2020 16th International Conference on Computational Intelligence and Security (CIS) | 978-1-6654-0445-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/CIS52066.2020.00041

Safety helmet detection method based on YOLO v4

Deng Benyang[1], Lei Xiaochun*[1,2] and Ye Miao[3]
1.School of computer and information security,
Guilin University of Electronic Technology, Guilin 541004, China;
2.Guangxi Key Laboratory of Image and Graphic Intelligent Processing,
Guilin 541004, China;
3.School of information and communication,
Guilin University of Electronic Technology, Guilin 541004, China;
*Corresponding author's e-mail: glleixiaochun@qq.com

***Abstract*—In the actual production process, safety accidents caused by workers not wearing safety helmets often occur. In order to reduce safety accidents caused by wearing helmets, a helmet detection method based on improved YOLO v4 is proposed. By collecting a self-made data set of on-site construction site video, using the K-means algorithm to cluster the data set in getting appropriately a priori frame dimensional center and obtaining more targeted edge information. Subsequently, a multi-scale training strategy is used in the network training process to improve the adaptability of the model from different scales of detection. The experimental results show that, in the helmet wearing detection task, the model mAP value reached 92.89%, the detection speed reached 15f/s, and its detection accuracy and detection speed were improved compared with YOLO v4, which satisfies the real-time requirements of the helmet detection task.**

***Index-Terms*—safety helmet detection; neural network; target detection; YOLO v4**

I. Introduction

Wearing helmets is an important condition to ensure the safety of workers. At present, there are two methods, manual inspection and video monitoring, for the detection of helmet wearing at construction sites. However, these supervision methods require a lot of cost on labor. Therefore, a real-time detection of whether workers in surveillance videos wear safety helmets is of great significance to safety supervision.

In recent years, deep learning is a new research direction in machine learning. In the field of target detection, Redmon J[1] proposed the YOLO detection algorithm, which discarded the intermediate steps of the candidate region and used a single convolutional neural network. Regressing each bounding box directly and predicting the probability of the corresponding category, the final detection speed reached 45f/s while ensuring the detection accuracy. On the basis of YOLO, Redmon successively proposed YOLO v2[2] and YOLO v3[3] detection algorithms which improved the detection effect and detection time. In 2020, Alexey A.B.[4] took over Redmon and made improvements on the basis of YOLO v3. The detection effect was further improved. It achieved 43.5% AP on the COCO data set, and the real-time speed can reach 65FPS. What's more, the detection accuracy increased nearly 10 percentage points compared with YOLO v3. Compared to the EfficientDet[77], the YOLO v4 spends half the time to achieve the same performance as 43AP. It can be seen that in the field of target detection, YOLO v4 can maintain accuracy and detection speed at the same time and achieve better detection results.

Safety helmet wearing detection includes a type of target detection problem. At present, there are not many researches related to helmet detection. Yan Rongrong[5] use a combination of skin color segmentation and Adaboost algorithm to locate human faces and detect helmets. Feng Guochen[6] used OpenCV color recognition related methods to study the automatic recognition of helmets. Zhang Bo[7] used OpenPose to locate the human head and neck position to generate sub-images, and used Faster R-CNN to detect the helmet in the sub-images. Most of the existing methods are based on traditional target detection methods, and most of them have problems such as slow detection speed and detection accuracy. What's more, they are greatly affected by the environment, making it difficult to meet the needs of the actual complex and changeable industrial environment.

This paper uses the video data of the construction site and the pictures obtained by the auxiliary web crawler to create a helmet detection data set. Firstly, based on the YOLO v4 model, performing K-means clustering analysis on the dimensions of the target frame obtains more accurate dimensional frame information, so that the model can obtain more edge information of the target object. Then using multi-scale images during the training process training enables the model to adapt images of different resolutions. Finally, train multiple times obtain the optimal parameters during model training. Experiments show that the improved YOLO v4 algorithm can improve the detection rate while ensuring the detection accuracy.

II. Safety helmet detection method

A. YOLO v4 Introduction

The YOLO v4 algorithm uses a new CSPDarknet53 neural network as the main model of the detector. YOLO v4 contains 29 convolutional layers of 3x3,725x725 receptive fields and 27.6M parameters. The SPP block is added to CSPDarknet53 which significantly increases the receptive field and separates contextual features. At last, PANET is used instead of FPN in YOLO v3 as parameter aggregation methods for different detector levels.

In addition, YOLO v4 also proposes some new improvement strategies:

1) Apply a new data augmentation method: mosaic and adversarial training;

2) Select the optimal hyperparameters through GA;

978-1-6654-0445-7/20/$31.00 ©2020 IEEE
DOI 10.1109/CIS52066.2020.00041

3) Improve existing methods to be more suitable for efficient training and inference: improve SAM and PAN.

