Power Equipment Image Recognition Based on Deep Learning

Xin Xie[1, a], Chunjiang Pang[1, b]

[1]School of computer science, North China Electric Power University, Baoding 071000, China;

[a]xiexin@ncepu.cn, [b]972158083@qq.com

Abstract

Aiming at the large amount of equipment data produced in the process of acceptance and maintenance of electric power equipment, a deep learning based image recognition algorithm for electric power equipment is proposed. The target detection model based on Faster R-CNN is used to extract three convolutional neural networks, ZF-Net, VGG-16 and ResNet-101, from the images of electric power equipment acquired from substation site. As the feature extraction network of Faster R-CNN target detection model, the ResNet-101 network with residual structure is selected, and the recognition accuracy is up to 93.4%.

Keywords

Power equipment identification, deep learning, Faster R-CNN.

1. Introduction

With the steady development of smart grid, many unattended substations and inspection robots have been built and installed in recent years [1]. The substation transmits the captured image of power equipment to the monitoring system through remote viewing function [2]. The intelligent real-time monitoring of the equipment has been completed, but the intelligent recognition points can not be achieved. Analysis. The received monitoring data are checked and analyzed by the inspectors one by one with the naked eye. The huge amount of data poses a great challenge to the inspectors. At the same time, the subjective judgment of the inspectors will lead to wrong judgment and missed judgment, as well as the fatigue of the eyes, which will seriously interfere with the safe and stable operation of power equipment [3].So it is necessary to use computer to recognize and analyze power equipment image intelligently. Most of the existing power equipment image recognition methods adopt pattern recognition or quadratic template matching [4]. The recognition accuracy is low, and the image needs complex preprocessing, and the recognition efficiency is greatly affected.

With the rapid development of computer vision technology and in-depth learning, great progress has been made in the field of image recognition both in recognition accuracy and recognition speed. It has been fully utilized in many fields such as medical image, industrial production, unmanned driving, face recognition and so on. In order to solve the problem of intelligent recognition and analysis of power equipment images and meet the requirements of high precision and high efficiency of power grid equipment, Faster R-CNN [5] target detection model based on deep learning is introduced into power equipment image recognition. End-to-end target detection is completed, and recognition speed is improved on the basis of improving recognition accuracy. According to the image characteristics of power equipment, three different convolution neural networks are selected to extract the features of power equipment image. After experimental comparison, the most suitable network model for power equipment recognition is selected.

2. Faster R-CNN Target Detection Model

After the accumulation of R-CNN [6] and Fast R-CNN [7], Faster R-CNN model integrates convolutional neural network layer, regional proposal network (RPN) [8], region of interest pooling, classification and border regression into a network. As shown in Figure 1, end-to-end learning is realized, which makes recognition accuracy and border regression more accurate. The detection speed has been greatly improved.

