



ARTICLE

An Improved Solov2 Based on Attention Mechanism and Weighted Loss Function for Electrical Equipment Instance Segmentation

Junpeng Wu^{1,2,*}, Zhenpeng Liu², Xingfan Jiang², Xinguang Tao² and Ye Zhang³

¹Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin, 132012, China

²School of Electrical Engineering, Northeast Electric Power University, Jilin, 132012, China

³Baishan Power Supply Company, State Grid Jilin Electric Power Co, Ltd., Jilin, 134300, China

*Corresponding Author: Junpeng Wu. Email: junpengwu80@163.com

Received: 06 September 2023 Accepted: 19 November 2023 Published: 30 January 2024

ABSTRACT

The current existing problem of deep learning framework for the detection and segmentation of electrical equipment is dominantly related to low precision. Because of the reliable, safe and easy-to-operate technology provided by deep learning-based video surveillance for unmanned inspection of electrical equipment, this paper uses the bottleneck attention module (BAM) attention mechanism to improve the Solov2 model and proposes a new electrical equipment segmentation mode. Firstly, the BAM attention mechanism is integrated into the feature extraction network to adaptively learn the correlation between feature channels, thereby improving the expression ability of the feature map; secondly, the weighted sum of CrossEntropy Loss and Dice loss is designed as the mask loss to improve the segmentation accuracy and robustness of the model; finally, the non-maximal suppression (NMS) algorithm to better handle the overlap problem in instance segmentation. Experimental results show that the proposed method achieves an average segmentation accuracy of mAP of 80.4% on three types of electrical equipment datasets, including transformers, insulators and voltage transformers, which improve the detection accuracy by more than 5.7% compared with the original Solov2 model. The segmentation model proposed can provide a focusing technical means for the intelligent management of power systems.

KEYWORDS

Deep learning; electrical equipment; attention mechanism; weighted loss function

1 Introduction

With the continuous development of modern industrial production, electrical equipment is the core component of industrial production, monitoring and maintaining its operating status to ensure production safety and efficiency is of great significance [1]. During the operation of electrical equipment, some faults that cannot be detected and processed promptly may lead to the shutdown of the electrical equipment, bringing huge economic losses to the enterprise, and accordingly the early-warning realization of online monitoring of electrical equipment and potential faults is important practically [2]. With the rapid development of computer vision and deep learning technology,



instance segmentation technology has provided a more effective means for electrical equipment fault detection and diagnosis. Through the realization of accurate positioning of electrical equipment, the performance of electrical equipment can be better analyzed. It can also provide strong support for the intelligent development of the power industry [3].

Unmanned inspection technology, utilizing automation and machine vision technology, has become increasingly popular in recent years, replacing manual inspection and monitoring of electrical equipment in regular inspections. This technology offers many advantages such as high efficiency, accuracy, safety, and reduced labor consumption [4]. The instance segmentation model can quickly and accurately recognize and segment electrical equipment to assist the unmanned inspection system in carrying out intelligent inspection. Compared with traditional manual inspection, unmanned inspection technology based on instance segmentation can greatly reduce inspection time and improve work efficiency, which can reduce the demand for human resources and lower the inspection cost. In addition, faults can be detected in advance and maintained promptly, reducing the downtime and losses caused by equipment failures and lowering operating costs.

Despite many advantages of instance segmentation techniques in unmanned inspection, there are still some challenges and shortcomings. In practical applications, instance segmentation of electrical equipment images faces the loss of detailed information in some regions [5], large differences in sizes [6], object overlapping [7] and other problems, which highly limit the application effect of traditional segmentation algorithms. To better adapt to the instance segmentation task of electrical equipment images, this paper proposes an improved Solov2 model. Firstly, the BAM attention mechanism is introduced into the backbone network to make the network pay more attention to the important areas in the image. Secondly, the weighted Dice Loss and CrossEntropy Loss, which are weighted by a 6:4 ratio, are used as the mask loss function to better segment electrical equipment of different sizes. Finally, an improved NMS post-processing algorithm is used so that the model can better distinguish different objects, thus improving the accuracy of image segmentation of electrical equipment.

The remainder of this paper is arranged as follows: [Section 2](#) introduces relevant literature. [Section 3](#) introduces the overall structure of the improved Solov2 segmentation model and the key functional modules involved in the proposed model. [Section 4](#) describes the dataset, experimental environment, and evaluation indicators used for the model. [Section 5](#) is the analysis of experimental results. Finally, [Section 6](#) summarizes the research of this article and provides prospects for future research.

2 Literature Review

Image segmentation algorithms have been proposed and applied to electrical equipment image segmentation with the continuation of deep learning in the field of image segmentation. Early image segmentation methods mainly rely on pixel similarity and color differences, making it difficult to segment objects with high accuracy, and usually requires manual adjustment of parameters.

With the rise of deep learning, image segmentation algorithms based on convolutional neural networks have gained importance. The regions with convolutional neural network features (R-CNN) algorithm proposed by Girshick et al. [8] combines target detection and image segmentation to detect objects in the image and segment them. Long et al. [9] proposed a fully convolutional neural network (FCN) for image segmentation, which replaces the fully connected layer in traditional convolutional neural networks with a convolutional layer, achieving end-to-end image segmentation. A deep learning-based target detection algorithm, Faster R-CNN, proposed by Ren et al. [10], introduces the Region Proposal Network (RPN), which uses the RPN network to generate candidate regions,

then uses the region of interest (RoI) pooling method to transform candidate regions into fixed-size feature maps, and finally performs classification and regression through the fully connected layer. The SegNet encoder-decoder structure-based image segmentation algorithm was proposed by Badrinarayanan et al. [11] which saves the index of the maximum value of the pooling layer in the convolutional neural network for subsequent upsampling operations. A symmetric encoder-decoder architecture network U-Net was proposed by Ronneberger et al. [12], which adds skip connections between the encoder and decoder, allowing the network to better utilize low-level and high-level features for segmentation. Dai et al. [13] proposed the Instance-aware Semantic Segmentation algorithm (IAS), which can more accurately identify and segment images by performing semantic segmentation on the instance segmentation results of objects. He et al. [14] proposed Mask R-CNN, an instance segmentation algorithm based on Faster R-CNN, which can achieve target detection and instance segmentation simultaneously. The Deep lab V3+ network was proposed by Chen et al. [15] that draws on the encoder decoder architecture of Feature Pyramid Networks (FPN) and other networks to achieve cross-regional fusion of feature matching and use group convolution to accelerate. The BlendMask algorithm from Chen et al. [16] improved the accuracy of object recognition and segmentation in images by blending the results of instance segmentation.

