applied sciences

MDPI

Article

Dark-Center Based Insulator Detection Method in Foggy Environment

Lisang Liu [1], Chengyang Ke [1,*] and He Lin [2]

[1] School of Electronic Electrical Engineering and Physics, Fujian University of Technology, Fuzhou 350118, China; liulisang@fjut.edu.cn
[2] State Grid Fujian Xiapu County Power Supply Company, Ningde 355100, China; zions07@outlook.com
* Correspondence: 2211905005@smail.fjut.edu.cn

Abstract: In foggy environments, outdoor insulator detection is always with low visibility and unclear targets. Meanwhile, the scale of haze simulation insulator datasets is insufficient. Aiming to solve these problems, this paper proposes a novel Dark-Center algorithm, which is a joint learning framework based on image defogging and target detection. Firstly, the dark channel prior algorithm is used to calculate the foggy sky image transmittance and then transpose it to the original image to generate a foggy-simulated insulator dataset; secondly, the defogging and restoration modules and an optimized defogging module are combined to improve the robustness of the defogging algorithm; then, for small insulator detection, the CenterNet network structure is improved to enhance the feature extraction capability for small targets; finally, the target detection accuracy in foggy environments is improved by jointly learning the structure details and color features recovered in image defogging via the defogging model and the target detection model, which effectively learn the structure details and color features recovered in image defogging. The experimental results on the CPILD dataset show that the proposed Dark-Center algorithm based on image defogging and target detection can effectively improve the performance of the target detector in foggy scenes, with a detection accuracy of 96.76%.

Keywords: image defogging; target detection; dark-center; insulators

check for updates

Citation: Liu, L.; Ke, C.; Lin, H. Dark-Center Based Insulator Detection Method in Foggy Environment. *Appl. Sci.* **2023**, *13*, 7264. https://doi.org/10.3390/app13127264

Academic Editor: Jan Egger

Received: 19 April 2023
Revised: 7 June 2023
Accepted: 13 June 2023
Published: 18 June 2023

Copyright: © 2023 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

Insulators [1–4] are an important part of transmission lines, and when insulators are subjected to temperature variations, their surface compressive stress increases, and they are prone to breakdown, resulting in insulation resistance values of 0 at both ends of the insulator string and loss of insulation function, which causes power supply interruption and triggers power outage accidents. Affected by natural disasters or human factors, insulators may be cracked, broken, and fouled [5–7], and the complex environment also affects image quality. For example, the foggy weather may make the image color grayish, contrast decrease, and unclear object features. Therefore, under the influence of complex transmission lines and foggy weather, clearer image processing and detection have become great challenges [8–11].

For target detection in a foggy environment, the image needs to be defogged first. Generally, defogging algorithms [12] can be divided into image-enhancement-based defogging algorithms [13] and image-restoration-based defogging algorithms [14]. Image-enhancement-based defogging algorithms are represented by Retinex, histogram equalization, and homomorphic filtering, while image-restoration-based defogging algorithms are represented by depth-information-based methods, polarization-characteristic-based methods, and prior-information-based methods. In addition, there are also algorithms for defogging using deep learning, such as DehazeNet [15], AODNet [16], and DCPDN [17], etc. Although the effectiveness of deep-learning-based defogging algorithms is improved

compared with traditional methods, they are time-consuming and often fail to meet the real-time requirements for target detection tasks. Among the above algorithms, the dark channel prior theory [18] based on the atmospheric scattering model is more effective, has higher detection accuracy, and fast running speed. Thus, it is adopted in this paper to achieve the effect of defogging by incorporating the method into the target detection network.

Target detection algorithms can be divided into two categories according to the presence or absence of a preset anchor frame as a benchmark: anchor-free methods and anchor-based methods. Anchor-free algorithms mainly include ConerNet [19], CenterNet [20], FCOS [21], etc., and anchor-based mainly include YoLo series [22], SSD [23], RetinaNet [24], etc. Since anchor-based methods require setting different anchor frames in different datasets, and the setting of anchor frames will have an impact on the accuracy, non-extreme value suppression is also required, resulting in slow running speed [25]. To address the above drawbacks, anchor-free target detection algorithms have been gradually developed. For CenterNet networks, anchor-free algorithms treat a target as a point, and one target is determined using one feature point. Its prediction result will divide the input image into different regions, and each region will have one feature point. Then, it will be determined whether an object exists or not in each feature point, as well as its type and confidence, according to the prediction result. The algorithms will also adjust the feature point to obtain the center coordinates and the width and height of the object.

Currently, target detection methods in foggy environments can be mainly divided into two categories: a two-stage method and an end-to-end approach. The two-stage method is an uncorrelated target detection method based on defogging detection. It defogs the fogged image using image enhancement and recovery methods and then detects the defogged image using target detection methods. The defogging process may bring problems such as artifacts, color distortion, and fog residue to the image, so not all images detected using defogging methods will improve their detection accuracy. The end-to-end approach, on the other hand, jointly optimizes the training of the defogging network and the target detection network to perform the defogging task and the detection task simultaneously [26–28] to solve the above problems. Therefore, this paper adopts an end-to-end approach to detect targets in foggy scenes to improve detection accuracy.

As we all know, it is essential to construct the high-resolution foggy insulator image datasets taken by UAVs, then to locate and detect the "self-exposure" areas of insulators with the help of image-processing and deep-learning technologies to realize intelligent power inspection. As stated above, to address the challenges of foggy environment target detection, this paper proposes a joint learning framework Dark-Center foggy insulator detection and identification algorithm based on CenterNet, which is called Dark-Center detection for short. Its characteristics are outlined as follows:

1. Due to the lack of large-scale foggy insulator datasets, this paper builds foggy image dataset CPILDs using the haze simulation method to construct insulator haze images;
2. To address the problem of poor robustness of most current defogging algorithms, this paper proposes a novel image-defogging algorithm with an optimized defogging and repairing process;
3. The Center algorithm is proposed to enhance the detection capability of small insulators to solve the problem of insufficient feature extraction capability of the detection network;
4. This paper jointly optimizes the training of the defogging algorithm and the target detection algorithm to overcome the problems of artifacts, color distortion, and fog residue.

2. Construction of Foggy Insulator Datasets CPILDs

Since there is no publicly available insulator dataset for foggy environments, this paper needs to construct a foggy insulator target detection dataset first, called CPILDs.

In the non-sky region of most color images, the image pixels are always low in one or more RGB channels, which is called the dark channel prior, as shown in Equation (1).

$$J^{dark}(x) = \min_{y \in \Omega(x)} \left(\min_{c \in \{r,g,b\}} J^c(y) \right) \to 0 \tag{1}$$

where J^c denotes each channel of the color image, J^{dark} denotes the output grayscale image, and $\Omega(x)$ denotes the pixel-centered filter window. The minimum values of the RGB component of each pixel were found and stored in a grayscale map of the same size as the original image. Then, a minimum value filter is applied to the grayscale image.

However, white fog would make the minimum values of its three channels larger. Defogging the image can make the value smaller and the image clearer. Therefore, a foggy image generation model is proposed as follows:

$$I(x) = J(x)t(x) + L(1 - t(x)) \tag{2}$$

where $J(x)$ is the original fog-free image, L is the brightness of the atmospheric light component, and $t(x)$ is the transmittance. The value of luminance L ranges from 0 to 1, indicating the color of the added fog, and the larger the luminance L is, the whiter the fog is. The transmittance $t(x)$ is within the range [0, 1], indicating the ratio of the original image and the fog in the output image. An appropriate transmittance $t(x)$ needs to be set for each pixel in the image as Equation (3).

$$t(x) = \exp\left[-D\left(-0.04\sqrt{(w - w_c)^2 + (h - h_c)^2} + s \right) \right] \tag{3}$$

where (w_c, h_c) is the image center, which is chosen as the fogging center. s is the fogging size, which is the square root of the maximum value of the image width and height. D is the fog thickness coefficient. From Equation (3), it can be seen that the larger D is and the smaller $t(x)$ is, the heavier the fog is and the smaller the light transmission is. Therefore, various fog environments can be simulated by setting different parameters of brightness L and the fog thickness coefficient.