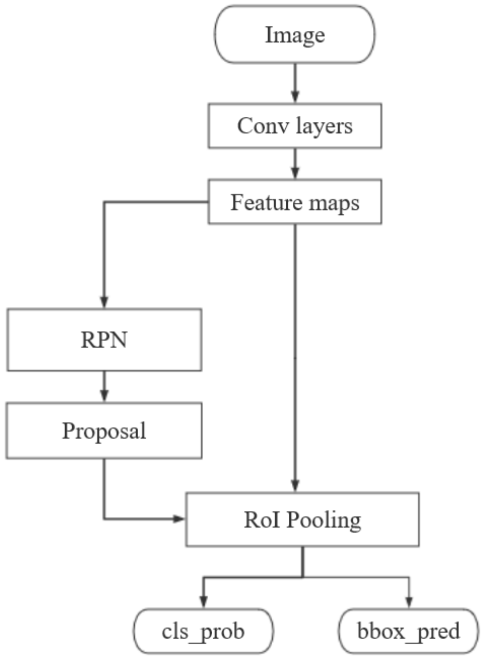

Fig. 1 Faster R-CNN target detection model structure

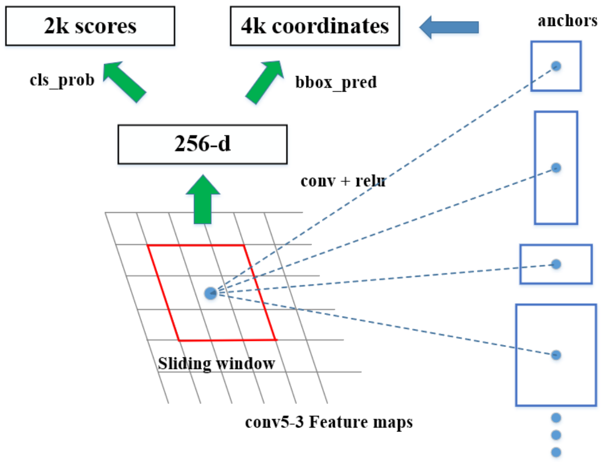

Fig.2 RPN propagation process

2.1 Convolutional Neural Network Layer

As a target detection method based on convolution neural network, Faster R-CNN first uses a set of classical convolution networks (such as ZF-Net [9], VGG16 [10], ResNet-101 [11]) which are pre-trained on the ImageNet data set to perform convolution and pooling operations for extracting image features and inputting them into an arbitrary size image. Pictures are output as feature maps, which are shared for subsequent RPN and full connection layer. This is also the key to accelerate the training process and improve the real-time performance of the model.

2.2 RPN

Traditional target detection methods take a long time to generate suggestion boxes, such as: AdaBoost algorithm uses sliding window + image pyramid method to generate detection boxes [12]; R-CNN and Fast R-CNN use selective search (SS) [13] method to generate suggestion boxes. Faster R-CNN uses RPN to generate regional suggestion boxes. It is a full convolution network, which can be completed synchronously on GPU, and greatly speeds up the generation of suggestion boxes.

Figure 2 shows the propagation process of RPN network. Taking ZF-Net as an example, the feature graph conv5-3 of w*h*256 is obtained. As the input of RPN, the sliding window of 3*3 is used to scan the feature map. Even if the convolution core of 3*3*256 is used to perform a convolution operation, the feature graph of w*h*256 is obtained. Each point can associate spatial information of local domain. At this time, each point on the feature map is mapped back to the original image one by one, and K anchor frames are obtained. In the original map, w * h * k anchor frames are generated. At the same time, the 256-dimensional feature maps obtained from the middle layer are deconvoluted by convolution kernels of 1*1*256*2k to get w*h*2k classification layer, and the boundary regression layer of w*h*4k is obtained by convolution kernels of 1*1*256*4k. The classified layer outputs the judged anchor frame with the score of foreground (object) and background through the function of softmax [14], each point is 2k; each anchor frame in the border regression layer corresponds to four offsets [x, y, w, h], so each point generates 4K position offsets.

Anchor frame is an important innovation in RPN. In order to get multi-scale detection frame, the traditional method of extracting candidate frame needs to establish image pyramid for multi-scale sampling. RPN scans the feature map with 3 *3 convolution core, maps the points corresponding to the center of convolution core back to the original image, and generates nine anchor frames with three scales {128, 256, 512} and three aspect ratios {1:1, 1:2, 2:1}, as shown in Figure 3. The size and number of the anchor frame are set by the actual detection target. The model usually adjusts the picture to 800*600, while the maximum aspect ratio of 1:2 in the anchor frame is 352*704, and the maximum aspect ratio of 2:1 is 736*384. Therefore, nine different aspect ratios and different sizes of the anchor frame basically cover the various sizes and shapes of the image.

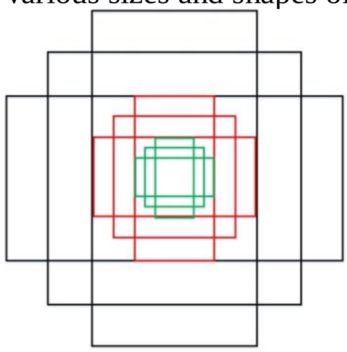

Fig.3 9 multi-scale anchors

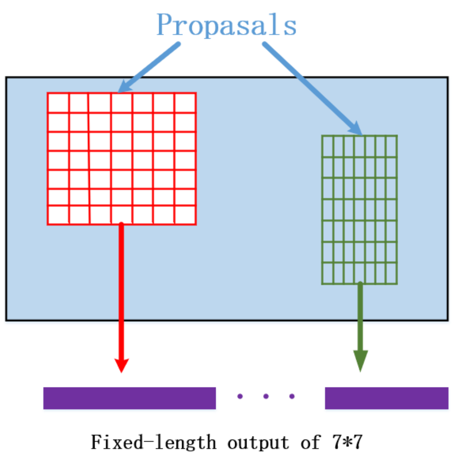

Fig.4 Region of interest pooling Schematic diagram

2.3 RoI Pooling

The pooling layer of region of interest receives the proposed region of RPN output and the feature map of convolution layer output, extracts the feature map of the proposed region and sends it to the follow-up network. Because the size of the proposed region is not fixed, it is impossible to input the fully connected network, so by improving the SPP-Net algorithm [18], the pooling of regions of interest is proposed. The principle is to map the proposed area corresponding to the original image to the feature map, then divide it horizontally and vertically into seven parts, and pool the maximum value of each part, so that the different size of the proposed area can be changed into a fixed size of 7*7 feature, and achieve a fixed size output, as shown in Figure 4.

