



Multi-Feature fatigue detection method for bus drivers based on YOLOv8n-SimAM with integrated temporal information

Keqi Wu^{1*}, Shuo Xu¹, Xin Zhang¹, Hui Xiao^{2, 3**}, Xiaoliang Zhang³, Pengfei Lin⁴, Xiaodong Wang⁵

1. Government Service Center of Beijing, Municipal Transport Commission (Beijing Boats Inspection Center), Beijing 100161, China

2. School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

3. Research and Development Center of Transport Industry of Big Data Processing Technologies and Application for Comprehensive Transport (ZHONG LU GAO KE), Beijing 100088, China

4. College of Computer Science, Beijing University of Technology, Beijing 100124, China

5. Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China

Abstract: Fatigue driving represents a major cause of road traffic accidents, with particularly severe consequences in passenger transport. Developing accurate fatigue detection methods resilient to in-vehicle environmental interference is crucial for enhancing road safety. This study proposes a multi-feature fatigue detection framework integrating temporal information based on YOLOv8n-SimAM for bus drivers. The SimAM attention mechanism enhances feature extraction from ocular and oral regions while suppressing cabin background interference. By incorporating the driver's eye and mouth states with temporal information, we established a multi-feature fatigue reasoning mechanism. Comparative experiments demonstrate that YOLOv8n-SimAM achieves superior performance in precision, recall, and $P_{\text{mAP}50-95}$ (mean average precision across IoU thresholds 0.5–0.95) over alternative models. Our approach achieves enhancements of 2.8% in precision, 5.4% in recall, and 7.5% in $F_1\text{-score}$ (harmonic mean of precision and recall) over the suboptimal benchmark, without a substantial increase in model parameters. Ablation studies confirm that integrating SimAM reduces false positive rate by 48%, while the temporal fatigue inference module decreases missed detection rate by 72%. The integrated model attains an $F_1\text{-score}$ of 0.906, an accuracy of 0.9, with corresponding missed detection and false positive rates of 0.111 and 0.087, respectively. This methodology enables reliable facial feature capture and fatigue assessment in challenging in-cabin environments, providing an effective solution for off-site monitoring of bus driver fatigue.

Keywords: traffic engineering; multi-feature fatigue reasoning; YOLOv8n; bus driver; object detection; SimAM attention mechanism; temporal information

1 Introduction

Statistics indicate that fatigue-related driving accounts for approximately 20% of all traffic accidents, yet it contributes to nearly 80% of traffic fatalities [1]. Fatigue impairs drivers' attention, slows reaction times, and increases the risk of operational errors [2]. Due to the high passenger capacity, accidents involving passenger transport vehicles often lead to more severe consequences. As a result, such vehicles are considered high-risk components in road traffic safety regulation. They have been identified as key targets in the national campaign against "over-speeding, over-loading, over-fatigue, and fatigue driving". Furthermore, the "Safety Management Regulations for Road Passenger Transport Enterprises" encourage companies to adopt intelligent video monitoring systems to detect and manage fatigue driving behaviors among professional drivers.

Current research on fatigue-driving detection primarily follows three technical approaches: facial-feature analysis, physiological-signal analysis, and vehicle-operation-based assessment [3]. Meth-

ods based on facial features acquire a driver's facial information through cameras or other imaging sensors and determine fatigue status using computer vision and deep learning techniques [4]. Zhu et al. [5] proposed a real-time fatigue detection algorithm based on facial video sequences and achieved competitive performance in both accuracy and processing speed. To address the low yawning-detection accuracy caused by complex facial movements, Yang et al. [6] integrated 3D convolutional neural networks with bidirectional long short-term memory networks and introduced a key-frame selection mechanism to enhance detection efficiency. Feng Xiaofeng et al. [7] adopted a multimodal fusion strategy that incorporates eye, mouth, and head-pose features, and further optimized the network architecture to improve detection speed. Physiological-signal-based fatigue detection methods evaluate fatigue by extracting electrooculogram (EOG), electroencephalogram (EEG), electrocardiogram (ECG), and other bio-signals, offering strong robustness and low false-alarm rates. Gao et al. [8] proposed a hybrid recurrent-CNN framework for fatigue detection using EEG signals, demonstrating the potential of nonlinear dynamic analysis in EEG-based modeling. Zhang et al. [9] developed a fatigue-detection

Received: 26 November 2025; **Revised:** 11 January 2026; **Accepted:** 2 February 2026

* E-mail address: wukeqi@bjjtw.gov.cn; ** Corresponding author: 1520030192@qq.com

DOI: 10.26599/HTRD.2026.9480101

2095-6215/© The Author(s) 2026. Published by Tsinghua University Press. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



method that integrates EEG and blink-related potentials. Zhou et al. [10] constructed a brain-fatigue recognition system based on Bayesian-optimized gcForest and highlighted the importance of incorporating ECG and other physiological signals.

Vehicle-operation-based fatigue detection methods infer a driver's alertness level by analyzing vehicle dynamic data collected through onboard sensors [11]. These approaches rely on the core assumption that driving precision and control stability deteriorate when a driver becomes fatigued. Among these indicators, the steering wheel angle (SWA) is the most commonly used metric. Under fatigue, a driver's steering operations become slower and less coherent, characterized by fewer small corrective adjustments and more sudden, large-amplitude corrections [12]. Li et al. [13] achieved fatigue detection by analyzing the operational characteristics of SWA and yaw angle, as well as statistical features such as approximate entropy. Zhang et al. [14] constructed a fatigue classification model by combining support vector machines with steering-angle features. Zhang et al. [15] developed a model for identifying driving-risk influencing factors based on vehicle speed and driver fatigue level.

In summary, current fatigue-driving detection methods have achieved substantial progress. However, physiological-feature-based approaches typically rely on highly intrusive devices, making continuous monitoring during driving difficult. For vehicle-operation-based detection, the trade-off between model lightweight and accuracy has not yet been effectively addressed. In terms of facial-feature-based detection, compared with private vehicles, bus drivers operate in a more complex in-cabin environment with stronger background interference, leading to reduced detection accuracy. Meanwhile, fatigue is generally a temporally evolving state, and most existing fatigue-detection methods rely on single-frame facial features, which tend to cause high false-alarm rates and lack effective mechanisms for temporal fatigue-state inference.

