Article

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

Justinas Lekavičius *[ID] and Valentas Gružauskas [ID]

Institute of Computer Science, Vilnius University, 08303 Vilnius, Lithuania; valentas.gruzauskas@mif.vu.lt
* Correspondence: me@justinaslekavicius.com; Tel.: +370-600-72586

Abstract: With the popularity of solar energy in the electricity market, demand rises for data such as precise locations of solar panels for efficient energy planning and management. However, these data are not easily accessible; information such as precise locations sometimes does not exist. Furthermore, existing datasets for training semantic segmentation models of photovoltaic (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 used, enriching the original resampled training data of varying ground sampling distances (GSDs) without compromising their integrity. Experiments with the DeepLabV3 model, ResNet-50 backbone, and pix2pix GAN architecture were conducted to discover the advantage of using GAN-based data augmentations for a more accurate RS imagery segmentation model. The result is a fine-tuned solar panel semantic segmentation model, trained using transfer learning and an optimal amount—60% of GAN-generated RS imagery for additional training data. The findings demonstrate the benefits of using GAN-generated images as additional training data, addressing the issue of limited datasets, and increasing IoU and F1 metrics by 2% and 1.46%, respectively, compared with classic augmentations.

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 methods of converting solar energy into electricity, which homeowners and electricity providers use. They are crucial to producing clean, renewable energy and are environment friendly and cheaper than before [1]. With the increasing popularity of solar power in the electricity market, demand rises for various data. Such data can include the locations of solar panels, their types, quantities, specifications, and power capacities. These data, especially the locations of solar panels, can be used for efficient policy-making, energy planning and distribution, and grid management. In this scenario, remote sensing data and machine learning are valuable tools. The remote sensing procedure, in particular by satellites, allows for measuring the characteristics of an observed area and collecting images. The collected images of the planet's surface can then be used to observe and detect patterns and objects or, in this case, solar panels. Convolutional neural networks (CNNs) are used, and semantic segmentation is performed to analyze visual data. For example, segmentation models like FCN [2] and U-Net [3] utilize end-to-end training and specialized architectures for feature combinations. Others, such as SegNet and DeepLabV3, prioritize performance [4] or multiscale object segmentation [5].

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 [6] has been utilized for the efficient solar panel detection performance and identification of rooftop solar panel locations, distributions, and surface areas in

images of 0.3 m/pixel spatial resolution. A hierarchical information extraction method for solar panels using multi-source satellite remote sensing images was also proposed [7] and tested in three selected provinces in China to locate as many solar panels as possible, reducing the number of false positives. The mask R-CNN [8] 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. Although 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. General-purpose vision transformer models such as SegFormer and Lawin Transformer have also been used for remote sensing image segmentation, demonstrating results comparable to other state-of-the-art models such as FCN and RemoteNet [9]. For building segmentation in remote sensing images, various models such as TransUNet, MiTNet, UNetFormer, and Segformer have been compared [10], presenting varying metrics such as F1 score, overall accuracy, and mean intersection over union when tested on the Global Cities WHU Satellite dataset. Semantic segmentation models and vision transformers also benefit from transferring the learning knowledge from one task to another, i.e., transfer learning. Such models as ResNet, FCN, and DeepLab were pre-trained on the ImageNet dataset or used weights from other models pre-trained on ImageNet [11,12]. Furthermore, transfer learning has also been used specifically for training remote sensing image semantic segmentation models. An improved U-Net model based on transfer learning was proposed for high-resolution remote sensing images, which exhibited better results for vehicle semantic segmentation than the regular U-Net model [13]. For the multiobject segmentation of remote sensing images when the issue of insufficient labeled data and imbalanced data classes is present, transfer learning has also improved the semantic segmentation model performance to some degree [14].

Large amounts of annotated data are needed to effectively train these semantic segmentation models, with even more benefits from more extensive and diverse datasets. Complications due to a lack of raw data, difficulty in dataset annotation, and limitations of sensor characteristics make this more difficult, as a considerable variation and amount of data are needed to create an effective model [15]. Unfortunately, regional data are sometimes not easily accessible due to privacy concerns or the unwillingness of solar panel installers to share them. Occasionally, these data do not include precise locations of solar panel installations and only include statistics such as power capacities, complicating the information-gathering process. Furthermore, solar panels can have different characteristics; for example, monocrystalline, polycrystalline, and thin film panels may have different colors, patterns, and sizes. Additionally, some solar panels may have distinctive grid lines, and some may even be incorporated into the building architecture for aesthetic reasons, making them more difficult to distinguish. The lack of such data hinders the efforts for accurate solar panel segmentation model training and regional solar panel map development. Data augmentations help expand the dataset for the machine learning model, especially when training a semantic segmentation model for object detection. Basic augmentations include, although 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 datasets 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 similarly. 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 lighting changes, while cropping and zooming of images manage scaling and background issues [18]. Augmentations are needed for extending deep learning model training datasets, and this methodology is

especially beneficial when the dataset is small. Sometimes, it reduces the chances of the model overfitting, i.e., performing too well on training data and poorly on new, unseen data. However, basic data augmentations may produce un-natural 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, generative adversarial networks, which are capable of producing new realistic samples from existing data [19], are beneficial in satisfying the demand for more diverse datasets.

When enhancing the diversity of limited datasets, generative adversarial networks have been successfully used for computed tomography imagery generation and classification improvements [20]. Furthermore, the advantage of GAN-generated imagery over classic augmentation is noted. Similarly, GAN-based data augmentations were used for defective photovoltaic module cell training sample creation [21]. Generative adversarial networks have also been used for remote sensing image reconstruction. The conditional discriminator PatchGAN was used for remote sensing imagery super-resolution, i.e., to create images that have not only high fidelity but also high perceptual quality [22]. The upscaled data are more realistic and feature higher fidelity, and when compared with 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.

Considering the novelty of the model and the potential of generating new material from an already-existing limited dataset, the accuracy of the trained semantic segmentation model can be considerably improved with realistic additional remote sensing images when compared with simply utilizing basic data augmentations. Therefore, we propose using generative adversarial networks (GANs) for data augmentation in solar panel segmentation. GANs can generate realistic images from limited data, thereby enhancing the diversity and quantity of training samples without manual annotation. Specifically, we utilize the pix2pix GAN for this purpose, as it is effective in not only experiments such as generating maps from aerial images [23] and architectural facades from labels [24], but also synthesizing photos from label maps [25].

The main objective of this work is to improve the semantic segmentation of solar panel installations by using pix2pix GAN-generated data for training data augmentation. The approach includes conducting experiments with datasets of various spatial resolutions and applying transfer learning and fine-tuning techniques. Furthermore, sensitivity analysis is performed to identify the optimal quantity of synthetic data to use for semantic segmentation model training to achieve better accuracy and reduced overfitting. Both basic data augmentations and GAN-generated data are used to train the semantic segmentation model, and their effectiveness is compared. The experiments reveal the impact of data augmentations on segmentation performance and provide insights into the most effective augmentation strategies. The resulting solar panel segmentation model is fine-tuned for remote sensing images, achieving improved detection accuracy.

