

Research on Detection Algorithm of Safety helmet based on improved YOLOX

Weichen Liao

Southwest University for Nationalities, Chengdu 610225, Sichuan, China

Abstract: In the industrial production, construction and other fields, safety helmet as an important safety protective equipment, the wearing situation of workers personal safety and property safety is of great significance. Therefore, this paper proposes a safety helmet detection method based on the improved YOLOX algorithm. First, the 5000 pictures of safety helmet wearing at construction sites are labelled. The Squeeze and-Excitation module is introduced in the YOLOX network structure. The original Loss function is replaced with varifocal Loss. After experimental verification, compared with the original YOLOX target detection algorithm, our algorithm improves by 2.13 percentage points, enhances the model's focus on key areas and optimizes the model training effect, while the number of model parameters does not increase significantly. In conclusion, our algorithm has a wide range of application prospects and research value.

Keywords: YOLOX; Helmet detection; Building construction; Target detection.

1. Introduction

Head injuries are one of the most common types of injuries in industrial accidents, according to data provided by the Occupational Safety and Health Administration (OSHA). Head injuries account for more than 25 percent of the total number of injury and death incidents in all industrial accidents, and about 75 percent of head injury incidents occur without wearing a safety helmet. Therefore, accurately detecting whether a safety helmet is worn is a task with practical application value and prospect. In helmet detection, a target detection algorithm is used [1] [2] [3] [4].

In recent years, object detection technology has been widely concerned and discussed in the field of computer vision. Neural network models based on deep learning, especially object detection algorithms based on convolutional neural networks (CNN), have made remarkable progress. Classical algorithms such as YOLO [5], SSD [6], CornerNet [8], etc., have the advantage of high detection speed on the premise of ensuring accuracy, and can meet the real-time requirements of video surveillance detection. However, the existing target detection algorithms are still faced with a series of challenges, such as low detection accuracy for small target objects, insufficient detection ability in complex scenes and so on. Therefore, how to quickly and accurately detect whether a person is wearing a safety helmet has become one of the research hotspots in recent years.

To solve the above problems, this paper proposes a helmet detection algorithm based on improved YOLOX. Specifically, the attention mechanism Squeest-and-Excitation module is introduced in this paper. The original Loss function is optimized as varifocal Loss to improve detection accuracy and robustness. The research results of this paper have a positive reference significance for the improvement of hard hat detection algorithm in practical application scenarios.

2. Research Methods

2.1. YOLOX network architecture

YOLOX algorithm is a target detection algorithm based on YOLO series algorithms. Its network structure mainly

includes four parts: backbone, neck, head and post-processing. It is shown in Figure 1.

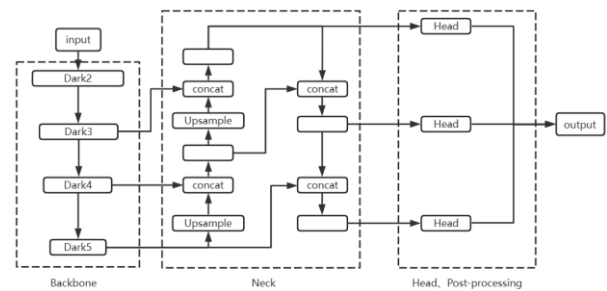


Figure 1. YOLOX network structure

The CSPDarknet53 Backbone network independently designed by YOLOX is used for backbone. This network not only guarantees high precision, but also has low computational complexity, which effectively improves the training and reasoning speed. The Neck part uses PAFPN [8] (Path Aggregation Feature Pyramid Network) feature pyramid network. Through multi-level feature fusion, target features at different scales and semantic levels can be captured better. The decoupling detection Head network independently designed by YOLOX is used in the head part, which introduces the idea of Anchor-free [9] target detection and adopts the point set prediction method for target detection, which avoids the uncertainty of Anchor selection in Anchor-based method and can locate small targets more accurately. Finally, the Post-processing module carries out post-processing on the output result of the model, including the steps of non-maximum suppression (NMS) and confidence screening, so as to filter the redundant information and get the final target detection result.

2.2. Hardhat detection algorithm based on improved YOLOX

For helmet wearing detection, the paper combines the existing YOLOX network model and adds the Squeeze and Excitation module. The original Loss function is optimized as varifocal Loss. The Anchor-free mechanism is adopted to improve the helmet wearing detection effect.

