Review of Scientific Instruments ARTICLE pubs.aip.org/aip/rsi

Infrared image target detection for substation electrical equipment based on improved faster region-based convolutional neural network algorithm

Cite as: Rev. Sci. Instrum. 95, 043702 (2024); doi: 10.1063/5.0200826
Submitted: 29 January 2024 • Accepted: 21 March 2024 •
Published Online: 10 April 2024

Changdong Wu,[a] Yanliang Wu, and Xu He

AFFILIATIONS

School of Electrical Engineering and Electronic Information, Xihua University, Chengdu 610039, China

[a]Author to whom correspondence should be addressed: ysx98042051@163.com

ABSTRACT

Substation electrical equipment generates a massive number of infrared images during operation. However, the overall quality of the infrared images is low and it lacks image detail information. When using traditional target detection algorithms for detection, feature extraction poses great difficulties. Therefore, to address this problem, this paper proposes a target detection algorithm based on the improved faster region-based convolutional neural network (Faster R-CNN). It achieves the correct identification of different types of electrical equipment in infrared images. First, the algorithm improves the backbone network of Faster R-CNN for feature extraction. An InResNet structure is proposed to replace the residual block structure of the original ResNet-34 network, which enhances the richness of feature extraction. Second, the rectified linear unit activation function in the original feature extraction network is replaced by the exponential linear unit activation function, and group normalization is used instead of batch normalization as the network normalization method. Then, the dense connection structure is introduced into the ResNet-34 network, and the whole network is called residual dense connection network. Finally, the improved Faster R-CNN is compared to the original Faster R-CNN, a single-shot multibox detector, and you only look once v3 plus spatial pyramid pooling. The experimental results show that the improved algorithm has the highest mean average precision and average recall for most of the substation electrical equipment in infrared images. Moreover, from the confidence level of the detected electrical equipment and the accuracy of the prediction box, the improved Faster R-CNN has the best performance.

I. INTRODUCTION

The reliable operation of substation electrical equipment has a direct impact on the stability of the power system. Therefore, it is important to maintain the safe and stable operation of the electrical equipment. Many types of faults in electrical equipment are manifested in the form of emitting high heat and rapid temperature rise. Infrared thermal imaging technology[1–3] is a non-contact temperature measurement technology, which can continuously generate a large number of infrared images. This enables timely and effective detection of equipment thermal faults. At present, this technology has been widely used in the online monitoring of electrical equipment in substations.[4,5] However, traditional target detection algorithms have difficulty in extracting features from infrared images. Therefore, it has become extremely important to use target detection algorithms based on deep learning to locate and identify electrical equipment in infrared images. At the same time, this lays the foundation for subsequent thermal fault diagnosis of electrical equipment.[6]

With the emergence and development of deep learning, more and more scholars apply deep learning-based target detection algorithms to the localization and identification of electrical equipment in infrared images. Compared with traditional target detection algorithms, this type of algorithm has the advantages of high detection accuracy and strong robustness. In addition, it can achieve multi-target detection. At present, target detection algorithms based on

deep learning are mainly divided into two categories. The first category is two-stage target detection algorithms. This type of algorithm is carried out in two steps. First, candidate boxes are generated using structures such as region proposal network (RPN). Then, the candidate boxes are classified and bounding box regression is performed by using a convolutional neural network. Representative algorithms include region-based convolutional neural network (R-CNN),[7] fast region-based convolutional neural network (Fast R-CNN),[8] and faster region-based convolutional neural network (Faster R-CNN).[9] The second category is one-stage target detection algorithms. This type of algorithm does not need to generate candidate boxes. It directly extracts features and identifies targets using convolutional neural networks. Representative algorithms include the single-shot multibox detector (SSD)[10] and you only look once (YOLO).[11] One-stage target detection algorithms have a simpler network structure and better real-time performance than two-stage target detection algorithms, but the latter have higher detection accuracy.

For the difference between target detection and image classification, Song and Pang[12] proposed a target detection backbone network. This target detection backbone network had good performance in terms of detection accuracy and efficiency. To solve the problem of inaccurate positioning of insulators in detection, Li *et al.*[13] proposed an insulator detection algorithm based on deep learning. The algorithm's directional detection framework with rotation angle could accurately localize insulator targets. Wu and Zuo[14] proposed a new deep convolutional network, which was used for small target detection in infrared images. The experimental results showed that this detection network outperformed many typical infrared small target detection algorithms. Fan and Chen[15] proposed a lightweight high accuracy target detection algorithm based on the YOLO framework. To further improve the detection accuracy of you only look once v3 (YOLOv3), Zhao *et al.*[16] added a convolutional layer module to the network structure of the original algorithm and roughly resized the anchor box of the feature map. The experimental results on the visual object class (VOC) dataset showed that the improved YOLOv3 algorithm had higher detection accuracy than the original algorithm. Tu *et al.*[17] improved the distinction between background and target in feature maps using a contrast learning strategy. This enhanced the target detection performance of the you only look once v5 (YOLOv5) network. The algorithm achieved better detection results in infrared images. Qiao *et al.*[18] proposed a target detection algorithm based on the improved feature extraction network. This target detection algorithm solved the problems of low target detection accuracy and inaccurate target position detection. The experimental results showed that the proposed algorithm was superior to the classical target detection algorithms. Zhang *et al.*[19] proposed a target detection algorithm based on multi-scale feature fusion and anchor box adaptation. They solved the problems of insufficient feature extraction, inaccurate detection box localization, and low detection accuracy in the Faster R-CNN algorithm. Compared with the Faster R-CNN algorithm based on ResNet-50, the overall detection results of the proposed algorithm had better performance. Xue and Wu[20] proposed an improved Faster R-CNN algorithm. They used the double-shortcut structure to enhance the feature extraction ability of the network. The experimental results showed that the improved algorithm had a high average accuracy for most substation equipment. For the complex substation background, Ou *et al.*[21] proposed an improved Faster R-CNN algorithm. The feature extraction network of this algorithm abandoned some deep convolutions and speeded up the training and testing. The algorithm had better performance in detection accuracy and speed. At the same time, the algorithm had strong noise immunity.

In the above literature, their improvements to the algorithm are mainly aimed at enhancing the feature extraction ability of the backbone network. The reason for this is that an infrared image is the pseudo-color image. It reflects the state of temperature distribution on the surface of the object. It is characterized by the low and concentrated overall gray scale distribution, low signal-to-noise ratio, and low contrast. As shown in Fig. 1, compared with the visible light image, the overall quality of the infrared image is low and lacks image detail information. However, the convolutional neural network-based target detection algorithm uses image data directly to learn the feature representation of the image. The lack of image quality and the lack of detail information will increase the difficulty of CNN in extracting features. Therefore, in order to enhance the feature extraction ability of the network, this paper proposes a target detection algorithm based on the improved Faster R-CNN. Comparing the proposed algorithm with other algorithms, the results show that the proposed algorithm has the highest detection accuracy and recall. In addition, the performance of the proposed algorithm is the best in terms of the confidence level on the detected electrical devices and the accuracy of the prediction boxes.

II. FASTER R-CNN TARGET DETECTION MODEL

Faster R-CNN is a typical representative of target detection algorithms based on region suggestions. It is improved on the basis of Fast R-CNN and introduces region proposal network (RPN) to generate target candidate regions exclusively. This improves the detection efficiency. Faster R-CNN mainly consists of feature extraction networks, region proposal networks, and target detection networks. The structure of the target detection model based on Faster R-CNN is shown in Fig. 2.