B. Improved Method

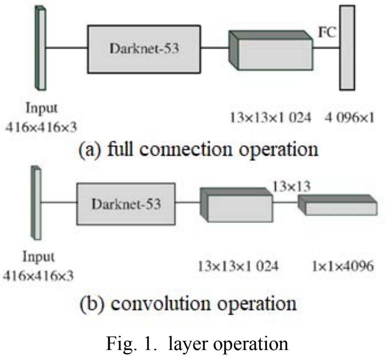
YOLO v4 has achieved good detection results in the field of target detection, while target detection for YOLO v4 was implemented on common data sets. As to specific helmet w-earing detection data sets, some improvements on YOLO v4 are needed. In order to adapt specific helmet inspection tasks, the improved method mainly includes the use of K-means algorithm for dimensional clustering and multi-scale training.

*1)K-means dimensional clustering algorithm:*YOLO v4 draws on the method of using a priori boxes to predict boundary coordinates in YOLO v3, using the K-means algorithm to predict 9 priori boxes, and larger-scale feature maps use a smaller priori boxes to obtain more target edge information.

The 9 sets of prior frame dimensions calculated by YOLO v4 based on the COCO data set are: (12, 16), (19, 36), (40, 28), (36, 75), (76, 55), (72, 146), (142, 110), (192, 243) , (459, 401). In the actual helmet wearing detection task, the priori frame dimension calculated by YOLO v4 is not suitable for the helmet wearing detection scenario. It is difficult to obtain accurate target frame information using the original prior frame dimensions of YOLO v4.

Therefore, in the helmet wearing detection scenario, the K-means algorithm is used to cluster analysis on the self-made helmet wearing data set, and 9 sets of priori box dimension centers are obtained, respectively: (7, 15), (9, 19), (11, 23), (15, 29), (21, 38), (30, 51), (44, 74), (69, 118), (128, 212).To use this cluster center Network training for helmet wearing detection can obtain better result.

2)Multi-scale Training: For the self-made helmet detection data set, the input images have different sizes. Therefore, in order to enhance the model's robustness to image size, a multi-scale training strategy can be adopted. Specifically, the fully connected layer is removed from the YOLO v4 network and changed to a fully convolution operation, because the existence of the fully connected layer makes the size of the input picture of the network, which must be fixed during the training process. Figure 1 shows the process of converting a fully connected layer into a convolutional layer.



(a) full connection operation

(b) convolution operation

Fig. 1. layer operation

Figure 1(a) uses the fully connected layer to make predictions. When the input image size is 416x416, after passing through the Darknet-53 network, the feature map of 13×13×1024 is output. Figure 1(a) passes through a fully connected layer containing 4096 neurons to obtain a set of 4096×1 Feature map; Figure 1(b) uses 4096 13×13 convolution kernels to finally get a 1×1×4096 feature map, which is essentially equivalent to 4096 neurons obtained by a full connection. For the above two network structures, when the picture input size is 416×416, the network can operate normally, but there are other sizes of pictures in the helmet self-made data set. For example, when a picture of 608×608 size is input, the network can be operated normally. After the network output a 19×19×1024 feature map, for the structure in Figure 1(a), the feature map needs to be fully connected with 4096 neurons. Since the size of the original architecture is 13×13 which is now replaced by 19 ×19, the network cannot use the previous parameter matrix during the propagation process, and cannot operate normally. As for the structure in Figure 1 (b), after the full connection is changed to convolution operation, the network can continue to run and get the correct output of 7×7×4 096. Therefore, after changing the fully connected layer to a full convolution operation and training with input images of different sizes, the improved algorithm can adapt to test images of different sizes.

Based on the above analysis, this article uses a multi-scale training strategy to train the self-made helmet data set. Since the entire network has 5 maximum pooling layers, the subsampled rate of the network is 32. During the training process, the input size of the training image of the helmet wearing dataset is divided into a series of 32 multiples. The size calculation formula is:

$$S_n = 32n, n \leq 9 \qquad (1)$$

Among them, S_n is the size of the input image. During network initialization, S_1 is 320 x 320.

