Analysis And Comparison on Image-Based Fatigue State Detection Methods

Hao Liu *

School of Information Science and Technology, Beijing University of Technology, Beijing, 100124, China

* Corresponding Author Email: Aurora@emails.bjut.edu.cn

Abstract. Nowadays, fatigue driving has become one of the main causes of road traffic accidents worldwide. To ensure people's safety, the academic community has continuously proposed and applied new fatigue detection methods, but there are still many limitations in current research. Manual extraction of fatigue features cannot obtain more detailed information; individual differences among different subjects may lead to reduced accuracy of fatigue detection; the detection delay is high, and the real-time performance needs to be improved...In response, this paper presents a comparative review of image-based fatigue detection methods. It systematically analyzes recent advances by comparing their core processes, including datasets, feature extraction techniques (both handcrafted and deep learning-based), and classification models. This research finds that Convolutional Neural Network(CNN) has gradually matured in the feature extracting domain. Many studies have built upon this foundation, introducing improvements and innovations. These studies extract features through machine learning and deep learning, and perform fatigue state classification using various combined models. The innovation of personalized thresholds allow flexible adjustment of judgment threshold according to each person's characteristics to increase accuracy, and the introduction of OpenCV, Dlib libraries and YOLO models has significantly enhanced real-time performance... The systematic summary provided in this study offers support for researchers to further optimize fatigue detection models, and also provides fatigue detection references for car manufacturers and high-risk operation industries. These contributions are conducive to ensuring the safety of relevant personnel.

Keywords: Fatigue detection; Computer vision; Deep learning; Feature fusion; Facial landmarks.

1. Introduction

Fatigued driving is widely recognized as the primary factor contributing to road accidents worldwide. According to statistics, the road traffic accidents directly caused by fatigue driving in one year account for 37% of the total traffic accidents in China, and the mortality rate is as high as 83%. The problem has become a major threat to public health and transport systems globally. To mitigate this problem, image-based fatigue detection methods have been actively developed. A prevalent approach in these methods leverages eye-state analysis, as when a person is in a state of fatigue, the frequency of blinking decreases. The opening and closing state of the eyes can be measured by calculating the Eye Aspect Ratio (EAR). In recent years, numerous advances have significantly enriched the field of image-based fatigue detection. In the process of image preprocessing and basic face region localization, OpenCV is commonly used to process video frames and adapt to image formats, and the dlib library is responsible for face region localization and cropping [1]. Feature extraction techniques are broadly categorized into handcrafted features (common in traditional machine learning) and automated feature learning (enabled by deep learning). Convolutional neural network (CNN) has become the core model for feature extraction [2]. These innovative measures include techniques that can enhance the model's adaptability to changes in input [3], the ability of deep learning to capture subtle fatigue-related features [4], and the integration of efficient architectures for real-time performance [5]. However, there is still a gap in the research on the systematic review of fatigue detection algorithm models, and few studies have compared and analyzed the feature extraction and classification mathods of different experiments. Given the plethora of innovative yet disparate methods, a clear framework for selecting the most appropriate

technique based on specific requirements is lacking. Addressing this gap is crucial for deploying reliable fatigue detection systems in diverse real-world scenarios to enhance public safety.

This study takes image analysis as its focus and aims to bridge this gap. We systematically summarize publicly available datasets, feature extraction methods, and classification models by reviewing a range of experimental approaches. While summarizing the frontier technology of image-based fatigue detection, this study provides a new perspective for the subsequent research to better guarantee the road traffic safety.

2. Analysis of Core Methods

Driver fatigue detection has attracted extensive research attention. It leads to the emergence of various image-based detection methods, such as eye-tracking technology [6], personalized threshold discrimination [7], the use of libraries like OpenCV and Dlib [8][9], and the YOLO model [10].