This study proposes a multi-feature fatigue detection method for commercial vehicle drivers based on YOLOv8n-SimAM (Simple

Attention Module) integrated with temporal information. Specifically, a lightweight parameter-free attention mechanism SimAM [16] and a multi-feature fatigue reasoning module are systematically integrated into the YOLOv8n baseline model. By suppressing background interference, SimAM enhances the comprehensive detection performance of the model for small-scale targets such as the eyes and mouths of commercial vehicle drivers. On the basis of drivers' eye and mouth states as well as temporal information, and drawing on the principle of PERCLOS (Percentage of Eyelid Closure over the Pupil over Time), the multi-feature fatigue reasoning module achieves more efficient and accurate fatigue detection. This method enables the off-site monitoring of fatigued driving behaviors of commercial vehicle drivers, thereby improving the safety level of road transportation.

2 Multi-feature fatigue-driving detection model design

The proposed multi-feature fatigue detection method incorporating temporal information is developed on the framework of the YOLOv8n object detection model. The SimAM attention mechanism is introduced to enhance facial-feature extraction and suppress background interference. By integrating temporal feature modeling and using blinking frequency and prolonged mouth opening during yawning as fatigue indicators, a multi-feature fatigue inference mechanism is constructed to detect fatigue driving among bus drivers. The technical workflow is illustrated in Fig. 1.

2.1 YOLOv8n object detection algorithm

This study employs YOLOv8n—selected for its favorable balance between accuracy and speed—as the object detection model for extracting fatigue-related features from the driver's mouth and eyes. YOLOv8, released by the Ultralytics team, represents the latest generation of YOLO-series object detection frameworks [17], supporting object detection, image classification, and instance segmentation tasks. It can be deployed efficiently on standard comput-

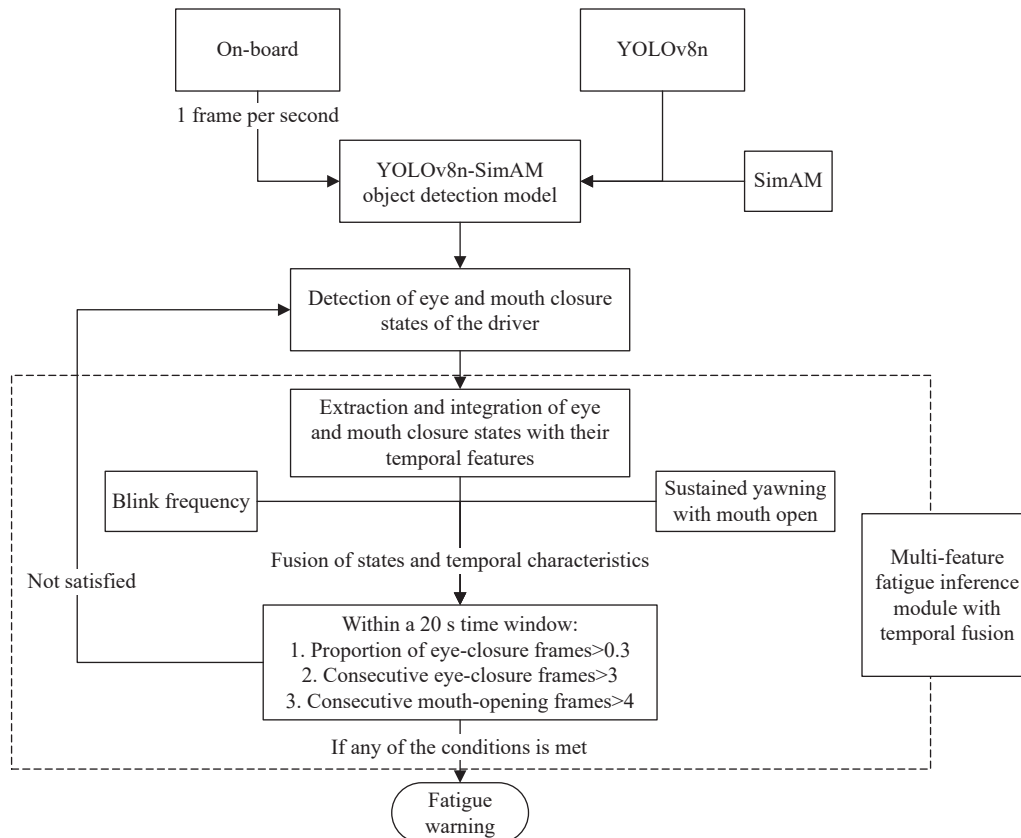


Fig. 1 Multi-feature fatigue driving detection model for passenger vehicle drivers

ing platforms. Building upon the strengths of previous YOLO versions, YOLOv8 further optimizes the trade-off between real-time performance and accuracy through network lightweighting, enhanced feature fusion, and innovative detection-head design.

The YOLOv8 model family includes YOLOv8n, YOLOv8s, YOLOv8m, and YOLOv8x, accommodating various application scenarios and performance requirements. YOLOv8n retains the classic three-stage architecture of the YOLO series (Backbone–Neck–Head) and introduces structural refinements to achieve efficient inference and lightweight design. Through simplified network depth, reduced channel dimensions, and optimized module structures, YOLOv8n preserves core YOLOv8 advantages while achieving extreme lightweighting and high-efficiency deployment, making it suitable for smooth operation on low-computing-power onboard devices in passenger vehicles.

As shown in Fig. 2, the Backbone of YOLOv8 adopts an improved CSP structure and replaces the C3 module with the C2f module. The C2f module maintains the cross-stage feature fusion capability of C3 while increasing branch connections to enrich feature representation, and uses lightweight design to reduce computational cost. At the end of the Backbone, an improved spatial pyramid pooling–fast (SPPF) module is introduced. Through multi-scale pooling operations (1×1 , 5×5 , 9×9 , 13×13), the SPPF enhances the model's ability to capture large-scale features while reducing time complexity via parallelized pooling.

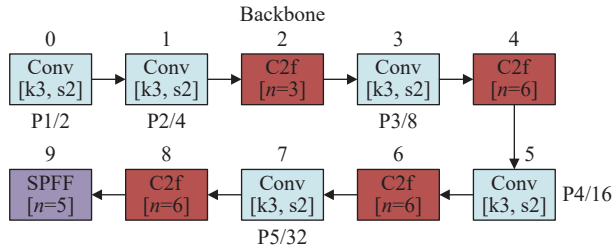


Fig. 2 Backbone network structure of YOLOv8

The Neck of YOLOv8 is optimized based on the PANet structure and adopts a bidirectional feature-fusion mechanism. High-level features are propagated top-down to enhance semantic information, while low-level features are propagated bottom-up to

supplement fine-grained details, enabling complementary cross-scale feature fusion. The C2f module is also employed in the Neck to replace traditional convolutional blocks, maintaining efficient feature transformation during feature fusion while avoiding excessive computational overhead. A 1×1 convolution is used to compress channel dimensions, reducing the computational cost of cross-layer connections and ensuring real-time performance during the feature-fusion process.

