MDPI

Article

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

Kinga Karwowska * and Damian Wierzbicki

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

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

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

1. Introduction

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

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

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

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

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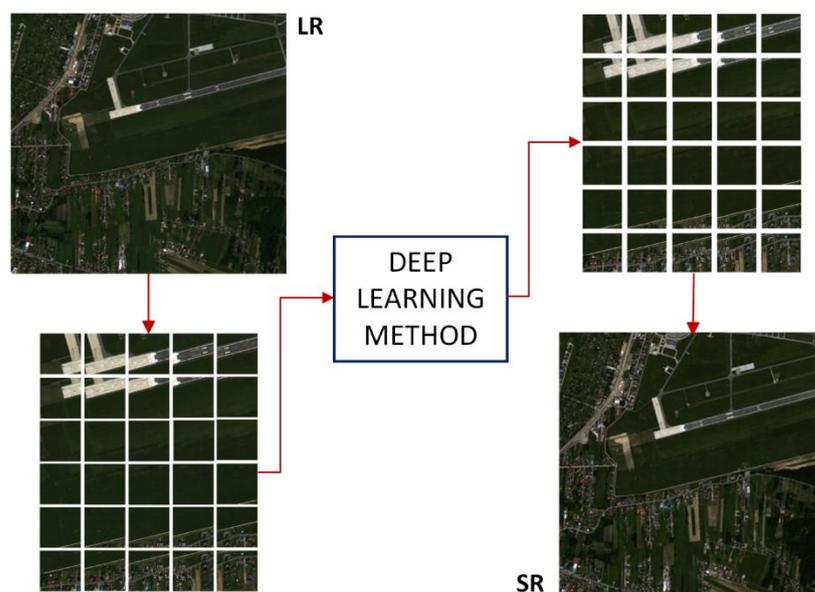

Figure 1. Diagram illustrating the process of improving the resolution of satellite imagery.

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

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

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

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

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

2. Related Works

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

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

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

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

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

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

3. Experiments and Results

3.1. The Proposed Method

The aim of the experiment is to combine super resolution (SR) images estimated with the use of a GAN network based on low resolution (LR) images. In the proposed methodology, the fragments of the scene are combined with the use of two-dimensional window functions. The application of the window function consists of preparing the adequate matrix of weights (which is symmetrical in relation to the center of the image) and then multiplying it by the super resolution (SR) image, estimated by the generative adversarial network. The specific stages of the process are presented in Figure 2.
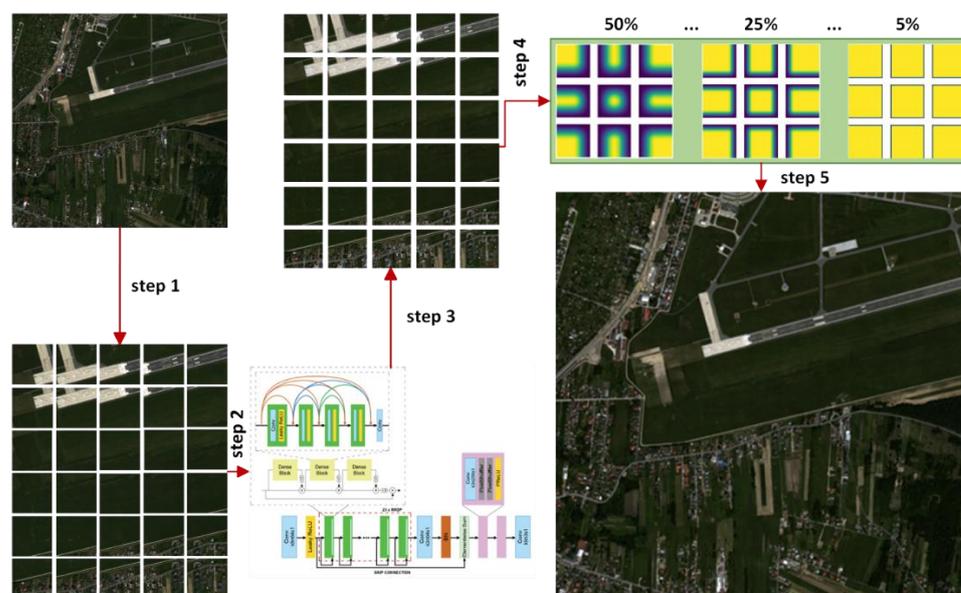

Figure 2. Diagram of the methodology to improve the resolution of whole satellite scenes. Step 1: divided of image, step 2: upload images to the network, step 3: use a method of improving spatial resolution using neural networks, step 4: select a window function to combine images, step 5: combine images.