2.1. Eye Fatigue Detection Characteristics and Algorithm Based on Eye Tracking Signal

The first method employed a self-developed desktop near-infrared eye tracker. Composed of an infrared camera and two sets of infrared fill lights, this device supports the acquisition of near-infrared images with a resolution of 3840×2160 pixels and a frame rate of 25 frames per second. During the signal collection process, participants are required to minimize head rotation. A nine-point calibration method is adopted to ensure signal accuracy, and stimuli are displayed on a 1920×1080 resolution electronic screen. The core of the feature extraction and fusion model lies in extracting fatigue-related features from eye-tracking signals and enhancing discriminative ability through fusion strategies. Feature extraction covers multiple dimensions, including pupil-related features and eye movement features. To combine the interpretability of manually extracted features and the comprehensiveness of automatically extracted features, this method proposes the Auto-Encoder Feature Union (AEFU) fusion model. The workflow of this model is as follows: first, an auto-encoder is used to automatically extract 128-dimensional latent features from eye-tracking signals; second, several types of manually calculated eye-tracking features are concatenated with these latent features; third, the combined features are input into a two-layer auto-encoder for further extraction of 64-dimensional fused features; finally, a fully connected layer is used to output the fatigue state classification result (binary classification: fatigued/non-fatigued). Manually extracted features provide indicators with clear physiological significance (e.g., pupil size reflecting ocular accommodation ability), while automatically extracted features capture complex signal patterns (e.g., subtle eye movement changes). Their fusion fully improves the comprehensiveness of feature representation. The AEFU algorithm optimizes feature fusion efficiency and offsets the cost disadvantage of using eye-tracking equipment. The classification algorithms in this method are divided into three categories: traditional machine learning algorithms based on manually extracted features; deep learning algorithms based on automatic feature extraction (including convolutional neural networks and auto-encoders); and fusion algorithms that combine manual features with auto-encoder-extracted automatic features.

After data augmentation, the experiment obtains 10,000 non-repetitive data points for both positive and negative samples, which are used for model training, validation, and testing. Accuracy is adopted as the algorithm performance evaluation metric, calculated by the formula: (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives). The final results showed that the average accuracy of the SVM model (a traditional machine learning algorithm) was 60.1%; the accuracy of the deep learning algorithm with a 5-layer encoder was 78.3%; and the accuracy of the fusion algorithm reached 87.9%. These results indicated that the combination of manually extracted eye-tracking features and deep learning-based automatically extracted latent features achieved the optimal performance.

2.2. Personalized Threshold and Multi-feature Fusion Driver Fatigue Detection Framework Based on Computer Vision

Compared with the first method, the second method achieves significant improvements in the dataset. It adopts the NTHU-DDD fatigue video dataset, which covers four driving scenarios: no glasses during the day, with glasses during the day, no glasses at night, and with glasses at night. This design greatly enhances data robustness. In the preprocessing stage, Histogram of Oriented Gradients (HOG) features are first extracted from images and input into a Support Vector Machine (SVM) to detect the driver's face. The facial position is recorded and a facial bounding box is marked. Then, an Ensemble of Regression Trees (ERT) is used to locate and annotate facial key points. After that, images are normalized through gamma correction, and the gradient of each pixel in the images is calculated to capture image edge and contour information. Figure 1 illustrates the feature extraction and classification workflow for fatigue detection. In this method, feature extraction is conducted from two perspectives. On one hand, appearance-based features are extracted by calculating the Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR). On the other hand, deep learning-based features are extracted using the MobileNetV3 small network to obtain deep-level features of images. These features are then used for driver fatigue state classification. Three fatigue discrimination parameters are employed in the experiment: F_{eye} (number of slow blinks per unit time), F_{mouth} (number of yawns per unit time), and PERCLOS (The percentage of frames in which the eyelid closes beyond the pupil within a unit of time, out of the total number of frames). A driver is determined to be in a fatigued driving state if $F_{eye} \geq 2$, $F_{mouth} \geq 2$, or PERCLOS≥ 0.4 within 30 seconds.

