An Improved Algorithm for Insulator and Defect Detection Based on YOLOv4

Gujing Han [1,2,*], Qiwei Yuan [1,2], Feng Zhao [3], Ruijie Wang [1,2], Liu Zhao [1,2], Saidian Li [1,2], Min He [4], Shiqi Yang [4] and Liang Qin [4]

1   School of Electronic and Electrical Engineering, Wuhan Textile University, Wuhan 430200, China
2   State Key Laboratory of New Textile Materials and Advanced Processing Technologies,
    Wuhan Textile University, Wuhan 430200, China
3   State Grid Information & Telecommunication Group Co., Ltd., Beijing 102211, China
4   School of Electrical and Automation, Wuhan University, Wuhan 430072, China
*   Correspondence: gjhan@wtu.edu.cn

Abstract: To further improve the accuracy and speed of UAV inspection of transmission line insulator defects, this paper proposes an insulator detection and defect identification algorithm based on YOLOv4, which is called DSMH-YOLOv4. In the feature extraction network of the YOLOv4 model, the improved algorithm improves the residual edges of the residual structure based on feature reuse and designs the backbone network D-CSPDarknet53, which greatly reduces the number of parameters and computation of the model. The SA-Net (Shuffle Attention Neural Networks) attention model is embedded in the feature fusion network to strengthen the attention of target features and improve the weight of the target. Multi-head output is added to the output layer to improve the ability of the model to recognize the small target of insulator damage. The experimental results show that the number of parameters of the improved algorithm model is only 25.98% of that of the original model, and the mAP (mean Average Precision) of the insulator and defect is increased from 92.44% to 96.14%, which provides an effective way for the implementation of edge end algorithm deployment.

Keywords: UAV inspection; insulator defect; DSMH-YOLOv4; feature reuse; SA-Net; multi-head

1. Introduction

As an important part of the overhead transmission line, insulators assume the role of mechanical support and electrical insulation of the line [1,2]. In addition, insulators work under high voltage and high load for a long time and are often eroded by all kinds of bad weather. Insulators are thus prone to damage, which seriously threatens the safety and stability of transmission lines [3,4]. Therefore, the detection of insulators and defects on high-voltage transmission lines has become an important task in power inspection. In recent years, due to its rapid rise, drone inspection technology has begun to gradually replace manual inspections [5,6]. To achieve real-time high-precision detection of insulators and their defects by UAVs [7], the ability to deploy detection algorithms on UAVs at the edge is a necessary prerequisite [8,9]. At the same time, it is necessary to further improve the recognition speed and accuracy of the algorithm for small targets such as insulator defects in complex backgrounds, to improve the algorithm performance and increase the detection efficiency [10,11].

With the continuous development of deep learning technology [12], methods related to the detection of insulators and their defects have been proposed one after another [13–15]. At present, the commonly used deep learning algorithms are mainly divided into two categories: the first category comprises the regression-based one-stage algorithms, such as SSD (Single Shot MultiBox Detector), YOLO (You Only Look Once), YOLOv2, YOLOv3, and YOLOv4 [16–20]. The second category comprises the two-stage algorithm based on region candidates, such as R-CNN (Region-Convolutional Neural Network), Fast R-CNN,

Faster R- CNN, and Mask R-CNN [21–24]. The one-stage algorithm directly zones the input image without generating candidate frames, providing faster detection than the two-stage algorithm, but with slightly lower detection accuracy. However, after several generations of development, the YOLO family of algorithms has the potential to combine accuracy and speed and is one of the main algorithms in current research into target detection applications.

YOLOv3, a classic algorithm in the first stage, is also quite good at recognizing small targets while maintaining its speed advantage. Jia et al. [25] proposed a YOLOv3 improvement algorithm for insulator detection, which can speed up the computation by improving the residual structure in the backbone feature extraction network to a depth-separable convolution; however, the accuracy is low. Liu et al. [26] replaced the YOLOv3 backbone network with DenseNet, which effectively improved the feature extraction capability of the images, but still used the original FPNs (Feature Pyramid Networks) network for the detection of broken small targets, which was not sufficient for the recognition of small targets. Zhu et al. [27] reduced the residual blocks in the YOLOv3 backbone network, which reduced the network depth and the number of operations, but lost some of the feature information, leading to a reduction in accuracy. Yang et al. [28] replaced the backbone feature extraction network of YOLOv3 with a lightweight Mobilenet network, which sped up the computation. However, the extraction of features by downsampling does not increase the number of channels sufficiently, the higher-level semantic information is lost, and the accuracy is low. YOLOv4 is a successor to YOLOV3 in terms of speed and accuracy, making it a valuable tool in the detection of insulators and defects. Yang et al. [29] incorporated an attention mechanism into the YOLOv4 feature extraction network, which enabled the model to better capture valid information, but with lower accuracy for target detection in complex contexts. He et al. [30] introduced a feature fusion structure in the YOLOv4 model to map shallow information to a feature pyramid and fuse deeper semantic information, but the increased number of model parameters and slower detection speeds led to poorer detection in UAV-embedded devices. Gao et al. [31] used migration learning and super-resolution generation networks based on YOLOv4. The overall performance of the model was improved and the network detection speed was accelerated, but the detection accuracy of small targets for insulator defects was insufficient. Han et al. [32] introduced the Self-Attention Mechanism and Channel Attention Mechanism modules in Tiny-YOLOv4, which greatly reduced the complexity of the model and allowed portability to embedded platforms; however, the model was slightly less accurate. Although the above algorithms can better detect insulators, it is difficult to meet the requirements of real-time recognition for small targets with broken insulators in complex contexts and these targets cannot be detected in real time during UAV inspections. Most techniques increase the size of the model while improving accuracy, making it difficult to perform embedded applications on UAVs.

Based on the above analysis, this paper proposes an improved algorithm for insulator and defect detection based on YOLOv4, called DSMH-YOLOv4, which further improves the recognition accuracy of small targets with broken insulators in complex backgrounds while increasing the insulator detection speed. The main improvements are as follows:

1. Replacing the original backbone feature extraction network with a lightweight backbone network, D-CSPDarknet53, through feature reuse, which greatly reduced the number of parameters and computation of the model.
2. Embedding the SA-Net attention module [33] between the backbone network and the feature fusion layer to improve the focus capability of the model for the complex background where the detection target is located.
3. Adding multiple outputs to the prediction module to improve the detection accuracy of the model for the small target of insulator defects.