This paper is structured as follows: Section 2 describes the materials and methods used in our experiments. Section 3 presents the results, including the training of the pix2pix generative adversarial network, sensitivity analysis, and solar panel segmentation model training outcomes. Section 4 discusses our research findings, compared with other authors' works. Finally, Section 5 concludes this paper and suggests directions for future work.

2. Materials and Methods

The data used for DeepLabV3 semantic segmentation and pix2pix generative adversarial network training are a collection of five solar panel aerial image datasets. The Provincial Geomatics Center of Jiangsu provides three datasets [26]. Two are sourced from Google Earth and the French National Institute of Geographical and Forestry Information (IGN) [27]. The datasets are presented and compared in Table 1. The main differences are the image formats, the ground sampling distance (or GSD, the distance between the centers of two neighboring pixels measured on the ground), and the image resolutions. For DeepLabV3 and pix2pix generative adversarial network training, a subset of 640 image-

mask pairs of each ground sampling distance is used, totaling 2560 images and 2560 solar panel binary semantic segmentation masks. Because there are two datasets of 0.1 m/pixel ground sampling distance, 320 image-mask pairs are used from each. Out of 640 image-mask pairs of each GSD, 80% are used for training the DeepLabV3 model and pix2pix GAN, 10% are used for validation, and 10% are used for testing the DeepLabV3 semantic segmentation model.

Table 1. A comparison of used solar panel semantic segmentation datasets.

Data Source	Image Format	Ground Sampling Distance	Image Resolution	Number of Used Images
Provincial Geomatics Center of Jiangsu	BMP	0.8 m/pixel	1024 × 1024	640 images/640 masks
Provincial Geomatics Center of Jiangsu	BMP	0.3 m/pixel	1024 × 1024	640 images/640 masks
French National Institute of Geographical and Forestry Information	PNG	0.2 m/pixel	400 × 400	640 images/640 masks
Provincial Geomatics Center of Jiangsu	BMP	0.1 m/pixel	256 × 256	320 images/320 masks
Google Earth	PNG	0.1 m/pixel	400 × 400	320 images/320 masks

Before semantic segmentation model training, width and height resampling is applied for all images and their binary semantic segmentation masks. This is needed to solve the issue of ground sampling distance and image resolution differences when several different datasets are used. In the experiments, a target GSD of 0.1 m/pixel and a target image resolution of 512 × 512 are used for resampling. This is done to keep as much detail in the resampled images as possible, while retaining computational efficiency and data uniformity to represent the scale of solar panels more accurately. Bringing all samples to the same "centimeters per pixel" ratio ensures a more accurate representation of the scale of solar panels in remote sensing images. Furthermore, this helps ensure better generalization of the solar panel segmentation model to remote sensing images where the solar panel installations are of different scales. The process for resampling the width and height of the remote sensing images to the target ground sampling distance is presented in Equations (1) and (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 original 0.1 m/pixel GSD images are not resampled to a new image resolution, as they are already of the target GSD, and the image-mask pairs of 0.8 m, 0.3 m, and 0.2 m 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 resampled at a target image resolution of 512 × 512. This is done by either cropping or padding the image-mask pairs. The process for padding image width and height is presented in Equations (3) and (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 of the same target GSD but with lower image resolution. It is done to bring the images and masks to the target image resolution without compromising the performed GSD resampling or image quality.

$$padding\_width = \lfloor \frac{target\_width - resampled\_width}{2} \rfloor \tag{3}$$

$$padding\_height = \lfloor \frac{target\_height - resampled\_height}{2} \rfloor \tag{4}$$

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 for 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 always include a solar panel. Additionally, this is done not just because the segmentation mask may locate multiple objects but because they may be of different sizes. First, the number of solar panel objects in a semantic segmentation mask is calculated. Then, the bounding box slices are located for each object, and the slice with the most significant 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. In addition, 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. The resampling process is illustrated in Figure 1.
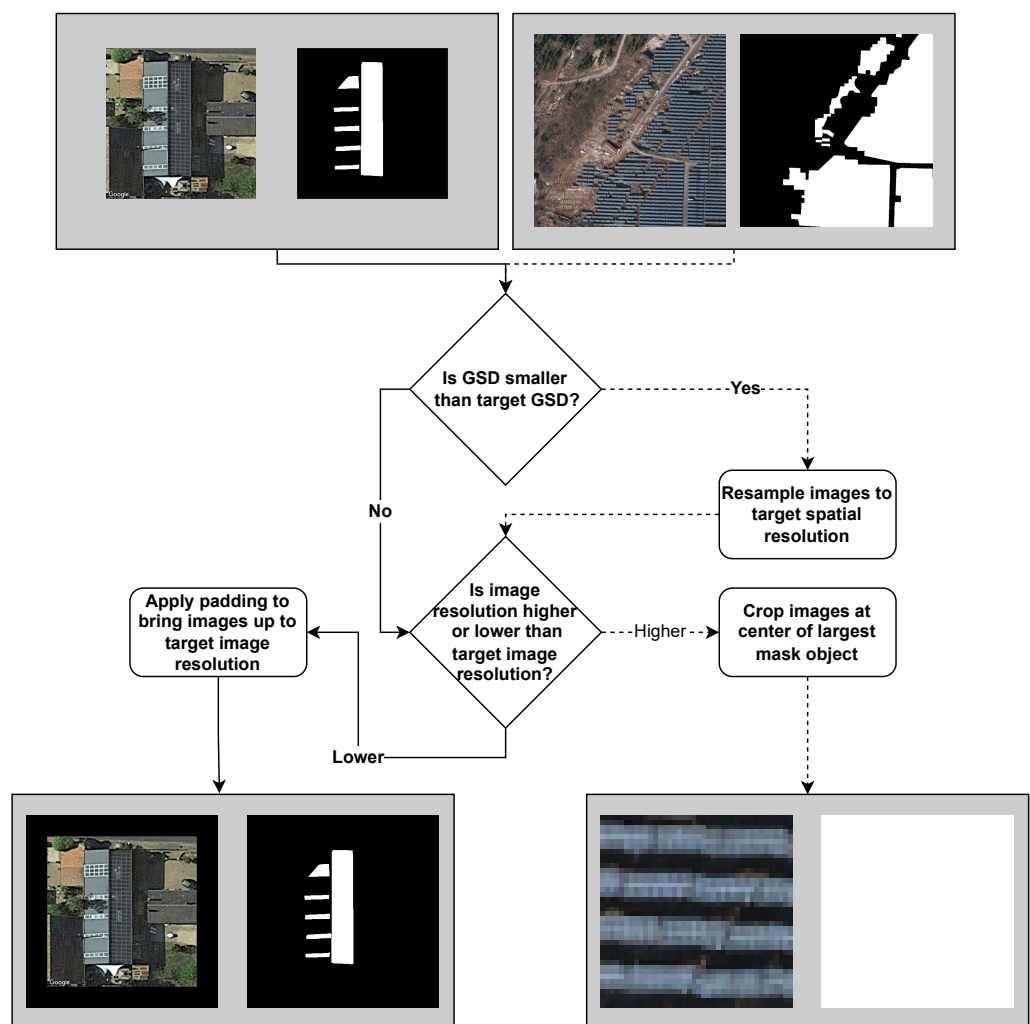
Figure 1. The process of image-mask pair resampling to target spatial and image resolution.