2.2.1. Introduction of the attention mechanism

Attention mechanism is a technique to improve the performance of deep neural networks and has been widely used in target detection tasks. Among them, the Squeeze-and-Excitation (SE) module is a lightweight, pluggable and efficient way to implement the attention mechanism. The SE module consists of two steps, as shown in Figure 2:

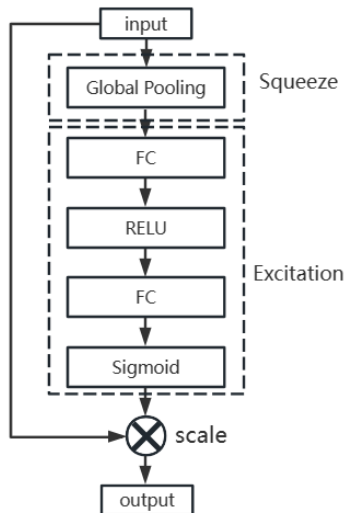


Figure 2. Attention mechanism structure

The first step is the squeeze operation, in which the input feature graph is globally average pooled in the channel dimension to obtain a vector with a small number of channels; The second step is the excitation operation. The vector is mapped to a vector with the same number of channels through two full connection layers. The sigmoid function is used to normalize it. The importance weights for each channel and the inputs are obtained.

There are two main ways to apply SE module in object detection tasks: one is to embed SE module into the target detection network as a part, the other is to apply SE module to the backbone network to improve the expression ability of features. The introduction of SE module can greatly improve the detection accuracy, while it has little influence on the computing cost of the model.

In this paper, SE module is inserted into neck network of detection network and connected with output of dark3, dark4 and dark5 in CSPDarknet53 feature extraction network. The up sampling and down sampling scaling operations of three feature graphs of different sizes are realized through FPN and PANet network structure. The feature maps of the same size are fused with each other, so as to enhance the feature representation ability of these layers. After fusion, the three-dimensional feature diagrams pass through the three YOLO convolution layers of large, medium and small, and enter the decoupling head of YOLOX for detection and recognition of helmet wearing. The network is shown in Figure 3.

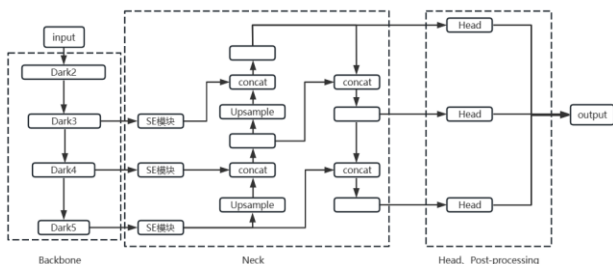


Figure 3. The structure of the improved YOLOx network

2.2.2. Varifocal Loss Optimization loss function

YOLOX improves the performance of target detection by introducing varifocal Loss optimization function. This Loss function is improved for the class imbalance in Focal Loss. By adjusting the weight coefficients of each positive and negative sample, the model pays more attention to the samples that are difficult to distinguish, so as to improve the detection accuracy.

Specifically, varifocal Loss replaces the square term index γ (2 by default) in Focal Loss with an adjustable parameter α , which is used to control the weight difference between positive and negative samples. At the same time, a new learnable parameter β is introduced to adjust the weight difference between the difficult and easy samples. That is, for the i -th sample,

$$L_i = -(1 - P_i)^\alpha (y_i \log(P_i) - \beta (y_i - P_i)^2 \log(1 - P_i)) - P_i^\alpha ((1 - y_i) \log(1 - P_i) - \beta (y_i - P_i)^2 \log(P_i)) \quad (1)$$

Where, P_i denotes the probability that the prediction is a positive sample, y_i denotes the true label, α and β are two tunable parameters.

In this paper, through experimental comparison, varifocal Loss showed better detection performance compared with traditional Focal Loss.

3. Experiment and results

3.1. Experimental data set

The experimental data set used in this paper is the construction scenes from various sites, including multiple scenes with different picture resolution, lighting conditions and personnel density. In the data set, there were situations where the scene was relatively complex, and there were also phenomena of wearing a hard hat but having a small target. The data set size was 5000 images. Of these, 2,800 images were used for the training model, 1,200 for the test set and 1,000 for the verification set.

