Mechatronics and Automation Technology
J. Xu (Ed.)

doi:10.3233/ATDE221150

Safety Helmet Monitoring of Power Grid Staff Based on Improved YOLOv3

Fanxing RAO[1], Xianchang XI, Shaobo WU, Jing SHENG and Yang ZHANG
School of Electrical Engineering, Nanchang Institute of Technology, Nanchang, China

Abstract. Aiming at the problem that the staff do not wear safety helmets or even wear safety helmets in the construction and maintenance of new energy charging stations, an improved YOLOv3 safety helmet detection method is designed in this paper. Firstly, the 128 * 128 feature map output is added based on the three feature map output of YOLOv3 algorithm, and then the FI module is added to fuse the four scale feature information to improve the detection accuracy of small targets. Firstly, on the basis of the 3 feature map outputs of the YOLOv3 algorithm, 128*128 feature map outputs are added, and then the FI module is added to fuse the information of the 4 scale feature information to improve the detection accuracy of small targets; Secondly, the DIoU function is used to optimize the boundary frame loss function, so that the boundary frame regression is more accurate. Compared with the original algorithm YOLOv3, the proposed algorithm has higher detection accuracy, shorter time consuming and meets the requirements of real-time detection. Compared with other advanced algorithms, the robustness, detection accuracy and detection speed of the proposed algorithm are better, which can provide technical reference for avoiding staff safety hazards.

Keywords. New energy security,target detection, safety helmet, improved YOLOv3.

1. Introduction

The wind power industry is an industry prone to accidents, and its safety in production has always been an important aspect of national safety supervision. For example, in the hoisting of wind turbines in wind farms, the construction of line foundations in photovoltaic fields, and the operation and maintenance of new energy charging stations, accidents may occur if the construction team leader and construction personnel relax a little. Therefore, the technical means of safety supervision need to be improved. As a commonly used personal protective equipment for industrial safety, the correct wearing of the safety helmet plays a vital role in the safety of construction workers. However, in a variety of complex work scenarios, staff often do not correctly wear safety purchases for some reason, or even don't wear safety helmets, which will not only threaten the personal safety of staff, but also affect the smooth development of construction site supervision. Therefore, the wearing supervision of the safety helmet is an indispensable part of the wind power industry.Second paragraph.

Most of the traditional supervision measures for safety helmets are carried out manually, but with the increase of monitoring time and the increase of monitoring scope, the traditional supervision method not only makes people very easy to become visually

[1] Corresponding Author, Fanxing RAO, B No. 289, Tianxiang Avenue, high tech Development Zone, Nanchang City, Jiangxi Province; E-mail: 7219958@qq.com.

fatigued and leads to misjudgment, but also It consumes a lot of material and human resources and cannot meet the current requirements for safety management in the wind power industry. Combined with the target detection technology in computer vision, the intelligent algorithm model is used to detect the wearing of the safety helmet, which can overcome the defects of manual detection. At present, the commonly used target detection network is generally divided into two categories. One is the target detection based on the classification idea. First, the candidate region is generated, and then the classifier is used to classify the candidate box. The commonly used algorithms are R-CNN, FastR-CNN and FasterR-CNN[1]. Another category is target detection based on regression prediction, one-stage detection network represented by YOLO and SSD[2] algorithms. The output results of the regression of the object detection box can be obtained by inputting the image. Reference[3] proposed R-CNN algorithm, which improves the detection rate of PASCAL VOC data set from 35.1% of the traditional detection rate to 53.7% through convolutional neural network. Reference[4] proposed SPPNet method, which is improved based on R-CNN, and the detection speed is 38 ~ 102 times faster than R-CNN. Reference[5] further optimized Fast R-CNN, using VGG16 network instead of AlexNet network. The training speed is 9 times higher than that of R-CNN, and the detection speed of each image reaches 0.3s. Literature[6] proposed YOLO algorithm based on CVPR, which greatly improves the detection speed by predicting and training end-to-end. Literature[7] proposed YOLOv2 based on YOLO algorithm, and added K-means clustering to further improve the detection speed, detection accuracy and recognition category. Literature[8] Based on YOLOv2 algorithm and combined with multi-scale prediction (FPN) network structure, proposed YOLOv3 algorithm after aggregation and convolution, which greatly improved its detection accuracy. Reference [9] uses ResNet50 network and deformable convolution module to improve SSD algorithm to detect safety helmets and reflective clothing, but there are multiple boundary boxes overlapping. By modifying the feature layer scale of YOLOv3, the detection accuracy of the model to the target is improved.