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

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

3.2. Equations

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

3.2.1. Peak Signal-to-Noise Ratio

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

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

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

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

3.2.2. Universal Quality Measure

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

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

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

3.2.3. Spatial Correlation Coefficient

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

3.2.4. Spectral Angle Mapper

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

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

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

3.2.5. Spectral Angle Mapper

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

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

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

3.2.6. VIFP

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

3.2.7. Normalized Root Mean-Squared Error

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

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

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

3.2.8. Mean Square Error

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

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

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

3.2.9. Root Mean Square Error

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

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

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

3.3. Preliminary Tests

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

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

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

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

stripes are visible in the resulting images. If the value of the sum of weights is lower than one, the DN values for those pixels are lower than the target value, which, in consequence, leads to the emergence of a dark grid. On the other hand, if the sum of weights is higher than 1, the DN value of the pixel is higher than the target value, which results in the emergence of lighter stripes (Figure 5).
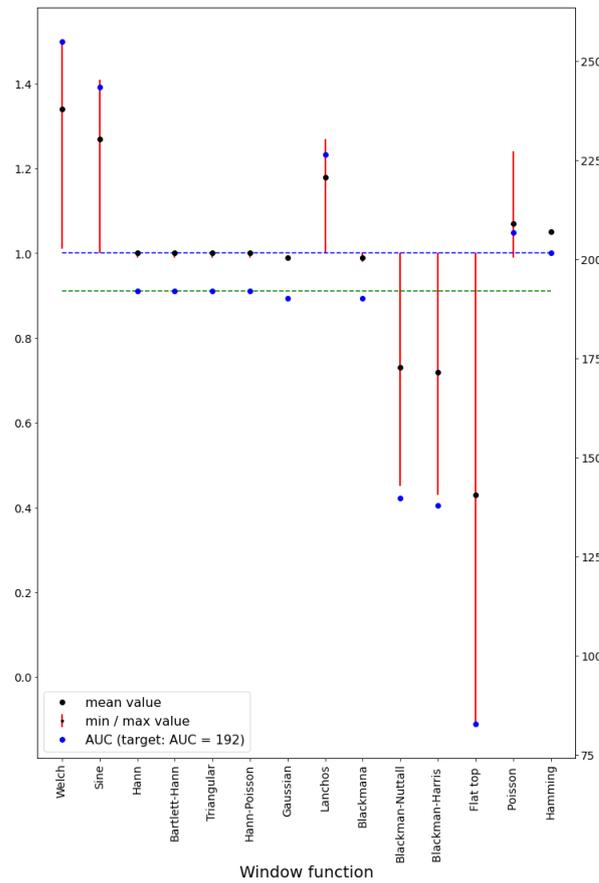

Figure 3. Presentation of the values of the average, minimum, and maximum values of the weights of the analyzed windows. It was assumed that the images overlapped in 50%. Numerical values were determined for the overlap area.

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

Window Function\Metrics	MSE	RMSE	PSNR	UQI	SCC	SAM	SSIM	RASE	VIFP	NRMSE
Overlap	0.00	0.00	-	1.00	1.00	0.00	1.00	0.00	1.00	0.00
Hann $a_0 = 0.5$	0.42	0.64	51.89	1.00	1.00	0.01	1.00	0.35	1.00	0.01
Bartlett-Hann	3.66	1.91	42.49	1.00	0.91	0.01	1.00	101.21	0.98	0.02
Triangular	3.84	1.95	42.29	1.00	0.90	0.01	1.00	104.21	0.98	0.02
Hann-Poisson	3.42	1.85	42.78	0.99	0.92	0.01	1.00	96.91	0.98	0.02
Gaussian	92.79	9.63	28.46	0.99	0.89	0.08	0.99	401.44	0.89	0.08
Gaussian *	75.74	8.70	29.34	1.00	0.88	0.07	0.99	354.06	0.87	0.07
Lanchos	1288.29	35.89	17.03	0.90	0.85	0.30	0.88	1658.52	0.72	0.32
Lanchos *	863.22	29.38	18.77	0.95	0.81	0.24	0.88	1294.12	0.59	0.24
Blackman	16.68	4.08	35.91	1.00	0.90	0.02	1.00	207.04	0.97	0.03
Blackman *	3.23	1.80	43.03	1.00	0.89	0.01	1.00	87.88	0.97	0.01

