Dean&Francis
ISSN 2959-6157

Research Progress on the Effectiveness of Deep Learning-Based Driver Fatigue Fusion Features

Bite Qiu*

Codgel Cranleigh School, Changsha, 410000, China
*Corresponding author: qzxmjb777@gmail.com

Abstract:
In modern society, automobiles are more popular around the world, and people pay more attention to the safety of automobiles According to the information provided by the National Transportation Safety Administration in 2024, there are many traffic accidents caused by fatigue driving. This paper introduces two fatigue recognition methods, and studies the fusion of the two types of features. Aiming at the problem that the existing facial recognition methods are sensitive to environmental interference, this study systematically reviews the application of deep learning models such as YOLO, RNN and ResNet3D in this field from the perspective of facial feature analysis, and proposes targeted optimization strategies. For the problem of strong invasiveness and limited comfort in traditional physiological monitoring methods, this study innovatively constructs a multi-modal feature fusion framework. By collaboratively analyzing facial features and physiological signals, it can effectively balance the accuracy and comfort of monitoring while maintaining the advantages of non-contact detection. At the end of the article, the fusion mechanism of facial information recognition and physiological information recognition is further discussed. The fusion mechanism gives full play to the complementary advantages of the two types of information, and effectively reduces the adverse effects of single information recognition on driving behavior. As mentioned above, this paper analyzes a fatigue recognition method that combines facial information recognition with physiological information recognition. In other literatures, it is also mentioned that this scheme has the advantages of portability and accuracy compared with the traditional fatigue identification method.

Keywords: Deep Learning; fatigue driving; feature fusion.

1. Introduction

Motor vehicles have been widely popularized in recent years, followed by driving safety problems caused by illegal driving by drivers. Some motor vehicle drivers turned a blind eye, resulting in a tragedy. One of the problems that is easy to be ignored is fatigue driving. According to statistics (Fig. 1), nearly 12 % of traffic accidents are caused by fatigue driving problems. From Fig. 2, it is not difficult to see that the night driving crowd and the long driving crowd account for 34 % of the total accident population. Most of the fatal accidents related to fatigue driving behavior are single-person driving accidents. Especially in the case of long-term high-intensity driving, lack of enough sleep, and long continuous driving distance, drivers are more prone to fatigue symptoms, resulting in decreased responsiveness and increased risk of accidents [1]. From the above analysis, it is evident that the issue of driver fatigue is highly serious.