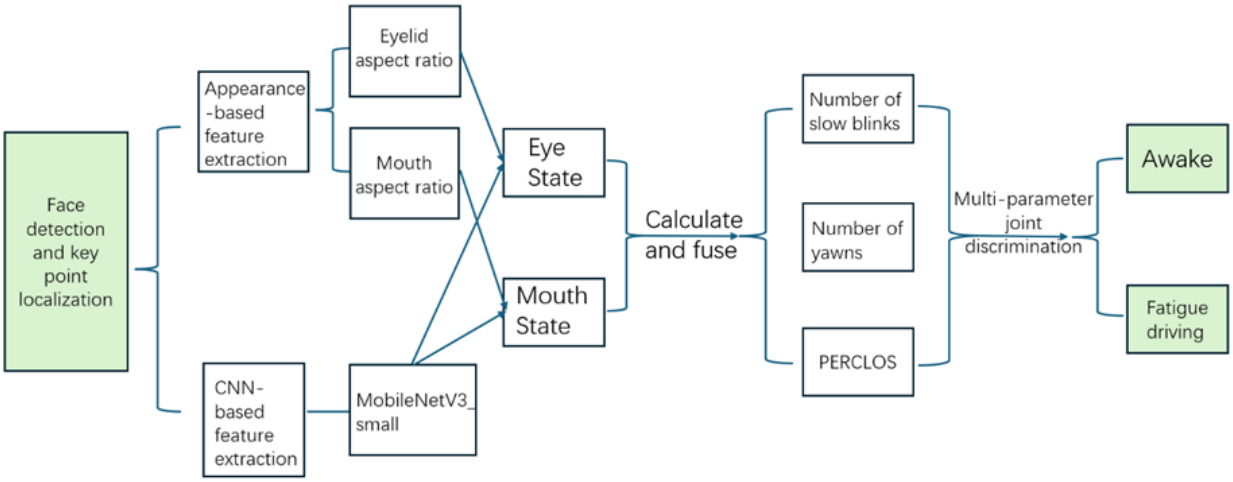

Fig. 1 The general process of fatigue detection

Based on the NTHU-DDD dataset, three sub-datasets are created: the slow blink dataset, the yawn dataset, and the driving dataset. The experiment uses miss rate, recall, precision, and F1 score as evaluation metrics. In slow blink detection, compared with the average threshold, the personalized threshold reduced the miss rate by 6.11%; compared with single-feature methods, the multi-feature fusion method reduced the miss rate by 43.26%. In yawn detection, the personalized threshold decreased the miss rate by 29.02% relative to the average threshold, and the fusion method lowered the miss rate by 87.13% compared with single-feature methods. In driver fatigue detection, the F1 score of the proposed framework reached 92.21%, outperforming various baseline methods such as 3D-DCNN, MT-DMF, and CNN-LSTM.

2.3. Vision-based driver fatigue detection: Stable facial keypoint detection using OpenCV and the Dlib library

The third method introduces innovations in facial landmark detection. It utilizes OpenCV and Dlib libraries for this task, focusing on locating key facial landmarks around the eyes and mouth. This lays the foundation for subsequent calculations of the Eye Aspect Ratio (EAR) and Mouth Aspect Ratio

(MAR). OpenCV leverages its built-in Haar cascade classifier to quickly detect facial regions and initially locate the approximate positions of key organs such as the eyes and mouth. The Dlib library, on the other hand, uses its pre-trained facial feature point detection model to accurately identify the coordinates of key facial landmarks. Even when the driver's head moves slightly or facial expressions change, it can stably track the positions of these landmarks, ensuring the continuity and accuracy of EAR and MAR calculations. However, compared with method 2, this method uses fixed EAR and MAR thresholds and does not account for inherent differences among drivers, such as variations in eye size. A Support Vector Machine (SVM) classifier is employed for classification in this method. The classifier is trained on data labeled with drowsy and non-drowsy states. It can distinguish between the driver's alert and drowsy states based on EAR, MAR, and their temporal variation patterns. Compared with other methods, this approach is relatively simple and incurs extremely low costs.

The approach uses benchmark datasets, including the NTHU Drowsy Driver Detection Dataset and YawDD. It also utilizes a custom dataset containing facial video recordings of participants in both drowsy and non-drowsy states. This custom dataset consideres multiple variations, such as lighting conditions, head positions, facial expressions, and whether glasses are worn.The final results showed that the overall accuracy of the method reached 92.3%, with a detection time of approximately 1.5 seconds, indicating good real-time performance. Under different lighting conditions, the system achieved an accuracy of 93.5% in daylight, 90.8% in dim light, and 88.2% in low light at night. Although its performance slightly decreased in low-light environments, it still maintained a high accuracy rate. Compared with traditional systems that only used threshold-based detection, the integration of the SVM classifier improved detection performance by approximately 8-10% under different lighting and facial conditions. It also achieved good detection results for subjects wearing glasses or with slight head movements.