2. YOLOv4 Basic Structure

The YOLOv4 model consists of a feature extraction module, a feature enhancement module, and a detection module, as shown in Figure 1. In the feature extraction module, the input image is first pre-processed to change the image size to $416 \times 416 \times 3$. The pre-processed images are passed through the CSPDarknet53 backbone network and three multi-scale feature maps containing different dimensional information are output: $13 \times 13 \times 1024$, $26 \times 26 \times 512$, and $52 \times 52 \times 256$. In the feature enhancement module, three different scales of feature maps need to be sampled in both directions and stacked. Then, these feature maps are fully fused to obtain an optimized feature layer with more generalizability. The detection module uses the resulting optimized feature layer to identify and localize targets at different scales and optimizes the output of the final prediction through non-maximum suppression.
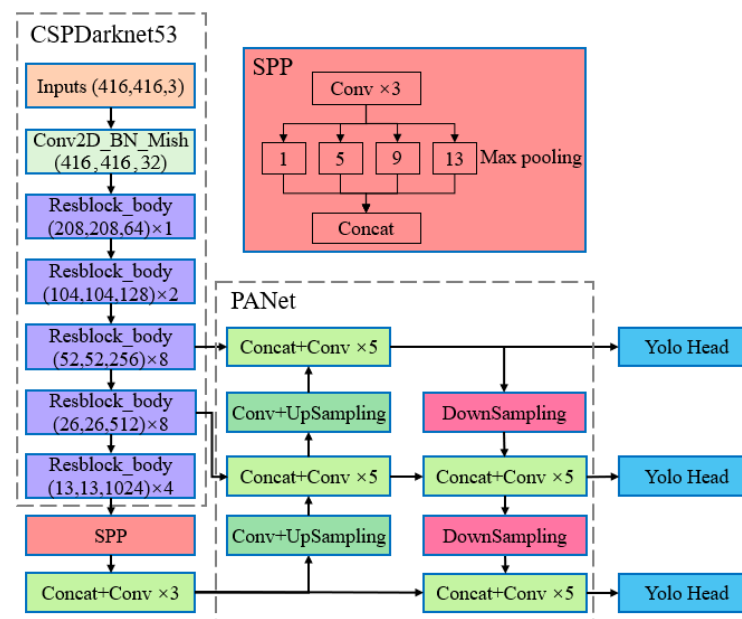

Figure 1. YOLOv4 structure diagram.

Figure 2 shows the structure of the residual module, The upper layer feature information Input enters the residual module and is directly extracted by the convolution units of the two paths to obtain the feature matrices X and Y. X will continue to enter the residual unit. After extracting features by two convolutions, X stacks itself to form the residual structure. In the figure, n indicates the number of stacks of residual units, and the function of the feature extraction module is mainly achieved by five residual modules, Resblock_body (n = 1, 2, 8, 8, 4). Y is transmitted to the bottom of the module via a residual path with a large span. It is then stacked and convolved with the residual unit output feature Z to output the extracted feature Output. This operation allows for maximum retention of the original information. As shown in Figure 1, the SPP (Spatial Pyramid Pooling) network after the 5 residual modules then processes the final output feature layer having a size of $13 \times 13 \times 1024$. This feature layer is first convolved to further extract features. Then, it is maximally pooled through 1×1, 5×5, 9×9, and 13×13 maximum pooling kernels. Finally, the results of each pooling are stacked and output.

The acquisition of deep image features through the series of residual structures described above leads to a large number of parameters in the backbone network, which adds significantly to the complexity of the model. Thus, lightweight improvements are necessary to suit the requirements of edge-end deployments, while maintaining the accuracy of the model detection.
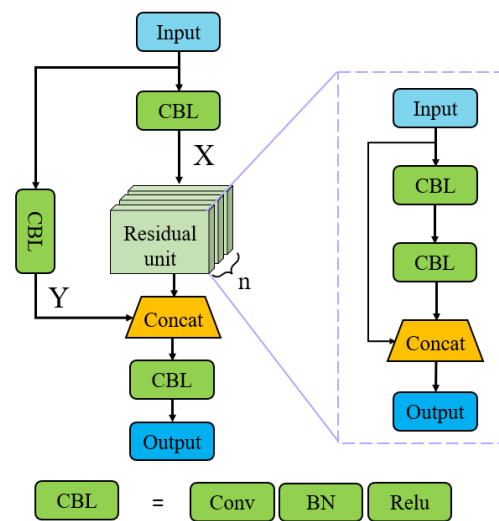
Figure 2. Residual module structure diagram. Conv: Convolution with kernel size 3×3, Strides = 2; BN: Batch Normalization; Relu: Rectified Linear Unit.

3. Improved YOLOV4 Model Network Structure

3.1. Subsection Feature Extraction Module Lightweighting Improvement

Convolutional neural networks can lead to degradation problems if the performance of the network is enhanced only by increasing the depth of the network. That is, the gradient disappears as the network deepens to a certain point, resulting in a loss of accuracy. However, residual neural networks can deepen the information fusion between different feature layers by establishing short-circuit links [34]. This allows a deeper model to be built that still has better recognition results.

The DenseNet network [35] is a modified residual structure. It is possible to create a tight fully connected structure between the upper and lower feature layers, thus effectively linking the features between the different feature layers. It can also be used to reduce the number of model parameters and computational effort by communicating through feature reuse in the process of establishing tight connections. The structure of its feature reuse DenseBlock is shown in Figure 3. The input feature T_0 outputs T_1 through the transfer function H_0; at the same time, T_0 is transmitted directly from the branch. T_0 and T_1 constitute the input of the transfer function H_1.
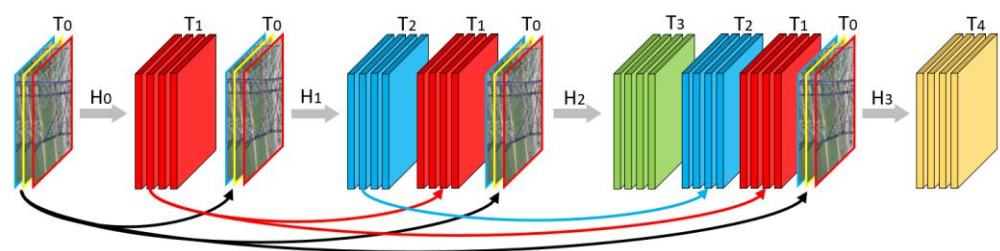

Figure 3. DenseBlock structure diagram.

The calculation process is shown in Equation (1):

$$T_L = H_{L-1}(T_0, T_1, \cdots, T_{L-1}) \tag{1}$$

In Equation (1), T_L denotes the input at layer L.

Compared with the original residual structure, DenseNet's residual edges directly connect the upper and lower features, using the original upper features combined with the lower features. This reduces the process of calculating the residual edges in the original residual structure and reduces the model volume. It reduces computation time by directly

using upper-level feature parameters and also allows for higher utilization of shallow-level information. It also prevents the problem of local information loss as the network deepens, speeds up convergence, and improves detection efficiency.