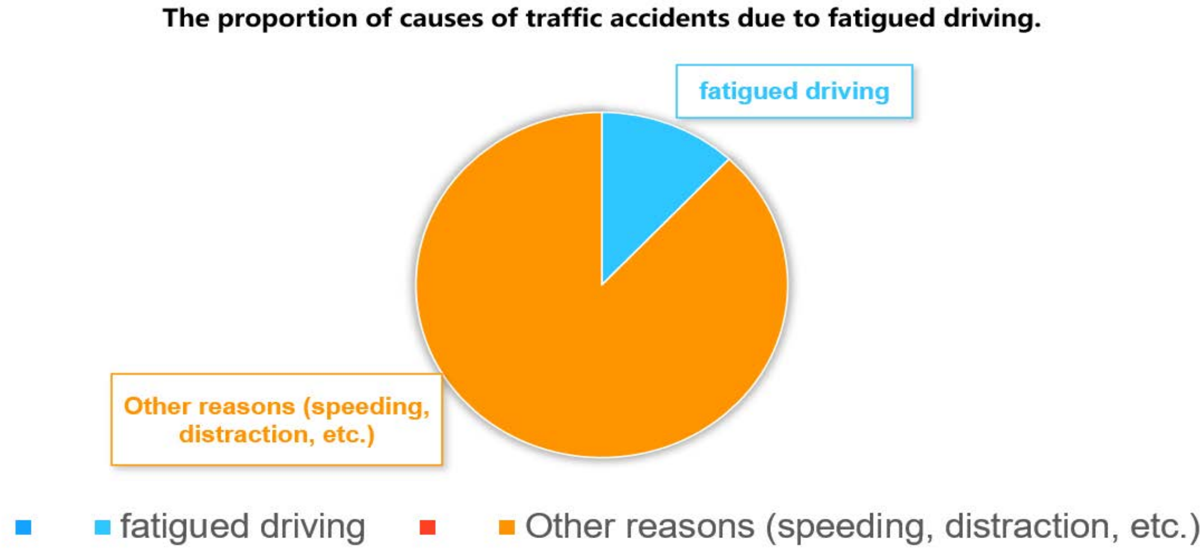

Fig. 1 The Proportion of Causes 0f Traffic Accidents due to Fatigued Driving.

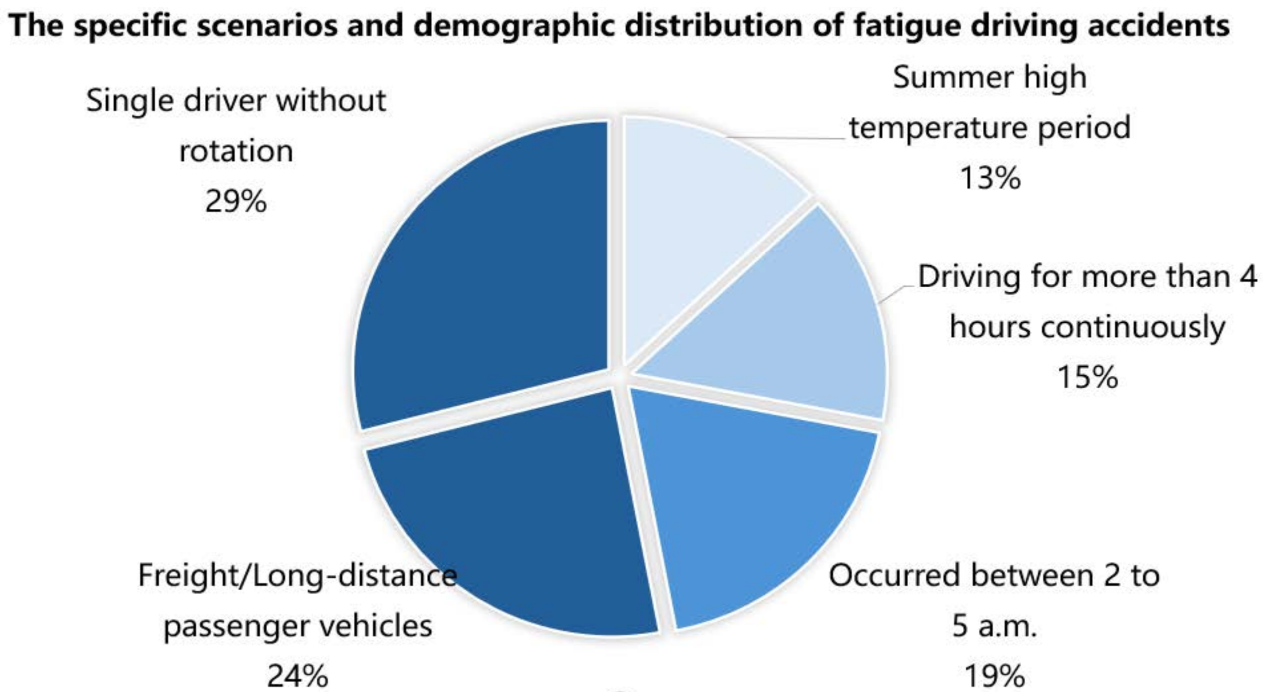

Fig. 2 The Specific Scenarios and Demographic Distribution of Fatigue Driving Accidents.

