Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 9
AIDML 2025

# Research Progress and Challenges of Computer Vision Object Detection Technology

Chun Ho Ma *

School of computer science and technology, Huazhong University of Science and Technology, Wuhan, 430074, China

* Corresponding Author Email: u202490042@hust.edu.cn

**Abstract.** With the continuous increase in the number of motor vehicles in the world, traffic accidents caused by drowsy driving have become a major social safety problem. This paper systematically analyzes the evolution path of the current fatigue driving detection technology, focusing on two major methodologies based on the fusion of single features and multiple features. Although the method based on a single feature has the advantages of low computing cost and easy deployment, it is limited by bottlenecks such as environmental interference, device dependence and high false alarm rate. Physiological signal detection can achieve good accuracy in a controlled environment, but it is necessary to solve the problems of individual differences and noise interference. Visual feature detection has significant advantages in non-contact, but the accuracy will be reduced in low-light and occlusion scenes. Although vehicle behavior detection is easy to integrate into the on-board system, the false positive rate is high. The multi-feature fusion method effectively integrates multi-modal data through feature-level, decision-level, and hybrid fusion strategies, which significantly improves the detection accuracy and robustness. In the future, it is necessary to combine lightweight models, edge computing and human-computer collaborative intervention technologies to promote the development of fatigue detection systems in the direction of intelligence and adaptation, so as to provide theoretical support for the realization of all-weather accurate monitoring and traffic safety assurance.

**Keywords:** Drowsy driving; feature-level fusion; Robustness; decision-level integration; Multimodal data.

## 1. Introduction

As the number of motor vehicles around the world continues to rise, intelligent traffic management is facing severe challenges. The rapid development of motorization has increased the number of serious traffic-related casualties. Drowsy driving has become a major cause of traffic accidents. The National Highway Traffic Safety Administration (NHTSA) estimates that 56,000 drowsiness accidents occur each year [1]. In addition, in Australia, fatigue accidents account for 15% of fatal accidents in heavy vehicles and 10% of all injuries, resulting in losses of more than US$250 million [2]. It highlights the urgency of developing intelligent driving monitoring technology, and object detection, as its core technology, has become the focus of research in the field of computer vision.

Significant breakthroughs have been made in object detection technology driven by deep learning in recent years, and CNN models represented by Simple linear iterative clustering (SLIC) and BiLSTM bidirectional network models [3] not only surpass traditional manual features (such as HOG, SIFT) method, and overcomes the shortcomings of the long time consuming of commonly used complex CNNs [4], thus greatly reducing the time-consuming segmentation process.

At present, there are two major paths in the evolution of technology, namely the single-feature method and the multi-feature fusion method (through feature cascade, attention mechanism and other strategies), both of which significantly improve the efficiency and accuracy of fatigue driving detection. The results show that optimizing the feature expression and fusion mechanism is an important means to break through the existing technical barriers, and effective multi-modal feature collaboration can not only enhance the adaptability of the model to extreme conditions such as low light and occlusion, but also meet the needs of edge computing through lightweight design, which is of great value for all-weather driving monitoring.

In this study, the mapping relationship between the fatigue driving detection method and its detection performance will be deconstructed systematically, and the theoretical framework of fusion with stronger interpretability will be established. The results can not only provide key technical support for autonomous driving, industrial testing and other fields, but also have significant social benefits for reducing the casualty rate of traffic accidents and building a smart transportation system.

## 2. Fatigue Driving Detection Method Based on a Single Feature

### 2.1. Physiological Signal Detection

The technique mentioned in this paper [5] is a single combination of EEG, ECG, and skin conductance (EDA) features, i.e., attention levels are monitored by the theta/δ energy ratio of EEG, autonomic states are assessed by HRV time-domain indicators of ECG, and changes in skin conductance are detected by EDA. The lightweight dual-channel EEG sensor and adaptive filtering algorithm are used to achieve 95% fatigue state recognition accuracy in the cockpit environment and provide real-time warning support for the vehicle safety system. In this paper [6], the limitations of physiological signal detection were studied, and a detailed summary of the advantages and disadvantages of physiological signal detection was finally obtained in Table 1

**Table 1.** Advantages and disadvantages of physiological signal detection.

| Advantages [5]. | Limitations [6]. |
|---|---|
| High accuracy: EEG can directly detect brain fatigue states, and the detection accuracy can reach 90% in theory. | High device dependence: EEG and ECG require electrode sensors, which are inconvenient for drivers to wear on a daily basis. |
| Non-subjective: Physiological signals do not rely on the driver's self-judgment. | Individual data differences: There are significant differences in physiological signals in different populations, which need to be individually calibrated. |
| Good performance in the experimental environment: EEG and ECG performed well in a controlled environment. | Noise interference: Factors such as motion and poor contact can easily lead to signal noise and affect the stability of detection. |