The Head of YOLOv8 abandons the traditional anchor-based design and adopts an anchor-free structure. It directly predicts object class probabilities and bounding-box coordinates on three feature maps of different scales (19×19 , 38×38 , and 76×76), eliminating the need for predefined anchor boxes, simplifying model design, and improving generalization to objects of previously unseen sizes. Two independent branches are used to handle object classification and bounding-box regression, respectively, avoiding gradient interference between tasks. The classification branch uses cross-entropy loss, while the regression branch employs Complete Intersection over Union (CIoU) loss to improve bounding-box localization accuracy.

2.2 SimAM attention mechanism

The attention mechanism is a powerful technique in deep learning that enables models to automatically focus on the most critical parts of the input data, thereby enhancing their ability to process salient information. Traditional neural networks typically increase expressiveness by stacking large numbers of parameters, which may instead reduce efficiency due to the resulting information overload. In contrast, attention modules dynamically assign weights to selectively strengthen features most relevant to the task while suppressing or discarding redundant information. This approach not only reduces unnecessary computation but also improves accuracy. The mechanism is analogous to human visual attention: the visual system scans an entire scene to locate regions of interest, then allocates more cognitive resources to those regions to obtain richer, task-relevant details, while ignoring irrelevant information.

As shown in Fig. 3, the computation process of the attention mechanism consists of three steps. First, the similarity between the Query and Key is calculated to obtain the attention scores. Second, the attention scores are normalized using the softmax function to

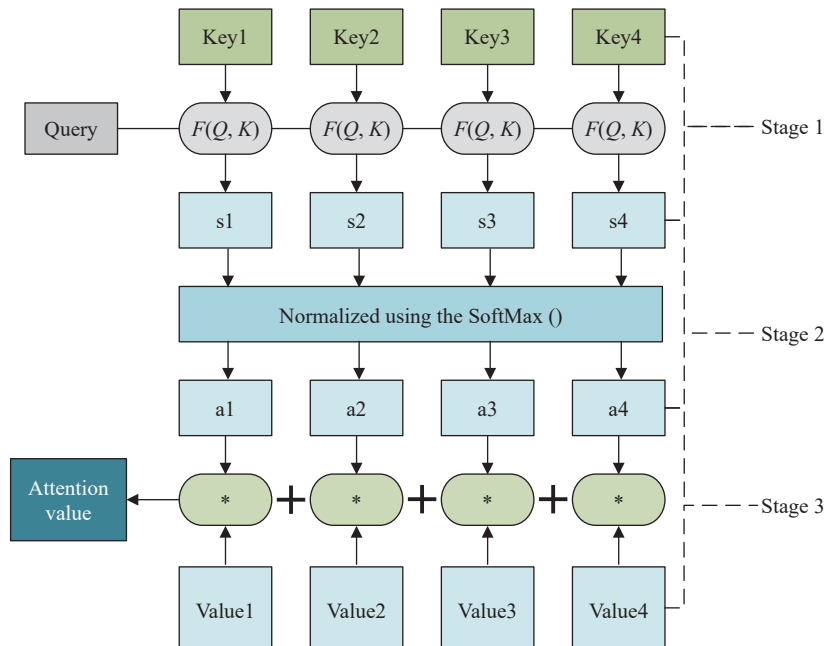


Fig. 3 Principle of attention mechanism

form a probability distribution in which all weight coefficients sum to 1, thereby determining the weights corresponding to important elements. Third, the Value vectors are weighted and summed according to these coefficients to produce the final attention output, denoted as the Attention Value. The above three steps can be concisely expressed by the following Eq. (1).

$$\begin{aligned} \text{att}((K, V), q) &= \sum_{i=1}^N \alpha_i v_i \\ &= \sum_{i=1}^N \frac{\exp(s(k_i, q))}{\sum_j \exp(s(k_j, q))} v_i \end{aligned} \quad (1)$$

where, α_i is the normalized attention weight corresponding to the i -th key-value pair; v_i is the i -th value vector extracted from feature maps; k_i is the i -th key vector extracted from feature maps; q is the single query vector to be matched.

In the field of object detection, attention mechanisms have become one of the core techniques for improving model performance. By introducing an appropriate attention mechanism, a model's feature representation capability can be enhanced, its ability to process multimodal data can be improved, and overall computational efficiency can be optimized. In this study, the SimAM (Simple Attention Module) attention mechanism is integrated into the YOLOv8n model. SimAM is a lightweight attention module that requires no additional parameters. While maintaining extremely low computational cost, it effectively strengthens the model's focus on target features, making it particularly suitable for embedded vision tasks. SimAM is inspired by the activation properties of biological neurons and designs an energy function based on spatial inhibition theory to evaluate the "energy consumption" of each neuron as a measure of its importance. Based on the computed attention importance, SimAM assigns corresponding attention weights to each neuron, thereby enhancing the representational capacity of the feature maps. By considering interactions across channel, spatial, and depth dimensions, SimAM can automatically focus on the internal structural characteristics of the target. The computation of SimAM's attention weights is based on the following energy function:

$$e_i = \frac{(x_i - \mu)^2}{2(\sigma^2 + \epsilon)} + \lambda \cdot (x_i - t)^2 \quad (2)$$

where, x_i is the activation value of the i -th neuron; μ and σ are the mean and standard deviation of the feature map, respectively (i.e., global statistics); t is the target value (typically set to 0 to suppress background activations); λ is a balancing parameter (default 1×10^{-4}), which controls the relative importance of the two terms.

The first term measures the deviation of the neuron activation from the mean, while the second term penalizes the deviation from the target value. The energy value is then converted into a weight within the range [0,1] through a Sigmoid function, where a larger value indicates greater importance of the corresponding neuron.

Compared with other commonly used attention mechanisms such as CBAM (Convolutional Block Attention Module), SE (Squeeze-and-Excitation), and ECA (Efficient Channel Attention), SimAM computes attention weights directly from statistical measures without introducing additional parameters. This makes it highly suitable for lightweight models and avoids deployment challenges caused by parameter expansion. SimAM simultaneously considers relationships across channels, spatial pixels, and depth features, making it particularly effective for capturing fine-grained structures of the eyes and mouth.