Ma et al. [1] designed a sample mask automatic labeling method based on thermal image guidance and a Progressive Optimization Model (POM), which achieved automatic labeling of instance masks and solved problems such as the difficulty of segmenting electrical equipment with complex structures. Yan et al. [2] studied a substation equipment instance segmentation model based on Mask R-CNN, which has a better segmentation effect on the substation equipment image dataset. Li et al. [17] proposed an IR insulator image segmentation algorithm based on dynamic mask and box annotation to achieve the overall segmentation of insulator strings by marking insulators with rectangular boxes in the infrared image. Zhou et al. [3] established the intrinsic connection between the parameters and the statistics of the image by optimizing the pulse-coupled neural network model, combining the information of the gradient of the image and the principle of great likelihood estimation, thus improving the precision of the fault segmentation of electrical equipment. Liu et al. [18] achieved instance segmentation of sparse radar detection points by using semantic information clustering and a visual multilayer perceptron. Liu et al. [19] proposed two innovative multiple kernel clustering algorithms to solve the problem of incomplete kernel matrices, which provide a new way to deal with incomplete data for instance segmentation tasks. Yu et al. [20] and Zhou et al. [21] addressed the image processing problem of water pollutant detection and infrared heat trace extraction by introducing the immune system correlation approach. An improved genetic algorithm based on immune thought was proposed by Dong et al. [22] for infrared image segmentation. Yıldız et al. [23] proposed a size-based instance pruning method for accurate segmentation of cell nuclei in histological images. These methods based on deep learning have provided many techniques for image recognition and segmentation. However, these methods still have some challenges in image segmentation of electrical devices, so we propose a new segmentation model for electrical device instances based on Solov2. This model can effectively improve the detection and segmentation accuracy of electrical equipment.

The contributions of this paper are as follows:

- 1) We incorporate the BAM attention mechanism in the feature extraction network, which improves the representation of the feature map by adaptively learning the correlation between feature channels.

2) The weighted sum of CrossEntropy loss and Dice loss is used as the mask loss function, and the model suitable for electrical equipment segmentation is better trained by reasonably weighing the weights of the two loss functions.

3) We improve the non-majority suppression algorithm to more accurately deal with the overlap problem in the segmentation of electrical device instances and improve the segmentation accuracy.

3 Image Instance Segmentation Framework Based on Improved Solov2

Solov2 is a neural network model for instance segmentation, which can be divided into three parts: backbone network, feature pyramid, and segmentation network. The backbone network usually uses the residual network (ResNet) [24] backbone network to extract image features. On top of the backbone network, Solov2 introduces the feature pyramid network (FPN) [25] to improve the model's detection ability for targets of different scales. In the segmentation network, Solov2 uses multiple parallel instance segmentation branches, with each branch responsible for generating a segmentation mask for an instance. Next, through an NMS post-processing step, Solov2 segments neighboring targets to avoid duplicate detection. Finally, the highest score segmentation result is mapped back to the original image space, and the predicted masks of each instance are combined to obtain the final segmentation image.

In the instance segmentation task of electrical equipment, the original Solov2 model has poor accuracy in detection and segmentation. The model cannot accurately identify the segmented electrical equipment regions in practical applications. Therefore, the Solov2 model is improved based on the original model to better apply it to the instance segmentation task of electrical equipment in this paper. The improved Solov2 framework is shown in Fig. 1. First, the electrical equipment image is processed by the backbone network to extract features for multiple feature maps. The output of the feature extraction network is sent to the FPN structure, where the FPN integrates features of different levels to obtain a multi-scale feature representation. Then, the output of the FPN is sent to the semantic branch network, which processes these feature maps to obtain semantic information for each pixel. Meanwhile, the output of the FPN is also sent to the mask branch, where the feature branch processes the feature maps of each scale to obtain a feature vector for each pixel. The results of mask prediction and semantic prediction are sent to a matching network, which matches them to obtain the mask segmentation result for each instance. Finally, the obtained mask segmentation results are fed into a post-processing module including non-maximal value suppression and binarisation of the mask segmentation to remove overlapping instances and further optimize the segmentation results.

3.1 BAM Module

In order to detect and distinguish different features of electrical equipment more accurately, in this paper, the BAM module is added to the last convolutional layer from stage 2 to stage 5 in the ResNet50 network [26], the structure of which is shown in Fig. 2. The BAM module can better capture the local and global features of the electrical equipment through the learnt attentional weights and the weighted fusion of the information from the different feature maps, which can increase the perception and generalization ability of the network, thus improving the accuracy of the model for segmentation of electrical equipment. The structure of the BAM module consists of two branches, the channel attention branch and the spatial attention branch.