To sum up, the existing algorithms sacrifice the detection speed to a certain extent while improving the accuracy of helmet small target detection. This paper proposes a helmet detection method based on improved YOLOv3 algorithm. Based on the original YOLOv3 network, the feature scale is added to extract shallow details, and the results are compared and analyzed through testing.

2. Principle of YOLOv3 Algorithm

In the network structure of YOLOv3, Darknet53 structure is used for feature extraction. The structure is composed of a series of 1 * 1 and 3 * 3 convolution layers. Through 5 times of down sampling, the output after 32, 16 and 8 times of down sampling is taken as the final extracted feature information, including feature maps of 13 * 13, 26 * 26 and 52 * 52 scales. Local feature interaction is realized in the form of convolution kernel in each scale[10]. In multi-scale feature detection, using FPN architecture for reference, the extracted feature information is fused by up sampling to obtain the prediction features of large, medium and small sizes. The total loss function of YOLOv3 is composed of three parts, and its expression is as follows.

$$L_{\text{box}} = \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{i,j}^{obj} (2 - \omega_i \times h_i)[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (\omega_i - \hat{\omega}_i)^2 + (h_i - \hat{h}_i)^2] \quad (1)$$

$$L_{cls} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{i,j}^{obj} \sum_{c \in class} p_i(c) \log(\hat{p}_i(c)) \quad (2)$$

$$L_{obj} = \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{i,j}^{noobj} (c_i - c_j)^2 + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{i,j}^{obj} (c_i - c_j)^2 \quad (3)$$

$$Loss = L_{box} + L_{cls} + L_{obj} \quad (4)$$

In the formula: indicates the size of the output feature map; represents the number of detection frames; represents the center coordinate value, height and width of the prediction detection frame; epresents the center coordinate value, height and width of the real detection frame; represent the prediction probability and real probability of category respectively; represents the total loss value.

3. Improved Algorithm of YOLOv3

3.1. Improved YOLOv3 Network Optimization

Aiming at the influence of complex working environment and the disappearance of small targets, an improved YOLOv3 network model is proposed, and the fourth characteristic scale is added: 104 * 104. The feature layers detected by four scales are fused. By fusing the feature layers of four scales, we can not only extract more shallow information, but also be compatible with targets at different scales[11].

3.2. Loss Function Optimization of YOLOv3

The specific method is to increase the original feature scale of 52*52 to 104*104 by upsampling the feature layer with the detection scale of 13*13 in YOLOv3 by 2 times. At the same time, the 109th, 85th, and 97th layers and the 11th, 61st, and 36th layers in the feature extraction network are respectively fused through the route layer, and the feature map values are shown in table 1.

Table 1. Value of Characteristic Diagram

layer	Size	Preset number of bounding boxes
layer1	13*13	13*13*3
layer2	26*26	26*26*3
layer3	52*52	52*52*3
layer4	104*104	104*104*3

When the input image gets the feature map through the feature extraction network, in order to avoid the loss of small target information in the image, this paper adds fi module (fusion information) before the input prediction of the four feature maps. For example, for the deep feature map of 13*13, the down sampling with scales of 104*104, 52*52 and 26*26 is used to 13*13, and then combined with its concat, so as to improve the positioning accuracy of small targets. The FI module structure is shown in the figure 1.

