Helmet Detection under the Power Construction Scene Based on Image Analysis

Yang Bo
Beijing Electric Power Research Institute
Beijing, China
email:2294759932@qq.com

Qin Huan
Beijing Electric Power Research Institute
Beijing, China
email:qinhuan89@126.com

Xie Huan
Beijing Electric Power Research Institute
Beijing, China
email:xiehuan1229@163.com

Zhu Rong
Beijing Electric Power Research Institute
Beijing, China
email:503157541@qq.com

Li Hongbin
State Grid Beijing Electric Power Company
Beijing, China
email:543648843@qq.com

Mu Kebin
State Grid Beijing Electric Power Company
Beijing, China
email:mukebin@163.com

Zhang Weizhong
State Grid Beijing Electric Power Company
Beijing, China
email:564365654@qq.com

Zhao Lei
Beijing Electric Power Research Institute
Beijing, China
email:1468083176@qq.com

***Abstract*—In this paper, an intelligent safety surveillance system is designed for the electric power construction system to detect whether workers wear helmets correctly. The YOLOv3 objective detection algorithm is used for fine-tuning the datasets for the electric power construction scene which is made by ourselves. The main targets for detection are the helmet and the head. The helmet and the head are tested to detect whether the worker wears a helmet. After our experiment, the accuracy of the helmet detection is over 90%. This method replaces the manual method with an automated way which saves a lot of human and material resources. This method can be applied to actual production.**

Keywords—Electric power; helmet; YOLOv3;objective detection

I. INTRODUCTION

The issues that the safety of worker in power construction systems have always been a major concern of the State Grid. A very important reason is that because workers do not wear helmets as required, their life safety is threatened. Wearing helmet not only prevents the damage of falling objects, but also makes the electric construction scene more orderly.

Traditional, people use manual methods to detect whether a worker wears a helmet in the power construction scenarios. The method not only wastes a lot of human and material resources, but also makes the efficiency of work slow. Therefore, this paper proposes an intelligent scheme based on image analysis, and designs an intelligent monitoring system for the power construction system. First we should install a monitoring device at the power construction site and then use the camera to capture high-definition live images and transfer them back to the computer. Then we should process the images, detect the target in the image and determine whether the workers wear helmets. Finally, if there is an image that a worker dose not wear a helmet, computer will alarm.

This work was supported in part by the State Grid Corporation Science and Technology Foundation 52022318001N, in part by the National Natural Science Foundation of China under Grant 61401113, in part by the Natural Science Foundation of Heilongjiang Province of China under Grant LC201426.

Before the rise of neural networks, objective detection methods mostly use the combination of features and classifiers to complete the detection of targets in different scenes. Common features are SIFT, HOG, SURF, color, texture features, etc.; common classifiers are Adaboost classifier, SVM classifier and so on. At present, in the field of objective detection, the algorithms are divided into a one-stage and two-stage method. There are algorithms such as SSD, YOLO, YOLOv2, YOLOv3 and so on in one-stage methods. There are algorithms such as RCNN, SPP-Net, Fast RCNN, Faster. RCNN, MR-CNN and so on in two-stage methods. The one-stage methods do not generate the bounding box, but solve the problem by regression directly. The two-stage methods are to generate a series of sample candidate frames by algorithm and then detect targets by convolutional neural network. The two-stage method is relatively high in detection accuracy and relatively insufficient in detection speed; the one-stage method is fast, but the accuracy is slightly inferior to the two stages. For the research project with the construction site as the project background, the two-stage target detection algorithm is more suitable for the detection of helmets.

In this paper, the worker's helmet wear is detected in the construction scenario. We use the target detection algorithm based on image analysis. This topic uses a one-stage YOLOv3 network to fine-tune and replace the traditional manual method in an automated way.

II. METHOD

YOLOv3 algorithm was proposed in 2018 which is improved based on YOLO and YOLOv2. Darknet-53 is the backbone of YOLOv3 which is used to feature extraction. The algorithm combines FPN ideas to fuse multiple scale feature maps. Figure 1 shows the basic schematic of the YOLOv3 algorithm.