A. Feature extraction network

Currently, the main convolutional neural networks used for feature extraction are Alexnet,[22] Googlenet,[23] VGG,[24] and ResNet.[25] Usually, the deeper the network, the better the fitting ability. However, as the number of layers of the convolutional neural network increases, gradient vanishing and gradient exploding may occur. In response to the degradation phenomenon, the shortcut connection structure of the ResNet network is a good solution to this problem. Meanwhile, the ResNet network can greatly accelerate the training speed of the network. Therefore, the ResNet network is chosen as the convolutional backbone network for feature extraction. The overall structure of ResNet-34 is shown in Fig. 3. The first building layer consists of one ordinary convolutional layer. The second building layer consists of one maximum pooling layer and two residual blocks. The third, fourth, and fifth building layers all start with a downsampling residual block. They are followed by three, five, and two residual blocks connected to them, respectively. The remaining individual layers are shown in Fig. 3. The specific structures of the residual block and downsampling residual block are shown in Figs. 4

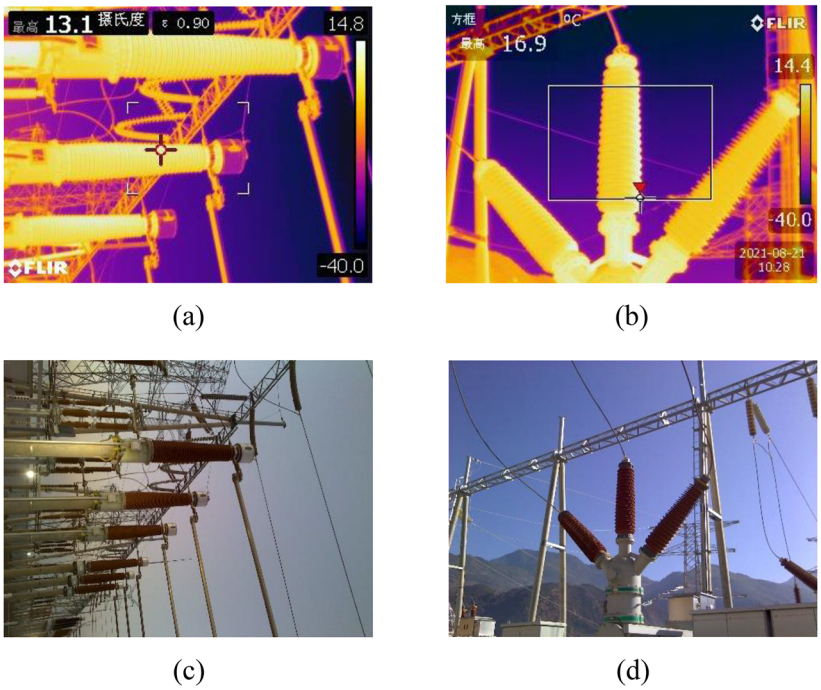

FIG. 1. Comparison of the infrared image and visible light image of electrical equipment. (a) and (b) Infrared image of electrical equipment. (c) and (d) Visible light image of electrical equipment.

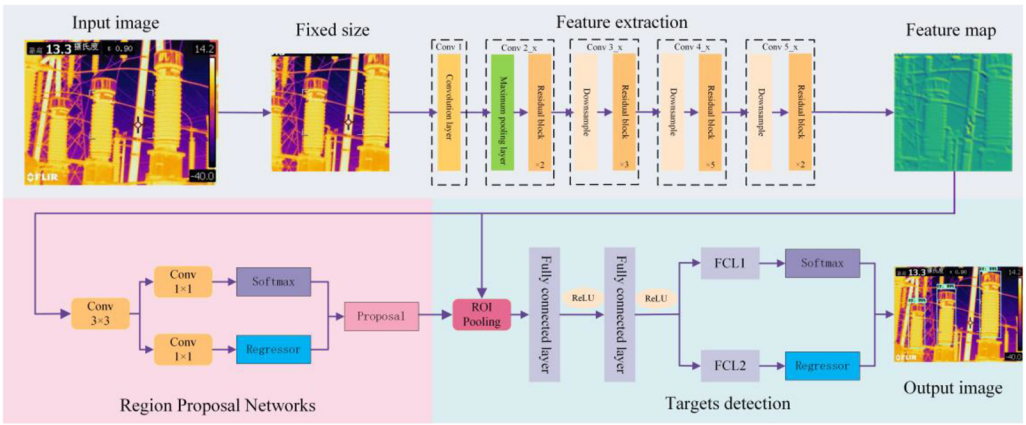

FIG. 2. Structure of the target detection model based on Faster R-CNN.

and 5, respectively. The purpose of using the downsampling residual block is to reduce the feature map size and increase the number of channels.

B. Region proposal networks

RPN is the key structure where the detection performance of Faster R-CNN is significantly better than Fast R-CNN. RPN takes the feature map generated by the convolutional backbone network as input. It also outputs a large number of rectangular boxes with scores. The structure of RPN is shown in Fig. 6. First, a 3 × 3 sliding window is used to perform convolution on the feature map. This generates k anchor boxes at the center of the sliding window. Then, through the classification layer and the regression layer, anchor boxes are classified and bounding box regression is performed. The anchor boxes scores and parameters are also output. The classification layer classifies whether each anchor box is a foreground target

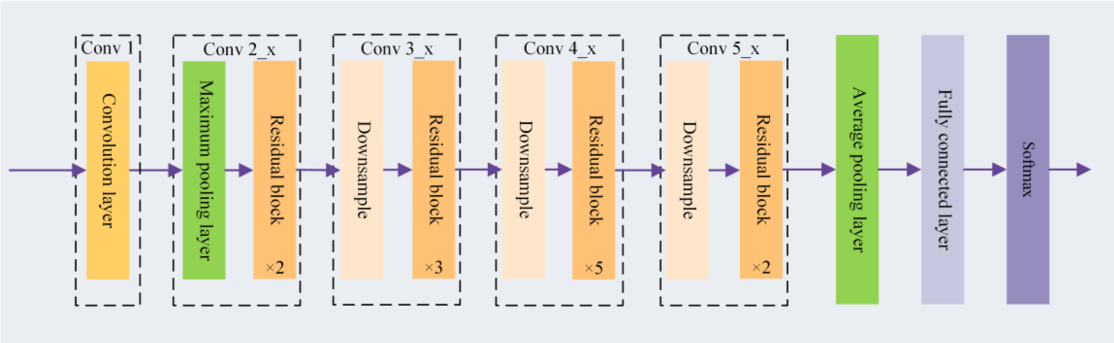

FIG. 3. ResNet-34 network structure.

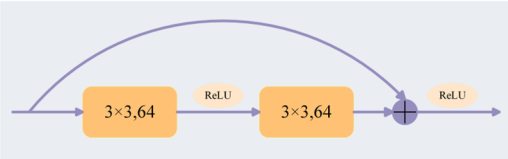

FIG. 4. Residual block.

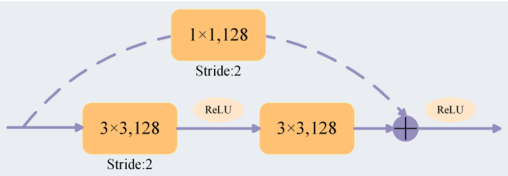

FIG. 5. Downsampling residual block.