As mentioned above, the CPILD public dataset is constructed by setting the brightness L between [0.7, 1.0] and the random fog thickness coefficient D. The CPILDs dataset contains 8480 fogged images and clear images, with two types of target detection label information. We divided training sets and test sets to 9:1 without additional validation sets. To further ensure randomness, the training and test sets were halved for the experiments, and a total of 4240 images were finally obtained, 3816 in the training set and 424 in the test set. A part of the generated foggy insulator images is shown in Figure 1.

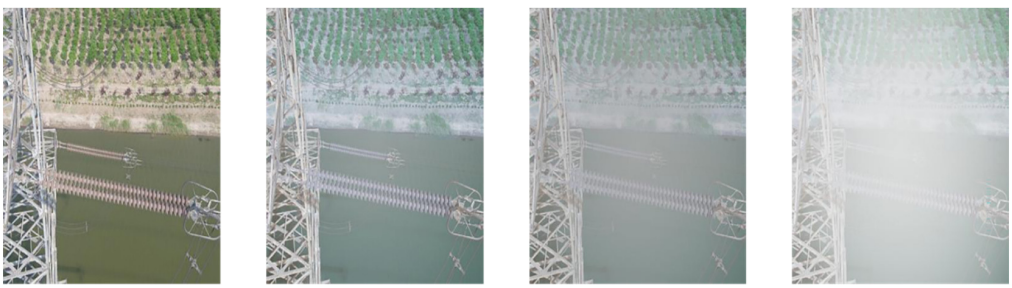


Figure 1. Fog simulation insulator image.

3. Dark-Center Fog Insulator Detection Algorithm

The defogging-detection-based uncorrelated target detection method would bring problems such as artifacts, missing details, and color distortion during image defogging,

which cause the detection accuracy to be lower. In order to improve the accuracy, the Dark-Center fog insulator detection algorithm is proposed in this paper by jointly optimizing the defogging network and the detection network, where the image recovery results are reconstructed, guided by the detection information, while the detection network can learn the structural detail features and color features recovered in image defogging. The specific process is shown in Figure 2.

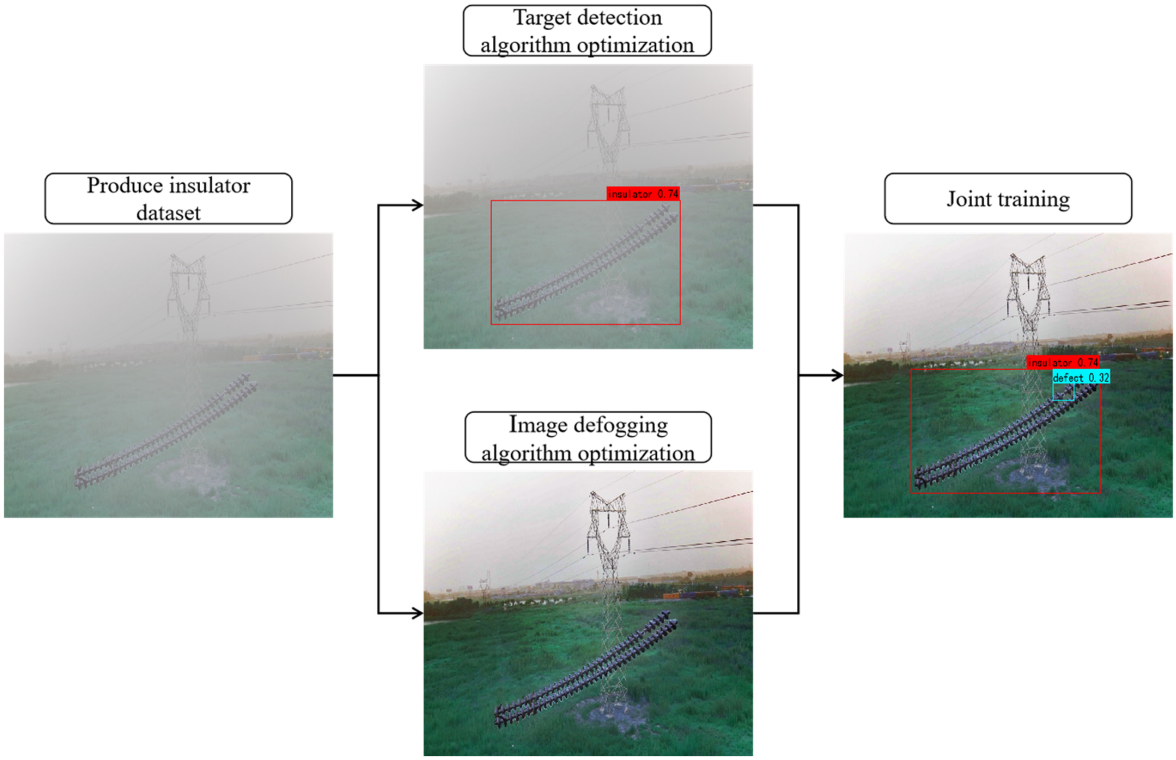


Figure 2. System flow block diagram.

3.1. Defogging Algorithm

The proposed defogging algorithm in this paper consists of two parts: defogging and restoration. Its overall process is shown in Figure 3. First, the fogged images are feature extracted and fused; then, the transmittance map and atmospheric light values are learned uniformly via the deformed atmospheric physical scattering model, and the defogged images are evolved, but the defogged images still have the problems of dark tones and unclear details; finally, the defogged images are repaired via the repair module using the contrast-constrained adaptive histogram equalization method to improve the image contrast and the algorithm's robustness.

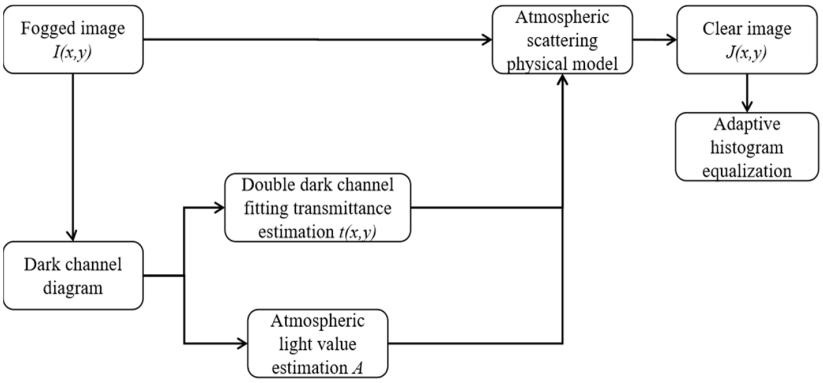


Figure 3. The overall flow of the defogging algorithm.

3.1.1. Optimization of Atmospheric Light Value Estimation

Due to the dark channel, a priori algorithm is not ideal for the sky area, and there are obvious transition areas and overexposure phenomena in the sky and non-sky parts. To solve this problem, this paper adjusts the atmospheric light threshold and designs a new parameter K to modify transmittance calculation.