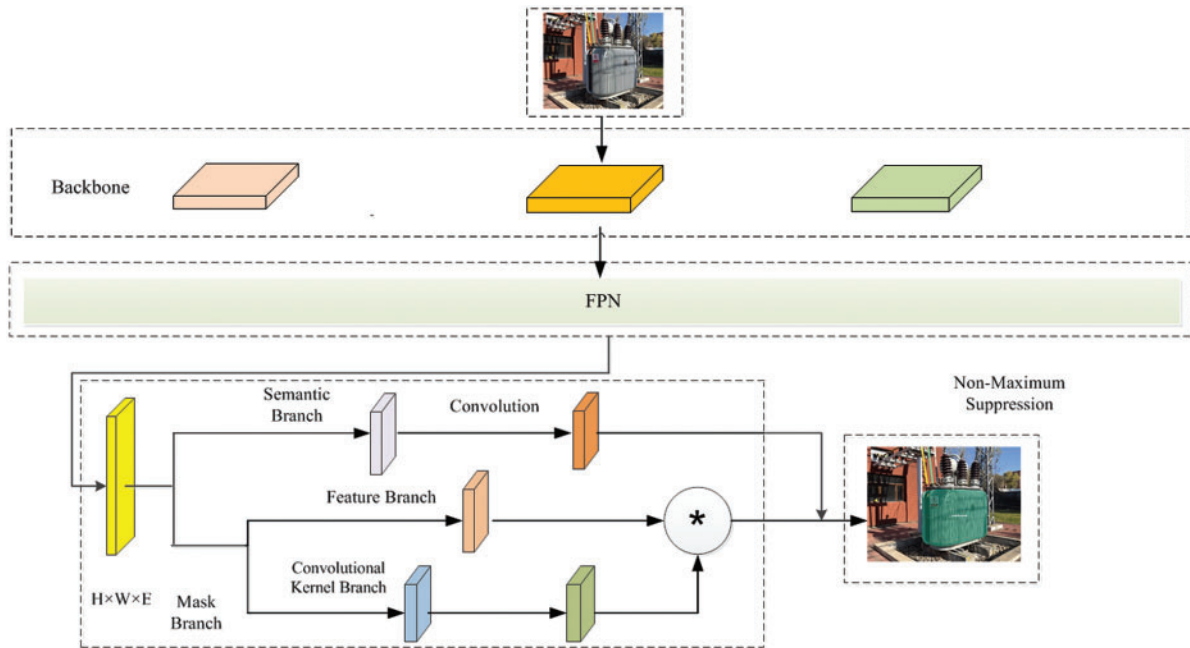


Figure 1: Improved structure diagram of Solov2 model

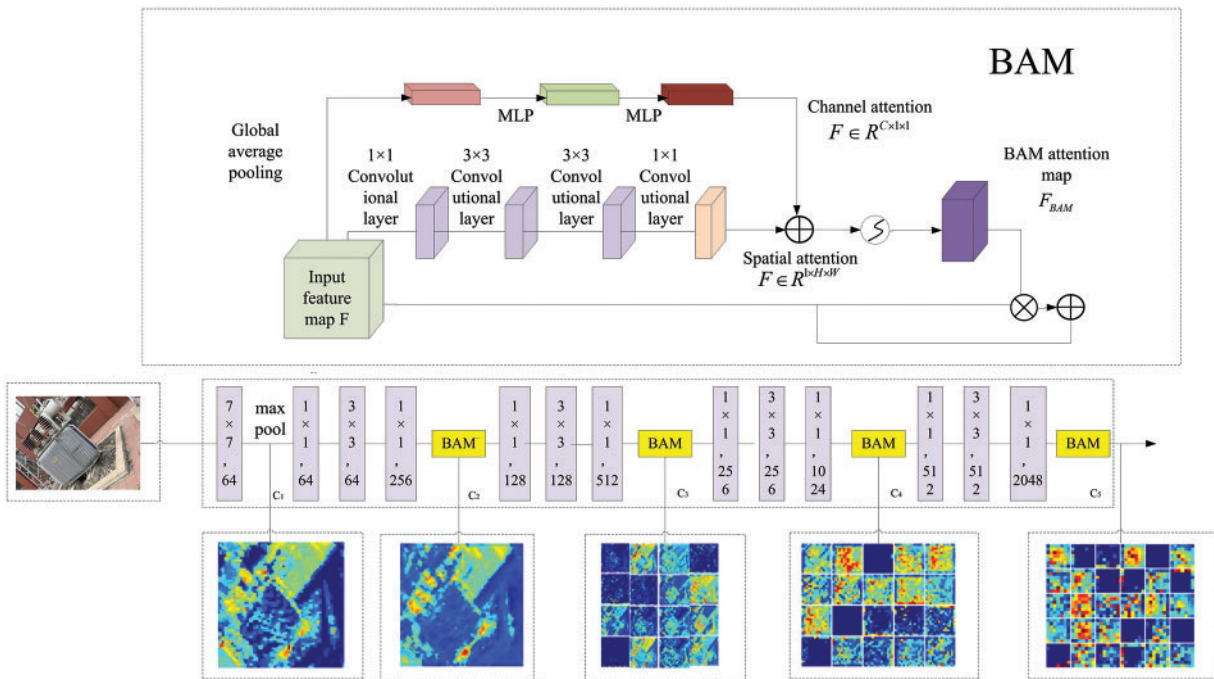


Figure 2: BAM module combined with ResNet50

The channel attention branch is mainly used to learn the correlation between different channels to control the importance of each channel. The calculation formula of the channel attention branch is shown in Eq. (1):

$$F_{Channel} = f_{BN}(W_1(W_0 F_{AvgPool} + b_0) + b_1) \quad (1)$$

where f_{BN} denotes batch normalization calculation, W_0 and b_0 represent the weight and bias of the hidden layer of MLP, and represent the weight and bias of the output layer of MLP, $W_0 \in R^{\frac{C}{r} \times C}$, $b_0 \in R^{\frac{C}{r}}$, $W_1 \in R^{\frac{C \times C}{r}}$, $b_1 \in R^C$.

The spatial attention branch can effectively improve the model's perception ability of the internal structure of electrical equipment, and can better distinguish the subtle details and features in electrical equipment, thereby improving the accuracy of detection and segmentation. The calculation formula of the spatial attention branch is shown in Eq. (2):

$$F_{Spatial} = f_{BN}(j_3^{1 \times 1}(j_2^{1 \times 1}(j_1^{1 \times 1}(j_0^{1 \times 1}(F)))))) \quad (2)$$

where j denotes convolution operation, f_{BN} denotes batch normalization operation, the superscript denotes the size of the convolution filter, and the subscript denotes the order of the operation.

After obtaining the channel attention map $F_{Channel}$ and spatial attention map $F_{Spatial}$ from the two attention branches, they are combined to generate the final three-dimensional attention map. The calculation formula is shown in Eq. (3):

$$\begin{aligned} F_{BAM} &= \sigma(F_{Channel} + F_{Spatial}) \\ F' &= F + F \otimes F_{BAM} \end{aligned} \quad (3)$$

where \otimes denotes element-wise multiplication, and σ denotes the sigmoid function.

3.2 Dice Loss and CrossEntropy Loss

The Solov2 algorithm uses category branching to achieve prediction of the target category and mask branching to achieve mask prediction of the corresponding instances under the category [27]. Thus, the total loss function L is:

$$L = L_{cate} + \lambda L_{mask} \quad (4)$$

where L_{cate} is the category loss function and L_{mask} is the mask loss function, λ is a hyperparameter and takes the value of 2.

Among them, the mask loss function is implemented by the dice loss function Dice Loss, to measure the prediction result of the model on the target instance mask. Dice loss mainly focuses on the area of overlap between predictions and true labels, but when the number of categories is high or the overlap of mini-objectives is low, it is prone to ignore a few categories or mini-objectives, leading to insufficient learning. CrossEntropy Loss (CE Loss) focuses on the accuracy of the overall prediction, can be fair in dealing with positive and negative samples, and the value of the loss is averaged over the whole, which can make the model for the entire dataset more accurate classification results. In this paper, the two loss functions are used at the same time, and the weighted sum of CrossEntropy Loss and Dice Loss can be introduced as the loss function, which can take into account the overlap of the pixel level and the accuracy of the overall prediction, to evaluate and optimize the performance of the model more comprehensively. Thus, the model can be better applied to the segmentation task of electrical equipment.