In the YOLOv8n object detection model, the SPPF module in the Backbone performs rapid spatial pyramid pooling to fuse feature maps of different scales and obtain richer contextual information. After multi-scale fusion through SPPF, further screening and

enhancement of these fused features are required to assign higher attention weights to small targets, enabling the network to better focus on task-relevant regions (such as the driver's mouth and eyes) while suppressing background noise. This improves the detection accuracy of small objects and reduces missed and false detections. In this study, the SimAM attention module is added after the SPPF in the Backbone of the YOLOv8n model. Since SimAM is a parameter-free lightweight attention mechanism, it is well aligned with the design goals of YOLOv8n, which emphasizes efficiency and compactness. SimAM enhances model performance without significantly increasing computational cost or model size, supporting lightweight deployment on onboard devices in passenger vehicles and meeting real-time operational requirements.

2.3 Multi-feature fatigue inference module with temporal information

2.3.1 Physiological fatigue features and detection indicators

The detection of fatigue relies on capturing physiological characteristics of the human body, which, once quantified through specific indicators, can be used as the basis for evaluating fatigue levels. Compared with behavioral features such as lane deviation or vehicle speed fluctuation, physiological features exhibit stronger objectivity and are less susceptible to intentional manipulation. According to existing research on fatigue-driving detection, typical physiological indicators include P_{clos} (percentage of eyelid closure over the pupil over time), the duration of a single yawning event, blink frequency, and EEG (electroencephalogram) band power ratios.

P_{clos} is widely used to assess driver fatigue [18]. Generally, when the driver is in an alert state, the eyes remain relatively stable and open; however, as fatigue accumulates, involuntary eyelid closure increases, leading to a higher P_{clos} value. Continuous monitoring of the driver's eye state enables quantification of eyelid closure degree and duration, thereby reflecting the driver's fatigue level. Borghi classified driver states into three categories based on P_{clos} values: alert (< 0.3), drowsy ($0.3 < P_{\text{clos}} < 0.7$) [19], and inattentive (> 0.7). In practical experiments, P_{80} is commonly adopted (i.e., the eye is considered closed when the eyelid covers more than 80% of the pupil area). The proportion of eye-closure duration within a given time window is then computed, as expressed in the following formula.

$$P_{\text{clos}} = \frac{t_3 - t_2}{t_4 - t_1} \quad (3)$$

where, $t_3 - t_2$ is the time required for the pupil to transition from a 20% open state to a 20% open state; $t_4 - t_1$ is the time required for the pupil to transition from an 80% open state to an 80% open state.

Yawning is a common physiological response during fatigue, characterized primarily by an involuntary mouth-opening action accompanied by deep inhalation and facial muscle stretching. It is widely regarded as one of the most distinguishable physiological features for fatigue-driving detection [20]. According to Sambale, a typical yawn lasts approximately 4–7 s [21] and consists of three phases: (1) an inhalation phase during which the mouth gradually opens; (2) a brief peak phase with strong muscle stretching; (3) a rapid exhalation phase with muscle relaxation. Moreover, as fatigue deepens, the number of yawns within a given period increases significantly.

Blink frequency and duration also exhibit notable changes under fatigue [22]. The blink frequency f_b is computed as shown in Eq. (4). Physiologically, eye-blink behavior is characterized as a transient process involving the rapid closure of the eyelids, whose occurrence frequency and duration characteristics exhibit notable alterations under drowsy conditions. However, the accurate measurement of such eye-blink dynamic characteristics typically relies on high-frame-rate video data. Limited by the 1 Hz sampling fre-

quency of the dataset adopted in this study, it is difficult to directly capture the transient process of a single eye blink. Therefore, this study does not perform rigorous calculations on the eye-blink frequency or the duration of a single eye blink. Instead, based on the changes in eyelid opening and closing states within discrete frame sequences, an approximate indicator reflecting the cumulative variation in eyelid closure degree over time is constructed and applied to drowsy state reasoning.

$$f_c = \frac{M}{T} \quad (4)$$

where, T is a time window; M is the total number of blinks within this time window.

EEG band-power ratio parameters can reflect variations in human reaction time and have demonstrated high accuracy in fatigue detection [23]. As they directly indicate brain activity, these parameters are capable of distinguishing between mild and severe fatigue levels. Among them, θ -wave power increases with rising fatigue levels, whereas α -wave power decreases; therefore, the θ/α ratio is positively correlated with fatigue severity. When the θ/α ratio $p \geq 1.2$, the driver's reaction time typically increases by more than 20%. However, in real driving scenarios, EEG acquisition requires contact electrodes (e.g., headband-type sensors), which can compromise driver comfort and are susceptible to motion artifacts caused by head movements, limiting their practical applicability.

2.3.2 Multi-feature fatigue inference mechanism with temporal information

Fatigue-driving detection based on a single feature or single-frame images presents clear limitations in passenger-vehicle driver monitoring. Individual differences among drivers and poor adaptability to varying facial characteristics often reduce detection accuracy and lead to higher false-alarm and missed-detection rates. To further enhance the reliability of fatigue detection for bus drivers, this study leverages the intrinsic abnormal variations exhibited in the eyes and mouth under fatigue. Using the YOLOv8n-SimAM model, eye and mouth states—as well as their temporal information—are detected and extracted. Drawing on the P_{clos} principle, temporal variations in blink frequency and yawning behavior under fatigue are incorporated to construct a multi-feature fatigue inference mechanism.

The driver's eye and mouth states are first initialized to “eyes open” and “mouth closed”, respectively. The trained YOLOv8n-SimAM model is then used to detect the real-time states of the eyes and mouth. A confidence threshold of 0.6 is introduced; detections below this threshold are discarded to avoid misclassification (e.g., preventing a blurred “open-eye” image from being misjudged as “closed-eye”). The dataset employed in this study was captured by vehicle-mounted cameras at a sampling rate of 1 frame per second and is presented in a segmented format. Thus, the drowsiness discrimination was implemented based on discrete frame sequences rather than continuous time signals. Within a temporal window of 20 frames, fatigue is determined if any of the following conditions are met:

(1) The traditional P_{clos} metric is typically derived from continuous video data with high temporal resolution or acquired via professional eye trackers. Given the constraints of low sampling rate in this study, a discrete approximation of this metric is implemented. Specifically, the drowsiness criterion is satisfied when a high proportion of eye closure states is observed within the unit time window and the frame-level P_{clos} value exceeds 0.3, as expressed in Eq. (5). That is, within a unit time window consisting of 20 frames, the driver is classified as drowsy if the proportion of frames corresponding to eye closure is greater than 0.3.

$$P_{\text{clos}} = \frac{t_{\text{ec}}}{T} \quad (5)$$

where, t_{ec} is the number of frames in which the driver's eyes are closed within a unit time window; T is the total number of frames in that time window.

