

Article

Not peer-reviewed version

---

# Data Augmentation with Generative Adversarial Network for Solar Panel Segmentation from Remote Sensing Images

---

[Justinas Lekavičius](#)<sup>\*</sup> and Valentas Gružas

Posted Date: 14 February 2024

doi: 10.20944/preprints202402.0821.v1

Keywords: deep learning; solar panels; semantic segmentation; data augmentation; generative adversarial network; remote sensing; transfer learning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# Data Augmentation with Generative Adversarial Network for Solar Panel Segmentation from Remote Sensing Images

Justinas Lekavičius \* and Valentas Gružasuskas \*

Institute of Computer Science, Vilnius University, Lithuania

\* Correspondence: justinas.lekavicius@mif.stud.vu.lt (J.L.); valentas.gruzauskas@mif.vu.lt (V.G.);  
Tel.: +370-600-72586 (J.L.); +370-685-76870 (V.G.)

**Abstract:** With the popularity of solar energy in the electricity market, demand arises for data such as precise locations of solar panels for efficient energy planning, management, and distribution. However, this data is not easily accessible and in some cases, information such as precise locations does not exist. Furthermore, existing data sets for training semantic segmentation models of PV installations are limited, and their annotation is time-consuming and labor-intensive. Therefore, for additional remote sensing (RS) data creation, the pix2pix generative adversarial network (GAN) is utilized, enriching the original resampled training data of varying GSDs without compromising its integrity. Experiments done with the DeepLabV3 model, ResNet-50 backbone, and pix2pix GAN architecture were conducted to find the optimal configuration for an accurate RS imagery segmentation model. The result is a fine-tuned solar panel semantic segmentation model, trained using transfer learning and utilizing an optimal amount – 60% of generated RS imagery for additional training data, increasing model accuracy. The findings demonstrate the benefits of using GAN-generated images as additional training data, increasing the size of small data sets, and improving the capabilities of the segmentation model for solar panel detection in RS images.

**Keywords:** deep learning; solar panels; semantic segmentation; data augmentation; generative adversarial network; remote sensing; transfer learning

## 1. Introduction

In this day and age, the use of solar energy is becoming increasingly popular due to the worldwide adoption of renewable resources. Solar panels are one of the primary means for converting solar energy to electricity and are used by homeowners and electricity suppliers. Solar panels are not only a key element to producing clean renewable energy, but they are also environment friendly and cheaper to afford than before [1]. With the increasing popularity of solar energy in the electricity market, demand arises for various data. Such data can include the locations of solar panels, their types, quantities, specifications, and power capacities. This data, precisely the locations of solar panels can be used for efficient policy-making, energy planning and distribution, and grid management. Unfortunately, this data is not easily accessible, due to either privacy concerns or unwillingness of solar panel installers to share this data. In some cases, this data does not include the precise locations of solar panel installations and only statistics such as the power capacities, which complicates the information-gathering process.

Fortunately, remote sensing data and machine learning are valuable tools in this scenario. The procedure of remote sensing, in particular by satellites, allows for measuring the characteristics of an observed area and collecting images. The collected images of the surface of the planet can then be used for observation and detection of patterns and objects, or in this case, solar panels. For analysis of visual data, convolutional neural networks (CNNs) are used and semantic segmentation is performed. Networks such as FCN, which uses end-to-end training and "skip" architecture for high-level and low-level feature combination [2], U-Net, which features a U-shaped encoder-decoder

structure and is efficient for small datasets [3], and others are often utilized. SegNet and DeepLabV3 are also commonly used, respectively either for performance-computational efficiency [4] or better multiscale object segmentation [5], with varying results. The current state-of-the-art in utilizing deep learning for solar panel detection spans various methods and tools, such as specific CNNs and ViTs, as well as their variations and iterations. RU-Net, which is a combination of ResNet34 U-Net, Atrous Spatial Pyramid Pooling, and Dense Conditional Random Field deep learning components, has been utilized for efficient solar panel detection performance and identifying rooftop solar panel locations, distributions, and surface areas in 0.3m/pixel spatial resolution images [6]. The proposed model demonstrates higher precision metrics when compared to classical models such as FCN and U-Net. A hierarchical information extraction method for solar panels using multi-source satellite remote sensing images was also proposed, tested in three selected provinces in China, and using the EfficientNet-B5 model for scene classification and the U2-Net model for the precise location of solar power panels, with the goal and purpose to locate as many solar panels as possible, reducing the number of false positives [7]. Mask R-CNN deep learning algorithm was used to identify solar photovoltaic panels in remote sensing images, and this method focuses on isolating solar panels from background objects [8]. While the usefulness of machine learning algorithms in object isolation and image segmentation is noted, the model was trained on a small dataset, featuring images taken with the same equipment under the same environmental conditions.

For remote sensing image segmentation, general-purpose vision transformers have also been used, with varying results. Models such as SegFormer and Lawin Transformer demonstrated results comparable to other state-of-the-art models such as FCN and RemoteNet, however, the results were achieved with careful image pre-processing, including cropping, contrast enhancement, and geometric transformations [9]. For building segmentation in remote sensing images, various models such as TransUNet, MiTNet, UNetFormer, and Segformer have been compared, presenting varying metrics such as F1 score, overall accuracy and mean intersection over Union when tested on Global Cities WHU Satellite dataset [10].

Semantic segmentation models and vision transformers also benefit from transfer learning, i.e. the process of transferring the learning knowledge from one task to another. Such models like ResNet, FCN, DeepLab, and more were either pre-trained on the ImageNet dataset or used weights from other models that were pre-trained on ImageNet [11,12]. Furthermore, transfer learning has also been used specifically for training remote sensing image semantic segmentation models. For high-resolution remote sensing images, in particular, an improved U-Net model based on transfer learning was proposed, showcasing better results for vehicle semantic segmentation when compared to the regular U-Net model [13]. For multiobject segmentation of remote sensing images when the issue of insufficient labeled data and imbalanced data classes is present, transfer learning has also proved to be beneficial in improving the semantic segmentation model performance to some degree [14].

Even though the RS images of solar panel installations are obtained, large amounts of annotated data are needed for model training, with even more benefits brought by larger and more diverse datasets. Complications due to a lack of raw data, difficulty in data set annotation, and limitations of sensor characteristics make this more challenging, as a large variation and amount of data are needed for creating an effective model [15]. Therefore, data augmentations are essential to expanding the data set for the machine learning model, especially when training a semantic segmentation model for object detection. Basic augmentations include, although are not limited to, flipping images, rotating and tilting them, and adjusting contrast or colors, to generate more images from already existing ones. Performing rotations and horizontal flipping is one of the possible classic data augmentations that can be performed on both the original image and its semantic segmentation mask, and the result is improved prediction performance of the segmentation network [16]. These augmentations are especially significant for small-scale data sets and their extension. Cropping, i.e. taking only a subset of an image is also a method for producing more samples, and the corresponding mask has to be cropped in the same way. Training data for semantic segmentation can also be increased by scaling,

brightness, and contrast adjustment of the input image for higher classification performance [17]. Some augmentations attempt to fix specific problems, for example, brightness adjustment addresses the issues of lighting changes, while cropping and zooming of images manage scaling and background issues [18]. Augmentations are needed for extending data sets when training deep learning models, and this methodology is not only beneficial when the data set is small, but also in some cases reduces model overfitting, i.e. when the model performs too well with training data, and poorly with new, unseen data. However, basic data augmentations may produce unnatural results when not used properly, as the model may have issues labeling objects in images that are too distorted or have their colors drastically changed. Therefore, to satisfy the demand for more realistic data from already limited data sets, generative adversarial networks are beneficial in this case.