The improved expression for the mask loss function proposed in this paper is shown in Eq. (5):

$$L_{mask} = \alpha \times DiceLoss + (1 - \alpha) \times CELoss \quad 0 < \alpha < 1 \quad (5)$$

where α is the weighting factor,

$$DiceLoss = 1 - \frac{2 \sum_{x,y} (p_{x,y}^2 \times q_{x,y}^2)}{\sum_{x,y} p_{x,y}^2 + \sum_{x,y} q_{x,y}^2} \quad (6)$$

where $p_{x,y}$ denotes the pixel value predicted by the mask at (x, y) ; $q_{x,y}$ denotes the pixel value of the true mask at (x, y) .

$$CELoss = - \sum_{i=1}^C p_i \log(q_i) \quad (7)$$

where p_i is the true probability distribution, q_i is the probability distribution of the model output and C is the number of categories.

The CrossEntropy loss ensures that the model performs accurate segmentation of different classes of electrical equipment, while the Dice loss can assess the degree of similarity between the model's predicted results and the true segmentation mask. The paper takes a weighted sum to combine the two loss functions as a mask loss function, which can optimize the model's performance in both electrical equipment detection segmentation and classification tasks, thereby improving the accuracy of the segmentation.

3.3 An Improved NMS Algorithm

Matrix NMS is used in Solov2 to screen out redundant detection frames. Matrix NMS uses a matrix calculation to process the detection results, representing each detection as a row vector and sorting them by confidence, calculating the overlap matrix between them, and using matrix operations to derive the final non-maximum suppression result. Matrix NMS is calculated as shown below:

$$S_f = \begin{cases} s_i & IoU(M, b_i) < Nt \\ s_i \times (1 - IoU(M, b_i)) & IoU(M, b_i) \geq Nt \end{cases} \quad (8)$$

where b_i is the i -th frame to be detected, S_i is the original score of b_i , S_f is the final score of b_i , M is the frame with the highest score in each round, N_t is a preset threshold, and $IoU(M, b_i)$ is the intersection ratio of b_i and M .

From the above equation, it can be seen that intersection over union (IoU) is the only consideration factor, but in electrical equipment application scenarios, when two detection frames have equal areas, the value of IoU can be affected by the distance between them, leading to the occurrence of missed detections. To address this issue, this paper calculates the overlap metric by considering the distance between the two detection frames, which avoids the traditional IoU misclassification caused by this distance. The steps for the improved non-maximum suppression algorithm are shown in [Fig. 3](#).

```

Input: Initial set of candidate boxes  $B=\{b_1,\dots,b_N\}$ 
Confidence set  $S=\{S_1,\dots,S_N\}$ 
Begin
Sort the candidate boxes according to their scores from highest to lowest
Define the empty set  $D$ 
While set  $B$  is not empty do
  Move the candidate box with the highest confidence level  $M$  from set  $B$  into set  $D$ 
  Define the empty set  $H$ 
  for  $b_j$  in  $B$  do
    Calculate the  $DIoU$  between candidate box  $b_j$  and candidate box  $M$ 
    Add the resulting  $DIoU$  to the set  $H=\{h_1,\dots,h_N\}$ 
  end for
  Define the empty set  $J$ 
  if  $h_i > 0.5$ 
    Add  $h_i$  to the set  $J$ 
  end if
  end for
  Remove the same elements from set  $B$  as set  $J$  to get a new set  $B$ 
end while
Output: Set  $D$ 
END

```

Figure 3: The NMS algorithm

The improved non-extreme value suppression function is shown in the following equation:

$$S_i = \begin{cases} s_i & DIoU(M, b_i) < Nt \\ s_i \times (1 - DIoU(M, b_i)), & DIoU(M, b_i) \geq Nt \end{cases} \quad (9)$$

of which

$$DIoU = IoU - \frac{P^2(b, b^{gt})}{c^2} = IoU - \frac{d^2}{c^2} \quad (10)$$

where b is the centre of the prediction box, b^{gt} is the centre of the real box, $P^2(b, b^{gt})$ represents the square d^2 of the distance between the real box and the center of the prediction box, and c represents the diagonal length of the minimum closure area of the two boxes.

In this paper, the distance between the centroids of two frames is added as an important factor in Matrix NMS to replace the traditional IoU calculation, which can reflect the overlap of two frames more accurately. Therefore, if the overlap of two boxes is large, but the center distance between the two boxes is also large, it will not be considered as overlapping boxes to avoid missed detection, making the remaining detection boxes more accurate and improving the accuracy of model segmentation.

4 Experimental Preparation

4.1 Datasets

The electrical equipment image dataset used in this paper consists of several sources, some of which are obtained from mobile phone photography and others from Chinese Power Line Insulator Dataset (CPLID) [28]. The captured images and the supplied images are saved in jpg format. The original dataset consists of a total of 600 images. To improve the generalization and robustness of the model, a data enhancement approach is adopted, including random flip, rotation and scaling operations. At the same time, a stratified sampling approach is adopted, and for the problem of uneven sample size in each category, sampling is weighted according to the number of categories in order to ensure that each category has approximately the same number of samples in the training set. After data augmentation and stratified sampling, a dataset of 2000 training sets, 400 validation sets and 200 test sets is finally obtained, thus greatly improving the training effect and generalization capability of the model. After data augmentation, the annotation method is pixel-level annotation, the annotation tool labelme is

used to add labels to the images, so that the corresponding annotation files are saved in json format, and the annotation files are saved in coco2017 format.

4.2 Experimental Environment

The operating system used for the experiments in this paper is Windows OS, the Pytorch framework in Python 3.8 environment, the graphics card model is NVIDIA RTX 3080GPU, the video memory size is 10 G and the memory size is 40 G. During the training process, the SGD optimizer is used, the learning rate is 0.01, the batch size is 8 and 120 epochs are trained. For testing, an NMS strategy with an IOU threshold of 0.5 is used.