### 2.2. Visual Feature Detection

The visual feature detection method uses the camera to analyze the driver's facial dynamic features to achieve fatigue monitoring, which mainly includes three indicators [7]: the proportion of eye closure time (PERCLOS>80% is determined to be fatigue), the frequency of yawning is significantly increased (reflecting physiological sleepiness), and the head posture is abnormal (leaning forward or deviating from direction). Through non-contact monitoring, this method captures in real time typical fatigue characteristics such as drooping eyelids, frequent yawning, and drooping heads. Table 2 summarizes the advantages and disadvantages of visual feature detection in a paper by University Pendidikan Sultan Idrts [7] and a paper by University Mohamed Khider [8].

**Table 2.** Advantages and disadvantages of visual feature detection.

| Advantages [7]. | Limitations [8]. |
|---|---|
| Non-contact detection: No need to wear a sensor, just a camera on board. | Strong environmental dependence: light changes and occlusion (such as sunglasses, masks) will affect the detection accuracy. |
| Easy to integrate: It can be combined with intelligent cockpit and driver assistance systems (ADAS) to improve driving safety. | Large individual differences: There are differences in facial features among different populations, and the generalization ability of the model is weak. |
| Low computational cost: Methods based on traditional computer vision methods, such as PERCLOS, have low computational requirements. | High rate of false alarms: Certain driving behaviors, such as looking down at the dashboard normally, can be misjudged as fatigue. |

In recent years, the application of deep learning in visual fatigue detection has gradually increased, for example, [7], YOLO+ResNet combines face detection and eye state analysis to improve the detection accuracy in low-light environments. The infrared camera + convolutional neural network (CNN) is used to reduce the influence of light changes and improve the stability of the system. However, even with advanced deep learning methods, visual inspection still struggles to completely overcome environmental interference.

### 2.3. Vehicle Behavior Detection

Dong Wang uses driving modes such as steering wheel angle, lane departure, and pedal operation to predict driver fatigue and perform vehicle behavior detection [9]. The results show that the decrease of steering wheel control ability leads to the increase of angle fluctuation, the frequency of lane departure increases significantly (detection success rate of 81.82%), and the brake and accelerator operation tend to be sluggish or unstable. This type of method extracts features through multi-sensor fusion (such as CAN bus, visual perception), and combines machine learning algorithms to achieve non-intrusive real-time monitoring. Table 3 summarizes the advantages and disadvantages of vehicle behavior detection.

**Table 3.** Advantages and Disadvantages of vehicle behavior detection:

| advantage | limitations |
|---|---|
| No additional sensors required: Built-in sensors (e.g. lane departure warning systems) can be used directly. | High false alarm rate: not only are there large individual differences in driving styles, but also different control modes are different for different drivers. |
| Computationally low: Rule-based approaches run efficiently on embedded systems. | Affected by the external environment: road conditions (curves, bumps) can lead to misjudgments. |
| Long-term monitoring effect is good: it is suitable for long-term monitoring of the driver's state. | Difficulty identifying the cause of fatigue: It is impossible to distinguish what causes fatigue. |

## 3. Research on fatigue driving detection method based on multi-feature fusion

### 3.1. Feature-level fusion

A paper from South China University [10] proposed a fatigue driving detection method based on feature-level fusion, which improves the detection accuracy by integrating visual features (PERCLOS) and physiological signals (heart rate). The researchers developed a fatigue detection framework using feature-fusion of PERCLOS and heart rate from RGB videos. PERCLOS was calculated via iris segmentation using K-means clustering to quantify eyelid closure, while heart rate was extracted by applying 2SR algorithm to facial skin pixels for rPPG generation, followed by Fourier transform. Two parallel 1D CNNs modeled temporal patterns: one tracking eyelid dynamics, the other analyzing heart rate variability. A boosting strategy adaptively adjusted model weights for decision fusion, enabling complementary integration of behavioral and physiological cues. Experiments confirmed enhanced accuracy over single-modality methods.