Therefore, this paper uses the idea of DenseNet using D-CSPDraknet53 as the backbone network of the model, as shown in Figure 4.
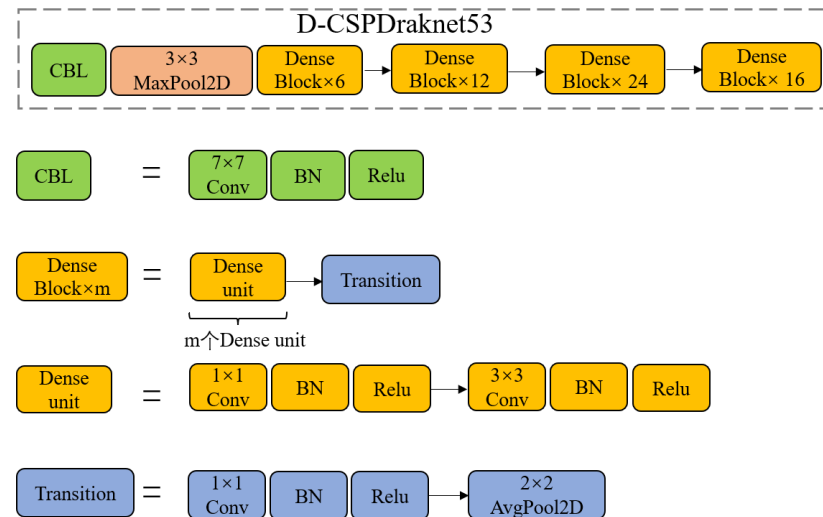

Figure 4. D-CSPDraknet53 structure diagram. 3×3 MaxPool2D: Max-Pooling with kernel size 3×3, Strides = 2; 7×7 Conv: Convolution with kernel size 7×7, Strides = 2; 7×7 Conv: Convolution with kernel size 7×7; 1×1 Conv: Convolution with kernel size 1×1; 2×2 AvgPool2D: Avg-Pooling with kernel size 2×2, Strides = 2.

The five original residual modules, Resblock_body, are replaced by a maximum pooling and four DenseBlocks (m = 6, 12, 24, 16), with m being the number of stacks of DenseBlocks. This design approach makes the connection between the layers of the network tighter, greatly simplifies the model complexity, reduces the number of operations, and improves the model detection speed. This type of connection mitigates gradient disappearance during network training. However, at the same time, the feature extraction effect of the model is reduced and the accuracy is slightly degraded. Thus, attention mechanism can be embedded to compensate and simultaneously enhance the model's detection accuracy in complex background situations.

3.2. Improvements to Enhance Feature Focus Based on Attention Mechanisms

The attention mechanism takes a reference from the human behavior of selectively focusing on the important parts of the received information and constructs a model that can redistribute the weight of target information from the irrelevant information in the information received by the network. The spatial attention mechanism is primarily designed to capture the correlation between pixel points in the input features. The channel attention mechanism aims to enhance the feature channels and amplify the target weights. Using them in combination results in better performance, but also inevitably increases the computational effort and number of parameters of the model. This paper introduces the SA-Net attention model using the idea of grouped convolution. SA-Net is not only a lightweight feature attention module but also allows feature attention in different dimensions. As shown in Figure 5, SA-Net receives the input features and splits the features into G sub-features in the channel dimension. The principle is shown in Equation (2). The feature weights are then updated in a parallel processing manner for each group of sub-features in multiple dimensions. To minimize the number of parameters in the attention model,

SA-Net further dichotomizes the groups of sub-features by the number of channels. The process is shown in Equation (3):

$$X = [X_1, \cdots, X_G], \ X_k \in R^{C/G \times H \times W} \tag{2}$$

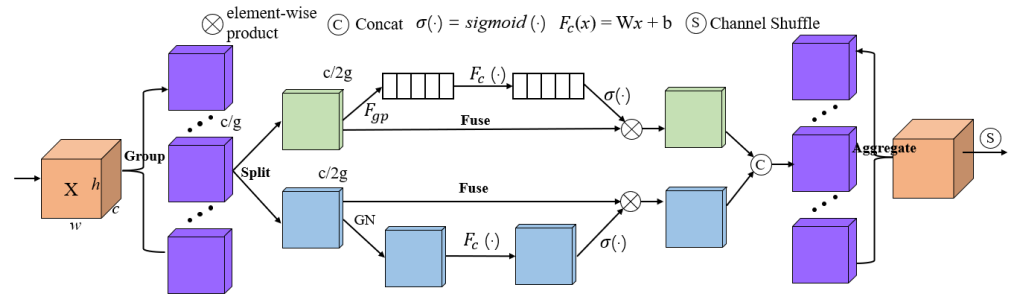$$X_k = [X_{k1}, X_{k2}], \ X_{k1}, X_{k2} \in R^{C/2G \times H \times W} \tag{3}$$

Figure 5. SA-Net structure diagram.

In Equations (2) and (3), X is the input feature and $X \in R^{C \times H \times W}$, C is the number of input channels, H is the input height, W is the input width, X_k is the kth sub-feature.

The dichotomized features are rescaled for channel and spatial attention weights so that the model focuses on channel and location information relevant to the target. The principle of the channel attention module is shown in Equations (4) and (5), and the principle of the spatial attention module is shown in Equation (6):

$$s = F_{gp}(X_{k1}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{k1}(i, j) \tag{4}$$

$$X'_{k1} = \sigma(F(s)) \cdot X_{k1} = \sigma(W_1 s + b_1) \cdot X_{k1} \tag{5}$$

$$X'_{k2} = \sigma(W_2 \cdot GN(X_{k2}) + b_2) \cdot X_{k2} \tag{6}$$

In Equations (4)–(6), F_{gp} is the global average pooling operation, σ is the sigmoid function, GN is the group normalization operation, W_1, W_2, b_1, b_2 are the parameters and $W_1, W_2, b_1, b_2 \in R^{C/2G \times 1 \times 1}$.

The two-dimensional branches are then connected to output a new feature layer after feature attention, as shown in Formula (7):

$$X'_k = [X'_{k1}, X'_{k2}] \in R^{C/G \times H \times W} \tag{7}$$

Finally, all updated sub-features are re-aggregated. However, there is no information communication between the feature layers. They are independent of each other and the feature updates in the channel dimension and the spatial dimension cannot be fused. Thus, it is also necessary to describe the joint relationship of features in space and channels in terms of channel interaction before output. This results in information fusion between different sub-features at the pixel level and makes the SA-Net module more efficient.