ISSN 2959-6157

At present, the research on driver fatigue driving has been paid attention to in the field of traffic safety. Now, many advanced technologies are used in the research of fatigue driving, involving multiple dimensions. The emergence and application of many high and new technologies have reduced the frequency of fatigue driving accidents. The current research mainly focuses on the detection of vehicle trajectory, the monitoring of driver's physiological characteristics and the computer vision method based on facial features.

In the literature, Hu et al. innovatively constructed a new development framework in their literature, which can continuously predict the trajectory of the vehicle. the model improves the recognition accuracy and robustness of the algorithm by integrating the 3D reconstruction sequence with the LSTM motion module [2]. In another paper, the Kyushu team ' s operation also gives me a glimpse. By introducing the top hat and bottom hat transformation in morphology, they combine this transformation with KLT tracker and K-means clustering algorithm. This wonderful operation has realized the analysis of the motion characteristics of the target area by the model. This method can continuously track the vehicle while improving the detection accuracy and efficiency, and provide effective guarantee for driving safety [3].

At present, there are many methods for detecting driver fatigue, such as: identifying driver behavior characteristics (such as blink frequency, limb movements, etc.) or detecting the characteristics of vehicle operation to output the driver 's current time status. Peng et al. proposed a highway vehicle trajectory reconstruction method based on automated vehicle (CAV) motion detection data. This method uses the IDM model to judge the motion of the vehicle, and inserts the predicted position data and speed data into the traffic flow, and then uses the IDM model to reconstruct the complete vehicle trajectory, thereby obtaining the driver 's driving state during this period [4].

The main research of this paper is a fusion driver fatigue recognition method based on facial information and physiological information. This paper systematically introduces the adaptive deep learning model, and proposes to analyze the dynamic characteristics of the driver 's facial information and the physiological information indicators extracted synchronously to achieve accurate judgment of the driver 's fatigue state. The purpose of this study is to promote the combination of driver fatigue recognition system and vehicle system. The combined system can provide real-time and portable status detection and early warning services for drivers, so as to effectively prevent the occurrence of fatigue driving behavior and reduce the incidence of related accidents.

2. Research Methods

2.1 Driver Facial Information Recognition Based On Deep Learning Models

2.1.1 Overview of basic content

When applying the deep learning model to analyze the driver's facial features, it is necessary to comprehensively consider factors such as individual differences and environmental interference. By comparing the output results of various models, the accurate and rapid recognition of the driver's state is finally realized. However, in the study, facial information is easy to be disturbed by external environment, and the driver is easy to control their own behavior to disguise their true state and other issues still exist. In order to prevent the above problems from affecting the data, this paper will focus on the YOLO model below, which can detect the driver's accurate fatigue by setting up a facial recognition module.

YOLO is a very advanced target detection algorithm. It only needs to perform a forward propagation on the image to predict the location and category information of all targets in the image. Different from the traditional method, the YOLO model can transform the target detection problem into a regression problem when processing data. It is the implementation of this operation that leads to the extremely fast processing speed of the YOLO model [5]. As shown in the structure diagram of the YOLO model in Fig. 3, the model is mainly composed of three parts: neck, backbone network and head.

Fig. 3 The Yolov10-S Model.

2.1.2 YOLOv7-tiny model

Because the YOLO model performs very well in the face of real-time video and single-subject video, this model shine in the field of driver fatigue detection. For example, in terms of feature extraction, Yao et al. mentioned the strategy of YOLOv7-tiny architecture to implement model compression, through channel pruning and quantitative perception training to achieve parameter reduction, and optimize the dynamic balance between detection efficiency and identification accuracy [6]. As a lightweight version of the series, YOLOv7-Tiny shows strong cross-platform deployment capabilities with its carefully optimized network structure and excellent computational efficiency. Through the collaborative optimization of training strategy and network architecture, the model significantly improves the performance of small target detection and multi-scale target recognition, and further enhances its adaptability in complex environments.