Figure 4. Sample image generated as a result of combining images with the use of selected windows. For methods marked with "*", histogram adjustment was applied before combining images after adjusting the histogram to the reference image.
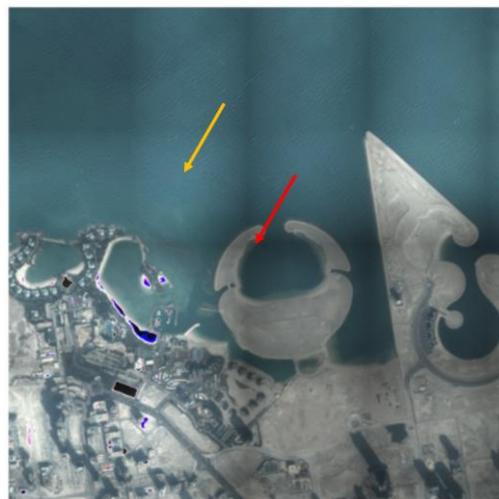

Figure 5. Fragment of the resulting image that was generated with the use of the Lanchos window. The lighter stripe of pixels that shows that the sum of weights is higher than 1 and is marked with the yellow arrow, while the red arrow indicates the stripe of darker pixels, where the sum of weights is lower than 1.

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

3.4. Results

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

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

3.4.1. Database

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

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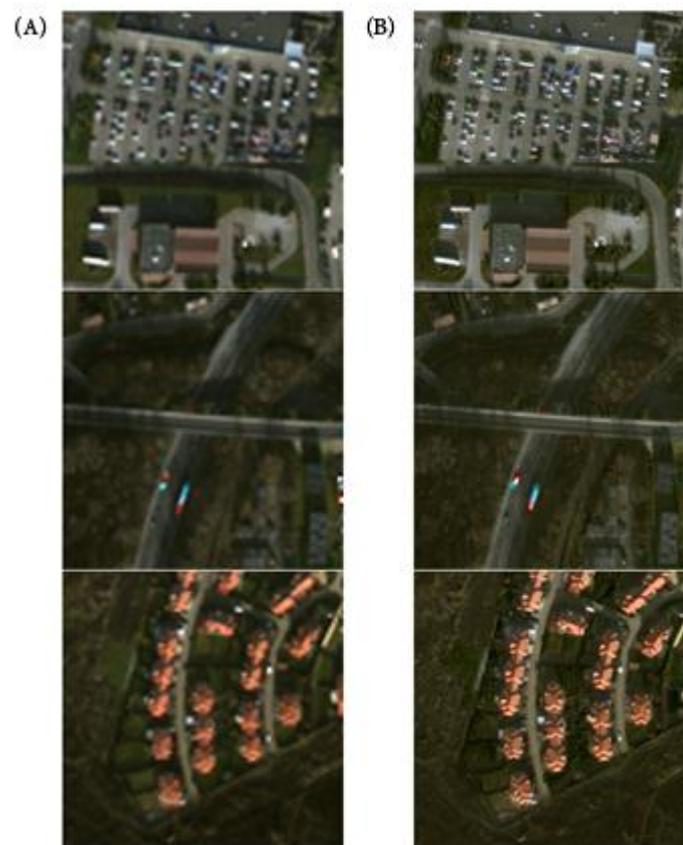

Figure 6. Sample pairs of (**A**) LR and (**B**) HR images that were used to train the ESRGAN network.