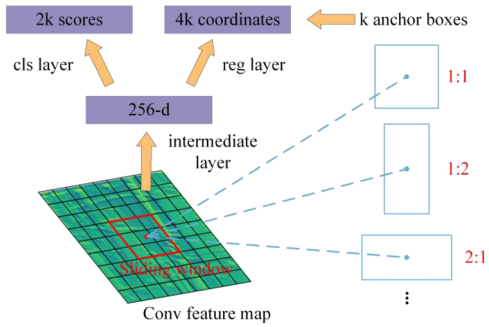

FIG. 6. Structure of RPN.

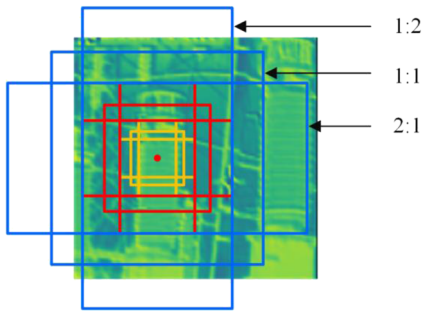

FIG. 7. Nine different anchor boxes.

or background. Hence, a total of 2k scores are output. These scores represent the probability that each anchor box is a target or background. The regression layer then regresses the position of each anchor box. The 4k parameters are the position coordinates of k anchor boxes. Then, it is subjected to the non-maximum suppression (NMS) method. At the same time, the threshold of intersection over union (IoU) is set to 0.7. Finally, only anchor boxes with coverage exceeding 0.7 are retained. After this treatment, the number of anchor boxes is less than one tenth of the original number, and the top N anchor boxes are chosen as candidate boxes to input into the Region of Interest (ROI) pooling layer.

The shape of substation electrical equipment in an infrared image is usually rectangular. When using image annotation software to label infrared images, it is found that the width and height of the labeled rectangular boxes are between tens and two hundred. Therefore, in this paper, the three sizes of the anchor box are set to [64 × 64], [128 × 128], and [256 × 256]. The three aspect ratios are set to [1:2, 1:1, 2:1]. Nine anchor boxes are generated in the center of each sliding window to detect all targets in the image. Adjusting the size of the anchor box makes it possible to more accurately locate substation electrical equipment in the infrared image. Figure 7 shows nine anchor boxes with different sizes and aspect ratios.

C. Target detection network

The target detection network mainly consists of a ROI Pooling layer, a fully connected layer, a classification layer, and a bounding box regression layer. The ROI pooling layer maps candidate regions of different sizes generated by RPN to fixed size output feature maps. Then, the Softmax classifier is used to calculate which

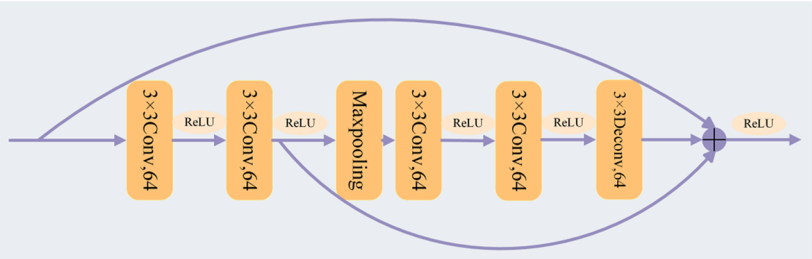

FIG. 8. InResNet structure.

specific category each proposal belongs to. Meanwhile, the bounding box regression layer is used to obtain the position offset for each proposal. This can obtain more accurate object detection boxes.

III. IMPROVED FASTER R-CNN TARGET DETECTION MODEL

In the practical application, it requires the accurate localization and identification of substation electrical equipment in an infrared image. However, using Faster R-CNN based on ResNet-34 for locating and identifying electrical equipment has problems such as low confidence in the equipment, missed detection of equipment, and inaccurate positioning. Hence, it should improve ResNet-34 to solve the above problems. In this paper, the backbone network is improved in the following three aspects.

A. InResNet structure

The main role of the backbone network is to extract features from the input infrared image. When it uses ResNet-34 as the backbone network of Faster R-CNN, it has the insufficient feature extraction ability. Therefore, in order to enhance its feature extraction ability, this paper improves the residual block. The improved residual block is shown in Fig. 8. The improved residual block is called InResNet structure.

As can be seen from Fig. 8, the InResNet structure adds a maxpooling layer, convolutional layers, and an anti-convolutional layer than the original residual block structure. The improved residual block can extract deep features by deepening the depth of the original residual block. The improved residual block has both shallow and deep features. In addition, then, the shallow and deep features are fused to increase the richness of feature extraction. At the same time, this structure is used for the entire ResNet-34 network. The convolution operation is performed again on the feature map that combines shallow and deep features. This greatly enhances the feature extraction ability.

B. Activation function and normalization

1. ELU activation function

The activation function used in the original ResNet-34 network is a rectified linear unit (ReLU) activation function.[26] The ReLU formula is given as follows:

$$f(x) = \begin{cases} x, & x > 0, \\ 0, & x \le 0. \end{cases} \tag{1}$$

Compared to the sigmoid and tanh activation functions, the ReLU activation function can achieve unilateral suppression. It is able to sparse the model. There is no problem of gradient saturating and gradient vanishing on the x > 0 region. Its computational complexity is low and convergence speed is fast. Although the ReLU activation function has many advantages, there are still shortcomings. When the input is close to zero or negative, the gradient of the function becomes zero. The network does not perform backpropagation and learning. To avoid issues with the ReLU activation function, the exponential linear unit (ELU) activation function[27] is used instead of the ReLU activation function. The ELU activation function is an exponential linear unit. The ELU formula is given as follows:

$$f(x) = \begin{cases} x, & x > 0, \\ \alpha\left(e^{x} - 1\right), & x \le 0, \end{cases} \tag{2}$$

where α is the hyperparameter, and it is set to 1.

The ELU activation function solves the problems with the ReLU activation function. As shown in Fig. 9, it corrects the negative part of the ReLU activation function with a corresponding exponential correction. This reduces the gap between the gradients so that the gradient of the function does not become zero when the input value is negative. The network continues to backpropagate and learn. It has the advantage of avoiding "neuron death" and has all the advantages of the ReLU activation function. At the same time, the average value of the output is close to zero, which can speed up the convergence of the network.

2. Group normalization

In a residual network, data are typically normalized using batch normalization (BN).[28] A larger learning rate can be used when BN is employed. The training process is more stable, and the training speed is greatly improved. At the same time, BN has a kind of regularization effect, which can reduce overfitting. However, BN is very much affected by the batch size. BN needs a sufficiently large batch. In addition, too small batch lead to higher inaccuracy in the batch statistics. This significantly increases the error rate of the model. Since our device cannot set a larger batch size, this paper uses group normalization (GN)[29] instead of BN as the network normalization method to eliminate the effect of batch size. As shown in Fig. 10,

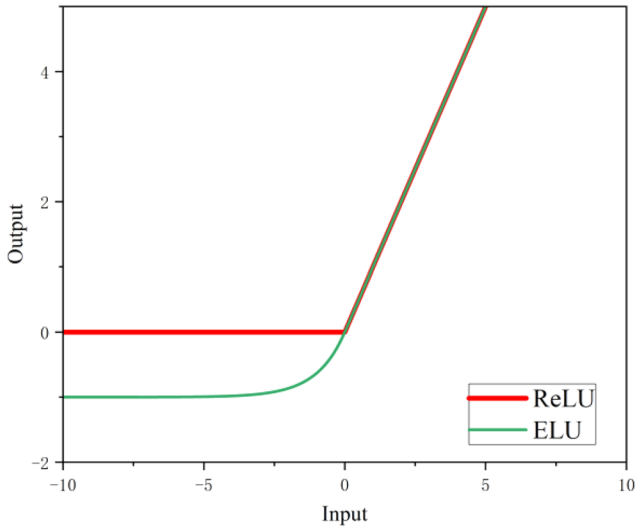

FIG. 9. Activation function curve.