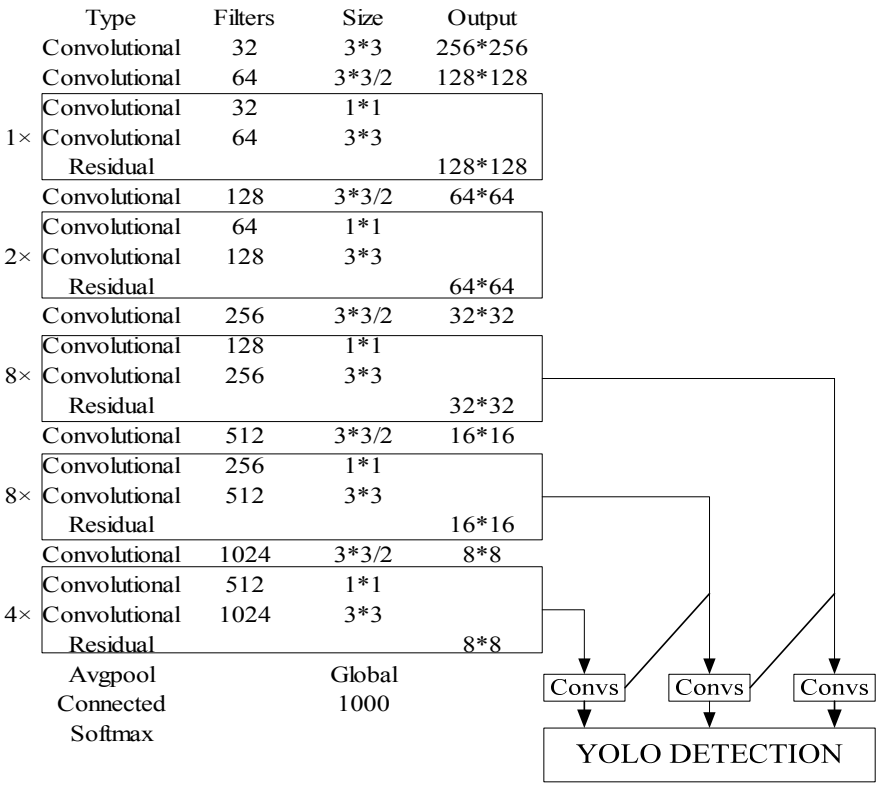


Fig. 1. YOLOv3 algorithm

A. Backbone

YOLOv3 algorithm use Darknet-53 as backbone to achieve objective detection. The backbone has 53 convolution layers, which are used to extract features. The Darknet-53 structure combines the idea of full convolution and also introduces the residual structure, which is effective for gradient problems in deep networks. The reason that the algorithm introduces the residual module is to ensure that the training models can still converge when the network is deep enough, and the deeper networks make the feature expression better. The residual module of the networks uses 1*1 convolution, which reduces the number of parameters and the amount of calculation during training.

The entire network has 5 subsamplings and 3 downsamplings. The network contact feature maps of the 32timesdownsampling and the 5th subsampling, the same as the 16 times downsampling and the 4th subsampling, 8 times downsampling and 3rd subsampling. In this way, the network learns both deep and shallow features, which is more perfect for model training. In the feature concat of 3 downsamplings, because the receptive field of 32 times downsampling is the deepest, it usually has a good effect on the detection of large targets. 16 times downsampling is suitable for objects of normal size; 8 times down sampling has the smallest receptive field and is suitable for detection of small targets.

B. Output

The YOLOv3 algorithm does not use the softmax method for regression operations, but uses logistic regression in order to avoid the same target having two categories that cannot be output at the same time. The prediction method of the bounding box refers to the idea of the SSD algorithm which uses the anchor method to predict the regression box. Then YOLOv3 uses K-means clustering method to select the regression box.

The loss function is selected in the way of YOLO, and the loss calculation is performed on the prediction frame length and width (w, h), coordinates (x, y), class and confidence.

Equation (1) is the loss function of length and width using the sum function.

$$SSE = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2 \tag{1}$$

The loss function of coordinates, category and confidence is equation (2) which uses the cross entropy loss function.

$$\begin{aligned} loss &= -\sum_{i=1}^{n} y_i \log \hat{y}_i + (1 - y_i) \log(1 - y_i) \\ \frac{\partial loss}{\partial y} &= -\sum_{i=1}^{n} \frac{y_i}{\hat{y}_i} - \frac{1 - y_i}{1 - \hat{y}_i} \end{aligned} \tag{2}$$

It can be inferred from the equation that the final loss is related to the difference between the predicted value and the true value.

The speed of YOLOv3 is far faster than other algorithms. For example, the detection time on the input resolution of 320*320 is only 22ms with the map value as 51.5.We compare the advanced methods in the field of target detection with YOLOv3 in Table 1.In terms of speed and accuracy, YOLOv3 has a good performance. Therefore, this paper use YOLOV3 algorithm as our main plan.