Herein, the frame-level P_{clos} is defined as the ratio of the number of frames classified as eye closure states to the total number of frames within a given time window.

(2) Under normal awake conditions, the duration of a single eye blink is typically much shorter than 1 s. Thus, under the 1 Hz sampling rate, eye-blink behaviors can rarely be fully captured across multiple consecutive frames. In contrast, when a driver is in a drowsy or fatigued state, the eyes tend to exhibit sustained closure or semi-closure behaviors with a duration of several seconds. Based on the aforementioned characteristics, the detection of eye closure states in three or more consecutive frames is defined as a prolonged eye-closure event. The probability of such events occurring under awake conditions is extremely low, which renders them one of the key criteria for drowsiness state discrimination.

(3) When the driver keeps the mouth open for four or more consecutive frames, this is considered a deep yawn and is treated as a fatigue condition. Moreover, if the mouth region cannot be detected (i.e., confidence score equals 0), it is assumed that the driver may be covering the mouth with a hand during yawning.

3 Experiments and results analysis

3.1 Data source and preprocessing

To better align with real-world fatigue detection scenarios for passenger vehicle drivers, this study constructs the foundational dataset using a publicly available dataset of bus drivers' facial features. The dataset was collected via in-vehicle cameras and contains the facial characteristics of bus drivers under various driving states, primarily focusing on fatigue-related features such as mouth and eye closures, including yawning and eye-closure behaviors. A total of 2,914 samples are included. These facial features under fatigue conditions are crucial indicators for determining whether a driver is experiencing fatigue. Examples of driver facial features corresponding to fatigue states are shown in Fig. 4.

However, the original dataset has several limitations, including limited variation in lighting conditions and a single camera angle. In practical fatigue detection scenarios, lighting conditions vary dynamically during vehicle operation—for instance, strong light under direct sunlight and insufficient illumination inside tunnels. Moreover, the mounting positions of in-vehicle cameras differ across passenger vehicles, resulting in diverse shooting angles. To address these issues, multiple data augmentation strategies were applied during the YOLOv8 model training process. These include



Fig. 4 Example of a bus driver fatigue driving dataset

geometric transformations, color transformations, and mixed augmentation techniques such as random flipping, random cropping, padding, random brightness and contrast adjustment, random hue adjustment, and mosaic augmentation.

Some augmented images generated during training are shown in Fig. 5, illustrating the visual effects produced by different augmentation operations. These augmentation strategies ensure that the training data sufficiently capture possible fatigue manifestations under diverse environments and viewpoints. By enriching the diversity of training samples, the model's robustness to variations in environment, driver posture, and target scale is significantly improved, ultimately enhancing its detection accuracy and generalization performance in real-world applications.

The facial physiological features of drivers in the original dataset were manually classified and annotated. To ensure the consistency between annotations and images, 10% of the samples were randomly selected for verification. The annotations include four states related to the driver's eyes and mouth: open eyes, closed eyes, open mouth, and closed mouth. Based on the timestamp watermark embedded in the images, it is evident that the dataset was recorded at a rate of one frame per second, capturing the facial physiological features of two drivers under different driving conditions in segmented sequences. Accordingly, the dataset was divided into a training set and a validation set at an approximate ratio of 7:3, following the temporal patterns of the two drivers during different driving periods. The validation set consists of a com-

plete driving segment, meaning a continuous time sequence of facial feature recordings from the same driver. This partitioning strategy enables more effective utilization of the YOLOv8 model during training, facilitating the learning of physiological features of the eyes and mouth under fatigue conditions. As a result, the model can achieve more accurate and efficient fatigue driving detection.

3.2 Experimental environment and parameter settings

The model training was conducted on a server equipped with an Intel(R) Xeon(R) Silver 4 216 CPU @ 2.10 GHz (dual processors), an NVIDIA GeForce RTX 2 080 Ti GPU (11 GB), and 64.0 GB of RAM (63.7 GB available). CUDA version 12.9 was used to accelerate the training process, and Python 3.8.5 served as the development environment.

During model training, a warm-up strategy was employed to prevent the initial high learning rate from destabilizing or damaging the model parameters. The SGD optimizer was used for adaptive parameter updates, with the initial learning rate set to 0.01. Validation was conducted simultaneously throughout the training process. An early stopping mechanism was adopted: if the performance on the validation set did not improve for 50 consecutive epochs, training was automatically terminated to prevent overfitting and unnecessary iterations. After validating with the early stopping mechanism, the number of training epochs was set to 150. Experimental environment and parameter settings are shown in Table 1.



Fig. 5 Datasets under the data enhancement strategy

Table 1 Experimental environment and parameter settings

Parameter settings	Specification / value
Processor	Intel(R) Xeon(R) Silver 4 216 CPU
GPU	NVIDIA GeForce RTX 2 080 Ti (11 GB)
CUDA	12.9
Python version	3.8.5
Input image size	640×640
Warm-up epochs	3
Training epochs	150
Batch size	16
Initial learning rate	0.01
Number of data loading workers	6
Early stopping patience	50

3.3 Evaluation metrics

The performance of the YOLOv8n-SimAM model is evaluated using Precision (P), Recall (R), P_{mA50} (mean average precision at 50% IoU), and $P_{mA50-95}$ (mean average precision across IoU thresholds from 50% to 95%) to quantify and compare the effectiveness of the improved model in detecting the driver's eye and mouth states. P_{mA50} reflects the model's ability to detect targets under a relatively lenient bounding-box matching criterion, whereas $P_{mA50-95}$ represents the overall detection accuracy across IoU thresholds of increasing strictness. Higher $P_{mA50-95}$ values indicate more accurate bounding-box predictions and stronger capability in distinguishing visually similar targets (e.g., slight mouth opening vs. closed mouth, half-closed eyes vs. fully closed eyes).

The multi-feature fatigue-driving inference mechanism is evaluated using P , R , $F_{1-score}$, A , R_M (miss rate), and R_{FA} (false alarm rate), which together assess the applicability and reliability of the proposed inference mechanism.

The $F_{1-score}$ represents the harmonic mean of precision and recall, providing a balanced evaluation of model performance and avoiding bias that may arise when examining a single metric in isolation.

The formulas for P , R , A , and R_M are given in Eqs. (6), (7), (8), and (9), respectively.

