Journal of Engineering Science and Technology Review 15 (1) (2022) 1 - 7

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Research Article

Detection Method of Helmet Wearing in Complex Scenes

Yuan An[1,*], Zhenyang Li[1], Lei Chen[1] and Haotian Chen[1,2]

[1]Xu Zhou University of Technology, Xuzhou 221008, China
[2]Illinois Institute of Technology, Chicago,IL60616,United States

Received 2 January 2022; Accepted 7 March 2022

Abstract

In complex scenes, the standard wearing of helmets is the basic key to ensuring personnel safety, its effective detection is affected by the small targets, dense targets, and mutual occlusion. This study proposed a detection algorithm based on the You Only Look Once version 5(YOLOv5) algorithm to detect the wearing of helmets in complex scenes effectively. Data enhancement and other processing on the data set were performed to determine the personnel area and obtain personnel location information. The complete downsampling information was preserved for subsequent feature extraction through Focus processing. Then, the head area of the subject person was extracted to expand the location area and improve the fault tolerance of the algorithm. Finally, given the overlapping area between the target frame and the detection frame, the Complete Intersection over Union (CIoU) loss function was selected for the center point distance and aspect ratio to improve the positioning accuracy. Moreover, a dynamic adjustment strategy was adopted for the learning rate during the training process. Results show that the algorithm can effectively detect and mark the wearing situation of the subject persons' helmets. The experiment is conducted in the test set of 2000 pictures. The accuracy rate of the experiment is 0.995, and the average response speed reaches 0.035 s. This study provides a technical reference for the detection of safety protective equipment among staff and the improvement of work safety factors in complex scene environments.

Keywords: Complex scene, YOLOv5,Helmet, Deep learning

1. Introduction

As a national pillar industry, coal mining has always been a high-risk industry. As a complex scene environment, the downhole environment has various complex conditions, such as poor lighting, undulating road surfaces, and water accumulation in roadways. Ensuring the life safety of underground workers and effectively reducing the probability of safety accidents have always been the focus of research in the technical field of this industry. In addition, postdisaster rescue is a complex scene environment. The disaster and rescue areas have various unpredictable and complex situations. Ensuring the life safety of rescuers is the primary task of a rescue mission. According to the analysis of safety accident cases in recent years, safety accidents or escalation of accidents caused by the improper wearing of helmets is not uncommon. The quick and effective detection and labeling of the wearing of helmets in complex scenes are the fundamental and direct measures to ensure personnel safety. However, the existing helmet detection models in complex scene environments gradually expose many problems, such as the slow execution speed of the model algorithm, the lack of labeling and alarm functions for the case of not wearing a helmet, and the high computing power requirements of hardware devices. Therefore, the study on a lightweight, easy-to-deploy, high-performance model for helmet detection algorithm that can simultaneously mark whether it is worn or not in a complex scene environment has great practical relevance and social application value.

The rise of neural networks has prompted scholars and research institutions to use the deep learning method for safety helmet detection. Target detection algorithms based on deep learning can be roughly divided into two categories. The first is the two-stage target detection algorithm, represented by the regional convolutional neural network (R-CNN) series of algorithms [1-5]. An algorithm in this category first generates candidate frames that may contain objects to be detected. Then, the algorithm performs position calibration and target classification on the candidate frame to obtain the final detection result. The algorithms in this category have high detection accuracy. However, the detection speed cannot meet the real-time requirements. The other category of algorithms is a single-stage target detection algorithm, represented by the Single Shot MultiBox Detector (SSD) and the You Only Look Once (YOLO) series of algorithms [6-7]. The algorithms in this category do not generate candidate frames but directly extract target features in the convolutional neural network (CNN) network to obtain target classification and location. The algorithms can detect rapidly, but the detection accuracy is poor. The target detection algorithms primarily focus on improving the detection accuracy of the single-stage target detection algorithms because the two-stage target detection algorithms cannot meet the real-time requirements. The YOLO algorithm was proposed in 2015 [7]. This algorithm abstracted the target detection task into a regression problem for the first time, which improved the detection speed. However, the problem of missed detection exists. YOLOv2, YOLOv3, YOLOv4, and YOLOv5 have been proposed successively to solve missed detection [8-11]. The YOLOv5 model added a Focus structure to the backbone network, thereby which achieved a new benchmark for the best

*E-mail address: anyuanbd@163.com
ISSN: 1791-2377 © 2022 School of Science, IHU. All rights reserved.
doi:10.25103/jestr.151.01

balance of detection speed and detection accuracy. The model was implemented in PyTorch and could be easily deployed.

