

Research on Sleep Detection Algorithm during Working Hours Based on Improved YOLOv8s Model

Wanpeng Qi

Tianjin University of Technology and Education, Tianjin, China

ABSTRACT

Sleeping during working hours is not only an uncivilized behavior, but also a serious dereliction of duty and irresponsibility towards the production position. In the production process, if workers sleep during work, it is easy to make mistakes, fail to adjust the working conditions of the machine in a timely manner, and fail to detect problems in the first time, which brings huge safety hazards to the entire production operation and affects production progress. It is indeed a practical problem that should not be ignored. This article proposes a work sleep detection algorithm based on the improved YOLOv8s model to improve detection accuracy and efficiency. After experiments, YOLOv8s was selected as the basis for the algorithm model in this paper. GAM attention mechanism was added to the original YOLOv8s algorithm model, and the loss function WIoUv1 was added. The mAP was increased from the initial 94% to the improved 95%, which provides certain reference value for future algorithm models for detecting sleep during work.

KEYWORDS

YOLOv8; GAM attention mechanism; WIoU loss function; mAP

1. INTRODUCTION

The phenomenon of sleeping at work refers to the occurrence of a state of sleep at the workplace. This phenomenon usually occurs in job positions that require continuous duty, as prolonged work can lead to employee fatigue and idleness, resulting in sleep problems that not only manifest as obvious sleep, but may also manifest as drowsiness. This phenomenon not only affects the quality and efficiency of individuals' work, but may also pose serious safety hazards. With the development of deep learning, object detection algorithms can be applied to detect in real-time whether workers are sleeping while working.

In recent years, deep learning is gradually applied to image detection and recognition, which greatly promotes the research in this field [1]. Object detection algorithms are an important component of computer vision, whose main task is to identify and locate objects of interest from images. With the development of deep learning technology, object detection algorithms have gradually become mainstream. These algorithms can automatically learn and extract image features, greatly improving the accuracy and efficiency of detection. At present, object detection algorithms are mainly divided into two types. One is the two-stage algorithm represented by R-CNN series algorithms. The other one is the one-stage algorithm represented by you only look once (YOLO) series algorithms [2]. The single-stage object detection algorithm transforms object detection into a classification problem, with the advantage of being fast and completing detection in just one step. However, due to excessive simplification, the accuracy is usually not as good as two-stage object detection algorithms. The two-stage object detection algorithm consists of two steps: first, generating candidate regions, and then

running classifiers on these regions. This method is more accurate than single-stage, but slower in speed [3].

Due to the significant impact of sleeping at work, it is not conducive to safety production in the long run. Therefore, we need a single object detection algorithm with faster detection speed to undertake the detection task. This article will make certain improvements to the YOLOv8 algorithm model to make the improved model more efficient in detection.

2. YOLOV8 ALGORITHM

2.1. YOLOv8 Algorithm Structure

YOLOv8 is released by Ultralytics, the company that developed YOLOv5. YOLOv8 is a new version of the YOLO object detection and image segmentation model, and its algorithm network structure mainly includes Backbone, Neck, and Head [4].

(1) Backbone: The backbone network is the foundation of the algorithm model, and the Backbone part is responsible for feature extraction. At the same time, residual connections and bottleneck structures are also used in the backbone network to reduce network size and improve performance. This section uses the C2f module as the basic building block, which has fewer parameters and better feature extraction capabilities compared to YOLOv5's C3 module.

(2) Neck: The Neck section is responsible for multi-scale feature fusion, enhancing feature representation ability by fusing feature maps from different stages of Backbone. It includes an SPPF module for pooling operations at different scales, concatenating feature maps of different scales together to improve detection ability for targets of different sizes. PAN module: used for path aggregation of features at different levels, enhancing the expressive power of feature maps through bottom-up and top-down paths.

(3) Head: It is responsible for processing the feature maps obtained from the feature extraction network and generating the final prediction results.

2.2. Different Models of YOLOv8 Algorithm

In the YOLOv8 algorithm, as the depth and width continue to increase, there are five different models: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. As the depth and width of the network continue to increase, the detection accuracy will continue to improve, but the detection speed will decrease. Therefore, using the YOLOv8 algorithm model as the basic network for identifying the phenomenon of sleeping at work requires multiple experiments and comparing experimental data to make a decision.