Although YOLOv7-tiny has the advantages of high efficiency, easy training and convenient deployment. However, in the actual research, it has obvious limitations in the extraction and detection of small target features. In the in-depth study, I found that this defect is due to the single target detection mechanism of the YOLO model. It is the existence of this mechanism that leads to the insufficient detection performance of the model in small object-intensive places. This is due to the limitation of grid division in the YOLO model. Each unit grid of the YOLO model can only predict a fixed number of subjects. Therefore, the YOLO model is often powerless in detecting multi-agent targets.

For the YOLOv-7tiny model, this model is a lightweight version of the YOLO model, the model's reduction in volume and capacity also directly leads to defects in the generalization ability and robustness of the model compared to the original version.

2.2 Driver Physiological Information Recognition Based on Deep Learning Models

2.2.1 Overview of basic content

The identification of driver 's physiological information needs to use the technology of biosensors. By allowing drivers to wear such devices directly to detect their own physiological indicators (EEG, electrocardiogram, and electrooculogram, etc.), drivers need to wear EEG caps, ECG patches, smart bracelets and other devices directly. The information identified by such devices can accurately determine whether the driver is in a fatigue state [7]. However, there are also defects, because the existence of the sensor is relatively low comfort and convenience and high cost limits the popularity of this method. Due to the low intensity and poor signal-to-noise ratio of physiological signals, This research needs to rely on a stronger analysis model for processing. Proper handling of the signal is crucial. The original physiological signals should first be preprocessed to remove high- and low-frequency noise. After preprocessing, the signals are then fed into 1D-CNN, RNN, or CRNN models to determine the driver's state.

2.2.2 The RNN model

The Recurrent Neural Network (RNN), which is a kind of neural network with cyclic connection structure. This network is designed for processing sequence data. Thanks to its internal loop connection structure, RNN can effectively capture the time dependence between sequence signals, so that the historical input information has a continuous impact on the subsequent output, forming an information transmission mechanism with memory effect. Compared with the CNN model where the input data and the output results are independent of each other, RNN can identify the correlation between the input data from time to time, thus generating accurate results.

In EEG signal processing tasks, RNN can make full use of its timing modeling advantages to effectively capture dynamic patterns in signals, thereby significantly improving the accuracy of recognition results. Xue mentioned the application of long short-term memory network in his research. As a variant of recurrent neural network, this network effectively alleviated the problem of gradient disappearance of RNN model [8]. In the study of driver fatigue identification, the advantages of RNN in timing are very suitable for such research. Because fatigue recognition is essentially a behavior analysis based on time series, the memory ability of RNN enables it to effectively extract time information in changing scenes, so it is highly consistent with the purpose of the task.

Deep recurrent neural network based on multi-task learning developed by Dong et al., which achieved accurate recognition of the subject's motor intention by extracting and integrating EEG signal features in different frequency bands [9]. Compared with traditional neural networks, RNN, with its unique memory mechanism, can continuously retain historical information through hidden states and act on subsequent calculation processes to form a dynamic information transmission link.

However, RNN has the following two technical limitations in practical applications:

1) Vanishing Gradient: this is one of the most serious problems in RNN. The specific reason is that when a small gradient value is continuously multiplied in the time step, the output result will infinitely approach 0. The emergence

ISSN 2959-6157

of this phenomenon also directly leads to the fact that the data retained long ago is difficult to affect the subsequent output results.

2) Excessive Memory Consumption: The RNN model needs to save the state of each time step in the forward propagation process when dealing with long time series information to ensure the use of back propagation. This operation takes up a lot of GPU memory and also limits the ability of RNN to process ultra-long time series.

2.3 Driver Physiological Information and Facial Information Detection Based on Deep Learning

2.3.1 Overview of Basic Content

The portability of facial feature recognition and the accuracy of physiological information recognition are combined to complement each other, so as to solve the invasive problem of physiological information recognition and the problem that facial information recognition is susceptible to environmental interference. The operation is to import the information into the appropriate model (such as ResNet3D) under the premise of ensuring time synchronization, and then obtain the probability that the driver is in a fatigue state. The schematic diagram is shown in Fig. 4.