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (6)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (7)$$

$$A = \frac{N_{TP}}{N_{TP} + N_{FN} + N_{FP} + N_{TN}} \quad (8)$$

$$M = \frac{N_{FP}}{N_{FP} + N_{TN}} \quad (9)$$

where, N_{TP} is the number of fatigue samples correctly predicted; N_{FN} is the number of fatigue samples incorrectly predicted; N_{FP} is the number of normal samples incorrectly predicted; N_{TN} is the number of normal samples correctly predicted.

The calculation formula for the P_{mA} is provided in Eq. (10).

$$P_{mA} = \frac{\sum_{i=1}^K P_{Ai}}{K} \quad (10)$$

where, P_{Ai} is the average precision of the model across different detection thresholds; K is the number of target categories.

3.4 Results analysis

3.4.1 Performance evaluation of YOLOv8n-SimAM

In this study, P , R , P_{mA50} , and $P_{mA50-95}$ were adopted as the evaluation metrics. To verify the superiority of the proposed model, controlled experiments were conducted on the identical dataset, where the YOLOv8n-SimAM model was compared with light-

weight models from the YOLO series (i.e., YOLOv3-tiny, YOLOv5nu, YOLOv8n) as well as their corresponding variants integrated with the SimAM attention mechanism (i.e., YOLOv3-tiny-SimAM, YOLOv5nu-SimAM).

As presented in Table 2, all six models exhibit high and comparable performance levels in terms of P_{mA50} . This indicates that under the lenient detection scenario with an IoU threshold of 0.5, the basic recognition coverage and accuracy of these models are sufficient to stably capture most positive samples for targets such as eye and mouth closure states. After the integration of the SimAM attention mechanism, YOLOv3-tiny-SimAM and YOLOv5nu-SimAM achieve certain improvements in Precision and Recall, yet with no notable improvements observed in their $P_{mA50-95}$ values. In contrast, the incorporation of the SimAM attention mechanism into YOLOv8n results in a distinct performance boost: its Precision is increased to 0.950, Recall to 0.962, and $P_{mA50-95}$ to 0.6065, all of which outperform the other five target detection models. Although the overall performance improvement margin is relatively modest, under the identical training and testing conditions, YOLOv8n demonstrates consistent upward trends in Precision, Recall and $P_{mA50-95}$ after the integration of SimAM. This finding verifies that the SimAM mechanism exerts a positive facilitation effect on the current task, enhancing the model's fine-grained recognition capability for eye and mouth closure states, and thus highlighting the comprehensive performance of the YOLOv8n-SimAM model.

Table 2 Comparison of object detection model performance

Model	P	R	P_{mA50}	$P_{mA50-95}$
YOLOv3-tiny	0.917	0.935	0.974	0.596
YOLOv5nu	0.935	0.950	0.978	0.604
YOLOv8n	0.94	0.947	0.9793	0.6047
YOLOv3-tiny-SimAM	0.935	0.948	0.976	0.597
YOLOv5nu-SimAM	0.945	0.955	0.977	0.604
YOLOv8n-SimAM	0.950	0.962	0.9795	0.6065

As shown in Fig. 6, the performance of the proposed YOLOv8n-SimAM model in driver facial state detection can be further illustrated using a confusion matrix. The model demonstrates strong robustness in detecting the closed mouth (1.00), opened mouth (0.97), and open-eyed (0.94) classes, indicating good learning and generalization capabilities for high-sample, highly distinguishable facial features. The closed eye class (0.83), although relatively lower than the other three classes, reflects the inherent difficulty of detecting eye closure states due to subtle eyelid deformations (e.g., partially closed eyes) causing feature ambiguity.

3.4.2 Performance evaluation of multi-feature fatigue driving detection with temporal information

To verify the effectiveness and reliability of the proposed multi-feature drowsy driving detection method based on the YOLOv8n-SimAM model, a sequence sample containing 1 000 consecutive frames was selected from the experimental dataset. On the basis of the existing annotation information of drivers' eye and mouth physiological states in the original dataset, the drowsy driving state corresponding to each frame of image in the sequence was labeled frame by frame. This annotation strategy effectively eliminates the annotation errors caused by transient non-drowsy behaviors such as unconscious mouth opening and eye blinking of drivers, thus ensuring the accuracy and consistency of drowsy state labels in the dataset. Representative models for facial-feature-based fatigue driving detection were used as benchmarks, including Faster-RCNN [24], VA-YOLO [25], and SMGA-Net [26].

The results, summarized in Table 3, compare the four models in terms of P , R , $F_{1-score}$, and P_m (model parameters). The proposed YOLOv8n-SimAM-based multi-feature fatigue driving detection method achieves the best performance, with a P of 0.923, R of 0.889, and $F_{1-score}$ of 0.906. Compared with the second-best model,

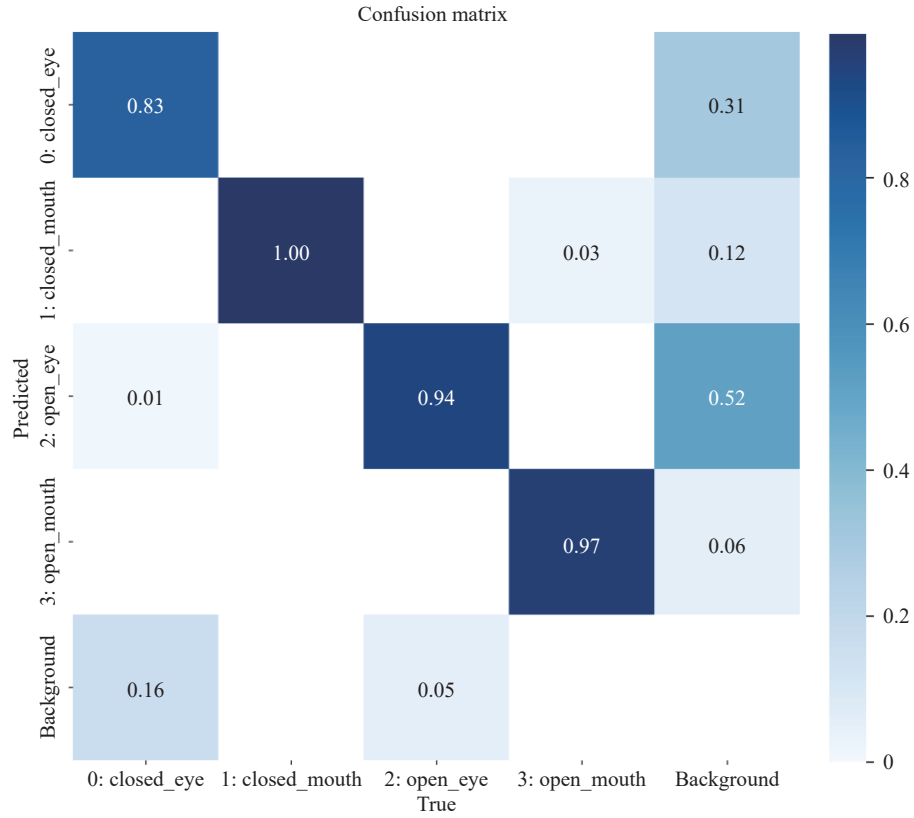


Fig. 6 Confusion matrix results

Table 3 Performance comparison of fatigue detection models

Model	P	R	$F_{1-score}$	$P_m (\times 10^6)$
Faster-RCNN	0.849	0.835	0.843	23.5
VA-YOLO	0.895	0.764	0.824	2.19
SMGA-Net	0.810	0.812	0.813	30.24
YOLOv8n-SimAM Multi-Feature Fatigue Detection	0.923	0.889	0.906	5.77

