Detection of Safety Helmet Wearing Based on Improved Faster R-CNN

Songbo Chen
Zhejiang University
College of Electrical Engineering
Hangzhou, China
18826076797@163.com

Wenbo Wang
Qingyuan Power Supply Bureau
Qingyuan, China
201821014299@email.scut.edu.cn

Ye Ouyang
Qingyuan Power Supply Bureau
Qingyuan, China
245860898@qq.com

Huiling Zhu
South China University of Technology
School of Electric Power Engineering
Guangzhou, China
821305167@qq.com
Corresponding Author

Tianyao Ji
South China University of Technology
School of Electric Power Engineering
Guangzhou, China
tyji@scut.edu.cn

Wenhu Tang
South China University of Technology
School of Electric Power Engineering
Guangzhou, China
wenhutang@scut.edu.cn

***Abstract*—In order to ensure the safety of workers and the stable operation of the power grid, the power grid companies in China have developed a very strict safety control system which contains many regulations, such as safety regulations and the two-ticket regulations. However, some workers are still lack of safety awareness in that they even do not wear safety helmets when carrying out construction or maintenance projects in substations. Safety helmet is an indispensable safety tool in electric power work, which can maintain the head safety of workers at all times and avoid fatal injuries such as electric shock and strike. Working without safety helmet is not only a violation of the safety control system, but also a manifestation of not being responsible for personal life and property. Nevertheless, the existing control means can not identify and prevent such behavior timely, efficiently and accurately. In order to better avoid this unsafe behavior, this paper proposes the Improved Faster R-CNN algorithm to inspect the wearing of safety helmet. Considering the real situation, the Retinex image enhancement is introduced to improve image quality for the outdoor complex scenes in substations. K-means++ algorithm is also adopted for better adaptation to the small size helmet. The experimental results show that compared with the Faster R-CNN algorithm, the mean average precision of the Improved Faster R-CNN is improved and the real-time automatic detection of the wearing of safety helmets is realized.**

***Index Terms*—Object Detection, Safty Helmet, Faster R-CNN, Retinex**

I. Introduction

Safety helmet is a kind of safety measures which can protect the head from the injury caused by falling objects and other specific factors. As the most basic personal protective equipment for electrical workers, it is very important to make sure that a worker wears the safety helmet all the time during a project. However, some workers lack of safety awareness and it frequently happens that they forget to or intend not to wear safety helmets. According to the statistics carried out by a power grid company in China, the vast majority of accidents are caused by the non-standard operation of workers, among which, not wearing a safety helmet properly is the biggest problem. Therefore, to supervise whether all the workers have worn safety helmets correctly is of great significance in ensuring that a project can be carried out safely and successfully, especially in substations.

At present, whether workers wear a safety helmet is mainly supervised by video monitoring. However, video monitoring relies much on manual inspection, which consumes a lot of manpower and time. It also depends on the physical and mental conditions of inspectors. Obviously, this control method can not identify the workers who do not wear the safety helmet timely and efficiently. In order to automatically identify whether a worker has worn a safety helmet even when he is in a crowd, it is necessary to make use of object detection methods. Object detection technology combines image classification and target location, aiming to locate the correct position of the target object on the image, and identify it.

Traditional object detection methods usually use manual design characteristics. These characteristics usually need the method of image processing. Scale Invariant Feature Transform (SIFT) [1], Histogram of Oriented Gradient (HOG) [2] and other methods can extract partial features, which is helpful to deal with the problem of local occlusion. Prewitt [3], Canny [4] and other methods can extract the obvious edge texture features with less calculation cost. Then, the features extracted by the above methods will be input into the classifiers, such as Support Vector Machine (SVM) [5] or AdaBoost algorithm to achieve the goal of object classification. For instance, R. Lienhart et al. introduce a novel method to compute rotated haar-like features faster and a new procedure for classifiers in [6], improving the performance by about 23.8%. Almeida et al. take advantage of digital image processing based on Watershed Transform to segment arresters [7]. Zou and Huang use K-means algorithm to cluster the infrared images of electrical equipment and SVM for classification in [8]. Rahmani realizes

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

the detection of electrical equipment fault by extracting the Zernike feature and SVM classifier in [9]. Only for specific categories, these methods are computationally complex and difficult to make full use of image information, as a result of which they only achieve low accuracy and show a poor generalization ability.