The DeepLabV3 architecture with the ResNet-50 backbone was used to train the solar panel semantic segmentation model. 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 and capabilities are suitable for the task of PV installation segmentation in remote sensing images of various GSDs. Due to the desired accuracy and computational efficiency balance, ResNet-50 was selected as the backbone over other options, such as MobileNet and ResNet-101. 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 compared with 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 10 consecutive 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). The average accuracy, precision, recall, F1 score, and intersect over union are calculated to evaluate the trained model. In this case, accuracy refers to pixel accuracy, i.e., the correctly classified pixels when comparing the ground truth mask and the predicted mask. This includes correctly predicting the pixels where the solar panel is segmented (white pixels) and 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.

Table 3. Metrics used to evaluate 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	$2 \times (\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})$
Intersection over Union	$\sum_{i=1}^{N} I(y_{\text{true},i} \cap y_{\text{pred},i}) / \sum_{i=1}^{N} I(y_{\text{true},i} \cup y_{\text{pred},i})$

The F1 score is a valuable metric for class imbalance (foreground and background), as it is the harmonic mean of precision and recall. Because of class imbalance (smaller percentage of white pixels, i.e., the PV objects, when compared with 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 have an IoU metric higher than or equal to 0.5, while poorly segmented images have an IoU metric lower than 0.5 but not equal to 0. Unsegmented images have an IoU metric of 0.

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. The parameters are displayed in Table 4.

Table 4. Hyperparameters used for pix2pix generative adversarial network training.

Epochs	Architecture		Lambda_L1	Image Channels		Filters	
	Generator	Discriminator		Input	Output	NGF	NDF
600	U-Net256	PatchGAN	75	1	3	128	64

The U-Net256 generator architecture was used; therefore, the images and masks were resized to 256×256 resolution. For the discriminator, the PatchGAN architecture with 3 convolutional layers was used. The lambda_L1 parameter was set to 75 instead of the default value of 100. This reduces the importance of L1 loss, encouraging the generator to produce images closer to the original input data, as the lambda_L1 value is used in the training objective. The reduction in lambda_L1 value in this case encourages the generation of 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 decreasing the learning rate to closer to zero. The 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 generated images). To maintain the balance of power between the generator and the discriminator, and so that the former does not outpower 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) was set at 128 (default value is 64). This allows the generator to capture more details and generate more convincing images. A batch size of 1 was maintained for higher stability and context preservation when only one image-mask pair was used for mapping learning at a time.

To compare the benefits of using classic data augmentations for the training dataset versus 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 a random horizontal flip with a 50% chance of it being applied, a random rotation of 5 degrees, a random perspective change with a 0.05 distortion scale and a 50% chance of it being applied, and a random application of Gaussian blur (5×5 kernel size and standard deviation of 0.1 min and 2.0 max) with a 50% change of it being used. The Gaussian blur application is intended to simulate the fogging of a satellite lens.

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

Notably, for the third and fourth scenarios, 25% of the remote sensing imagery generated using the generative adversarial network is used to prove that the trained semantic segmentation model benefits from additional synthetic remote sensing image-mask pairs. In contrast, the fifth and sixth experiment variants display the 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:

1. Training the model from scratch;
2. Training the model using transfer learning;
3. Fine-tuning the last layer of the previous model;
4. Fine-tuning the remaining layers of the previous model.