2.4. Fatigue Detection Algorithm Based on Deep Learning: Yawning and Eye Analysis

Unlike the third method, the fourth method adopts a pure CNN (Convolutional Neural Network) architecture, which exhibits strong capability in autonomously learning facial patterns. This method is implemented with libraries such as OpenCV. OpenCV is used for real-time computer vision processing, camera access, and object detection using Haar feature classifiers. The Haar cascade classifier is applied to detect face and eye regions in images. The extracted regions are then resized and normalized to match the input size required by the CNN model. For eye detection, the Eye Aspect Ratio (EAR) technique is also used to assist in classifying eye states (open or closed). Figure 2 shows the flow chart of the detection algorithm in this experiment. First, the convolutional layers in the CNN extract features for eye states and yawning separately, capturing features such as shapes, edges, and motion patterns. Next, the features extracted from the eye state processing branch and the yawning detection branch are concatenated into a single feature vector. This allows the network to consider both fatigue indicators simultaneously. The fused feature vector then passes through fully connected layers to learn high-level representations. Finally, the output layer, activated by the SoftMax function, outputs the probability distribution for four categories: "yawning", "not yawning", "eyes closed", and "eyes open", thereby realizing the classification of fatigue states.

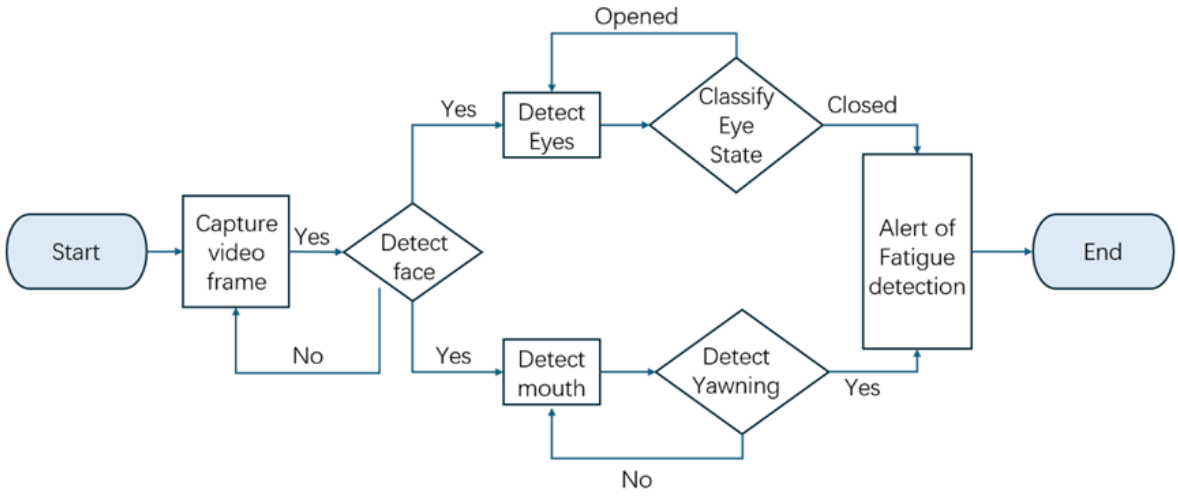

Fig. 2 Flowchart of fatigue detection algorithm

Two public datasets are used in the experiment: the YawDD dataset and the MRL Eye dataset, containing a total of 2,900 images. Among them, 70% of the dataset is used for training and 30% for testing. Different from method 2 and method 3, data augmentation techniques (such as scaling, flipping, and rotation) are applied to the training set to handle image variations and noise. This improves adaptability and optimized scene robustness. Precision, recall, F1 score, and support are used as metrics to evaluate the model's classification performance. Meanwhile, accuracy and loss value are employed to measure the overall performance of the model. The results showed that the overall test accuracy of the CNN model reached 96.54%, with a precision of 0.90 for the "yawning" category and 0.97 for the "eyes closed" category. For the VGG16 model, the overall test accuracy was 95.85%, with a precision of 0.80 for "yawning" and 0.92 for "eyes closed". In summary, the CNN model outperformed the VGG16 model in all evaluation metrics. This indicates that it has better performance in fatigue state classification tasks and can identify different fatigue states more accurately.