The first step is to set the atmospheric light threshold. When calculating the atmospheric light, A is the value averaged by taking the top 0.1% of all luminance points, but the value is not proper due to the excessive value of atmospheric light in the sky part. Therefore, we consider setting the maximum global atmospheric light value as the threshold value for all pixel points, which is equal to the threshold value if the value of the pixel point is greater than this threshold value. In this paper, the threshold value is set to 220, which means the pixel points with gray values in the interval [220, 255] are taken in the dark channel map.

The next step is to design the controllable parameter K for recalculating the transmittance. In normal cases, the atmospheric light value of the sky part and the A are relatively close, while the dark channel a priori algorithm is only valid for the area far from the sky. Therefore, we subtract A from the eligible pixel points, and if the result is greater than the threshold, the region is considered to be a sky region, and the transmittance needs to be recalculated; if the result is less than the threshold, the region is not considered to be a sky region, and the transmittance does not need to be recalculated. The threshold value varies by different K, i.e., the transmittance at each pixel point is adjusted by the K. The modified transmittance calculation formula is expressed as follows:

$$J(x) = \frac{I(x) - A}{\min\left(\max\left(\frac{K}{|I(x)-A|}, 1\right) \cdot \max(t(x), t_0), 1\right)} + A \tag{4}$$

where the K is set to 0.22, and the t_0 is set to 0.1 to achieve the optimal effect.

3.1.2. Double Dark Channel Fitting Transmittance Estimation

The transmittance estimation uses guided filtering instead of soft matting to optimize transmittance and suppress contour effects, reducing the high computational complexity, though the efficiency of estimating transmittance is still low and does not solve the distortion problem in principle [29,30].

Improper template selection may distort detailed information such as boundaries. The larger the template, the larger the dark channel converges to 0. To enhance the detailed information in the recovered image and avoid increasing the computational complexity of the transmittance estimation, this paper adopts a double dark channel fitting approach to optimize the transmittance using the image's detail information, avoiding high computational complexity such as soft matting and bootstrap filtering. We selected a field of $\Omega_{35}(x)$ of 35 × 35. A small field of $\Omega_{25}(x)$ is selected for the minimum filtering of the original image as well, and the best experimental result showed its size of 25 × 25. The algorithm flow is outlined as follows.

Step 1: Use the 25 × 25 of $\Omega_{25}(x)$ as the field to obtain one dark channel $J_{25}^{dark}(x)$.

$$J_{25}^{dark}(x) = \min_{y \in \Omega_{25}(x)} \left\{ \min_{c \in \{r,g,b\}} (J^c(x)) \right\} \tag{5}$$

Then, use this $J_{25}^{dark}(x)$ to estimate $t_{25}(x)$.

$$t_{25}(x) = 1 - \omega_{25} \times J_{25}^{dark}(x) \tag{6}$$

Step 2: Using the 35 × 35 of $\Omega_{35}(x)$ as the field, obtain the other dark channel $J_{35}^{dark}(x)$.

$$J_{35}^{dark}(x) = \min_{y \in \Omega_{35}(x)} \left\{ \min_{c \in \{r,g,b\}} (J^c(x)) \right\} \tag{7}$$

Then, use this $J_{35}^{dark}(x)$ to estimate $t_{35}(x)$.

$$t_{35}(x) = 1 - \omega_{35} \times J_{35}^{dark}(x) \tag{8}$$

Step 3: Obtain the final transmittance estimate.

$$t(x) = \alpha t_{25}(x) + (1 - \alpha) t_{35}(x) \tag{9}$$

where α is the adjustment factor, $0 \leq \alpha \leq 1$. In this paper, α is selected as 0.1.

The above method optimizes the transmittance with the image's information, enriches the detail information of the transmittance map, better suppresses the contour distortion of the image, significantly enhances the detail information, and reduces computational complexity.

3.1.3. Adaptive Histogram Equalization

To further improve the robustness of the algorithm, the restoration module uses the defogging feature of image enhancement to repair the defogged images in a targeted manner. Since the side effects of the defogged images are mainly reflected in the dark image brightness, the restoration module needs to enhance the image brightness and contrast to eliminate the side effects of the defogging module as much as possible, while the image enhancement also has a certain defogging effect, which can compensate for the degradation of the defogging effect brought by the compressed multiscale convolutional network [31].

The commonly used image enhancement methods for improving image contrast and brightness are Retinex and HE. Both of them are complex and difficult to adjust parameters, which will seriously affect the defogging speed. The CLAHE algorithm has a faster computing speed compared with Retinex and HE. Therefore, in this paper, the restoration module uses histogram equalization for image processing.

The CLAHE algorithm clips the part of the histogram that exceeds the limit by limiting the local contrast intensity and distributes the clipped part evenly to other regions of the histogram to avoid noise amplification. The algorithm flow is outlined as follows:

1. The image region is divided into M × N non-overlapping sub-regions.
2. The grayscale histogram is calculated for all sub-regions.
3. Given a limit value of T, $H(i)$ is calculated as follows.

$$H(i) = \begin{cases} H(i) + L, H(i) < T \\ H_{\max} \quad , H(i) \geq T \end{cases} \tag{10}$$

where $H_{\max}$, T, and L are related to $H_{\max} = L + T$. The value of L is calculated as follows:

$$L = \frac{N_{\sum shadow}}{N_g} \tag{11}$$

where N_g is the number of gray levels in the subregion, and $N_{\sum shadow}$ is the total number of pixels in the shaded part.

4. Histogram equalization is performed for all subregions.

3.2. CenterNet Structure

CenterNet is an anchor-free target detection method, which has a better detection effect than the target detection network with a prior frame setting and only needs to extract the centroid of each target in the target detection process without post-processing, which improves the detection speed compared with the traditional target detection network. CenterNet uses the residual network (ResNet50) as the backbone feature extraction network, and the residual network structure consists of Conv Block and Identity Block, and upsampling is performed using Transpose Convolution, and its specific network framework is shown in Figure 4.

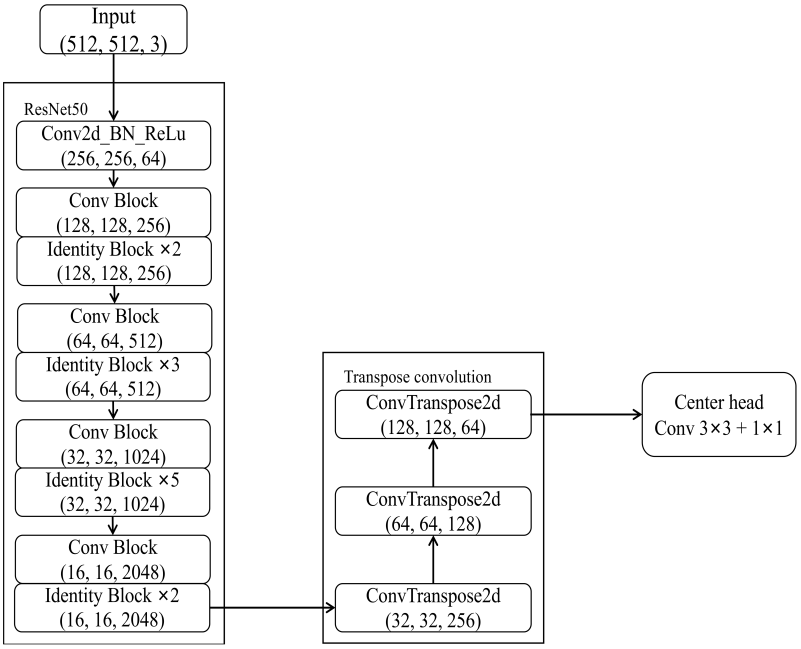


Figure 4. CenterNet network structure.