In recent years, deep learning has made a breakthrough in image classification and object detection, and becomes the most effective automatic feature learning method. It can automatically learn the feature representation of the target layer by layer, and represent the data as abstract features through a series of nonlinear transformations of the original data and avoid the inefficient and tedious manual design features. Thanks to the rapid development of deep learning, object detection technology based on deep learning has also made great progress. Girshick et al. proposed Regions with CNN (R-CNN) [10] in 2004, which impelled object detection to develop greatly. It extracts regions where target objects may be contained, called region proposals, through Selective Search method. These region proposals will be separately input into neural networks such as AlexNet [11] and VGGNet [12] for feature extraction, and is finally discriminated by SVM. Not long after that, Fast R-CNN [13] as well as Faster R-CNN [14] are introduced to reach higher speed. Fast R-CNN no longer repeatedly extracts features from each region proposal, but directly from the entire image, which significantly reduces computation time. Nevertheless, it continues to use Selective Search of R-CNN to extract region proposals, wasting a lot of time. On this basis, Faster R-CNN introduces Region Proposal Network (RPN) [15] to extract region proposals, which further shorten the time of region extraction. Later, truly state-of-the-art detection algorithms such as You Only Look Once (YOLO) [16], [17] show up prominently. These algorithms can attain the location and classification of the target directly through one network. Li et al. propose a novel Scale-Aware Fast R-CNN model, which performs excellently in detecting small pedestrian [18]. Guo proposes a DisturbIoU algorithm to prevent overfitting, and apply the Faster R-CNN algorithm to recognize the electric equipments in [19]. Shi et al. propose a pruning method for a reduced Single Shot Multi-Box Detector (SSD) [21] and obtain a higher mean value of average precision (mAP) than the original one in [20].

The Faster R-CNN can achieve higher accuracy compared with the YOLO algorithm. In the real-time monitoring system of substations, workers who do not wear safety helmets need to be identified correctly and efficiently. Therefore, the detection of safety helmet is achieved based on the Faster R-CNN algorithm, called the Improved Faster R-CNN. And K-means++ and Retinex [22], [23] are introduced in order to improve the detection accuracy on the premise of ensuring the detection speed.

II. Detection of Safty Helmets Wearing Based on the Improved Faster R-CNN

The detection model of safety helmet based on the Improved Faster R-CNN is shown in Fig. 1. First, we constructed a

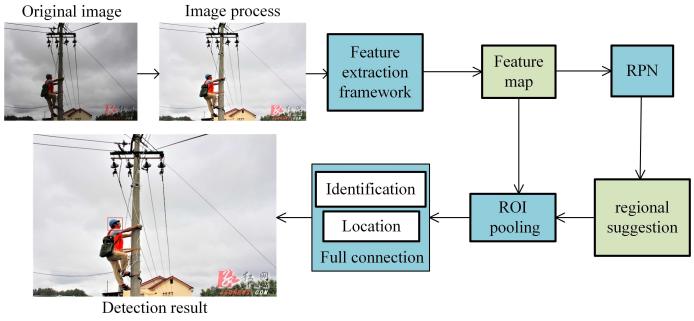

Fig. 1. Detection model.

data set of safety helmets. After Multi-Scale Retinex with Chromaticity Preservation (MSRCP) processing, these image of our data set is input into the feature extraction framework for feature extraction. And then the output of the feature extraction framework, called feature map, is fed into the RPN to generate region proposals and their scores. In order to accurately inspect the wearing situation of safety helmets, K-means++ algorithm is introduced to analyze the size of safety helmets in our data set, which is much more helpful to generate an appropriate size of region proposals. Afterwards, the region proposals and the feature map are input into the Region of Interest (RoI) pooling layer simultaneously to collect features of corresponding region proposals. Finally, the features of the proposed regions are passed through to the full connection layer, which calculates the classification scores and coordinates of the region proposals.