3. NETWORK IMPROVEMENT OF YOLOV8

3.1. GAM Attention Mechanism

The GAM attention module is a global attention mechanism that improves the performance of neural networks by reducing information diffusion and amplifying global interactive representations. The attention mechanisms of GAM and CBAM are quite similar, both using channel attention mechanism and spatial attention mechanism. But the difference lies in the processing of channel attention and spatial attention. The structure of GAM attention mechanism is shown in Figure 1:

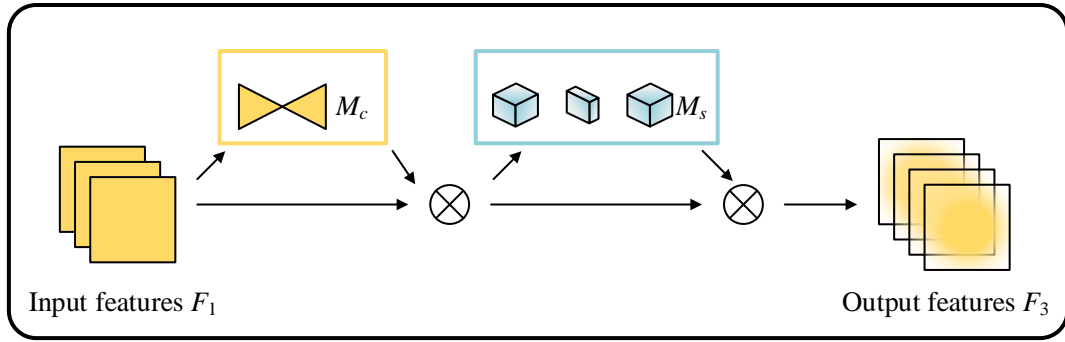


Figure 1. GAM attention mechanism module

The relationship between input feature F_1 , intermediate state F_2 , and output feature F_3 is:

$$\begin{cases} F_2 = M_c(F_1) \otimes F_1 \\ F_3 = M_s(F_2) \otimes F_2 \end{cases}$$

In the formula: M_c is the channel attention module; M_s is a spatial attention module

The channel attention submodule uses 3D permutation to retain information across three dimensions. It then magnifies cross-dimension channel-spatial dependencies with a two-layer MLP [5].

The spatial attention mechanism assumes that the size of the input feature map F_2 is $C \times H \times W$. Firstly, the input features are spatially fused through two 7×7 convolutional layers, and then the output is activated by the sigmoid function. Finally, a new feature map $M_s(F_2)$ is obtained.

The structural diagrams of the channel attention submodule and spatial attention submodule are shown in Figure 2.

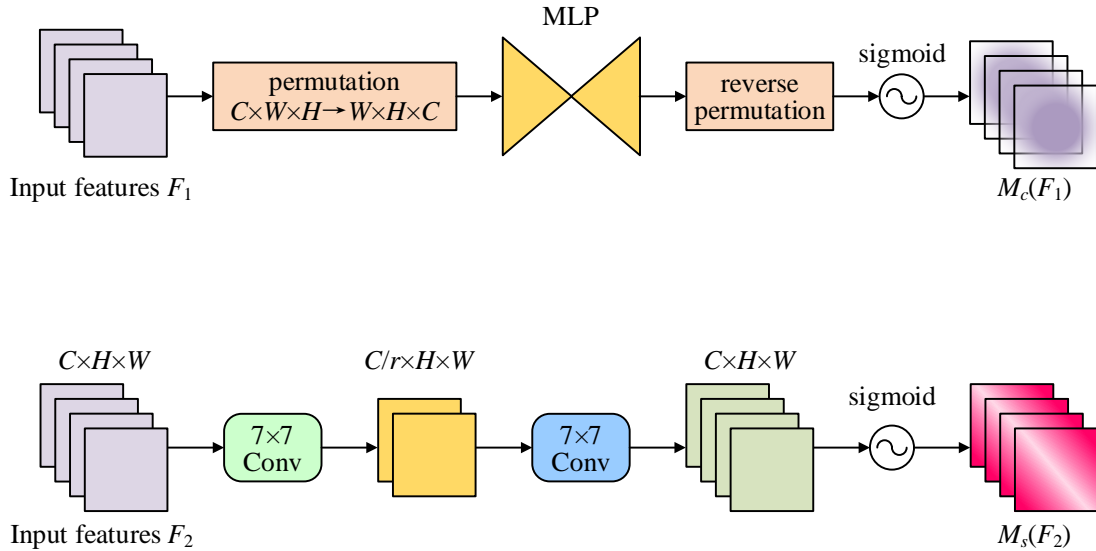


Figure 2. Channel attention submodule and spatial attention submodule

3.2. WIOU Loss Function

In object detection tasks, the similarity between predicted boxes and real boxes is an important evaluation metric. Although the traditional IoU loss function can intuitively reflect the degree of matching between the target detection results and the real situation, it has limitations in some cases, such as when the predicted box and the real box do not intersect, the IoU loss function cannot be optimized. Therefore, we introduce the Wise IoU loss function to solve the gradient vanishing problem in bounding box regression. By using a dynamic non monotonic focusing mechanism, it is

possible to better focus on anchor boxes of ordinary quality, thereby improving the performance of the object detection model.