The proposal of the YOLO algorithm has triggered scholars to advance target detection algorithms further. Xu et al. added feature maps in the YOLOv3 algorithm and improved the loss function, thereby enhancing the indicators of the algorithm to a certain extent [12]. Cheng et al. improved the convolutional layer of the YOLOv3 algorithm, and all indicators were considerably improved [13]. Xu et al. proposed a data augmentation algorithm based on the YOLOv4 algorithm, thereby improving the ability of the model to detect small targets [14]. Zhang et al. added a multispectral attention module to the neck of the YOLOv5 network to improve the generalization ability of the model [15]. The above research improved the performance of the algorithm model to a certain extent, but it is only employed in general scenarios. No relevant experimental research has been carried out in complex scene environments. Moreover, full labeling of helmet-wearing conditions has not yet been realized. Thus, coping with dense environments in complex scenarios is difficult.

Based on the above analysis, this study makes the following improvements to the YOLOv5 algorithm to achieve the full labeling of helmets in complex scene environments. The DenseBlock module to retain the complete information of the small target to the greatest extent replaces the slicing operation. The detection layer is added to improve the learning ability of the algorithm model in dense scenes. The algorithm fully marks the wearing of helmets to facilitate timely alarm and reduce the accident rate.

2. State of the art

Scholars have done a series of related research work on helmet detection and recognition, which is one of the application fields of target recognition. In 2013, Kelm et al. used mobile radio frequency to identify whether the protective equipment of passing personnel is standardized [16]. However, confirming whether the inspection personnel wore a helmet was impossible because of the limited working range of the radio frequency device reader. Target detection algorithms based on deep learning were proposed one after another with the introduction of CNNs. In 2014, Girshick et al. proposed the R-CNN, which employs the candidate region method instead of manually setting features to make great progress in the target detection algorithm [17]. In 2015, Girshick [3] and Ren [4] proposed Fast R-CNN and Faster R-CNN algorithms. These algorithms improved the detection accuracy and detection speed of the algorithm. In particular, the Faster R-CNN algorithm realized the end-to-end network training. The target detection algorithm represented by the R-CNN series of algorithms needs to generate a large number of candidate regions. It also has high detection accuracy. However, the detection speed cannot meet the real-time requirements.

Many scholars have done a series of related research work on helmet detection and recognition, which is one of the application fields of target recognition. In 2013, Kelm et al. used mobile radio frequency to identify whether the protective equipment of passing personnel is standardized [16]. However, confirming whether the inspection personnel wore a helmet was impossible because of the limited working range of the radio frequency device reader. Target detection algorithms based on deep learning were proposed one after another with the introduction of CNNs. In 2014, Girshick et al. proposed the R-CNN, which employs the candidate region method instead of manually setting features to make great progress in the target detection algorithm [17]. In 2015, Girshick [3] and Ren [4] proposed Fast R-CNN and Faster R-CNN algorithms. These algorithms improve the detection accuracy and detection speed of the algorithm. In particular, the Faster R-CNN algorithm realizes the end-to-end training of the network. The target detection algorithm represented by the R-CNN series of algorithms needs to generate a large number of candidate regions. It also has a high detection accuracy. However, the detection speed cannot meet the real-time requirements.