After the data input, the CenterNet network first performs one convolution, batch normalization, activation function, and maximum pooling; after that, it goes through four groups of residual structures, each group of which consists of Conv Block to change the network dimension and Identity Block to deepen the network, and the number of each group of after the fourth stacking of Conv Blocks and Identity Blocks; the initial extraction of features is completed, and the effective feature layers with dimensions (16, 16, 2048) are obtained; subsequently, the effective feature layers are up-sampled three times using transposed convolution to obtain the high-resolution feature maps, and finally, the results are predicted using the CenterNet detection layer used for result prediction.

3.3. Dark-Center Insulator Detection Algorithm

On the basis of the CenterNet structure mentioned in Section 3.2, this paper optimizes it and proposes a Center structure to improve the detection effect. Firstly, the feature extraction network ResNet50 output layer using the M-SPP model, followed by the addition of the E-ASPP module after transposed convolutional upsampling, and finally, the attention mechanism CA is added to the Center head network to obtain more effective local features and improve the detection accuracy. The Center network framework is shown in Figure 5.

3.3.1. CenterNet Fusion M-SPP

Since insulator sizes vary greatly in realistic shooting scenes, which undoubtedly increases the requirements of the network for extracting target features at different scales, this paper adds the M-SPP module after the backbone network to enhance the perceptual field range and separate important feature information to improve the feature extraction capability.

The proposed M-SPP model uses a spatial pyramid pooling idea based on the SPP principle [32]. The structure of spatial pyramid pooling is shown in Figure 6. To be more versatile in practical scenarios and better obtain insulator feature information with smaller sizes, this paper adopts a residual block instead of a residual edge to improve the small-size insulator detection accuracy. Residual block consists of the convolution, normalization, and activation function, firstly, after convolution, normalization, and activation function, then reduce the operation amount via dw convolution, and finally, 1×1 convolution. In addition, the convolution module is added in front of the pooling layer to further expand the perceptual field range, then the three different pooling modules obtain different feature

information, and the three obtained feature information is spliced and fused with the residual block so that the output feature map obtains both global feature information and different local feature information after pooling. It has been tested that adding a large convolutional layer for the large pooling layer can play a role in enriching the feature information.

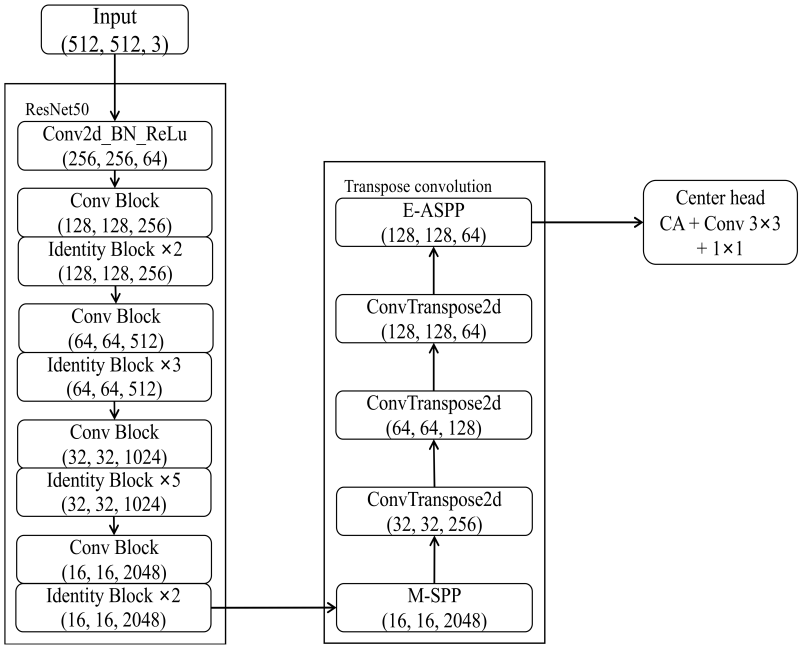


Figure 5. Center network structure.

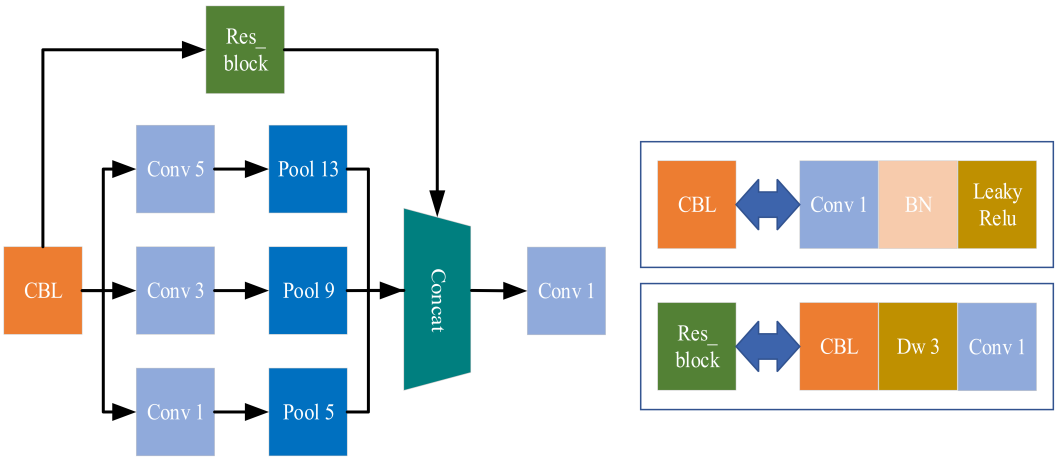


Figure 6. Structure of spatial pyramid pooling.

3.3.2. CenterNet Fusion CA Attention Mechanism

Although the channel attention mechanism SENet has a significant effect on improving the model performance, it usually ignores the location information, which is actually crucial for target detection. Therefore, this paper introduces a lightweight CA attention mechanism, which embeds spatial location information on top of SENet, and the network structure is shown in Figure 7.

This mechanism decomposes channel attention into two one-dimensional feature encoding processes. First, the aggregated features are pooled with a global average of (1 × W) and (H × 1) along the W and H directions, respectively. This captures remote dependencies along one spatial direction while retaining accurate location information along the other spatial direction and encodes the resulting feature maps into a pair of

orientation-aware and location-sensitive features, respectively. Then, the weights of the W-dimensional and H-dimensional features are reassigned using the Sigmoid activation function. Finally, the input is multiplied by the weights of these two dimensions to obtain the output, which is complementarily applied to the input feature maps to enhance the representation of the object of interest.

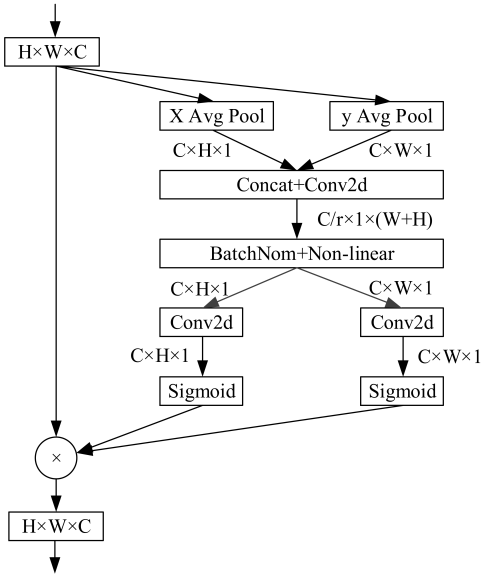


Figure 7. CA attention mechanism.