Generative adversarial networks can be used for transforming images from one domain to another. This is done by utilizing generative networks, which aim to create believable images from a source domain, and discriminator networks, which are trained to identify whether the created images are genuine or created by the generator [19]. The rivalry of the generator and discriminator plays a key part in generative adversarial network training, ensuring that the generated synthetic images are believable compared to the original images. Generative adversarial networks have also been used for remote sensing image reconstruction. Conditional discriminator PatchGAN was used for remote sensing imagery super-resolution, i.e., for creating images that have not only high fidelity but high perceptual quality as well [20]. The upscaled data is more realistic and features higher fidelity, and when compared to a simulated dataset, the evaluation metrics of state-of-the-art algorithms such as SRGAN, SRCNN, and SRResNet are higher, particularly peak signal-to-noise and structural similarity. In the context of paired image-to-image translation, pix2pix is a type of generative adversarial network suitable for synthesizing remote sensing images from labels, e.g., semantic segmentation masks of solar panels. First introduced in 2018, it is an approach effective at synthesizing photos from label maps [21]. Several experiments have been conducted using this solution, such as generating maps from aerial photos [22] and architectural facades from labels [23]. Considering the potential of generating new material from an already existing limited data set, the accuracy of the trained semantic segmentation model can be considerably improved with fake but realistic additional remote sensing images when compared to simply utilizing basic data augmentations. Therefore, using GAN-generated data alongside the original training data can be an effective method to improve solar panel installation detection in RS images.

The main aim of the work is to enhance the model for solar panel installation semantic segmentation, utilizing pix2pix GAN for data augmentation to get the best results with a limited amount of data. To achieve this goal, a series of experiments are performed, utilizing a selected data set and algorithms, transfer learning, fine-tuning, and data augmentations using both basic augmentations and generative adversarial network-generated data. The experiments reveal the results of solar panel semantic augmentation without and with data augmentations and the optimal amount of synthetic data needed for better results. Inspecting the training process, parameter optimizations, various criteria, and combinations can determine the best results (based on metrics such as accuracy, precision, recall, etc.). This, combined with the findings and results of several experiments, can make solar panel identification much more efficient and less labor-intensive. The final result is a solar panel semantic segmentation model, fine-tuned for the task of solar panel segmentation in remote sensing images, and trained with data augmented using the pix2pix GAN for enhanced solar panel detection accuracy.

## 2. Materials and Methods

The data used for semantic segmentation and GAN model training is a collection of five solar panel aerial image data sets. The Provincial Geomatics Center of Jiangsu provides three data sets [24], and two are sourced from Google Earth and the French National Institute of Geographical and Forestry Information (IGN) [25]. The data sets are presented and compared in Table 1. The main differences are the types of files, the ground sampling distance (or GSD, i.e., the distance between centers of two

neighboring pixels measured on the ground), and image resolutions. For DeepLabV3 and pix2pix GAN training, 640 image-mask pairs of each GSD were used, a total being 2560 images and 2560 semantic segmentation masks. Out of 640 image-mask pairs of each GSD, 80% are used for training of DeepLabV3 model and pix2pix GAN, while for validation and testing of the DeepLabV3 model, 10% are used for each.

**Table 1.** A comparison of used solar panel semantic segmentation data sets.

Data Source	Image Format	Ground Sampling Distance	Image Resolution
Provincial Geomatics Center of Jiangsu	BMP	0.8m/pixel	1024x1024
Provincial Geomatics Center of Jiangsu	BMP	0.3m/pixel	1024x1024
French National Institute of Geographical and Forestry Information	PNG	0.2m/pixel	400x400
Provincial Geomatics Center of Jiangsu	BMP	0.1m/pixel	256x256
Google Earth	PNG	0.1m/pixel	400x400

To solve the issue of GSD and image resolution differences, resampling is applied for all images and their masks. In the experiments, a target GSD of 0.1m/pixel and a target image resolution of 512x512 are used for resampling. The reason for this is to keep as much detail in the resampled images as possible while retaining computational efficiency, as well as data uniformity so that the scale of solar panels is accurately represented. Bringing all images to the same "centimeters per pixel" ratio ensures a more accurate actual scale of solar panels in remote sensing images. The process for resampling image width and height to target GSD is presented in Equation 1 and Equation 2.

$$resampled\_width = original\_width \times \left( \frac{spatial\_resolution}{target\_spatial\_resolution} \right) \quad (1)$$

$$resampled\_height = original\_height \times \left( \frac{spatial\_resolution}{target\_spatial\_resolution} \right) \quad (2)$$

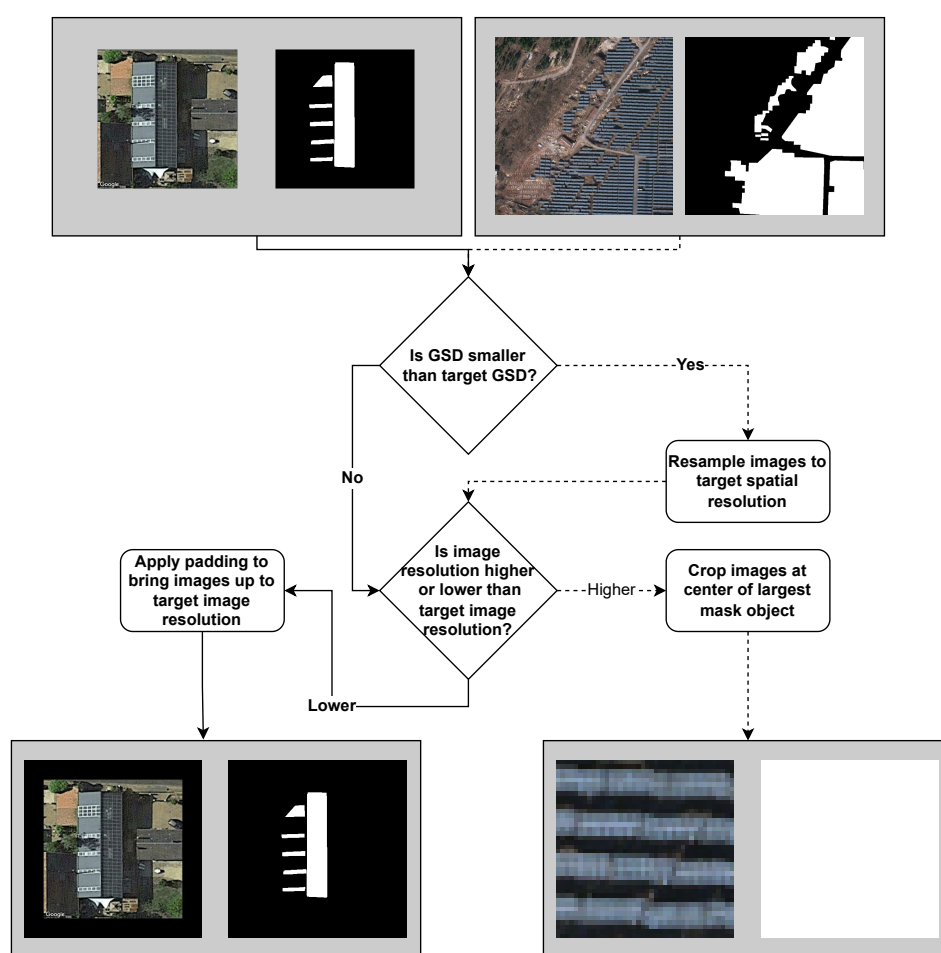
The result is that the images of the original 0.1m/pixel GSD are not resampled to a new image resolution, as they are already of the target GSD, and the image-mask pairs of 0.8m, 0.3m, and 0.2m are resampled. Lanczos resampling is used for remote sensing image quality preservation during upsampling, although at a higher computational cost. In contrast, nearest neighbor resampling is used for the binary semantic segmentation masks to avoid artifacts and retain the sharp edges of the mask objects. After GSD resampling, the image-mask pairs are then resampled to the target image resolution of 512x512. This is done by either cropping or padding the image-mask pairs. The process for padding image width and height is presented in Equation 3 and Equation 4, and is applied with black pixels to both the image and segmentation mask (horizontal left-right padding and vertical top-bottom padding). The padding is performed for image-mask pairs that are of the same target GSD, but lower image resolution to bring the images and masks to the target image resolution without compromising the performed GSD resampling or the quality of the original remote sensing images.