In the first scenario, the model is trained without pre-trained weights. This serves as a baseline for a comparison with other scenarios and shows how a model trained from the ground up for the specific dataset performs compared with fine-tuned models and those 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) [28] subset using 20 categories in the Pascal VOC [29] dataset, such as bicycle, aeroplane, bottle, and dining table. While solar panels are not one of these 20 categories, the learned general features like shapes and edges are still beneficial, providing a 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 them for 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 builds upon the previous scenario, allowing the whole 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 solar panel semantic segmentation task and make it even more adapted. The result is a total of 24 experiments. The extra samples are only used for the training dataset when experiments use the additional remote sensing images generated utilizing the pix2pix generative adversarial network. At the same time, the validation and testing datasets 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 consist of high-quality, real-life data, and the synthetic remote sensing images, although realistic, 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 dataset for each GSD (total of 2560 pairs), and an 80/10/10 split was used for training, validation,
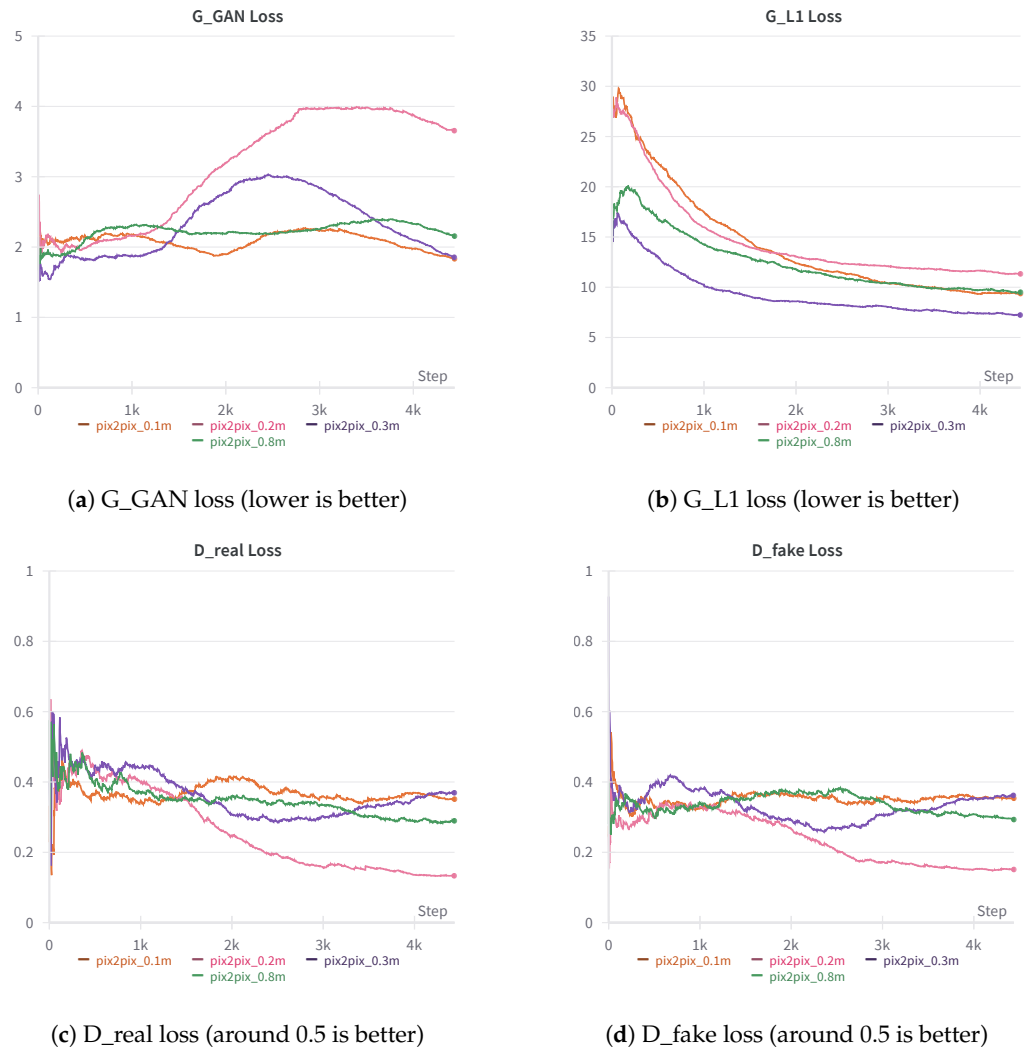
and testing data. The result was 512 image-mask pairs of each GSD used for training, 64 image-mask pairs for model validation, and 64 image-mask pairs for model testing (2048, 256, and 256 pairs, respectively). Because there were two datasets of 0.1 m/pixel GSD, 256 image-mask pairs made up 512 image-mask samples for the 0.1 m GSD training dataset, 32 pairs for validation, and 32 pairs for testing. To ensure reproducibility, consistent sampling across runs, and consistent shuffling and randomizing when running the code multiple times, the random seed was set to 35 and applied to PyTorch, random, and NumPy modules. Where needed, GAN-generated image-mask pairs were appended to the training dataset. The batch size was set to 48, and 12 workers were used for the data loaders. During data loader creation, the image-mask pairs were resampled to a target GSD and target image resolution of 0.1 m/pixel and 512 × 512, 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 generative adversarial network 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) and the number of epochs for training. For training, 512 image-mask pairs of each ground sampling distance were used, and four separate pix2pix GAN models were trained (one for each GSD) to generate remote sensing images of different resolutions and different solar panel scales. Most importantly, the image-mask pairs were the same ones used in the original dataset for DeepLabV3 semantic segmentation model training, so new data were generated from existing image-mask pairs. 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 the G_L1 loss to be as low as possible, indicating the generated image's closer resemblance to the original data, and for the 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 0.8 m/pixel and 0.3 m/pixel GSDs displays lower G_L1 loss, signaling the generated image's close resemblance to the source material. 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 0.8 m/pixel and 0.3 m/pixel GSDs. These images also originally contain fewer details when compared with images of higher GSDs, which may explain 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 0.2 m/pixel and 0.1 m/pixel GSD shows that, in the case of training with 0.2 m 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., having more fine details that the generator had difficulty recreating, consequently not fooling the discriminator. There are fluctuations in discriminator losses, particularly in the 0.2 m/pixel GSD dataset, which lowers to values of 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, and vehicles, which 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. However, compared with images generated from 0.8 m and 0.3 m GSD segmentation mask data, they can be seen as lower quality, in some cases featuring unrealistic road formations or building shapes, although retaining mostly

correct solar panel installation generation. In worst cases, the solar panel installations are also not generated convincingly, resulting in inferior quality samples. Examples are noisy formations, distorted shapes, and inconsistent colors.

(**a**) G_GAN loss (lower is better)

(**b**) G_L1 loss (lower is better)

(**c**) D_real loss (around 0.5 is better)

(**d**) D_fake loss (around 0.5 is better)

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

After the pix2pix GAN training is complete, the semantic segmentation masks used for GAN training are used for model testing, i.e., generating new image-mask pairs. A total of 512 binary semantic segmentation masks are used for each GSD 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 but with subtleties that can differentiate them. Figure 3 illustrates that the generated images feature realistic solar panel installations and sufficiently believable environments. Upon closer inspection, details such as roads leading nowhere or inconsistent building layouts can be observed; however, the most crucial aspect, i.e., the solar panels, is generated with satisfactory quality. Some samples, however, are of worse quality, mainly 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 with satisfactory-quality

samples is insignificant. In total, 2048 new remote sensing images were generated, i.e., 512 images for each GSD.
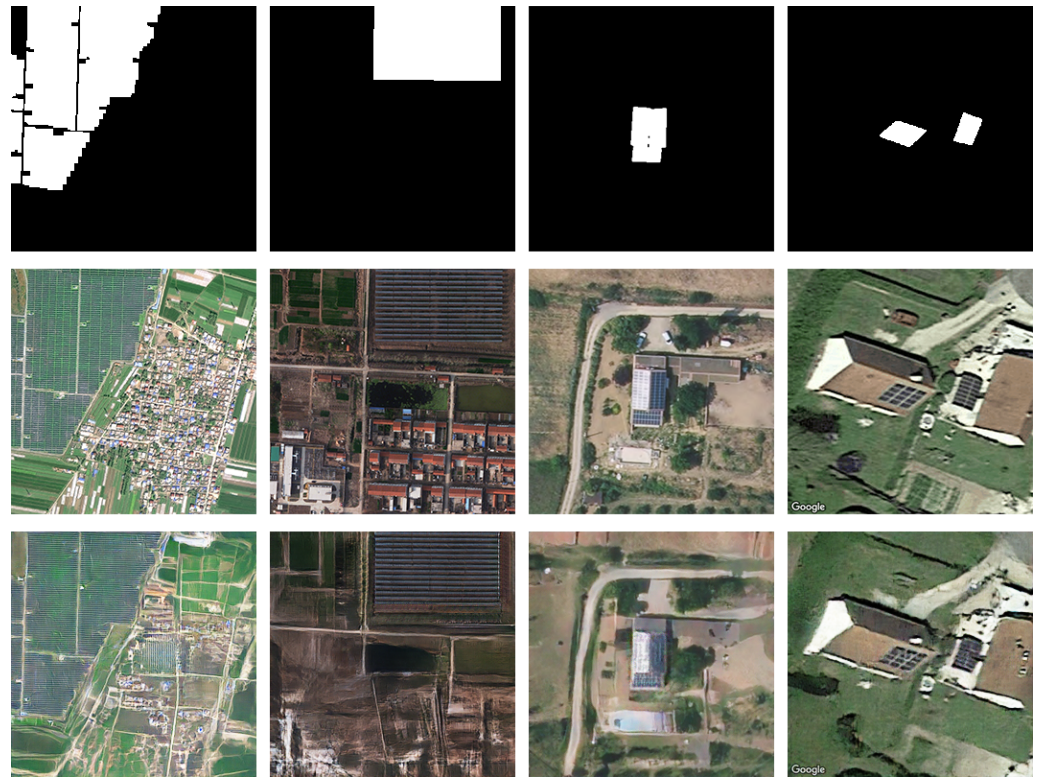