For the model to predict the target category information while also focusing on the location information, this paper introduces it to the detection module of Center (Center–Head) before 3 × 3 convolution and 1 × 1 convolution to construct the CA–Center–Head detection module; its network structure is shown in Figure 8.

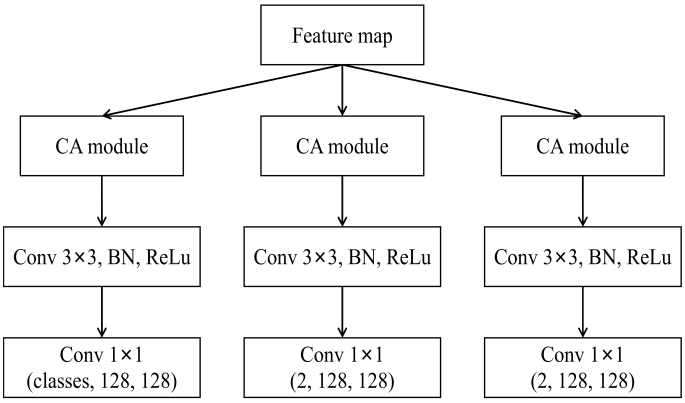


Figure 8. CA–Center–Head module structure.

3.3.3. CenterNet Fusion E-ASPP

The ASPP module in the Center network is shown in Figure 9, within the red box. The ASPP module [33] has unique advantages in extracting multi-scale information from images and is widely used in image processing. The ASPP module consists of multiple convolutional operations in parallel and global average pooling operations. Except for the 1 × 1 convolution, each convolution kernel has a different void rate. Convolution kernels with large void rates facilitate the segmentation of large targets; convolution kernels with small void rates facilitate the segmentation of small targets. ASPPs fail to take full advantage of the perceptual field brought via the dilated convolution; thus, a more efficient feature

pyramid module E-ASPP is designed using the Dense ASPP dense connectivity and the ECA attention mechanism.

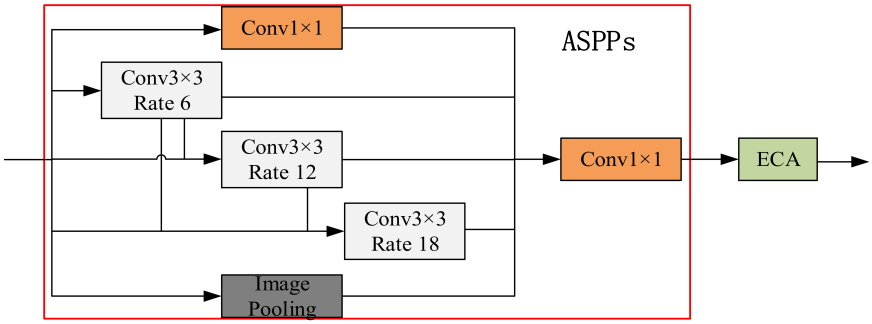


Figure 9. E-ASPP feature pyramid module.

The input of the ASPPs module is a 1/4 size feature map obtained via transposed convolution with 64 channels. In the E-ASPP module, the feature maps are convolved with 1×1 convolution, global average pooling, and null convolution with dilation rates of 6, 12, and 18, respectively. The difference from ASPP is the way of fusion for different dilation rate convolutions. First, the results obtained from the branch with a dilation rate of 6 are added to the branch with a dilation rate of 12 and 18, and second, the integrated 12 branches are added to the 18 branches to obtain five feature maps of a size of 1/4 and several channels of 64. The resulting feature maps are stitched and fused in the channel dimension, and then the ECA attention mechanism is added to obtain the feature information generated using the E-ASPP module.

4. Experimental Results and Analysis

The experimental hardware configuration is AMD Ryzen 5 1600 CPU, 8 GB RAM, NVIDIA GeForce GTX 1650 DDR6 with 4 Gbyte video memory, Ubuntu 18.04, 64-bit software environment, Pytorch 1.2.0 deep-learning framework, and CUDA 11 parallel computing framework.

4.1. Defogging Experiments

Generally, peak signal-to-noise ratio (*PSNR*), structural similarity (*SSIM*), and information entropy (*S*) are adopted to evaluate the performance of the defogging algorithm.

PSNR is evaluated by calculating the mean squared error (*MSE*) of the fogged and unfogged images to assess the model. The smaller the *MSE*, the larger the *PSNR*, representing the greater similarity between images. For an $H \times W$ image, the *MSE* and *PSNR* are obtained using Equation (12) and Equation (13), respectively.

$$MSE = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X(i,j) - Y(i,j))^2 \tag{12}$$

$$PSNR = 10\lg \frac{(2^n - 1)^2}{MSE} \tag{13}$$

SSIM indicates the similarity of the two images and evaluates the similarity between the images in three aspects: brightness, contrast, and structure, respectively. The closer the result is to 1 means the more similar the graphics are and the less distorted the images are. The expressions are expressed as follows:

$$SSIM(X,Y) = l(X,Y) \cdot c(X,Y) \cdot s(X,Y) \tag{14}$$

$$l(X,Y) = \frac{2u_X u_Y + C_1}{u_X^2 + u_Y^2 + C_1} \tag{15}$$

$$c(X,Y) = \frac{2\partial_X \partial_Y + C_2}{\partial_X^2 + \partial_Y^2 + C_2} \tag{16}$$

$$s(X,Y) = \frac{\partial_{XY} + C_3}{\partial_X \partial_Y + C_3} \tag{17}$$

where ∂_X and ∂_Y denote the variance of image X and image Y; u_X and u_Y denote the mean of image X and Y; C_1, C_2, and C_3 are constant terms; ∂_{XY} denotes the covariance of image X and Y.

Calculating the magnitude of the information entropy evaluation value is in finding the total expectation of the image grayscale value. The amount of detailed information in an image can be characterized by the result of the information entropy evaluation. The more information the image contains, the larger the value of information entropy is. Its calculation formula is expressed as follows:

$$S = \sum_{i=0}^{255} \frac{f(i,j)}{(M \times N)^2} \log \frac{f(i,j)}{(M \times N)^2} \tag{18}$$

where I is the gray value of a pixel, j is the average of the grayness of a small region centered on that pixel value, $f(i,j)$ is the number of occurrences of the binary, and M and N denote the length and width of the image, respectively.

To verify the effectiveness of the improved dark channel prior algorithm, this paper compares the effects of the three improvement strategies on the model performance with experimental results derived from the CPILD test set, as shown in Table 1. As seen in the table, each improvement strategy on the model does not make SSIM values change greatly. However, the PSNR is significantly improved over the original model with a value of 16.27 dB after fitting the transmittance estimation using the dual dark channel; meanwhile, after optimizing the atmospheric light value estimation, the PSNR model is improved by 0.36 dB, and the PSNR metric is slightly improved when CLAHE is used as the restoration module. Combining these three improvements, the PSNR improves by 0.55 dB compared to the original model. These experiments show that the improved model has improved both the defogging effect and the image quality.

Table 1. Improved dark channel prior defogging algorithm for ablation experiments.

t(x)	J	clahe	PSNR [dB]	SSIM
			15.80	0.78
✔			16.27	0.80
	✔		16.16	0.78
		✔	15.89	0.78
✔	✔	✔	16.35	0.79

Note: t(x) is the fitted transmittance estimate using the dual dark channel; J is the optimization of the atmospheric light value estimate; clahe is the CLAHE with repair module. ✔ means with.

