A Review of Fatigue Driving Detection Methods Based on Drivers' Facial Features

Sirui Min*
Chang'an Dublin International College of Transportation at Chang'an University, Chang'an University, Xi'an, Shaanxi, China
**Corresponding Author*

Abstracts: Research in the field of fatigue driving detection technology is undergoing a paradigm shift towards multimodal sensing and intelligent decision-making. At this stage, the mainstream methods include three categories: physiological signal monitoring, driving behavior analysis, and facial feature recognition. Among these, the facial feature-based detecting method has become the focus of research due to its non-contact and low-cost advantages. Despite the notable advancements witnessed in this domain, its potential remains constrained by several limitations. These include a paucity of dataset diversity, constrained model generalization capabilities, and an inability to effectively adapt to complex driving environments. In order to enhance the robustness and practicality of the detection system, future research should be initiated with multimodal information fusion, personalized model training, and edge computing deployment. In this paper, the methods and operations of traditional image processing and deep learning techniques are systematically examined. The existing research results are compared and analyzed, and the aim is to provide theoretical support and practical guidance for the optimization of fatigue driving detection technology.

Keywords: Fatigue Driving Detection; Deep Learning; Facial Features

1. Introduction

Fatigue would cause the driver's head drooping, blurred vision, and other such symptoms. These issues pose a serious threat to the safety of the individual driver, as well as to other road users. Indeed, fatigue is one of the main reasons for the frequent occurrence of traffic accidents and it has been reported that the probability of traffic accidents is higher when the driver is fatigued. The National Highway Traffic Safety Administration (NHTSA) reports that approximately 100,000 police-recorded crashes in the U.S. are related to drowsy driving each year, resulting in more than 1,500 deaths; however, a study by the AAA Foundation estimates that the actual number may be even higher, with approximately 328,000 crashes related to fatigued driving each year, resulting in approximately 6,400 deaths. Enhancing road traffic safety and reducing traffic accidents need to be solved urgently, and driver behavior detection technology has gradually become an important topic in traffic safety research. The monitoring of driver behavior has been shown to facilitate the early detection of potential fatigue, distracted driving, and other dangerous behaviors. This, in turn, enables the implementation of intervention measures with a view to protecting the safety of drivers and other road users. This paper first summarizes the basic approaches and tools in fatigue driving detection and then draws current challenges and future trends through comparison and research. The research can be applied to intelligent transportation systems and self-driving cars, helping to improve the safety of the systems effectively. The technology has the potential to assist both society and the state in mitigating economic losses and casualties resulting from traffic accidents, improving overall traffic management.

2. Literature Review

In recent years, a significant number of scholars and research teams have conducted research in the field of driver feature detection. The existing research in this area primarily focuses on driver visual features, such as the face and head. A plethora of pertinent studies have demonstrated that the eye movement, mouth activity, and head posture of drivers in a state of fatigue will undergo regular changes. The objective of this study is to determine the fatigue state through real-time acquisition of driver's facial images by camera and extraction of visual features such as eye, mouth, head posture, etc. To this end, the traditional image processing method is employed, whereby indicators such as PERCLOS (Percentage of Time with Eyes Closed) and blinking frequency/velocity are extracted by manual design. Thereafter, the traditional classifiers (e.g., SVM, Decision Tree, KNN, etc.) are adopted to determine whether the fatigue is fatigue or not. With the progression of technology, an increasing number of researchers are turning to deep learning methodologies for the automated identification, categorization, and learning of features of diverse signals and images, leveraging the capabilities of deep neural networks. The employment of such algorithms enables the automatic extraction of complex features from high-dimensional data, thereby significantly enhancing the accuracy of detection.

3. Technical Approaches to Fatigue Driving Detection Based on Facial Features

In the context of research into driver fatigue detection, the acquisition and analysis of facial features are of critical importance. As research into fatigue driving continues to deepen, the development of technology for detecting fatigue in drivers based on facial features has evolved from conventional image processing methods to deep learning methods. Each technology has its own advantages and areas for improvement. The following will analyze and compare two typical methods, conventional image processing methods and deep learning methods.

Despite the differences in the fatigue state recognition mechanism between traditional methods and deep learning methods, they demonstrate a high degree of consistency and complementarity in the pre-processes, such as data acquisition, face detection, key point extraction, and image pre-processing. Firstly, it is important to note that data acquisition constitutes the fundamental stage of the detection process. Researchers typically utilize in-vehicle cameras to collect driver facial images or video sequences, thereby constructing fatigue datasets comprising behaviors such as blinking, yawning, and head nodding. A number of these studies utilize publicly available datasets, including NTHU-DDD and YawDD, and employ facial key regions and fatigue state markers for image labeling. However, datasets like NTHU- DDD and YawDD include participants from a narrow range of ages, genders, and ethnic backgrounds and were collected under controlled lab conditions, so they may not capture the lighting and environmental variations of real-world driving, which can limit an algorithm’s generalizability in the field.