Figure 3. An example of binary semantic segmentation mask (**top row**), original remote sensing image (**middle row**), and GAN-generated image (**bottom row**) using pix2pix trained model for each GSD, respectively, from left to right: 0.8 m, 0.3 m, 0.2 m, 0.1 m.

3.2. Sensitivity Analysis

A sensitivity analysis was performed to find the optimal amount of GAN-generated data for use as additional training data. This was done to examine how much additional data are needed to obtain 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 ten times using transfer learning and with an additional 10% of generated remote sensing images incrementally added to the training dataset (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 original data without basic augmentations). For sensitivity analysis, validation and testing subset average IoU and average loss metrics were compared, and the changes with additional GAN data usage are detailed in Figure 4.

Observing the data in the figure reveals that the best average validation and testing IoU values are when 60% and 90% additional GAN data are used, respectively. Furthermore, the lowest average 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 the average validation IoU is at 60%, while the peak of the average 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 begins losing the ability to generalize to new data.
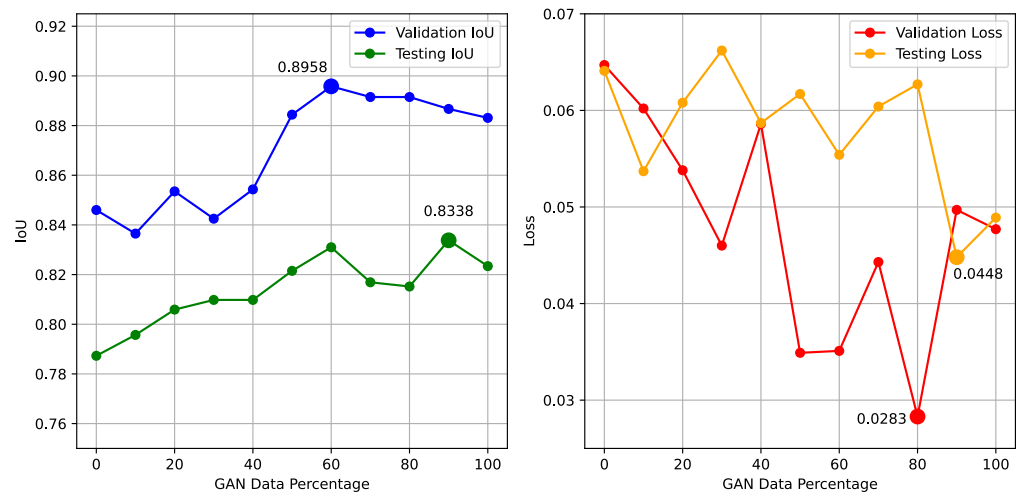
Figure 4. Sensitivity analysis of DeepLabV3 model training using transfer learning and various percentages of additional GAN-generated training data, comparing average validation IoU, average testing IoU (**left**), and average loss (**right**) metrics.

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

Therefore, the total training subset size becomes 3276 (2048 original image-mask pairs plus 1228 additional generated pairs), while the overall dataset increases to 3788 samples (3276 training, 256 validation, and 256 testing 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) is evaluated, as they are arguably the most optimized for the solar panel installation semantic segmentation task. The training and testing results of the 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 dataset (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 5, the gan60 scenario (training the model with 60% additional GAN remote sensing image data) features the best metrics. When comparing the results of training without any augmentations, the benefits of using the generative adversarial network for additional training data synthesis are evident, especially when comparing the improvements of all metrics. Comparing the results of the experiment gan60 with those of the baseline experiment no_aug, average accuracy increased by 0.78%, average precision by 3.41%, average recall by 2.49%, average F1 score by 2.71%, and average IoU by 3.19%, while average loss decreased by 0.0282. Furthermore, more images that feature solar panels are successfully segmented. The number of images that do not have segmented solar panels are the lowest (only comparable to the scenario basic_aug), and the sum of images with correctly and poorly segmented photovoltaic panels is higher than in other scenarios. Although the scenarios gan25_aug and gan60_aug have a higher count of correctly segmented solar panels, the amounts of unsegmented images are higher, and the overall sums of correctly and

poorly segmented panels are lower. Although the sum of correctly and poorly segmented images is equal for the basic_aug and gan60 scenarios, the former has a lower count of correctly segmented panels when compared with the latter. Compared with the scenario basic_aug, which features only the usage of basic data augmentations, the improvements and benefits are still visible. When using GAN-based data augmentations compared with basic augmentations, the average pixel accuracy increased by 0.79%, average precision by 0.5%, average recall by 1.6%, average F1 score by 1.46%, and average IoU by 2%, while average loss decreased by 0.0179.

Table 5. Comparison of testing results for all six scenarios, with transfer learning and fine-tuning applied. The best values are written in bold.

Experiment	Avg Acc (%)	Avg Prec (%)	Avg Rec (%)	Avg F1 (%)	Avg IoU (%)	Avg Loss	Solar Panel Segmentation (IoU)		
							Correct $\geq$ 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	**98.67**	**90.13**	**88.11**	**87.96**	**83.32**	**0.0368**	237	11	8
gan60_aug	98.04	89.82	87.69	87.77	82.90	0.0611	238	9	9

Based on the model testing results, it was observed that the best outcome is achieved by training the semantic segmentation model while increasing the original training dataset by 60% using the synthetic remote sensing images generated by the generative adversarial network pix2pix. However, applying additional basic image augmentations did not yield significant benefits based on the testing results and displayed slightly worse outcomes. This lack of improvement can be attributed to the challenging nature of the generated RS images. Despite their realistic appearance, these images still contain some noise and artifacts in certain samples, which may have contributed to the limited effectiveness of the basic image augmentations. The generated RS images, while visually convincing, present complexities that impact their suitability for semantic segmentation model training when basic augmentations are additionally applied. Therefore, it is important to consider the possible negative outcome of applying additional classic augmentations to synthetic data.

