



# An Improved Detection Method of Safety Helmet Wearing Based on CenterNet

Bo Wang<sup>1</sup>, Yong Zhang<sup>1</sup>, Qinjun Zhao<sup>1</sup>(✉), and Shengjun Shi<sup>2</sup>

<sup>1</sup> School of Electrical Engineering, University of Jinan, Jinan 250022, China  
cse.zhaoqj@ujn.edu.cn

<sup>2</sup> Harbin Engineering University Robot Group Co., Ltd., Harbin 150000, China

**Abstract.** In some factories or construction sites, accidents occur because workers do not wearing safety helmets correctly. In order to reduce the accident rate, an improved detection method of safety helmet wearing based on CenterNet algorithm is proposed. The original IOU method is optimized by combining with GIoU, and debug the training model Res/DLA framework in the training process. At the same time, various parameters are adjusted by experiments. In the safety helmet wearing test task, theoretical analysis and experimental results show that mAP (Mean Average Precision) is up to 42.6%, detection rate is increased to 30.3%. Compared with CenterNet, the detection accuracy and detection rate are slightly improved. The proposed algorithm not only meets the real-time performance of detection task in safety helmet wearing detection but also has higher detection accuracy.

**Keywords:** Improved detection · Safety helmet · CenterNet

## 1 Introduction

Helmets recognition systems play a crucial role in the safe production. Through such real-time supervision of the construction operation plant area, it sounds the safety alarm for the workers, and improves the safety awareness of the workers while reducing the occurrence of safety accidents. In the future trend, with the continuous development of the industry and continues to segment demand, helmet identification system will further optimize related functions, leading to a more convenient management for the enterprise.

Domestic and foreign scholars had some research: Girshick and Ren [2] proposed Fast R-CNN and Fast R-CNN respectively, which improved the accuracy and detection speed, and the frame rate reached 5 f/s. In 2015, Redmon J [3] proposed the YOLO detection algorithm, which can detect video at a speed of 45 f/s. Based on YOLO, Redmon also proposed YOLOv2 [4] and YOLOv3 [5]

Supported by the National Key R & D Program of China (2018AAA0101703) and the key research and development project of Shandong province (2019GNC106093).

detection algorithms. In April 2019, Zhou et al. proposed CenterNet: objects as points [6], and proposed a new anchor free algorithm based on key points.

Based on the CenterNet algorithm, this paper optimizes it when it is applied to the safety helmet wearing detection that combining with GIoU to optimize the original IoU method. By adding C detection frame (C detection frame is the smallest rectangular box including detection box and real box) to make up for the defect that IOU's loss is the same when the detection frame and real box do not overlap, we also debug the training model Res/DLA framework in the training process. At the same time, various parameters are adjusted by experiments. To some extent, it eliminates the occurrence of coincidence detection and false detection of non target objects. Experiments show that the improved algorithm based on CenterNet can guarantee the detection rate and improve the detection accuracy.

## 2 CenterNet Principle

CenterNet takes the target as the centre of BBox when building the model. In the detection, the center point is determined by key point estimation, and other attributes, such as size and position, are regressed.

### 2.1 Frame

As shown in the Fig. 1.

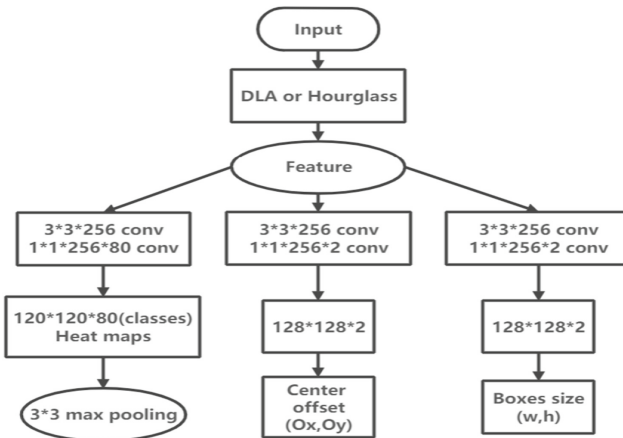


Fig. 1. Flow chart of framework.

### 2.2 Prediction of Key Point

Set  $A \subseteq B^{W*H*3}$  as input image, the goal is to generate a heat map for key point prediction  $E \subseteq [0, 1]^{\frac{W}{B}*\frac{H}{B}*C}$ , the R is the size scaling, and C is the number of feature maps. For the key point c of Ground Truth, its coordinate position is  $\theta \in B^2$ , after the network processing, the position on the feature map is

$$\tilde{\theta} = \left\lfloor \frac{\theta}{B} \right\rfloor \tag{1}$$

The paper passes the key points of Ground Truth through the Gaussian core

$$E_{d,e,c} = \exp\left(-\frac{(d - \tilde{\theta}_d)^2 + (e - \tilde{\theta}_e)^2}{2\sigma^2\theta}\right) \tag{2}$$

disperse to  $E \subseteq [0, 1]^{\frac{W}{B}*\frac{H}{B}*C}$ , the loss function is

$$L_k = \frac{1}{N} \sum_{dec} \begin{cases} (1 - \hat{E}_{dec})^\alpha \log(\hat{E}_{dec}) \text{ if } E_{dec} = 1 \\ (1 - E_{dec})^\beta (\hat{E}_{dec})^\alpha, \text{ otherwise} \\ \log(1 - \hat{E}_{dec}) \end{cases} \tag{3}$$