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

3.4.2. The ESRGAN Network

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

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

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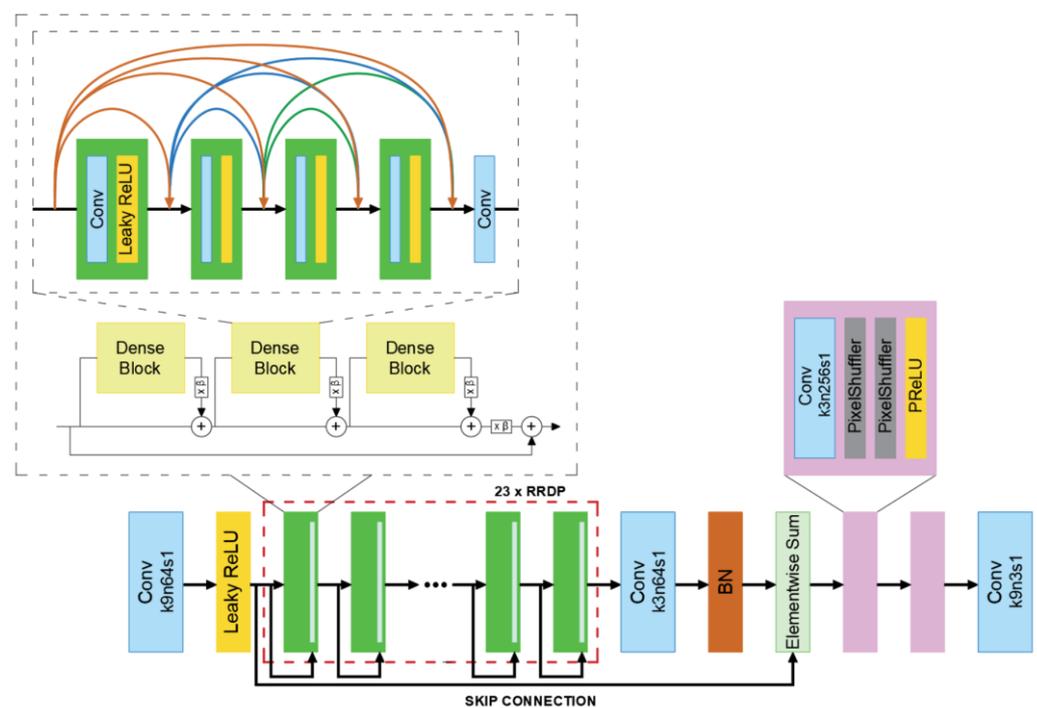

Figure 7. Model of the ESRGAN network generator (base on [41]).

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

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

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

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

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

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

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

where l^{SR}—Perceptual loss function.

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

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

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

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

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

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

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

3.4.3. Network Training

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

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

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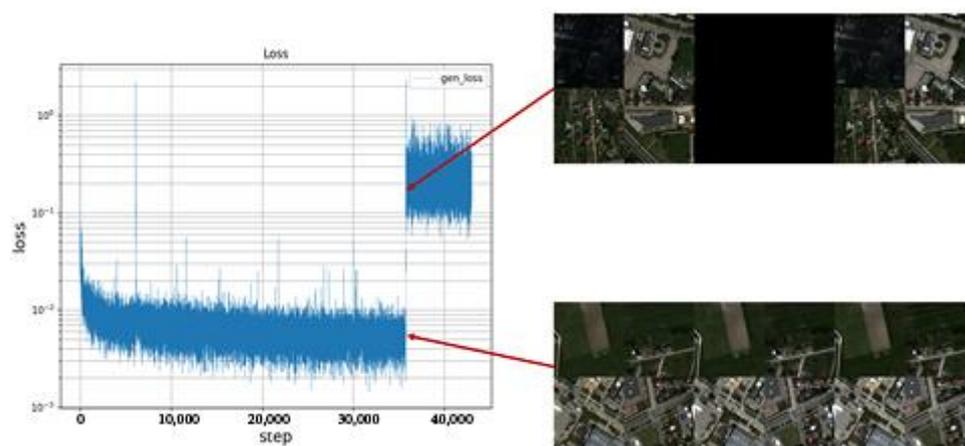
Figure 8. Sample pairs of LR and HR images that were used to train the ESRGAN network.

Table 3. Learning rate values used in network training.