The final trained semantic segmentation model can segment solar panel installations at different scales, shapes, and shades. The model was tested with images taken from Google Maps. The photos consist of random locations throughout Lithuania, containing small solar panel installations and solar power stations. The original images and their segmentation results are displayed in Figure 5. Notably, using the semantic segmentation model trained using 60% of additional GAN-generated samples (as seen in row 6 of the figure) yields some of the best results. Solar panels are accurately segmented, and various solar panel installations are generally well detected, as they are also distinct at different scales due to their grid lines and rectangular shape. The model also performs well with larger solar panel installations, especially compared with the model trained with the original data, without augmentations (row 2), where the solar panel farms in the first and second columns are not segmented correctly. When tested with images of solar power stations, either the model predicted the array of solar panels as a single object or the entire semantic segmentation mask was white, depending on the scale.
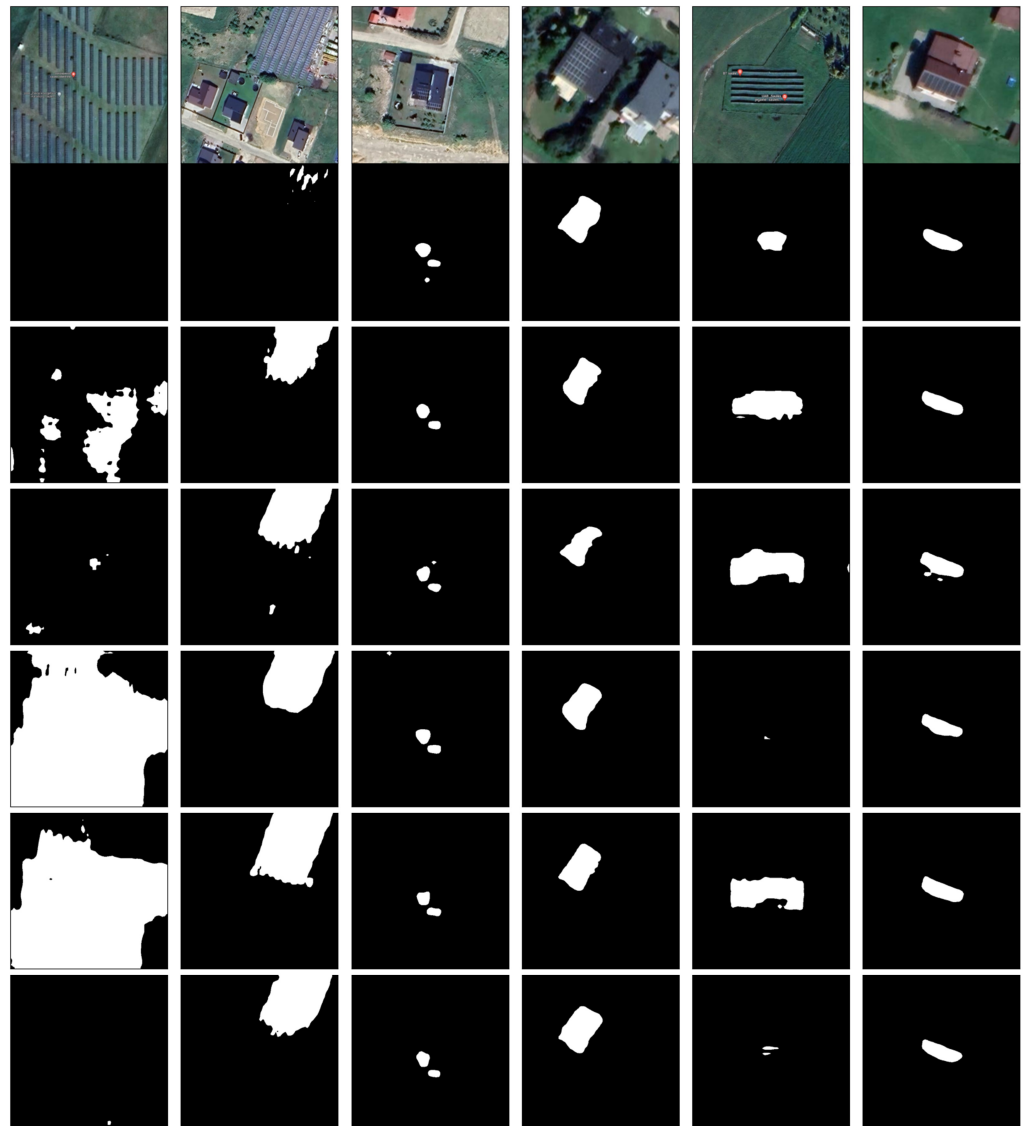
Figure 5. An example of several RS images of locations throughout Lithuania (**row 1**) and segmentation results using a model trained without data augmentations (**row 2**), with basic data augmentations (**row 3**), using 25% of additional GAN training data (**row 4**), using 25% GAN data + basic augmentations (**row 5**), using 60% of additional GAN training data (**row 6**), and using 60% of additional GAN training data + basic augmentations (**row 7**).

4. Discussion

Using the pix2pix generative adversarial network for data augmentation, the semantic segmentation model's accuracy is improved, and the issue of the need for manual 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 are needed. For instance, datasets such as thermal images and their respective segmentation masks could be expanded with new synthetic data, especially when expertise in photovoltaic farm fault detection is needed [30]. Pamungkas et al. explored the use of generative adversarial networks for a more efficient solar panel fault classification, and the advantages of GAN augmentations were compared with the utilization of geometric augmentations [31]. The authors note that combining the classic augmentations with GAN augmentations resulted in varying effectiveness, possibly due to false positives or negatives caused by GAN augmentations. Using GAN augmentation can also be applied to improving solar farm capacity estimation, as either an alternative or an additional solution

to exploring other data sources [32]. 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 this work focuses on improving solar panel segmentation from RS imagery using GAN-based data augmentations instead of segmentation model architecture optimizations and improvements, this method may be 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 [33], 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 [34]. 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 dataset, the study would likely benefit from an even more diverse training dataset featuring newly generated images using the generative adversarial network. This would, however, depend on the nature of the original dataset and potential points of caution, such as data quality and class imbalances.

The solar panel semantic segmentation model's performance on the testing dataset was compared with the works of other authors, which use either the same datasets or a similar combination of the model architecture and backbone. The solar panel segmentation model "gan60" was used for comparisons, which was trained with 3788 dataset samples (3276 training, 256 validation, and 256 testing image-mask pairs). Because our work utilizes several types of data from different spatial resolutions and sources and the data were resampled to 0.1 m GSD, objectively comparing our results with others was challenging due to dataset combination differences, the contrasts in the used number of samples for training and testing the models, and the overall nature of this work, i.e., the improvement of the model with newly generated data instead of parameters or architecture optimizations. However, even though the comparability of the other works is limited due to differences in architectures and datasets, the comparison with the state of the art is important in demonstrating the performance of the semantic segmentation model when GAN-augmented data are used for training. An overview of similar datasets and architecture solutions was made, and the comparison results are presented in Table 6.