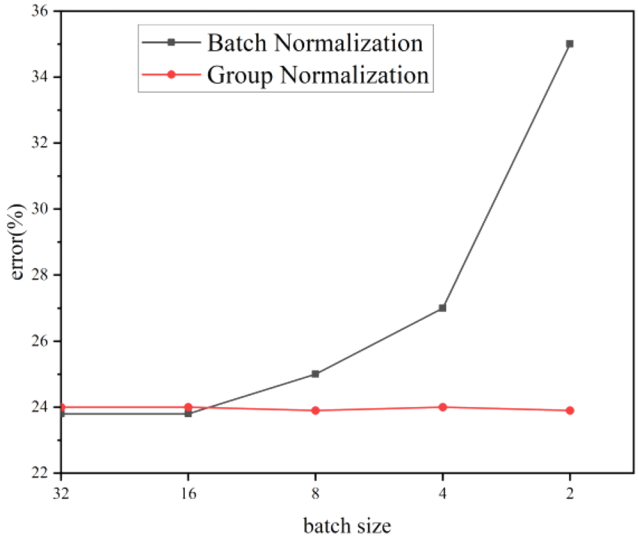

FIG. 10. Comparison of error rates between GN and BN on different batches.

GN is not affected by the batch size. In addition, GN performs better than BN when the batch is very small.

C. Residual dense connection network

In the neural network structure, different levels of the network have different sensory domains and semantic levels. When dealing with the scale of an image, the deepening of the network has resulted in retaining more semantic information in the feature map while losing some detailed features. To further enhance the feature extraction ability of the backbone network, a dense connection structure is introduced. The dense connection structure[30] is one that interconnects all the layers. Specifically, each layer accepts all the layers in front of it as its additional inputs. In this paper, the dense connection structure is introduced into the backbone network to form a residual dense connection network. Its structure is shown in Fig. 11. The purpose of this network is to fuse the feature maps at various scales and enhance the feature extraction ability of the whole backbone network.

The residual dense connection network consists of two parts: a residual network consisting of the InResNet structure and a dense connection structure. The movement of feature information in each layer is enhanced without extending the depth of the network. In addition, the representation of detailed information in the shallow layer is improved. The residual dense connection network takes the shallow feature maps as input. In addition, it fuses the features extracted from different depths. It makes the multi-scale context information richer and further enhances the feature extraction ability of the backbone network.

Figure 12 shows a comparison of the feature maps extracted from the Conv2_4 convolutional layer of the original ResNet-34 network and the improved ResNet-34 network. From Fig. 12, it can be seen that the improved feature maps are more capable of displaying the characteristic information of the electrical equipment. It is more conducive to the localization and identification of electrical equipment.

The following conclusions can be drawn from Sec. III:

(1) Using the InResNet structure increases the richness of feature extraction.
(2) Using the ELU activation function and GN, the network convergence speed can be accelerated and device requirements can be reduced.
(3) The dense connection structure enhances the multi-scale fusion of feature maps and further enhances the feature extraction ability of the network.

For the convenience of later description, the network with the improved residual block structure is referred to as ResNet-34+In. The network using GN and ELU activation functions on top of ResNet-34+In is referred to as ResNet-34+In+GE. The network that introduces a dense connection structure to the ResNet-34+In+GE network is referred to as ResNet-34+In+GE+C. When using a series of ResNet networks as the backbone of a Faster R-CNN, the Faster R-CNN is abbreviated as the ResNet network name.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Dataset creation

In this paper, these infrared images are 320 × 640 and 240 × 320 pixels, JPEG format. The infrared images and names of the six electrical equipment to be identified are shown in Fig. 13.

The amount of infrared image data in this paper is relatively small. However, the deep learning network structure often has a large number of parameters. A large amount of data is needed to learn the parameters in the network to avoid overfitting in the trained model. Therefore, it may use dataset augmentation methods to prevent overfitting of the network. Common data augmentation methods include geometric operations, color transformations, random erasure, and noise addition. In this paper, 87 infrared images are first randomly selected as the test sample set. Then, the dataset

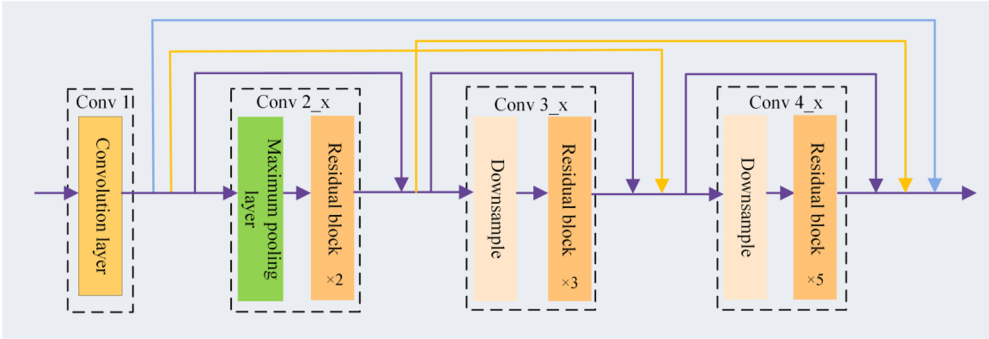

FIG. 11. Residual dense connection network.

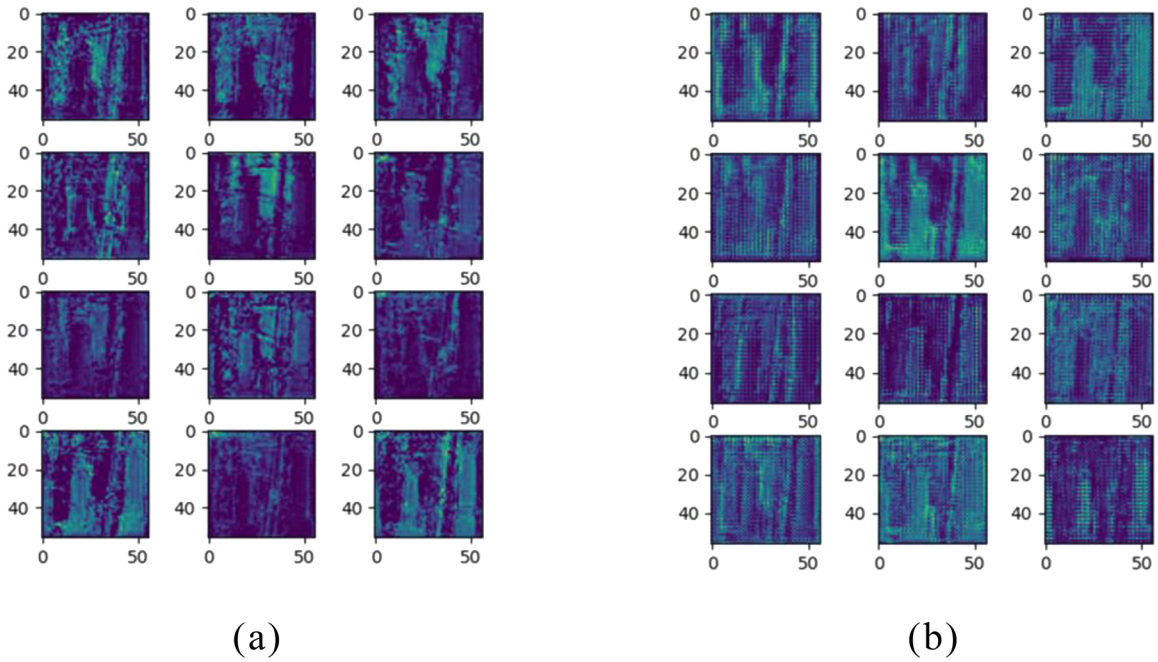

FIG. 12. Comparison of feature maps before and after network improvement. (a) Feature maps of Conv2_4 for the pre-improvement network. (b) Feature maps of Conv2_4 for the improved network.