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

3.4.4. Combining Images

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

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

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

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

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

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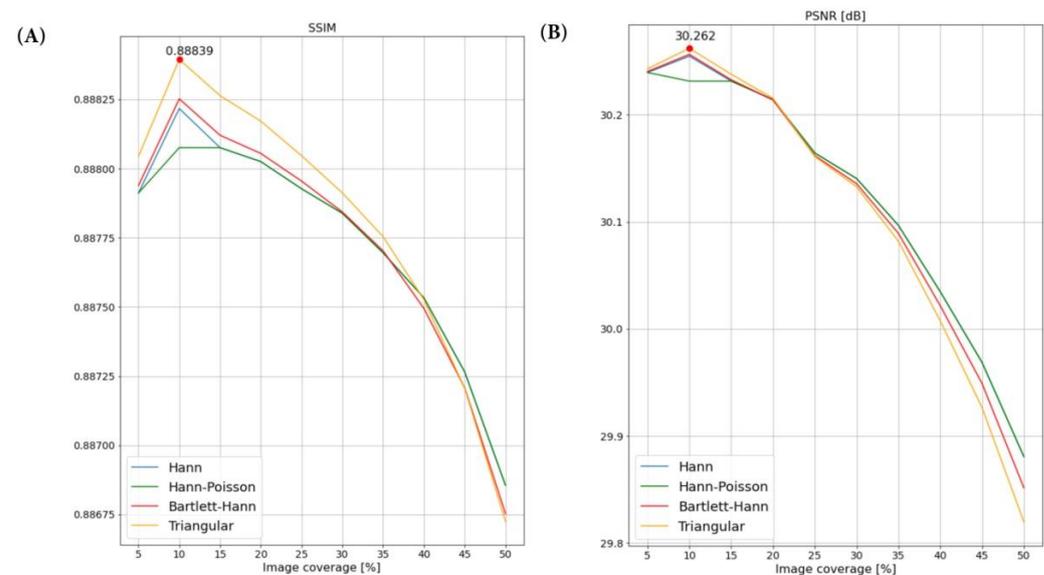

Figure 10. The values of the (**A**) SSIM and (**B**) PSNR metrics for the image presented in Figure 9.

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

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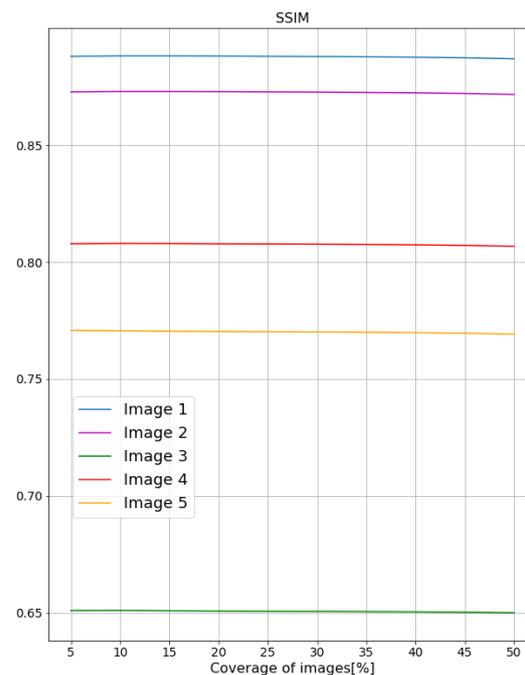
Figure 11. SSIM values for the analyzed images.

4. Discussion

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

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

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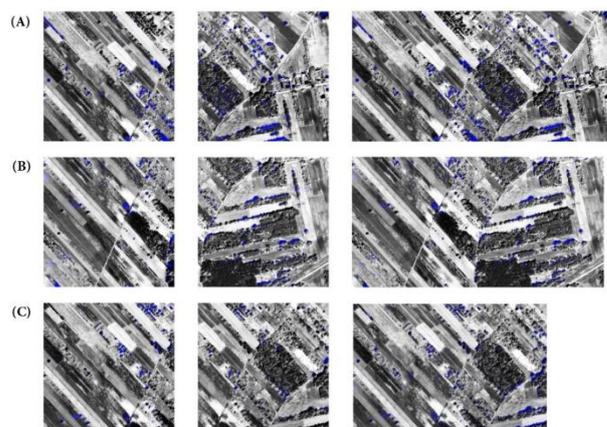
Figure 12. An example of the application of window function to combine shadow masks that were detected with the use of the UNet network. The images shown have an overlap of: (**A**) 5%, (**B**) 10%, (**C**) 50%. For panchromatic images (where histogram equalization was applied) window functions were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