3.3. Improve Detection Module for Small Target Recognition

While considering the model's light weight and targeting of complex backgrounds, enhancing the algorithm's ability to recognize small targets is also key to evaluating the algorithm's performance.

As can be seen from Figure 6, the insulator defect targets occupy too small an area within the insulator as a whole. A large amount of shallow information is lost when deep features are used for prediction, resulting in deep features accounting for more global
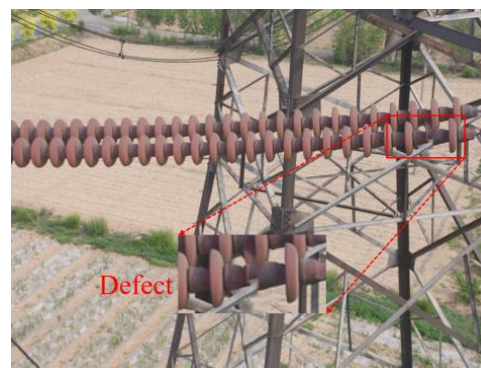
information and less detailed information. This is not conducive to the identification of small targets for insulator defects.
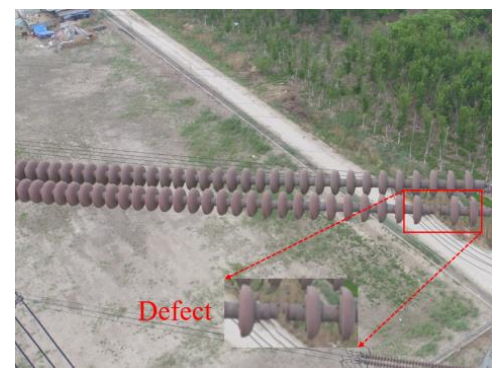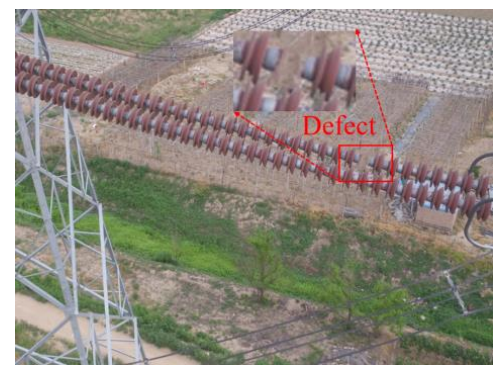

Figure 6. (**a**–**f**) Insulator defect pictures.

The principle of convolution operation is shown in Figure 7; when a 5 × 5 feature map is convolved once, a new 3 × 3 feature map is obtained. The S26-pixel point in this new feature map will contain the feature information from the 9 pixel points from S1 to S9. After another convolution, the pixel S27 will contain the feature information of the 25 pixel points from S1 to S25. Each detail of information in the 5 × 5 feature layer is given less weight in S27, which has higher overall information content. The network is deepened and the shallow detail information is compressed so that small target features cannot be easily identified.
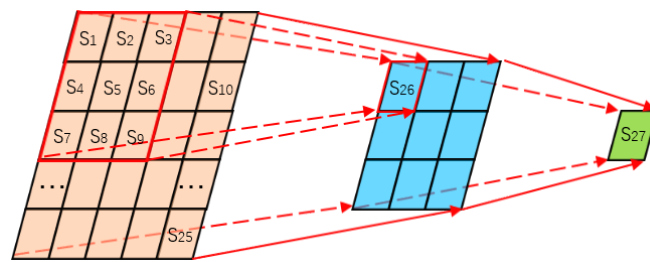
Figure 7. Convolution operation principle.

Therefore, in this paper, we add a shallow feature layer with a higher resolution of 104×104 to the original three feature detection layers of the YOLOv4 network. This operation facilitates the identification and localization of small broken targets by the detection module and improves the model's ability to identify small broken targets.

3.4. DSMH-YOLOv4 Model

The improved algorithm proposed in this paper is manifested in the following three aspects: firstly, the original YOLOv4 backbone feature extraction network is replaced by D-CSPDraknet53; secondly, the SA-Net attention mechanism is added between the backbone feature extraction network and the feature fusion layer; thirdly, a shallow feature layer is added to the feature detection layer to obtain more small target information. The structure diagram of the DSMH-YOLOv4 model is shown in Figure 8.
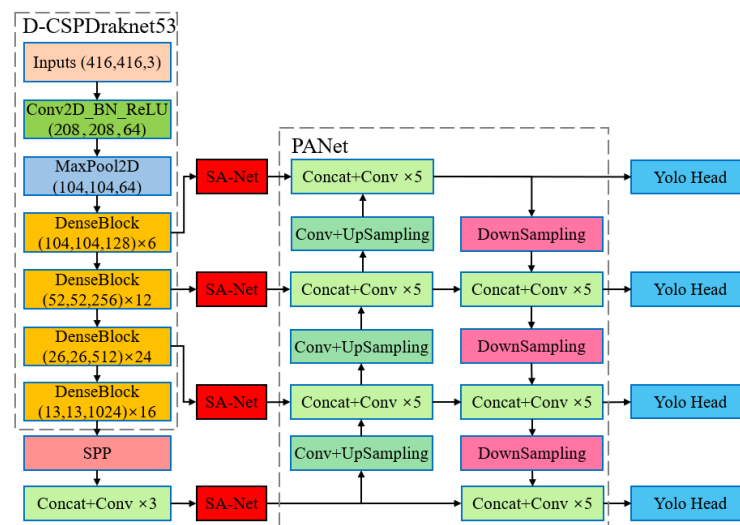

Figure 8. DSMH-YOLOv4 structure diagram.

The input image is extracted by the backbone network to obtain four effective feature layers of different sizes. In addition, the SA-Net rescales the initial effective feature layers from the miscellaneous ones to amplify the target weights and update the information interaction in the feature layers. PANet performs bi-directional sampling fusion using a feature layer with the target weights amplified by an attention mechanism, combining the information in the different scale feature layers to provide a database for improving the accuracy of the detection module.

4. Experimental Results and Analysis

4.1. Experimental Environment

This study used a deep learning framework based on the PyTorch 1.6 environment, with Ubuntu 18.08, Python 3.8.0, and CUDA = 11.2, where the training graphics card configuration was an RTX A6000/48G graphics card.

As this experimental dataset is partly from a publicly available dataset and partly collected in the field, the data were amplified and screened for labeling. The total number of datasets was 1588, of which the ratio of the training set, validation set, and test set was 8:1:1. This resulted in a training set of 1272, a validation set of 158, and a test set of 158. The number of positive and negative samples in the dataset was the number of insulators and insulator breakage samples. The numbers were 2908 and 715, respectively, with an overall ratio of around 4:1.