is increased by performing data augmentation operations, such as rotation, flipping, histogram equalization, and white noise on all remaining infrared images. A total of 2167 infrared images are obtained after the data augmentation methods. In addition, we randomly divide them into training and validation sample sets in the ratio of 8:2. Finally, all the infrared images in the training set and validation sample set are labeled using LabelImg software. Since the target detection algorithm used is Faster R-CNN, the format of the label is XML. The number of labels for six types of electrical equipment is shown in Table I.

B. Evaluation metrics

There are many metrics in the field of target detection that can be used to evaluate the performance of a target detection model. The evaluation metrics we usually use include average precision (AP), mean average precision (mAP), and average recall (AR). These evaluation metrics are mainly defined by the following four parameters.

- True positive (TP): Predicted as a positive sample, actual as a positive sample.

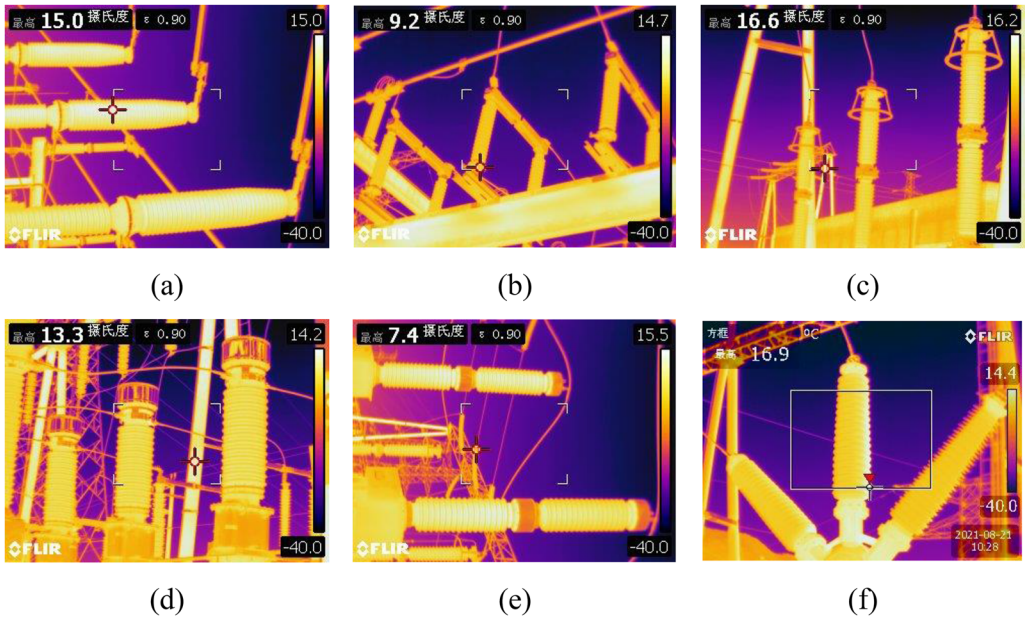

FIG. 13. Infrared images of six types of electrical equipment. (a) Circuit breaker (QF). (b) Disconnect switches (QSs). (c) Lightning arrester (LA). (d) Current transformer (CT). (e) Potential transformer (PT). (f) Bushing (BS).

TABLE I. Number of labels for six types of electrical equipment.

Label name	QF	QS	LA	CT	PT	BS
Number of labels	933	1075	1104	1214	314	543

- False positive (FP): Predicted as a positive sample, actual as a negative sample.
- True negative (TN): Predicted as a negative sample, actual as a negative sample.
- False negative (FN): Predicted as a negative sample, actual as a positive sample.

The precision rate is an effective evaluation metric for the predicted results. It represents the proportion of correctly predicted samples out of the predicted positive samples. The precision formula is given as follows:

$$P = \frac{TP}{TP + FP}. \tag{3}$$

The recall rate is an evaluation metric for the original sample. It represents the proportion of predicted positive samples out of the actual positive samples. The recall formula is given as follows:

$$R = \frac{TP}{TP + FN}. \tag{4}$$

The AP and mAP are shown in Eqs. (5) and (6), respectively,

$$AP = \int_0^1 P(R)\,\mathrm{d}R, \tag{5}$$

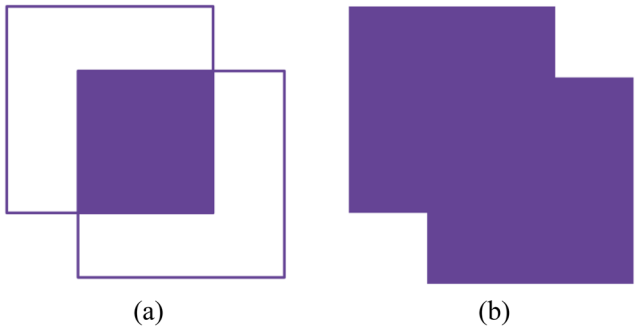

FIG. 14. Intersection and union. (a) Area of intersection. (b) Area of union.

$$mAP = \frac{\Sigma AP}{N}, \tag{6}$$

where N is the number of categories.

In the target detection task, whether a sample is positive or negative needs to be determined using intersection over union (IoU). IoU represents the ratio of the intersecting part between the predicted box and the actual box to the merging part. The IoU formula is given as follows:

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union}, \tag{7}$$

where the Area of Intersection and Area of Union represent the portion, as shown in Fig. 14.

TABLE II. Comparison of mAP and AR for different networks.