VA-YOLO, P is improved by 2.8%, while R and $F_{1-score}$ are improved by 5.4% and 7.5% relative to Faster-RCNN, respectively.

In terms of model lightweight design, the proposed method has 5.77×10^6 parameters, slightly higher than VA-YOLO's 2.19×10^6 , but significantly lower than Faster-RCNN's 23.5×10^6 and SMGA-Net's 30.24×10^6 , while maintaining clear advantages in fatigue detection accuracy. Combining the above analysis, the proposed model leverages the SimAM attention module to enhance the extraction of facial features under fatigue conditions and integrates a temporal reasoning mechanism based on multi-feature fusion. This approach ensures high precision and low false alarm rates in fatigue state recognition while maintaining lightweight deployability.

3.4.3 Analysis of ablation experiments

An ablation study was conducted to evaluate the effectiveness of the proposed YOLOv8n-SimAM-based multi-feature fatigue detection method for bus drivers, which incorporates temporal information. P , R , $F_{1-score}$, A , R_M , and R_{FA} (false alarm rate) were used as evaluation metrics. Four experimental configurations were designed to analyze the contributions of the SimAM attention mech-

anism and the multi-feature fatigue inference module with temporal information:

- (1) YOLOv8n.
- (2) YOLOv8n + SimAM attention mechanism.
- (3) YOLOv8n + multi-feature fatigue inference module with temporal information.
- (4) YOLOv8n + SimAM attention mechanism + multi-feature fatigue inference module with temporal information.

These configurations allowed for a systematic investigation of the impact of each component on fatigue detection performance.

The results, shown in Table 4, indicate that when using YOLOv8n alone as the feature detector, the model achieves a P of 0.857, but R is only 0.462, $F_{1-score}$ is 0.6, and the R_M reaches 0.538. This demonstrates that a single-object detection model is insufficient for capturing multi-modal fatigue features, such as consecutive eye closures and yawning, making it difficult to cover the manifestations of fatigue in complex driving scenarios and resulting in a significant risk of missed detections. After introducing the SimAM attention mechanism into YOLOv8n, the model's P increases to 0.923, and the R_{FA} decreases by 48%, from 0.083 to 0.043. This indicates that the parameter-free spatial suppression mechanism in SimAM effectively enhances the model's focus on critical eye and mouth features, reducing misclassification of non-fatigue features. When YOLOv8n is integrated with the multi-feature fatigue inference module that incorporates temporal information, R increases by 84%, from 0.462 to 0.852, the R_M decreases by 72%, from 0.538 to 0.148, and the $F_{1-score}$ improves by 44.7%, from 0.6 to 0.868. The

Table 4 Ablation experimental results of multi-feature fatigue driving detection model

Model Configuration			Evaluation Metrics					
YOLOv8n	SimAM	Multi-Feature temporal inference module	P	R	$F_{1-score}$	A	R_M	R_{FA}
√	×	×	0.857	0.462	0.6	0.68	0.538	0.083
√	√	×	0.923	0.444	0.6	0.68	0.556	0.043
√	×	√	0.885	0.852	0.868	0.86	0.148	0.13
√	√	√	0.923	0.889	0.906	0.9	0.111	0.087

multi-feature fatigue inference module reduces missed detections by analyzing the driver's eye and mouth states and extracting corresponding temporal features for inference. P also increases from 0.857 to 0.885, indicating that multi-feature constraints enhance classification accuracy and reduce misjudgments caused by single-feature detection. When the SimAM attention mechanism is combined with the multi-feature temporal inference module, the over-

all model achieves an $F_{1\text{-score}}$ of 0.906, A of 0.9, R_M of 0.111, and R_{FA} of 0.087, with all metrics at optimal levels. This demonstrates the SimAM mechanism's ability to precisely capture key features, reducing misclassification, while the multi-feature module effectively lowers missed detection of fatigue states. Together, they achieve high-precision, low-miss, and low-false-alarm fatigue detection performance, as illustrated in Fig. 7.



Fig. 7 Detection effect of multi-feature fatigue inference mechanism

4 Conclusion and future work

This study proposes a multi-feature drowsiness detection method for commercial vehicle drivers based on YOLOv8n-SimAM integrated with temporal information. Specifically, the SimAM attention mechanism is incorporated to enhance the model's capability of detecting drivers' eye and mouth features while suppressing background interference in the driving environment. Furthermore, by fusing the drivers' eye and mouth states with temporal information, a multi-feature drowsy driving reasoning mechanism is established. Compared with single-frame image-based drowsiness detection methods, the proposed approach improves the practical applicability of the model in the scenario of commercial vehicle driver drowsiness detection. For the target detection model, the integration of the SimAM attention mechanism yields slight improvements in P , R and $P_{mAS0-95}$ metrics, thereby achieving a certain enhancement in the comprehensive detection accuracy of the model. Comparative experiments demonstrate that, under the condition of fewer model parameters, the proposed model outperforms the baseline model with P , R and $F_{1\text{-score}}$ improved by 2.8%, 5.4% and 7.5%, respectively. Ablation experiments verify that the overall performance of the drowsy driving detection model is improved to varying degrees under the synergistic effect of the SimAM attention mechanism and the multi-feature drowsy driving reasoning mechanism integrated with temporal information, with the $F_{1\text{-score}}$ reaching 0.906, P reaching 0.9, R_M reaching 0.111, and R_{FA} reaching 0.087. In particular, the missed detection rate of drowsiness detection is reduced by 72%, and the $F_{1\text{-score}}$ is increased by 44.7%.