Redmon J et al. proposed the YOLOv1 algorithm in 2015 [7]. This algorithm greatly improves the detection speed. However, some missed detections exist. In 2016, the SSD algorithm was proposed [6]. The performance of the SSD algorithm is better than that of the YOLOv1 algorithm. The SSD algorithm can maintain the detection accuracy at the level of the two-stage target detection algorithm while ensuring the detection speed. Xu et al. used the MobileNet network structure to replace the visual geometry group(VGG) feature extraction network structure in the SSD algorithm. They also formed an improved SSD algorithm for the helmet-wearing detection method [18]. However, the depth of target feature extraction is not suitable for complex scene environments because of the limited number of SSD network layers. In 2016, the YOLOv2 algorithm [8] and the YOLOv3 algorithm [9] were proposed, and the YOLOv3 algorithm performed the best in the target detection process. According to the author, the YOLOv3 algorithm can achieve a mean average precision (mAP) of 57.9% in 51 ms during the training process of the MS COCO data set. Compared with RetinaNet, the YOLOv3 algorithm improves the detection speed by nearly 3.8 times while guaranteeing similar detection. This finding shows that the YOLOv3 algorithm has good detection accuracy and detection rate in the field of target detection. Given the advantages of the YOLOv3 algorithm in target detection, many scholars have conducted in-depth research. Lin et al. transformed helmet detection into a single classification detection task and set the output as an 18-dimensional tensor to establish a new helmet detection model [19]. Wang et al. improved the detection accuracy by improving the generalized intersection over union (GIoU) calculation method in YOLOv3 for helmet detection [20]. Shi et al. constructed a new helmet-wearing detection method by adding an image feature pyramid structure to YOLOv3 to extract the multiscale features of helmets [21]. The above algorithm models can realize the automatic detection of helmets, and the detection accuracy meets the requirements. However, the related algorithm models are all large and complex networks, and the algorithms need to be deployed on mobile terminals or embedded devices in complex scenarios. The computing power of mobile terminals is not enough to support this large and complex network structure.

The YOLOv5 algorithm was launched by Ultralytics in 2020. The network weight file size of this algorithm is only one-ninth of the YOLOv3 network model, and the performance requirements on the mobile terminal are far lower than those of YOLOv3 [22]. The YOLOv5 model added a Focus structure to the backbone network, achieving a new benchmark with the best balance of speed and accuracy. After the YOLOv5 algorithm was proposed, many scholars have performed related algorithm verification and improvement work. Zhang et al. [15] proposed an improved YOLOv5 helmet detection method for the needs of construction sites by using the K-means++ algorithm to redesign the a priori frame size and introducing a spectral channel attention module into the feature extraction network. The average accuracy of the algorithm is improved by 3.3%. Jiang et al. [23] proposed a lightweight target detection network HourGlass-YOLO, suitable for deployment in mobile devices. The volume of the YOLOv5 model was reduced by 87% under the premise of ensuring accuracy. Zhao et al. [24] used the DenseBlock module in replacing the overall Focus structure of the backbone network to improve the detection ability of the YOLOv5 algorithm for small targets. They also introduced the Squeeze and Excitation(SE-Net) channel attention module in the network neck, which increases the average accuracy of the algorithm by 6.57%.

The above scholars have done much work in various aspects, such as improving the accuracy of the algorithm, reducing the size of the algorithm, and improving the generalization of the algorithm for common application scenarios. However, very few studies have been done on related work for complex scene environments. Moreover, most of the research work only marked the staff who wore helmets and did not mark the staff who did not wear helmets. In a complex scene environment, the focus of the algorithm should be the staffs who do not wear helmets as required. Therefore, based on the above research, the YOLOv5 algorithm network model is used in identifying the helmet-wearing conditions among staff in complex scene environments and marking them all to reduce fundamentally the incidence of safety accidents.

The remainder of this study is organized as follows. Section 3 describes the network model of the YOLOv5 algorithm. Section 4 analyzes the experimental results. Section 5 summarizes the article and makes an outlook for the follow-up work.

3. Methodology

3.1 General framework

The YOLO algorithm was proposed by Joseph Redmon and colleagues in 2015. This algorithm uses a neural network to convert target detection into a regression problem. It only needs to input the image into the neural network during training, obtain the algorithm model through the neural network, and then input the algorithm model into the image. The YOLO algorithm has been widely employed in the field of target detection. The YOLOv5 algorithm is the most advanced detection network in the current versions of YOLO algorithms. It is obtained by algorithm optimization based on the YOLOv3 and YOLOv4 algorithms, thereby improving the detection speed and detection accuracy. Launched in 2020, the YOLOv5 algorithm is implemented using PyTorch and can be easily deployed and implemented.