4.3 Model Evaluation Criteria

In this paper, the performance of the instance segmentation model is evaluated using the metrics of precision (P), recall (R), average precision (AP), mean average precision (mAP), frames per second (Fps), and the reconciled mean of precision and recall (F1). $AP_{0.5}$ and $AP_{0.75}$ denote the AP metrics when the IoU threshold is taken as 0.5 and 0.75, respectively; mAP denotes the average value of the AP corresponding to the IoU threshold from 0.5 to 0.95 (with a step size of 0.05), which is computed as shown in the following equation:

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN} \quad (11)$$

$$AP = \int_0^1 p(R)dR \quad (12)$$

$$mAP = \frac{\sum_{i=0}^n AP(i)}{n} \quad (13)$$

$$F_1 = 2 \frac{P \times R}{P + R} \quad (14)$$

where TP (True positives) is the number of positive samples accurately identified as positive by the model, FP (False positives) is the number of negative samples incorrectly identified as positive, TN (True negatives) is the number of negative samples accurately identified as negative by the model, and FN (False negatives) is the number of positive samples incorrectly identified as negative by the model.

5 Experimental Comparison and Analysis

Fig. 4 shows the P-R curves of the Solov2 model before and after the improvement respectively. The area under the P-R curve of the improved Solov2 model is significantly larger than the original Solov2 model. This indicates that the improved model has higher average accuracy, higher classification performance and better robustness than the original model at each recall rate.

Fig. 5 shows the comparison of the F1 values of the Solov2 model before and after the improvement. The results show that the improved model has a significant improvement in the F1 values compared with the original model, where the F1 values of the transformer, insulator, voltage transformer and total have increased by 0.13%, 0.13%, 0.07% and 0.11%, respectively.

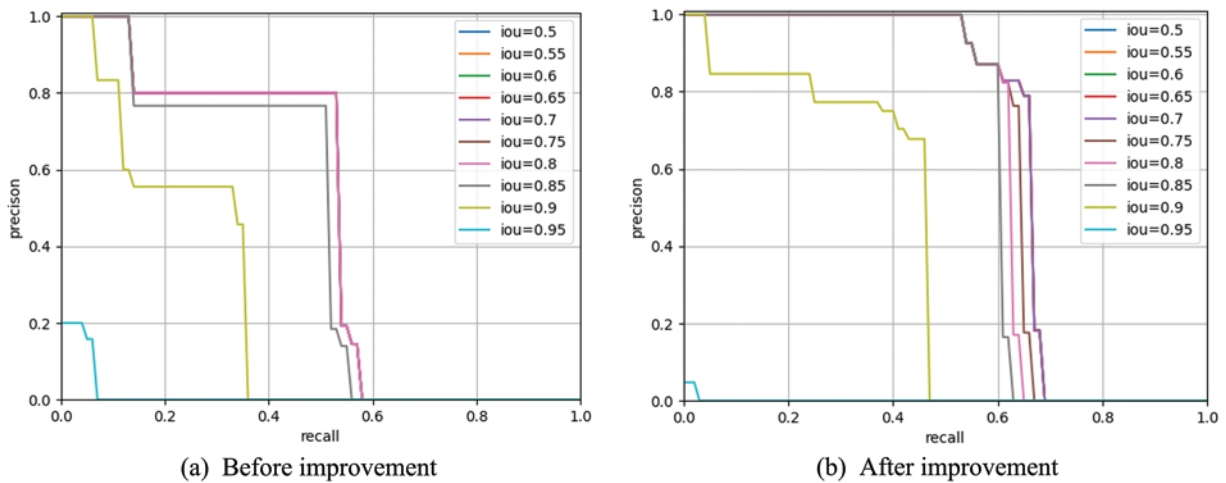


Figure 4: Comparison of P-R curves before and after improvement

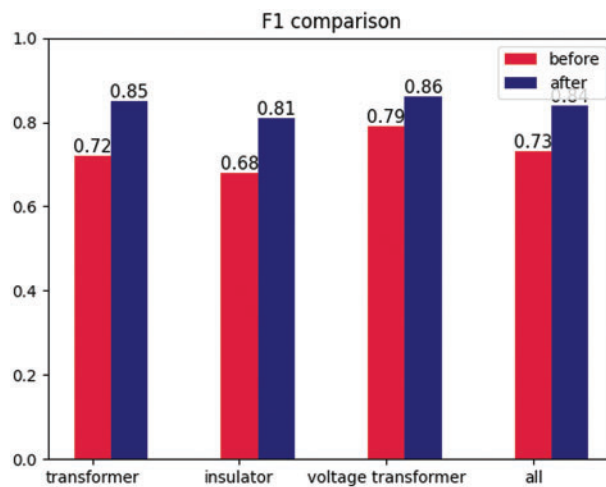


Figure 5: Comparison of F1 values before and after improvement

In the experiments in this paper, electrical equipment apparatus such as insulators, transformers and voltage mutual inductors as the test set to compare the segmentation effect of the traditional Solov2 model and the improved Solov2 model. The experimental results are shown in Fig. 6. The experimental results show that the improved Solov2 model performs more superiorly in the electrical equipment instance segmentation task. In the segmentation comparison graph, it can be seen that the improved Solov2 has significantly improved the accuracy of segmentation compared with the original Solov2, which can detect and segment the target object more accurately. The segmentation boundary of the improved Solov2 is clearer and the segmentation area is more accurate. For complex scenes and objects, the segmentation effect of the improved Solov2 is more outstanding, which can extract the boundaries and contours of the target objects better, and has higher precision and accuracy in the segmentation task.

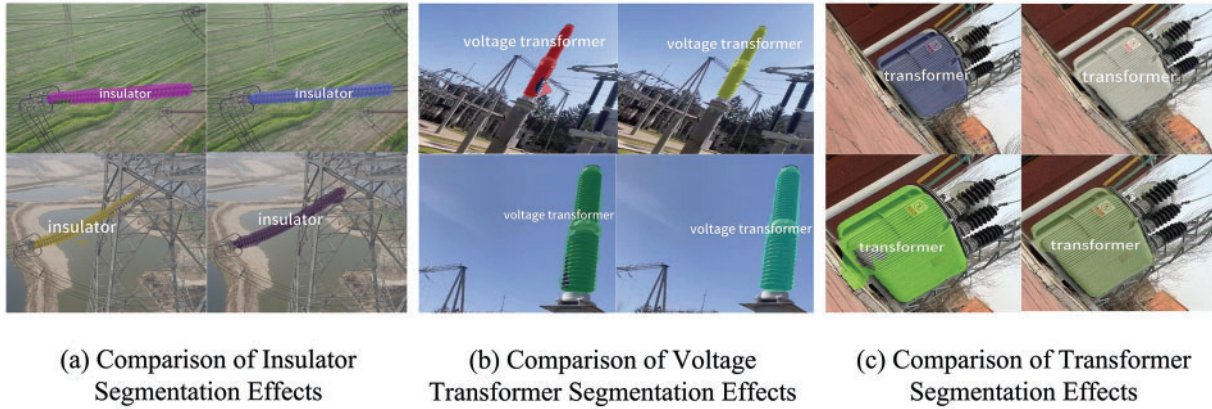


Figure 6: Effect of segmenting electrical equipment

5.1 Backbone Network Performance Comparison

In order to select the best backbone network and to improve the generalization ability of the model, a comparison is made on the segmentation ability of four different backbone networks on the dataset, and the results are shown in Table 1. The training time of the model with ResNet-50-BAM as the backbone network is 136 min and an average accuracy mean of 77.6%, which is about 30 min less than the training time of ResNet-101 and ResNet-101-BAM backbone networks, which is 2.9% higher than the average accuracy mean of ResNet-50 and 2.5% higher than the average accuracy mean of ResNet-101 backbone networks.

