Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD27793

A Pilot Fatigue Detection Method Based on the YOLOv8 Model

Wenyue Zhang

Department of Humanities and Social Sciences, Harbin Institute of Technology, Harbin, China
zhh_0123@126.com

Abstract. Pilot action and behavior detection are one of the important means to evaluate pilots' workload. A detection method with high accuracy and low latency is a key challenge in current pilot fatigue driving detection. This paper proposes an improved YOLOv8 model-based method for detecting typical pilot actions and behaviors, aiming to achieve efficient and accurate detection of pilots' fatigue driving states. First, this study introduces the Coordinate Attention (CA) mechanism to enhance the model, establishing an improved YOLOv8 detection model that realizes the extraction of key features of typical actions and behaviors. Second, a lightweight detection head optimization strategy based on Partial Convolution (PConv) is adopted, addressing the issues of lightweight design and generalization ability of the improved model. Finally, this study constructs a dataset of typical actions and behaviors during the flight climb phase and conducts simulation experiments for training and testing. The experimental results show that the improved model achieves an accuracy of 97.7% in pilot fatigue feature detection, with the detection latency of 1.3ms, meeting the requirements of accuracy and real-time performance. This provides reliable technical support for pilot fatigue driving detection.

Keywords: YOLOv8, Pilot Fatigue Detection, Coordinate Attention Mechanism, Lightweight Detection Head

1. Introduction

In the field of fatigue detection, numerous studies have been conducted by previous researchers. Early studies mostly relied on physiological signals (e.g., electroencephalogram, electrocardiogram) for fatigue judgment. Although these methods demonstrate considerable accuracy, they require contact-based signal collection, potentially interfering with pilots' operations [1]. With the development of computer vision technology, non-contact detection methods based on facial features have become a research focus—for example, judging fatigue states by detecting features such as blink frequency and head posture. However, traditional machine learning algorithms (e.g., Support Vector Machine, AdaBoost) have poor generalization ability under complex lighting conditions and posture changes [2].

The existing methods exhibit two primary limitations: first, their feature extraction capability is limited, hindering accurate capture of subtle fatigue features; second, the model's substantial computational load and inadequate real-time performance fail to meet the stringent requirements of

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD27793

aviation scenarios. As a representative of single-stage object detection algorithms, the YOLOv8 model has the characteristics of fast detection speed and high accuracy, and has shown good adaptability in driver state detection [3]. However, when directly applied to pilot fatigue detection, it still has problems such as insufficient attention to key features and high model latency.

This study introduces the Coordinate Attention (CA) mechanism to enhance the model's capability for extracting key fatigue features (e.g., eyes, mouth) and employs a lightweight detection head optimization strategy based on PConv to reduce model parameters and computational load, thereby improving detection accuracy and real-time performance. The significance of this study lies in: providing an efficient and accurate technical solution for pilot fatigue detection, which helps reduce the risk of flight accidents and ensure aviation safety; meanwhile, the improved model architecture can provide a reference for other high-precision real-time detection scenarios.

2. Research methods

2.1. Analysis of the YOLOv8 model architecture

The YOLOv8 model adopts a three-stage architecture consisting of "backbone network - feature fusion network - detection head" [4]. The backbone network utilizes the CSPDarknet structure, transforming input images (640×640) into 16×16 deep feature maps via 5 downsampling operations. Each layer includes a Cross-Stage Partial (CSP) block and a convolution layer to realize the initial extraction of multi-scale features. The feature fusion network integrates the Feature Pyramid Network (FPN) with the Path Aggregation Network (PAN). The FPN fuses high-level features with middle-level features through upsampling, while the PAN fuses middle-level features with low-level features through downsampling. Finally, three scales of feature maps are output (P3: 80×80, P4: 40×40, P5: 20×20), corresponding to the detection of small-, medium-, and large-sized targets respectively. The detection head employs a coupled head structure, implementing bounding box regression and category prediction via 3×3 and 1×1 convolutions respectively. The output format is (x, y, w, h, conf, cls), where (x, y) are the coordinates of the target center, (w, h) are the width and height of the target, conf is the confidence, and cls is the category probability [5].

In the context of pilot fatigue detection, YOLOv8 demonstrates adaptability in two key aspects: first, its single-stage detection architecture circumvents the candidate region generation process inherent in two-stage algorithms, ensuring detection speeds that meet real-time requirements; second, multi-scale feature fusion effectively captures both eye features (small targets) and facial contours (large targets). However, its limitations are as follows: insufficient attention to key regions (e.g., eyelids, mouth corners) during the feature fusion process, resulting in poor robustness of fatigue feature extraction; redundant parameters in the 3×3 convolution layer of the detection head (the total number of parameters in each layer is Cin×Cout×3×3, where Cin and Cout are the number of input and output channels respectively), leading to high inference latency.