We constructed distance attention based on distance metrics and obtained WIoU v1 with a two-layer attention mechanism. The calculation formula is as follows:

$$L_{WIoUv1} = R_{WIoU} L_{IoU}$$

$$R_{WIoU} = \exp \left(\frac{(x-x_{gt})^2 + (y-y_{gt})^2}{(W_g^2 + H_g^2)^*} \right)$$

4. ANALYSIS OF EXPERIMENTAL RESULTS

4.1. Dataset Processing

The dataset used in this experiment is partly derived from screenshots of videos of sleeping at work, and partly from searching for images of sleeping at work on Baidu, and selecting clear or scene diverse images from them. Expand the dataset to 800 images through rotation, increasing brightness, and other methods. Among them, there are 560 training set images, 160 validation set images, and 80 test set images, and the Labeling annotation tool is used to carefully label the parts of the images that appear to be sleeping at work, with the keyword "Sleeping Hill" labeled.

4.2. Experimental Environment

After preparing the dataset, choose to create an RTX 4090 (24GB) server in Autodl, select the most widely used YOLOv8 community image, and run PyTorch version 2.0 and CUDA version 11.8.

4.3. Evaluation Criteria

In the YOLO series of algorithms, mAP can be used to evaluate the performance of models in different categories and to compare the performance between different models. In object detection tasks, a high mAP usually indicates that the model has good performance.

The metric used in this article to evaluate the effectiveness of algorithm improvement is the average accuracy mAP, as shown below:

$$mAP = \frac{\sum AP}{N}$$

mAP stands for average accuracy, usually referring to the average AP of all categories in all images.

4.4. Comparison and Analysis of YOLOv8 Network Model Experiments

After 200 rounds of training, the running time and mAP of each YOLOv8 network model are shown below:

Table 1. The running time and mAP of each YOLOv8 network model

model data \ Network model	YOLOv8n	YOLOv8s	YOLOv8m	YOLOv8l	YOLOv8x
time	0.12h	0.154h	0.258h	0.377h	0.557h
mAP	0.924	0.94	0.917	0.923	0.91

The running time of YOLOv8s is only 0.034h longer than YOLOv8n, but the mAP is 0.016 higher. Moreover, YOLOv8m, YOLOv8l, and YOLOv8x have longer running times, and their mAP is not as high as YOLOv8s. Therefore, we use YOLOv8s as the basic network for identifying sleep during work.

4.5. Result Analysis

YOLOv8s is the basic algorithm framework for identifying sleeping at work in this article. Based on it, the GAM attention module is inserted into the Backbone, and the WIoUv1 loss function is introduced. After 200 epochs of training, the improved mAP is significantly improved, and the specific data improvement is shown below:

Table 2. The improved mAP is significantly improved and the specific data improvement

Algorithm model	The original YOLOv8s model	Improved YOLOv8s model
mAP	0.94	0.95

After adding the GAM attention module and introducing the WIoUv1 loss function, the mAP increased from 94% to 95%, and the algorithm model for detecting sleep during work showed significant improvement. Figure 3 shows the insertion position of the GAM attention module into the Backbone.

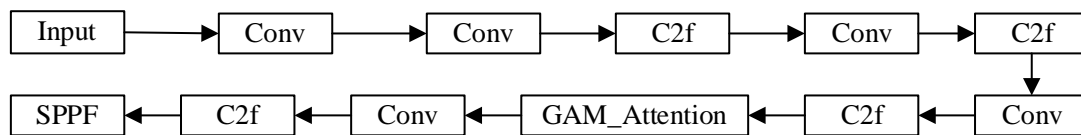


Figure 3. The position of GAM attention module in the backbone network

5. SUMMARY

This article proposes a work sleep detection algorithm based on an improved YOLOv8s model. After multiple experiments, GAM attention mechanism was added to the original YOLOv8s algorithm model, and WIoUv1 loss function was introduced. The improved mAP increased from the initial 94% to the improved 95%, with a 1% increase in mAP. This provides certain reference value for future algorithm models for detecting sleep during work.

REFERENCES

- [1] Li M, Zhu H, Chen H, et al. Research on object detection algorithm based on deep learning[C]//Journal of Physics: Conference Series. IOP Publishing, 2021, 1995(1): 012046.
- [2] Liu Y, Chen H. A Lightweight Container Detection Network Model Based on YOLOv5s[C]//Technical Committee on Control Theory, Chinese Association of Automation), (Chinese Association of Automation), (Systems Engineering Society of China), (Proceedings of the 42nd Chinese Control Conference). (8). School of Artificial Intelligence, Hebei University of Technology; Control Engineering Technology Innovation Center of Hebei Province, Hebei University of Technology; 2023:6. DOI:10.26914/c.cnkihy.2023.036061.
- [3] Du L, Zhang R, Wang X. Overview of two-stage object detection algorithms[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1544(1): 012033.
- [4] Wang F, Wang H, Qin Z, et al. UAV target detection algorithm based on improved YOLOv8 [J]. IEEE Access, 2023.
- [5] Liu Y, Shao Z, Hoffmann N. Global attention mechanism: Retain information to enhance channel-spatial interactions [J]. arxiv preprint arxiv:2112.05561, 2021.