Network name	mAP@0.5	AR@0.5:0.95 (medium)	AR@0.5:0.95 (large)
ResNet-34	0.8460	0.6786	0.7472
ResNet-34+In	0.8716	0.6851	0.7730
ResNet-34+In+GE	0.8887	0.6985	0.7840
ResNet-34+In+GE+C	0.9029	0.7126	0.7925

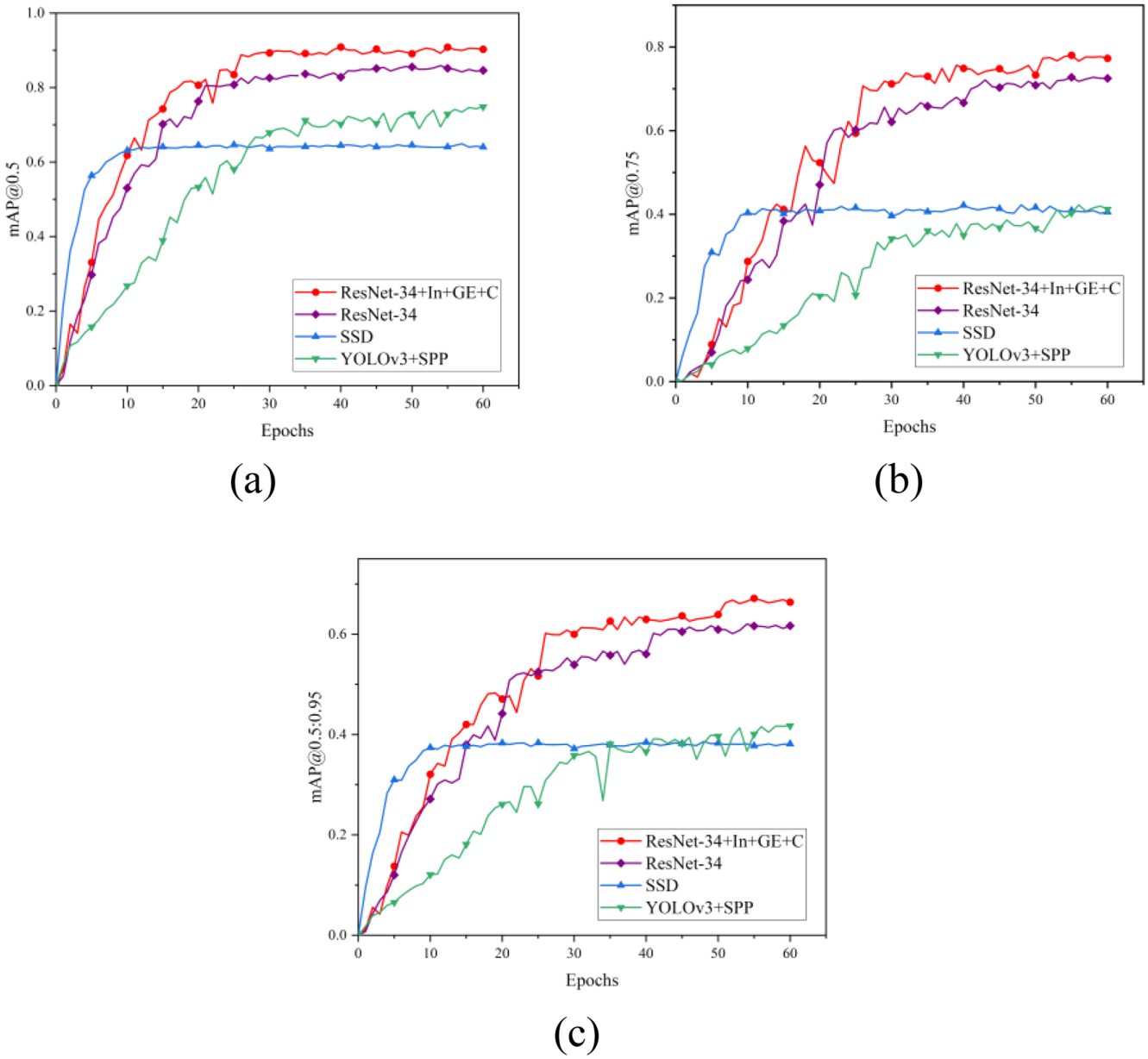

FIG. 15. mAP of the four algorithms for different IoU thresholds. (a) mAP@0.5. (b) mAP@0.75. (c) mAP@0.5:0.95.

TABLE III. Comparison of electrical equipment detection results under different algorithms. Boldface denotes the maximum value of AP for different electrical equipment in different algorithms.

	AP@0.5 (%)						
Model	QF	QS	LA	CT	PT	BS	mAP@0.5 (%)
ResNet-34	88.45	**95.88**	77.04	89.57	84.97	71.69	84.60
SSD	74.11	84.58	47.60	72.47	64.21	41.23	64.03
YOLOv3+SPP	87.10	79.41	68.84	80.87	74.14	58.68	74.84
ResNet-34+In+GE+C	**95.36**	95.79	**87.22**	**93.95**	**90.56**	**78.86**	**90.29**

In this paper, AP represents the average accuracy of each type of electrical equipment. mAP represents the mean of the average accuracy of six different electrical equipment. In COCO metrics, mAP is often used in combination with IoU. For example, mAP@0.5 indicates the mAP when the IoU threshold is set to 0.5. In general, mAP represents the accuracy of the algorithm to detect the target. The higher the mAP, the higher the accuracy of the target type. AR represents the recall rate of the target detected by the algorithm. The higher the AR, the lower the missing rate and the stronger the ability of successfully detecting the target.

C. Experimental platform

The experimental platform is Windows 11 64-bit system. Python 3.8 is chosen as the programming language. The hardware conation for the experimental part includes Gen Intel(R) Core (TM) i5-1135G7 CPU 2.40 GHz, 43 GB of RAM, and a graphics card with 24 GB of video memory RTX3090. The graphics processing unit (GPU) is utilized for acceleration of the training and testing process. A learning rate decay training method is used for network training. The initial learning rate is set to 0.01 and decays by a factor of 0.5 every 20 epochs. Momentum is set to 0.9, and the size of batch size is set to 4. A total of 60 epochs are trained. The optimizer uses stochastic gradient descent (SGD) algorithm.

D. Experimental results

The ResNet-34, ResNet-34+In, ResNet-34+In+GE, and ResNet-34+In+GE+C networks are trained using the above experimental environment and training methods. Their training results are shown in Table II.

As can be seen from Table II, the mAP of the improved network is larger than the mAP of ResNet-34. Through continuous improvement of ResNet-34, its corresponding mAP is also gradually increasing. It shows that the improvement of ResNet-34 enhances its feature extraction ability. This makes the proposed algorithm more accurate in detecting electrical equipment in an infrared image.

In the actual shooting scene, most of the substation electrical equipment in the infrared image belongs to large- and medium-sized targets. Table II shows AR@0.5:0.95 of medium targets (targets with a pixel size of more than 32 × 32 but less than 96 × 96 in the image defined in the COCO evaluation metric) and large targets (targets with a pixel size of more than 96 × 96) in different networks. AR@0.5:0.95 is the average AR over the IoU from 0.5 to 0.95 (in steps of 0.05). The higher the AR value, the lower the miss detection probability. As can be seen from Table II, with the continuous improvement of ResNet-34, AR is constantly increasing, while mAP is also constantly improving. It shows that the improved network can detect more electrical equipment while maintaining higher accuracy. It fully shows that the performance of the improved network is better.

In order to effectively evaluate the proposed algorithm, it is compared with other target detection algorithms, such as Faster R-CNN with ResNet-34, SSD, and you only look once v3 plus spatial pyramid pooling (YOLOv3+SPP). Due to the poor performance of training SSD and YOLOv3+SPP directly, we use a migration learning approach to train these two target detection algorithms. The mAP for training these four different types of target detection algorithms is shown in Fig. 15.

As can be seen from Fig. 15, the proposed algorithm has the highest mAP. It is 5.69% higher than ResNet-34. It is also much higher than SSD and YOLOv3+SPP. Moreover, even on the more stringent requirements of mAP@0.75 and mAP@0.5:0.95, the mAP of the proposed algorithm is the highest. The results show that regardless of the IoU threshold, the proposed algorithm has the highest mAP.

Table III compares the detection results of electrical equipment under different algorithms. As can be seen from Table III, the proposed algorithm has the highest AP@0.5 (%) on all electrical equipment except QS equipment. The highest AP@0.5 (%) of QS equipment is ResNet-34. However, it is only 0.09% higher than our proposed algorithm. The results show that the proposed algorithm has the highest AP and mAP for most of substation electrical equipment.