4.2. Experimental Process

During the training process, freeze training and thaw training were performed for the disclosed target detection model by loading the pre-trained model using the idea of migration learning. The frozen layer was trained with 50 epoch rounds and 300 rounds of thawing training. For the improved model, retraining was used, again set for 300 rounds. The initial learning rate of the model was 0.001, the batch size was set to 16, the momentum was 0.9, and the weight decay was 0.0005.

The loss curve obtained from the training is shown in Figure 9, where the horizontal coordinate represents the number of training rounds and the vertical coordinate is the loss change value during the training process. As shown in Figure 9, the loss values are greatly reduced at the beginning of the training process, indicating that the learning rate is appropriate. The curve levels off when the iteration reaches 200 epochs, indicating that the model had converged. The training loss and the validation loss of DMSH-YOLOv4 are lower than those of YOLOv4, respectively, and DMSH-YOLOv4 converges faster than YOLOv4. It is concluded that our model has lower overfitting and higher accuracy.
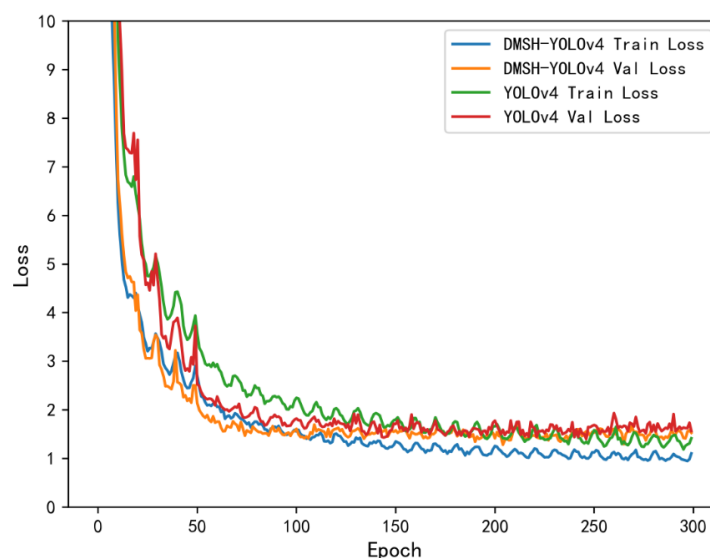

Figure 9. Loss curve during training.

4.3. Experimental Evaluation Indicators

To compare the performance of different algorithms, the performance of the algorithms was evaluated using average precision (AP), mean average precision (mAP), parameters, model size, FLOPs, and detection speed.

The average precision is the sum of the area of the curve enclosed by precision (P) and recall (R). The P-R graph is shown in Figure 10. The mAP is the average of the APs of the targets to be classified. Accuracy refers to the probability that a positive sample is predicted out of all the samples predicted to be positive, and recall refers to the probability that a positive sample is predicted out of the positive samples. The number of parameters affects the size of the model and the amount of computation affects the speed of the model. FLOPs refer to the number of floating point operations, i.e., the amount of computation, and are often used to measure the complexity of the model. FPS is the number of images

per second detected by the model and is often used to measure the detection speed of the model. Accuracy, recall, *AP* value, and *mAP* are calculated as shown in Equations (8)–(11):

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$AP = \int_{0}^{R} P(r)dr \tag{10}$$

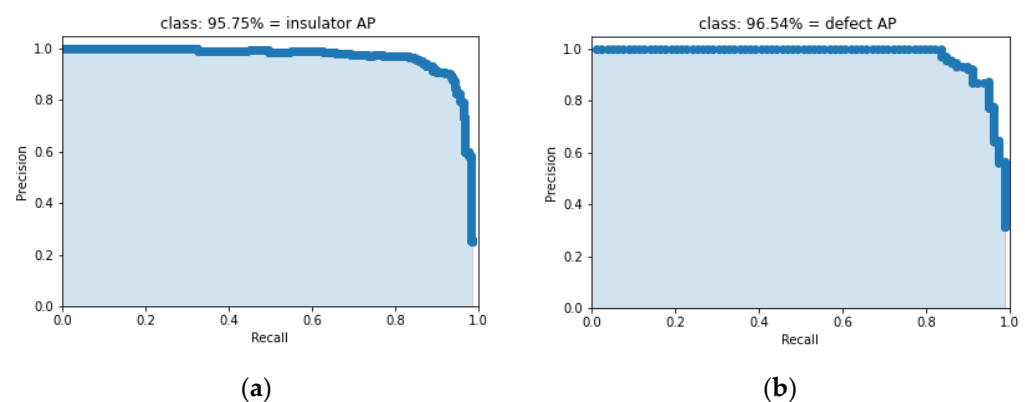$$mAP = \frac{AP_{\text{insulator}} + AP_{\text{defect}}}{2} \tag{11}$$

Figure 10. (**a**) Insulator P-R curve diagram; (**b**) defect P-R curve diagram.

In Equations (8) and (9), *TP* is the number of samples that the model considers to be positive and are indeed positive. *FP* is the number of samples that the model considers to be positive but which are actually negative. *FN* is the number of samples that the model considers to be negative but which are actually positive.

4.4. Comparison of Experimental Results

Table 1 shows the test results of comparison experiments between the algorithm in this paper and the current mainstream algorithms Fast R-CNN, SSD, YOLOv3, YOLOv4, YOLOv5, and YOLOX, with the same dataset and experimental parameters in this paper.

Table 1. Comparison of evaluation indicators of mainstream detection algorithms.

Model	Insulators AP (%)	Defects AP (%)	mAP (%)	Parameters	Model Size (MB)	FLOPs (G)	Speed (FPS)
Faster R-CNN	94.36	75.74	85.06	137,098,724	522.99	370.406	9.68
SSD	90.18	84.95	87.57	26,285,486	100.27	62.798	35.62
YOLOv3	92.15	91.27	91.71	61,949,149	236.32	66.096	23.58
YOLOv4	93.32	91.56	92.44	64,363,101	245.53	60.334	19.97
YOLOv5s	93.30	90.63	91.96	7,276,605	27.76	17.060	58.18
YOLOXs	96.35	91.88	94.11	8,968,255	34.21	26.806	46.18
DSMH-YOLOv4	95.75	96.54	96.14	16,718,580	63.78	29.347	29.33

As can be seen in Table 1:

(1) The mAP values of DSMH-YOLOv4 showed different degrees of improvement compared to the other six models, with a significant improvement of 11.08% and 8.57% compared to Faster R-CNN and SSD, respectively, and 4.43% and 3.7% compared to the YOLOv3 and original YOLOv4, and 4.18% and 2.03% compared to YOLOv5s and YOLOXs, respectively. The algorithm in this paper improved the detection accuracy of

insulator breakage even more, reaching 96.54%, compared to YOLOv4 and YOLOXs; the improvement was 4.98% and 4.66%, respectively, indicating that the improved algorithm in this paper has a more accurate localization effect on insulators and their defects.