TABLE I. COMPARATIVE EXPERIMENT OF YOLOV3 AND OTHER METHODS ON VOC DATASETS

	mAp	time（ms）
SSD321	45.4	61
DSSD321	46.1	85
R-FCN	51.9	85
SSD513	50.4	125
DSSD513	53.3	156
FPN FRCN	59.1	172
RetinaNet-101-800	57.5	198
YOLOv3-320	51.5	22
YOLOv3-416	55.3	29
YOLOv3-608	57.9	51

III. EXPERIMENTS

Based on the consideration of speed and accuracy, this thesis designs the experimental scheme based on the idea of image analysis using YOLOv3 algorithm. As shown in Figure 2, the YOLOv3 algorithm is fine-tuned through a self-made datasets to realize real-time detection of the helmet under power construction scene.

(1) Training process: First, we should prepare the training dataset which is marked by a certain format, and then use the image preprocessing to the training data. At last, we use the YOLOv3 algorithm to fine-tune the processed data and obtain a final detection model.

(2) Testing process: First, we create a testing data, and the testing dataset and the training dataset cannot be the same. And then we use the image preprocessing to the testing dataset. At last, we use the model obtained from the training process to obtain the final test result.

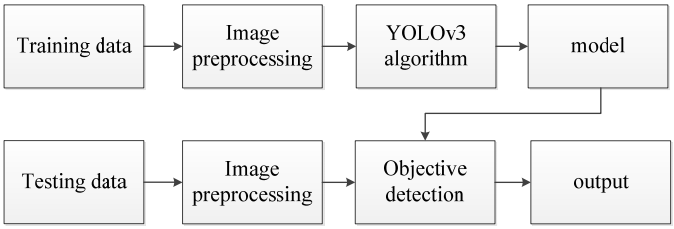
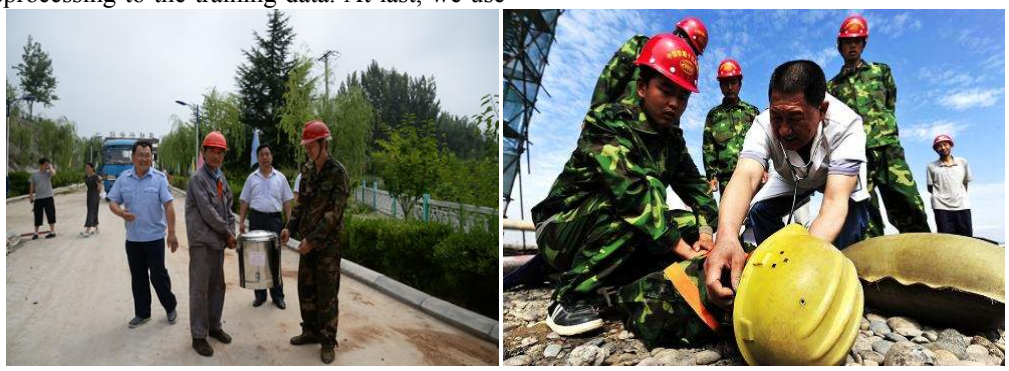


Fig. 2. Method of objective detection

A. Our datasets

Our dataset is for power construction scenarios, including the basic goals of helmets and heads. There are 1555 images in the dataset which are labeled in VOC format. The number of helmets is 2,253, and the number of heads is 2,346, the samples are in equilibrium. The image is marked with a rectangular frame, and the category information, position information, and coordinate information of the target are recorded in the xml files.



(a) (b)

Fig. 3. Our dataset

B. Data amplification

The training process of the network requires a large amount of data to support, otherwise it may be over-fitting during training. Unlike the ImageNet dataset, there are not enough images in our dataset for training the model. So it is necessary for us to make data amplification. In this paper, we use the rotation, saturation, exposure, and tonal adjustment of the image for data expansion. Data amplification means not only facilitates the optimization of the model, but also prevents over-fitting due to little data.

C. Model fine-tuning

The experiment used a self-made dataset for fine-tuning. According to the image annotation made by the self-made data, using the k-means algorithm to cluster, we get the value of 9 anchors of the construction scene hard hat dataset. In the experiment given by YOLOv3, the input resolution of the image was set to 416*416 on the public dataset, and the dataset used was the COCO dataset. Then the algorithm gets nine results: (10*13); (16*30); (33*23); (30*61); (62*45); (59*119); (116*90); (156*198); (373*326). Because there are small targets in our data, the input size of the image is changed to 608*608, and the anchor value obtained according to the actual data set is (31*38); (49*52); (60*71) ; (69*112); (87*119); (88*115); (115*113); (140*162); (204*296). Finally, we adjust the category parameters in the network structure, divide the training data set, and use the program to realize the training set, the verification set, the test set with a specific gravity of 3:1:1, the iteration number is 30,000 times, the batchsize is 64, and the learning rate is 0.0001.