Table 6. Model testing results compared with other authors' works with similar models and datasets.

Author	Avg Acc (%)	Avg Prec (%)	Avg Rec (%)	Avg F1 (%)	Avg IoU (%)	Image GSD	Image Size	Dataset Samples
Guo et al. [35]	84.06	86.78	84.06	85.40	74.52	0.2 m	400 × 400	7686
Mayer et al. [36]	-	87.30	87.50	87.40	74.10	0.1 m	320 × 320	4028
Yang et al. [37]	99.30	91.40	92.60	85.30	84.10	0.1 m	256 × 256	14,349
Zhu et al. [38]	94.77	89.51	94.77	92.06	85.30	0.1 m	512 × 512	4012
Jiang et al. [39]	91.80	90.20	73.40	80.10	77.80	0.8 m	256 × 256	12,000
Ours	98.67	90.13	88.11	87.96	83.32	0.1 m	512 × 512	3788

The TransPV vision transformer-based model by Guo et al., when validated on a subset of the BDPV dataset (a combination of IGN and Google Earth datasets also used in this work), demonstrated the generalization capabilities with an IoU of 74.52%, an accuracy of 84.06%, an F1-score of 85.40%, a precision of 86.78%, and a recall of 84.06%. Although the BDPV dataset is used only for generalization validation, the model's performance was improved with random classic augmentations during preprocessing for training data, using augmentations such as rotations, scaling, color shifts, and application of Gaussian

blur to prevent overfitting. While the results may not be objectively comparable to our solution because of training dataset differences, using the BDPV dataset for GAN training to create more samples is encouraged, as it benefits the model more when compared with the usage of similar classic augmentations. The 3D-PV-Locator by Mayer et al. relies on DeepLabV3 with a ResNet-101 backbone, in contrast with the ResNet-50 backbone used in this work. The segmentation dataset has a similar number of training and testing samples (3222 and 403, respectively) compared with ours (3276 and 256, respectively), although the used dataset differs from ours. While the GSD is the same, the image size is smaller (320 × 320 compared with 512 × 512). Our work demonstrates a higher average IoU (83.32% compared with 74.10%) and average precision (90.13% compared with 87.30%), while the recall and F1 score metrics are only slightly higher. However, this comparison also lacks some objectiveness, as augmentation solutions are not mentioned. Additionally, Mayer et al. note that their comparison of related approaches that use different datasets makes the comparability limited, but displays the model's performance in a comparable range to that of the state of the art [36]. Yang et al. also used the same dataset sourced from Google Earth and IGN and the architecture based on DeepLabV3+; however, their approach combines weakly supervised and semi-supervised learning. Therefore, their results from fully supervised experiments are used to compare our work. In this instance, only the F1 score is higher by 2.66%, while other metrics are lower. The superior metrics may also be impacted by the authors producing more accurate annotations for each image in the dataset and performing manual screenings. Zhu et al. used the DeepLabV3+ architecture with the ResNetV1c and ResNeSt backbone for their detail-oriented deep learning network for refined segmentation, and the generalization capability was tested on the Jiangsu province dataset. The detail-oriented network features higher recall, F1 score, and IoU metrics but lower accuracy and precision. While data augmentations are not acknowledged, the high dataset quality, compared with the one gathered by the authors, is noted, mainly due to higher IoU metrics compared with test results with the authors' dataset. Jiang et al. also trained their segmentation model using the Jiangsu province dataset and applied basic data augmentations such as image rotations and flipping to randomly selected 30% of the original samples. The model achieved an average accuracy of 91.80%, precision of 90.20%, recall of 73.40%, F1 score of 80.10%, and IoU of 77.80%. The authors' model outperformed ours only in terms of a slight increase in precision; however, the nature of their work differs from ours, as Jiang et al. focused on estimations of rooftop solar panel power generation. Overall, although the compared works of other authors utilize various techniques that are not easily comparable, the common goal of solar panel semantic segmentation, as well as the same used datasets and similar architectures, brings insight into how various solutions for the same datasets produce varying results. Because this work focuses on applying an optimal amount of GAN-generated data determined by the sensitivity analysis results for better semantic segmentation model generalization and accuracy, comparing its results with other works straightforwardly is challenging. However, it could encourage combining this technique with semantic segmentation architecture improvements and parameter optimizations. Performing a sensitivity analysis for GAN-generated images would bring even more insight into how much data are needed for effective model improvements.

The class imbalance issue is relevant when training the GAN for data augmentations and 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 than pixel accuracy. This is relevant when segmenting the solar panels and performing other critical analyses such as fault detections [40]. Additionally, due to the nature of resampling datasets to a standard spatial resolution and resizing the images to a target image resolution (by either 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, paying more attention to the validation loss

metric and the validation IoU metric would be recommended. 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. Higher-quality datasets and more careful labeling are required to combat this issue.

Although the result of this work is the successful improvement of the semantic segmentation model using additional data generated using the GAN, it can be improved even further with additional computational and time resources. If resource limitations were absent, a more detailed sensitivity analysis could be performed for future work. With more time and computational resources, a more thorough sensitivity analysis with stochastic simulations and the 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 number of generated synthetic remote sensing images to use for model training. 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 adjusting the learning rate and training epoch amount, 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 datasets. A precise comparison of the model's performance with the works of other authors was also limited due to a lack of a unified dataset for solar panel segmentation model benchmarking. Although the comparisons were made with the models trained using similar datasets and similar architectures, the differences in training/validation subsets and architectures across the other works make the comparison more limited. However, this does not impact the novelty of this work, and the additional benchmarking dataset would be primarily beneficial for a model comparison with other proposed state-of-the-art solutions.

In addition to using the pix2pix generative adversarial network for additional training data synthesis, its variety can be improved with style transfer techniques, for example, using the CycleGAN [41] model. Varying environmental conditions are common in remote sensing data, and for the semantic segmentation model to perform better with such a variety of imagery, it would be beneficial to apply unpaired image-to-image translation. For example, translating both existing and newly created remote sensing images of solar panel installations into such that display different lighting or weather conditions (such as rain or snow) would not only introduce a larger variety of training data but also increase their quantity even further. Nevertheless, the application of an additional generative adversarial network would have to be done carefully to maintain the quality of the generated samples. Furthermore, it would require more time and additional computational resources to train, and the impact of generated additional training samples on the performance of the semantic segmentation would have to be researched–ideally with a thorough sensitivity analysis.