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

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

5. Conclusions

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

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

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

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

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

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

Table A1. *Cont.*

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

Table A1. *Cont.*

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

Appendix B

Table A2. Results.
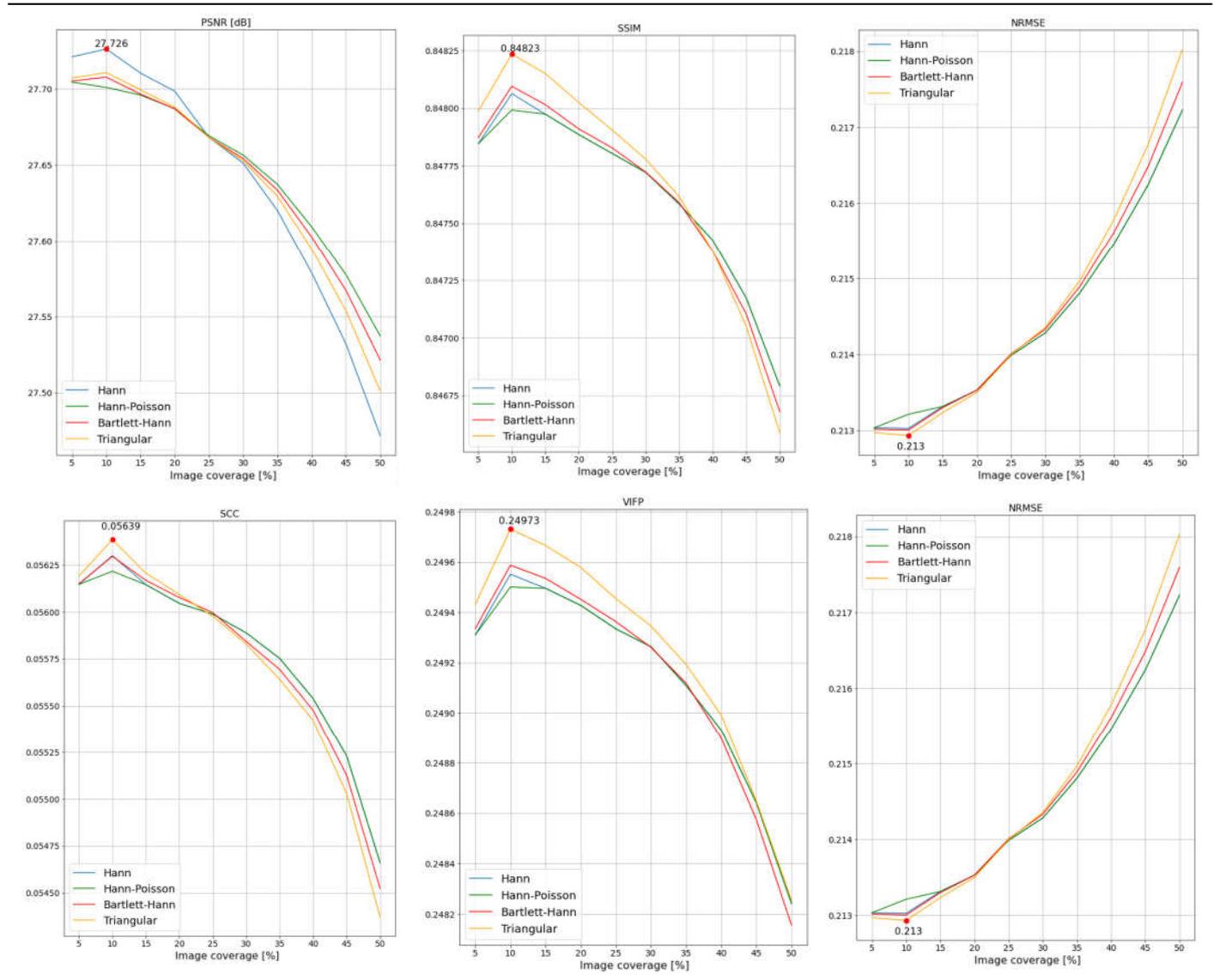
References

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

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

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

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

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

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

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

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

15. Scarpa, G.; Gargiulo, M.; Mazza, A.; Gaetano, R. A CNN-Based Fusion Method for Feature Extraction from Sentinel Data. *Remote Sens.* **2018**, *10*, 236. [CrossRef]