2.5. Driver Real-time Fatigue Detection Based on Face Analysis and Machine Learning Technology

The fifth method focuses on real-time performance and conducts real-time driver fatigue detection research based on face analysis and machine learning technologies. Figure 3 shows the architecture of this experimental method. Firstly, during data preprocessing, for the machine learning model, this experiment uses the Haar Cascade classifier to detect the facial area, and then applies the Histogram of Oriented Gradients (HOG) to extract key features such as texture and shape; for the processing of the deep learning object detection model, it marks the key areas such as eyes in the frame with bounding boxes and formats the annotations according to the model's requirements. Compared with the single EAR and MAR features of method 3, this method has a more comprehensive feature extraction dimension. During the training stage, the machine learning model extracts features from the detected facial area using HOG to capture texture and shape information, providing input for classification models such as KNN and SVM; YOLOv5 and YOLOv8 extract image features through multi-layer convolution operations and predict bounding box coordinates and confidence to detect fatigue-related features. As an advanced computer vision model, YOLOv5 and YOLOv8 can simultaneously complete target localization (generate bounding boxes for key facial areas) and state classification (determine "sleepy" or "not sleepy") through a single neural network forward propagation, without the need for staged processing, thus having efficient real-time performance and being suitable for real-time requirements in driver monitoring scenarios.

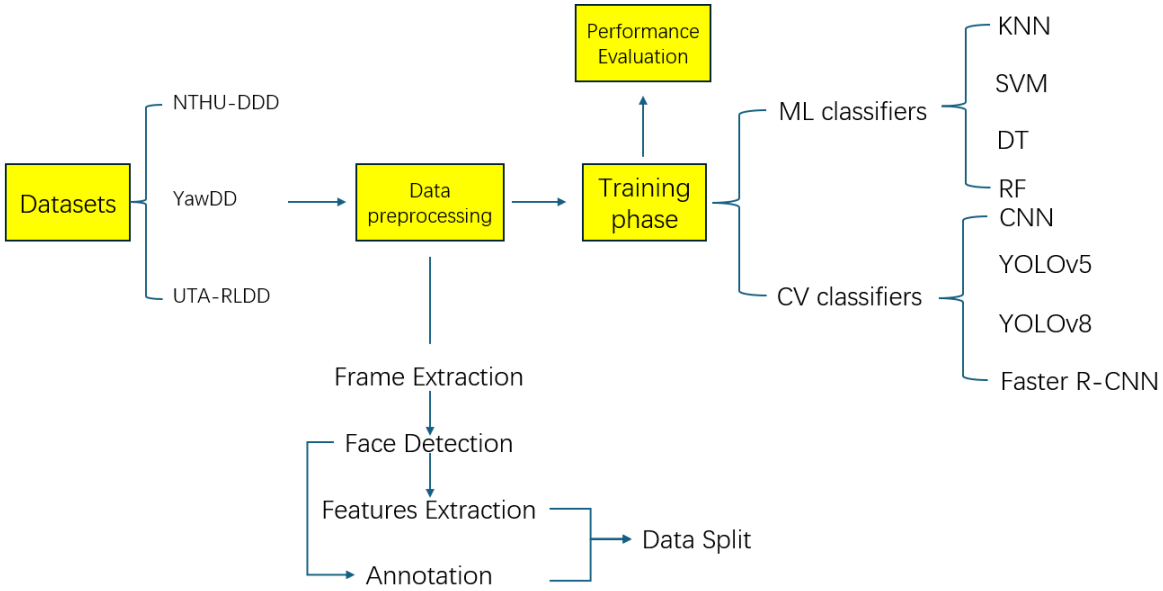

Fig. 3 The architecture of the fatigue detection recommendation method

This method analyzes multiple datasets. Among them, NTHUDDD contains 36 infrared videos covering normal driving, yawning, etc., recorded under different lighting conditions during day and night; YawDD contains 322 normal facial expression videos and 29 yawning videos, with diverse participant characteristics (including different genders, races, and whether wearing glasses); UTA-RLDD is 30-hour RGB videos recorded by 60 healthy participants, focusing on capturing fine micro-expressions related to fatigue. Among the machine learning models on UTA-RLDD, the KNN accuracy rate was 99.27%, the SVM accuracy rate was 97.45%, and the random forest accuracy rate was 99.58%; in the deep learning models on UTA-RLDD, the CNN accuracy rate was 100%, the accuracy rates of YOLOv5 and YOLOv8 were 99.9%, and the accuracy rate of Faster R-CNN was 63.4%. From the data in the comprehensive table, it can be seen that in the field of machine learning, KNN stands out; while in computer vision algorithms, YOLOv5 performs particularly well.

3. Methods Comparison

This section systematically compares the five reviewed methods from three core dimensions: preprocessing and feature extraction techniques, unique innovations, and optimal application scenarios. The relevant comparison summary is shown in Table1.