2.4 Classification and Border Regression

As shown in Fig. 5, the feature maps of 7*7 size for each proposed region are obtained from the pooling layer of the region of interest, and then flattened into one-dimensional vectors. The probability vectors of each category are output through the full-connection layer with N+1 neuron units and the Softmax layer, where N is the total number of categories, and at the same time through the full-connection of 4 N neuron units. At the next layer, the position offset of each proposed region is obtained by the border regression, and the final target detection frame is obtained.

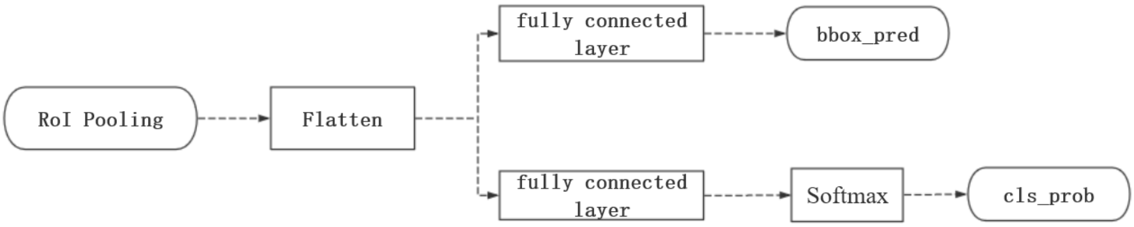

Fig.5 Classification and border regression process

3. Experimental Data Processing and Result Analysis

3.1 Experimental Environment

This experiment is based on tensorflow deep learning framework under Ubuntu 16.04 operating system to train and test the target detection model. The hardware configuration of the computer is: CPU is Intel Core i7-8700k, GPU is NVIDIA GTX1080Ti, memory is 16G, solid-state hard disk 2TB.

3.2 Experimental Data Set and Preprocessing

The data set of this experiment is based on the power equipment pictures collected in the substation field. The resolution of the image is converted to 1200 x 800. Among them, there are 2000 round pressing plates, terminal rows, air switches and square pressing plates, 80% of each category is training data, 10% is validation data set and 10% is test data set.

LabelImg is used to label the collected data sets manually. As shown in Figure 8, the object can be labeled directly in the form of a rectangular box through LabelImg, then the category of the object can be input, and the label file in XML format will be generated by clicking the Save button, including each labeled file in the image. Information about other target objects, such as the category of the object, the coordinate position of the lower left corner and the upper right corner of the rectangular box, etc.

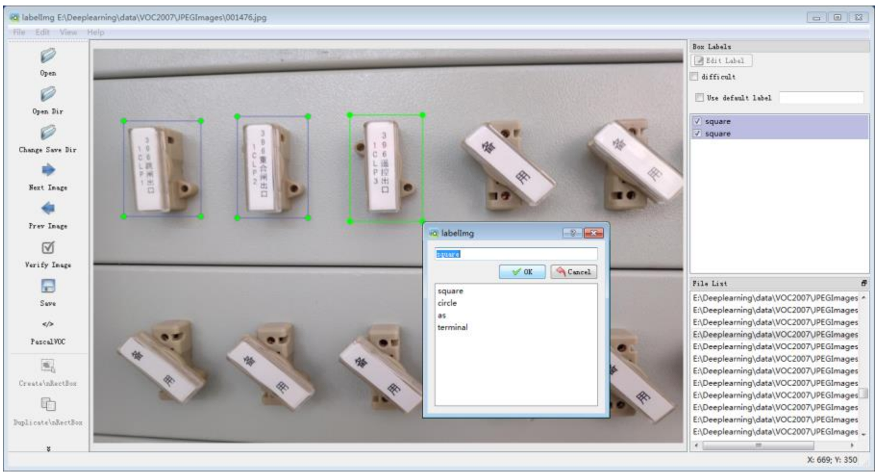

Fig.6 Annotated data set

3.3 Experimental Results and Analysis

In this experiment, the training batch size is 128, the learning rate is 0.001, the weight attenuation rate is 0.0005, the number of candidate areas before non-maximum suppression is 6000, and the number of candidate areas after non-maximum suppression is 300. In this experiment, mean Average Precision (MAP) and Frame Per Second (FPS) are used as the evaluation criteria of the model.