A. Feature Extraction

First of all, images will be normalized before being input into the feature extraction framework. For example, if we set the maximum image size as 600×1000, then images whose width is larger than 600 will be cut to 600, and its height will be clipped to the scale of the width. After scaling, if the height is still larger than 1000, the height is further scaled to 1000 and the width is scaled in proportion accordingly, so as to guarantee that the maximum image size does not exceed 600×1000.

The Improved Faster R-CNN algorithm for feature extraction is based on the ResNet-101 (Res101) [24], the entire framework of which is depicted in Fig. 2.

As shown in Fig. 2, the feature extraction framework includes 5 main blocks, i.e., Topblock, Block1, Block2, Block3, Block4 as the standard ones. The Topblock involves 2 layers, namely, a convolution layer and a pooling layer. Except the Topblock, there are 33 blocks in the whole framework, which are simply stacked Block1, Block2, Block3 and Block4 repeatedly. For each block of the rest 33 blocks, it uses a 1×1 convolution block and a 3×3 convolution block to reduce the dimension of the input, and then makes use of a 1×1 convolution block to restore its dimension. Moreover, the input of each block is added to the output of the stacked convolution blocks, which avoids the problem of vanishing

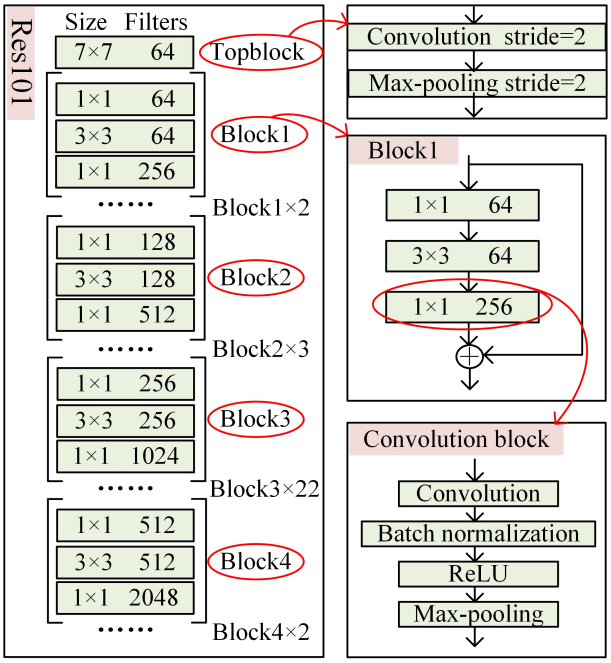

Fig. 2. Feature extraction framework.

gradient and degeneration caused by the over depth of the network. In each convolution block, right after a convolution layer, there is one batch normalization layer, one ReLU layer and one max-pooling layer respectively. If the stride of operation is 2, the size of the image will be reduced from $M \times N$ to $(M/2) \times (N/2)$. Since only the stride of the first convolution layer, the first pooling layer in Topblock, the last Block1 and the last Block2 is 2, the image size changes to $(M/2^4) \times (N/2^4) = (M/16) \times (N/16)$ after passing through the feature extraction network.

B. Region Proposal Networks

Instead of traditional sliding window and Selective Search method, RPN is used to generate region proposals directly, which greatly improves the generation speed of region proposals. The RPN model based on region proposals is illustrated in Fig. 3.

After passing through the feature extraction framework, a $(M/16) \times (N/16)$ feature map would go through a convolution layer with a window of 3×3 to further centralize feature information. Then, each grid cell of the $(M/16) \times (N/16)$ feature map is assigned 9 anchors.