$$padding\_width = \left\lfloor \frac{target\_width - resampled\_width}{2} \right\rfloor \quad (3)$$

$$padding\_height = \left\lfloor \frac{target\_height - resampled\_height}{2} \right\rfloor \quad (4)$$



Alternatively, cropping is applied when the image-mask pair image resolution is higher than the target resolution after GSD resampling. However, this introduces the issue of information loss when cropping a large image. Therefore, instead of applying center cropping, the procedure of identifying the largest segmentation object in the mask is performed, and the cropping is done to focus on the object. This ensures that the cropped image and mask will always include a solar panel. Also, this is done not just because the segmentation mask may locate multiple objects, but because they may be of different sizes. Firstly, the number of solar panel objects in a semantic segmentation mask is calculated. Then, for each object, the bounding box slices are located and the slice with the largest area is selected for coordinate extraction. The center of the largest object's bounding box is then calculated, and the coordinates for image-mask pair cropping are calculated. Also, consideration is taken that the cropping dimensions do not exceed the actual image's dimensions and that the cropping does not occur beyond the boundaries of the image (for example, if the largest segmentation object is in the corner of the image). The resampling process is illustrated in Figure 1.



**Figure 1.** The process of image-mask pair resampling to target spatial and image resolution.

For solar panel semantic segmentation model training, DeepLabV3 architecture with ResNet-50 backbone was utilized. The DeepLabV3 pre-built model from PyTorch was chosen as it is one of the more recent semantic segmentation convolutional neural networks, and its complexity, as well as capabilities, are suitable for the task of PV installation segmentation in remote sensing images of various GSDs. ResNet-50 was selected as the backbone over other options such as MobileNet and ResNet-101 due to the desired balance of accuracy and computational efficiency. The hyperparameters used for model training are detailed in Table 2 and were not changed across runs. The parameters were kept the same throughout all experiments, ensuring that the benefits of data augmentations

were observed instead of different parameter optimizations. Early stopping is implemented to stop model training after 10 epochs if the target metric stops improving when compared to the average target metric. In this case, because the highest model accuracy is desired, the target metric is validation intersection over union (IoU). Therefore, during training, if the validation IoU metric stops improving and is lower than the average validation IoU for consecutive 10 epochs (in this case, the number of epochs is "patience"), early stopping is initiated, halting the training.

**Table 2.** Hyperparameters used for DeepLabV3 semantic segmentation model training.

Epochs	Early Stop Patience	Adam Optimizer		Scheduler StepLR	
		Learning Rate	Weight Decay	Step Size	Gamma
100	10 epochs	0.001	0.0001	20	0.1

After training, the DeepLabV3 model for solar panel segmentation is tested using the testing subset, and the average evaluation metrics are calculated (Table 3). To evaluate the trained model, average accuracy, precision, recall, F1 score, and intersect over union are calculated. In this case, accuracy refers to pixel accuracy, i.e. the correctly classified pixels when comparing the ground truth mask and predicted mask. This includes correctly predicting not only the pixels where the solar panel is segmented (white pixels) but also the background (black pixels). While this provides a good insight into pixel-wise correctness, the IoU metric is arguably more relevant in testing the model accuracy, as the IoU metric indicates how well the predicted mask overlaps with the ground truth. F1 score is a useful metric when there is a class imbalance (foreground and background), as it is the harmonic mean of precision and recall. Because of the issue of class imbalance (smaller percentage of white pixels, i.e. the PV objects, when compared to black background pixels), inspecting the IoU and F1 metrics provides the most insight into the model accuracy. Furthermore, the semantic segmentation capabilities of the model are tested by counting the correctly segmented images, poorly segmented images, and unsegmented images. Correctly segmented images are considered those with an IoU metric higher than or equal to 0.5, while poorly segmented images are those that have an IoU metric lower than 0.5, but not equal to 0. Unsegmented images are those having the IoU metric of 0.

**Table 3.** Metrics used for evaluation of trained semantic segmentation DeepLabV3 models using testing dataset.

Metric	Formula
Accuracy	$\frac{1}{N} \sum_{i=1}^N I(\text{predicted}_i = \text{true}_i)$
Precision	$TP / (TP + FP)$
Recall	$TP / (TP + FN)$
F1 Score	$2 \times (Precision \times Recall) / (Precision + Recall)$
Intersection over Union	$\sum_{i=1}^N I(y_{\text{true},i} \cap y_{\text{pred},i}) / \sum_{i=1}^N I(y_{\text{true},i} \cup y_{\text{pred},i})$

For data augmentation with the generative adversarial network, pix2pix implementation was used. Experimenting with different setups and parameters, an optimal setup of parameters for training the pix2pix GAN for image-to-image translation from domain A (binary semantic segmentation mask) to domain B (remote sensing image) was determined. U-Net256 generator architecture was used, therefore the images and masks were resized to 256x256 resolution. For the discriminator, PatchGAN architecture with 3 convolutional layers was used. The lambda\_L1 parameter was set to 75 instead of the default value of 100, This is done to reduce the importance of L1 loss, which encourages the generator to produce images closer to the original input data, as the lambda\_L1 value is used in the training objective. The reduction of lambda\_L1 value in this case encourages generating more realistic outputs. The pix2pix GAN was trained for 600 epochs – 300 epochs of training with a consistent learning date of 0.0001, and 300 epochs later, gradually decaying the learning rate to closer to zero.

The number of channels for input and output images (input\_nc and output\_nc parameters) were set to 1 and 3 respectively (1 for input grayscale semantic segmentation masks and 3 for output RGB remote sensing synthetic generated images). To maintain the balance of power between the generator and discriminator, and so that the former does not overpower the latter and vice versa, the number of discriminator filters in the first convolutional layer (ndf parameter) was set at 64 (default value), while the number of generator filters in the last convolutional layer (ngf parameter) set at 128 (default value is 64). This allows the generator to capture more detail and generate more convincing images. A batch size of 1 was maintained for higher stability and context preservation when only one image-mask pair is used for mapping learning at a time.