The YOLOv5 algorithm has four network models with different sizes, such as YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The larger the size of the algorithm model is, the higher the detection accuracy is and the slower the detection speed is. The YOLOv5s network has the smallest volume, whereas YOLOv5m, YOLOv5l, and YOLOv5x continue to deepen and widen based on YOLOv5s. In a complex scene environment, detecting whether the staff wears helmets in real time is necessary. Thus, the study chooses YOLOv5s as the backbone network. The four network structures of the YOLOv5 algorithm are shown in Fig. 1.

Fig. 1. Schematic diagram of the YOLOv5 network model structure

Based on the idea of anchor boxes, the YOLOv5 algorithm abandons the manual selection of anchor boxes, uses the K-means clustering algorithm to obtain good prior values on the bounding boxes, and automatically updates the anchor boxes. The YOLOv5 algorithm adopts the PyTorch framework. The network structure is divided into input, backbone network, neck, and prediction.

The input of YOLOv5 uses mosaic data enhancement in performing a series of processing on the data, including random scaling, random cropping, and random arrangement, to enhance the small target detection ability and improve the robustness of the network. The YOLOv5 algorithm obtains the predicted frame based on the initial anchor frame and then compares the predicted frame with the real frame. The algorithm updates the anchor frame on the basis of the difference between the predicted frame and the real frame. It iterates the network parameters to obtain the best anchor frame value.

The backbone is the backbone network of the YOLOv5 algorithm. The YOLOv5 algorithm adds the Focus module. Before entered into the backbone, the image is entered to the Focus module to slice the image and slice the original 608*608*3 image to a 304*304*12 image feature map. Then, a 32-convolution kernel feature map with complete information is formed through the convolution operation. The spatial pyramid pooling(SPP) is employed for maximum pooling of feature maps, and feature maps of different scales are spliced together to form a CNN of image features. Compared with the previous YOLO version, the YOLOv5s algorithm adopts two cross stage partial connection (CSP) structures in the backbone network, namely, CSP1 and CSP2 structures. CSP Darknet53 adopts CSP1_1 structure and CSP1_3 structure to enhance the learning ability of CNN, reduce memory cost, break through computational bottlenecks, and enhance CSP accuracy. The CSP2_1 structure is adopted in the fusion feature layer to strengthen feature fusion between networks and enhance network feature fusion capabilities.

The neck is the network fusion layer, which is between the backbone and head. It adopts a feature pyramid network (FPN) + pyramid attention network (PAN) structure. FPN is top-down, and PAN is the bottom pyramid. FPN uses upsampling to mix and combine the acquired top feature information and backbone network feature information. The combined information is passed onto the prediction layer.

At the prediction, the YOLOv5 algorithm predicts image features and generates bounding boxes and predicted classes. In this part, the YOLOv5 algorithm uses GIoU as the loss function of the bounding box. In overlapping target detection, the effect of GIoU-non maximum suppression (GIoU-nms)is stronger than that of the traditional nms.

The YOLOv5s network structure is shown in Fig. 2.