Table 1: Improved backbone network performance vs. original model

Backbone network	AP _{0.5} /%	AP _{0.75} /%	mAP/%	Training time/min
ResNet-50	95.6	81.9	74.7	134
ResNet-101	95.9	85.3	75.1	168
ResNet-50-BAM	96.7	90.6	77.6	136
ResNet-101-BAM	96.8	92.1	78.1	169

Therefore, considering the overall performance of the backbone network model, using ResNet-50-BAM as the model for the backbone network can provide more rapid and accurate analysis and identification of electrical equipment conditions, thus improving the efficiency of equipment fault diagnosis and maintenance. In addition, the advantages of ResNet-50-BAM in terms of training time also mean that the model can be updated more quickly to accommodate changes in electrical equipment conditions and to identify and resolve problems promptly.

5.2 Loss Function Performance Analysis

In the Solov2 model, the weighted sum of Dice Loss and CE Loss (DCE-Loss) is used as the mask loss function. The model is trained and tested on the test set by adjusting different weighting coefficients α and different thresholds, where the weighting coefficients α are taken as 0.1 to 0.9 and the thresholds are taken as 0.1 to 0.3, and the results are shown in Fig. 7. It can be seen that the segmentation accuracy of the model gradually improves the weight coefficient α increases, but starts

to decline after the weight coefficient α is greater than 0.6. When the weight coefficient α is 0.6, the highest mAP is found in the three sets of results, which are 78.9%, 79.1% and 79.2%, and this indicates that the best segmentation accuracy can be obtained by choosing a weighting factor α of 0.6 on this dataset. However, the segmentation accuracy of the model starts to drop when the weighting factor α is greater than 0.6, because choosing too high a weighting factor α causes the model to focus too much on Dice Loss and ignore CE Loss. The results of the other three groups of experiments do not show any significant difference under different thresholds, indicating that choosing different thresholds can ensure that the model can obtain better segmentation accuracy with suitable loss function weights. The results of the other three groups of experiments do not differ significantly under different thresholds. Based on the results of these experiments, the weighting factor α is 0.6 to obtain the best segmentation accuracy.

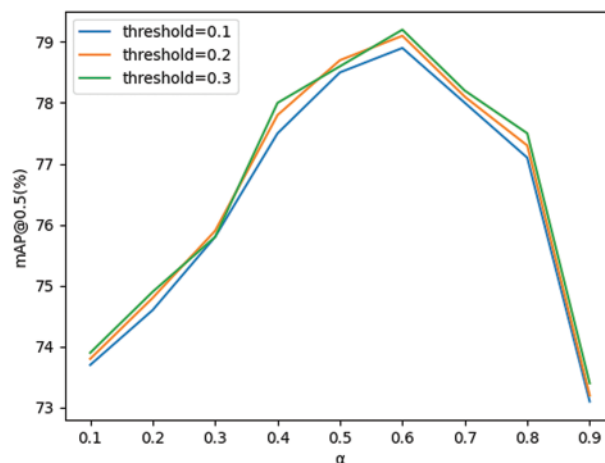


Figure 7: Accuracy of the model at different α

In the original Solov2 model, the DCE-Solov2 model with the improved mask loss function is trained separately from the original Solov2 model to obtain the total loss value change curves shown in Fig. 8. By observing the total loss value variation curve in Fig. 8, it can be seen that under the same number of training iterations, the DCE-Solov2 model with an improved mask loss function converges faster than the original Solov2 model. The DCE-Solov2 model achieves lower loss values than the original Solov2 in the early stages of training and maintains relatively low loss values throughout the subsequent training process. At 120 epochs of training, the total loss value of the DCE-Solov2 model is 0.08732, which is 0.04288 lower than that of the original Solov2, indicating that the mask loss function used in the DCE-Solov2 model can better measure the difference between the model predictions and the true labels, thus more accurately guiding the training of the model and improving the robustness and stability of the model.

The performance results of the Solov2 model using different mask loss functions are shown in Table 2. Compared to Solov2 using other mask loss functions, DCE-Solov2 improves $AP_{0.5}$ by 11.6%, 6.2%, 3.2%, and 1.8%, respectively; improves $AP_{0.75}$ by 13.8%, 16.2%, 11.0%, and 8.9%, respectively; improves mAP by 12.2%, 8.9%, 7.1%, and 4.5%, respectively. This is because DCE-Solov2 can consider the overlap between targets when calculating mask loss as well as better handle class imbalance problems, to avoid excessive focus on classes with more samples, which allows the model to better capture the details and shapes of electrical equipment. The introduction of the improved mask loss function resulted in only a slight decrease in the frame rate of the model. This indicates that while

improving the performance of the model, its speed is not significantly affected. Therefore, the DCE-Solov2 model is more suitable for electrical equipment segmentation tasks.