In order to improve the accuracy of helmet wearing detection in term of size, K-means++ clustering analysis is introduced to cluster the actual size of safety helmet. It is widely acknowledged that K-means clustering is a simple and practical algorithm. It randomly selects *K* objects as the initial clustering center, then calculates the distance between each object and each clustering center, and assigns each object to the nearest clustering center. However, the classification result can be affected by the selection of initial points which was

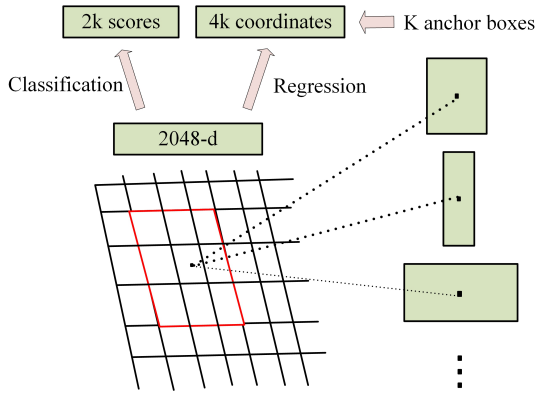

Fig. 3. Region proposal network.

randomly selected. In this connection, K-means++ clustering algorithm improves the selection of initial points rather than simply choosing randomly. Moreover, the purpose of the algorithm is to make the prediction box closer to the ground-truth, regardless of its size. The larger the intersection-Over-Union (IOU) between anchor and ground-truth is, the closer the anchor is to the ground-truth. Thus we replace Euclidean distance with IOU between anchor and ground-truth to automatically generate anchors and set the number *K* of anchor box to 9. The calculation of the IOU and the distance metric for K-means++ is as follows:

$$\mathrm{IOU(A,B)} = \frac{\text{Intersection of A and B}}{\text{Union of A and B}} \tag{1}$$

$$\begin{aligned} &d(\text{anchor,ground-truth}) \\ &\quad = 1 - \mathrm{IOU(anchor, ground\text{-}truth)} \end{aligned} \tag{2}$$

That is to say, each grid cell of the feature map generates 9 anchors with different aspect ratios such as 1:1, 1:2 and 2:1 and different areas such as 64^2, 128^2 and 256^2. These different sizes of anchors are actually equivalent to the multi-scale methods commonly used in object detection. If the size of feature map is 40×60, RPN will totally generate $40 \times 60 \times 9 = 21600$ anchor boxes.

Afterwards, these anchors are preliminary positioned and classified through two branches. One of the branches involves a convolution layer with a 1×1 window, and it is used to classify anchor boxes into two categories, namely, foreground and background. The other branch has a convolution window, whose size is also 1×1, and it outputs four corresponding coordinate values to initially acquire the location of the target objects.

In order to further correct the location of these anchors, bounding-box regression is performed to make them closer to the ground-truth. Fig. 4 depicts the principle of bounding-box regression. The box of dotted line represents the anchor box, and the box of solid line indicates the ground-truth. The translation (t_x, t_y) and scale factor (t_w, t_h) between the anchor box and the ground-truth can be calculate as:

$$t_x = (x - x_a)/w_a \tag{3}$$

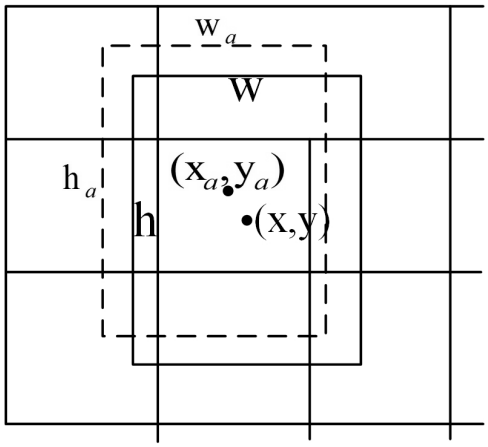

Fig. 4. Coordinate regression.

$$t_y = (y - y_a)/h_a \tag{4}$$

$$t_w = \log(w/w_a) \tag{5}$$

$$t_h = \log(h/h_a) \tag{6}$$

In order to force the network to pay attention to the most interesting area, Non-Maximum Suppression (NMS) is adopted to further filter out the overlapped anchors in foreground anchors. And eventually the top 300 boxes, which is called the region proposals, will be sifted to input into the RoI Pooling layer.