Since there are several variants of the method based on physical and non-physical models, and most of the image evaluation results are evaluated using the full reference method on a synthetic dataset containing pairs of clear and haze images, the haze images with different concentrations are not necessarily consistent with the results of different method variants. The best defogging method on thin haze images does not have the best defogging effect on thick haze images, and the best method in the full-reference image quality evaluation does not necessarily perform consistently in the no-reference image quality evaluation. Therefore, the five most representative image-defogging methods, namely dark channel prior, automatic color equalization, single-scale Retinex, DehazeNet network, and multi-scale Retinex, were selected for comparison tests, and the experimental results are shown in Figure 10.

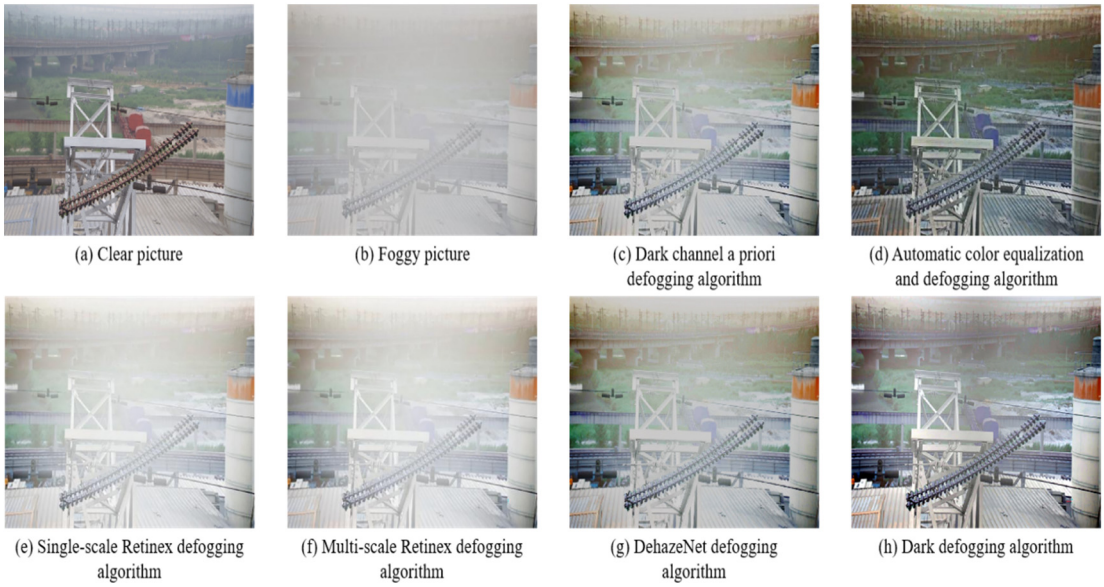


Figure 10. Comparison of the results of different defogging algorithms.

As seen in Figure 10, the dark-channel-based prior method (DCP) can make the image clearer in visual effect after processing, but the image becomes darker overall, as shown in Figure 10c; the automatic-color-balance-based defogging algorithm (ACE) enhances the contrast of the image but produces more noise, which has a greater impact on the subsequent image recognition work, as shown obviously in Figure 10d; while the single-Retinex-model-based scale and multi-scale defogging algorithms (SSR) for foggy images are less effective, as shown in Figure 10e,f; deep-learning-based DehazeNet defogging algorithm does not obtain good results demonstrated for this scene, as shown in Figure 10g; the improved defogging algorithm (Dark) performs better compared with the previous five algorithms, and the image quality has been significantly improved. The image saturation and color are enhanced, too. Table 2 shows the performances of the six defogging algorithms on the CPILD test set.

Table 2. Performance comparison of 6 different defogging algorithms.

Algorithms	PSNR [dB]	SSIM	S [bit]
DCP	15.80	0.78	7.04
ACE	16.11	0.79	7.23
SSR	11.62	0.76	6.41
MSR	12.53	0.77	6.54
DehazeNet	16.23	0.80	7.34
Dark	16.35	0.80	7.55

As seen in Table 2, the proposed defogging algorithm has the best PSNR and SSIM metrics, with PSNR values of 0.55, 0.24, 4.73, 3.82, and 0.12 dB higher relative to the dark channel prior (DCP), automatic color equalization (ACE), single-scale Retinex (SSR), multi-scale Retinex (MSR), and DehazeNet, respectively. As for SSIM values, the proposed algorithm is 0.8, among which is the closest to 1. The best PSNR and SSIM values indicate that the proposed algorithm can process foggy images with the least image distortion, best defogging effect, and highest image quality. For the comparison of information entropy, the SSR algorithm has the lowest recovered information in the test set, which is only 6.41 bits. ACE obtains the next best information entropy result, with a value of 7.23 bits. The proposed algorithm has the best-recovered result, with an information entropy of 7.55 bits, which indicates that the proposed algorithm has a better image-defogging effect than other algorithms, the best image quality, rich information carried, and less image distortion.

4.2. Target Detection Experiments

4.2.1. VOC Dataset

The VOC dataset is one of the most commonly used standard datasets in the field of target detection. Therefore, the VOC2012 dataset is selected to validate the reliability of the model. This dataset has 20 categories with a total of 17125 images, 13870 images in the training set, 1542 images in the validation set, and 1713 images in the test set, where the ratio of the training set to the validation set is 9:1.

4.2.2. Evaluation Indicators

In this paper, the mean average accuracy (*mAP*), the number of frames per second transmitted (*Speed*), and the total number of parameters (*Params*) are used as the evaluation indexes of the algorithm detection accuracy and detection speed and model size, respectively. In the target detection task, *mAP* is the mean value of the average accuracy (*AP*) for each type of target, as shown in Equation (19).

$$mAP = \frac{1}{N}\sum_{\mathrm{i}=1}^{N} AP_{\mathrm{i}} \tag{19}$$

where N is the number of categories, and i denotes a particular category. The AP for a particular category i is calculated as follows:

$$AP_{\mathrm{i}} = \int_0^1 \mathrm{P}(R)dR \tag{20}$$

where $\mathrm{P}(R)$ is the mapping relationship between precision (P) and recall (R), which is often represented by a P–R curve, and the area of the region below the curve is the AP value for that category. Precision and recall are calculated as follows:

$$P_{\mathrm{i}} = TP/(TP + FP) \tag{21}$$

$$R_{\mathrm{i}} = TP/(TP + FN) \tag{22}$$

where TP denotes the number of samples where both the detection category and true label are i; FP denotes the number of samples where the detection category is i, and the true label is not i; and FN denotes the number of samples where the detection category is not i, but the true label is i.

4.2.3. Model Training Strategy

To perform migration learning when training the model, this experiment converts all the input images of the network to 512 × 512 pixels size. At the same time, to speed up the training and prevent the initial training weights from being destroyed, the training was subjected to a freezing process for the first 50 iterations, during which every eight images in a Bach size would be frozen and then thawed after a 50-epoch training is completed. Then, after thawing every four images to a Bach size, the weights were updated once for each completed epoch and were then saved. The initialized learning rate is set as 0.0005, the adaptive optimizer Adam is used, with momentum as 0.9 and weights decay as 0, and the learning rate decreasing method chooses cos.

4.2.4. Experimental Results

To verify the superiority of the method in this paper, five representative and well-performing feature pyramid methods were selected for comparison in the experiments, including SPP (He, 2015), ASPP (Chen, 2017), PPM (Zhao, 2017), RFB and RFB-s (Liu, 2018), and Strip Pooling (Hou, 2020). The number of input and output channels was controlled to be 2048 in the experiments, and both input channels were first transformed to 128 via a 1 ×

1 convolution to ensure that the number of parameters of each model was kept to (33, 35). The experimental results are shown in Table 3.