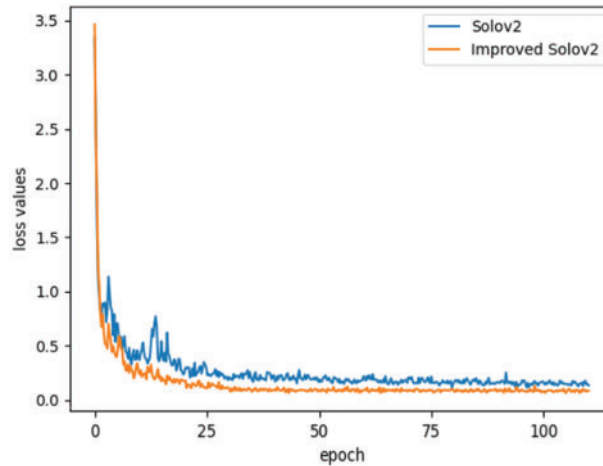


Figure 8: Comparison of loss function curves for different masks

Table 2: Performance comparison of networks using different mask loss functions

Mask loss function	AP _{0.5} /%	AP _{0.75} /%	mAP/%	Fps
MAE loss	85.8%	77.0%	67.4%	16
Lovasz loss	91.2	74.6	70.3	19
CrossEntropy loss	94.2	79.8	72.1	16
Dice loss	95.6	81.9	74.7	16
DCE-loss	97.4	90.8	79.2	15

5.3 Comparison of NMS Algorithm Results

In the original Solov2 model, different non-maximal suppression algorithms are used and then trained and tested separately, the experimental results are shown in Table 3. As can be seen from the results, the NMS algorithm used in this paper has a better segmentation effect on the electrical device, compared with Matrix NMS and Matrix GIoU-NMS, AP_{0.5} improves by 1.5% and 0.3%, respectively, AP_{0.75} improves by 4.2% and 1.8%, mAP improves by 4.1% and 1.2%, improving in terms of segmentation accuracy, segmentation stability. Because the NMS algorithm used in this paper takes into account not only the overlapping area between target frames but also the positional relationship and size differences between target frames, allowing for more accurate detection and segmentation of various components and parts in electrical equipment. Meanwhile, Matrix DIoU-NMS is more computationally complex compared with Matrix NMS, so there may be a slight decrease in computational efficiency, but this effect is negligible in comparison to the improvement in segmentation accuracy.

Table 3: Comparison of network performance before and after improving the NMS algorithm

NMS algorithm	AP _{0.5} /%	AP _{0.75} /%	mAP/%	Fps
Matrix NMS	95.6	81.9	74.7	16
Matrix GIoU-NMS	96.8	84.3	77.6	16
Matrix DIoU-NMS	97.1	86.1	78.8	15

5.4 Ablation Experiments

In order to validate the effectiveness of the Solov2 improvement method proposed in this paper for the electrical equipment segmentation task, an ablation experiment approach is used. Specifically, the improvement mechanisms are added separately in Solov2, and then the models are trained separately using the electrical equipment dataset, and tested on the electrical equipment segmentation task to verify the impact of each improvement mechanism on the overall model performance. The results are shown in [Table 4](#).

Table 4: Comparison table for ablation experiments

Serial number	ResNet-50-BAM	Matrix DIoU-NMS	DCE-loss	mAP/%
1				74.7
2	✓			77.6
3		✓		78.8
4			✓	79.2
5	✓	✓		79.4
6	✓		✓	79.7
7		✓	✓	80.1
8	✓	✓	✓	80.4

The results in [Table 4](#) demonstrate that adding the BAM module to ResNet-50 improved mAP by 2.9% compared to the original Solov2. This is because the BAM module's ability to aggregate global information enhances the feature map representation, resulting in a more accurate feature extraction of the target object. The BAM module also weakens background interference, improving detection accuracy. When the Matrix DIoU-NMS is used instead of the original NMS algorithm, mAP can be further improved by 4.1%. The improved NMS algorithm can better measure the degree of overlap between two frames, accurately selecting the best detection result and reducing the number of redundant frames. When using DCE-Loss as a mask loss function improves the mAP by 4.5%. This is because DCE-Loss helps the network to learn the contours and pixel classification of electrical devices and segment the shape and location of lines or cables more accurately. Combining these two aspects of information, the network can better understand the structure and content of the electrical devices, improving the segmentation accuracy of the network. Two different improvement mechanisms are added to the original Solov2 network, and the mAP of the network improved by 4.7%, 5.0%, and 5.4%, respectively. The experiments show that the BAM attention mechanism module, Matrix DIoU-NMS, and DCE-Loss all have different degrees of improvement in the segmentation accuracy of the

network. The final fusion of the BAM attention mechanism module, Matrix DIoU-NMS, and DCE-Loss to the network yields optimal detection results, with a 5.7% improvement in mAP compared with the original Solov2 network. The algorithm in this paper uses a combination of the BAM attention mechanism module, Matrix DIoU-NMS, and DCE-Loss. After experimental testing, the mAP metric of this algorithm reach 80.4% on the test dataset, achieving accurate segmentation of electrical devices.

5.5 Comparison of Segmentation Results of Different Models

After determining the backbone network, NMS algorithm, and improved mask loss function, the improved Solov2 model is obtained. To verify the performance of the improved Solov2 model, we compare it with other mainstream instance segmentation algorithms on the electrical equipment dataset, and the results are shown in Table 5. According to the table below, it can be seen that the improved Solov2 achieved significant improvements in $AP_{0.5}/\%$, $AP_{0.75}/\%$, and $mAP/\%$, reaching 97.5%, 91.3%, and 80.4%, respectively, which are 1.9%~10.9% higher than other algorithms in $AP_{0.5}/\%$, 9.4%~15.2% higher in $AP_{0.75}/\%$, and 5.7%~13.8% higher in mAP . At the same time, the improved Solov2 has a frame rate of 15, which is slightly lower than the original Solov2 model. This result shows that our improvement method has enhanced the detection segmentation performance of Solov2, which can provide effective support in areas such as automated detection and segmentation of electrical equipment.

Table 5: Comparison of different models for electrical equipment segmentation performance

Network model	$AP_{0.5}/\%$	$AP_{0.75}/\%$	$mAP/\%$	Fps
Mask R-CNN	93.4	80.9	73.8	7
YOLACT	86.6	76.1	66.6	28
Solo	92.9	80.7	73.5	8
Solov2	95.6	81.9	74.7	16
Improved Solov2	97.5	91.3	80.4	15

5.6 Predictive Masks Error Analysis

In order to quantitatively assess the performance of the model in mask prediction, we perform an error analysis of the model using the real mask and the predicted mask, respectively [29], and the results are shown in Table 6. According to the experimental data, the mAP obtained from the true mask test is 8.3% higher than the mAP obtained from the predictive mask test.

Table 6: Comparison of segmentation experimental results for predictive and true masks

	$AP_{0.5}/\%$	$AP_{0.75}/\%$	$mAP/\%$
Predictive mask	97.5	91.3	80.4
True mask	98.4	92.6	88.6

This difference is partly due to the model overfitting on the training set during the training process, resulting in the model with excessive attention on the limited features and patterns of the training set, and thus the poor performance on test sets. Additionally, differences in data distribution between

different training and testing sets, such as image quality, illumination conditions, and target category distributions, affect performance of the model on the test set.