(2) From the comparison of the number of parameters and computation, we can see that the computation and number of parameters of the original YOLOv4 model are 60.334 G and 64,363,101 respectively, and the model size is 245.53 MB. There is a degree of difficulty in deploying at the edge section. DSMH-YOLOv4 was lightened and improved to reduce the number of calculations and parameters by 51.36% and 74.02%, respectively, and the model size was reduced from 245.53 MB to 63.78 MB, indicating that the algorithm in this paper can achieve the requirements of lightening and real-time detection.

(3) The detection speed of DSMH-YOLOv4 was 29.33 FPS, which is 46.87% higher than the 19.97 FPS of the original YOLOv4 model. Although the detection speed is not as fast as that of SSD, YOLOv5s, and YOLOXs, the detection accuracy has a clear advantage. This indicates that DSMH-YOLOv4 can meet the requirements of deploying embedded devices with a combination of both detection accuracy and speed.

Table 2 shows the results of the experiments for the current mainstream lightweight YOLOv4 network.

Table 2. Comparison of evaluation indicators of mainstream lightweight YOLOv4 network.

Model	Insulators AP (%)	Defects AP (%)	mAP (%)	Parameters	Model Size (MB)	FLOPs (G)	Speed (FPS)
MobilenetV1-YOLOv4	91.00	83.21	87.11	12,692,029	53.51	10,540	53.12
MobilenetV2-YOLOv4	90.66	87.68	89.71	10,801,149	48.70	8153	51.74
MobilenetV3-YOLOv4	91.36	87.12	89.25	11,729,069	56.3	7599	46.55
D-CSPDarknet53-YOLOv4	92.92	90.70	91.81	16,483,909	62.71	26.315	32.34
Xu et al. [6]	91.21	93.00	92.11	14,963,517	57.08	10,905	44.25
DSMH-YOLOv4	95.75	96.54	96.14	16,718,580	63.78	29,347	29.33

As can be seen in Table 2, based on the lightweight model, the YOLOv4 model using D-CSPDarknet53 as the backbone network has the highest mAP value compared with the YOLOv4 model under the Mobilenet series backbone network. The mAP of D-CSPDarknet53-YOLOv4 is 4.7%, 2.1%, and 2.56% higher than that of MobilenetV1-YOLOv4, MobilenetV2-YOLOv4, and MobilenetV3-YOLOv4, respectively. Although the detection speed of D-CSPDraknet53-YOLOv4 is inferior to that of MobilenetV1-YOLOv4, MobilenetV2-YOLOv4, and MobilenetV3-YOLOv4, the detection accuracy of D-CSPDraknet53-YOLOv4 has obvious advantages. The improved algorithm proposed in [6] effectively improves the detection accuracy of insulator defects, but the average detection accuracy is still low, and the mAP value of the DSMH-YOLOv4 model is 4.03% higher than that of the improved algorithm proposed in literature 6.

Table 3 shows the results of the ablation experiments for the three modified methods.

Table 3. Comparison of evaluation indicators of various improved algorithms.

Model	D-CSPDarknet53	SA	4Head	Insulators AP (%)	Defects AP (%)	mAP (%)	Parameters	Model Size (MB)	FLOPs (G)	Speed (FPS)
YOLOv4				93.32	91.56	92.44	64,363,101	245.53	60.334	19.97
Algorithm 1	√			92.92	90.70	91.81	16,438,909	62.71	26.315	32.34
Algorithm 2		√		95.64	93.16	94.40	64,363,213	245.53	60.344	19.20
Algorithm 3			√	94.65	94.14	94.39	65,423,452	249.57	70.762	15.60
Algorithm 4	√	√		95.04	93.58	94.31	16,439,021	62.71	26.316	31.74
Algorithm 5	√		√	94.53	94.82	94.67	16,718,460	63.78	29.346	30.97
DSMH-YOLOv4	√	√	√	95.75	96.54	96.14	16,718,580	63.78	29.347	29.33

SA: Shuffle Attention Neural Networks; 4Head: add a shallow feature layer with 104 × 104 to the original three feature detection layers. The position of "√" in the table indicates that the improved algorithm adopts the corresponding strategy.

As can be seen in Table 3:

(1) Algorithm 1 replaces the backbone network with D-DSPDarknet53 only; the model size is compressed from 245.53 MB to 62.71 MB, the number of parameters is only 25.54%

of the original model, the computation is only 43.62% of the original model, the detection speed is improved by 12.37 FPS, and the compression effect is significant, but the insulator detection accuracy is decreased compared to that of the original YOLOv4 model by 0.4 % and the recognition accuracy of the defect decreased by 0.86%.

(2) Algorithm 2 embeds the SA-Net module in the original YOLOv4 model, with almost no increase in the number of parameters and computation, and no significant decrease in detection speed. The insulator detection accuracy increases by 3.06% compared to the original YOLOv4 model, and the recognition accuracy of the defect increases by 2.8%.

(3) Algorithm 3 adds a shallow feature layer to the detection layer of the original YOLOv4 model, increasing the number of parameters by 1.65% of the original model and decreasing the detection speed by 4.37 FPS, but the insulator detection accuracy increases by 1.33% compared to the original YOLOv4 model, and the defect recognition accuracy increases by 2.58%.

(4) The results of ablation experiments with Algorithms 3–5 show that D-DSPDarknet53 compresses a large number of parameters of the original model, and SA-Net improves the detection accuracy of the model and, while adding more detection layers, further improves the model's detection of small targets. The improved DSMH-YOLOv4 shows a 2.43% improvement in the detection accuracy of the insulator itself and a 4.98% improvement in the identification of small targets for insulator defects. The model compression is 25.98% of the original model, and the detection speed is improved by 46.87% compared to 19.97 FPS of YOLOv4, which is feasible for high-accuracy detection of edge-end insulator faults.

Table 4 further demonstrates the prediction results of each improved algorithm by selecting three typical aerial images of insulators. In Table 3, the red boxes are insulators and the blue boxes are defects. It can be seen from the recognition of Image 1 that when there are obscured insulators in the image, only Algorithm 2 with the fusion model can identify them, indicating that SA-Net is helpful in the recognition process for the obscured insulator targets. From the recognition of image 2, it can be seen that when there are multiple insulators in the image and the size difference is too large, only Algorithm 1 with the fusion improvement can identify the insulators of small targets, indicating that D-CSPDarknet53 does help in the recognition of small targets. From image 3 it can be seen that the addition of a multiple output detection layer can be of great help in the identification of broken targets.