In the field of target detection, it requires high detection accuracy and high recall is required. Hence, we compare the AR of different target detection algorithms for medium and large targets. The results are shown in Fig. 16.

As can be seen from Fig. 16, the value of AR@0.5:0.95 on the medium target of the proposed algorithm is 71.26%. It is 3.40% higher than ResNet-34 and much higher than AR@0.5:0.95 for the other two algorithms. The value of AR@0.5:0.95 on the large target of the proposed algorithm is 79.25%. In addition, AR@0.5:0.95 of ResNet-34 is 74.72%. The proposed algorithm AR@0.5:0.95 is 4.53% higher than ResNet-34 and much higher than SSD and YOLOv3+SPP. Therefore, the proposed algorithm can detect more electrical equipment.

To further evaluate the proposed algorithm, two infrared images from the test-set are randomly selected and put into the four

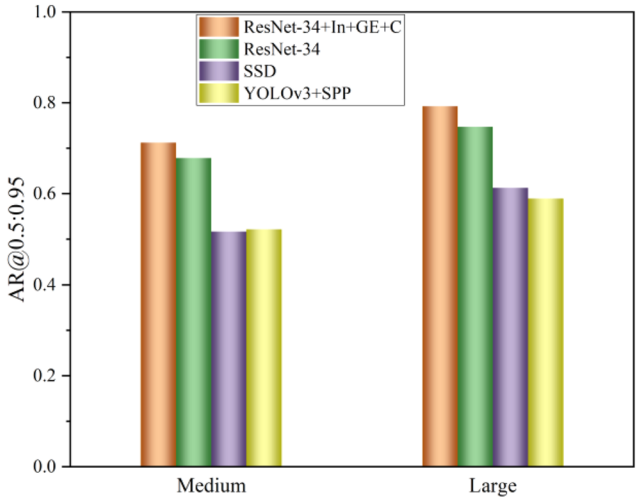

FIG. 16. AR of four target detection algorithms on medium and large targets.

trained target detection algorithms for prediction. The predicted results are shown in Figs. 17 and 18.

Figure 17 shows the detection results of a randomly selected infrared image. The category confidence indicates the probability that the algorithm detects a target in an image belonging to a certain category. Three QS equipment are detected in Fig. 17(a) with confidence levels of 99%, 99%, and 98%, respectively. Compared with Fig. 17(a), the confidence level of the QS equipment detected in Fig. 17(b) is slightly lower. Compared with Fig. 17(a), there is a situation of missed detections in Fig. 17(c). The confidence level of the QS equipment is 14% lower than the lowest confidence level detected in Fig. 17(a). In addition, its prediction box fails to accurately frame the size of the QS equipment. Figure 17(d) detects two QS equipment but also has one missed detection. Meanwhile, the confidence level of its detected QS equipment is much lower than that of the detected QS equipment in Fig. 17(a). Overall, the proposed algorithm has better performance. It can accurately detect more electrical equipment.

Figure 18(a) shows the detection results using the proposed algorithm. It can be seen that QS and QF equipment are successfully detected. The minimum confidence levels for the two types of electrical equipment are 99% and 98%, respectively. Figure 18(b) shows the detection results of ResNet-34. Compared with Fig. 18(a), it misses QF equipment and its confidence level of detected QS equipment is slightly lower. Figure 18(c) shows the detection results of SSD. There are missed detections for both QS and QF equipment, and the confidence level of its detected QS equipment is much lower than that of Fig. 18(a). Figure 18(d) shows the detection results of YOLOv3+SPP. Although two types of electrical devices are successfully detected, there is missed detection of QF equipment, and the confidence level for the detected QS equipment is very low. Compared to other algorithms, the proposed algorithm has good performance. Figures 17 and 18 show that the proposed algorithm can accurately detect electrical equipment while ensuring that electrical equipment is not missed. In terms of confidence and accuracy

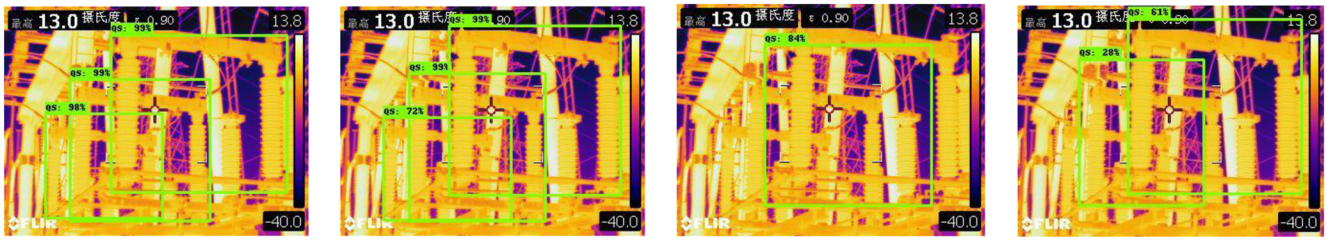

FIG. 17. Detection results of the four algorithms. (a) The proposed algorithm. (b) ResNet-34. (c) SSD. (d) YOLOv3+SPP.

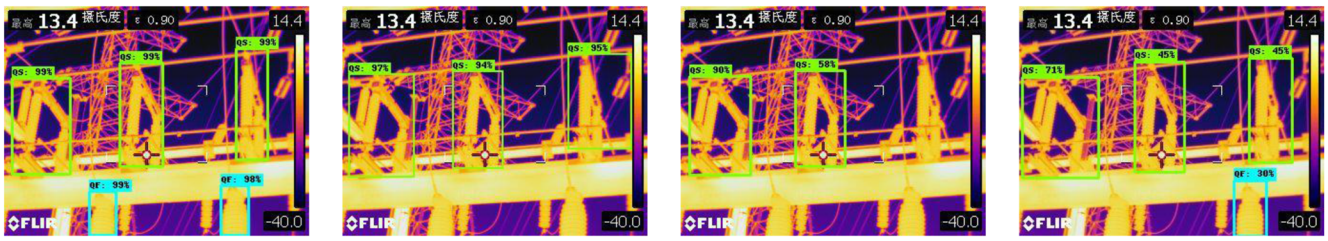

FIG. 18. Detection results of the four algorithms. (a) The proposed algorithm. (b) ResNet-34. (c) SSD. (d) YOLOv3+SPP.

of prediction boxes, the performance of the proposed algorithm is the best.

Moreover, it can obtain the following conclusions.

(1) The comparison of mAP and AR for different networks shows that the improved network can detect more electrical equipment while maintaining higher accuracy.
(2) The comparison of mAP, AP, and AR for four target detection algorithms shows that the proposed algorithm has the highest mAP, AP, and AR.
(3) The detection results of the four algorithms show that the proposed algorithm can accurately detect electrical equipment while ensuring that electrical equipment is not missed. In addition, in terms of confidence and accuracy of prediction boxes, the performance of the proposed algorithm is the best.

V. CONCLUSION

Aiming at the problem of traditional target detection algorithms in extracting features from an infrared image, this paper proposes a target detection algorithm based on the improved Faster R-CNN. In this paper, the improvement of the backbone network of Faster R-CNN is as follows:

(1) Use the InResNet structure as the residual block structure for the ResNet-34 network.
(2) Replace the ReLU activation function and BN with the ELU activation function and GN, respectively.
(3) Introduce the dense connection structure in the ResNet-34 network to form the residual dense connection network.