Table 1. Methods Comparison

Core methods of literature	Algorithm model	Data set	Feature extraction	Method characteristics	Summary of innovation points
Eye Fatigue Detection Characteristics and Algorithm Based on Eye Tracking Signal	SVM, etc. (Traditional machine learning) +AEFU	19 subjects (electronic screen use) + data enhancement, near infrared images	Pupil characteristics and eye movement characteristics; Subtle characteristics of deep learning	Near infrared camera, applicable day and night, and can identify occlusion information	AEFU Algorithm: Fusion of Manual Features and Automatic Coding Features Correlate pupil dynamics

					with fatigue
Computer vision-based driven fatigue detection framework with personalized threshold and multi-feature fusion	SVM etc. (Machine learning) +MobileNetV3	NTHU-DDD, daytime and nighttime driving, glasses or not, RGB images	Whether to close eyes and yawn (EAR, MAR); Characteristics of deep learning	Personalization threshold; Classification analysis was carried out for different illumination and glasses	Personalized threshold: Adjust EAR/MAR threshold for different drivers to reduce misjudgment
Vision-based driver fatigue detection: Combining eye aspect ratio, mouth aspect ratio with machine learning techniques	CNN (Extract Facial Species Sign) +SVM(Classification)	NTHU-DDD, daytime and nighttime driving, glasses or not, RGB images	Whether close eyes, yawn (EAR. MAR)	Use OpenCV/Dlib to track facial keys with low latency	Using OpenCV/Dlib Tracking Facial Key Points, Lightweight Deployment
Fatigue Detection Algorithm Based on Machine Learning	CNN (core model), VGG16, OpenCV (image, video processing, etc.), Haar (detecting face, eye area)	YawDD, MRL Eye Dataset, RGB images	Whether close eyes, yawn (Deep learning, automatic extraction)	With pure CNN architecture, the ability to learn facial patterns independently is strong	Data enhancement
Driver Real-time Fatigue Detection Based on Face Analysis and Machine Learning Technology	KNN, SVM, YOLOv5, YOLOv8 (extracting image features by multi-layer convolution), etc.	NTHU-DDD, daytime and nighttime driving, glasses or not, YawDD,UTA-RLDD, RGB images	Facial texture and shape features; Characteristics of deep learning	Applications with high real-time requirements	High accuracy target detection (YOLO)

Table 1 provides a comparative overview of five image-based fatigue detection methods, focusing on their key features, innovations, and main application scenarios. In terms of preprocessing and

feature extraction, methods 3 and 4 use OpenCV for initial eye and mouth localization, followed by Dlib for accurate facial landmark detection. In contrast, Methods 2 and 5 extract HOG (Histogram of Oriented Gradients) features from the images. Regarding the types of features used, Methods 1, 2, and 5 involve both handcrafted features (traditional machine learning) and automatically learned features (deep learning). Method 1 combines both types for analysis. Method 3 manually extracts EAR (Eye Aspect Ratio) and MAR (Mouth Aspect Ratio) for analysis. Method 4 uses deep learning to automatically extract subtle fatigue features. All the above methods have unique innovations in feature extraction or classification models. The innovation of method 1 lies in the use of near-infrared cameras, which can well alleviate problems such as unclear facial recognition at night and facial occlusion by eyes. The AEFU algorithm it uses integrates manually and automatically extracted features, enhancing the experimental robustness. The core contribution of method 2 is the proposal of personalized thresholds. It adjusts the eye aspect ratio and mouth aspect ratio thresholds for different drivers, which can reduce misjudgments. The innovation of method 3 is the use of OpenCV/Dlib to track facial key points, resulting in low latency. The innovation of method 4 is the adoption of a pure CNN architecture. It uses deep learning to automatically learn and extract facial features, with strong ability to independently learn facial patterns. Moreover, the data generalization is greatly improved after data augmentation. The advantage of method 5 is the use of the YOLO model, which plays an important role in applications with high real-time requirements.

This analysis indicates that there is no "optimal method" applicable to all scenarios, and the selection of methods must be closely aligned with application requirements: Methods 3 and 5 are suitable for scenarios with high real-time requirements. Methods 1 and 2 are appropriate for application scenarios that allow individual calibration and have high accuracy requirements. Method 4, however, achieves the best performance when trained in a controlled environment with sufficient data.

