Volume 9 Issue 12, 2023
DOI: 10.6919/ICJE.202312_9(12).0014

Based on Improved Yolov8 Safety Helmet Detection

Fushun Shan, Jin Wang

School of Tianjin University of Technology and Education, Tianjin 300222, China

Abstract

In order to solve the problem that the original Yolov8 model often misses and mistake in detecting whether workers wear safety helmets, this paper proposes a safety helmet target detection algorithm based on improved Yolov8N. The attention modules of CBAM, SimAM and SA are added to the backbone of the original model to enhance the ability of the network to extract effective features of the target. Moreover, the weighted bidirectional feature pyramid BiFPN is introduced to enhance feature fusion and improve the accuracy of target detection. Experimental data show that using the improved model in the homemade dataset, the mAP of detecting safety helmets increased by 1.3% and the recall increased by 4%.

Keywords

Yolov8 Model; Safety Helmets; Attention Module; Feature Pyramid BiFPN.

1. Introduction

With the rapid development of artificial intelligence, Computer Vision(CV) and Deep Learning research is also becoming more and more mature. Many researchers have applied machine learning and CV to real-time monitoring and identification of whether workplace personnel are wearing safety helmets correctly, timely detection of improper wearing and corresponding measures, so as to reduce the occurrence of accidental injuries, all of which have achieved relatively good results. Liqiong Yang et al.[1] applied machine learning algorithms to helmet wearing detection, Zejia Han et al.[2] introduced ResNet50 network and deformable convolution based on the original SSD[3], and Ding Tian et al.[4] introduced ECA coordinate attention mechanism in the YOLOX model, and improved the loss function using CLOU, which improved the Mean Average Precision (mAP) of the improved YOLOX by 1.2%. However, the construction site scene is complex, the crowd is dense, the background of various operating materials and the types of safety helmets, there are many colors, and it is easy to be fooled by people wearing ordinary hats, and it is related to the safety of people's lives, and the lower the helmet detection tolerance rate, the better.

In view of the above problems, this paper proposes an improved detection algorithm based on the idea of the original YOLOv8 algorithm, which detects whether the staff wears safety helmets in complex backgrounds, and improves the detection mAP of targets, hoping to provide certain reference value for the subsequent application of safety helmet object detection in the field of deep learning.

2. Description of the Yolov8 Algorithm

2.1 Basic Network Structure

YOLOv8 is a new YOLO model structure upgraded by Ultralytics from YOLOv5, Yolov8 is much higher than YOLOv5 in terms of accuracy, but the speed is slightly reduced. In the Backbone part, the concept of CSP is still used, but the C3 module in YOLOv5 is replaced with the C2f module to achieve a further lightweight effect, while the SPPF module is still used in yolov8; In the feature fusion network, yolov8 still uses the PAN-FPN architecture, but yolov8 removes the convolutional structure in the PAN-FPN upsampling stage in YOLOv5, and also replaces the C3 module with the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202312_9(12).0014

C2f module. In the head part, a decoupled detection head similar to yolov6 and PP-YOLOE is used, and the output of cls and reg is output by the two heads, and the convergence speed and model performance can be better improved by using the decoupled head; In the loss function part, YOLOv8 uses VFL Loss as categorical loss and DFL Loss + CIOU Loss as regression loss. YOLOv8 also abandoned the previous Anchor-Base idea and used the Anchor-Free idea. This paper uses the yolov8n model, and its architecture is shown in Figure 1.