C. RoI Pooling

For traditional CNN, when the network is trained, the input image size must be a fixed value, and the output is also a fixed size. If the size of image is uncertain, it is necessary to cut the image to the required size through wrapping, cropping or other operation, which is much more complicated and may damage the integrity of the image itself. Inspired by [15], Faster R-CNN introduced RoI Pooling to solve this problem.

Inputting the 300 region proposals generated by RPN and the feature map produced by the feature extraction framework to the RoI pooling layer, RoI pooling first maps the region proposals to the corresponding position of the feature map. Then it divides the mapped area into sections of the same size, and perform the max-pooling operation. Greatly improving processing speed, this method finally obtains output with fixed size from proposals of different sizes. These outputs will be input into the full connection layer to complete the function of identification and more accurate positioning.

D. Calculation of Loss Function

In the entire network, the loss function can be divided into two parts, calculated by:

$$\text{loss} = \frac{1}{N_{\text{cls}}} \sum_i L_{cls}\left(p_i, p_i^*\right) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{reg}\left(t_i, t_i^*\right) \tag{7}$$

in which, the first term is generated by classification, and the second term is produced by bounding-box regression. The detailes of classification loss and regression loss are calculated by:

$$L_{\text{cls}}\left(p_i, p_i^*\right) = -\log\left[p_i^* p_i + \left(1 - p_i\right)\left(1 - p_i^*\right)\right] \tag{8}$$

$$\mathrm{L}_{\text{reg}}(t_i, t_i^*) = \begin{cases} 0.5(t_i - t_i^*)^2, & \text{if } |t_i - t_i^*| < 1 \\ |t_i - t_i^*| - 0.5, & \text{otherwise} \end{cases} \tag{9}$$

where i indicates the index of anchors.

The classification loss can be divided into two parts. One is generated by RPN classifying foreground and background anchors, where if the anchor *i* is regarded as negative, $p_i^* = 0$; otherwise, $p_i^* = 1$. The other one is generated by the last full connection layer for precise identification, where p_i represents the probability of anchor i predicting as helmet or no helmet while p_i^* represents its corresponding reality.

Similarly, the regression loss is also composed of RPN loss and the loss of the last full connection layer for precise location. t_i embodies each coordinate of the predicted bounding-box while t_i^* embodies each coordinate of the corresponding ground-truth box.

In our experiment, the value of N_{cls} is about 256, and the value of N_{reg} is nearly 2400, as a result of which λ is simply set as 10 in order to balance the influence of the two terms.

III. Experiments

In this section, the existing data and its preprocessing method are briefly introduced. Afterwards, the processed images are fed into the proposed algorithm to obtain the required detection model. Finally, the performance of the proposed algorithm is evaluated through a series of experiments.

A. Experiment Preparation

Since no public data set has been established in the research on helmet wearing detection, the data used in this experiment are collected from the Internet and a substation located in Qingyuan City on-site, with a total of 1065 images, including images with and without helmets of various background and different construction sites, substations and other places. Some typical samples in our data set are shown in Fig. 5.

Some images in our data set are too dark or backlit. Therefore, in order to improve the quality of these images and make the learning of the network more conducive, the Retinex based image enhancement technology is adopted. The Retinex theory holds that the image observed by human is the product of light image *L* and object reflection image *R*. *R* is the image of real constancy, because *R* depends on the reflection property of the object itself, basically unchanged, while *L* changes with the brightness of the environment.

As shown in Fig. 6, original images are shown on the left side, and the images on the right is the effect after MSRCP processing. From the figure it can be found that when the weather is bad or the image is dark, MSRCP processing enhances the image while maintaining the color of the original image and retaining the details to some extent.

Fig. 5. Samples of the data set.

Fig. 6. MSRCP operation comparison.

The experimental images have been already processed. In order to realize the supervised learning of the helmet wearing detection, the images need to be manually marked to indicate whether people wear safety helmets. Therefore, the head of everyone in the images is labeled by a rectangular box with its top left vertex coordinate $(x_{\min}, y_{\min})$, the bottom right vertex coordinate $(x_{\max}, y_{\max})$ and its category information (helmet or no helmet), which is made into a standard PASCAL VOC data set format and stored in XML files.