In recent years, multimodal learning methods based on Transformer have become a research hotspot. For example, a research paper [11] proposed that Multimodal Transformer (MMT) is susceptible to light/occlusion interference due to traditional visual inspection (e.g., PERCLOS, 97.06% accuracy) and EEG detection (95-97%) needs to invade the device. Therefore, MMT provides an innovative solution for fatigue detection by combining EEG and vision. MMT fuses EEG theta wave and visual blink frequency with cross-modal attention, achieving an accuracy rate of 98.5% (+6.2% compared with single-modality), enhancing illumination robustness, breaking through the limitations of manual feature fusion, and providing a real-time early warning scheme

### 3.2. Decision-level convergence

Decision-level fusion calculates the fatigue state of each modality separately, and then combines different detection results to finally determine the fatigue level of the driver. In a research paper [11] on decision-level fusion, a driving state monitoring system was proposed, which was mainly reflected in the fusion of fatigue and emotion detection, and the real-time detection of 68 points of human face by Dlib was used to construct a dual-threshold model (PERCLOS+ dynamic threshold) to evaluate fatigue. The RM-Xception network was used to identify seven types of emotions and quantify risks. Finally, a weighted scoring model (0.5×fatigue + 0.5×emotion) was established based on temporal fusion, and a four-level early warning was triggered. Experiments show that fatigue detection is better than traditional methods, and the accuracy rate of emotion recognition is 73.32% (Fer2013), which promotes intelligent driving to enter the stage of multimodal fusion. At present, there are many decision-level fusion methods in practical applications, such as Mercedes-Benz's "Attention Assist" system, which combines steering wheel behavior and visual information, and uses decision-level fusion to determine whether the driver is fatigued.

### 3.3. Mixed methods

Hybrid fusion combines feature-level fusion and decision-level fusion to make full use of multimodal data while reducing computational costs. Driver fatigue detection is an important research direction of intelligent transportation systems, and the traditional methods mainly rely on physiological signals, visual features and vehicle behaviors, but single feature detection is very susceptible to environmental influences and has a high false alarm rate. In recent years, hybrid methods (feature-level fusion + decision-level fusion) have become an important strategy to improve detection accuracy and robustness. In a research paper on mixed methods [12], feature-level fusion unifies different modal features such as EEG, PERCLOS, and vehicle behavior data in the data input stage, and uses deep learning models (such as CNNs and LSTMs) to extract key information to improve detection accuracy. Decision-level fusion calculates the fatigue state of different subsystems (such as RGB-D camera detection and driving behavior analysis), and then makes the final judgment through voting mechanism and Bayesian reasoning. The combination of RGB-D camera and deep learning technology can further enhance the stability of fatigue detection and reduce interference such as lighting and occlusion. Future research can optimize the lightweight fusion model, use self-supervised learning to improve the generalization ability, and combine the human-machine collaboration system to achieve more intelligent fatigue detection and intervention, so as to provide more reliable technical support for road safety. In recent years, the multimodal fusion method based on LSTM + Transformer has shown high accuracy in hybrid fusion.

## 4. Current limitations and future prospects: the development direction of fatigue driving detection technology

### 4.1. Analysis of existing limitations

#### 4.1.1. Data bias

At present, the driving fatigue detection system faces the bottleneck of training data bias [9], first of all, poor environmental adaptability (+37% distortion rate of eyelid features caused by low light/strong backlight), and then insufficient biometric compatibility (data biased towards specific races/youths, dark skin color pupil error of 15-20%, ignoring drooping eyelids in the elderly), and the scene coverage is low (90% of the data comes from the laboratory/closed roads, lack of data on complex working conditions such as heavy rain/plateau, and the model response delay is 2.3 seconds). As a result, the actual false alarm rate of the drive test reached 12.7%, and the cross-regional accuracy fluctuated by more than 18%.

#### 4.1.2. Real-time bottleneck - computing resources and latency

Fatigue detection systems need to process large-scale sensor data (such as EEG, cameras, vehicle behavior data) in real time, but there are computational bottlenecks in current methods, and real-time performance is one of the key factors affecting the detection effect and user experience in fatigue detection systems. Existing studies [12] suggest that real-time monitoring systems may face limited computing resources and latency when performing complex data processing. For example, deep learning-based visual inspection methods require high-performance GPUs for computing, while most in-vehicle computing units have limited computing power, making it difficult to support real-time inference at high frame rates. In addition, the acquisition and processing of physiological signals (such as EEG and ECG) require high-precision synchronization, and multimodal data fusion further exacerbates the data delay problem. This limits the application of fatigue detection systems in highly dynamic driving scenarios, such as highway driving.

#### 4.1.3. User Privacy and Ethics Disputes

Autonomous driving faces fatigue and privacy ethics conflicts, with studies showing that [13] in-vehicle surveillance (facial/physiological sensing) can reduce miscalculations but lead to a privacy crisis (biometric data leakage leading to commercial abuse), and that professional driver surveillance throughout the day may constitute a privacy violation, leading to a realistic paradox that ensuring safety requires more in-depth behavioral monitoring, while ethical principles call for minimizing data collection. There is an urgent need to establish industry data standards and multi-party collaboration to achieve technical ethics harmonization.