To compare the benefits of using classic data augmentations for the training dataset versus using GAN-generated data, the DeepLabv3 semantic segmentation model was also trained with basic augmentations performed on the training dataset. In the case of training a semantic segmentation model with RS images of PV installations, it is crucial to ensure that the performed augmentations are not too drastic and produce realistic data. For example, because of the nature of how solar panels are usually installed and positioned, drastic image perspective alterations may result in unrealistic data and influence the model training process negatively. The regular data augmentations performed were random horizontal flip with 50% chance of it being applied, random rotation of 5 degrees, random perspective change with 0.05 distortion scale and 50% chance of it being applied, and random application of Gaussian blur (5x5 kernel size and standard deviation of 0.1 min and 2.0 max) with 50% change of it being applied. The Gaussian blur application is intended to simulate the fogging of a satellite lens.

The RS images are also normalized to mean (0.485, 0.456, 0.406) and standard deviation (0.229, 0.224, 0.225), based on ImageNet common values. However, in the image-mask pairs, only the image is normalized, as the segmentation mask does not require it. Furthermore, Gaussian blur is applied only to the RS image and not its mask, to avoid artifacts. Other augmentations, such as horizontal flip, rotation, and perspective change are applied both to the image and its mask, and the usage of manual seed ensures that although the augmentations are chosen at random, they are applied identically.

The experiments with the DeepLabV3 semantic segmentation model were performed in six ways, and they are as follows:

- Training the model without performing basic data augmentations
- Training the model with performed basic data augmentations
- Training the model with additional 25% GAN generated training data and without performing basic data augmentations
- Training the model with additional 25% GAN generated training data and with performed basic data augmentations
- Training the model with the additional optimal amount of GAN-generated training data (60% in this case) and without performing basic data augmentations
- Training the model with the additional optimal amount of GAN-generated training data (60% in this case) and with performed basic data augmentations

It is important to note that for the third and fourth scenarios, 25% of the remote sensing imagery generated using the generative adversarial network is used as a proof-of-concept that the trained semantic segmentation model benefits from additional synthetic remote sensing image-mask pairs. In contrast, the fifth and sixth experiment variants display further improved performance of the semantic segmentation model trained with an optimal amount of GAN-generated data.

For each experiment, the model is trained in four scenarios. In the first scenario, the model is simply trained from scratch without pre-trained weights. This serves as a baseline for comparison with other scenarios, as well as shows how a model trained from the ground up for the specific dataset performs in comparison to fine-tuned models and the ones that utilize transfer learning. In the second scenario, transfer learning is applied. The model is trained using weights that were trained on the Common Objects in Context (COCO) subset using 20 categories in the Pascal VOC dataset, such as



bicycle, aeroplane, bottle, dining table, and more. While solar panels are not one of these 20 categories, the learned general features like shapes and edges are still beneficial, providing a good foundation for quicker convergence and better generalization to new data. The third scenario involves fine-tuning the model from the second scenario by freezing all layers except the final one. This way only the weights of the last layer are updated, and the other layers are frozen, better adapting it for the task of solar panel installation semantic segmentation in remote sensing images. At the same time, features learned by the earlier layers are retained as they are frozen. Using previously learned knowledge, this approach can be beneficial in reducing overfitting and training times. In the fourth scenario, the model from the third scenario is fine-tuned by unfreezing all layers and training the entire network. This is done to build upon the previous scenario, allowing the entire model to be fine-tuned more extensively for the best performance. The approach of incrementally fine-tuning the model that utilized transfer learning is a way to develop the deep learning model for the specific task of solar panel semantic segmentation and make it even more adapted to this task. The result is a total of 24 experiments. When experiments are performed with additional remote sensing images generated using the pix2pix generative adversarial network, the additional samples are only used for the training dataset, while the validation and testing data sets are kept intact. This is done to keep the validation and testing subsets consistent across all experiments and for more objective evaluations. Furthermore, the validation and testing data consists of real-life data that is of high quality, and the synthetic remote sensing images, although realistic to a degree, are less suitable for testing and validating the model.

### 3. Results

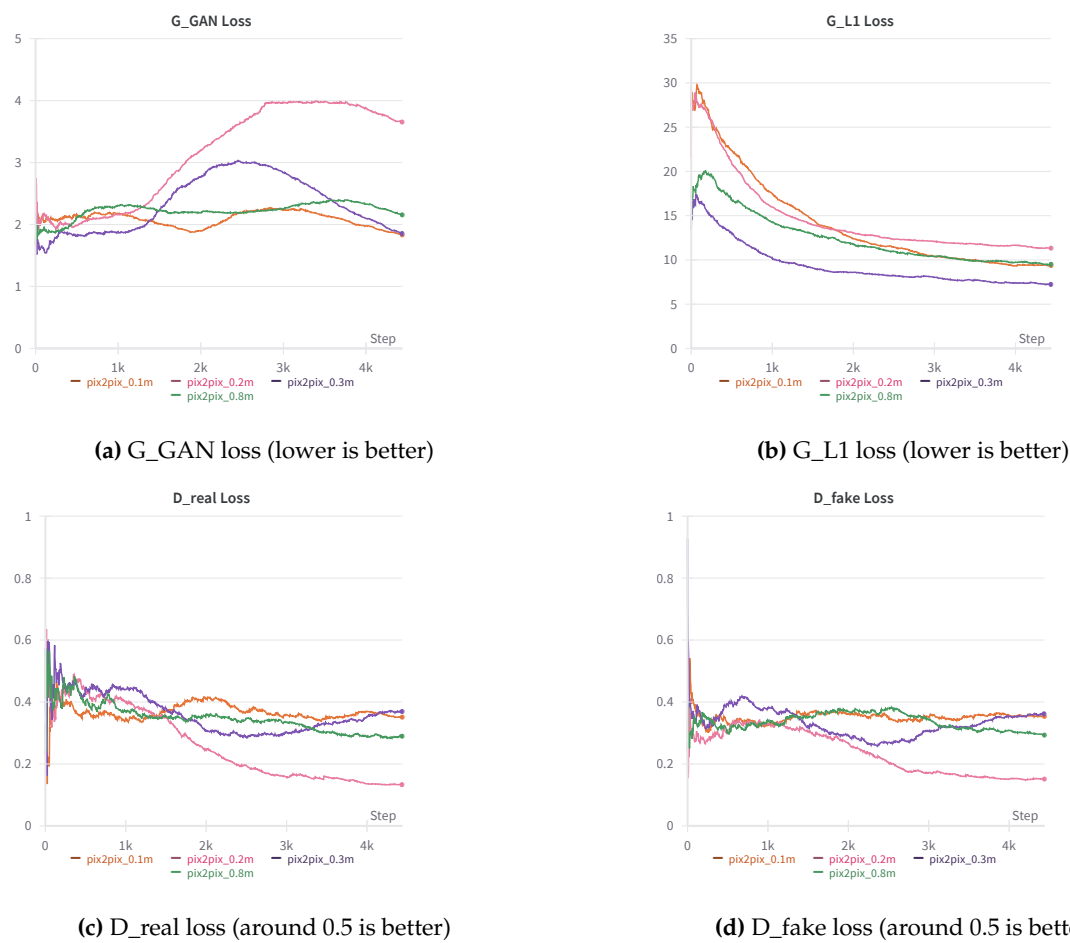
The experiments were performed in a Google Colab environment, and the model training was done on an NVIDIA A100 GPU. To balance computational efficiency and time resources, 640 image-mask pairs were randomly selected from the data set for each GSD (total of 2560 pairs), and an 80/10/10 split was used for training, validation, and testing data. The result is 512 image-mask pairs of each GSD used for training, 64 image-mask pairs for validation of the model, and 64 image-mask pairs for the testing of the model (a total of 2048, 256, and 256 pairs respectively). Because there are two datasets of 0.1m/pixel GSD, 256 image-mask pairs are used from each to make up a total of 512 image-mask samples for the 0.1m GSD training dataset, 32 pairs from each for validation, and 32 pairs from each for testing. To ensure reproducibility and consistent sampling across runs, as well as consistent shuffling and randomizing when running the code multiple times, the random seed is set to 35 and applied to PyTorch, random, and NumPy modules. Where needed, GAN-generated image-mask pairs were appended to the training data set. For the data loaders, a size batch of 48 was used and 12 workers. During data loader creation, the image-mask pairs were resampled to a target GSD and target image resolution of 0.1m/pixel and 512x512 respectively, attempting to retain as much information and image quality as possible, and ensuring scale and "centimeter per pixel" ratio consistency.