Table 4. Test image comparison.

Model	Image 1	Image 2	Image 3
YOLOv4			
Algorithm 1			

Table 4. *Cont.*

Model	Image 1	Image 2	Image 3
Algorithm 2			
Ours			

To further visualize the attention region in the model, this paper presents a fractional heat map visualization analysis of the predicted results of the DSMH-YOLOv4 model. As can be seen from Figure 11, the heat map focuses on the center of the target detection box, with the focal area at the center of the target in red and a decreasing proportion of attention spreading outwards. This further validates the model improvements. The green boxes represent detected insulators and the orange boxes represent detected defects.
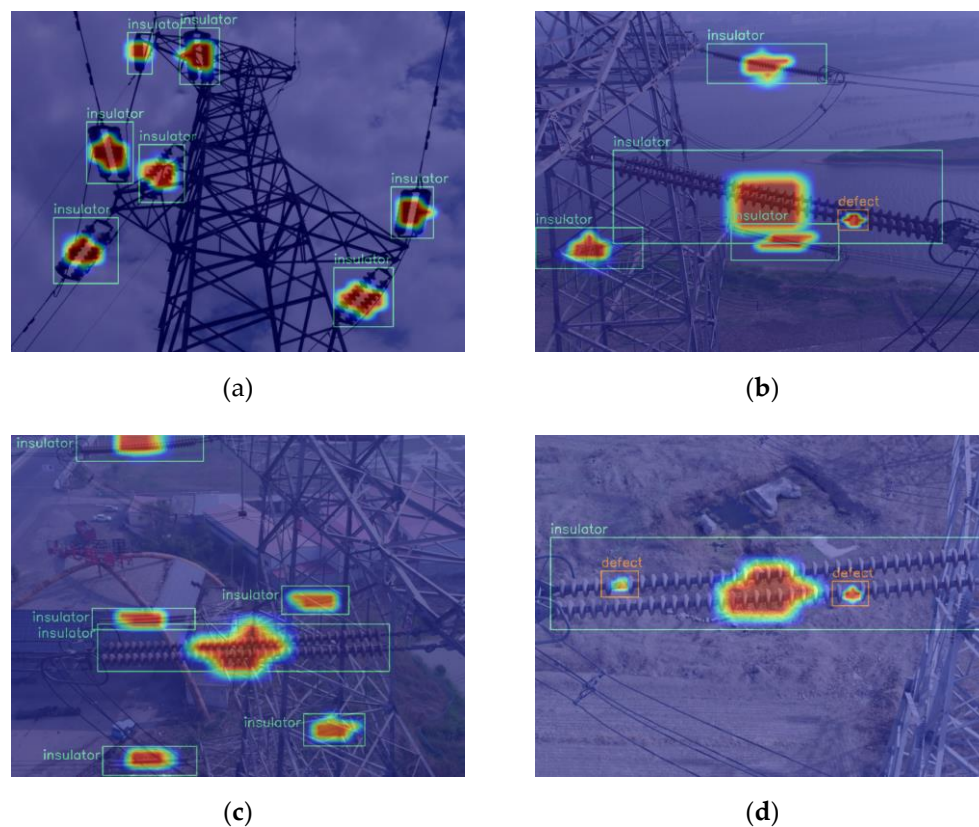

(a)

(b)

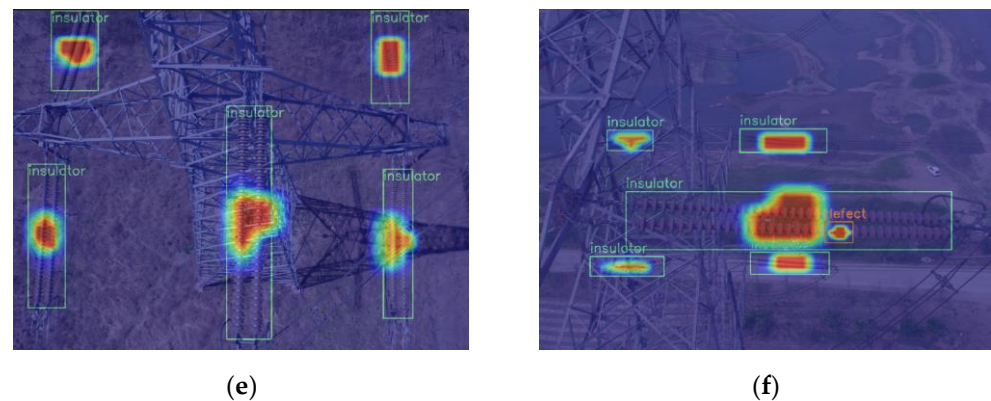(c)

(d)

Figure 11. *Cont.*

(**e**)　　　　　　　　　　　　　　　　　　(**f**)

Figure 11. Heat map of the detection of the DSMH-YOLOv4 model. (**a**) Seven insulators have been detected; (**b**) four insulators and a defect have been detected; (**c**) six insulators have been detected; (**d**) one insulator and two defects have been detected; (**e**) five insulators have been detected; (**f**) five insulators and a defect have been detected.

5. Conclusions

This paper proposes an improved algorithm, DSMH-YOLOv4, for insulator and defect detection based on YOLOv4. Firstly, the backbone feature extraction network is replaced with D-CSPDarknet53 to significantly reduce the number of parameters and computational effort in the model. In addition, an attention mechanism is embedded before the feature fusion and the promiscuous feature layers are used to enhance the feature extraction capability of the network. Finally, a large-size detection layer is added to enhance the network's ability to identify small targets for insulator defects.

The experimental results show that the number of parameters of the improved network model is only 25.98% of that of the original model; the detection speed is improved by 9.36 FPS; the recognition accuracy of insulator defect is 96.54%, which is 4.98% higher than that of the original YOLOv4 network; and the mean average accuracy of mAP for insulators and their defects is improved from 92.44% to 96.14%. The improved algorithm significantly improves the performance and provides an effective way to realize the embedded application of real-time transmission line insulator location and defect recognition.

Author Contributions: Conceptualization, G.H. and Q.Y.; methodology, G.H.; software, Q.Y.; validation, Q.Y., R.W. and L.Z.; formal analysis, G.H.; investigation, F.Z. and S.L.; resources, S.Y.; data curation, M.H., F.Z. and S.L.; writing—original draft preparation, Q.Y.; writing—review and editing, G.H., Q.Y. and L.Q.; visualization, R.W.; supervision, S.Y. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Dataset link: https://github.com/InsulatorData/InsulatorDataSet (accessed on 11 April 2022).