In data preprocessing, face detection and alignment methods such as detectors of Haar features, Dlib face recognizer, and MTCNN are widely used to obtain the locations of key parts such as eyes, mouth, and nose, which provide the basis for subsequent analysis. Among them, Liu Xing et al. introduced an improvement strategy based on the MTCNN structure to enhance the detection accuracy and speed, combined with R-Net, and improved the MTCNN network structure by optimizing the sub-network, reducing the number of parameters by using average pooling, and replacing the fully-connected layer with mean-value pooling, which significantly improved the model efficiency.

In the preprocessing stage, researchers often eliminate environmental interference through operations such as grayscaling and histogram equalization and use integral projection to locate the five sensory regions. For example, after acquiring the image with a CCD camera, Guo Huili's team enhanced the contrast of the image and strengthened the details of the eyes through grayscale transformation and binarization; in response to the misjudgment problem caused by the wearing of glasses, Zou Xintong and others innovatively propose methods such as horizontal integral projection method and histogram statistical features to replace the traditional binarization of area ratios, which improves the detection accuracy of 18.7% in comparison with that in the eyeglasses scenario.

In the study of fatigue driving detection based on facial features, traditional image analysis methods, and deep learning methods have some commonality in the overall process, but there are significant differences in the key technical aspects.

3.1 Conventional Image Processing Methods

Most studies in this field rely on manually designed features, with a primary focus on quantifying eye, mouth, and head behaviors. Eye-related parameters such as PERCLOS (percentage of eyelid closure over time) and blink frequency (BF) are commonly used, while mouth aspect ratio (MAR) and yawn duration have also shown strong correlations with changes in the head's Euler angles. These features are typically extracted through quantitative analysis based on facial key point coordinates or the results of region segmentation using computer vision techniques.

Many approaches involve expert-designed thresholds to determine fatigue levels. Notably, Ding Fusheng's team made significant progress by addressing the limitations of relying on single indicators. They established a mapping between multidimensional behavioral features and the Karolinska Sleepiness Scale (KSS), enabling the integration of subjective and objective measures. Their model achieved a fatigue recognition rate of 89.3% in highway driving scenarios.

Guo Huili's team combined the Adaboost algorithm with grayscale integral projection to precisely locate eye and mouth regions. They then applied the Dempster-Shafer (D-S) evidence theory to fuse eye, mouth, and head features and developed a hierarchical warning mechanism. To address the challenge of dynamic occlusion, Li Yongda's team proposed a hybrid AdaBoost-CamShift algorithm, improving the accuracy of eye detection from 72.1% to 86.4% by enabling rapid face localization and continuous tracking.

In many existing systems, support vector machines (SVMs) are employed along with preset thresholds (e.g., PERCLOS > 25%) to classify fatigue states. SVMs are advantageous for handling small sample sizes and high-dimensional data. Additionally, Kalman filtering is often integrated to handle sudden lighting changes, helping to maintain system stability. These systems typically achieve response times within 200 milliseconds, meeting real-time performance requirements.

3.2 Deep Learning Methods

In comparison with conventional image processing methodologies, deep learning methods exhibit superior capabilities in feature extraction and enhanced robustness, thereby facilitating more precise driver state recognition. The model design and feature extraction methods can be categorized into three types as follows:

3.2.1 Image-based classification model

Classical convolutional neural networks (CNNs), such as ResNet-50 and VGGNet, have been widely used for classifying facial regions to determine fatigue status. To improve deployment efficiency on embedded devices, several studies have explored the use of lightweight network architectures-such as MobileNet and ShuffleNet V2-for real-time detection.

For instance, Xiong Qunfang applied CNNs to extract global facial features, combined with the Landmark algorithm to localize key facial points. Fatigue indicators were then quantified using the eye aspect ratio (EAR) and mouth opening angle (α). Similarly, Hu Xizhi et al. replaced the VGG-16 backbone in the original SSD framework with ResNet-50 for target detection and feature extraction, achieving a significant improvement in average face detection accuracy (AP increased from 90.7% to 97.7%).

