiūnas and Dalia Čalnerytė; validation, Arminas Šidlauskas, Andrius Kriščiūnas and Dalia Čalnerytė; formal analysis, Arminas Šidlauskas, Andrius Kriščiūnas and Dalia Čalnerytė; investigation, Arminas Šidlauskas, Andrius Kriščiūnas and Dalia Čalnerytė; resources, Arminas Šidlauskas, Andrius Kriščiūnas and Dalia Čalnerytė; data curation, Arminas Šidlauskas and Andrius Kriščiūnas; writing—original draft preparation, Arminas Šidlauskas and Andrius Kriščiūnas; writing—review and editing, Andrius Kriščiūnas and Dalia Čalnerytė; visualization, Arminas Šidlauskas, Andrius Kriščiūnas and Dalia Čalnerytė; supervision, Andrius Kriščiūnas; project administration, Andrius Kriščiūnas; funding acquisition, Arminas Šidlauskas, Andrius Kriščiūnas and Dalia Čalnerytė. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The OSM data are freely available from the Open Street Map website https://wiki.openstreetmap.org (accessed on 10 September 2024). The Sentinel-2 data is available on the Copernicus Data Space Ecosystem https://dataspace.copernicus.eu (accessed on 10 of September 2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Radočaj, D.; Obhođaš, J.; Jurišić, M.; Gašparović, M. Global open data remote sensing satellite missions for land monitoring and conservation: A review. *Land* **2020**, *9*, 402. [CrossRef]
2. Hoang, T.D.; Pham, M.T.; Vu, T.T.; Nguyen, T.H.; Huynh, Q.-T.; Jo, J. Monitoring agriculture areas with satellite images and deep learning. *Appl. Soft Comput.* **2020**, *95*, 106565. [CrossRef]
3. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
4. Notti, D.; Cignetti, M.; Godone, D.; Giordan, D. Semi-automatic mapping of shallow landslides using free Sentinel-2 images and Google Earth Engine. *Nat. Hazards Earth Syst. Sci.* **2023**, *23*, 2625–2648. [CrossRef]
5. Nhangumbe, M.; Nascetti, A.; Ban, Y. Multi-Temporal Sentinel-1 SAR and Sentinel-2 MSI Data for Flood Mapping and Damage Assessment in Mozambique. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 53. [CrossRef]

6. Maximov, M.; Elezi, I.; Leal-Taixé, L. CIAGAN: Conditional Identity Anonymization Generative Adversarial Networks. 2020. Available online: https://github.com/dvl-tum/ciagan (accessed on 10 September 2024).

7. Li, T.; Lin, L. AnonymousNet: Natural Face De-Identification With Measurable Privacy. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; pp. 56–65.

8. Shin, H.-C.; Tenenholtz, N.A.; Rogers, J.K.; Schwarz, C.G.; Senjem, M.L.; Gunter, J.L.; Andriole, K.P.; Michalski, M. Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2018; pp. 1–11.

9. Cannas, E.D.; Mandelli, S.; Bestagini, P.; Tubaro, S.; Delp, E.J. Deep Image Prior Amplitude SAR Image Anonymization. *Remote Sens.* **2023**, *15*, 3750. [CrossRef]

10. Yao, Z.; Liu, Q.; Yang, J.; Chen, Y.; Wu, Z. PPUP-GAN: A GAN-based privacy-protecting method for aerial photography. *Future Gener. Comput. Syst.* **2023**, *145*, 284–292. [CrossRef]

11. Townshend, J.R.; Masek, J.G.; Huang, C.; Vermote, E.F.; Gao, F.; Channan, S.; Sexton, J.O.; Feng, M.; Narasimhan, R.; Kim, D.; et al. Global characterization and monitoring of forest cover using Landsat data: Opportunities and challenges. *Int. J. Digit. Earth* **2012**, *5*, 373–397. [CrossRef]

12. Sudmanns, M.; Tiede, D.; Augustin, H.; Lang, S. Assessing global Sentinel-2 coverage dynamics and data availability for operational Earth observation (EO) applications using the EO-Compass. *Int. J. Digit. Earth* **2020**, *13*, 768–784. [CrossRef]

13. Alavipanah, S.K.; Firozjaei, M.K.; Sedighi, A.; Fathololoumi, S.; Naghadehi, S.Z.; Saleh, S.; Naghdizadegan, M.; Gomeh, Z.; Arsanjani, J.J.; Makki, M.; et al. The Shadow Effect on Surface Biophysical Variables Derived from Remote Sensing: A Review. *Land* **2022**, *11*, 2025. [CrossRef]

14. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. Available online: http://www.github.com/goodfeli/adversarial (accessed on 10 September 2024).

15. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv* **2016**, arXiv:1611.07004.

16. Shaham, T.R.; Gharbi, M.; Zhang, R.; Shechtman, E.; Michaeli, T. Spatially-Adaptive Pixelwise Net-Works for Fast Image Translation. 2021. Available online: https://tamarott.github.io/ASAPNet_web (accessed on 10 September 2024).

17. Wang, T.-C.; Liu, M.-Y.; Zhu, J.-Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8798–8807.

18. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A.; Research, B.A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks Monet Photos. 2017. Available online: https://github.com/junyanz/CycleGAN (accessed on 10 September 2024).

19. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597. Available online: https://arxiv.org/abs/1505.04597 (accessed on 10 September 2024).

20. Park, T.; Liu, M.-Y.; Wang, T.-C.; Zhu, J.-Y. Semantic Image Synthesis with Spatially-Adaptive Normalization. 2019. Available online: http://arxiv.org/abs/1903.07291 (accessed on 10 September 2024).

21. Duan, C.; Pan, J.; Li, R. Thick cloud removal of remote sensing images using temporal smoothness and sparsity regularized tensor optimization. *Remote Sens.* **2020**, *12*, 3446. [CrossRef]

22. Anandakrishnan, J.; Sundaram, V.M.; Paneer, P. CERMF-Net: A SAR-Optical Feature Fusion for Cloud Elimination From Sentinel-2 Imagery Using Residual Multiscale Dilated Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 11741–11749. [CrossRef]

23. Darbaghshahi, F.N.; Mohammadi, M.R.; Soryani, M. Cloud Removal in Remote Sensing Images Using Generative Adversarial Networks and SAR-to-Optical Image Translation. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–9. [CrossRef]

24. Singh Praveer and Komodakis Nikos, Cloud-GAN: Cloud removal for Sentinel-2 imagery using a cyclic consistent generative adversarial networks. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain,, 22–27 July 2018.

25. Zaytar, M.A.; El Amrani, C. Satellite image inpainting with deep generative adversarial neural networks. *IAES Int. J. Artif. Intell. (IJ-AI)* **2021**, *10*, 121–130. [CrossRef]

26. Oluwadare, T.S.; Chen, D.; Oluwafemi, O.; Babadi, M.; Hossain, M.; Ibukun, O. Reconstructing Snow-Free Sentinel-2 Satellite Imagery: A Generative Adversarial Network (GAN) Approach. *Remote Sens.* **2024**, *16*, 2352. [CrossRef]

27. Zheng, J.; Liu, X.-Y.; Wang, X. Single Image Cloud Removal Using U-Net and Generative Adversarial Networks. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6371–6385. [CrossRef]

28. Hukkelås, H.; Lindseth, F. Does Image Anonymization Impact Computer Vision Training? *Anonymize Dataset Training & Evaluation Model Training Performance Drop on Original Validation Set? 2023.* Available online: https://github.com/hukkelas/deep_privacy2 (accessed on 10 September 2024).

29. Juknelienė, D.; Kazanavičiūtė, V.; Valčiukienė, J.; Atkocevičienė, V.; Mozgeris, G. Spatiotemporal patterns of land-use changes in Lithuania. *Land* **2021**, *10*, 619. [CrossRef]

30. Mariani, G.; Scheidegger, F.; Istrate, R.; Bekas, C.; Malossi, C. BAGAN: Data Augmentation with Balancing GAN. 2018. Available online: http://arxiv.org/abs/1803.09655 (accessed on 10 September 2024).

31. Anaissi, A.; Jia, Y.; Braytee, A.; Naji, M.; Alyassine, W. Damage GAN: A Generative Model for Imbalanced Data. In *Data Science and Machine Learning*; Springer: Singapore, 2023. [CrossRef]

32. Kraiem, M.S.; Sánchez-Hernández, F.; Moreno-García, M.N. Selecting the suitable resampling strategy for imbalanced data classification regarding dataset properties. an approach based on association models. *Appl. Sci.* **2021**, *11*, 8546. [CrossRef]

33. Koziarski, M.; Woźniak, M. CCR: A combined cleaning and resampling algorithm for imbalanced data classification. *Int. J. Appl. Math. Comput. Sci.* **2017**, *27*, 727–736. [CrossRef]

34. Liao, L.; Xiao, J.; Wang, Z.; Lin, C.-W.; Satoh, S. Uncertainty-Aware Semantic Guidance and Estimation for Image Inpainting. *IEEE J. Sel. Top. Signal Process.* **2020**, *15*, 310–323. [CrossRef]

35. Zheng, H.; Lin, Z.; Lu, J.; Cohen, S.; Zhang, J.; Xu, N.; Luo, J. Semantic Layout Manipulation with High-Resolution Sparse Attention. 2020. Available online: http://arxiv.org/abs/2012.07288 (accessed on 10 September 2024).

36. Salehi, P.; Chalechale, A.; Taghizadeh, M. Generative Adversarial Networks (GANs): An Overview of Theoretical Model, Evaluation Metrics, and Recent Developments. *arXiv* **2020**, arXiv:2005.13178.

37. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. 2017. Available online: http://arxiv.org/abs/1706.08500 (accessed on 10 September 2024).

38. Nilsson, J.; Akenine-Möller, T. Understanding SSIM. 2020. Available online: http://arxiv.org/abs/2006.13846 (accessed on 10 September 2024).

39. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multi-Scale Structural Similarity For Image Quality Assessment. In Proceedings of the 37th Asilomar Conference of Signals, Systems & Computers, 2003, Pacific Grove, CA, USA; 2003. [CrossRef]

40. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*; Pearson: London, UK, 2018.