



## Article

# Weakly Supervised Learning for Transmission Line Detection Using Unpaired Image-to-Image Translation

Jiho Choi and Sang Jun Lee \*

Division of Electronic Engineering, Jeonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju 54896, Korea; jihochoi@jbnu.ac.kr

\* Correspondence: sj.lee@jbnu.ac.kr; Tel.: +82-63-270-2463

**Abstract:** To achieve full autonomy of unmanned aerial vehicles (UAVs), obstacle detection and avoidance are indispensable parts of visual recognition systems. In particular, detecting transmission lines is an important topic due to the potential risk of accidents while operating at low altitude. Even though many studies have been conducted to detect transmission lines, there still remains many challenges due to their thin shapes in diverse backgrounds. Moreover, most previous methods require a significant level of human involvement to generate pixel-level ground truth data. In this paper, we propose a transmission line detection algorithm based on weakly supervised learning and unpaired image-to-image translation. The proposed algorithm only requires image-level labels, and a novel attention module, which is called parallel dilated attention (PDA), improves the detection accuracy by recalibrating channel importance based on the information from various receptive fields. Finally, we construct a refinement network based on unpaired image-to-image translation in order that the prediction map is guided to detect line-shaped objects. The proposed algorithm outperforms the state-of-the-art method by 2.74% in terms of F1-score, and experimental results demonstrate that the proposed method is effective for detecting transmission lines in both quantitative and qualitative aspects.

**Keywords:** weakly supervised learning; image-to-image translation; transmission line detection; unmanned aerial vehicle; channel attention mechanism



**Citation:** Choi, J.; Lee, S.J. Weakly Supervised Learning for Transmission Line Detection Using Unpaired Image-to-Image Translation. *Remote Sens.* **2022**, *14*, 3421. <https://doi.org/10.3390/rs14143421>

Academic Editors: Yingying Dong, Chenghai Yang, Giovanni Laneve and Wenjiang Huang

Received: 7 June 2022

Accepted: 13 July 2022

Published: 16 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Recently, unmanned aerial vehicles (UAVs) have been widely utilized in many industrial fields. In the applications of construction site monitoring and infrastructure inspection, UAVs reduce the cost and inspection time while ensuring the safety of inspectors [1]. Furthermore, UAVs contribute to increasing the efficiency of precision agriculture by spreading seeds and monitoring the conditions of crops more effectively than human workers [2]. Beyond this, UAVs are also applied to military surveillance, aerial photography, search and rescue, and product delivery [3]. As technology advances, UAVs with high-level autonomy have been utilized in applications such as forest fire monitoring and security system.

For the reliable operation of autonomous UAVs, obstacle detection and avoidance are important functions. Most autonomous UAVs and drones equip cameras for visual recognition, and path planning and control are successfully conducted based on the accurate recognition of surrounding environments. Previous studies proposed several obstacle detection methods. In particular, Huang et al. [4] proposed an obstacle avoidance algorithm using a monocular camera and millimeter-wave radar together. Similarly, a deep learning-based recognition algorithm was employed to detect multiple obstacles in [5]. On the other hand, many researchers have conducted research on obstacle avoidance and path planning based on deep learning approaches [6–11]. Ou et al. [9] suggested a framework based on deep reinforcement learning to plan feasible global paths with an obstacle map. Yuan et al. [10] presented a path planning method based on a convolutional neural network (CNN) model that can detect and localize obstacles such as buildings.

Data collection has become more straightforward than ever, and with the support of extensive public datasets, deep learning techniques have shown promising results in various industrial fields. However, supervised learning methods require ground truth data, which cause expensive costs for manual labor and time-consuming tasks, especially for large datasets. To address these limitations, weakly supervised learning has received attention to train deep neural networks with weak supervision. In recent years, weakly supervised learning has been applied to various tasks such as object detection [12–14], segmentation [15–19], and localization [20–22]. Wang et al. [16] suggested a segmentation algorithm based on a combination of U-Net and class activation map and trained using only image-level labels. They demonstrated that training a CNN model with weak supervision can also segment cropland accurately.

Among several types of obstacles, the transmission line is a critical obstacle that must be avoided. A collision with transmission lines can damage the stable supply of electricity, and furthermore, crashed UAVs can cause secondary accidents. Deep learning methods have been successfully applied to the detection and localization of transmission lines [21–25]. In the case of transmission line datasets, class imbalance occurs due to the background area that occupies most of the aerial image. In [23], they proposed a generalized focal loss function to handle class imbalance in the transmission line detection task. Lee et al. [21] introduced a weakly supervised learning method for detecting transmission lines, and they employed the *VisualBackProp* algorithm proposed by Bojarski et al. [26] to localize transmission lines. After this study, Choi et al. [22] proposed an extended study for transmission line detection. They utilized only patch-level labels to reduce the cost for collecting pixel-level ground truth data. In [22], they used the assumption that transmission lines are partially straight, and their proposed method connects broken lines by utilizing the orientations of the line segments. However, the class labels of the patch images require rough location information of transmission lines in images.

In this paper, we propose a transmission line detection method based on weakly supervised learning and unpaired image-to-image translation. The main contribution of this paper is three-fold as follows.

- We develop a weakly supervised algorithm for detecting transmission lines in UAV images. Unlike the previous methods, which require pixel-level labels, our proposed method requires minimal labeling work for preparing training data, and therefore, it is easily applicable to real-world problems.
- We integrate a novel attention module into the classification network to obtain a robust localization mask. To incorporate the information from various receptive fields, we introduce a parallel dilated attention (PDA) module.
- For the training of the refinement network, we generate pseudo-line data and employ the cycle consistency loss, which was proposed in [27]. The refinement network enhances the line-shaped property of transmission lines, and therefore, the localization result is significantly improved in both quantitative and qualitative aspects.

The remainder of this paper is structured as follows: the related work is summarized in Section 2. The proposed method is presented in Section 3. Results and conclusions are presented in Sections 4 and 5, respectively.

## 2. Related Work

### 2.1. Attention Mechanism

The attention mechanism has attracted many deep learning researchers, and they have proposed the following mechanisms. The bottleneck attention module (BAM) [28] is an attention module that computes attention maps from the two separated spatial and channel attention branches. Unlike the parallel structure of BAM, the spatial and channel attention of convolutional block attention modules (CBAM) [29] are sequentially configured. Yang et al. [30] proposed SimAM, which is a parameter-free attention module to calculate 3D attention weights in the channel and spatial dimensions. Hu et al. [31] tried to reduce the amount of computation while performing recalibration with a squeeze and excitation

(SE) module, although this operation destroys the relationship between channels and weights due to channel reduction. The SE module utilizes the global average pooling (GAP) and two fully connected layers, which can be integrated into CNN architectures such as VGGNet, ResNet, and GoogLeNet. Wang et al. [32] proposed an efficient channel attention (ECA) to compute local cross-channel interaction by applying 1D convolution. The ECA module has shown significant improvements in state-of-the-art object detection, image classification, and object segmentation along with lightweight parameters. We compare the ECA module and other channel attention methods [30,31] with the proposed attention module and demonstrate the effectiveness of our proposed attention module in Section 4.5.

It has been used in a variety of applications concurrently with the progress of many studies on the attention mechanism. From the perspective of time series data, attention mechanisms can be implemented with sequence-to-sequence models with encoder and decoder architectures, to make models that pay attention to specific sequence data [33–35]. Furthermore, the attention module was applied to the CNN model which uses time-series data to estimate the blood pressure [36] and classifies the sleep stage in [37]. Many studies attempted to solve diverse problems on remote sensing image data such as classifications [38,39], ship detection [40,41], and semantic segmentation [42]. Ma et al. [39] implemented the channel and spatial attention module and integrated it into CNN architecture for the classification of the remote sensing scene images. Detecting small-scale ships is a challenging task in optical remote sensing images. Hu et al. [40] proposed detection models with the attention mechanism to suppress background while focusing on small ships to improve detection accuracy. Moreover, refs. [43,44] integrate attention modules in their deep neural network to solve segmentation tasks such as esophagus and lungs in medical images. Motivated by channel attention, we improved the localization mask by adopting an attention mechanism to focus on the important channels.