4. Deficiency and Prospect

There are some deficiencies in the above experiments. The first is the problem of small sample size: In method 1 and method 4, the training samples are small, and lack data on different lighting conditions and real road environment; The YOLO series in method 5 showed a slight performance reduction in YawDD, indicating the risk of over-fitting a particular data set. In this way, we can combine the literature data for training, use StyleGAN algorithm to generate multi-skin color and light face, and combine vehicle enterprises for actual road detection. Second, environmental sensitivity issues, method 2 and method 3 exhibit reduced accuracy at night or when covering the face, and can be changed to thermal imaging cameras, 3D face modeling, or head pose estimation; For the problem of high cost of using high precision eye movement instrument in method 1, a monocular eye movement tracking algorithm based on smart phone camera can be developed.

5. Summary

This study focuses on image-based fatigue detection and systematically reviews several typical technical methods, analyzing their core characteristics in data acquisition, feature extraction, model construction, and performance. Method 3 emphasizes the optimization of machine learning techniques, offering a lightweight implementation solution. It has been demonstrated that under stable lighting and unobstructed conditions, traditional machine learning can achieve over 90% detection accuracy. Method 4 adopts a deep learning approach and proposes a “dual-branch CNN feature fusion” architecture to extract and integrate eye and mouth features separately. This addresses the issue of high false detection rates in single-modality models and improves performance in complex environments. Methods 1, 2, and 5 incorporate both machine learning and deep learning strategies. Method 1 introduces a hybrid model that combines machine learning and deep learning, enhancing experimental robustness. Method 2 improves detection accuracy through weighted decision fusion.

Method 5 implements modifications to the CNN architecture, providing effective solutions to increase detection precision.

Our study addresses the issues of technical fragmentation and unclear scene adaptability by systematically reviewing and contrasting five representative technical paths. The primary contribution of this paper is that it fills the significant gap between numerous isolated studies and the lack of comparative analysis, and provides a basis for future technology selection. Furthermore, the study offers practical insights for overcoming existing research bottlenecks, thereby aiding researchers and industry practitioners in driving relevant technological innovation.

At present, the core bottleneck of image-based fatigue detection technology lies in environmental adaptability and data generalization, etc. Future efforts should focus on strategies such as multi-modal sensor fusion, dataset expansion, and hardware-algorithm co-design to break through these bottlenecks. These advancements are expected to facilitate the widespread deployment of robust fatigue detection systems in intelligent vehicles and beyond.

References

[1] Cao S, Feng P, Kang W, et al. Optimized driver fatigue detection method using multimodal neural networks[J]. Scientific Reports, 2025, 15(1): 12240.

[2] Shang L, Si H, Wang H, et al. Research on fatigue detection of flight trainees based on face EMF feature model combination with PSO-CNN algorithm[J]. Scientific Reports, 2024, 14(1): 20641.

[3] Huang Y, Liu C, Chang F, et al. Self-supervised multi-granularity graph attention network for vision-based driver fatigue detection[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, 8(4): 3067-3080.

[4] Ji Z, Xie X, Jiang E, et al. Integrating DRN-RF with computer vision for detection of control room operator's mental fatigue[J]. PloS one, 2025, 20(4): e0320780.

[5] Ma B, Fu Z, Rakheja S, et al. Distracted driving behavior and driver's emotion detection based on improved YOLOv8 with attention mechanism[J]. IEEE Access, 2024, 12: 37983-37994.

[6] Sun W, Wang Y, Hu B, et al.Exploration of eye fatigue detection features and algorithm based on eye-tracking signal[J]. Electronics, 2024, 13(10): 1798.

[7] Li X, Lin H, Du J, et al.Computer vision-based driver fatigue detection framework with personalization threshold and multi-feature fusion[J]. Signal, Image and Video Processing, 2024, 18(1): 505-514.

[8] NARAYANA G V. Vision-Based Driver Fatigue Detection Using Eye and Mouth Aspect Ratios with Machine Learning[J]. INTERNATIONAL JOURNAL, 2025, 9(3).

[9] Makhmudov F, Turimov D, Xamidov M, et al.Real-time fatigue detection algorithms using machine learning for yawning and eye state[J]. Sensors, 2024, 24(23): 7810.

[10] Essahraui S, Lamaakal I, El Hamly I, et al. Real-time driver drowsiness detection using facial analysis and machine learning techniques[J]. Sensors, 2025, 25(3): 812.