Future research efforts could focus on constructing self-built datasets or collecting additional datasets that include diverse driver cohorts with higher temporal resolution and continuous temporal information, so as to improve the generalization capability of the proposed multi-feature drowsiness detection method integrated with temporal information. In addition, the drowsiness reasoning mechanism can be further enriched by fusing multi-modal data such as drivers' head pose parameters and vehicle steering wheel angle data, which is expected to enhance the robustness of the method in complex scenarios such as commercial vehicle driver monitoring. Furthermore, it is necessary to optimize the network architecture of the target detection model in depth and systematically evaluate the stability of the model performance through statist-

ical analysis, so as to realize more accurate and reliable drowsy driving detection.

Acknowledgements

None.

Funding

This research was sponsored by the Science and Technology Project of Beijing Municipal Commission of Transport (Grant No. BJJTW-2024-DSJ-16).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Author contribution statement

Keqi Wu: conceptualization, data curation, funding acquisition, methodology, software, writing—original draft, writing—review & editing. Shuo Xu: data curation, methodology, investigation. Xin Zhang: data curation, investigation, methodology, software, writing—original draft. Hui Xiao: formal analysis, project administration, supervision, writing—review & editing. Xiaoliang Zhang: formal analysis, investigation, resources. Pengfei Lin: investigation, validation, writing—original draft. Xiaodong Wang: investigation, validation, writing—editing.

Data availability statement

The authors do not have the authority to share them.

Use of AI Statement

During the preparation of this manuscript, ChatGPT was adopted to polish English grammar, sentence logic and overall readability. All text, figures, experimental analysis and conclusions in this paper were independently designed, written and verified by the authors. The authors thoroughly revised and proofread the full text after AI-assisted language optimization, and bear full responsibility.

ity for all academic content and viewpoints of this publication.

References

- [1] Zeng J, Wang K, Hu X, et al. Study on driver fatigue detection system in-loop test technology. *Journal of Highway and Transportation Research and Development*, 2021, 38(5): 123–128,137.
- [2] Li B, Wang M, Wang L, et al. Experimental research on the impact of driving time on commercial driver's driving ability. *Journal of Highway and Transportation Research and Development (English Edition)*, 2007, 2(2): 89–93.
- [3] Yang JC, Wei F, Lin L, et al. A research survey of driver drowsiness driving detection. *Journal of Shandong University(Engineering Science)*, 2024, 54(2): 1–12.
- [4] Pei YL, Cai XX, Liu YH. Study on method for convenient detecting long-distance bus driver's fatigue state before driving. *Journal of Highway and Transportation Research and Development*, 2020, 37(1): 131–140.
- [5] Zhu TJ, Zhang C, Wu T, et al. Research on a real-time driver fatigue detection algorithm based on facial video sequences. *Applied Sciences*, 2022, 12(4): 2224.
- [6] Yang H, Liu L, Min W, et al. Driver yawning detection based on subtle facial action recognition. *IEEE Transactions on Multimedia*, 2020, 23: 572–583.
- [7] Feng XF, Fang B. Online fatigue detection of vehicle drivers based on facial features. *Mechanical Science and Technology for Aerospace Engineering*, 2021, 40(11): 1767–1772.
- [8] Gao ZK, Li YL, Yang YX, et al. A recurrence network-based convolutional neural network for fatigue driving detection from EEG. *Chaos*, 2019, 29(11): 113126.
- [9] Zhang BT, Yan DC, Wang XM, et al. Fatigue driving detection method based on the fusion of EEG and blink potential features. *Journal of Railway Science and Engineering*, 2025, 22(8): 3738–3747.
- [10] Zhou Y, Chen PK, Fan YF, et al. A multimodal feature fusion brain fatigue recognition system based on Bayes-gcForest. *Sensors*, 2024, 24(9): 2910.
- [11] Chu WL, Guo P, Zhang J, et al. A review of fatigue driving detection methods for motor vehicle drivers. *Electronic Design Engineering*, 2025, 33(4): 36–41.
- [12] Xi JF, Wang SQ, Ding TQ, et al. Detection model on fatigue driving behaviors based on the operating parameters of freight vehicles. *Applied Sciences*, 2021, 11(15): 7132.
- [13] Li ZJ, Chen LK, Peng J, et al. Automatic detection of driver fatigue using driving operation information for transportation safety. *Sensors*, 2017, 17(6): 1212.
- [14] Zhang L, Yang DM, Ni HF, et al. Driver fatigue detection based on SVM and steering-wheel angle characteristics. *Society of Automotive Engineers (SAE)-China Congress*. Singapore: SpringerSingapore, 2017: 729-738.
- [15] Zhang WH, Zhao JF, Luo WW, et al. Identification model of driving risk factors on accident-prone road sections based on DEMATEL-ISM. *Journal of Traffic and Transportation Engineering(Information Technology)*, 2015, 13(1): 13–17.
- [16] Yang J, Jiang YX, Xiong XY. Combining Transformer and SimAM lightweight pavement damage detection algorithms. *Journal of Railway Science and Engineering*, 2024, 21(9): 3911–3920.
- [17] Varghese R, Sambath M. Yolov8: A novel object detection algorithm with enhanced performance and robustness. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*. IEEE, 2024: 1-6.
- [18] Xu LM. Design of a fatigue driving detection system based on improved PERCLOS. *Modern Electronics Technique*, 2023, 46(22): 41–45.
- [19] Borghi G, Pini S, Vezzani R, et al. Mercury: a vision-based framework for driver monitoring. *Intelligent Human Systems Integration 2020*. Cham: Springer International Publishing, 2020: 104-110.
- [20] Qian LB, Xu JB, Xu HL, et al. A fatigue detection fusion method based on facial features of industrial vehicle drivers. *High Technology Letters*, 2025, 35(4): 409–418.
- [21] Sambale J, Koehler U, Hildebrandt W, et al. Yawning: a review of the literature. *Somnologie*, 2025, 29: 114–118.
- [22] Tu Z, Pang Y. The effect of blink rate and visual setting on the symptoms of digital eye strain. *Investigative Ophthalmology & Visual Science*, 2023, 64(8): 4000.
- [23] Lin XF, Huang ZQ, Ma WF, et al. EEG-based driver drowsiness detection based on simulated driving environment. *Neurocomputing*, 2025, 616: 128961.
- [24] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(6): 1137–1149.
- [25] Yin LF, Ding ZY. Lightweight research on fatigue driving face detection based on YOLOv8. *Recent Advances in Computer Science and Communications*, 2026, 19(2): e26662558315127.
- [26] Huang YM, Liu CS, Chang FL, et al. Self-supervised multi-granularity graph attention network for vision-based driver fatigue detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024, 8(4): 3067–3080.