## 2.2. Image-to-Image Translation

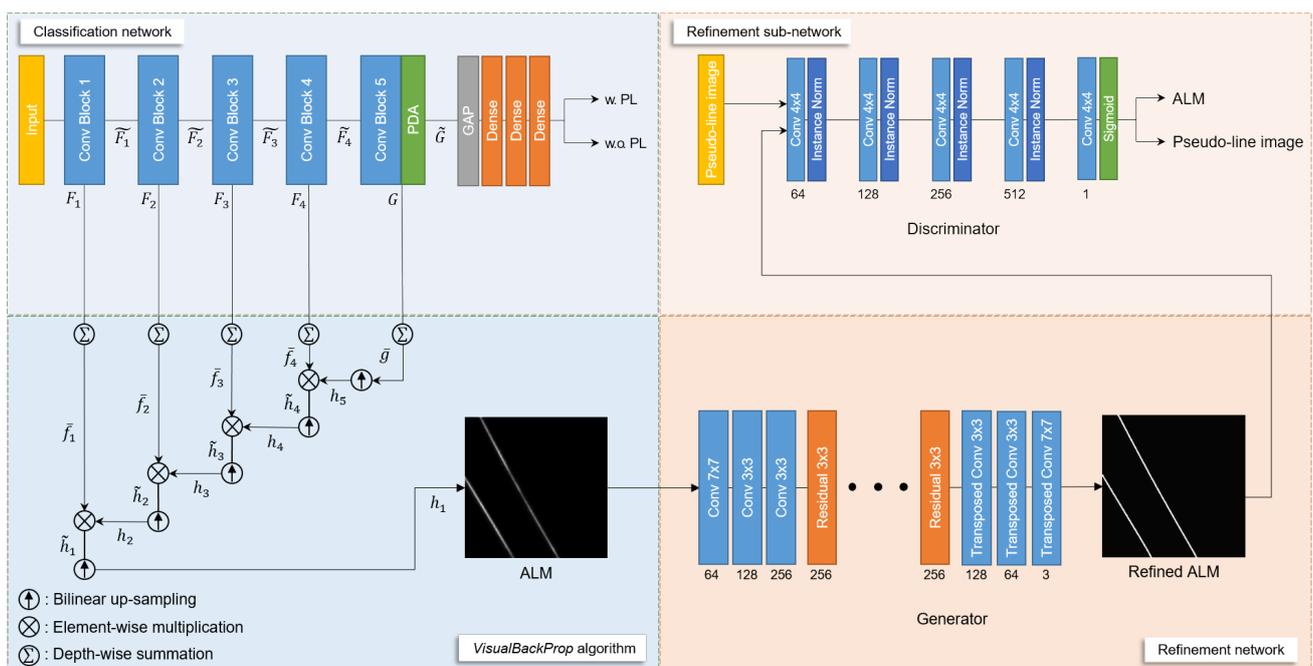
Generative adversarial networks (GAN) [45] have made great strides in deep learning, and subsequent algorithms such as deep convolutional GAN (DCGANs) [46], conditional GAN (CGAN) [47], and InfoGAN [48] have been proposed. Isola et al. [47] proposed a supervised learning method that performs image-to-image translation using a paired dataset. Paired image-to-image translation is restrictive in its applications to real-world problems because it requires data correspondences. On the other hand, unpaired image-to-image translation techniques can address the limitations of paired image-to-image translation methods. There are several style transfer networks including CycleGAN [27], DiscoGAN [49], and DualGAN [50], and these methods translate the style of input images based on the unpaired datasets. Zhu et al. [27] proposed CycleGAN, which translates images from the source domain to the target domain with the cycle consistency loss, and this method shows remarkable results in the style transfer task. Researchers have extended unpaired image-to-image translation for several applications [51–58] to address issues such as data imbalance, lack of diversity, and limitation in collecting real paired dataset.

Zi et al. [51] constructed a modified CycleGAN to effectively generate clear images from cloudy images by utilizing unpaired image datasets. Furthermore, CycleGAN has been applied for data augmentation by translating synthetic images into realistic images in [52,53]. In particular, Mao et al. [53] improved the performance of classifying the actual wildfire smoke by utilizing images that were artificially generated images by utilizing [27]. To safely drive autonomous vehicles, road surface detection is essential as knowledge of road surface conditions (e.g., dry, wet, snowy) affects autonomous driving control [54]. Dry conditions can be collected more frequently, resulting in unbalanced data problems. To address this lack, they generated images of wet and snowy roads through an unpaired image-to-image translation method. Although biometric systems (e.g., fingerprint-based, and face-based) are widely used for security purposes, these recognition systems can be vulnerable to presentation attacks. A presentation attack aims to interfere with the normal functioning of a biometric recognition by presenting artifacts or biometric characteristics.

To prepare for such a presentation attack detection (PAD), Nguyen et al. [55] generalized the model to fake presentation attack face images obtained via CycleGAN. Inspired by these studies, we employed the image-to-image translation to refine the localization mask using pseudo-line data, and the effectiveness of a refinement network is shown in an ablation study later on.

### 3. Proposed Method

This section presents the proposed transmission line detection method. The weakly supervised learning framework proposed by Bojarski et al. [26] is employed to generate a localization mask for the transmission lines. Different from the previous work, we introduce a novel attention mechanism called PDA to improve the quality of the localization mask, and it is called attention localization mask (ALM). Furthermore, we develop a refinement network by utilizing an unpaired image-to-image translation technique between ALM and pseudo-line data. An overview of the proposed framework is shown in Figure 1. The upper and lower left parts present the backbone network for classifying images with and without transmission lines and the process for generating ALM from the hierarchical feature maps, respectively. The lower right part is the generator of the refinement network that produces the refined image, and the upper right part is a discriminator for adversarial training of the refinement network.



**Figure 1.** Overview of the proposed method. The (left) part shows the base network for extracting power line (PL) features and obtaining localization mask. The (right) part presents the refinement network and a sub-network for the adversarial learning of the refinement network.

#### 3.1. Classification Network and VisualBackProp Algorithm

We constructed a classification network to implement the *VisualBackProp* algorithm [26], which can obtain localization masks using feature maps. The classification network was constructed based on the VGG16 architecture, which consists of five convolution blocks to classify images with and without transmission lines. A convolution block contains convolution layers, a rectified linear unit (ReLU), and a max pooling operation. Although similar models were utilized in [21,22] for localizing transmission lines in patch images, our proposed model is different from the previous methods in that transmission lines can be localized in the original images. We employed image-level labels of the same size as the original size of the  $512 \times 512$  in the power line dataset.