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

References

1. Ru, C.; Zhang, S.; Zhang, Z.; Zhu, Y.; Liang, Y. Fault Identification Method for High Voltage Power Grid Insulator Based on lightweight Mobilenet-SSD and MobileNetV2-DeeplabV3+ Network. *High Volt. Eng.* **2022**, *48*, 3670–3679.
2. Hao, S.; Ma, R.; Zhao, X.; Ma, X.; Wen, H.; An, B. Self-Explosion Fault Detection Algorithm for Glass Insulator Based on Super-Resolution Deep Residual Network. *High Volt. Eng.* **2022**, *48*, 1817–1825.
3. Hao, Y.; Liang, W.; Yang, L.; He, J.; Wu, J. Methods of image recognition of overhead power line insulators and ice types based on deep weakly-supervised and transfer learning. *IET Gener. Transm. Distrib.* **2022**, *16*, 2140–2153. [CrossRef]
4. Han, G.; He, M.; Gao, M.; Yu, J.; Liu, K.; Qin, L. Insulator Breakage Detection Based on Improved YOLOv5. *Sustainability* **2022**, *14*, 6066. [CrossRef]

5. Sadykova, D.; Pernebayeva, D.; Bagheri, M.; James, A. IN-YOLO: Real-Time Detection of Outdoor High Voltage Insulators Using UAV Imaging. *IEEE Trans. Power Deliv.* **2020**, *35*, 1599–1601. [CrossRef]

6. Xu, S.; Deng, J.; Huang, Y.; Ling, L.; Han, T. Research on Insulator Defect Detection Based on an Improved MobilenetV1-YOLOv4. *Entropy* **2022**, *24*, 1588. [CrossRef]

7. Han, G.; Li, T.; Li, Q.; Zhao, F.; Zhang, M.; Wang, R.; Yuan, Q.; Liu, K.; Qin, L. Improved Algorithm for Insulator and Its Defect Detection Based on YOLOX. *Sensors* **2022**, *22*, 6186. [CrossRef]

8. Hao, K.; Chen, G.; Zhao, L.; Li, Z.; Liu, Y.; Wang, C. An Insulator Defect Detection Model in Aerial Images Based on Multiscale Feature Pyramid Network. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3522412. [CrossRef]

9. Waleed, D.; Mukhopadhyay, S.; Tariq, U.; El-Hag, A.H. Drone-Based Ceramic Insulators Condition Monitoring. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 6007312. [CrossRef]

10. Li, B.; Zeng, J.; Zhu, X.; Wang, S.; Guo, Z.; Liu, H. Insulator defect detection network based on multi-scale context awareness. *High Volt. Eng.* **2022**, *48*, 2905–2914.

11. Zhai, Y.; Yang, K.; Wang, Q.; Wang, Y. Disc Insulator Defect Detection Based on Mixed Sample Transfer Learning. Proceedings of the CSEE. Available online: https://kns.cnki.net/kcms2/article/abstract?v=iJQOcxL8JgxGpwh6a5pyNkiV4NA4bKr6 Q3GvlsIcuvTz5BpQfttspv3LImqMB3lnLXSpS8_PW3DPTxm-wewaGlJFHjkUNfLnIhZWWaVPnwHbAiHgRMCK6Q==& uniplatform=NZKPT&language=CHS. (accessed on 1 May 2022).

12. Saini, V.K.; Kumar, R.; Mathur, A.; Saxena, A. Short Term Forecasting Based on Hourly Wind Speed Data Using Deep Learning Algorithms. In Proceedings of the 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things ICETCE 2020, Jaipur, India, 7–8 February 2020; pp. 30–35.

13. Yang, Z.; Xu, Z.; Wang, Y. Bidirection-Fusion-YOLOv3: An Improved Method for Insulator Defect Detection Using UAV Image. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3521408. [CrossRef]

14. Qiu, Z.; Zhu, X.; Liao, C.; Shi, D.; Qu, W. Detection of Transmission Line Insulator Defects Based on an Improved Lightweight YOLOv4 Model. *Appl. Sci.* **2022**, *12*, 1207. [CrossRef]

15. Zhang, Z.; Zhang, B.; Lan, C.; Liu, H.; Li, D.; Pei, L.; Yu, W. FINet: An Insulator Dataset and Detection Benchmark Based on Synthetic Fog and Improved YOLOv5. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 6006508. [CrossRef]

16. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Cheng-Yang, F.; Alexander, C. SSD: Single Shot Multibox Detector. In Proceedings of the European Conference on Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37.

17. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

18. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.

19. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

20. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.

21. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition IEEE Computer Society, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

22. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef]

24. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.

25. Jia, X.; Yu, Y.; Guo, Y.; Huang, Y.; Zhao, B. Lightweight detection method of self-explosion defect of aerial photo insulator. *High Volt. Eng.* **2022**. [CrossRef]

26. Liu, J.; Liu, C.; Wu, Y.; Xu, H.; Sun, Z. An improved method based on deep learning for insulator fault detection in diverse aerial images. *Energies* **2021**, *14*, 4365. [CrossRef]

27. Zhu, Y.; Zheng, Y.; Qin, J. Insulator target detection based on improved YOLOv3. *Insul. Surge Arresters* **2022**, *3*, 166–171.

28. Yang, L.; Fan, J.; Song, S.; Liu, Y. A light defect detection algorithm of power insulators from aerial images for power inspection. *Neural Comput. Applic.* **2022**, *34*, 17951–17961. [CrossRef]

29. Yang, Z.; Xu, X.; Wang, K.; Li, X.; Ma, C. Multitarget detection of transmission lines based on DANet and YOLOv4. *Sci. Program.* **2021**, *2021*, 6235452. [CrossRef]

30. He, H.; Huang, X.; Song, Y.; Zhang, Z.; Wang, M.; Chen, B.; Yan, G. An insulator self-blast detection method based on YOLOv4 with aerial images. *Energy Rep.* **2022**, *8*, 448–452. [CrossRef]

31. Gao, W.; Zhou, C.; Guo, M. Insulator defect identification via improved YOLOv4 and SR-GAN algorithm. *Electr. Mach. Control* **2021**, *25*, 93–104.

32. Han, G.; He, M.; Zhao, F.; Xu, Z.; Zhang, M.; Qin, L. Insulator detection and damage identification based on improved lightweight YOLOv4 network. *Energy Rep.* **2021**, *7*, 187–197. [CrossRef]

33. Zhang, Q.; Yang, Y. SA-Net: Shuffle attention for deep convolutional neural networks. *arXiv* **2021**, arXiv:2102.00240.

34.   He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *arXiv* **2015**, arXiv:1512.03358.
35.   Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.