Ding Fusheng used MobileNetV3-56, a lightweight deep-learning model, to detect 68 facial landmarks. The SCRPD-0.5GF model was employed for face detection, providing foundational data for further feature extraction and analysis. ShuffleNet V2's lightweight design also proves suitable for the computational constraints of in-vehicle systems. Zheng Han et al. utilized an enhanced version of ShuffleNet V2k16 for facial keypoint detection and localization. Comparative experiments showed that the improved ShuffleNet outperformed conventional CNN models such as ResNet-50 and VGG in both accuracy and real-time performance, further confirming the effectiveness of lightweight deep learning models in fatigue detection tasks.

This methodology is predicated on the analysis of a single frame image, which benefits from a relatively simple structure and fast response speed, and is thus well-suited to real-time detection. However, it is prone to misjudgment due to its inability to capture temporal changes in facial movements.

3.2.2 Target-detection-based methods

In certain works, target detection algorithms such as YOLOv5 and SSD are utilized to directly detect critical regions, including eyes and mouths, and to perform state identification. YOLOv5 combines the Mosaic data enhancement strategy with an improved loss function, such as EIOU, which significantly improves the accuracy and speed of detection. It also has a good prospect for real-world in-vehicle applications.

Research by Zhang Zhenli and Jin Yunfeng et al. has demonstrated significant performance gains in YOLOv5 through a series of targeted modifications. These include replacing the original GIOU loss with EIOU, optimizing bounding box regression, addressing the issue of hard sample balancing, and accelerating model convergence.

Despite their shared goal of improving YOLOv5 performance, the two studies differ in their specific approaches. Jin Yunfeng et al. incorporated the CBAM module-which combines channel and spatial attention mechanisms-alongside Mosaic augmentation. In contrast, Zhang Zhenli et al. applied the SE attention mechanism, which focuses solely on channel attention, and introduced a Mosaic-8 augmentation strategy. Their modifications aimed to strengthen feature extraction and improve robustness under complex driving conditions.

Experimental results from Zhang Zhenli et al. confirmed that their improved YOLOv5 model achieved a detection accuracy of 97.8%, outperforming both the traditional SSD (92.2%) and the unmodified convolutional YOLOv4 (97.29%). The YOLO series of target detection methods has been shown to localize the face, eyes, mouth, and other regions with strong real-time performance, making it suitable for practical application deployment. However, it has been observed that these methods are not sufficiently sensitive to some subtle fatigue feature recognition, and rely on high-quality training data.

3.2.3 Timing-based modeling approach

The integration of CNN and LSTM architectures has been widely used to model temporal dynamics in image sequences, aiming to detect fatigue-related behaviors such as prolonged eye closure and frequent yawning. For example, Jingyi Xu et al. proposed a framework that combines CNNs for spatial feature extraction with LSTMs for learning the temporal sequence of feature vectors. An attention mechanism was incorporated to assign probabilistic weights to the hidden states of the LSTM, thereby emphasizing key temporal cues that are critical for fatigue detection.

In a different approach, Tencze Zhang et al. computed the information entropy and discretization degree of fatigue-related features using a sliding window technique, followed by LSTM-based classification of the resulting temporal data. This method addressed the limitations of single-frame detection by capturing temporal dependencies. Additionally, they introduced an improved stacking-based multi-model fusion strategy, which integrated temporal features from the face, head, and body posture. The CNN-LSTM model served as a meta-classifier, enabling end-to-end spatiotemporal modeling.

Further research has also explored end-to-end video-based spatiotemporal modeling using 3D CNNs or Transformer-based architectures such as C3D and TimeSformer, which are capable of capturing more complex behavioral patterns associated with driver fatigue.

Temporal modeling methods (e.g., LSTM, GRU) capture dynamic behaviors such as eye closure duration, yawn frequency, etc., through consecutive image frames, which have higher detection stability, but the model structure is complex and computationally intensive and is prone to response delay.

4. Method Comparison and Analysis

Table 1 below is a comparative analysis of traditional methods and deep learning methods.

Table 1. Comparison of Traditional Methods and Deep Learning Methods

	Conventional image processing methods	Deep learning method
Feature extraction approach	Manual design (e.g. eye closure rate, blink frequency)	Automatically extracted by neural network convolutional layer
Accuracy	Stable performance in simple scenes, easily disturbed in complex scenes	High accuracy, especially in dynamic, occluded, or changing light scenes
Calculation Difficulty	Low computational effort, suitable for embedded devices (e.g., in-vehicle systems)	High computational difficulty, relies on GPU acceleration, needs to optimize model lightweight
Real-time (processing delays below 100 ms are designated as “strict real-time,” whereas	High real-time (single-frame detection typically takes 150–180 ms. While this exceeds the strict 100 ms target, its simple algorithms, low computational	Real-time performance depends on model complexity; only lightweight models-via pruning, quantization, or compact

delays of up to 200 ms are deemed acceptable under prevailing industry standards.)	load, and minimal hardware requirements keep latency within acceptable limits.)	architectures-can achieve inference latency of 100–200 ms.
Robustness	Sensitive to illumination and attitude, requires strict constraints	More robust to noise and interference

5. Current Challenges and Research Trends

Currently, conventional image processing approaches to detecting driver fatigue are rule-based and have relatively low computational complexity compared to methods that rely on facial feature analysis. As a result, they are commonly used in embedded systems. However, these approaches struggle to adapt to dynamic environments, such as varying lighting conditions or partial occlusions, and their generalizability is limited due to significant individual differences in eye shape, facial expressions, and yawning behavior. Fixed thresholds are often insufficient to accommodate such variability.