The results of this test indicate that there are some performance differences in the model in terms of mask prediction, and that further research and improvements are needed to ensure higher accuracy in the instance segmentation task.

6 Conclusion

In this paper, an improved Solov2 model is proposed to address the problems of low segmentation accuracy and poor robustness when segmenting electrical equipment images. Firstly, the BAM module is added to ResNet to improve the detection capability of the model for objects of different scales; then, the weighted sum of Dice Loss and CrossEntropy Loss 6:4 is used as the mask loss to solve the problems of different target sizes and complex shapes in the electrical equipment segmentation task; finally, the improved NMS algorithm is used to enable the model to more accurately calculate the similarity between objects and improve the segmentation accuracy of the model. The experiments show that the improved Solov2 model has higher accuracy and more detailed segmentation effect on the segmentation of electrical equipment images. However, the model is computationally intensive and requires high computational resources when processing large-scale images. In the future, the processing speed and efficiency of Solov2 will be further improved so that it can be better applied to the segmentation of electrical equipment images.

Acknowledgement: We would like to thank the valuable comments and suggestions of the reviewers, which have improved the presentation.

Funding Statement: Jilin Science and Technology Development Plan Project (No. 20200403075SF), Doctoral Research Start-Up Fund of Northeast Electric Power University (No. BSJXM-2018202).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Junpeng Wu, Zhenpeng Liu; data collection: Xingfan Jiang, Xinguang Tao, Ye Zhang; analysis and interpretation of results: Junpeng Wu, Zhenpeng Liu; draft manuscript preparation: Junpeng Wu, Zhenpeng Liu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The datasets generated during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] J. Ma, K. Qian, X. Zhang and X. Ma, "Weakly supervised instance segmentation of electrical equipment based on RGB-T automatic annotation," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9720–9731, 2020.
- [2] N. Yan, T. Zhou, C. Gu, A. Jiang and W. Lu, "Instance segmentation model for substation equipment based on mask R-CNN," in *2020 Int. Conf. on Electrical Engineering and Control Technologies (CEECT)*, Melbourne, VIC, Australia, pp. 1–7, 2020.
- [3] D. Zhou and M. Chi, "Pulse-coupled neural network and its optimization for segmentation of electrical faults with infrared thermography," *Applied Soft Computing*, vol. 77, pp. 252–260, 2019.

- [4] Y. Wang, Y. Xu, T. Xu, J. Han and B. Yang, "Unmanned inspection system of substation based on multi-dimensional perception robot," in *2022 IEEE 2nd Int. Conf. on Electronic Technology, Communication and Information (ICETCI)*, Changchun, China, pp. 939–942, 2022.
- [5] X. Xu, T. Wu and L. Su, "Fault region extraction of electric equipment based on PCNN," in *Proc. of the 2019 4th Int. Conf. on Electromechanical Control Technology and Transportation (ICECTT)*, Guilin, China, pp. 45–48, 2019.
- [6] D. W. Doody, "NFPA 70E: Reducing and eliminating electrical hazards through electrical equipment design considerations," in *2013 IEEE IAS Electrical Safety Workshop*, Dallas, TX, USA, pp. 199–204, 2013.
- [7] A. Shpiganovich, A. Shpiganovich and A. Boychevskiy, "Assessment of reliability of individual units electrical equipment by characteristics power supply systems," in *2021 3rd Int. Conf. on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)*, Lipetsk, Russian Federation, pp. 1088–1090, 2021.
- [8] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 580–587, 2014.
- [9] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 3431–3440, 2015.
- [10] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [11] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [12] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, vol. 9531, pp. 234–241, 2015.
- [13] J. Dai, K. He and J. Sun, "Instance-aware semantic segmentation via multi-task network cascades," in *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 3150–3158, 2016.
- [14] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," in *2017 IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, pp. 2980–2988, 2017.
- [15] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.
- [16] H. Chen, K. Sun, Z. Tian, C. Shen, Y. Huang *et al.*, "BlendMask: Top-down meets bottom-up for instance segmentation," in *2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 8570–8578, 2020.
- [17] T. Li, J. Zhou, G. Song, Y. Wen, Y. Ye *et al.*, "Insulator infrared image segmentation algorithm based on dynamic mask and box annotation," in *2021 11th Int. Conf. on Power and Energy Systems (ICPES)*, Shanghai, China, pp. 432–435, 2021.
- [18] J. Liu, W. Xiong, L. Bai, Y. Xia, T. Huang *et al.*, "Deep instance segmentation with automotive radar detection points," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 84–94, 2023.
- [19] X. Liu, X. Zhu, M. Li, L. Wang, E. Zhu *et al.*, "Multiple kernel k -means with incomplete kernels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 5, pp. 1191–1204, 2020.
- [20] X. Yu, X. Ye and S. Zhang, "Floating pollutant image target extraction algorithm based on immune extremum region," *Digital Signal Processing*, vol. 123, no. 103442, pp. 1–14, 2022.
- [21] Z. Zhou, B. Zhang and X. Yu, "Immune coordination deep network for hand heat trace extraction," *Infrared Physics & Technology*, vol. 127, no. 104400, pp. 1–10, 2022.

- [22] E. Dong, H. Jian and J. Tong, "An improved genetic algorithm based on immune theory in target segmentation of infrared images," in *2019 IEEE Int. Conf. on Mechatronics and Automation (ICMA)*, Tianjin, China, pp. 1096–1101, 2019.
- [23] S. Yıldız, A. Memiş and S. Varlı, "Size-based adaptive instance pruning for refined segmentation of cell nuclei in histology images," in *2023 31st Signal Proc. and Communications Applications Conf. (SIU)*, Istanbul, Turkiye, pp. 1–4, 2023.
- [24] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770–778, 2016.
- [25] T. -Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan *et al.*, "Feature pyramid networks for object detection," in *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 936–944, 2017.
- [26] J. Park, S. Woo, J. Y. Lee and I. S. Kweon, "BAM: Bottleneck attention module," arXiv: 1807.06541, 2018.
- [27] X. Wang, R. Zhang, T. Kong, L. Li, C. Shen *et al.*, "Solov2: Dynamic and fast instance segmentation," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17721–17732, 2020.
- [28] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang *et al.*, "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, pp. 1486–1498, 2020.
- [29] X. Wang, T. Kong, C. Shen, Y. Jiang and L. Li, "SOLO: Segmenting objects by locations," in *European Conf. on Computer Vision*, vol. 12363, pp. 649–665, 2020.