Table 3. Performance comparison of different methods on the VOC dataset.

SPP	ASPP	PPM	RFB	RFB-s	Strip Pooling	M-SPP	Params [MB]	Speed [FPS]	mAP [%]
✔							33.98	30.86	81.35
	✔						34.72	30.55	81.71
		✔					33.18	31.06	81.24
			✔				34.45	30.69	81.54
				✔			33.52	30.60	81.65
					✔		33.22	31.13	81.75
						✔	34.95	28.49	82.28

Note: ✔ means with.

As seen in Table 3, by comparing the existing SPP-like models, the M-SPP model achieves the best detection accuracy with an mAP value of 82.28% when the number of parameters and the detection speed are relatively similar. This is sufficient to show that it can obtain local feature information of different sizes as well as global information, expand the sensory field, enrich the network expression, and effectively improve the network accuracy.

The feature extraction network with a cavity convolution operation can solve the problem of feature resolution reduction due to the increase in the perceptual field. Different expansion rates in the ASPP structure can process the input feature maps in parallel and extract multi-scale target information. However, since the expansion rates (6, 12, and 18) do not fully utilize the perceptual field to extract effective information, the E-ASPP module is proposed in this paper to enhance the model effect, and the results are shown in Table 4. Compared with the original ASPP, the ASPPs can improve the detection accuracy up to 82.06%, which means it can perform feature extraction more effectively. Moreover, with the subsequent addition of the ECA attention mechanism to re-integrate the feature information, the mAP of the proposed E-ASPP was further improved to 82.21%, which is 0.15% higher in comparison to ASPPs.

Table 4. Characteristic pyramid E-ASPP.

ASPP	ASPPs	ECA	Params [MB]	Speed [FPS]	mAP [%]
✔			32.95	26.22	81.78
	✔		33.46	21.06	82.06
	✔	✔	33.46	20.99	82.21

Note: ✔ means with.

For comparing the effects of the three improvement strategies, M-SPP, E-ASPP, and CA, on the Center model, experiments were carried out, and the results are shown in Table 5. The mAP value of the original model was 79.64%. After taking the above strategies, the mAP values were all improved compared to the original model. When combining the three improvement strategies, the mAP value is improved to 83.42%, which is 3.78% higher than the original model.

Table 5. Center ablation experiment.

M-SPP	E-ASPP	CA	Params [MB]	Speed [FPS]	mAP [%]
			32.67	31.40	79.64
✔			34.95	28.49	82.28
	✔		33.46	20.99	82.21
		✔	32.67	29.75	81.27
✔	✔	✔	35.75	18.88	83.42

Note: ✔ means with.

To verify the detection effect of the Center model, YoLov5, YoLov7, Faster R-CNN, SSD, and CenterNet models are adopted for comparison, and the models have been converged when the same training parameters are used. The mAP values were calculated for all categories in the VOC2012 dataset after detection, and the results are shown in Table 6.

Table 6. Comparison of target detection results of different algorithms.

Model	Network	Params [MB]	Speed [FPS]	mAP [%]
YoLov5	CSPDarknet53	46.73	26.72	80.49
YoLov7	----	37.30	28.95	83.24
Faster R-CNN	ResNet50	28.47	10.50	79.31
SSD	VGG16	26.15	46.89	78.59
CenterNet	ResNet50	32.67	31.40	79.64
Center	ResNet50	35.75	18.88	83.42

Note: ---- indicates no specific name for the YoLov7 backbone feature extraction network.

For the YoLo family of algorithms, the Center model outperforms YoLov5 and YoLov7 in terms of mAP values, Speed, and Params. The faster R-CNN has the slowest detection speed of 10.50 FPS due to the shortcomings of two-stage detection. The SSD, in contrast, has the fastest detection speed but has the lowest detection accuracy 78.59%.

4.3. Dark-Center Target Detection

From Section 3.1, it can be seen that the images after the defogging operation have more prominent insulator contours, richer information, and higher recognizability as well as contrast. Therefore, the defogging algorithm is jointly trained with the target detection algorithm, which we called Dark-Center, to achieve insulator detection in foggy environments. The comparison experiments were carried out on the CPILD dataset, and the experimental results are shown in Table 7.

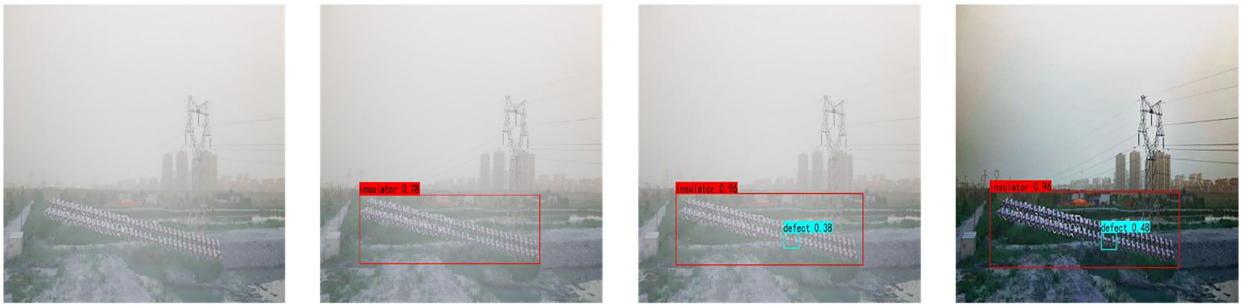
Table 7. Foggy day environment experiment comparison.

Model	Insulator AP [%]	Defect AP [%]	mAP [%]
CenterNet	95.89	59.64	77.77
Center	98.45	90.21	94.33
Dark-Center	98.85	94.67	96.76

For the CenterNet target detection algorithm, the accuracy of detecting insulators is only 77.77%, but the improved Center model has a substantial improvement for the recognition of insulator defects, with an increase of 16.56% in mAP value and up to 94.33%. The addition of the defogging algorithm in the preprocessing module further improves the mAP value to 96.76%. Moreover, the Dark-Center model can also improve the detection accuracy of both insulators and defects, with an increase in Insulator AP from 95.89% to 98.85% and that of Defect AP from 59.64% to 94.67%.

To visualize the differences between the original CenterNet algorithm, the proposed Center model, and the Dark-Center model with the defogging process, a detection image is selected for comparison and analysis, and the detection results are shown in Figure 11.

As seen in Figure 11, the CenterNet detection algorithm only detects the overall insulator orientation at 0.78 and fails to accurately identify the defective part (Figure 11b), while the improved Center algorithm not only detects the insulator with an accuracy of 0.96 but also accurately determines the defects at 0.38 (Figure 11c). For the Dark-Center model with the additional defogging algorithm, its defogging effect is obvious, the insulator identification accuracy is up to 0.96, and the defect identification is 0.48 (Figure 11d).



(a) Original image (b) CenterNet target detection algorithm (c) Center target detection algorithm (d) Dark-Center target detection algorithm

Figure 11. Performance comparison of different algorithms in a foggy environment.