From equation (1), it can be concluded that the size of the input image is: {320, 353..., 608}. In the training process, an input image size is randomly selected every 10 rounds to achieve the model that can adapt to different sizes of images effect. The multi-scale training detection process is shown in Figure 2。

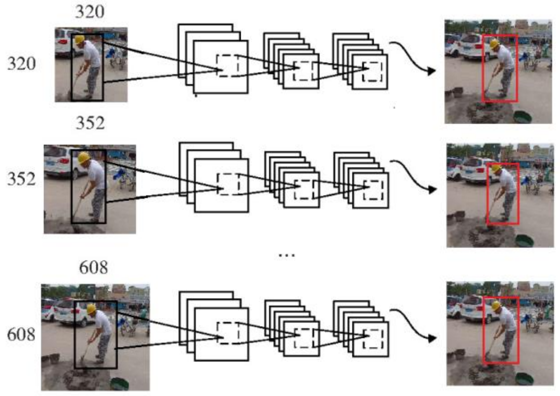


Fig. 2. Multiscale training detection

III. Safety helmet detection method

A. Experiment Platform

TABLE I System Experiment Environment

CPU	Intel(R)i7-9750H CPU@2.60GHz
RAM	16G
GPU	Nvidia GeForce GTX 1660Ti 6GB
Programing Language	Python 3
Deep Learning Framework	Tensorflow 2

B. Model Training

This paper uses the weight parameters provided by the YOLO v4 official website as the initialization parameters for network training. It also uses the construction site video and supplemented by the web crawler pictures as the data set for multiple training. The fine-tune network parameters optimize the overall network training effect. Some experimental parameter settings are shown in Table 2. The entire model training process is shown in Figure 3. When the number of iterations reaches 15,000, the average loss function avg loss value is 3.6782 and the model training effect is at the best.

TABLE Ⅱ Network Parameter Description

Learning Rate	0.001
Epoch	100
Iterations	16000
Batch Size	16
Lr_step	80
Lr_factor	0.1

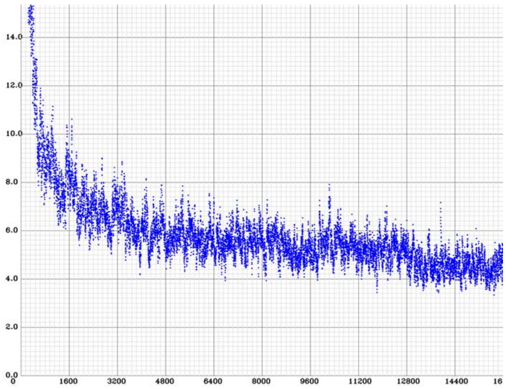


Fig. 3. Model Training Process

C. Model Performance Analysis

1) Performance test under different target sizes: In order to verify the detection effect on different target detection performance, the test data set is sorted according to the target size. The targets with 0-25%, 25% - 50%, 50% - 75% and 75% - 100% of the target size are divided into four sub categories A, B, C and D, which represent the sizes of different targets. As shown in Figure 4, the performance data of the original and improved Yolo V4 algorithm for different size targets are listed.

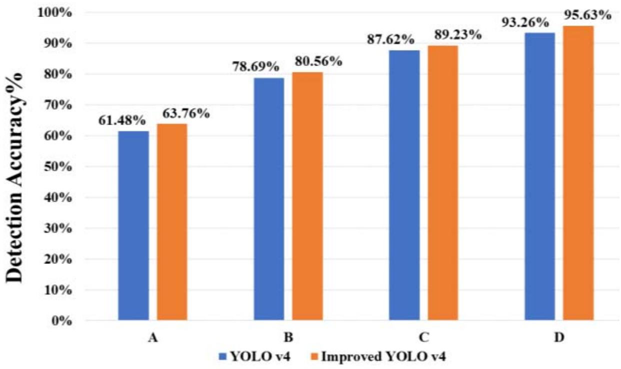


Fig. 4. Multiscale detection effect of two algorithms

As seen from Figure 4, the detection accuracy of the improved Yolo V4 algorithm for different target sizes is higher than that of the original Yolo V4 algorithm. Therefore, the use of multi-scale training strategy makes the whole algorithm achieve good results in the detection of small targets.

2) Comparative experiment of safety helmet wearing detection model: In terms of model performance analysis, this paper uses the mean Average Precision (mAP) of each category as the evaluation index for the quality of the model, and conducts experiments on the improved YOLO v4 algorithm. At the same time, Faster R-CNN, YOLO v3 and the original YOLO v4 algorithms are used to compare with the improved YOLO v4. mAP and Frames Per Second (FPS) are used as the detection effect evaluation indicators. The experiment results are shown in Table 3.

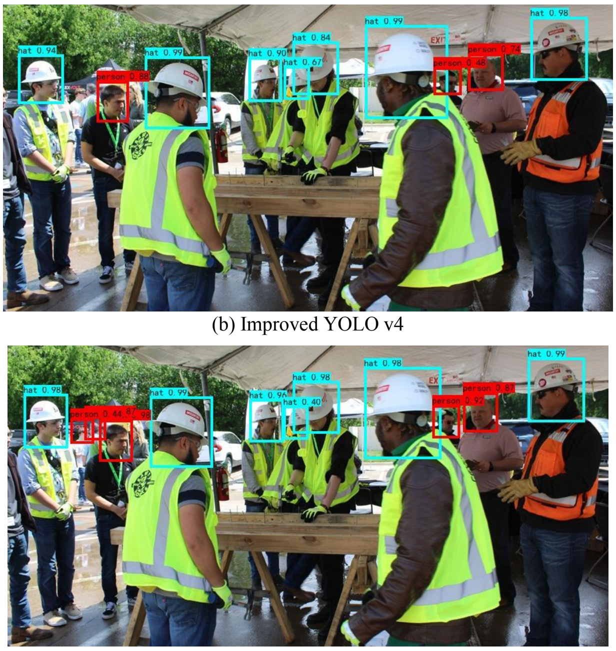
TABLE III Experimental Results Comparison