After binary classification, the localization mask is generated by the *VisualBackProp* algorithm. In order to obtain the mask for localizing transmission lines, we employed the *VisualBackProp* algorithm for the last feature maps  $\mathbf{F}_i \in \mathbb{R}^{H_i \times W_i \times C_i}$  of the convolution blocks. The  $i$ -th feature map  $\mathbf{F}_i$  consists of  $C_i$  feature maps  $\mathbf{f}_i^1, \dots, \mathbf{f}_i^{C_i}$ . The first process of the *VisualBackProp* algorithm is to compute a single feature map  $\tilde{\mathbf{f}}_i$  by accumulating the feature maps  $\mathbf{f}_i^k \in \mathbb{R}^{H_i \times W_i}$  in the depth direction as (1).

$$\tilde{\mathbf{f}}_i = \sum_{k=1}^{C_i} \mathbf{f}_i^k, \quad (1)$$

where  $i$  and  $c_i$  denote the number of convolution blocks and channels, respectively. The accumulated feature map is upsampled via bilinear interpolation to generate  $\mathbf{h}_i$ , and it is multiplied with the previous feature map  $\tilde{\mathbf{f}}_{i-1}$  to compute  $\tilde{\mathbf{h}}_{i-1}$  as (2).

$$\tilde{\mathbf{h}}_{i-1} = \tilde{\mathbf{f}}_{i-1} \otimes \mathbf{h}_i, \quad (2)$$

where  $\otimes$  is the elementwise multiplication operation. Although the *VisualBackProp* algorithm provides reasonable localization masks in many cases, it fails when transmission lines have weak visual properties. To address this problem, we propose an attention mechanism for weakly supervised learning to enhance the responses of transmission lines in the localization map. In Figure 1,  $\tilde{\mathbf{F}}_i \in \mathbb{R}_{+}^{\frac{H_i}{2} \times \frac{W_i}{2} \times C_i}$  is the output of  $i$ -th convolution block in VGG16, which can be obtained after the ReLU and max pooling operations for  $\mathbf{F}_i$ , and it is utilized as the input for the following convolution block.

### 3.2. Parallel Dilated Attention Module

Inspired by SE-Net [31] and ECA-Net [32], we introduce a novel channel attention module called PDA. SE-Net [31] conducts dimensionality reduction in fully connected layers to reduce computational load, and Wang et al. [32] utilizes 1D convolution instead of fully connected layers to reduce model complexity without dimensionality reduction. Although ECA-Net employs a kernel size, which can be adaptively determined through a mapping function, this module still has the limitation that its receptive field is fixed. Figure 2 presents the structure of the proposed PDA module. Compared to the previous methods, the PDA module consists of three lightweight 1D convolutions with different dilation ratios, and therefore, the proposed attention mechanism can merge the information from various receptive fields.

In PDA, the GAP is applied to the feature map to acquire a vector with the size of  $1 \times 1 \times C$ , where  $C$  denotes the number of channels. A feature vector of the identical length to the channel size of the previous feature map is diverged and provided as parallelized 1D convolutions. In Figure 2,  $D$  indicates the dilation ratio of the 1D convolution. The parallel structure of 1D convolutions is applied to obtain information from various receptive fields depending on dilation ratios. The various dilation ratios are advantageous for obtaining abundant features as they can collect information from narrow to broad receptive fields.

In this experiment, we set the dilation ratios to 1, 2, and 4, respectively, and padding was set equal to the dilation ratio to acquire an output vector of the same length as the input. The output vectors of the 1D convolution, which contain different locally correlated channel information, are concatenated together to aggregate the information of the parallel operation with a vector size of  $1 \times 1 \times 3C$ . A fully connected layer compresses meaningful information and learns interdependencies between channels. A sigmoid function is utilized to obtain an attention vector  $\mathbf{a}$  that contains values between 0 and 1, and it is expressed as follows:

$$\mathbf{a} = [a_1, \dots, a_{c_5}], \quad 0 \leq a_k \leq 1. \quad (3)$$

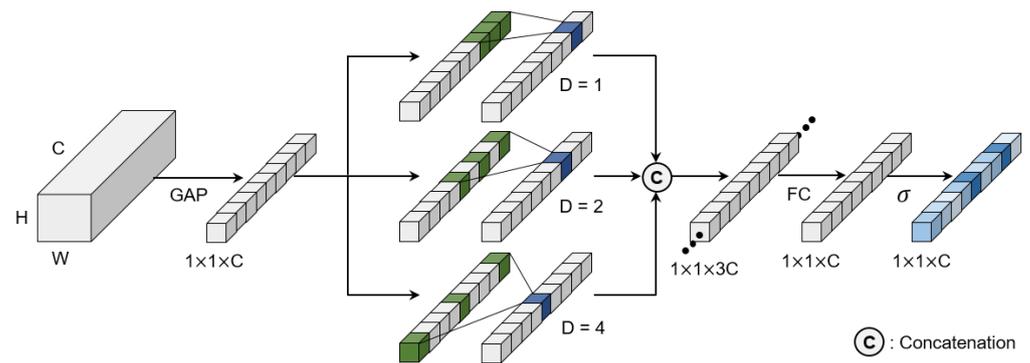


Figure 2. Structure of PDA module.

The attention vector  $\mathbf{a}$  is computed from the last convolution block, and the length of the attention vector is equal to the number of channels. Each component of  $\mathbf{a}$  indicates the significance of the corresponding channel, and therefore, PDA guides the network to focus on important features for classifying and localizing transmission lines. The localization mask based on PDA is called ALM.

Figure 3 shows details of the PDA module, which is placed between the last convolution feature map and ReLU operation. The last convolution map  $\mathbf{F}_5$  consists of  $\mathbf{f}_5^1, \dots, \mathbf{f}_5^{c_5}$ , and the  $k$ -th channel  $\mathbf{g}^k$  of the weighted feature map  $\mathbf{G} \in \mathbb{R}_+^{\frac{H_5}{2} \times \frac{W_5}{2} \times c_5}$  can be computed as

$$\mathbf{g}^k = a_k \otimes \mathbf{f}_5^k, \tag{4}$$

where  $\otimes$  is the elementwise multiplication.

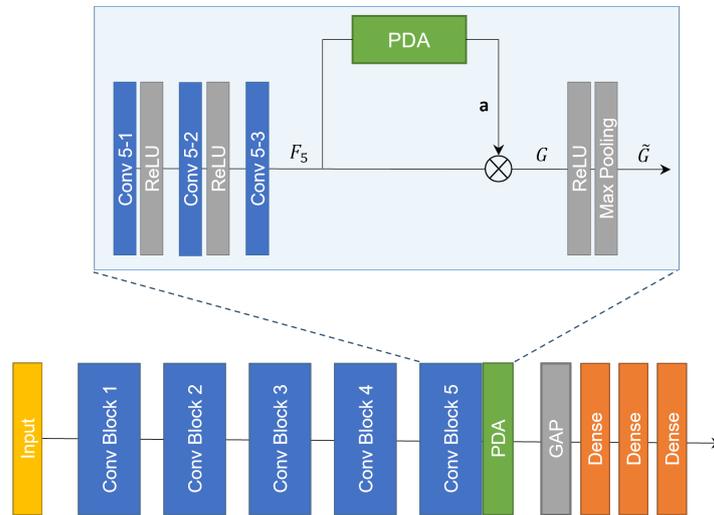


Figure 3. PDA module in the classification network.

The attention-weighted feature map is utilized for transmission line localization. Different from the summation operation in the *VisualBackProp* algorithm, we conducted the weighted summation with the components of the attention vector computed from the PDA module. The attention-weighted feature map is beneficial to focus on important features for localizing transmission lines. In the same way as (1), the weighted feature map  $\mathbf{G}$  is accumulated in the depth direction to compute the accumulated feature map  $\tilde{\mathbf{g}}$  as

$$\tilde{\mathbf{g}} = \sum_{k=1}^{c_5} \mathbf{g}^k. \tag{5}$$

Whereas feature maps are added in the channel direction in the original VBP algorithm, our PDA module computes a channel attention vector to more effectively aggregate the

information in the feature maps. To incorporate  $\bar{\mathbf{g}}$  with other feature maps,  $\bar{\mathbf{g}}$  is upsampled to the identical size as the previous feature map  $\bar{\mathbf{f}}_4$ , and this procedure can be expressed as follows:

$$\mathbf{h}_5 = \text{Upsampling}(\bar{\mathbf{g}}), \quad (6)$$

where *Upsampling* indicates the bilinear interpolation, and the result is denoted by  $\mathbf{h}_5$ . Through the equation of (2), the elementwise multiplication is conducted between  $\mathbf{h}_5$  and  $\bar{\mathbf{f}}_4$ , and the process is repeated until obtaining  $\mathbf{h}_1$ , which is called ALM.