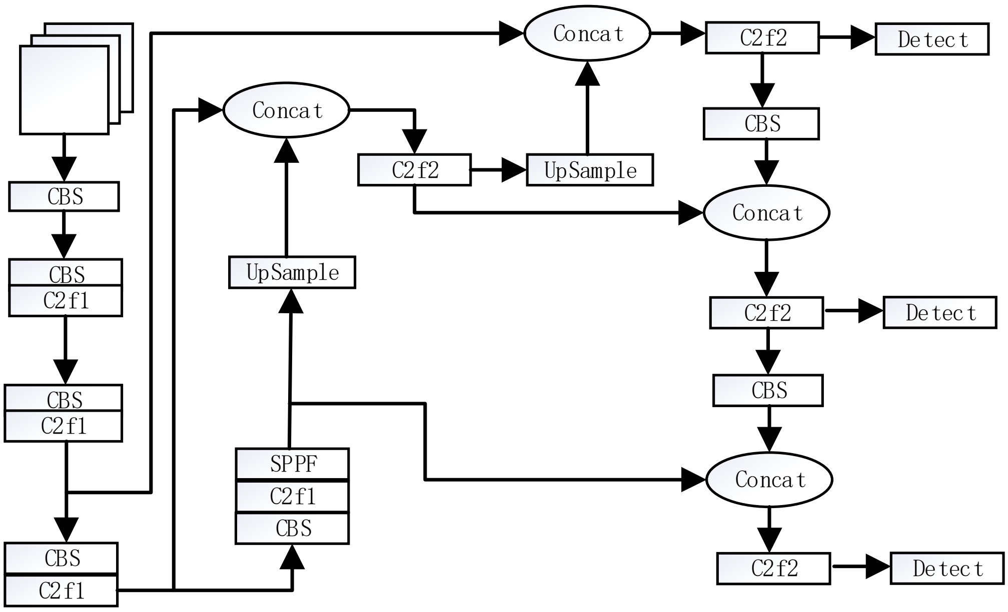

Figure 1. The original yolov8 model architecture

2.2 Module Parsing

CBS module is composed of three parts: two-dimensional convolution (Conv2d), batch normalization (BN layer) and SiLU activation function. C2f is similar to the CSP structure of yolov5, it has two types: C2f1 and C2f2. C2f1 Bottleneck has a residual architecture, but C2f2 does not. The SPPF module is a continuous three consecutive serial Maxpooling with residual structure, and the feature map without Maxpooling and the feature map that does Maxpooling every additional time are connected spliced to achieve feature fusion.

3. Improvements of the Yolov8 Algorithm

3.1 Backbone Network Structure Improvements

Adding an attention mechanism to the neural network can improve the utilization rate of the effective features of the target and enhance the ability of the network to extract the effective features of the target. In this experiment, the CBAM[5] attention module and SimAM[6] attention module were added behind the fourth layer of the original model, and the Shuffle Attention (SA)[7] attention module was added behind the ninth layer.

The CBAM and SA attention mechanisms can model the channel and spatial information of the image at the same time, which can better capture the important information in the image, thereby improving the performance and accuracy of the model. However, it makes the computational complexity of the model high and requires more computing resources. Therefore, the SimAM attention module is later introduced, SimAM is a similarity-based attention mechanism, which does not require additional parameters for learning, but determines the importance of each key vector by calculating the similarity score, thus realizing a lightweight and efficient attention mechanism.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202312_9(12).0014

3.2 Head Network Structure Improvements

In the Head module of the original model, the BiFPN-weighted bidirectional feature pyramid network and SA attention mechanism are introduced. The BiFPN structure is used multiple times and weighted feature fusion is performed. The SA module calculates the features of each subset and the features in other subsets, and then fuses the features to obtain the feature representation. The improved Yolov8 model architecture is shown in Figure 2.