#### 3.1. Pix2pix GAN Training

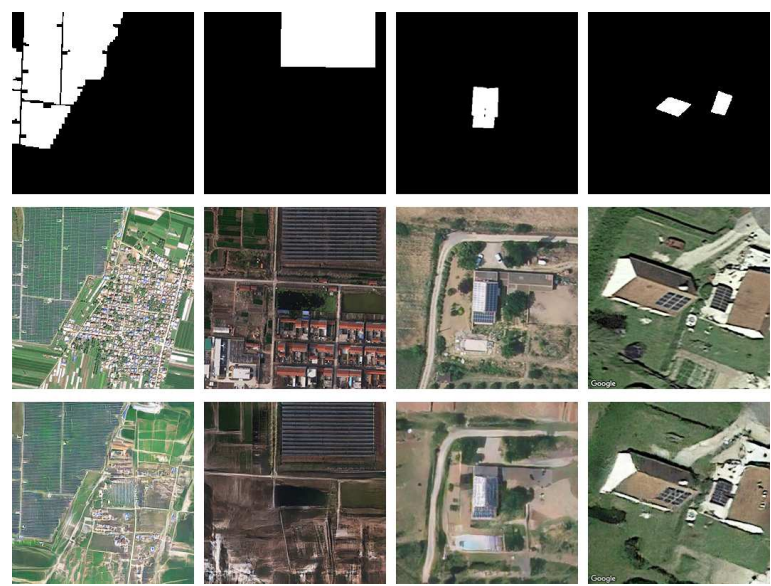
Before performing the main experiments and training the semantic segmentation model variants, the pix2pix GAN was trained for remote sensing data augmentation. To best fit the task of generating new remote sensing images from binary semantic segmentation masks, different parameter combinations were tested, such as the generator and discriminator parameters for balance (so that the discriminator does not overpower the generator, and vice versa), number of epochs for training, etc. For training, 512 image-mask pairs of each GSD were used, and four separate pix2pix GAN models were trained (one for each GSD), to generate remote sensing images of different GSDs and different solar panel scales. Most importantly, the image-mask pairs are the same ones that are used in the original dataset for DeepLabV3 semantic segmentation model training, so that new data from already existing image-mask pairs is generated.

The training progress was closely examined and visualized in graphs using the Weights & Biases API, detailing the changes in generator (G\_GAN and G\_L1) and discriminator (D\_real and D\_fake) losses, which are displayed in Figure 2. The desired outcome is for G\_L1 loss to be as low as possible, indicating the generated image's closer resemblance to the original data, and G\_GAN loss to decrease over time, indicating the generator's ability to learn the mappings between domains A and B more effectively and generate more convincing images that are challenging for the discriminator to evaluate. Looking at the graphs, observations can be made that training with remote sensing images of 0.8m/pixel and 0.3m/pixel GSDs display lower G\_L1 loss, signaling the generated image's close resemblance to the source material, however, still creating new and convincing results. Furthermore, the discriminator losses reach values closer to 0.5, indicating a fair challenge for the discriminator. This is also visible when inspecting the generated images from masks that originally belonged to remote sensing images of 0.8m/pixel and 0.3m/pixel GSDs. These images also originally contain fewer details when compared to images of higher GSDs, and this may be an explanation of why the generative adversarial network performed better with these images, i.e. the lower difficulty of recreation and reconstruction. The training process visualization with images of 0.2m/pixel and 0.1m/pixel GSD shows that in the case of training with 0.2m GSD images, the G\_GAN loss increased over time, and the G\_L1 final loss was the highest among all four. This may be due to the nature of these images, i.e. them having more fine details that the generator had difficulty recreating, consequently not fooling the discriminator. There are fluctuations in discriminator losses, particularly in the 0.2m/pixel GSD dataset which lowers to values less than 0.2. As established, this is likely due to the nature of these images, i.e. having more details such as roads, buildings, vehicles, etc. that are harder to replicate convincingly. The final result confirms this, with the newly generated images appearing less realistic on closer inspection, but fairly convincing when looked at from afar. Nevertheless, when compared to images generated from 0.8m and 0.3m GSD segmentation mask data, they can be considered lower quality, in some cases featuring unrealistic road formations or building shapes, although retaining mostly correct solar panel installation generating. In worst cases, the solar panel installations are not generated convincingly as well, resulting in inferior quality samples. Examples are noisy formations, distorted shapes, and inconsistent colors.

After the pix2pix GAN training is complete, the same semantic segmentation masks that were used for GAN training are used for model testing, i.e., generating new image-mask pairs. For each GSD, 512 binary semantic segmentation masks are used to generate respective remote-sensing images, resulting in new synthetic data. The final output quality varies based on the binary segmentation mask and the GSD of images used originally for testing. Upon visual inspection, it can be determined that the synthetic remote sensing images closely resemble the original data, however with subtleties that can differentiate them. As illustrated in Figure 3, the generated images feature realistic solar panel installations and sufficiently believable environments around them. Upon closer inspection, details such as roads leading to nowhere or inconsistent building layout can be observed, however, the most important aspect, i.e. the solar panels are generated in satisfactory quality. There are some samples, however, that are of worse quality, particularly when the original semantic segmentation masks feature small objects. In this case, the solar panels are generated with artifacts such as noise and different colors. Nevertheless, the number of inferior-quality samples compared to satisfactory-quality samples is insignificant. In total, 2048 new remote sensing images were generated, i.e. 512 images for each GSD.



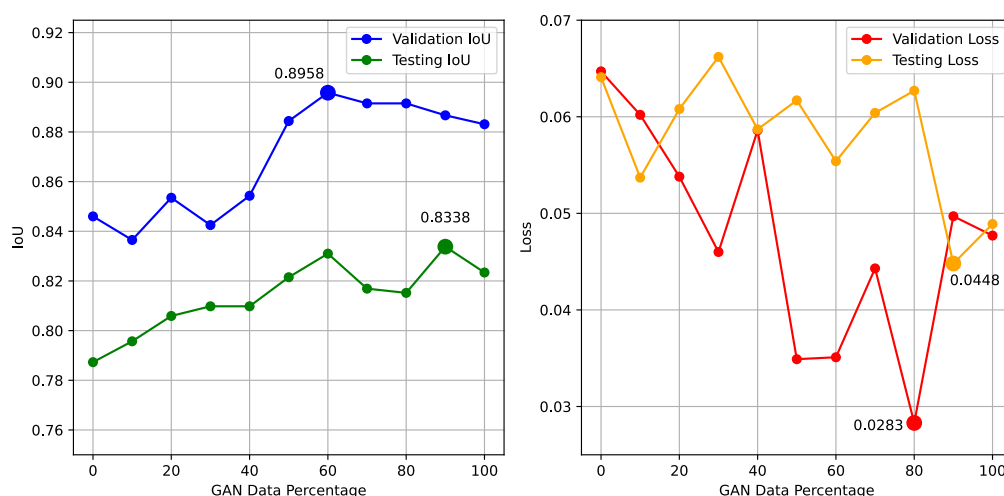
**Figure 2.** Line graphs displaying changes in pix2pix losses during training with data sets of various spatial resolutions. The highest GSD images (0.1m/pixel) have the most detail, and the lowest GSD images (0.8m/pixel) have the least detail.