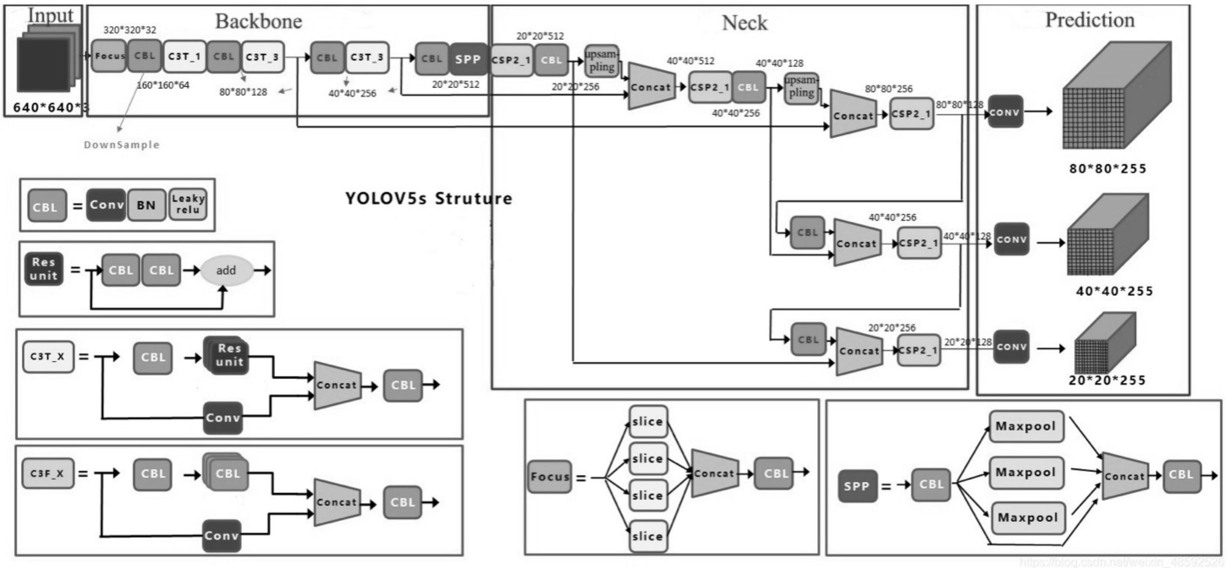

Fig. 2. The network structure of YOLOv5s algorithm

3.2 Particle representation

The loss function employed by YOLOv5 consists of GIoU loss function, target confidence loss function LossObj, and classification loss function. GIoU is calculated by intersection over union (IoU). IoU ratio is an indicator for the performance evaluation of the target detection algorithm. It is the ratio of the intersection and union between the predicted frame and the real frame. GIoU introduces a penalty term based on IoU to reflect the intersection of the detection frame and the real frame accurately. The calculation formulas are as follows:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

$$GIoU = IoU - \frac{|A_C - (A \cup B)|}{|A_C|} \tag{2}$$

In the formula, A is the detection frame, B is the real frame, and A_C is the minimum circumscribed rectangle area between the predicted frame and the real frame.

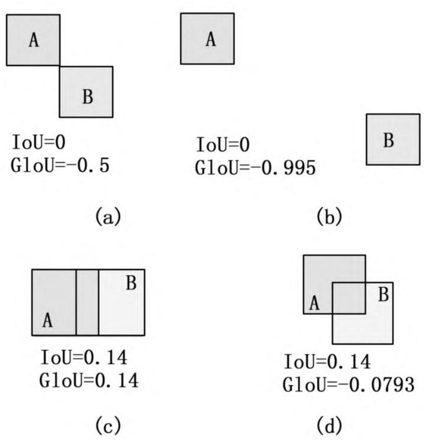

Fig. 3. The positional relationship between the real box and the predicted box

As shown in Figures 3(a) and 3(b), GIoU can measure the distance between boxes A and B when the two boxes do not intersect. Figures 3(c) and 3(d) show that GIoU can also reflect the intersection of boxes A and B when the two boxes intersect. Therefore, GIoU can better reflect the positional relationship between the detection frame and the real frame than IoU can. However, GIoU degenerates into IoU when a special relationship of inclusion exists and is included between the detection box and the real box. Therefore, this study uses CIoU_Loss as the loss function of the bounding box to make the predicted box fit the real box closely.

The calculation formula of CIoU is as follows:

$$CIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - av \tag{3}$$

In the formula, b and b^{gt} represent the center point of the predicted frame and the real frame, respectively, $d = \rho(b, b^{gt})$ represents the distance between the center points of the two frames, and c represents the diagonal distance between the predicted frame and the minimum circumscribed rectangle of the real frame, as shown in Fig. 4. a is the trade-off parameter, and v is employed to measure the consistency of the aspect ratio. The calculation formula of the two parameters is as follows:

$$a = \frac{v}{1 - IoU + v} \tag{4}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{5}$$

In the formula, w and w^{gt} represent the width of the predicted box and the ground truth box, respectively. h and h^{gt} represent the heights of the predicted and ground truth boxes, respectively.