2.2. Embedding design of the Coordinate Attention (CA) mechanism

The Coordinate Attention (CA) mechanism enhances the positional sensitivity of features by explicitly modeling spatial coordinate information, and its core lies in decomposing channel attention into two independent directional attention maps (horizontal and vertical) [6]. Within the YOLOv8 model, the CA mechanism is integrated into the three feature fusion nodes (P3, P4, P5) of the FPN. The specific implementation process is as follows:

DOI: 10.54254/2755-2721/2025.LD27793

First, feature compression is performed. Global average pooling is applied to the input feature map F∈RC×H×W in the horizontal and vertical directions respectively, obtaining a horizontal feature vector F_h in C×H×1 and a vertical feature vector F_w in C×1×W. The calculation formulas are:

$$Fh(c,i,1) = W1\sum j = 1WF(c,i,j) \tag{1}$$

$$Fw(c,1,j) = H1\sum i = 1HF(c,i,j) \tag{2}$$

where c is the channel index, and i, j are spatial coordinates.

Second, feature fusion and encoding are conducted. The F_h and F_w are concatenated along the spatial dimension to form

$$Fcat \in RC \times (H+W) \times 1 \tag{3}$$

The number of channels is compressed to C/r (where r is the reduction ratio, set to 32 in the experiment) through 1×1 convolution, and the ReLU activation function is applied to obtain:

$$Fenc = ReLU(Conv1 \times 1(Fcat)) \tag{4}$$

Third, attention generation is implemented. The encoded features are split into F_h' and F_w' along the spatial dimension. The number of channels is restored to the original number through 1×1 convolution respectively, and then the Sigmoid function is used to generate horizontal attention weights and vertical attention weights

$$Ah \in RC \times H \times 1 \tag{5}$$

$$Ah = Sigmoid(Conv1 \times 1(Fh')) \tag{6}$$

$$Aw \in RC \times 1 \times W \tag{7}$$

$$Aw = Sigmoid(Conv1 \times 1(Fw')) \tag{8}$$

Finally, feature enhancement is carried out. The original feature map is multiplied with the attention weights in an element-wise manner to obtain the enhanced feature map:

$$Fout(c,i,j) = F(c,i,j) \times Ah(c,i,1) \times Aw(c,1,j) \tag{9}$$

2.3. Lightweight optimization of the detection head based on PConv

Partial Convolution (PConv) decomposes the standard 3×3 convolution into 1×3 and 3×1 separable convolutions, which reduces the total number of parameters while maintaining the receptive field [7]. The optimization process of the YOLOv8 detection head in the present study is as follows:

First, convolution decomposition is performed. Each 3×3 convolution layer in the detection head is replaced with a cascaded structure of "1×3 convolution + 3×1 convolution". Let Cin be the number of input channels and Cout be the number of output channels. The total number of parameters of the original 3×3 convolution is:

$$Pstd = Cin \times Cout \times 3 \times 3 \tag{10}$$

DOI: 10.54254/2755-2721/2025.LD27793

The total number of parameters of the decomposed PConv is:

$$Ppconv = Cin \times Cout \times (1 \times 3 + 3 \times 1) = 6 \times Cin \times Cout \tag{11}$$

The parameter compression ratio is:

$$\frac{Ppconv}{Pstd} = \frac{6}{9} = 2/3 \tag{12}$$

Second, to further reduce the computational load, a depthwise separable convolution layer is added after PConv. This operation decomposes the standard convolution into depthwise convolution (channel-wise convolution) and pointwise convolution (1×1 convolution), reducing the computational load from $Cin \times Cout \times H \times W \times k \times K$ (where k is the convolution kernel size) to $Cin \times H \times W \times k \times k + Cin \times Cout \times H \times W$ .Finally, the reconstructed detection head structure is "PConv layer + depthwise separable convolution layer + BN layer + SiLU activation function + 1×1 prediction layer", where the 1×1 convolution is used to output the bounding box parameters and category probabilities required for object detection.

2.4. Experimental design

This study employs a two-stage verification strategy: first, evaluating the improved model's feature extraction capabilities on a general dataset; second, assessing its performance in fatigue detection using a specialized driver dataset; and finally, utilizing transfer learning to determine its applicability to pilot fatigue detection.

The general dataset used is VOC2007. For the driver fatigue detection task, the publicly available YawnDD dataset (foreign) is adopted. This dataset comprises facial video frames captured during driving scenarios, annotated into two categories: "Fatigued" (characterized by typical fatigue indicators including continuous eye closure and frequent yawning) and "Non-fatigued" (defined as a stable facial state without evident fatigue features).

The experimental environment is configured as follows: 13th Gen Intel(R) Core(TM) i7-13650HX 2.60 GHz CPU, NVIDIA RTX 4060 (notebook) GPU, and the deep learning framework is PyTorch 1.10.1. The input image size is uniformly adjusted to 512×512, the batch size is set to 8, and the number of iterations is 100.