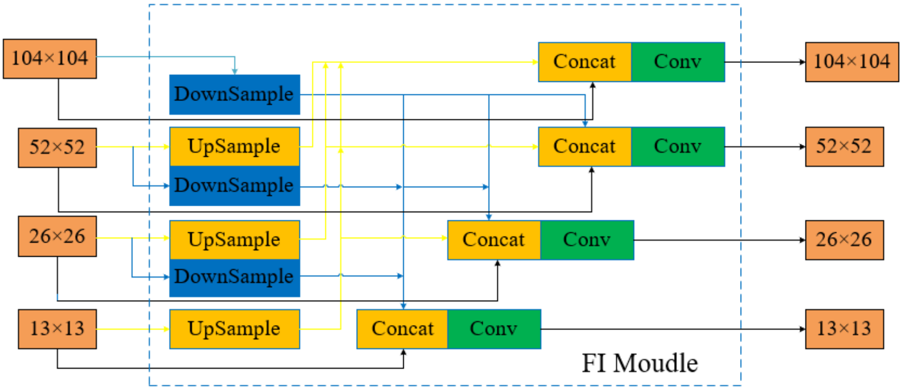

Figure 1. FI Moudle

In feature fusion, self-information should be used as the main information, and other scale information should be supplemented as auxiliary information. Therefore, this paper introduces the Spatial Pyramid (SPP) to extract feature information by block pooling of the feature map. By using the SPP layer to realize the fusion of local features and global features, the ability to extract feature maps is enriched and the phenomenon of overfitting is effectively prevented. The improved network structure is shown in figure 2.

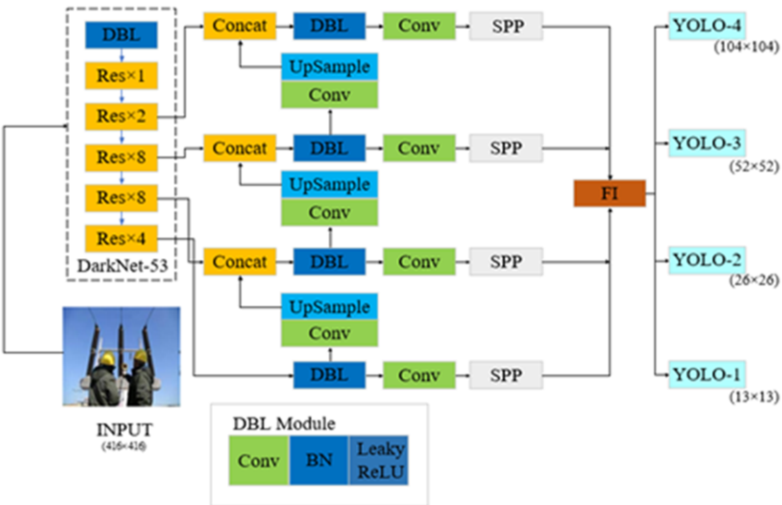

Figure 2. The improved model structure of YOLOv3

From the improved network structure shown in figure 2, it can be seen that the output detection layer of the four features of the improved network structure contains deep semantic information and shallow feature information, and compared with the original YOLOv3 network, it can make better use of shallow Layer information, while increasing the detection ability of small targets, preserves the detection ability of large targets and medium targets.

3.3. Loss Function Optimization

At present, most target detection algorithms use the intersection union ratio (IOU) as the measurement standard, but the loss function of the original YOLOv3 uses the root mean square error (MSE) function, but the MSE function only represents the distance relationship between the two frames and cannot represent their interaction. At the same time, because IOU has the characteristics of scale invariance, it can not only measure the distance between targets, but also characterize the degree of overlap of targets. Therefore, taking IOU as a loss function can better reflect the quality of the bounding box than MES.