### 3.3. Refinement Network via Image-to-Image Translation

The ALM in the previous step still has room for improvement due to blurred line responses and weak connections of transmission lines in the localization mask. Therefore, we constructed a refinement network to transfer visual characteristics of transmission lines. The refinement network employs a generator of an image-to-image transformation architecture. To transfer line-shaped properties, we generated a dataset based on a rule-based algorithm, and it is called the pseudo-line dataset. The pseudo-line dataset includes 250 binary pseudo-line images with the image size of  $512 \times 512$ , which is identical to the size of ALM. Figure 4 shows examples of pseudo-line data.

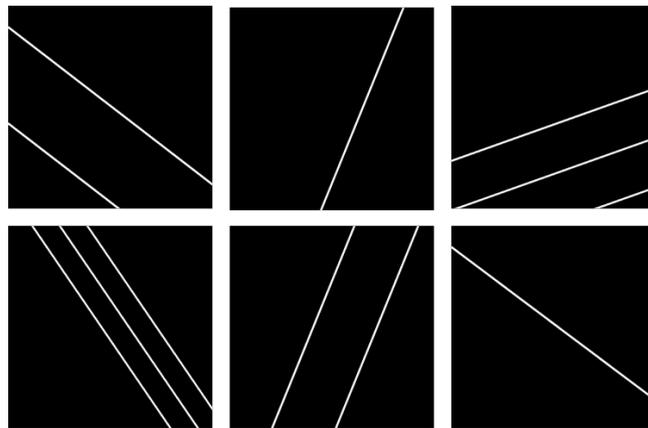


Figure 4. Examples of pseudo-line dataset.

Figure 5 presents the procedure for generating the pseudo-line dataset. We prepared an image filled with zeros and generated pseudo-lines with randomly selected pixel coordinates  $a$  and  $b$ , which are integers ranging from 0 to 512. Because multiple lines which are too close to each other are not desirable, we generated pseudo-lines that connect opposite sides of images. Since the power line dataset contains multiple transmission lines, we also generated the pseudo-line dataset with variable numbers of transmission lines. In Figure 5,  $i$  denotes an arbitrary line interval, and multiple lines are generated with the distance of  $i$ . By utilizing the properties of the pseudo-line dataset, we refine the ALM to improve the localization accuracy.

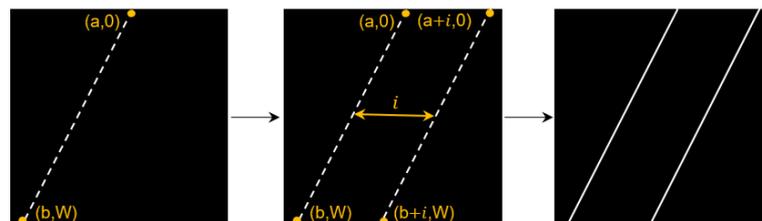


Figure 5. Procedure for generating pseudo-line dataset.

To construct a refinement network, we adopt the structure of CycleGAN [27]. We defined the ALM as a source domain  $S$ , and the pseudo-line dataset as a target domain  $T$ , respectively. The purpose of the refinement network is to obtain a mapping

function from the source domain to the target domain. To train the refinement network, we constructed two generators and two discriminators for the adversarial training of the unpaired image-to-image translation framework. The first generator that maps from the source to the target domain is denoted by  $G : S \rightarrow T$ , and it is presented in the lower right part of Figure 1. In some images, ALMs show weak responses and missing parts for localizing transmission lines. To address these limitations, the generator  $G$  allows the ALMs to contain the properties of pseudo-line images, and it is utilized as the refinement network. The refinement network restores weak responses and missing parts of ALMs based on the line-shaped properties of pseudo-line images. The first discriminator  $D_T$  is trained to distinguish between  $G(s)$  and pseudo-line dataset in the target domain  $T$ , where  $G(s)$  is generated images from the source domain. As the adversarial training proceeds, the generator  $G$  becomes creating realistic line images, and these result images are called refined ALMs. The discriminator  $D_T$  is presented at the upper right part in Figure 1, and the adversarial loss for training  $G$  and  $D_T$  is defined as

$$L_{adv}(G, D_T, S, T) = E_{t \sim p_{data}(t)}[\log(D_T(t))] + E_{s \sim p_{data}(s)}[\log(1 - D_T(G(s)))] \tag{7}$$

where  $p_{data}(t)$  and  $p_{data}(s)$  are the distributions of the source and target domains.

Similarly, the second generator  $F : T \rightarrow S$  maps target domain data into the source domain, and the second discriminator  $D_S$  distinguishes ALMs and reconstructed ALMs in the source domain. Another adversarial loss for the training of  $F$  and  $D_S$  is defined as

$$L_{adv}(F, D_S, T, S) = E_{s \sim p_{data}(s)}[\log(D_S(s))] + E_{t \sim p_{data}(t)}[\log(1 - D_S(G(t)))] \tag{8}$$

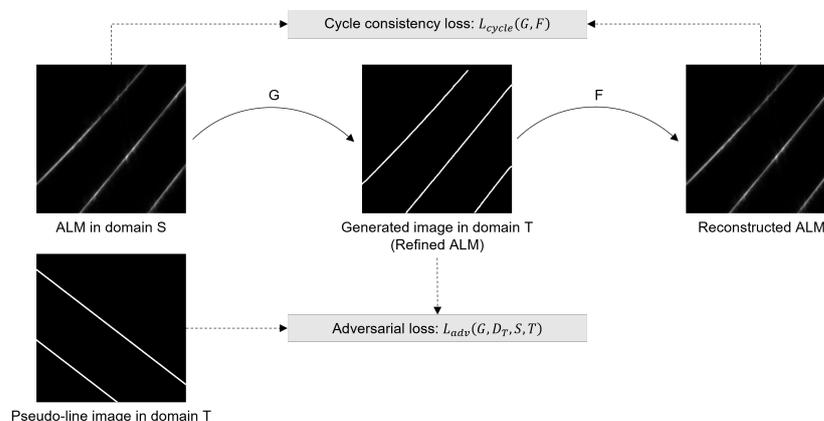
Since using only adversarial losses can cause mode collapse, the cycle consistency loss is employed to train the generators in a constrained space. Based on the cycle consistency loss, which is defined as (9), the generated images can be translated back to the original images as shown in Figure 6.

$$L_{cycle}(G, F) = E_{s \sim p_{data}(s)}[\|F(G(s)) - s\|] + E_{t \sim p_{data}(t)}[\|G(F(t)) - t\|] \tag{9}$$

The total loss function for training the refinement network is formulated as the combination of the adversarial losses and the cycle consistent loss, and it is defined as

$$L(G, F, D_S, D_T) = L_{adv}(G, D_T, S, T) + L_{adv}(F, D_S, T, S) + \lambda L_{cycle}(G, F) \tag{10}$$

where  $\lambda$  is a hyper-parameter for controlling the effect the cycle consistency loss, and we utilized  $\lambda = 10$  in experiments.



