Violation Monitoring System for Power Construction Site

Xianghua Zheng, Jianguang Yao and Xingchun Xu
State Grid Taizhou Power Supply Company, Taizhou, Jiangsu, China
zhengxianghua_sg@163.com

Abstract. Automatic detection of violations will improve the efficiency of enterprise management and reduce the workload of employees. In this paper, a construction site monitoring system based on Tiny-yolo-dense is developed for three common violations of electric power construction site: no safety helmet, safety gloves and no labour suits. The experimental results show that the system can meet the requirements of real-time and accuracy of power construction site monitoring.

1. Introduction
Safety management is an important part of electric power enterprise management. It is a comprehensive system science. The object of safe management is the state management and control of all people, things, environment in production. Hence, it is a dynamic management. The level and success of safety management is directly related to the social reputation and economic benefits of enterprises, the state and collective property and the safety of workers' lives.

Due to the lack of qualified power construction enterprises, low quality and ability of some enterprises leads to the level of safety supervision is not in place, failing to manage, resulting in a lot of inappropriate mistakes. At present, the company has established a construction site safety monitoring system based on mobile video surveillance, which can realize remote video surveillance in the company's headquarters. However, due to the large number of construction sites and multiple concurrent video, it needs to invest a lot of manual work to carry out the surveillance, identification and processing of illegal operations. In case of shortage of persons, the efficiency is relatively low and some wrongs may be happened.

At present, with the development of artificial intelligence technology, image and video detection and recognition technology has been applied to all aspects of social life [1]. Therefore, based on artificial intelligence technology, it is urgent to carry out intelligent safety supervision on the construction site of power network. Through intelligent processing of construction site monitoring video, we achieve the automatic detection and identification of illegal operations. Three detection objects include: whether to wear a helmet, whether to wear insulating gloves, whether to wear labour suits. The above functions of monitoring system can greatly reduce the workload of safety supervisors and improve work efficiency to provide effective support for safety supervision work.

2. Related Work
Since AlexNet acquired a celebrated victory at the ImageNet Competition in 2012, convolutional neural network has attracted a lot of attention.

The biggest difference between deep learning and traditional pattern recognition methods is that features are learned automatically from large data rather than designed by hand[2]. Good features can
greatly improve the performance of the pattern recognition system. In the various applications of pattern recognition in the past decades, the characteristics designed by hand are in the dominant position. It usually takes five to ten years to produce a well recognized feature. Deep learning can quickly learn new and effective features from training data for new applications. In the framework of neural network, feature representation and classifier are optimized jointly, which can maximize the performance of their joint cooperation. In some famous competitions, convolutional network models improve the detection rate of objects by 20%.

3. Design of Detection System

3.1. Implementation Route of Detection System

The specific implementation route of our application is shown in Figure 1. For the intelligent detection of illegal behaviors in the construction site of distribution network, firstly, it is necessary to collect image samples of helmet, safety gloves and labour suits in the multi-form and multi-angle from the environment of the site. This step is called as video and image data acquisition. Video or images of operations which meet with safety regulations or be illegal to safety regulations are collected by the workers on the site of the distribution network. The images about different types of helmet, safety gloves and labour suits were collected to form a standard data set.

Then, the collected image and video data are pre-processed (including image enhancement and image amplification), and the image information is extracted by annotating the image. Because the size of the image data set is not uniform, the image is normalized to the same size, and then the normalized image data set is annotated. The annotation format meets the PASCAL VOC data set format. The VOC data set provides the standard image data set format for target recognition and provides for visiting. A common set of tools for querying data sets and annotations, where the XML file in the annotation folder of the data set is the interpretation of picture information.

Finally, the image information is trained by deep learning algorithm. The iterative optimization algorithm is used to adjust the detection model of illegal behavior which can meet the requirements of the construction site environment. According to the collected images, deep learning model is selected to train and optimize the illegal operation model. The core idea of deep network is to use the whole graph as the input of the network and directly return the bounding box (boundary box) position and its category in the output layer.

![Figure 1. The framework of detection system](image)

3.2. Data Amplification

The training samples amplified in this paper. Considering the difference of images with helmet, safety gloves and labour suits, this paper rotates all the samples of training samples by ±10° and ±20°. Moreover, we flip the image of safety gloves horizontally. The center of the image is intercepted. After rotation processing, if the object near the edge of the image is incomplete or completely lost, the labeling is abandoned.
3.3. Network Selection

At present, the popular target detection algorithms can be divided into two categories. One is R-CNN (R-CNN, Fast R-CNN, Faster R-CNN) based on region Proposal, which is two-stage. It is necessary to use selective search or CNN (RPN) to generate region proposal firstly. Then, classification and regression are done on region proposal. Differently, the other methods are one-stage algorithm. It uses only one CNN network to predict the categories and locations of different targets directly. This method is fast and can meet the real-time requirements of construction site monitoring.

In this paper, we use YOLO (You Only Look Once) algorithm in one-stage algorithm to detect the mobile phone target. This series of algorithms uses a single CNN model to achieve end-to-end target detection. The whole algorithm is processed in the following steps: firstly, the input image resize to 448x448. Then, they are put into training of CNN network. Finally, the network is processed to acquire the prediction results and get the detection target. Compared with the R-CNN algorithm, it is a unified framework. Moreover, its speed is faster. The Yolo’s training process is also end-to-end.