MAP: Represents the mean of the correct average accuracy for each category.

FPS: Represents the number of pictures that can be processed per second.

In training Faster R-CNN target detection model, three different convolution neural networks ZF-Net, VGG16 and ResNet101 are used as shared convolution layer. The recognition accuracy of three different network combinations on power equipment is tested and compared to find the best network structure for identifying power equipment image. Table 1 shows the average accuracy of power equipment identification and the average detection speed.

Table 1 Effect comparison of different convolution neural networks

Network Structure	MAP	FPS(frame/s)
ZF-Net	0.893	19
VGG16	0.917	15
ResNet-101	0.934	13

Observing the experimental results in the table above, under the same training data set and test data set, the maximum MAP of ResNet101 is 93.4%, followed by VGG16 model, whose MAP is 1.7% lower than ResNet101, and the lowest is 89.3% for ZF-Net. Because of the residual structure in ResNet101 model, it has deeper network and extracts deeper features. The generalization ability of the model is stronger. Secondly, VGG16 network with 16 layers has deeper network depth than ZF network, so the MAP is higher. But with the deepening of network layers, it needs training and testing. As can be seen from the table, there is a significant difference in the number of pictures recognized per second. ResNet101 recognizes at least 13 frames per second, while ZF-Net recognizes 19 frames per second.

4. Conclusion

In this paper, Faster R-CNN target detection model is studied to identify and detect power equipment pictures. It is verified that the algorithm model can automatically extract the features of different power equipment pictures through convolutional neural network after a lot of data training, and can quickly and accurately identify and mark which power equipment belongs to. Prepare. According to the characteristics of power equipment image and the requirement of real-time detection, ResNet101, a deep residual network with residual structure, is selected as the shared convolution layer by experiment comparison. It is used for feature extraction, and the recognition accuracy is up to 93.4%.

References

[1] LIU Yuefan,YE Haibin,LIU Kai,et al: Application research of patrol robot in 500 kV unmanned substation[J]. Power supply, 2016,33(9):69-72.

[2] LIN Yunzhen: Application of substation remote vision technology in power system automation[J]. Electronic test, 2018,401 (20):71-72.

[3] LI Junfeng,WANG Qinruo,LI Min: Image recognition of power equipment combined with deep learning and random forest[J]. High voltage technolo-gy,2017,43(11):3705-3 711.

[4] ZHANG Hao,WANG Wei,XU Lijie,et al: Application of Image Recognition Tech-nology in Power Equipment Monitoring[J]. Power system protection and control, 2010, 38(6).

[5] REN S, HE K, GIRSHICK R,SUN J: Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6):1137-1149

[6] GIRSHICK R, Donahue J, Darrell T, et al: Rich feature hierarchies for accurate object detection and semantic segmenta-tion[C]. Proceedings of the IEEE con-ference on computer vision and pattern recognition, Columbus, OH, United states, June,23-28,2014.USA: IEEE,2014:580-587.

[7] GIRSHICK R, Fast R-CNN [C]. IEEE Inter-national Conference on Computer Vision. Washington, USA; IEEE Computer Society, 2015:1440-1448

[8] Shelhamer E , Long J , Darrell T : Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016:1-1.

[9] Zeiler M D, Fergus R: Visualizing and Understanding Convolutional Networks[C]. European conference on computer vision, Zurich, Switzerland. September,6-12,2014. Zurich: Springer, Cham,2014:818-833.

[10] Simonyan K, Zisserman A: Very Deep Con-volutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[11] He K, Zhang X, Ren S, et al: Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2016.

[12] Peng Wang; Chunhua Shen; Barnes, N; Hong Zheng: "Fast and Robust Object Detection Using Asymmetric Totally Corrective Boosting," IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no.1, pp. 33-46,2012.

[13] Uijlings J R R, Van De Sande K E A. Gevers T, et al: Selective search for object recognition[J]. International journal of computer vision,2013,104(2)

[14] Memisevic R, Zach C, Pollefeys M, et al: Gated Softmax Classification[J]. Advances in Neural Information Processing Systems, 2010:1603-1611.