5. Conclusions

In summary, this study demonstrates the significant advantages of using generative adversarial networks for augmenting training datasets with synthetic remote sensing images, particularly for the semantic segmentation of solar panel installations. Our sensitivity analysis identified that augmenting the training dataset by 60% with GAN-generated images produces optimal results. This augmentation improved all metrics, including average accuracy, precision, recall, F1 score, and intersection over union (IoU), while reducing the average loss. The findings highlight that expanding a limited dataset with GAN-generated images improves the performance of the semantic segmentation model when tested with real-world data. This approach proves more beneficial than traditional data augmentations,

as evidenced by the improved segmentation of solar panels in various scenarios. The application of the pix2pix GAN reduces the need for extensive manual data labeling and introduces a novel method for generating diverse and realistic training data.

For future work, we recommend further optimization of both the segmentation model and the GAN parameters. This includes fine-tuning hyperparameters, exploring different GAN architectures like CycleGAN for style transfer, and conducting more comprehensive sensitivity analyses with multiple runs to ensure the robustness of the results. Additionally, addressing the class imbalance issue and improving the quality of generated images are crucial steps for enhancing the overall performance and reliability of the segmentation model. Finally, we recommend the creation of a universal benchmark dataset for an even more accurate evaluation of solar panel segmentation models and their performance comparisons.

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.

Data Availability Statement: The original contributions presented in this study are included in this article; further inquiries can be directed to the corresponding author.

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

1. 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), Muscat, Oman, 15–16 January 2019; pp. 1–5.
2. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015, Boston, MA, USA, 7–12 June 2015.
3. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015.
4. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 2481–2495. [CrossRef] [PubMed]
5. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
6. 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), Long Beach, CA, USA, 7–8 November 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6.
7. 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 Sens.* **2022**, *14*, 4211. [CrossRef]
8. Sait, M.; Erguzen, A.; Erdal, E. Using Mask R-CNN to Isolate PV Panels from Background Object in Images. *Int. J. Trend Sci. Res. Dev.* **2020**, *5*, 1191–1195.
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 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, Hamburg, Germany, 13 November 2023; ACM: Hamburg, Germany, 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), Naples, Italy, 5–7 December 2022; IEEE: Piscataway, NJ, USA, 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), Montreal, QC, Canada, 21–25 August 2022; IEEE: Piscataway, NJ, USA, 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), Dharan, Nepal, 10–12 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. [CrossRef]

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. [CrossRef]

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 Obs. Remote Sens.* **2021**, *14*, 2387–2402. [CrossRef]

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), Chengdu, China, 19–21 August 2022; IEEE: Piscataway, NJ, USA, 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. *J. Crit. Rev.* **2020**, *7*, 2864–2872.

18. Alomar, K.; Aysel, H.I.; Cai, X. Data Augmentation in Classification and Segmentation: A Survey and New Strategies. *J. Imaging* **2023**, *9*, 46. [CrossRef] [PubMed]

19. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]

20. Frid-Adar, M.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. Synthetic data augmentation using GAN for improved liver lesion classification. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA , 4–7 April 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 289–293. [CrossRef]

21. Luo, Z.; Cheng, S.Y.; Zheng, Q.Y. GAN-Based Augmentation for Improving CNN Performance of Classification of Defective Photovoltaic Module Cells in Electroluminescence Images. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Macao, China, 21–24 July 2019; Volume 354, p. 012106. [CrossRef]

22. 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**.

23. Henry, J.; Natalie, T.; Madsen, D. *Pix2Pix GAN for Image-to-Image Translation*; ResearchGate Publisher: Berlin, Germany, 2021; pp. 1–5.

24. 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), Bangkok, Thailand, 30 October–1 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 310–316.

25. 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 (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.

26. 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 Syst. Sci. Data* **2021**, *13*, 5389–5401. [CrossRef]

27. 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. [CrossRef] [PubMed]

28. Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2015**, arXiv:1405.0312.

29. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]

30. 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. [CrossRef]

31. Pamungkas, R.F.; Utama, I.B.K.Y.; Jang, Y.M. A Novel Approach for Efficient Solar Panel Fault Classification Using Coupled UDenseNet. *Sensors* **2023**, *23*, 4918. [CrossRef] [PubMed]

32. Ravishankar, R.; AlMahmoud, E.; Habib, A.; de Weck, O.L. Capacity Estimation of Solar Farms Using Deep Learning on High-Resolution Satellite Imagery. *Remote Sens.* **2023**, *15*, 210. [CrossRef]

33. 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 Sens.* **2023**, *15*, 5232. [CrossRef]

34. da Costa, M.V.C.V.; de Carvalho, O.L.F.; Orlandi, A.G.; Hirata, I.; de Albuquerque, A.O.; e Silva, F.V.; Guimarães, R.F.; Gomes, R.A.T.; Júnior, O.A.d.C. Remote Sensing for Monitoring Photovoltaic Solar Plants in Brazil Using Deep Semantic Segmentation. *Energies* **2021**, *14*, 2960. [CrossRef]

35. Guo, Z.; Lu, J.; Chen, Q.; Liu, Z.; Song, C.; Tan, H.; Zhang, H.; Yan, J. TransPV: Refining photovoltaic panel detection accuracy through a vision transformer-based deep learning model. *Appl. Energy* **2024**, *355*, 122282. [CrossRef]

36. Mayer, K.; Rausch, B.; Arlt, M.-L.; Gust, G.; Wang, Z.; Neumann, D.; Rajagopal, R. 3D-PV-Locator: Large-scale detection of rooftop-mounted photovoltaic systems in 3D. *Appl. Energy* **2022**, *310*, 118469. [CrossRef]

37. Yang, R.; He, G.; Yin, R.; Wang, G.; Zhang, Z.; Long, T.; Peng, Y. Weakly-semi supervised extraction of rooftop photovoltaics from high-resolution images based on segment anything model and class activation map. *Appl. Energy* **2024**, *361*, 122964. [CrossRef]

38. Zhu, R.; Guo, D.; Wong, M.S.; Qian, Z.; Chen, M.; Yang, B.; Chen, B.; Zhang, H.; You, L.; Heo, J.; et al. Deep solar PV refiner: A detail-oriented deep learning network for refined segmentation of photovoltaic areas from satellite imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *116*, 103134. [CrossRef]

39. Jiang, H.; Yao, L.; Lu, N.; Qin, J.; Liu, T.; Liu, Y.; Zhou, C. Geospatial assessment of rooftop solar photovoltaic potential using multi-source remote sensing data. *Energy AI* **2022**, *10*, 100185. [CrossRef]

40. Jumaboev, S.; Jurakuziev, D.; Lee, M. Photovoltaics Plant Fault Detection Using Deep Learning Techniques. *Remote Sens.* **2022**, *14*, 3728. [CrossRef]

41. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2242–2251.