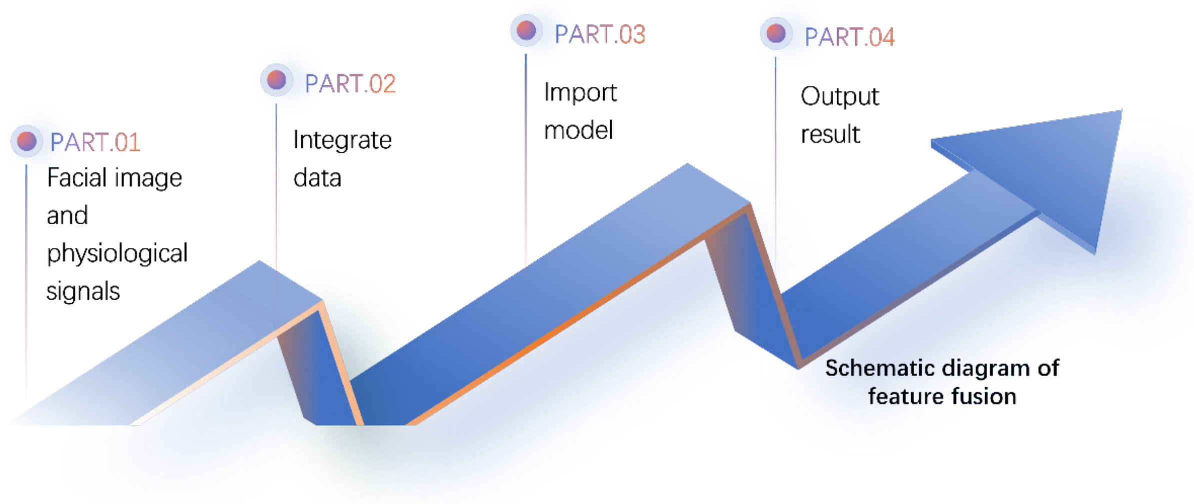

Fig.4 Schematic diagram of feature fusion.

2.3.2 ResNet3D Model

In order to process this information, Han et al. proposed the application of three-dimensional residual network (Residual Network 3D) as a deep convolutional neural network, which inherited the characteristics of residual learning and shortcut linking of ResNet (Residual Network), and replaced the two-dimensional convolution kernel of ResNet with three-dimensional [10]. This ingenious improvement enabled the model to have the ability to extract temporal features, which is particularly suitable for processing time-dimensional data analysis tasks such as video sequences. This key improvement enables the model to effectively analyze the dynamic changes and motion patterns in video sequences, and significantly improves its ability to capture temporal features. that the ResNet3D model performs well in many subsequent performance evaluations, especially in key indicators such as memory footprint, computing speed, and parameter scale. As mentioned above, the three-dimensional convolutional network can effectively extract the temporal features in the eyelid closure sequence, so as to further understand the content in the video.

Because the ResNet3D model has the ability to process time, and the processing of time is very important for this study. In fusion information processing, effectively leveraging the advantages of the ResNet3D model enables accurate identification of temporal relationships among different information sources. By integrating these features with their temporal dependencies, the driver's state can be precisely recognized across different time periods.

However, there are two limitations in the practical application of the model:

1) Excessively Complex Computations: The ResNet3D model will perform additional calculations in the time dimension during use. this operation will cause the cost of calculation to increase from square to cubic, resulting in a surge in computation.

2) Limited time modeling capability: This is because the sensory field of the ResNet3D model in the time dimension is local. Therefore, in order to capture the information of long time series, therefore, multi-layer networks

need to be superimposed, and such operations will cause a series of problems, including the explosion of computing power.