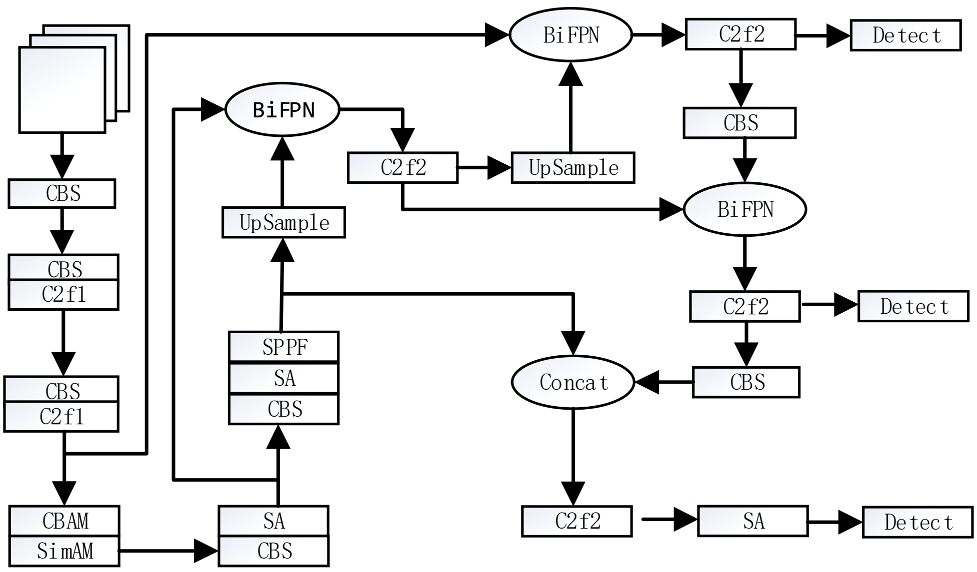

Figure 2. Improved Yolov8 model architecture

4. Experimental Results and Analysis

4.1 Datasets Authoring

The images of this experiment came from the open-source SHWD dataset, from which a total of 1,000 images with and without safety helmets were screened out with two categories of labels, and 810 images were divided into training sets, 90 images as validation sets, and 100 images as test sets. The image label labeling of SHWD dataset does not meet the type requirements of this experiment, so the labeling image labeling software is used to annotate and modify it by yourself. The two types of target components marked basically meet the requirements of this experiment.

4.2 Experimental Environment and Evaluation Indicators

This experiment is built using the Pytorch1.9.0 learning framework, the operating platform is ubuntu 18.04, using Nvidia GeForce RTX 3090 24GB, CUDA11.1 acceleration library and miniaconda3.

In order to objectively evaluate the performance of experimental improvement, this paper uses three indicators: accuracy, recall and average precision mAP to evaluate the model effect.

Recall represents the proportion of all that predicted positive correctly to all that are actually positive, also known as recall rate. The calculation formula is as follows.

$$R = \frac{TP}{TP + FN} \tag{1}$$

Accuracy indicates the proportion of correct predictions that are positive to all predictions that are positive, also known as precision. The calculation formula is as follows.

Volume 9 Issue 12, 2023
DOI: 10.6919/ICJE.202312_9(12).0014

$$P = \frac{TP}{TP+FP} \tag{2}$$

where TP represents the number of correctly divided places among all positive cases, that is, True Positives; FP represents the number of incorrectly classified cases among all positive examples, that is, False Positives; FN represents the number of misclassified examples among all negative examples, that is, False Negatives.

mAP is the mean of the average accuracy (AP) of each class. The calculation formula is as follows.

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{3}$$

where N is the total number of object classes detected, and APi represents the actual accuracy of the i-th category object.

4.3 Analysis of Experimental Results

In this experiment, the target detection model was trained a total of 150 times, and the comparison data between the original YOLOv8 model and the improved model are shown in Table 1. The train Precision-Recall Curve are shown in Figure 3 and Figure 4.

Table 1. Comparison of the data of the two models

Model Architecture	mAP@0.5/%	mAp@0.5:0.95/%	Recall/%	Precision/%
Original Yolov8	88.63%	55.75%	80.11%	90.61%
Improved Yolov8	89.86%	56.63%	84.12%	87.46%