The predecessor of YOLOv2 network in this paper is one YOLO network [4]. YOLO network does not need region generation, and directly regress the prediction target in the whole input image, which greatly improves the detection speed. The network divides the reconstructed input image into S*S(S=7) meshes, predicts two boundary frames and their confidence in each mesh, and C(C=2) types. Conditional probability is used to filter the boundary frames with low thresholds by non-maximum suppression method to get the final target prediction frame.

In order to solve the problem of missing and inaccurate location in YOLO network detection, anchor frame, anchor frame height and width dimension clustering, high resolution classifier and improved position coordinate prediction methods are introduced in YOLOv2. In YOLOv2 training, a multi-scale training strategy [6] is adopted. In order to improve the detection speed, YOLOv2 designs Darknet-19 and Tiny-yolo network structures which only contain convolution layer and pool layer. On VOC2007 test set, YOLOv2 in 416 *416 is superior to Faster RCNN [3], SSD [5] and YOLO. However, the accuracy of YOLOv2 in detecting gloves and people, especially occlusion or overlapping targets, needs to be improved. In this paper, we propose: 1) Tiny-yolo network structure combined with dense connection; 2) labeling training samples to occlude overlapping targets’ foreground region to enhance the learning of foreground region characteristics.

In order to realize on-line illegal behaviors detection, Tiny-yolo with simple structure and low computational complexity is selected as the basic network of YOLOv2. Tiny-yolo consists of 9 layers of coiling and 6 layers of maximum pool.

The network leads to the loss of information layer by layer during transmission, which cannot make better use of multi-layer feature information and reduces the detection accuracy. In order to realize multi-layer feature reuse and fusion, and avoid the computational complexity caused by the new structure, this paper uses the idea of intensive connection proposed by Huang et al. [7], and embeds intensive modules only in the deep layer of Tiny-yolo network feature map with low resolution, that is, replacing Tiny-yolo’s 7th convolution layer with intensive modules (see dotted frame in Figure 2). A Tiny-yolo-dense network with dense connections is constructed (resolution 16*16) so that the eighth convolution layer can receive multilayer convolutions with dense connection block outputs.

![Figure 2. Network structure of Tiny-yolo-dense](image)
Table 1. Detection results of our network

| Class of scene | Table Column Head | Precision(%) | Recall(%) | Detection Speed(fram·s\(^{-1}\)) |
|----------------|-------------------|--------------|-----------|-------------------------------|
| safety glove   |                   | 94.79        | 91.22     |                               |
| safety helmet  |                   | 98.90        | 93.67     | 33                            |
| labour suit    |                   | 94.74        | 91.10     |                               |

4. Experiments and Outcome

4.1. Test Platform

In this paper, YOLOv2 algorithm[8] is implemented on the Darknet framework, and the server platform is configured as Intel R Xeon (R), CPU E3-1245 v3@3.40GHz processor, 16GB runs memory, 2T hard disk capacity, 12GB GTX 1080Ti GPU. The system used in this paper is Ubuntu 16.04.

4.2. Network Training Parameter Settings

In the training phase, the momentum term was 0.9. The asynchronous stochastic gradient descends, the initial learning rate is 0.001, and the attenuation coefficient is 0.0005.

4.3. Outcome of Experiments

In this experiment, we has 211 test images for the scene with safety glove, 182 test images for scene with safety helmet and 133 test images for labour suit. As shown in Figure 3, we can find our model can handle the detection of safety glove, safety helmet and labour suit. The detection results of our network is shown in table I.

![Detection samples of our model in different scenes](image)

(a) Detection samples of safety glove; (b) Detection samples of safety helmet; (c) Detection samples of labour suit;

5. Conclusion

The construction site of distribution network has strict requirements on whether to wear safety helmet, safety gloves and labour suits. Therefore, the detection requires high accuracy and real-time. In view of the differences in the detection mechanism of safety helmet, work gloves and labour suits, specific models should be trained separately. Moreover, the models also need to ensure the real-time detection by network camera. Hence, the accuracy and real-time of the training model is the key and difficult
points to ensure the applicability of the monitoring system.
In this paper, Tiny-yolo-dense is used to realize the monitoring system of violations in electric power construction site. Tiny-yolo-dense is one-stage algorithm, which has better real-time performance in deep learning. The violations in this paper include three states: no gloves, helmet and no labour suits. And these three conditions are transformed into the detection of gloves, helmet and labour suits commonly used in power construction. The system achieves good detection results and ensures the real-time detection for the availability of system.

6. References
[1] Bo Li, Cheng Chen,"Transmission line detection based on a hierarchical and contextual model for aerial images," Journal of Electronic Imaging 27(4), 043054, (2018). doi:10.1117/1.JEI.27.4.043054.
[2] Bo Li, Cheng Chen, First-Order Sensitivity Analysis for Hidden Neuron Selection in Layer-Wise Training of Networks. Neural Processing Letters, 2018, Vol.48, no.2, pp. 1105–1121.
[3] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
[4] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]// IEEE conference on Computer Vision and Pattern Recognition. 2016: 779-788.
[5] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]// European conference on computer vision. 2016: 21-37.
[6] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6517-6525
[7] Huang G, Liu Z, Weinberger K Q, et al. Densely connected convolutional networks[C]// IEEE conference on Computer Vision and Pattern Recognition. 2016: 4700-4708.
[8] Redmon J. YOLO: real-time object detection [EB/OL]. https://pjreddie.com/darknet/yolo/. 2017-10-24