$\alpha$  and  $\beta$  are set to 2 and 4; N is the number of key points. Without considering  $(1 - E_{dec})^\beta$ , the new formula is

$$(1 - Z_t)^\alpha * \log Z_t \tag{4}$$

$$Z_t = \begin{cases} \hat{E}_{dec}, \text{ if } E_{dec} = 1 \\ 1 - \hat{E}_{dec}, \text{ otherwise} \end{cases} \tag{5}$$

$\log Z_t$  is the standard cross-entropy loss function.

And that is adding the prediction of the local offset to each key point  $\tilde{T} \subseteq B^{\frac{W}{B}*\frac{H}{B}*2}$ .

$$L_{off} = \frac{1}{N} \sum_{\theta} \left| \hat{T}_{\tilde{\theta}} - \left(\frac{\theta}{B} - \tilde{\theta}\right) \right| \tag{6}$$

### 2.3 IOU(Intersection over Union)

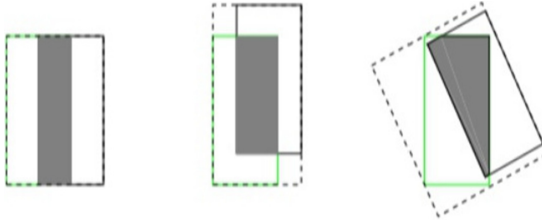
IoU is what we call cross ratio.

$$IOU = \frac{M \cap N}{M \cup N} \tag{7}$$

- (1) It can reflect the effect of the predicted detection frame and the real detection frame;
- (2) Scale invariance, that is scale invariant. In regression tasks, IoU is a direct indicator of distance between predict box and ground-truth (Satisfy non-negativity; Identity; Symmetry; Triangle inequality).

Problems (disadvantages) that may occur as a loss function:

- (1) If the prediction frame and the detection frame do not intersect,  $IOU = 0$ . Moreover, when  $loss = 0$ , the gradient can not be returned, resulting in the inability to train;
- (2) As shown in the Fig. 2, the IOU of the three cases are equal, but it can be seen that their coincidence degree is not the same. The IOU value can not reflect how the two boxes intersect. The effect is good to bad from left to right.



**Fig. 2.** Chart of comparison.

As shown in the figure above, three boxes with different relative positions have the same  $IoU = 0.33$ , but different  $GIoU = 0.33, 0.24, -0.1$ . When the alignment direction of the box is better, the value of  $GIoU$  is higher.

### 2.4 Improvement Direction

The regression loss (MSE loss, L1-smooth loss, et al.) optimization of BBox in detection task is not completely equivalent to IOU optimization, and Ln norm is also sensitive to object scale, IOU can not directly optimize the non overlapping part. Therefore, IOU can be directly set as the loss of regression. This method is called  $GIoU$  Loss(Generalized Intersection over Union) [7].

$$GIoU = IOU - \frac{P - (M \cup N)}{P} \tag{8}$$

The meaning of the above formula is: firstly, calculate the minimum closure area  $C$  of the two boxes (popular understanding: including the area of the smallest box of the prediction box and the real box), and then calculate the  $IoU$ , and then calculate the proportion of the closure area that does not belong to the two boxes in the closure area, and finally use  $IoU$  to subtract this proportion to get  $GIoU$ .

Characteristic:

- (1) GIoU as distance,  $L_{GIoU} = 1 - \text{GIoU}$ , which satisfy nonnegativity, identity, symmetry, trigonometric inequality;
- (2) Scale invariance;
- (3) When two frames overlap wirelessly,  $\text{IoU} = \text{GIoU}$ ;
- (4) When a and B coincide,  $\text{GIoU} = 1$ . When the intersection of a and B is very small or there is no intersection,  $\text{GIoU} = -1$ ;
- (5) GIoU takes into account non overlapping areas not considered by IoU.

GIoU is used to optimize the original IoU in CenterNet. Finally, the improved CenterNet algorithm is used to test the wearing of safety helmet.

### 3 Experimental Data Set Making

Data Collection: In this paper uses the same specification but different color helmet wearing pictures collected on the Internet to make the datasets.

Data Filtering: The background image without subject is deleted from the datasets, and the photo of human body wearing safety helmet is needed in the experiment.

Data Marking: In the annotation, we select the target in the image by using lableImg. The interface is shown in Fig. 3.