The calculation formula of CIoU_Loss is as follows:

$$CIoU_{Loss} = 1 - CIoU \tag{6}$$

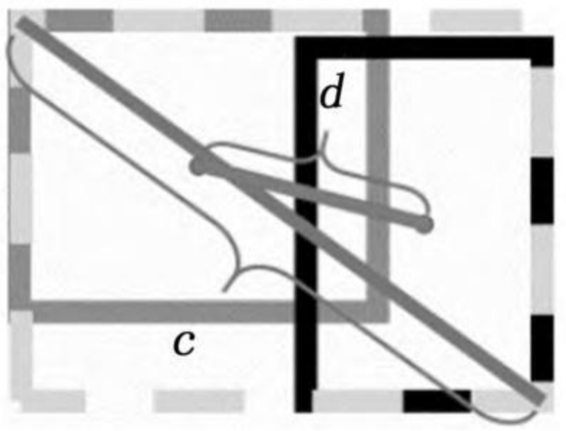

Fig. 4. Schematic diagram of CIoU parameters

4 Result Analysis and Discussion

Given the particularity of the complex scene environment, the evaluation indicators employed in this study are the average precision(AP), recall(R), and detection rate. Detection rate is the number of frames of processed images per second (frame per second [FPS]).

The calculation formula of the average accuracy is as follows:

$$AP = \int_0^1 P(R)dR \tag{7}$$

The calculation formulas of precision(P) and recall(R) are shown in Equations (8) and (9).

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

TP (true positive) represents the number of targets that are predicted to be positive and are actually positive. FP (false positive) represents the number of targets that are predicted to be positive but are actually negative. FN (false negative) represents the number of targets that are predicted to be negative but are actually positive.

The analysis of the YOLOv5 loss function indicates that the loss value of the algorithm consists of three parts, which represent the loss value of the object position, the object category, and whether the target contains the three categories of loss values, as shown in Formula (10).

$$loss = box\_loss + cls\_loss + obj\_loss \tag{10}$$

Fig. 5 shows the loss value convergence of function curve. The algorithm model has obtained a good fitting effect.

The initial learning rate of the network model in this study is set to 0.01, the learning rate period is 0.2, decay weight of the learning rate is 0.0005, the training number is 20, and the data set size is 2000.

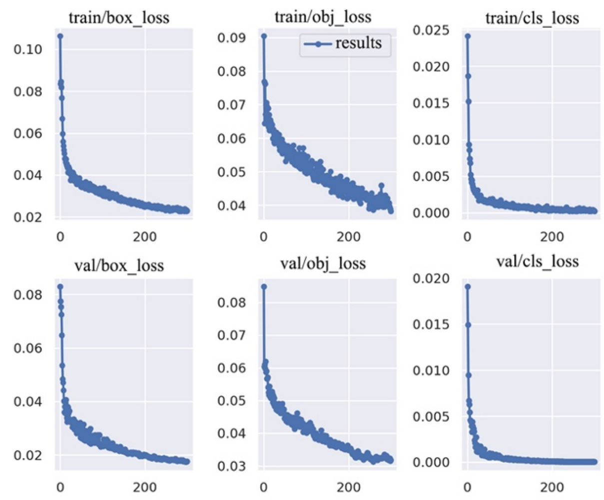

Fig. 5. Convergence curve for loss value of algorithm model

The regression curve of P rate, R rate, and mAP value obtained by the algorithm model is shown in Fig. 6.

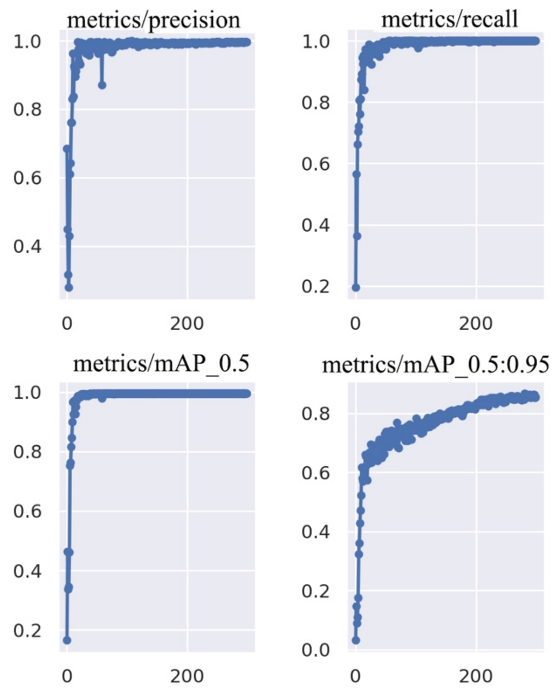

Fig. 6. Regression curve of algorithm evaluation index

The parameter of mAP can be employed to measure the recognition accuracy of the algorithm model. This value is obtained by averaging the AP values of all categories. The corresponding P-R curve is drawn by calculating the highest accuracy rate under different R values. The area enclosed by the curve and the horizontal axis is the AP value for this category. Figure 6 shows that the AP values of the P, R, mAP_0.5, and mAP_0.5:0.95 can reach up to 0.999, 0.998, 0.999, and 0.968, respectively, meeting the recognition accuracy requirements in complex scenes.