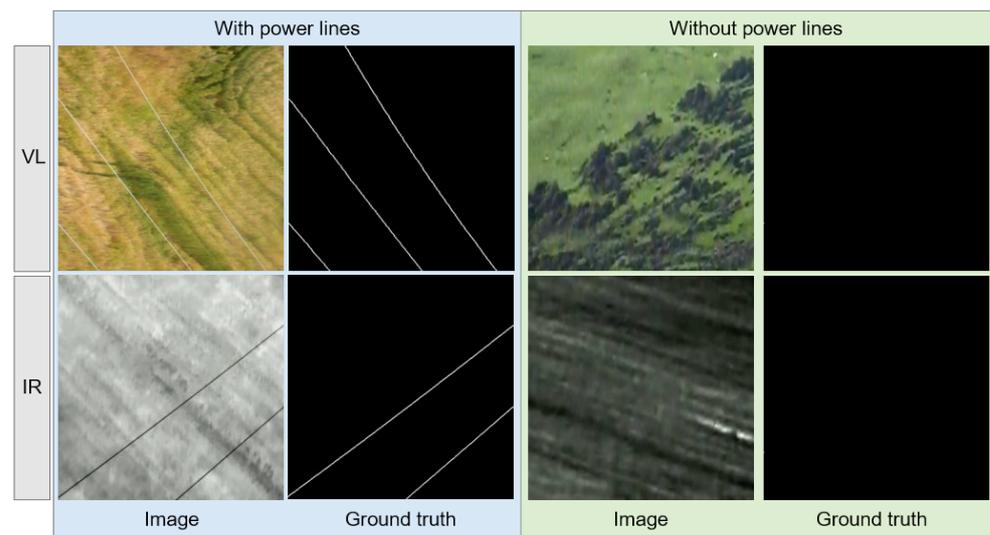
**Figure 6.** The process for generating refined ALMs. The generator  $G$  maps ALMs into the target domain, and therefore, responses of transmission lines are enhanced based on the line-shaped properties of the target domain data.

## 4. Experimental Results

We conducted experiments with the hardware environment including Intel Core i9-10900K CPU, 64 GB DDR4 RAM, and NVIDIA RTX 3090. The proposed algorithm was implemented based on PyTorch and OpenCV. To train the classification network, we utilized the Adam optimizer with the initial learning rate of 0.0001 and weight decay of 0.05. The refinement network was trained in the identical training setting, excluding the initial learning of 0.0002.

### 4.1. Dataset Description

We employed the public power line dataset consisting of 400 infrared (IR) and 400 visual light (VL) images. Figure 7 presents example images with and without transmission lines in the first and third columns, respectively. The ground truth corresponds to each image in the second and last columns. The dataset was collected for seasonal days in 21 different regions across Turkey in cooperation with the Turkish Electricity Transmission Company (TEIAS). The dataset is collected under diverse conditions, and therefore, it is challenging to recognize transmission lines in the image due to different backgrounds and illumination. In this study, 400 VL images with the size of  $512 \times 512$  were used, of which 200 VL images contain transmission lines, and the others do not. For the training of the base classification network, we split the dataset into 300, 50, and 50 images for the training, validation, and test sets, respectively.



**Figure 7.** Example images and their corresponding ground truth data.

### 4.2. Evaluation Measure

The proposed method was evaluated based on the criterion proposed by Choi et al. [22]. Recall, precision, and F1-score were computed from the true positive (TP), false positive (FP), and false negative (FN). In [22], TP is defined as the number of cases where more than 50% of line pixels are correctly detected. To define false responses that occur near a transmission line, the FP is defined based on a tolerance range. By setting the tolerance range as 10 pixels on both sides of a transmission line, FP includes incorrect responses occurring in background regions and thick predictions on transmission lines. If a predicted line is composed of less than 10 pixels, then it is regarded as noise.

### 4.3. Quantitative Evaluation

Table 1 presents the quantitative comparison of the proposed method with several previous algorithms. To represent differences in approaches, we categorized the results based on the learning types and annotation levels, and more specifically, annotation levels are divided into pixel, patch, and image levels. The learning type of weakly supervised

manner can be divided into patch and image-level annotations, which is a difference between the previous and our methods. In the previous methods, patch-level annotations were utilized by dividing the original images into  $128 \times 128$  sub-images and assigning class labels for these patch images.

**Table 1.** Quantitative comparison with other methods. The best result is highlighted in bold, and the second-best result is underlined. S and WS indicate supervised learning and weakly supervised learning, respectively.

Methods	Learning Type	Annotation Level	Recall (%)	Precision (%)	F1-Score (%)
Long et al. (2015) [59]	S	Pixel	98.17	93.04	95.54
Li et al. (2019) [60]	S	Pixel	97.25	95.50	96.36
Bojarski et al. (2016) [26]	WS	Patch	41.28	28.13	33.46
Lee et al. (2017) [21]	WS	Patch	86.24	<b>100</b>	92.61
Choi et al. (2021) [22]	WS	Patch	<b>98.17</b>	90.68	<u>94.27</u>
Ours	WS	Image	<u>97.90</u>	<u>96.15</u>	<b>97.01</b>

Although annotating patch images require less burden compared to pixel-level labels, patch-level annotations have limitations with respect to it requiring the location information of transmission lines in the original images. On the other hand, we utilized the entire size of the image for the training of the classification network. In transmission line images, lines are composed of small numbers of pixels and most of the remaining area consists of the background, and therefore, the larger the image size, the more difficult it is to localize and detect transmission lines. Nevertheless, we achieved quantitatively significant improvements by utilizing the entire images.

Compared with Choi et al. [22], our proposed method improves the performance by 5.47% and 2.75% in terms of precision and F1-score, respectively, while the recall is slightly lower by 0.27%. We utilized the entire image size from the beginning to the end of the algorithm, while [21,22] divide the image into patches once in the middle and merge them back together. In addition, in the process of dividing into patch images, approximate location information of transmission lines is required to train the classification network. It is worth noting that our proposed method does not require any location information. Furthermore, our method is meaningful in that the localization masks obtained from the feature maps can be improved by the refinement network. In another weakly supervised learning method, Lee et al. [21] achieved substantial performance in precision and F1-score, but there was a trade-off between precision and recall. By contrast, our proposed method has satisfactory results in both recall and precision for detecting transmission lines.

Table 1 also presents the accuracy of the segmentation algorithms proposed in [59,60]. Although the supervised learning methods show satisfactory performance in both recall and precision, these algorithms require time-consuming work for generating pixel-level annotations. To reduce the cost for preparing ground truth data, our proposed method adopts a weakly supervised learning framework. It is noteworthy that even though we only utilized image-level annotations, the proposed algorithm shows higher accuracy in terms of precision and F1-score.

#### 4.4. Ablation Study

We conducted an ablation study to demonstrate the effectiveness of each step of the proposed algorithm, and the results are presented in Table 2. We employed the algorithm proposed by Bojarski et al. [26], which is effective to represent clues in the feature map of a convolutional network, and it is considered as the baseline model and denoted as localization mask in Table 2. The attention vector obtained from the PDA contains the scores for each channel of the feature map, and the scores range from 0 to 1. This calculated score is multiplied by the feature map for channel-wise to give importance to each channel, and when performing binary classification of the input image, it is localized by focusing

on the learned features. By adding the PDA to the baseline, we reached an F1-score of 92.35%, which improved performance by 1.22% compared to the existing localization mask. The localization mask performance after the refinement process showed that the refinement network, which is part of the proposed method, is meaningful. In Table 2, the performance improved by 3.59% and 2.96%, respectively, in terms of recall and precision. These results show that the adoption of CycleGAN is adequate as a way to improve the localization mask through the process of transferring the characteristics of an ideal line fully connected from one point to another. The best performance is the result of applying both steps, and the proposed model outperforms the baseline with recall and precision of 97.90% and 96.15%, respectively. In other words, the performance improved by 7.53% in recall and 4.25% in precision, and finally F1-score improved by 5.88% compared to the localization mask. We have shown that every step we suggest is beneficial in successfully detecting transmission lines.