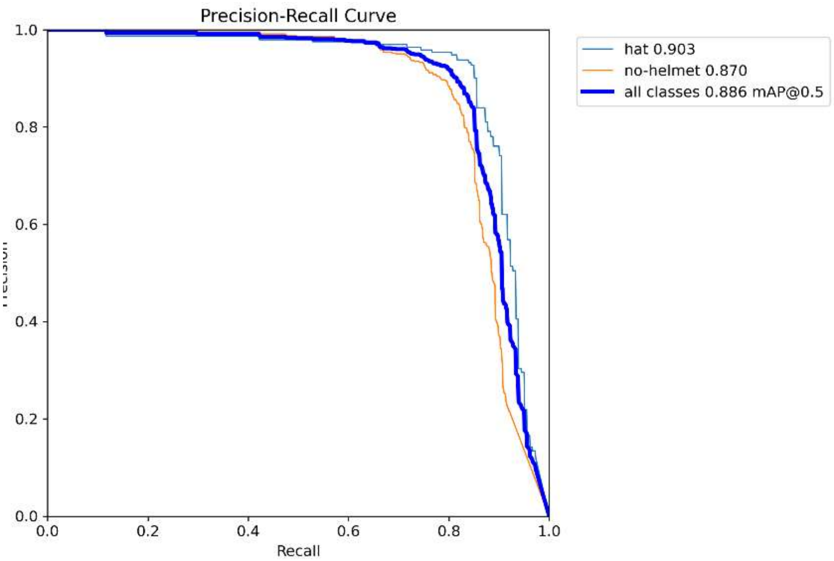

Figure 3. PR curve of the Original Yolov8 model

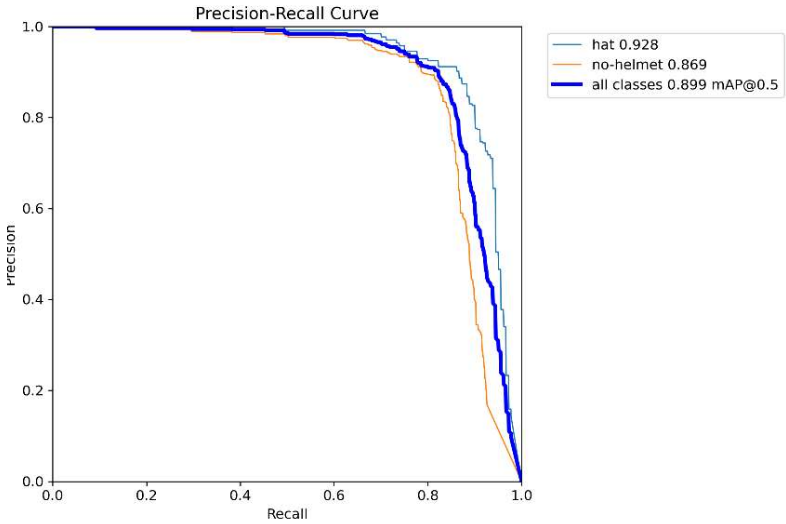

Figure 4. PR curve of the Improved Yolov8 model

It can be seen from Table 1 that Yolov8 with attention mechanism and feature fusion increased mAP by 1.2% when the IOU threshold was set to 0.5 compared with the original model. When the IOU threshold is 0.5~0.95 (from 0.5 to 0.95, step size 0.05), mAP increases by 0.9%; The recall rate improved by 4%. While Precise has declined, the model has indeed improved overall within control. As shown in Figures 5 and Figures 6, the improved YOLOv8 model not only improves the missed detection problem but also improves the accuracy.

Figure 5. Original model Detection Renderings

Figure 6. Improved model Detection Renderings

ISSN: 2414-1895 DOI: 10.6919/ICJE.202312_9(12).0014

5. Conclusion

In this paper, a method to add multiple attention mechanisms to the backbone network and detection head network part of the original Yolov8 is proposed to enhance the feature extraction ability of safety helmets. After the comparison of various data indicators after the experiment, the improved algorithm has indeed improved most of the indicators of target recognition of safety helmets, and the detection effect has indeed improved, which provides certain reference value for the research of safety helmet detection.

References

[1] Liqiong Yang, Liqiang Cai and Song GU. Safety helmet wearing behavior detection based on machine learning method[J]. China Safety Science and Technology,2019,15(10):152-157.

[2] Zejia HAN, Qingkun XIAO, Liqi Zhang. Reflective Coat Detection Algorithm Based on Improved SSD Safety Helmet [J]. Automation and Instrumentation,2021,36(09):63-68.

[3] Wei L , Dragomir A , Dumitru E , et al. SSD: Single Shot MultiBox Detector[J]. ECCV 2016: 14th European Conference on Computer Vision, 2016: 21-37.

[4] Tian Ding, Xiangyang Chen, Qiang Zhou, et al. Real-time detection of helmet wearing based on improved YOLOX[J]. Electronic Measurement Technology,2022,45(17):72-78.).

[5] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV), 2018:3-19.

[6] Yang L, Zhang R Y, Li L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks[C]//International conference on machine learning. PMLR, 2021: 11863-11874.

[7] Zhang Q L, Yang Y B. Sa-net: Shuffle attention for deep convolutional neural networks[C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021: 2235-2239.