The values of mAP@0.5, AR@0.5:0.95 (medium), and AR@0.5:0.95 (large) have been improved by improving the ResNet-34 network. They increase by 5.69%, 3.4%, and 4.53%, respectively. This indicates that the ResNet-34+In+GE+C network has a stronger feature extraction ability.

The improved Faster R-CNN is compared and analyzed with other target detection algorithms, such as Faster R-CNN with ResNet-34, SSD, and YOLOv3+SPP. The following conclusions are obtained:

(1) The value of mAP@0.5 of the proposed algorithm is 90.29%, which is 5.69%, 26.26% and 15.45% higher than that of Faster R-CNN with ResNet-34, SSD, and YOLOv3+SPP, respectively.
(2) AP@0.5 for QF, QS, LA, CT, PT, and BS using the proposed algorithm is 95.36%, 95.79%, 87.22%, 93.95%, 90.56%, and 78.86%, respectively.
(3) The proposed algorithm is the best in terms of confidence as well as prediction boxes compared to the other algorithms.

The improved Faster R-CNN can better identify and locate substation electrical equipment, which lays the foundation for thermal fault diagnosis of electrical equipment. In future work, we will aim to improve the algorithm performance by reducing the backbone network parameters while maintaining its high detection accuracy, thus improving the detection efficiency.

ACKNOWLEDGMENTS

This paper was supported by the Xihua University Talent Introduction Project (Grant No. Z222014), the 2023 Higher Education Science Research Plan Project of the Chinese Higher Education Association (Grant No. 23XJH0103), the Quality Project of Graduate Education, Xihua University (Grant No. YJG202308), the Image Detection Technology for Power Equipment (Grant No. H242065), and the Research Project on Education and Teaching Reform, Xihua University (Grant No. Xjjg2023005).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Changdong Wu: Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Resources (equal); Writing – review & editing (equal). **Yanliang Wu**: Data curation (equal); Resources (equal); Software (equal); Writing – original draft (equal). **Xu He**: Investigation (equal); Methodology (equal); Supervision (equal).

DATA AVAILABILITY

The partial data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]R. A. Epperly, G. E. Heberlein, and L. G. Eads, "Thermography, a tool for reliability and safety," IEEE Ind. Appl. Mag. **5**(1), 28–36 (1999).
[2]M. S. Jadin and S. Taib, "Recent progress in diagnosing the reliability of electrical equipment by using infrared thermography," Infrared Phys. Technol. **55**(4), 236–245 (2012).
[3]S. Bagavathiappan, B. B. Lahiri, T. Saravanan, J. Philip, and T. Jayakumar, "Infrared thermography for condition monitoring—A review," Infrared Phys. Technol. **60**, 35–55 (2013).
[4]S. Han, R. Hao, and J. Lee, "Inspection of insulators on high-voltage power transmission lines," IEEE Trans. Power Delivery **24**(4), 2319–2327 (2009).
[5]J. Wang, X. Xiao, Y. Fan, L. Cai, Y. Tong, Z. Rao, and Z. Huang, "Interface defect detection for composite insulators based on infrared thermography axial temperature method," Infrared Phys. Technol. **93**, 232–239 (2018).
[6]C. L. Wen and F. Y. Lü, "Review on deep learning based fault diagnosis," J. Electron. Inf. Technol. **42**(1), 234–248 (2020).
[7]R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, United States, 23–28 June 2014* (IEEE Computer Society, 2014), pp. 580–587.
[8]R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision* (ICCV, 2015), pp. 1440–1448.
[9]S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017).
[10]W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *14th European Conference on Computer Vision, ECCV 2016, Amsterdam, Netherlands, 8–16 October 2016*, *LNCS Vol. 9905* (Springer Verlag, 2016), pp. 21–37.

[11]J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *29th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, United States, 26 June to 1 July 2016* (IEEE Computer Society, 2016), pp. 779–788.
[12]Y. L. Song and Y. W. Pang, "Backbone network for object detection task," Laser Optoelectron. Prog. **57**(4), 041021 (2020).
[13]C. L. Li, Q. H. Zhang, W. H. Chen, X. B. Jiang, B. Yuan, and C. L. Yang, "Insulator orientation detection based on deep learning," J. Electron. Inf. Technol. **42**(4), 1033–1040 (2020).
[14]S. C. Wu and Z. R. Zuo, "Small target detection in infrared images using deep convolutional neural networks," J Infrared Millimeter Waves **38**(3), 371–380 (2019).
[15]X. C. Fan and C. M. Chen, "Lightweight and high-precision object detection algorithm based on YOLO framework," Chin. J. Liq. Cryst. Disp. **38**(7), 945–954 (2023).
[16]Q. Zhao, B. Q. Li, and T. W. Li, "Target detection algorithm based on improved YOLO v3," Laser Optoelectron. Prog. **57**, 121502 (12) (2020).
[17]X. G. Tu, Z. H. Yuan, B. K. Liu, J. H. Liu, Y. Hu, H. Q. Hua, and L. Wei, "An improved YOLOv5 for object detection in visible and thermal infrared images based on contrastive learning," Front. Phys. **11**, 1193245 (2023).
[18]T. Qiao, H. S. Su, G. H. Liu, and M. Wang, "Object detection algorithm based on improved feature extraction network," Laser Optoelectron. Prog. **56**(23), 231008 (2019).
[19]X. Zhang, Z. Lv, Y. Sun, B. Huang, Z. Niu, G. Liu, and K. Mu, "Intelligent detection technology of infrared image of substation equipment based on deep learning algorithm," in *2021 IEEE Sustainable Power and Energy Conference, iSPEC 2021, Nanjing China, 22–24 Dec 2021* (Institute of Electrical and Electronics Engineers, Inc., 2021), pp. 3855–3860.
[20]T. Xue and C. D. Wu, "Target detection of substation electrical equipment from infrared images using an improved faster regions with convolutional neural network features algorithm," Insight **65**(8), 423–432 (2023).
[21]J. H. Ou, J. G. Wang, J. Xue, J. P. Wang, X. Zhou, L. C. She, and Y. D. Fan, "Infrared image target detection of substation electrical equipment using an improved faster R-CNN," IEEE Trans. Power Delivery **38**(1), 387–396 (2023).
[22]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM **60**(6), 84–90 (2017).
[23]C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, United States, 7–12 June 2015* (IEEE Computer Society, 2015), pp. 1–9.
[24]K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, United States, 7–9 May 2015* (ICLR, 2015).
[25]K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *29th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, United States, 26 June to 1 July 2016* (IEEE Computer Society, 2016), pp. 770–778.
[26]X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks" in *14th International Conference on Artificial Intelligence and Statistics, AISTATS 2011, Fort Lauderdale, FL, United States, 11–13 April 2011* (Microtome Publishing, 2011), Vol. 15, pp. 315–323.
[27]D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)" in *4th International Conference on Learning Representations, ICLR* 2016, 2-4 May, 2016 (ICLR, 2016).
[28]S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift" in *32nd International Conference on Machine Learning, ICML 2015, Lile, France, 6–11 July 2015* (International Machine Learning Society; 2015), Vol. 1, pp. 448–456.
[29]Y. X. Wu and K. M. He, "Group normalization," Int. J. Comput. Vision **128**(3), 742–755 (2020).
[30]G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "*Densely connected convolutional networks*," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society, 2017), pp. 2261–2269.