Recent progress in deep learning-particularly in convolutional neural networks (CNNs), time-series modeling, and other architectural innovations-has greatly improved the ability to capture subtle facial movements. Despite these advancements, meeting real-time requirements remains a major challenge. Vehicle systems typically require a response time of under 100 milliseconds, but complex models like LSTMs are computationally intensive and struggle to deliver results within this timeframe. Developing lightweight models is another critical concern.

Future research is expected to focus on multimodal feature fusion-for instance, using CNNs for feature extraction, LSTMs for temporal modeling, and YOLO for key point localization. By integrating these techniques, it may be possible to overcome the limitations of individual methods and achieve accurate, robust, and real-time fatigue detection.

6. Conclusion

Facial feature-based fatigue driving detection methods have made significant progress in recent years, and this study provides a systematic review of facial feature-based fatigue driving detection techniques, focusing on comparing the approach of conventional image processing and deep learning techniques. In the future, with the continuous development of sensor technology and artificial intelligence algorithms, these methods are expected to be widely used in automatic driving and intelligent driver assistance systems, and it is necessary to seek a balance between algorithmic innovation and engineering landing, and effectively promote the transformation of this technology from laboratory validation to the large-scale application of road safety protection systems, so as to provide effective protection for road safety.

References

[1] Zheng, H. (2024). Research on driver fatigue detection algorithm based on deep learning. Information Technology and Informatization, (12), 49–52.

[2] Ding, F., Qin, Y., Zhang, L., & Lü, H. (2024). Driver fatigue state detection method based on multi-feature empirical fusion. Foreign Electronic Measurement Technology, 43(09), 190–196. https://doi.org/10.19652/j.cnki.femt.2406164

[3] Liu, X., Wen, L., Cheng, K., Chen, B., Zhang, Y., & Yu, J. (2024). Driver fatigue detection using improved multi-task cascaded convolutional neural network. Journal of Yibin University, 24(12), 7–11, 68. https://doi.org/10.19504/j.cnki.issn1671-5365.2024.12.02

[4] Xu, J., Tian, J., Liu, X., & Gong, L. (2024). Fatigue driving detection method based on attention mechanism improvement. Sensors and Microsystems, 43(04), 115–118. https://doi.org/10.13873/J.1000-9787(2024)04-0115-04

[5] Zhang, Z., Li, L., Song, J., Fu, H., & Chen, Y. (2024). Research on driver fatigue detection based on deep learning. Electromechanical Engineering Technology, 53(01), 193–197.

[6] Jin, Y., Lu, Z., Wang, R., & Liang, C. (2024). Fatigue driving detection based on improved YOLOv5s. Journal of Beihua University (Natural Science Edition), 25(02), 255–261.

[7] Zhang, T. (2023). Research on driver fatigue detection based on facial and body behavior features (Master's thesis, Lanzhou Jiaotong University). https://doi.org/10.27205/d.cnki.gltec.2023.001004

[8] Hu, X., & Huang, B. (2021). Fatigue driving

detection method based on facial feature analysis. Science Technology and Engineering, 21(04), 1629–1636.

[9] Xiong, Q., Lin, J., & Yue, W. (2018). Fatigue driving state detection method based on deep learning. Control and Information Technology, (06), 91–95. https://doi.org/10.13889/j.issn.2096-5427.2018.06.400

[10] Guo, H., Wang, N., & Guo, H. (2018). Research on fatigue driving warning system based on multiple facial features. Journal on Communications, 39(S1), 22–29.

[11] Zou, X., Wang, S., Zhao, W., Zhao, X., & Li, T. (2017). Fatigue driving detection based on eye and mouth state recognition. Journal of Jilin University (Information Science Edition), 35(02), 204–211. https://doi.org/10.19292/j.cnki.jdxxp.2017.02.015

[12] Li, Y., Zhang, C., & Meng, L. (2014). Research on fatigue driving detection system for train drivers based on head pose features. Journal of Transport Information and Safety, 32(05), 114–119.