The P–R diagram of the detection algorithm employed in this study is shown in Figure 7, where hat indicates that the algorithm detects a target wearing a helmet, and person indicates that the algorithm detects a target that does not wear a helmet. Figure 7 shows that the detection accuracy reaches 0.995, which meets the needs of complex scene environments.

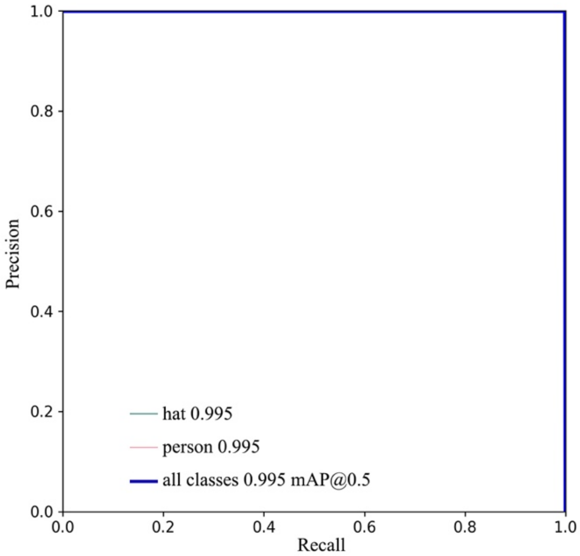

Fig. 7. P-R curve

In this study, the algorithm detection rate adopts the FPS parameter, which can measure the number of picture frames transmitted per second. The brain considers the image to be coherent when the frame rate of the image observed by the human eye is greater than 16 FPS, which is based on the special structure of the human senses. This phenomenon is called persistence of vision. The maximum number of detected picture frames per second can reach 30.5 FPS under this algorithm.

Fig. 8 shows the algorithm' s detection and labeling of personnel wearing helmets in complex scenes. The algorithm model can clearly label whether an individual wears a helmet. Thus, a reference for subsequent personnel safety status monitoring can be provided.

Fig. 8. Example for the labeling of helmet wearing

In summary, the algorithm model employed in this study can meet the real-time detection of helmet wearing in complex scene environments in terms of detection accuracy and detection rate.

5. Conclusions

This study started with the characteristics of a complex scene environment and applied the YOLOv5 algorithm to helmet-wearing detection to realize real-time detection and marking of helmet wearing in a complex scene environment. Then, the algorithm detection results from the algorithm model detection accuracy and detection speed were analyzed. The following conclusions could be drawn:

(1) The convergence function curve of algorithm loss value obtains a good fitting effect, proving that the algorithm has good performance.

(2) The algorithm has good performance in the evaluation indicators, such as the accuracy rate and R rate, obtained by detecting the wearing of helmets in a complex scene environment. Thus, the requirements of the complex scene environment are met.

(3) During the experiment, the detection speed can reach 30.5 FPS, which meets the real-time requirements of video images.

In this study, the YOLOv5 algorithm is applied to the detection of helmets in complex scene environments, which provides certain technical guidance for improving the safety of workers in complex scene environments. This application can be extended to other target detection in complex scene environments. The pictures employed in this experiment are all well lit in complex scenes, and the occluders do not block the target too much. In the actual application environment, too many occluders and insufficient light may exist. Thus, the algorithm needs to be further improved, and the follow-up work will do further research in related aspects.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License.