When the target frame intersects with the prediction frame, the value of IOU can reflect the prediction effect. The higher the value, the higher the prediction accuracy. However, when the target value does not intersect with the prediction frame, its value is displayed as 0, which cannot show the distance between them. In this paper, IOU is optimized by the ratio of the lead out distance to the diagonal length of the minimum bounding box, and a new penalty function DIou[12] is proposed. Its calculation formula is as follows:

$$DIoU = \frac{\rho^2(d, d^{gt})}{c^2} \tag{5}$$

In the formula: ρ represents the Euclidean distance between two frames; d, d^{gt} respectively represent the distance between the predicted frame and the midpoint of the boundary of the real frame.At this time, the formula of its loss function is as follows:

$$Loss = 1 - DIoU \tag{6}$$

4. Experiment and Analysis

4.1. Preparation of Experimental Environment

The experimental platform built in this paper: the server CPU is Intel Xeon gold 6240r, and the installation system is Ubuntu 20 04 operating system, the GPU is NVIDIA tasla M40 * 2, and is equipped with two 12gb graphics cards. CUDA, cudnn and python3 are installed 8 and pytorch framework. The depth framework is based on the darknet-53 framework of the original YOLOv3.

4.2. Preparation of Data Sets

The data set is made through field collection and network crawling, and a total of 2154 helmet wearing sample databases are jointly constructed, including "wearing helmet", "not wearing helmet" and "not wearing helmet correctly". The corresponding labels and images are 617 "y_helmet", "1084" nohelmet "and 453" f_helmet ", as shown in the figure 3. Among them, for the problem of whether to wear the helmet correctly, the lower impeachment belt is detected to determine whether to wear the helmet correctly.

Figure 3. Safety helmet classification detection

4.3. Figures and Tables

There are many evaluation criteria for the effect of target detection, and the more common ones are intersection and merger ratio, recall, accuracy, average accuracy (AP) [13] and detection speed (FPS). In order to evaluate the improved model more accurately and in real time, the average accuracy (AP) and detection speed (FPS) are selected as the evaluation indexes. The definition of AP evaluation index is shown in the formula:

$$AP = \frac{\sum Precision_C}{images_C} \tag{7}$$

In the formula: $\sum Precision_C$ is the sum of all accuracy rates of this class in the validation set. $images_C$ is the number of images that contain targets in this category. FPS is the number of pictures that can be processed per second, as shown in the formula:

$$FPS = \frac{N}{T_n - T_s} \tag{8}$$

In the formula: N indicates the total number of pictures processed; T_n represents the time at the end of the algorithm; T_s represents the time when the algorithm starts running.

4.4. Analysis of Experimental Results

In order to verify the effect of the improved YOLOv3 algorithm in helmet wearing detection, fast CNN and YOLOv3 algorithm models are constructed for comparative analysis. The test is carried out under the same equipment, parameters and samples. The comparison results are shown in table 2.

Table 2. The training results of the 3 algorithms

Algorithm	AP(100%)			FPS
	Y_helmet	F_helmet	nohelmet	
YOLOv_3	90.23	92.42	97.21	20
Faster R_CNN	92.45	93.37	95.56	1.6
Optimize YOLOv3	92.40	93.45	96.71	12

It can be seen from table 2 that the YOLOv3 algorithm has a higher detection AP value on the helmet class, but a lower AP value on the other two categories; while the Faster R-CNN algorithm has a lower detection AP value on the F_helmet class than the other two. high. According to table 2, combined with the detection speed (FPS) of the model, the FPS value of the improved YOLOv3 algorithm is 12, and the AP value of the response is also high, which meets the real-time detection requirements of helmets during construction in the wind power industry. Therefore, considering the detection accuracy (AP) and detection speed (FPS) comprehensively, the detection effect of the improved YOLOv3 algorithm is the best.