**Table 2.** Ablation study for demonstrating the effectiveness of PDA module and refinement network.

Localization Mask	PDA	Refinement	Recall Rate (%)	Precision (%)	F1-Score (%)
✓			90.37	91.90	91.13
✓	✓		92.03	92.67	92.35
✓		✓	93.96	94.86	94.41
✓	✓	✓	97.90	96.15	97.01

#### 4.5. Comparison with Other Attention Modules

In this section, the proposed attention module is compared with previous attention mechanisms including SimAM [30], SE [31], and ECA [32]. Table 3 summarizes the results of comparative experiments. An identical base network of VGG16 was utilized to obtain localization masks based on the previous attention modules. The localization accuracy with the use of PDA is compared with the accuracies based on the previous methods. SE showed less than 90% accuracies in all evaluation metrics, and PDA was higher than SE by 3.15% in terms of precision. In addition, another channel attention module ECA showed better performance than SE, but it was still lower than the accuracy of ours. ECA utilizes only one selective kernel for 1D convolution, and it is limited to properly representing features of transmission lines due to small receptive field. As shown in Table 3, SimAM, which generates 3D weights, was unsuitable for our task, and this experiment showed that not all attention modules are effective in localizing transmission lines. The PDA module showed plausible performance in all evaluation metrics compared to the previous methods. The performance improvement is attributed to the advantage that PDA captures abundant feature representations and broadens the receptive field by utilizing three different 1D convolutions.

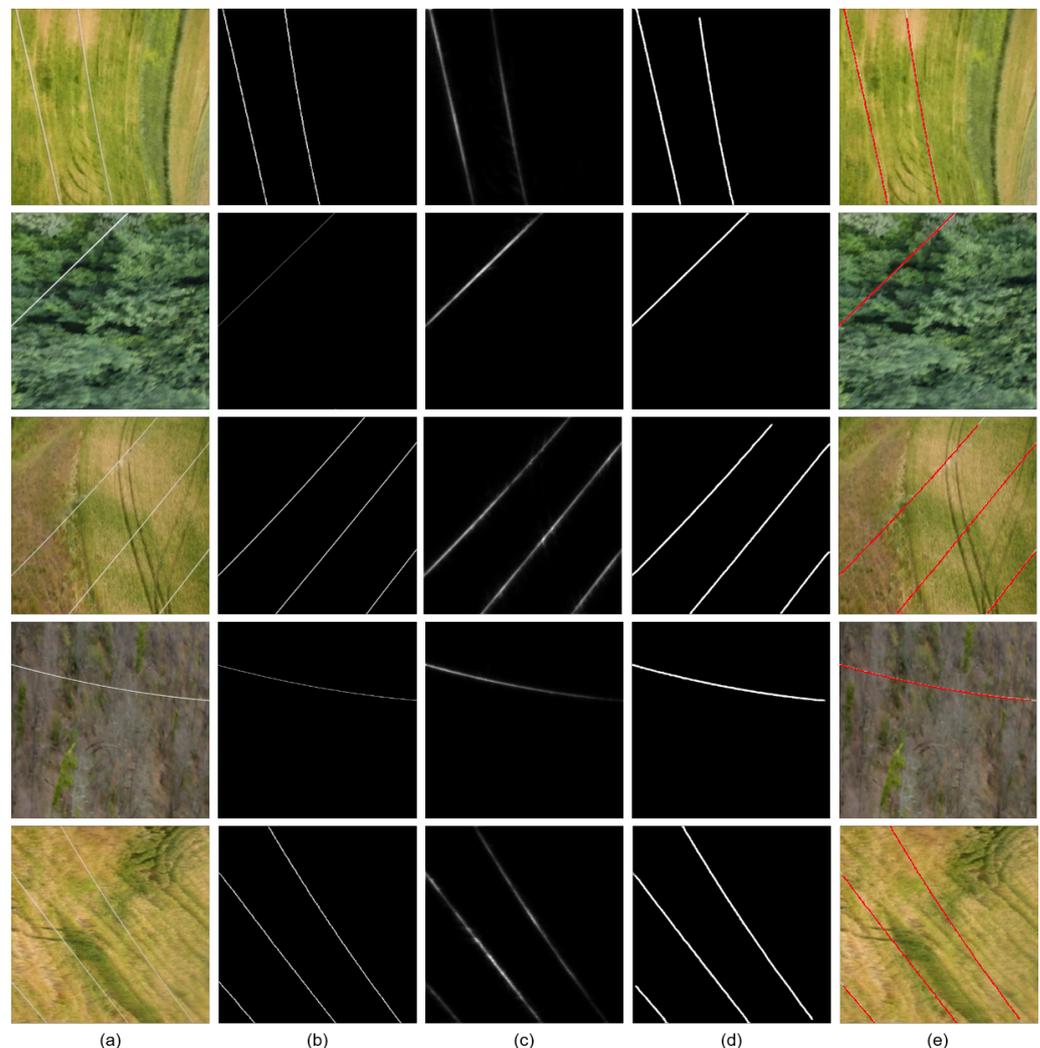
**Table 3.** Comparison with existing attention modules. These results indicate the performance before applying the refinement network.

Methods	Recall Rate (%)	Precision (%)	F1-Score (%)
SimAM [30]	89.88	88.97	89.42
SE [31]	89.89	89.52	89.71
ECA [32]	90.97	92.08	91.52
PDA (Ours)	92.03	92.67	92.35

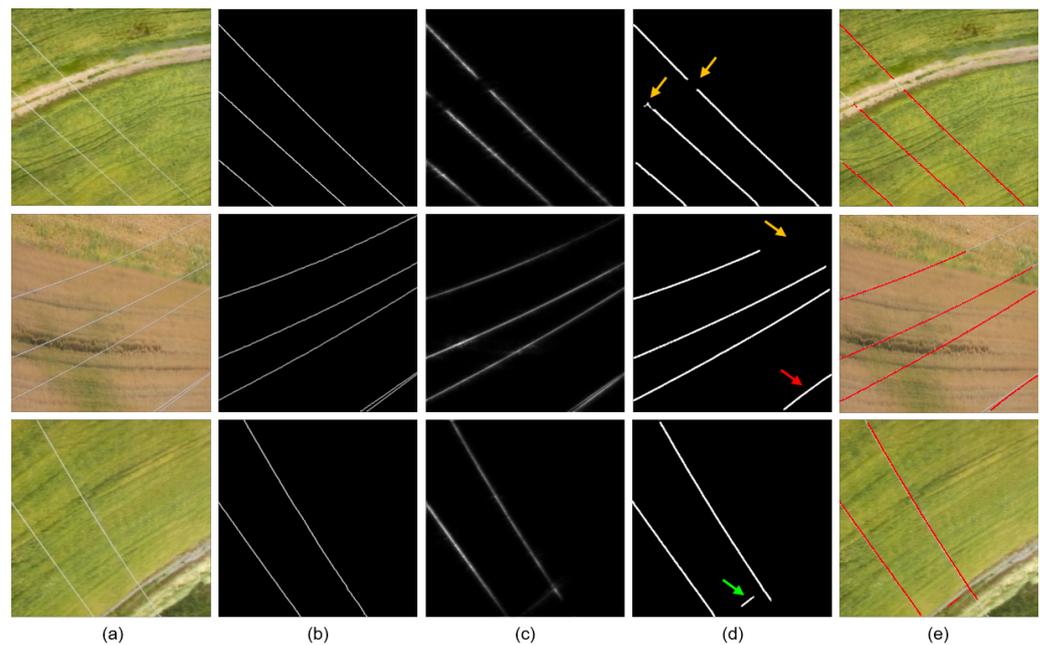
#### 4.6. Qualitative Evaluation

Figure 8 presents result images of the proposed algorithm for detecting transmission lines. Figure 8a,b show input images and the corresponding ground truth data. The ALM is the localization mask with the application of the attention vector obtained from the PDA, for focusing more on the important channels. As shown in Figure 8c, PDA is effective in localizing the lines in the input images, and the predicted structure and number of lines are similar to the ground truth even before applying the refinement network. However, most