As mentioned before, the K-means++ clustering algorithm is also introduced to generate anchors. In order to make anchor closer to our target, the following results as shown in Table I can be obtained by applying (2). It can be find out that most of the anchors are small targets, and their aspect ratio is less than 1. Hence, the basic parameters are modified that the 9 anchors are composed of three aspect ratios, including 0.8, 0.9 and 1, and three areas, including 32^2, 64^2 and 128^2.

In the data set, about 90% of the images are randomly selected as the training set and the rest 10% for testing. The details of the data set are shown in Table II.

TABLE I
DETAILS OF ANCHORS

anchor	25×28	37×42	51×57
Aspect ratio	0.89	0.88	0.89
anchor	67×77	84×100	110×128
Aspect ratio	0.87	0.84	0.86
anchor	137×166	191×226	310×365
Aspect ratio	0.83	0.85	0.85

TABLE II
DETAILS OF DATA SET

Category	Amount for training	Amount for testing	Total
Helmet	421	111	532
No helmet	431	102	533
Total	852	213	1065

After using aforementioned MSRCP algorithm, the quality of our data is improved, and then training images and their labeled information are used to train the entire network through 20000 iterations. The loss is optimized by Momentum, where $\text{momentum} = 0.9$ and its weight decay is 0.0005, with a batch size of 16 and a learning rate of 0.001. The experiment is conducted on the computer system with Ubuntu 18.04, and the hardware configuration is Nvidia GeForce GTX2080 GPU, 11G memory.

B. Training process

Fig. 7 shows the plot of the average loss against the number of iterations during the training process. The total loss includes regression loss of RPN, classification loss of RPN and precise regression and identification loss of the last full connection layer. At the beginning of training, the total loss decreases faster and then gets slower as the number of iterations increases.

As can be seen from the second figure, after about 10500 iterations, the validation loss is lower than the training loss, which indicates that the model can perfectly predict the training data but perform poor on new data. To prevent this overfitting phenomenon, and take the time cost and economic benefit into account simultaneously, the model of the 10000th iteration is selected as the detection model.

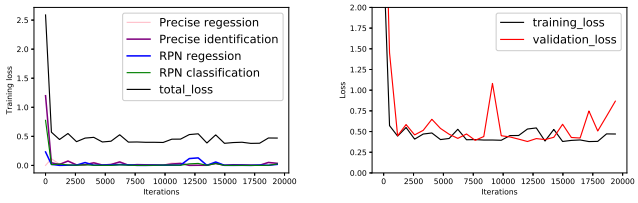

Fig. 7. Training and validation loss.

C. Experimental results

In order to evaluate the effectiveness of the proposed method, Precision, Recall, mAP and Frames Per Second (FPS) are used in the experiment. The definitions of the indexes are given as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

$$\text{FPS} = \text{Total Frames}/\text{Elapsed Time} \tag{12}$$

where TP indicates correct prediction of a positive sample, FP indicates erroneous prediction of a negative sample and FN represents erroneous prediction of a positive sample; Precision changes with recall values vary. The average value of precision under different recall values is called average precision (AP), and the average accuracy of all categories is called mean Average Precision (mAP).

Fig. 8 presents some detection results, which demonstrates that this method can accurately locate the targets in the image, regardless of the image size or light intensity. Even in the extreme case, as shown in the fourth image, it can still detect that the person is wearing a safety helmet, which shows the generalization and robustness of the proposed algorithm. In this experiment, its mAP achieves 94.3%, and the average detection speed is about 11 images per second, which ensures the high accuracy to meet the real-time requirements.

Table III compares the test results of YOLO, the Faster R-CNN algorithm and the Improved Faster R-CNN algorithm using the data set collected in this research. Through the comparative analysis, we can find that the mAP of the improved Faster R-CNN model reaches 94.3%, and the average precision of helmet wearing is 93%, while the average precision without helmet is relatively lower, 93%. In general, the Improved Faster R-CNN is the only breakthrough of 90% among three models.