3.2. Configuration of the experimental environment

The experimental environment of this paper is python3.7.11, the deep learning framework is PyTorch1.8.1, the training process is completed on NVIDIA GeForce RTX 2080 Ti, 11018iBx3 GPU, CUDA version is 11.4. During the training, the initial learning rate is set as 0.0001, and the learning rate is adjusted every 30 epochs. The adjustment strategy is that the learning rate varies according to the size of the training batch. When the weight file is used, yolox_tiny.

3.3. Model performance evaluation index

In order to objectively evaluate the performance of the algorithm in helmet wearing detection, mean Average Precision (mAP[10]) should be calculated. mAP index can objectively reflect the overall performance of the algorithm, and the calculation formula can be expressed as

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (2)$$

AP is the area value under the PR curve, which can be calculated using a definite integral

$$AP = \int_0^1 P(R) dR \quad (3)$$

3.4. Analysis of experimental results

Through the experiment, mAP data of different models were obtained, as shown in Table 1.

Table 1. Experimental results data

	mAP@.5	mAP@.5:.95	Parameters	GFLOPs
YOLOX-tiny	90.01	63.2	5.03M	6.44
+SE	90.30	64.51	5.03M	6.44
+varifocalLoss	89.19	64.67	5.03M	6.44
Our	89.53	65.33	5.03M	6.44

It can be seen from Table 1 that, on the basis of YOLOX-tiny, the mAP of experimental results (mAP@.5:.95) is more than 1% higher than the original algorithm no matter SE module is added or the optimization Loss function is changed into varifocal Loss. If SE module is added at the same time, When the Loss function is changed into varifocal Loss, mAP (mAP@.5:.95) increases by 2.13%, while the number of model parameters does not increase. It can be seen from the comparison experiment in Table 2 that by comparing SE module with CBAM module and Focal Loss with varifocal Loss, it is proved that SE module and varifocal Loss can significantly optimize the performance of the original YOLOX algorithm and improve the accuracy.

Table 2. Comparison experiment

	mAP@.5	mAP@.5:.95
YOLOX-tiny+CBAM	89.83	63.03
YOLOX-tiny+SE	90.30	64.51
YOLOX-tiny+focalLoss	89.19	63.70
YOLOX-tiny+varifocalLoss	89.89	64.67

4. Epilogue

Therefore, an improved algorithm based on YOLOX is proposed in this paper, which introduces the Squeeze and Excitation module and the Varifocal Loss function, and is experimentally verified on the hard hat detection mission. The experimental results show that the introduction of the Squeeze and Excitation module to focus on the helmet characteristics and optimize the Varifocal Loss function can significantly improve the performance of the helmet detection algorithm.

In the future, we will continue to further study the target detection technology and explore a more efficient real-time helmet detection algorithm.

References

- [1] Wang, Y., Zhang, D., Liu, S., Hou, Q., Zhang, C., Chen, B., ... & Huang, F. (2021). YOLOX: Exceeding YOLO Series in 2021. arXiv preprint arXiv:2107.08430
- [2] Jie H, Li S , Gang S , et al. Squeeze-and-Excitation Networks[C]// IEEE. IEEE, 2017.
- [3] Zhang H, Wang Y , Dayoub F , et al. VarifocalNet: An IoU-aware Dense Object Detector[J]. 2020.
- [4] irshick Ross,Donahue Jeff,Darrell Trevor,Malik Jitendra. RegionBased Convolutional Networks for Accurate Object Detection and Segmentation. [J]. IEEE transactions on pattern analysis and machine intelligence,2016,38(1).
- [5] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer.
- [7] Law H, Deng J . CornerNet: Detecting Objects as Paired Keypoints[J]. International Journal of Computer Vision, 2020, 128(3):642-656.
- [8] Liu S, Qi L , Qin H , et al. Path Aggregation Network for Instance Segmentation[C]// IEEE. IEEE, 2018.
- [9] Huang L, Yang Y , Deng Y , et al. DenseBox: Unifying Landmark Localization with End to End Object Detection[J]. Computer Science, 2015.
- [10] Everingham M, Winn J . The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Development Kit Contents. Springer-Verlag.