transmission lines in the ALMs are blurry, and several predictions contain disconnected or omitted parts. To address these limitations, the refinement network is applied to the ALMs, and the results are presented at Figure 8d. The refinement network generates refined ALMs, which complement smudged or indistinct lines to make the lines sharp and bold, based on the characteristics of the target domain data. The refinement network also connects broken lines to generate intersecting lines of images, providing qualitatively plausible results even if compared with ground truth data. In Figure 8e, original images are overlaid with the refined ALM. Figure 9 represents failure cases of the proposed algorithm. In the refined ALM of the second example, the red arrow indicates a merged prediction of two transmission lines close to each other. However, even though two close transmission lines are recognized as a single line, it does not critically affect UAVs to operate a collision avoidance function. The yellow arrows in the refined ALMs indicate missing parts of transmission lines. When the responses of transmission lines in ALM are too weak, the refinement network could not restore the corresponding parts of transmission lines. The green arrow in the last example in Figure 9d indicates a false positive case, and such a failure is usually isolated and occurs in a local area. Therefore, we expect that these types of failures can be recovered by applying a post-processing based on the properties of transmission lines.



**Figure 8.** Results of transmission line detection using our proposed method. (a) Original image. (b) Ground truth. (c) ALM. (d) Refined ALM. (e) Overlaid image with the refined ALM.



**Figure 9.** Examples of failure cases of the proposed method. The yellow and green arrows indicate missing parts of transmission lines and false positive responses, respectively. The red arrow indicates a false negative case for two transmission lines that are close to each other. (a) Original. (b) Ground truth. (c) ALM. (d) Refined ALM. (e) Overlaid image with the refined ALM.

## 5. Conclusions

In this paper, we propose a transmission line detection algorithm based on weakly supervised learning and image-to-image translation. By only utilizing image-level labels, the proposed algorithm can be trained with the minimal human involvement. The proposed method consists of two steps: (1) localization of transmission lines based on PDA and (2) refinement via image-to-image translation. The PDA module computes a score vector based on the information from various receptive fields. The attention vector provides the channel importance of object features, and it is utilized for generating ALM. Furthermore, we constructed a refinement network that transfers line-shaped properties of transmission lines to improve weak responses and disconnected components in the ALM. We demonstrated that the PDA module outperforms the previous attention methods for localizing transmission lines. Moreover, the refinement network significantly improved the accuracy of transmission line detection in both quantitative and qualitative aspects.

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

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1G1A1009792).

**Data Availability Statement:** Publicly available datasets were analyzed in this study. These data can be found here: <https://data.mendeley.com/datasets/twxp8xccsw/6>, (accessed on 6 June 2022).

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

## References

1. Mohamed, N.; Al-Jaroodi, J.; Jawhar, I.; Idries, A.; Mohammed, F. Unmanned aerial vehicles applications in future smart cities. *Technol. Forecast. Soc. Chang.* **2020**, *153*, 119293. [CrossRef]
2. Radoglou-Grammatikis, P.; Sarigiannidis, P.; Lagkas, T.; Moscholios, I. A compilation of UAV applications for precision agriculture. *Comput. Netw.* **2020**, *172*, 107148. [CrossRef]

3. Shakhathreh, H.; Sawalmeh, A.; Al-Fuqaha, A.; Dou, Z.; Almaita, E.; Khalil, I.; Othman, N.S.; Khreishah, A.; Guizani, M. Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges. *arXiv* **2018**, arXiv:1805.00881.
4. Huang, X.; Dong, X.; Ma, J.; Liu, K.; Ahmed, S.; Lin, J.; Qiu, B. The Improved A\* Obstacle Avoidance Algorithm for the Plant Protection UAV with Millimeter Wave Radar and Monocular Camera Data Fusion. *Remote Sens.* **2021**, *13*, 3364. [[CrossRef](#)]
5. She, X.; Huang, D.; Song, C.; Qin, N.; Zhou, T. Multi-obstacle detection based on monocular vision for UAV. In Proceedings of the 2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA), Chengdu, China, 1–4 August 2021; pp. 1067–1072. [[CrossRef](#)]
6. Pedro, D.; Matos-Carvalho, J.P.; Fonseca, J.M.; Mora, A. Collision avoidance on unmanned aerial vehicles using neural network pipelines and flow clustering techniques. *Remote Sens.* **2021**, *13*, 2643. [[CrossRef](#)]
7. González de Santos, L.M.; Frías Nores, E.; Martínez Sánchez, J.; González Jorge, H. Indoor path-planning algorithm for UAV-based contact inspection. *Sensors* **2021**, *21*, 642. [[CrossRef](#)]
8. Dai, X.; Mao, Y.; Huang, T.; Qin, N.; Huang, D.; Li, Y. Automatic obstacle avoidance of quadrotor UAV via CNN-based learning. *Neurocomputing* **2020**, *402*, 346–358. [[CrossRef](#)]
9. Ou, J.; Guo, X.; Lou, W.; Zhu, M. Quadrotor Autonomous Navigation in Semi-Known Environments Based on Deep Reinforcement Learning. *Remote Sens.* **2021**, *13*, 4330. [[CrossRef](#)]
10. Yuan, S.; Ota, K.; Dong, M.; Zhao, J. A Path Planning Method with Perception Optimization Based on Sky Scanning for UAVs. *Sensors* **2022**, *22*, 891. [[CrossRef](#)]
11. Wang, D.; Li, W.; Liu, X.; Li, N.; Zhang, C. UAV environmental perception and autonomous obstacle avoidance: A deep learning and depth camera combined solution. *Comput. Electron. Agric.* **2020**, *175*, 105523. [[CrossRef](#)]
12. Ge, C.; Wang, J.; Wang, J.; Qi, Q.; Sun, H.; Liao, J. Towards automatic visual inspection: A weakly supervised learning method for industrial applicable object detection. *Comput. Ind.* **2020**, *121*, 103232. [[CrossRef](#)]
13. Huang, B.; Chen, R.; Zhou, Q.; Xu, W. Eye landmarks detection via weakly supervised learning. *Pattern Recognit.* **2020**, *98*, 107076. [[CrossRef](#)]
14. Zhang, F.; Du, B.; Zhang, L.; Xu, M. Weakly supervised learning based on coupled convolutional neural networks for aircraft detection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5553–5563. [[CrossRef](#)]
15. Fu, K.; Lu, W.; Diao, W.; Yan, M.; Sun, H.; Zhang, Y.; Sun, X. WSF-NET: Weakly supervised feature-fusion network for binary segmentation in remote sensing image. *Remote Sens.* **2018**, *10*, 1970. [[CrossRef](#)]
16. Wang, S.; Chen, W.; Xie, S.M.; Azzari, G.; Lobell, D.B. Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sens.* **2020**, *12*, 207. [[CrossRef](#)]
17. Kim, W.S.; Lee, D.H.; Kim, T.; Kim, H.; Sim, T.; Kim, Y.J. Weakly supervised crop area segmentation for an autonomous combine harvester. *Sensors* **2021**, *21*, 4801. [[CrossRef](#)]
18. Wang, P.; Yao, W. Weakly Supervised Pseudo-Label assisted Learning for ALS Point Cloud Semantic Segmentation. *arXiv* **2021**, arXiv:2105.01919.
19. Blaga, B.-C.-Z.; Nedevschi, S. Weakly Supervised Semantic Segmentation Learning on UAV Video Sequences. In Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 23–27 August 2021; pp. 731–735. [[CrossRef](#)]
20. Yang, Z.; Zhao, L.; Wu, S.; Chen, C.Y.C. Lung lesion localization of COVID-19 from chest CT image: A novel weakly supervised learning method. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1864–1872. [[CrossRef](#)]
21. Lee, S.J.; Yun, J.P.; Choi, H.; Kwon, W.; Koo, G.; Kim, S.W. Weakly supervised learning with convolutional neural networks for power line localization. In Proceedings of 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017; pp.1–8. [[CrossRef](#)]
22. Choi, H.; Koo, G.; Kim, B.J.; Kim, S.W. Weakly supervised power line detection algorithm using a recursive noisy label update with refined broken line segments. *Expert Syst. Appl.* **2021**, *165*, 113895. [[CrossRef](#)]
23. Jaffari, R.; Hashmani, M.A.; Reyes-Aldasoro, C.C. A Novel Focal Phi Loss for Power Line Segmentation with Auxiliary Classifier U-Net. *Sensors* **2021**, *21*, 2803. [[CrossRef](#)]
24. Hota, M.; Kumar, U. Power Lines Detection and Segmentation In Multi-Spectral Uav Images Using Convolutional Neural Network. In Proceedings of the 2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS), Ahmedabad, India, 1–4 December 2020; pp.154–157. [[CrossRef](#)]
25. Vemula, S.; Frye, M. Mask R-CNN Powerline Detector: A Deep Learning approach with applications to a UAV. In Proceedings of the 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 11–15 October 2020; pp. 1–6. [[CrossRef](#)]
26. Bojarski, M.; Choromanska, A.; Choromanski, K.; Firner, B.; Jackel, L.; Muller, U.; Zieba, K. VisualBackProp: visualizing CNNs for autonomous driving. *arXiv* **2016**, arXiv:1611.05418.
27. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232. [[CrossRef](#)]
28. Park, J.; Woo, S.; Lee, J.Y.; Kweon, I.S. Bam: Bottleneck attention module. *arXiv* **2018**, arXiv:1807.06514.
29. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 August 2018; pp. 3–19. [[CrossRef](#)]