3. Discussion and Analysis

This paper systematically discusses the application of deep learning models such as Res Net3 D, GNN and YOLO in fatigue driving recognition, and then focuses on the problems and improvement methods of these models. For the YOLO model, it can improve its ability to identify small targets through data enhancement; the GNN model can accelerate the training and improve the stability of the model by standardizing or normalizing the node features. Some problems of the ResNet3D model originate from the 3D convolution kernel, which can be decomposed into 2D spatial convolution kernel and 1D temporal convolution kernel, thus significantly reducing the amount of computation and memory consumption. In addition, this paper summarizes the contents of such models (Table 1).

Table 1. Summary of Three Deep Learning Models

Model	Advantages	Disadvantage	Improvement method
YOLOv7-tiny	Strong generalization ability and simple structure	Difficult to recognize small objects	Data augmentation Structural improvement
RNN	Have memory capability and flexible model architecture	Vanishing gradients and Memory usage is too high	Feature standardization and normalization and application of Long Short-Term Memory Network
Res Net 3D	Time information can be extracted	The computational load and memory consumption are too high	Decompose the 3D convolution kernel into 2D and 1D

4. Summary

This paper focuses on the identification and research of fatigue driving, systematically introduces three deep learning models, respectively shows the efficiency and accuracy of the three models in driver information extraction, and analyzes the limitations and optimization directions of the three models in detail. This study is committed to building a more accurate driver fatigue identification technology system, and promoting the public's scientific understanding of fatigue driving hazards through empirical research results, so as to provide theoretical support and practical path for road traffic safety management. In the face of the urgent problem of fatigue driving, the effective use of deep learning model is of great significance to reduce the incidence of related accidents. The real-time monitoring system based on deep learning can continuously track the changes of the driver 's mental state, and trigger the early warning mechanism in time when the fatigue symptoms are detected. Through the combination of real-time early warning and active intervention, the safety accidents caused by fatigue driving can be effectively controlled.

References

[1] Zhang Zhenli, Li Liang, Song Jingjing, et al. Research on Driver Fatigue Detection Based on Deep Learning. Mechanical & Electrical Engineering Technology, 2024, 53(01): 193-197.

[2] Hu H N, Cai Q Z, Wang D Q, et al. Joint Monocular 3D Vehicle Detection and Tracking. Computing Research Repository, 2019: 5389-5398.

[3] Kyushu U,Tanaka K, Yamamoto T, et al. Efficient vehicle detection and tracking strategy in aerial videos by employing morphological operations and feature points motion analysis. Multimedia Tools and Applications, 2020, 79(35): 26023-26043.

[4] Chen Peng, Wang Tong, Zheng Nan. Reconstructing Vehicle Trajectories on Freeways Based on Motion Detection Data of Connected and Automated Vehicles. Journal of Intelligent Transportation Systems, 2021, 26: 639-654.

[5] Shu Liangjun, Ding Fei, Wang Bingkun, et al. Mobile Target Detection with an Integrated Attention Mechanism and Lightweight Feature Fusion Network. Flight Control & Detection, Aarly access, Oct. 2025.

[6] Yao Yining. Research on Driver Fatigue Recognition Based on Mouth and Eye Movement. Xi'an: Xi'an Technological University, 2022.

[7] Zhou You. Research and Implementation of a Cognitive Fatigue Recognition System Based on Multimodal Feature Fusion. Nanjing: Nanjing Forestry University, 2024.

[8] Xue Huixia. Research on Driver Steering Intention Recognition by Integrating Internal and External Cockpit Features. Chongqing: Chongqing University of Technology, 2025.

[9] Dong Yan-jiao, Han Yuan-jie, Liu Jie-li. A Target Identification Method Based on Information Fusion of GNN-DS. Guilin University of Electronic Technology, 2009, 25(22): 193-194+172.

[10] Han Xiangyu. Research on Fatigue Recognition of Operators Based on Fusion of Multiple Eyelid Features. Hengyang: University of South China, 2023.