Figure 4. YOLOv3 detection result graph

Figure 5. Faster R-CNN detection result

Figure 6. Optimize YOLOv3 detection result

The detection effect of the above three algorithms is verified by using a single picture, and the detection effect is shown in the figure 4-6. According to the detection results shown in the figure, the improved YOLOv3 algorithm has higher accuracy and less false detection and missed detection.

4.5. Robustness Verification

To verify the robustness of the model. Carry out experimental verification in multiple scenarios. Select but not limited to pictures of high-altitude operation, equipment maintenance, building construction and other aspects to verify the detection performance of the improved YOLOv3. The results are shown in the figure 7.

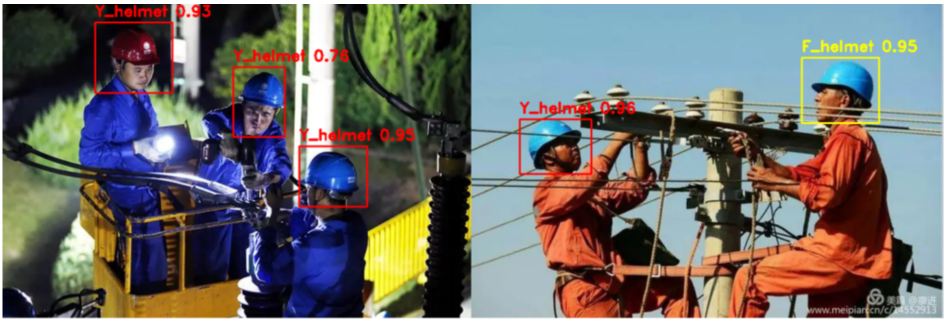

Figure 7. Improved detection results of YOLOv3 in different scenarios

It can be seen from the figure 7 that the improved YOLOv3 has better recognition accuracy in multiple scenes. It is proved that the improvement effect of the detection model is good.

5. Conclusion

In this paper, an improved YOLOv3 algorithm is proposed for the behavior of workers who do not wear helmets and do not wear helmets regularly. Through the comparative experiment, the horizontal comparison of the difference between the improved YOLOv3 and the standard YOLOv3 and fast r-cnn algorithm proves that the proposed improved algorithm has high accuracy and detection accuracy. Through the robustness verification, it proves the robustness of the algorithm and achieves the purpose of enhancing the image detection accuracy, but its speed improvement is limited. Therefore, it is necessary to make reasonable adjustments to the samples to further improve the recognition ability of the algorithm in complex situations.

Acknowledgments

This work was supported by Jiangxi Provincial Department of Education Science and Technology Project(GJJ211918), National Natural Science Foundation of China(51865031), Funded by the Natural Science Foundation of Jiangxi Province 20212BAB204040 (20212BAB204040), National Natural Science Foundation of China Regional Project(51867019).

References

[1] 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, 2017, 39(6):1137-1149.
[2] B, Crooks VC, Buckwalter JG, Chiu V. Blood pressure levels before dementia. Arch Neurol. 2005 Jan;62(1):112-6.
[3] Girshick R , Donahue J , Darrell T , et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. IEEE Computer Society, 2013.
[4] 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.
[5] 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, 2017, 39(6):1137-1149.
[6] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[J]. IEEE, 2016.
[7] Redmon J, Farhadi A . YOLO9000: Better, Faster, Stronger[J]. IEEE, 2017: 6517-6525.
[8] Redmon J, Farhadi A . YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.
[9] Han, et al Detection algorithm of safety helmet reflective clothing based on improved SSD [J] Automation and instrumentation, 2021, 36(09):63-68. DOI:10.19557/j.cnki.1001-9944.2021.09.014.
[10] He K , Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. IEEE, 2016.
[11] Wu Zhe.Research on ship detection, recognition and tracking in channel under dynamic background based on deep learning [D] Three Gorges University.
[12] Zheng Z, Wang P, Liu W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[J]. arXiv, 2019.
[13] Everingham M, Gool L V, Williams C, et al. The Pascal Visual Object Classes (VOC) Challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.