**Figure 3.** An example of binary semantic segmentation mask (**top row**), original remote sensing image (**middle row**), and GAN-generated image (**bottom row**) using pix2pix trained model for each GSD respectively, from left to right: 0.8m, 0.3m, 0.2m, 0.1m.

### 3.2. Sensitivity Analysis

To find the optimal amount of GAN-generated data for usage as additional training data, sensitivity analysis was performed. This was done to examine how much additional data is needed to get the best results before the model stops benefiting from extra samples, and in the worst case, starts to overfit. To determine the best threshold of GAN data usage, the DeepLabV3 semantic segmentation model was trained 10 times using transfer learning and with an additional 10% of generated remote sensing images incrementally added to the training data set (10% added, 20% added, and so on). This was compared with the baseline of 0% additional data (the results of when the model was trained with transfer learning, and with original data without basic augmentations). For sensitivity analysis, validation, and testing subset IoU and loss metrics were compared, and the changes with additional GAN data usage are detailed in Figure 4.



**Figure 4.** Sensitivity analysis of DeepLabV3 model training using transfer learning and various percentages of additional GAN generated training data, comparing validation and testing IoU (left) and loss (right) metrics.

Observing the data in the figure reveals that the best validation and testing IoU values are when 60% and 90% additional GAN data is used respectively. Furthermore, the lowest loss values of validation and testing subsets are in the 80%-90% range. This indicates that when the percentage of additional images generated by the GAN is between 60% and 90%, the best results are achieved. In this case, because the desired outcome is the best model semantic segmentation accuracy, the IoU values are more relevant in the sensitivity analysis. Upon visual inspection, it can be observed that the peak of validation IoU is at 60%, while the peak of testing IoU is at 90% (IoU being 83.38%), although it is barely higher than it was at 60% (IoU being 83.10%). After the 60% threshold, the IoU values generally start to decrease (except the peak testing IoU at 90%), indicating potential overfitting as the model starts losing the ability to generalize to new data.

Based on the findings of sensitivity analysis, it was decided that 60% of additional GAN-generated remote sensing images for the training dataset is the optimal amount for more beneficial model training. This means that an additional 307 remote sensing images and their respective masks will be used for each GSD, a total being 1228 additional image-mask pairs added to the original training data set of 2048 image-mask pairs.

### 3.3. Solar Panel Semantic Segmentation Results

To compare the final results of all six scenarios, the fourth iteration of the trained models (using transfer learning and fine-tuned remaining layers after previously fine-tuning only the final one) are evaluated, as they are arguably the most optimized for the solar panel installation semantic

segmentation task. The training and testing results of models are inspected in all six scenarios, i.e. training without augmentations (abbr. no\_aug), training with basic augmentations (abbr. basic\_aug), training with additional 25% of GAN generated remote sensing images (abbr. gan25), training with additional 25% of GAN generated remote sensing images plus basic data augmentations for training data set (abbr. gan25\_aug), training with optimal amount (60%) of GAN generated remote sensing images (abbr. gan60), and training with optimal amount (60%) of GAN generated remote sensing images plus using basic data augmentations (abbr. gan60\_aug).

Looking at the testing results across all six scenarios displayed in Table 4, the gan60 scenario (training the model with 60% additional GAN remote sensing image data) shows the best metrics. When comparing with results of training without any augmentations, the benefits of using the generative adversarial network for additional training data synthesis is evident, especially when comparing the improvements of metrics such as average precision, average recall, average F1 score, average IoU, and average loss. Furthermore, more images are successfully segmented. Compared with the scenario basic\_aug, which features only the usage of basic data augmentations, the improvements, and benefits are still visible, especially in better IoU and loss metrics.

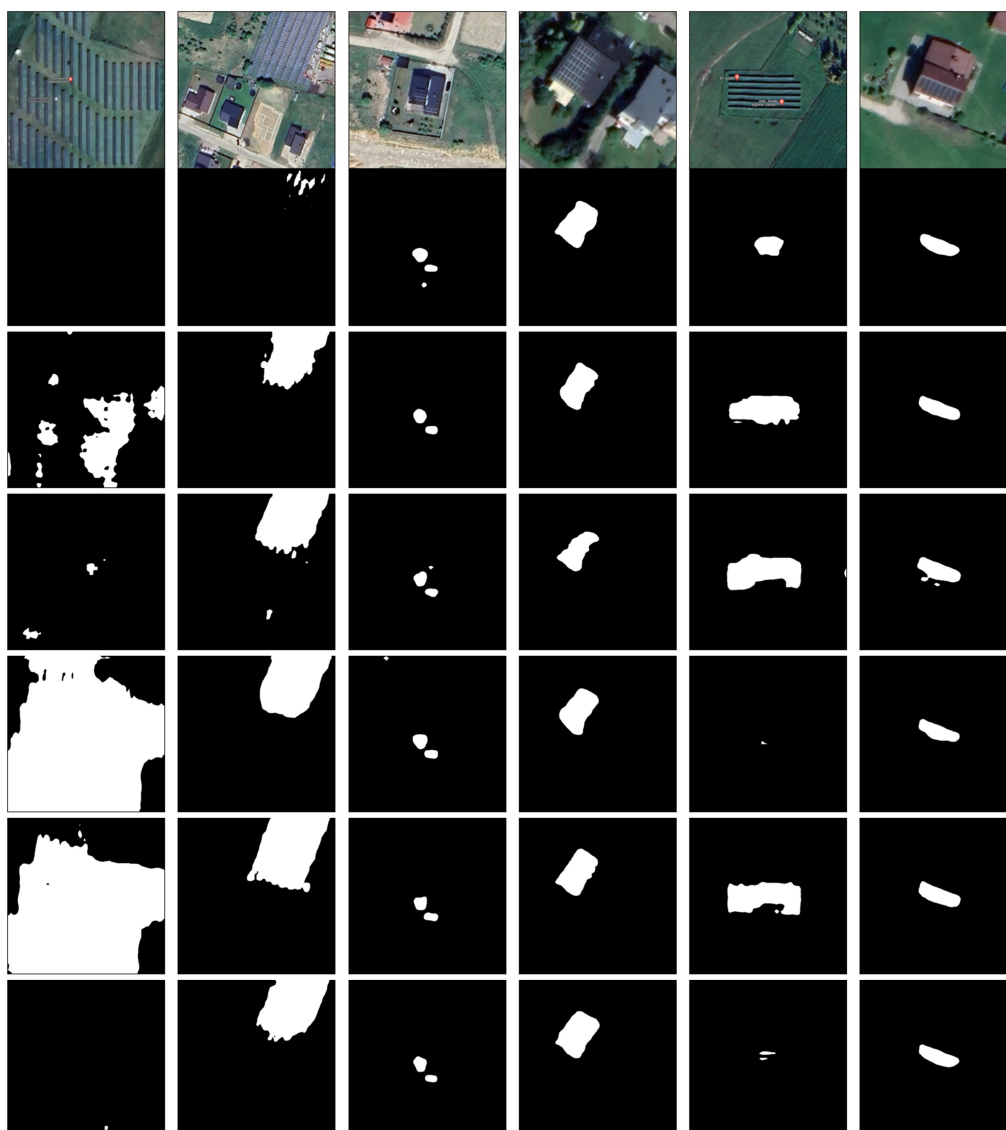
**Table 4.** Comparison of testing results for all six scenarios, with transfer learning and fine-tuning applied. The best values are written in bold.