Algorithm	mAP/%	FPS/(f/s)
Faster R-CNN	**94.82**	0.2
YOLO v3	86.24	12.0
YOLO v4	90.66	15.0
Improved YOLO v4	92.89	**15.0**

D. Result Analysis

It can be seen from the experimental results that Faster R-CNN has the highest mAP value which is reaching 94.82%, and the improved YOLO v4 algorithm mAP value is 92.89%. Although the detection accuracy is slightly lower than the Faster R-CNN, its detection speed is 75 times faster than it. Both YOLO v3 and the original YOLO v4 are inferior to the improved YOLO v4 algorithm in detection accuracy. It can also be seen that the improved YOLO v4 takes into account both the detection accuracy and the detection speed, and can be better applied in the actual environment to achieve the task of detecting helmet wearing.

In addition, in order to intuitively feeling the detection differences between different algorithms, this paper selects a set of detection pictures for analysis. Figure 5 is the detection effect of the algorithm on the actual construction site environment. In this figure, the blue box indicates the person wearing the helmet and the red box indicates the person without wearing the helmet. It can be seen that there are some missed detection objects in Picture a, and the confidence value is low. The improved YOLO v4 algorithm in Picture b has a good detection effect and a high confidence. In summary, the improved YOLO v4 algorithm proposed in this paper can meet the real-time detection requirements while maintaining a high detection rate.



(a) YOLO v4

(b) Improved YOLO v4

Fig. 5. Test Scene

Summary

This paper proposes a method for detecting helmet wearing based on the improved YOLO v4. It uses the construction site video supplemented by web crawler pictures as a data set to carry out the helmet wearing detection test. Then improve the YOLO v4 network by using target frame dimensional clustering and multi-scale training methods to ensure higher accuracy. Meanwhile, this still has a fast speed, which can meet the accuracy and real-time requirements of the helmet wearing detection task in the actual industrial environment.

Acknowledgement

This work is supported by Project to Improve the scienti -fc research basic ability of Middle-aged and Young Teachers (No.2019ky0222,2020KY05033),the Open Funds Guang -xi Key Laboratory of Image and Graphic Intelligent Processing, No.GIIP2004, Student’s Platform for Innovation and Entrepreneurship Training Program under Grant (No.201910595093，201910595142).

References

[1] 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, 2015:779-788.

[2] Redmon J, Farhadi A.YOLO9000:better, faster, stronger[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017:6517-6525.

[3] Redmon J, Farhadi A. YOLOv3: an incremental improvement[J].IEEE Conference on Computer Vision and Pattern Recognition, 2018:89-95.

[4] YOLOv4: Optimal Speed and Accuracy of Object Detection. BOCHKOVSKIY A,WANG C Y,LIAO H Y M. https://arxiv.org/abs/2004.10934.2020

[5] Yan Rongrong. Research on safety helmet detection algorithm for industrial scene [D]. Xi'an University of Technology, 2019.

[6] Feng Guochen, Chen Yanyan, Chen Ning, Li Xin, Song Chengcheng. Research on Automatic Identification Technology of Safety Helmet Based on Machine Vision[J]. Machine Design and Manufacturing Engineering, 2015, 44(10): 39-42.

[7] Zhang Bo, Song Yuanbin, Xiong Ruoxin, Zhang Shichao. Detection of wearing helmets fused with human joint points[J]. Chinese Journal of Safety Science, 2020, 30(02): 177-182.

[8] Liu Xiaohui, Ye Xining. Application of skin color detection and Hu moment in helmet recognition[J]. Journal of East China University of Science and Technology, 2014, 40(3).

[9] Research on target vehicle detection based on improved SSD algorithm[J].Chen Bingqu, Deng Tao. Journal of Chongqing University of Technology (Natural Science). 2019(01)

[10] Application of improved convolutional neural network in pedestrian detection[J]. Xie Linjiang, Ji Guishu, Peng Qing, Luo Entao. Computer Science and Exploration. 2018(05)

[11] Safety helmet recognition based on parallel two-way convolutional neural network [J]. Huang Yuwen, Pan Difu. Enterprise Technology Development. 2018(03)