IV. RESULTS

A. Experimental Results

After several adjustments to the hyper parameters, we obtained a stable model and used the model to test the data to obtain the final experimental results. In the test procedure, the thresholds that can be set are the confidence and NMS thresholds, where the confidence is set to 0.3 and the NMS threshold is set to 0.4.

(a) Small target detection (b) Helmet and head

(c)Big target detection (d) Occlusion detection

Fig. 4. Experimental result

It can be seen from the experimental results in Fig. 4 that the helmets in the scene can be correctly detected, and the algorithm can not only correctly detect multiple targets, but also have certain anti-occlusion characteristics. Figure (a) shows the detection of the targets when they are small, and Figure (b) shows the detection of the helmet and head. Figure(c) shows the detection results of big targets, we can see the results are good. Figure (d) shows the targets when they are blocked by others.

B. Statistical Results

Based on the final version of the test model, we used the test data for testing. We take the final experimental results obtained into mathematical statistics and analyze the final test results. Equation (3) is the basic formula of model statistics. Precission is the accuracy rate, False Alarm is the false alarm rate, and Missing Alarm is the missed alarm rate. TP is the positive sample that can be correctly detected in the experiment. FP is the positive sample that cannot be detected correctly. Sample, FN is the negative sample of error detection.

$$precission = \frac{TP}{TP + FP}$$
$$FalseAlarm = \frac{FP}{TP + FP} \quad (3)$$
$$MissingAlarm = \frac{FN}{TP + FN}$$

As shown in Table 2, it is the final accuracy rate, missed alarm rate and false alarm rate of the model. The test data has a total of 1326 images, of which the number of targets of the helmet is 2,423 and the number of heads is 2,541.

TABLE II. STATISTICAL RESULTS

	Precission	FalseAlarm	MissingAlarm
Helmet	96.58%	0.89%	3.42%
Head	97.12%	0.69%	2.88%

According to the statistical results in the table, The accuracy rate of the helmet is 96.58%, the false alarm rate is 0.89%, the miss alarm rate is 3.42%. The accuracy rate of head is 97.12, the false alarm rate is 0.69%, and the miss alarm rate is 2.88%. It can be seen that the accuracy of the YOLOv3 model fine-tuned under the self-made data of the construction site can reach 95% or more, and it can be applied to the actual construction scene for wearing the worker's helmet.

V. CONCLUSION

In this paper, we use the target detection algorithm YOLOv3 in deep learning to make a stable and reliable helmet detection model by making a self-made construction scene helmet data set and using this data set for fine adjustment. According to the experimental results, the detector can be applied to the site construction site for on-site inspection, and the detection rate of the helmet can be over 95%, meeting the real life needs.

REFERENCES

[1] Histograms of Oriented Gradients for Human Detection[C]// null. IEEE Computer Society, 2005.

[2] Viola P, Jones M. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2):137-154.

[3] Lowe D G. Object Recognition from Local Scale-Invariant Features[C]// iccv. IEEE Computer Society, 1999:1150.

[4] Freund, Yoav, Schapire, Robert E. A decision-theoretic generalization of on-line learning and an application to boosting[C]// European Conference on Computational Learning Theory. Springer, Berlin, Heidelberg, 1995:23-37.

[5] Vapnik V N. The Nature of Statistical Learning Theory[J]. IEEE Transactions on Neural Networks, 1997, 8(6):1564.

[6] W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]. European Conference on Computer Vision. Springer, Cham, 2016:21-37.

[7] 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. IEEE Computer Society, 2016:779-788.

[8] Redmon J , Farhadi A . [IEEE 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Honolulu, HI (2017.7.21-

2017.7.26)] 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - YOLO9000: Better, Faster, Stronger[J]. 2017:6517-6525.

[9] Redmon J, Farhadi A . YOLOv3: An Incremental Improvement[J]. 2018.

[10] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// Computer Vision and Pattern Recognition. IEEE, 2013:580-587.

[11] He K , Zhang X , Ren S , et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904-16.

[12] Girshick R. Fast R-CNN[J]. Computer Science, 2015.

[13] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015:1-1.

[14] Gidaris S, Komodakis N . [IEEE 2015 IEEE International Conference on Computer Vision (ICCV) - Santiago, Chile (2015.12.7-2015.12.13)] 2015 IEEE International Conference on Computer Vision (ICCV) - Object Detection via a Multi-region and Semantic Segmentation-Aware CNN Model[J]. 2015:1134-1142.