#### 4.1.4. Future research directions

The paper study [13] shows that the current intra-individual accuracy rate of EEG+ eye movement cross-modal detection is over 99%, but the cross-individual generalization is only 75%, so in the future breakthrough direction, first, the establishment of dynamic adaptive models (transfer learning + continuous learning to achieve personalized evaluation), and the second is the development of embedded edge architecture (micro EEG/visual sensor + lightweight network up to millisecond response), and third, establish a hierarchical intervention mechanism system (haptic/environmental warning→ assisted driving takeover →emergency braking), and it is necessary to optimize the intervention timing algorithm to balance safety and experience.

## 5. Conclusion

This paper systematically deconstructs the core methods and performance bottlenecks of fatigue driving detection technology, and reveals the key differences and complementarities between single feature and multi-feature fusion methods. The method based on a single feature is difficult to meet the needs of complex driving scenarios due to its limitations (such as physiological signal-dependent equipment, visual detection susceptible to environmental interference, and high false alarm rate of vehicle behavior). The multi-feature fusion method significantly improves the detection accuracy through cross-modal feature interaction (such as the boosting strategy in feature-level fusion and the time series scoring model in decision-level fusion), but it is limited by data synchronization, computing power requirements, and privacy protection. The current technology faces challenges such as insufficient cross-scenario generalization ability due to training data bias, delay in real-time processing of multimodal data, and ethical conflicts between biometric collection and privacy protection.

The future direction of development should focus on the following aspects. In terms of algorithm optimization, a lightweight multimodal fusion model (such as TinyML architecture) can be developed, combined with transfer learning and self-supervision technology to solve the problem of data bias and individual differences, and at the same time, the hardware collaboration method can be used to build an edge-cloud collaborative computing framework, and miniaturized non-contact sensors (such as infrared cameras, dry electrode EEG) can be integrated to achieve low-power real-time detectionIn terms of human-computer interaction, we should design hierarchical intervention mechanisms (such as haptic feedback→ environmental regulation→ autonomous driving takeover), balance safety response and driving experience, and finally try to establish multimodal data anonymization standards and industry norms in ethical governance to ensure that the application of technology complies with the principle of privacy protection. Through multidisciplinary interdisciplinary and technological innovation, fatigue driving detection technology will gradually break through the existing barriers, thereby providing key support for the construction of an intelligent traffic safety system, and ultimately realizing the vision of "zero accident" driving

## References


[1] Driver I. Driver Fatigue and Road Accidents. The Royal Society for the Prevention of Accidents, Birmingham, 2011.Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.

[2] Australian Transport Safety Bureau. Working Safer, Not Less. Information Bulletin,2006

[3] Wang J L, et al. Research on Fatigue Driving Detection Algorithm Based on Deep Learning. Chongqing Normal University, 2021. DOI: 10.27672/d.cnki.gcsfc.2021.000657.

[4] Liu M Z, Xu X, Hu J, et al. Real-time detection of driver fatigue based on CNN-LSTM. IET Image Processing, 2022, 16(2): 576-595.

[5] Wang L, Song F, Zhou T H, et al. EEG and ECG-based multi-sensor fusion computing for real-time fatigue driving recognition based on feedback mechanism. Sensors, 2023, 23(20): 8386.

[6] Pan H, Hu Y, Wang Y, et al. Fatigue Detection in Air Traffic Controllers: A Comprehensive Review. IEEE Access, 2024.

[7] Yan X, bin Abas A. Advancements and Perspectives in Fatigue Driving Detection: A Comprehensive Review. IECE Transactions on Intelligent Unmanned Systems, 2024, 1(1): 4-15.

[8] Sen A, Das A, Karmakar P, et al. mmAssist: Passive Monitoring of Driver's Attentiveness Using mmWave Sensors. 15th International Conference on Communication Systems & Networks (COMSNETS), IEEE, 2023: 545-553.

[9] Wang D, Wu P. A Review of State-of-Art Methods for Passive Driver Fatigue Detection: Progress and Prospect. 3rd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS), IEEE, 2024: 787-794.

[10] Yao Y, Wang Y, Wang C, et al. Facial Feature Extraction Algorithm Based on Improved YOLOv7-Tiny. IEEE Access, 2025.

[11] Hang Y, Yang M, Cui J, et al. Driver Emotion and Fatigue State Detection Based on Time Series Fusion. Electronics, 2022, 12(1): 26.

[12] Ahmed S G, Verbert K, Siedahmed N, et al. AI Innovations in rPPG Systems for Driver Monitoring: Comprehensive Systematic Review and Future Prospects. IEEE Access, 2025.

[13] Sood J, Deb S, Kan C, et al. Driver Fatigue in Prolonged Automated Driving: Research Gaps and Future Directions. IISE Annual Conference Proceedings, Institute of Industrial and Systems Engineers (IISE), 2023: 1-6.