3. Research results and analysis

3.1. Experimental data results and analysis

Table 1 presents a comparative evaluation of the original YOLO model and the improved YOLO model across four key metrics: recall, precision, facial landmark detection mean Average Precision (mAP), and detection latency, highlighting the enhancements in accuracy and real-time performance achieved by the optimized model. The training results show that the improved model achieves a mean Average Precision (mAP) of 97.9% for facial landmark detection, an accuracy of 97.7%, a recall of 96.1%, and a detection latency controlled at 1.3ms, which meets the real-time requirement of ≤30ms for the aviation scenario [8]. This indicates that the CA mechanism enhances the ability to capture local features, and the PConv optimization effectively improves the computational efficiency.

DOI: 10.54254/2755-2721/2025.LD27793

Table 1. Performance comparison between original and improved YOLO models

Metric	Original YOLO Model	Improved YOLO Model
Recall	95.4%	96.1%
Precision	97.4%	97.7%
Facial Landmark Detection (mAP)	96.6%	97.9%
Detection Latency	3.1ms	1.3ms

Specifically, the improved model shows significant improvements in recall, precision, mAP of facial landmark detection, and detection latency. The recall increases from 95.4% to 96.1%, indicating that the model has a stronger ability to identify fatigue states and reduces missed detections. The precision rises from 97.4% to 97.7%, which means false detections are further suppressed. The mAP of facial landmark detection reaches 97.9%, an increase of 1.3 percentage points compared with the original model, reflecting that the model's ability to extract features from key regions (e.g., eyelids, mouth corners) is significantly enhanced. Notably, the detection latency is greatly reduced from 3.1ms to 1.3ms, with a reduction rate of 58%, which fully verifies the effectiveness of the lightweight design.

These performance improvements are mainly attributed to two improvements: first, the introduction of the Coordinate Attention (CA) mechanism enhances the model's ability to perceive spatial position information, making it more focused on fatigue-related regions (e.g., eyes, mouth), thereby improving the feature recognition ability; second, the reconstruction of the detection head based on PConv, through the combined use of convolution decomposition and depthwise separable convolution, significantly reduces the number of parameters and computational complexity while maintaining the receptive field, achieving a balance between speed and accuracy.

Further analysis shows that the improved model not only meets the high real-time requirement (latency $\leq$30ms) of the aviation scenario but also maintains excellent detection accuracy, demonstrating potential for deployment in actual flight environments. Its low-latency feature is particularly suitable for cockpit monitoring systems requiring high-frequency and low-latency responses, providing reliable technical support for the real-time evaluation of pilots' states.

4. Conclusions

The pilot fatigue detection method based on the improved YOLOv8 proposed in this paper effectively enhances detection accuracy and real-time performance by incorporating the Coordinate Attention mechanism and optimizing the lightweight detection head. The experimental results demonstrate that the improved model achieves an accuracy of 97.7% and a latency of 1.3ms in pilot fatigue feature detection, sufficiently meeting the stringent requirements of the aviation scenario.

This study has three limitations: first, the dataset does not encompass real flight environments characterized by high-altitude strong radiation and extreme temperature variations; second, the model exhibits insufficient robustness against occlusions such as pilots wearing oxygen masks or sunglasses; third, the impact of individual differences among pilots (e.g., baseline blink rate) on detection outcomes is not considered.

Future research can be carried out in the following aspects: constructing a multi-modal dataset under real flight scenarios (integrating infrared images and physiological signals); introducing an attention mask mechanism to address occlusion issues and combining Generative Adversarial

DOI: 10.54254/2755-2721/2025.LD27793

Networks (GANs) to enhance data diversity; designing an adaptive threshold adjustment algorithm to realize personalized fatigue judgment.

References

[1] Wang, L., & Zhang, J. (2020). A review of driver fatigue detection based on physiological signals. Biomedical Signal Processing and Control, 58, 101815.
[2] Zhao, Y., et al. (2019). Driver fatigue detection using facial landmarks and support vector machine. IEEE Access, 7, 173264-173273.
[3] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv: 1804.02767.
[4] Ultralytics. (2023). YOLOv8: Real-Time Object Detection. https: //github.com/ultralytics/ultralytics
[5] Li, S., et al. (2022). Application of YOLOv5 in driver distraction detection. Journal of Intelligent & Fuzzy Systems, 42(3), 2345-2356.
[6] Hou, Q., et al. (2021). Coordinate attention for efficient mobile network design. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 13713-13722.
[7] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1251-1258.
[8] Zhu, X., et al. (2020). PERCLOS-based driver fatigue detection using improved YOLOv4. IEEE Transactions on Intelligent Transportation Systems, 21(12), 5181-5191.