5. Conclusions

In this paper, a joint learning framework that combines image defogging with target detection, Dark-Center, is designed based on the DCP defogging algorithm and CenterNet target detection algorithm. First, a foggy simulation insulator dataset CPILDs is constructed using the dark channel of a prior algorithm. Then, the Dark algorithm is proposed to improve the robustness of the defogging algorithm by combining defogging and restoration modules. Meanwhile, the Center model is proposed to improve the feature extraction capability of the model for small targets on the basis of the CenterNet fused with M-SPP and E-ASPP modules. The target detection accuracy in foggy scenes is greatly improved via the joint learning of the defogging algorithm (Dark) and the target detection model (Center). The experimental results showed that the Dark defogging algorithm can not only remove most of the fog in the image but also retain the original image detail information and improve the image quality. The Center target detection model incorporating the M-SPP and E-ASPP modules can achieve the target detection of insulators in the foggy environment. The insulator detection accuracy reached 98.45%, and the defect detection accuracy reached 90.21% on the CPILD dataset. Moreover, the joint learning framework Dark-Center based on image defogging and target detection further improved the insulator detection accuracy to 98.85% and the defect detection accuracy to 94.67%. The experimental results confirmed the feasibility of Dark-Center-based insulator detection for foggy weather and the generalization of the model. It also has excellent detection accuracy and real-time response speed even in small target detection tasks. Considering the current model is only capable of dealing with foggy environments, the authors' future work will focus on intelligent algorithms that can learn insulator features in a wide range of complex environments for image recovery.

Author Contributions: Conceptualization, C.K., L.L., and H.L.; methodology, C.K.; software, C.K.; validation, C.K., L.L., and H.L.; formal analysis, C.K.; investigation, C.K.; resources, C.K.; data curation, C.K.; writing—original draft, C.K.; writing—review and editing, C.K. and H.L.; visualization, C.K.; supervision, C.K.; project administration, C.K.; funding acquisition, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the University–Industry Cooperation Project "Research and Application of Intelligent Traveling Technology for Steel Logistics Based on Industrial Internet", grant number 2022H6005, Natural Science Foundation of Fujian Provincial Science and Technology Department, grant number 2022J01952, and Research Start-up Projects, grant number GY-Z12079.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding authors.

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

References

1. Mahmood, A.; Beltramelli, L.; Abedin, S.F.; Zeb, S.; Mowla, N.; Hassan, S.A.; Sisinni, E.; Gidlund, M. Industrial IoT in 5G-and-beyond networks: Vsion, architecture, and design trends. *IEEE Trans. Ind. Inform.* **2022**, *18*, 4122–4137. [CrossRef]
2. Zhang, H.H.; Zou, J.; Tian, B.; Wu, M.R. Analysis on insulator flashover of overhead transmission and distribution line under electromagnetic pulse excitation of high altitude nuclear explosion. *J. Electr. Eng. Technol.* **2020**, *35*, 435–443.
3. Cheng, L.; Zhang, H.Y.; Yi, T.Q.; Xie, Y.Z.; Feng, S.M.; Liu, X. Isolation switch insulation defect detection method based on 3D electric field time-frequency analysis. *High Volt. Technol.* **2020**, *46*, 1417–1423.
4. Asprou, M.; Kyriakides, E.; Albu, M.M. Uncertainty bounds of transmission line parameters estimated from synchronized measurements. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 2808–2818. [CrossRef]
5. Liu, Z.Q.; Wang, H.F. Automatic detection of transformer components in inspection images based on improved Faster R-CNN. *Energies* **2018**, *11*, 3496. [CrossRef]
6. Lu, S.Y.; Zhang, Y.; Li, J.X.; Mu, S.Y. Application of mobile robot in high voltage substation. *High Volt. Technol.* **2017**, *43*, 276–284.
7. Xia, H.; Yang, B.; Li, Y.; Wang, B. An improved CenterNet model for insulator defect detection using aerial imagery. *Sensors* **2022**, *22*, 2850. [CrossRef]
8. Wang, M.; Du, Y.; Zhang, Z.R. Research on UAV aided patrol and insulator defect image recognition. *J. Electron. Meas. Instrum.* **2015**, *29*, 1862–1869.
9. Shan, C.; Wu, H.T.; Shi, C.L.; Chen, Y.Y. Method of insulator defect detection in image processing. *J. China Acad. Metrol.* **2010**, *21*, 297–304.
10. Tang, B.; Qin, Q.; Huang, L. Transmission line insulator string aerial image recognition based on color model and texture feature. *J. Electr. Power Sci. Technol.* **2020**, *35*, 13–19.
11. Zhao, Z.B.; Xu, G.Z.; Qi, Y.C.; Liu, N.; Zhang, T.F. Multi-patch deep features for power line insulator status classification from aerial images. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, Canada, 24–29 July 2016; pp. 3187–3194.
12. Guo, F.; Cai, Z.X.; Xie, B.; Tang, J. Review and prospect of image defogging technology. *Comput. Appl. Softw.* **2010**, *30*, 2417–2421.
13. Zhu, P.; Zhu, H.; Qian, X.M.; Li, H. A sharpening method for scene images with foggy weather images. *Chin. J. Graph.* **2004**, *1*, 126–130.
14. Stockham, T.G. Image processing in the context of a visual model. *Proc. IEEE* **1972**, *60*, 828–842. [CrossRef]
15. Cai, B.; Xu, X.; Jia, K.; Qing, C.; Tao, D. DehazeNet: An end-to-end system for single image haze removal. *IEEE Trans. Image Process.* **2016**, *25*, 5187–5198. [CrossRef] [PubMed]
16. Li, B.Y.; Peng, X.L.; Wang, Z.Y.; Xu, J.Z.; Feng, D. AOD-Net: All-in-one dehazing network. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4780–4788.
17. Zhang, H.; Patel, V.M. Densely connected pyramid dehazing network. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
18. He, K.M.; Sun, J.; Tang, X.O. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353. [PubMed]
19. Law, H.; Deng, J. CornerNet: Detecting objects as paired keypoints. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
20. Zhou, X.Y.; Wang, D.Q.; Krahenbuhl, P. Objects as points. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–19 June 2019.
21. Tian, Z.; Shen, C.H.; Chen, H.; He, T. FCOS: Fully convolutional one-stage object detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
22. Rahman, E.; Zhang, Y.; Ahmad, S.; Ahmad, H.; Jobaer, S. Autonomous vision-based primary distribution systems porcelain insulators inspection using UAVs. *Sensors* **2021**, *21*, 974. [CrossRef]
23. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016.
24. Lin, T.; Goyal, P.; Girshick, R.; He, K.M.; Dollar, P. Focal loss for dense object detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
25. Ge, Z.; Liu, S.T.; Wang, F.; Li, Z.M.; Sun, J. YOLOX: Exceeding YOLO series in 2021. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Online, 15–19 June 2021.
26. Zhou, M.F.; Wang, X.L. Weak supervised deep neural network target detection model in remote sensing images. *Chin. Sci. Inf. Sci.* **2018**, *48*, 1022–1034.
27. Wang, Y.D.; Guo, J.C.; Wang, T.B. An improved pedestrian and vehicle detection algorithm for foggy images. *J. Xi'an Univ. Electron. Sci. Technol.* **2020**, *47*, 70–77.
28. Li, W.B.; He, R. Aircraft target detection in remote sensing images based on deep neural network. *Comput. Eng.* **2020**, *46*, 268–276.
29. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
30. Ji, X.Q.; Dai, M.; Sun, L.N.; Lang, X.L.; Wang, H. Research on dark channel prior image defogging algorithm. *Photonics-Laser* **2011**, *22*, 926–929.

31. Zhang, Z.H.; Zhou, W.X. Image defogging algorithm based on full convolutional regression network. *Adv. Lasers Optoelectron.* **2019**, *56*, 252–261.
32. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
33. Chen, L.; Zhu, Y.K.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.

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