Experiment	Avg	Avg	Avg	Avg	Avg	Avg	Segmented Panels (IoU)		
	Acc (%)	Prec (%)	Rec (%)	F1 (%)	IoU (%)	Loss	Correct ≥ 0.5	Poor < 0.5	None = 0
no_aug	97.89	86.72	85.62	85.25	80.13	0.0650	229	12	15
basic_aug	97.88	89.63	86.51	86.50	81.32	0.0547	235	13	8
gan25	98.09	89.11	85.94	85.71	80.42	0.0586	229	16	11
gan25_aug	97.91	88.84	87.25	87.08	81.41	0.0550	238	8	10
gan60	<b>98.67</b>	<b>90.13</b>	<b>88.11</b>	<b>87.96</b>	<b>83.32</b>	<b>0.0368</b>	<b>237</b>	<b>11</b>	<b>8</b>
gan60_aug	98.04	89.82	87.69	87.77	82.90	0.0611	238	9	9

Based on the model testing results, it can be determined that by training the semantic segmentation model while increasing the original training data set by 60% using synthetic remote sensing images generated by the generative adversarial network pix2pix, the best result is achieved. Applying additional basic image augmentations did not provide benefits based on the testing results, and in fact, displayed slightly worse results. This is likely due to the already challenging nature of generated synthetic remote sensing images, which although look realistic and believable, still contain some noise and artifacts in some of the samples.

The final trained semantic segmentation model is capable of segmenting solar panel installations at different scales, shapes, and shades. The model was tested with RGB images taken from Google Maps. The images consist of random locations throughout Lithuania, containing small solar panel installations, as well as solar power stations. The original images and their segmentation masks are displayed in Figure 5. Monocrystalline and polycrystalline solar panel installations are generally well detected, as they are also distinct at various scales due to their grid lines and rectangular shape. When tested with images of solar power stations, the model either predicted the array of solar panels as a single object or the entire semantic segmentation mask was white, depending on the scale.





**Figure 5.** An example of several RS images of locations throughout Lithuania (**row 1**), and solar panel segmentation masks, respectively using a model trained without augmentations (**row 2**), with basic augmentations (**row 3**), using 25% of additional GAN training data (**row 4**), using 25% GAN data + basic augmentations (**row 5**), using 60% of additional GAN training data (**row 6**), and using 25% of additional GAN training data + basic augmentations (**row 7**).

#### 4. Discussion

Using the pix2pix generative adversarial network for data augmentation, not only the accuracy of the semantic segmentation model is improved, but also the issue of data labeling is addressed. Because new images are generated from already existing data, this can be an alternative to manual annotation, a time-consuming and labor-intensive process when more diverse data is needed. For instance, data sets such as thermal images and their respective segmentation masks could be expanded with new synthetic data, especially when expertise in the field of photovoltaic farm fault detection is needed [26]. Furthermore, this can also be applied to the improvement of solar farm capacity estimation, either as an alternative or an additional solution to exploring other data sources [27]. Although classic data augmentations such as contrast adjustments, random rotations, and flips are utilized, brand-new remote sensing images may benefit the process of solar farm detection and energy generation capacities with even more potential in terms of accuracy.

Because the focus of this work is the improvement of solar panel segmentation from RS imagery using GAN-based data augmentations instead of segmentation model architecture optimizations and improvements, this method may be potentially combined with other segmentation solutions, designed specifically for PV installation detection. Other works propose new models as improvements in solar panel segmentation, such as better detection of small-scale installations in the form of a size-aware network [28], and note the potential of even better applications with broader data sources. However, the performance of various semantic segmentation models may also depend on the nature of the training data, demonstrated by the comparison of U-Net, DeepLabv3+, PSPNet, and FPN architectures and the fact that U-Net outperformed the newer DeepLabv3+ architecture [29]. Likewise, the problem of the limited amount of samples is also mentioned, although mitigated to an extent with two classic augmentations, i.e. random horizontal and vertical flips with 50% probability. Nevertheless, although these augmentations introduce variety to the data set, the study would likely benefit from an even more diverse training data set, featuring newly generated images using the generative adversarial network. This would, however, depend on the nature of the original data set, and potential points of caution such as data quality and class imbalances.

Although the result of this work is the successful improvement of the semantic segmentation model using additional data generated using the GAN, it has the potential to be improved even further with additional computational and time resources. For future work, if the issue of resource limitations was absent, a more detailed sensitivity analysis could be performed. With more time and computational resources, a more thorough sensitivity analysis with stochastic simulations and application of the Central Limit Theorem can be performed for more valid and consistent results. Running the sensitivity analysis experiments multiple times (e.g. 30 runs) would be beneficial for determining an even more accurate optimal amount of generated synthetic remote sensing image to use for model training more accurately. Furthermore, the issue of generated synthetic remote sensing images' inferior quality when the input semantic segmentation masks have class imbalances (a lot of black background and small white objects) can potentially be addressed with pix2pix optimizations for this specific task. The optimizations would include the learning rate and training epoch amount adjustments, tuning the generator, and discriminator parameters such as the number of filters. Although this would likely result in even longer and more computationally intensive training sessions, the result would more than likely allow for generating even more realistic remote sensing images from existing limited data sets.

The issue of class imbalance is relevant when training not only the GAN for data augmentations, but also the solar panel semantic segmentation model, and is mentioned in several works. The pixel accuracy metric may be unreliable when the class imbalance issue is present, due to the dominant background pixels being correctly evaluated. Therefore, the IoU metric should be examined more closely in contrast to pixel accuracy. This is relevant not only when segmenting the solar panels themselves, but also when performing other critical analysis such as fault detections [30]. Also, due to the nature of resampling data sets to a common spatial resolution and resizing the images to a target image resolution (either by cropping or padding), some remote sensing images may appear annotated in a simplified manner, e.g. a zoomed-in array of several solar panel installations may appear annotated as a single large segmentation object. In that case, when training the model, it would be recommended to pay more attention to the validation loss metric in addition to the validation IoU metric. The reason for that is that the latter may be misleading when evaluating the model on the testing dataset due to correct segmentation falsely being marked as poor segmentation. This may cause problems during model training, validation, and testing, when the model segments the objects more accurately than they are labeled in the original data, resulting in a falsely lower IoU metric. To combat this issue, higher-quality data sets are required, along with more careful labeling.

## 5. Conclusions

To sum up, the usage of the generative adversarial network is beneficial for training dataset augmentation with brand new remote sensing images from already existing data. Expanding a limited

data set with newly generated images proved to be beneficial for the training of solar panel installation semantic segmentation model in terms of accuracy and segmentation quality. Conducted experiments and tests display the benefits of using artificial data over performing basic data augmentations and the potential of improving the model for more accurate solar panel detection. With even more improvements to not only the segmentation model hyperparameters but also the generative adversarial network parameters, the fine-tuned semantic segmentation model may be used for even more precise solar panel detection in RS images and potentially the development of a comprehensive solar panel map, allowing for easier solar panel market and usage analysis.