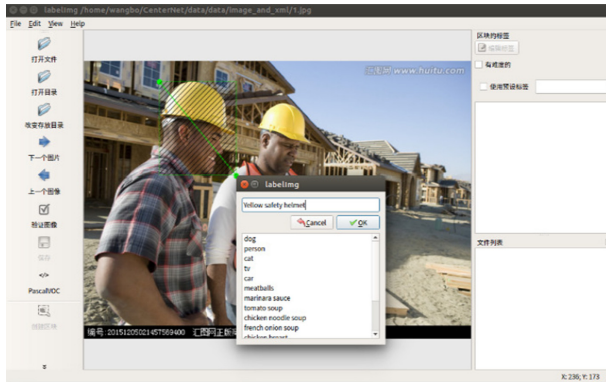


Fig. 3. Diagram of annotation.

### 4 Experimental Analysis

As shown in Fig. 4.

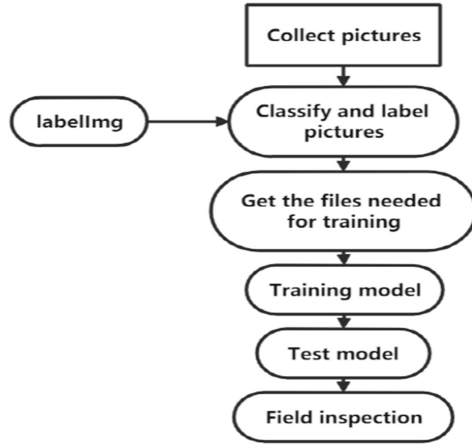


Fig. 4. Chart of experimental flow.

### 4.1 Experimental Scheme

In this paper, The datasets are divided into five categories: A to E. The classification is shown in Fig. 5.



Fig. 5. Diagram of sample.

### 4.2 Network Training

The parameters are adjusted according to many experiments, as shown in Table 1.

**Table 1.** Description of network parameters.

Parameter	Parameter value
Learning rate	1.25e-4
Epoch	140
Batch size	5
NMS	False
$lr\_step$	[90, 120]
Arch	$DLA_{34}$
$head\_conv$	256

### 4.3 Experimental Platform and Network Training

In this paper, the improved CenterNet algorithm is used to test. The improvement point is: the IOU in the original CenterNet algorithm is improved by using the GIoU which calculates the minimum closure region. At the same time, Faster R-CNN in reference [1], RetinaNet in [6], Yolo V3 in reference [5] and CenterNet in reference [7] are used for comparison, the results shown in Table 2.

**Table 2.** Comparison of experimental results.

Algorithm	mAP(%)	Frame rate recognition
<i>FasterR - CNN</i>	42.5	5.3
RetinaNet	30.7	5.3
YOLO v3	31.3	20.2
CenterNet	39.3	28.4
Improved CenterNet	42.6	30.3

### 4.4 Result Analysis

The experimental results show that although the accuracy rate of improved CenterNet is lower than that of Faster R-CNN, the detection rate ratio is several times of that of Faster R-CNN. Therefore, the improved CenterNet performs better. However, the other methods are not as good as the improved CenterNet algorithm in two aspects. The improved CenterNet algorithm takes into account the detection accuracy and detection rate at the same time, and can complete the safety helmet wearing detection task better.

In addition, in order to feel the detection differences more intuitively between different algorithms, this paper selects some detection images for comparative analysis, in which Fig. 6 is the detection effect of CenterNet algorithm, and Fig. 7 is the detection effect of improved CenterNet algorithm.



Fig. 6. The detection effect of CenterNet algorithm.



Fig. 7. The detection effect of improved CenterNet algorithm.

By observing Fig. 6 and Fig. 7, it can be found that the coincidence frame display and the false recognition of non target objects are solved to a certain extent. Therefore, in this paper that the improved CenterNet algorithm can maintain a high detection rate and meet the requirements of real-time detection for helmet wearing detection and classification tasks.

## 5 Conclusion

In this paper, an improved method of helmet wearing detection based on CenterNet algorithm is proposed. The safety helmet wearing detection experiment is carried out by using the video taken by mobile phone in the construction site as the data set, which can ensure the high detection accuracy and still have a fast detection speed, and meet the requirements of the accuracy and real-time of the safety helmet wearing detection in the work environment monitoring video basically.

## References

1. Tingting Z., Jianwu Z., et al.: A survey of image object detection algorithm based on deep learning. *Telecommun. Sci.* **36**(7), 92–106 (2020)
2. Girshick, R.: Fast R-CNN. In: *IEEE International Conference on Computer Vision*, pp. 1440–1448 (2015)



3. Redmon J, Divvala, S., Girshick, R., et al.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2015)
4. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 6517–6525 (2017)
5. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 89–95 (2018)
6. Zhou, X.Y, Wang, D., Philipp, K.: Objects as Points. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
7. Rezatofghi, H., Tsoi, N., Gwak, J.Y., et al.: Generalized intersection over union: a metric and a loss for bounding box regression. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2020)