16. Oveis, A.H.; Giusti, E.; Ghio, S.; Martorella, M. CNN for Radial Velocity and Range Components Estimation of Ground Moving Targets in SAR. In Proceedings of the 2021 IEEE Radar Conference (RadarConf21), Atlanta, GA, USA, 7–14 May 2021; pp. 1–6.

17. Wang, J.; Lu, C.; Jiang, W. Simultaneous Ship Detection and Orientation Estimation in SAR Images Based on Attention Module and Angle Regression. *Sensors* **2018**, *18*, 2851. [CrossRef]

18. Kuuste, H.; Eenmäe, T.; Allik, V.; Agu, A.; Vendt, R.; Ansko, I.; Laizans, K.; Sünter, I.; Lätt, S.; Noorma, M. Imaging System for Nanosatellite Proximity Operations. *Proc. Est. Acad. Sci.* **2014**, *63*, 250. [CrossRef]

19. Blommaert, J.; Delauré, B.; Livens, S.; Nuyts, D.; Moreau, V.; Callut, E.; Habay, G.; Vanhoof, K.; Caubo, M.; Vandenbussche, J.; et al. CHIEM: A New Compact Camera for Hyperspectral Imaging. 2017. Available online: https://www.researchgate.net/publication/321214165_CHIEM_A_new_compact_camera_for_hyperspectral_imaging (accessed on 18 October 2022).

20. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.

21. Karwowska, K.; Wierzbicki, D. Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3292–3312. [CrossRef]

22. Lu, T.; Wang, J.; Zhang, Y.; Wang, Z.; Jiang, J. Satellite Image Super-Resolution via Multi-Scale Residual Deep Neural Network. *Remote Sens.* **2019**, *11*, 1588. [CrossRef]

23. Schmidt, R. Multiple Emitter Location and Signal Parameter Estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [CrossRef]

24. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.

25. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.

26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.

27. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv* **2015**, arXiv:1512.00567.

28. Han, J.; Zhang, D.; Cheng, G.; Guo, L.; Ren, J. Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3325–3337. [CrossRef]

29. Wang, Z.; Zhang, Y.; Yu, Y.; Zhang, L.; Min, J.; Lai, G. Prior-Information Auxiliary Module: An Injector to a Deep Learning Bridge Detection Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6270–6278. [CrossRef]

30. Yu, D.; Ji, S. A New Spatial-Oriented Object Detection Framework for Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [CrossRef]

31. Kemker, R.; Luu, R.; Kanan, C. Low-Shot Learning for the Semantic Segmentation of Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6214–6223. [CrossRef]

32. Vinayaraj, P.; Sugimoto, R.; Nakamura, R.; Yamaguchi, Y. Transfer Learning With CNNs for Segmentation of PALSAR-2 Power Decomposition Components. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6352–6361. [CrossRef]

33. Šćepanović, S.; Antropov, O.; Laurila, P.; Rauste, Y.; Ignatenko, V.; Praks, J. Wide-Area Land Cover Mapping With Sentinel-1 Imagery Using Deep Learning Semantic Segmentation Models. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10357–10374. [CrossRef]

34. Feng, Y.; Sun, X.; Diao, W.; Li, J.; Gao, X.; Fu, K. Continual Learning with Structured Inheritance for Semantic Segmentation in Aerial Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [CrossRef]

35. Zuo, Z.; Li, Y. A SAR-to-Optical Image Translation Method Based on PIX2PIX. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Kuala Lumpur, Malaysia, 17–22 July 2021; pp. 3026–3029.