30. Yang, L.; Zhang, R. Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the 38th International Conference on Machine Learning (PMLR), Virtual, 18–24 July 2021; pp. 11863–11874.
31. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [[CrossRef](#)]
32. Wang, Q.L.; Wu, B.G.; Zhu, P.F.; Li, P.H.; Zuo, W.M.; Hu, Q.H. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11531–11539. [[CrossRef](#)]
33. Aguirre, N.; Grall-Maës, E.; Cymberknop, L.J.; Armentano, R.L. Blood pressure morphology assessment from photoplethysmogram and demographic information using deep learning with attention mechanism. *Sensors* **2021**, *21*, 2167. [[CrossRef](#)] [[PubMed](#)]
34. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2017; pp. 5998–6008.
36. Eom, H.; Lee, D.; Han, S.; Hariyani, Y.S.; Lim, Y.; Sohn, I.; Park, K.; Park, C. End-to-end deep learning architecture for continuous blood pressure estimation using attention mechanism. *Sensors* **2020**, *20*, 2338. [[CrossRef](#)]
37. Eldele, E.; Chen, Z.; Liu, C.; Wu, M.; Kwoh, C.K.; Li, X.; Guan, C. An attention-based deep learning approach for sleep stage classification with single-channel eeg. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 809–818. [[CrossRef](#)]
38. Shi, C.; Zhang, X.; Sun, J.; Wang, L. A Lightweight Convolutional Neural Network Based on Group-Wise Hybrid Attention for Remote Sensing Scene Classification. *Remote Sens.* **2022**, *14*, 161. [[CrossRef](#)]
39. Ma, W.; Zhao, J.; Zhu, H.; Shen, J.; Jiao, L.; Wu, Y.; Hou, B. A spatial-channel collaborative attention network for enhancement of multiresolution classification. *Remote Sens.* **2021**, *13*, 106. [[CrossRef](#)]
40. Hu, J.; Zhi, X.; Shi, T.; Zhang, W.; Cui, Y.; Zhao, S. PAG-YOLO: A portable attention-guided YOLO network for small ship detection. *Remote Sens.* **2021**, *13*, 3059. [[CrossRef](#)]
41. Chen, L.; Shi, W.; Deng, D. Improved YOLOv3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images. *Remote Sens.* **2021**, *13*, 660. [[CrossRef](#)]
42. Seong, S.; Choi, J. Semantic segmentation of urban buildings using a high-resolution network (HRNet) with channel and spatial attention gates. *Remote Sens.* **2021**, *13*, 3087. [[CrossRef](#)]
43. Tran, M.T.; Kim, S.H.; Yang, H.J.; Lee, G.S.; Oh, I.J.; Kang, S.R. Esophagus segmentation in CT images via spatial attention network and STAPLE algorithm. *Sensors* **2021**, *21*, 4556. [[CrossRef](#)] [[PubMed](#)]
44. Kim, M.; Lee, B.D. Automatic lung segmentation on chest X-rays using self-attention deep neural network. *Sensors* **2021**, *21*, 369. [[CrossRef](#)] [[PubMed](#)]
45. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
46. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
47. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134. [[CrossRef](#)]
48. Chen, X.; Duan, Y.; Houthoofd, R.; Schulman, J.; Sutskever, I.; Abbeel, P. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems*; MIT Press: Barcelona, Spain, 2016; pp. 2172–2180.
49. Kim, T.; Cha, M.; Kim, H.; Lee, J.; Kim, J. Learning to discover cross-domain relations with generative adversarial networks. In Proceedings of the International Conference on Machine Learning (PMLR), Sydney, Australia, 6–11 August 2017; pp. 1857–1865.
50. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2868–2876. [[CrossRef](#)]
51. Zi, Y.; Xie, F.; Song, X.; Jiang, Z.; Zhang, H. Thin Cloud Removal for Remote Sensing Images Using a Physical Model-Based CycleGAN with Unpaired Data. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [[CrossRef](#)]
52. Liu, W.; Luo, B.; Liu, J. Synthetic Data Augmentation Using Multiscale Attention CycleGAN for Aircraft Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
53. Mao, J.; Zheng, C.; Yin, J.; Tian, Y.; Cui, W. Wildfire Smoke Classification Based on Synthetic Images and Pixel-and Feature-Level Domain Adaptation. *Sensors* **2021**, *21*, 7785. [[CrossRef](#)]
54. Choi, W.; Heo, J.; Ahn, C. Development of Road Surface Detection Algorithm Using CycleGAN-Augmented Dataset. *Sensors* **2021**, *21*, 7769. [[CrossRef](#)]
55. Nguyen, D.T.; Pham, T.D.; Batchuluun, G.; Noh, K.J.; Park, K.R. Presentation attack face image generation based on a deep generative adversarial network. *Sensors* **2020**, *20*, 1810. [[CrossRef](#)]
56. Sandouka, S.B.; Bazi, Y.; Alajlan, N. Transformers and Generative Adversarial Networks for Liveness Detection in Multitarget Fingerprint Sensors. *Sensors* **2021**, *21*, 699. [[CrossRef](#)]
57. Gao, P.; Tian, T.; Li, L.; Ma, J.; Tian, J. DE-CycleGAN: An object enhancement network for weak vehicle detection in satellite images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3403–3414. [[CrossRef](#)]

58. Noh, K.J.; Choi, J.; Hong, J.S.; Park, K.R. Finger-vein recognition using heterogeneous databases by domain adaption based on a cycle-consistent adversarial network. *Sensors* **2021**, *21*, 524. [[CrossRef](#)] [[PubMed](#)]
59. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [[CrossRef](#)]
60. Li, Y.; Xiao, Z.; Zhen, X.; Cao, X. Attentional information fusion networks for cross-scene power line detection. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1635–1639. [[CrossRef](#)]