**Author Contributions:** Conceptualization, J.L. and V.G.; methodology, J.L. and V.G.; software, J.L.; validation, J.L.; formal analysis, J.L.; investigation, J.L.; resources, J.L. and V.G.; data curation, J.L.; writing—original draft preparation, J.L.; writing—review and editing, J.L. and V.G.; visualization, J.L.; supervision, V.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

**Conflicts of Interest:** The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
FCN	Fully Convolutional Network
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
GSD	Ground Sampling Distance
IoU	Intersection Over Union
PV	Photovoltaic
RS	Remote Sensing
SRCNN	Super-Resolution Convolutional Neural Network
SRGAN	Super-Resolution Generative Adversarial Network
ViT	Vision Transformer

References

- Guangul, F.M.; Chala, G.T. Solar Energy as Renewable Energy Source: SWOT Analysis. In Proceedings of the 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC); January 2019; pp. 1–5.
- Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation 2015.
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation 2015.
- Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation 2016.
- Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation 2017.
- Li, L.; Lau, E. RU-Net: Solar Panel Detection From Remote Sensing Image. In Proceedings of the 2022 IEEE Green Energy and Smart System Systems (IGESSC); IEEE: Long Beach, CA, USA, November 7 2022; pp. 1–6.
- Ge, F.; Wang, G.; He, G.; Zhou, D.; Yin, R.; Tong, L. A Hierarchical Information Extraction Method for Large-Scale Centralized Photovoltaic Power Plants Based on Multi-Source Remote Sensing Images. Remote Sensing 2022, 14, 4211, doi:10.3390/rs14174211.
- Sait, M.; Erguzen, A.; Erdal, E. Using Mask R-CNN to Isolate PV Panels from Background Object in Images.

9. Gonçalves, M.; Martins, B.; Estima, J. A Detailed Analysis on the Use of General-Purpose Vision Transformers for Remote Sensing Image Segmentation. In Proceedings of the Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery; ACM: Hamburg Germany, November 13 2023; pp. 20–29.
10. Angelis, G.-E.; Domi, A.; Zamichos, A.; Tsourma, M.; Drosou, A.; Tzovaras, D. On The Exploration of Vision Transformers in Remote Sensing Building Extraction. In Proceedings of the 2022 IEEE International Symposium on Multimedia (ISM); IEEE: Italy, December 2022; pp. 208–215.
11. Sahoo, P.; Saha, S.; Mondal, S.; Sharma, N. COVID-19 Detection from Lung Ultrasound Images Using a Fuzzy Ensemble-Based Transfer Learning Technique. In Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR); IEEE: Montreal, QC, Canada, August 21 2022; pp. 5170–5176.
12. Patel, A.; Degadwala, S.; Vyas, D. Lung Respiratory Audio Prediction Using Transfer Learning Models. In Proceedings of the 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC); November 2022; pp. 1107–1114.
13. Zhang, H.; Jiang, Z.; Zheng, G.; Yao, X. Semantic Segmentation of High-Resolution Remote Sensing Images with Improved U-Net Based on Transfer Learning. *Int J Comput Intell Syst* 2023, 16, 181, doi:10.1007/s44196-023-00364-w.
14. Cui, B.; Chen, X.; Lu, Y. Semantic Segmentation of Remote Sensing Images Using Transfer Learning and Deep Convolutional Neural Network With Dense Connection. *IEEE Access* 2020, 8, 116744–116755, doi:10.1109/ACCESS.2020.3003914.
15. Sun, X.; Wang, B.; Wang, Z.; Li, H.; Li, H.; Fu, K. Research Progress on Few-Shot Learning for Remote Sensing Image Interpretation. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 2021, 14, 2387–2402, doi:10.1109/JSTARS.2021.3052869.
16. Liu, J.; Liu, G.; Zhao, Y. Improve Semantic Segmentation of High-Resolution Remote Sensing Imagery with RS-TTA. In Proceedings of the 2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI); IEEE: Chengdu, China, August 19 2022; pp. 817–825.
17. Kumar, A.; Thakur, A.K.; Kumar, P.; Mandal, M.; Kundu, S. A CNN Based Efficient Brain Tumor Detection Using MRI. *JOURNAL OF CRITICAL REVIEWS* 2020, 7.
18. Alomar, K.; Aysel, H.I.; Cai, X. Data Augmentation in Classification and Segmentation: A Survey and New Strategies. *Journal of Imaging* 2023, 9, 46, doi:10.3390/jimaging9020046.
19. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks 2014.
20. Wang, J.; Gao, K.; Zhang, Z.; Ni, C.; Hu, Z.; Chen, D.; Wu, Q. Multisensor Remote Sensing Imagery Super-Resolution with Conditional GAN. *J Remote Sens* 2021, 2021, 2021/9829706, doi:10.34133/2021/9829706.
21. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks 2018.
22. Henry, J.; Natalie, T.; Madsen, D. Pix2Pix GAN for Image-to-Image Translation; 2021;
23. Yu, Q.; Malaeb, J.; Ma, W. Architectural Facade Recognition and Generation through Generative Adversarial Networks. In Proceedings of the 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE); IEEE: Bangkok, Thailand, October 2020; pp. 310–316.
24. Jiang, H.; Yao, L.; Lu, N.; Qin, J.; Liu, T.; Liu, Y.; Zhou, C. Multi-Resolution Dataset for Photovoltaic Panel Segmentation from Satellite and Aerial Imagery. *Earth System Science Data* 2021, 13, 5389–5401, doi:10.5194/essd-13-5389-2021.
25. Kasmi, G.; Saint-Drenan, Y.-M.; Trebosc, D.; Jolivet, R.; Leloux, J.; Sarr, B.; Dubus, L. A Crowdsourced Dataset of Aerial Images with Annotated Solar Photovoltaic Arrays and Installation Metadata. *Sci Data* 2023, 10, 59, doi:10.1038/s41597-023-01951-4.
26. Pierdicca, R.; Paolanti, M.; Felicetti, A.; Piccinini, F.; Zingaretti, P. Automatic Faults Detection of Photovoltaic Farms: solAIr, a Deep Learning-Based System for Thermal Images. *Energies* 2020, 13, 6496, doi:10.3390/en13246496.
27. Ravishankar, R.; AlMahmoud, E.; Habib, A.; de Weck, O.L. Capacity Estimation of Solar Farms Using Deep Learning on High-Resolution Satellite Imagery. *Remote Sensing* 2023, 15, 210, doi:10.3390/rs15010210.

28. Wang, J.; Chen, X.; Shi, W.; Jiang, W.; Zhang, X.; Hua, L.; Liu, J.; Sui, H. Rooftop PV Segmenter: A Size-Aware Network for Segmenting Rooftop Photovoltaic Systems from High-Resolution Imagery. *Remote Sensing* 2023, 15, 5232, doi:10.3390/rs15215232.
29. Costa, M.V.C.V. da; Carvalho, O.L.F. de; Orlandi, A.G.; Hirata, I.; Albuquerque, A.O. de; Silva, F.V. e; Guimarães, R.F.; Gomes, R.A.T.; Júnior, O.A. de C. Remote Sensing for Monitoring Photovoltaic Solar Plants in Brazil Using Deep Semantic Segmentation. *Energies* 2021, 14, 2960, doi:10.3390/en14102960.
30. Jumaboev, S.; Jurakuziev, D.; Lee, M. Photovoltaics Plant Fault Detection Using Deep Learning Techniques. *Remote Sensing* 2022, 14, 3728, doi:10.3390/rs14153728.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.