TABLE III
COMPARISON OF DIFFERENT METHODS

Detection Method	AP for helmet	AP for no helmet	mAP
YOLO	82.0%	89.1%	85.6%
Faster R-CNN	85.3%	88.1%	86.2%
Improved Faster R-CNN	**94.7%**	**94.0%**	**94.3%**

IV. Conclusion

In order to detect the personnel without safety helmet in the construction of substation effectively and timely, the Faster R-CNN algorithm is improved to accurately detect the safety helmets. By applying this method to the on-side operation of substations, the interference of light, distance and other factors can be overcome and the wearing situation of multiple people can be identified at the same time. The high accuracy of this method has been verified by our data set, and the experimental results show that the mAP of the model can reach 94.3%, and the detection speed is 11.62 images per second, which proves that the model has a good detection ability. It is believed that this method can get better results by including more images of different backgrounds in the data set. However, the detection speed is relatively slow. The next step is to improve the detection speed while ensuring high detection accuracy.

Fig. 8. Detection results.

Acknowledgment

This work is funded by Research Project of Qingyuan Power Supply Bureau, China Southern Grid (GDKJXM20183511).

References

[1] Christopher J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2):121-167.

[2] Wei Hao, Jiebo Luo. Generalized Multiclass AdaBoost and Its Applications to Multimedia Classification[C]// Computer Vision and Pattern Recognition Workshop, 2006 Conference on. IEEE, 2006.

[3] Canny J.A computational approach to edge detection.IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, PAMI-8 (6) :679-698.

[4] Dalal N, Triggs B.Histograms of oriented gradients for human detection.In:Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) .San Diego, CA, USA:IEEE, 2005, 1:886-893.

[5] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[J]. CVPR (1), 2001, 1(511-518): 3.

[6] Lienhart R , Maydt J . An Extended Set of Haar-like Features for Rapid Object Detection[C]// Image Processing. 2002. Proceedings. 2002 International Conference on. IEEE, 2002.

[7] C. A. L. Almeida et al., Intelligent thermographic diagnostic applied to surge arresters: A new approach, IEEE Trans. Power Del., vol. 24, no. 2,pp. 751C757, Apr. 2009.

[8] H. Zou and F. Huang, A novel intelligent fault diagnosis method for electrical equipment using infrared thermography, Infr. Phys. Technol., vol. 73, pp. 29C35, Nov. 2015.

[9] A. Rahmani, J. Haddadnia, and O. Seryasat, Intelligent fault detection of electrical equipment in ground substations using thermo vision technique, in Proc. 2nd Int. Conf. Mech. Electron. Eng., vol. 2, Aug. 2010, pp. V2-150CV2-154.

[10] Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(1): 142-158.

[11] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[13] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[14] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.

[15] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[16] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2016: 779-788.

[17] Redmon J, Farhadi A. YOLO9000: Better,Faster,Stronger[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[18] Li J, Liang X, Shen S M, et al. Scale-aware fast R-CNN for pedestrian detection[J]. IEEE Transactions on Multimedia, 2017, 20(4): 985-996.

[19] Jianlong G, Weixia F, Manhua W. An Improved Faster R-CNN Algorithm for Electric Equipment Detection[C]//Proceedings of the 2019 11th International Conference on Computer and Automation Engineering. ACM, 2019: 138-141.

[20] Shi G , Shi G , Shi G , et al. Visualization and Pruning of SSD with the base network VGG16[C]// International Conference on Deep Learning Technologies. ACM, 2017.

[21] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[22] Jobson D J,Rahman Z,Woodell G A. A multiscale retinex for bridging the gap between color images and the human observation of scenes.[J]. IEEE Transactions on Image Processing,1997,6(7).

[23] Jobson D J,Rahman Z,Woodell G A. Properties and performance of a center/surround retinex.[J]. IEEE Transactions on Image Processing,1997,6(3).

[24] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[25] Simonyan, Karen, Zisserman, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.