36. Chen, X.; Chen, S.; Xu, T.; Yin, B.; Peng, J.; Mei, X.; Li, H. SMAPGAN: Generative Adversarial Network-Based Semisupervised Styled Map Tile Generation Method. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4388–4406. [CrossRef]

37.   Kaiser, P.; Wegner, J.D.; Lucchi, A.; Jaggi, M.; Hofmann, T.; Schindler, K. Learning Aerial Image Segmentation from Online Maps. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6054–6068. [CrossRef]

38.   Fu, Y.; Liang, S.; Chen, D.; Chen, Z. Translation of Aerial Image Into Digital Map via Discriminative Segmentation and Creative Generation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [CrossRef]

39.   Vandal, T.J.; McDuff, D.; Wang, W.; Duffy, K.; Michaelis, A.; Nemani, R.R. Spectral Synthesis for Geostationary Satellite-to-Satellite Translation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [CrossRef]

40.   Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *arXiv* **2017**, arXiv:1609.04802.

41.   Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Loy, C.C.; Qiao, Y.; Tang, X. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. *arXiv* **2018**, arXiv:1809.00219.

42.   Choi, J.-S.; Kim, Y.; Kim, M. S3: A Spectral-Spatial Structure Loss for Pan-Sharpening Networks. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 829–833. [CrossRef]

43.   Ji, H.; Gao, Z.; Mei, T.; Ramesh, B. Vehicle Detection in Remote Sensing Images Leveraging on Simultaneous Super-Resolution. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 676–680. [CrossRef]

44.   Tang, W.; Deng, C.; Han, Y.; Huang, Y.; Zhao, B. SRARNet: A Unified Framework for Joint Superresolution and Aircraft Recognition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 327–336. [CrossRef]

45.   Shen, C.; Ji, X.; Miao, C. Real-Time Image Stitching with Convolutional Neural Networks. In Proceedings of the 2019 IEEE International Conference on Real-time Computing and Robotics (RCAR), Irkutsk, Russia, 4–9 August 2019; pp. 192–197.

46.   He, X.; He, L.; Li, X. Image Stitching via Convolutional Neural Network. In Proceedings of the 2021 7th International Conference on Computer and Communications (ICCC), Chengdu, China, 9–12 December 2021; pp. 709–713.

47.   Lin, M.; Liu, T.; Li, Y.; Miao, X.; He, C. Image Stitching by Disparity-Guided Multi-Plane Alignment. *Signal Process.* **2022**, *197*, 108534. [CrossRef]

48.   Pielawski, N.; Wählby, C. Introducing Hann Windows for Reducing Edge-Effects in Patch-Based Image Segmentation. *PLoS ONE* **2020**, *15*, e0229839. [CrossRef] [PubMed]

49.   Keelan, B. *Handbook of Image Quality: Characterization and Prediction*; CRC Press: Boca Raton, FL, USA, 2002; ISBN 978-0-429-22280-1.

50.   Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

51.   Wang, Z.; Bovik, A.C. A Universal Image Quality Index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]

52.   Goetz, A.; Boardman, W.; Yunas, R. Discrimination among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm. In *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop*; AVIRIS Workshop: Pasadena, CA, USA, 1992.

53.   Sheikh, H.R.; Bovik, A.C. Image Information and Visual Quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444. [CrossRef]

54.   Prabhu, K.M.M. *Window Functions and Their Applications in Signal Processing*; CRC Press: Boca Raton, FL, USA, 2017; ISBN 978-1-315-21638-6.

55.   Li, H.; Zhang, Y.; Gao, Y.; Yue, S. Using Guided Filtering to Improve Gram-Schmidt Based Pansharpening Method for GeoEye-1 Satellite Images. In Proceedings of the 4th International Conference on Information Systems and Computing Technology, Shanghai, China, 22–23 December 2016; pp. 33–37.

56.   Sekrecka, A.; Kedzierski, M. Integration of Satellite Data with High Resolution Ratio: Improvement of Spectral Quality with Preserving Spatial Details. *Sensors* **2018**, *18*, 4418. [CrossRef] [PubMed]

57.   Dong, X.; Sun, X.; Jia, X.; Xi, Z.; Gao, L.; Zhang, B. Remote Sensing Image Super-Resolution Using Novel Dense-Sampling Networks. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1618–1633. [CrossRef]

58.   Cui, B.; Jing, W.; Huang, L.; Li, Z.; Lu, Y. SANet: A Sea–Land Segmentation Network Via Adaptive Multiscale Feature Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 116–126. [CrossRef]