References

1. Girshick R., Donahue J., Darrell T., Malik J., "Rich feature hierarchies for accurate object detection and semantic segmentation". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*,2014, pp.580-587.
2. He K., Zhang X.Y., Ren S.Q., Sun J., "Spatial pyramid pooling in deep convolutional networks for visual recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 2015, pp.1904-1916.
3. Girshick R., "Fast R-CNN". *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp.1440-1448.
4. Ren S.Q., He K.M., Girshick R, Sun J., "Faster R-CNN: Towards real-time object detection with region proposal networks". *Advances in Neural Information Processing Systems 28*. 2015, pp.91-99.
5. He K.M., Gkioxari G., Dollar P,Girshick R., "Mask R-CNN". *Proceedings of the IEEE International Conference on Computer Vision*,2017, pp.2980-2988.
6. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.Y., Berg A.C., "SSD:single shot multibox detector". *European Conference on Computer Vision.Cham:Springer*, 2016, pp.21-37.
7. Redmon J., Divvala S., Girshick R., Farhadi A., "You only look once ：unified，real- time object detection". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp.779-788.
8. Redmon J., Farhadi A., "YOLO9000: better, faster, stronger". *IEEE Conference on Computer Vision and Pattern Recognition*,2017, pp.6517-6525.

9. Redmon J., Farhadi A., "YOLOv3: an incremental improvement". *IEEE Conference on Computer Vision and Pattern Recognition*,2018, pp.89-95.
10. Bochkovskiy A., Wang C.Y., Liao H.Y.M., " YOLOv4: Optimal speed and accuracy of object detection" . arXiv: 2004.10934,2020.
11. Jocher G., "Yolov5" [EB/OL] . Retrieved from https://github.com/ultralyc-s/yolov5.2020-08-09.
12. Xu K.,Deng C., "Safety helmet wearing recognition algorithm based on improved YOLOv3". *Laser & Optoelectronics Progress*, 58(6), 2021, 615002(in Chinese).
13. Cheng R., He X. W., Zheng Z. L.,Wang Z.T., "Multi-Scale Safety Helmet Detection Based on SAS-YOLOv3-Tiny". *Applied Sciences*, 11(8), 2021, pp.3652-3668.
14. Xu C. Y., Yuan H. Y., Li G., Zheng Y., Lliu H., "Study on safety helmet wearing detection method based on scene enhancement". *Computer Engineering and Application*, 5(07), 2021,pp.1-8(in Chinese).
15. Zhang J., Qu P. Q., Sun C., et al. "Safety helmet wearing detection method based on improved YOLOv5". *Computer Application*, (6) 2021,pp.1-11(in Chinese).
16. Kelm A., Laussat L., Meins-Becker A., Platz D., Khazaee M.J., Costin A.M., Helmus M., Teizer J., "Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites". *Automation in Construction*, 36(08), 2013, pp.38-52.
17. Girshick R., Donahue J., Darrell T., Malik J., "Rich feature hierarchies for accurate object detection and semantic segmentation". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp.580-587.
18. Xu X.F., Zhao W.F., Zou H.Q., Zhang L., Pan Z.Y., "Detection algorithm of safety helmet wear based on MobileNet-SSD". *Computer Engineering*, https://doi.org/10.19678/j.issn.1000-3428.0058733.
19. Lin J., Dang W.C., Pan L.H., Bai S.W., Zhang R., "Helmet detection method based on YOLO". *Application of Computing System*, 28(9), 2019,pp.174-179.
20. Wang B., Li W.J., Tang H., "Improved YOLOv3 algorithm and its application in helmet detection". *Computer Engineering and Applications*, 56(9), 2020, pp.33-40.
21. Shi H., Chen X.Q., Yang Y., "Safety helmet wearing detection method of improved YOLOv3". *Computer Engineering and Applications*,55(11), 2019, pp.213-220.
22. Arcos-Garcia A., Alvarez-Garcia J. A., Soriamorillo L., M. "Evaluation of deep neural networks for traffic sign detection systems". *Neurocomputing*, 316, 2018, pp.332-344.
23. Jiang R.X., Arifu Kurban, Geng L.T., "Hardhat Detection Algorithm for Lightweight Networks". *Computer Engineering and Applications*, 57(20),2021, pp.263-270.
24. Zhao R., Liu H., Liu P.L., Lei Y., Li D., "Hard hat detection algorithm based on improved YOLOv5s". *Journal of Beihang University*, https://doi.org/10.13700/j.bh.1001-5965.2021.0595.