Journal of computer science and software applications
https://www.mfacademia.org/
ISSN:2377-0430
Vol. 4, No. 4, 2024

Multimodal Fusion-Based Analysis and Detection of Fatigue Driving

Mary Clark

University of Alabama at Birmingham

mary.clark@uab.edu

Abstract:

Fatigue driving is a major hazard to road safety and a common cause of traffic accidents. Developing effective methods to monitor and identify driver fatigue for early warnings is a critical research focus. Traditional detection methods often rely on monitoring specific physiological parameters, but these can be less effective due to environmental interferences like poor image quality, complex lighting, or unstable equipment. In contrast, computer vision-based methods offer accurate, non-invasive detection and a better user experience, though they still face challenges such as high error rates and limited robustness due to individual variations and the ambiguous nature of fatigue. Addressing these issues, this paper introduces a fusion model that combines facial detection and head posture analysis using a multi-task convolutional neural network (MTCNN). This model, utilizing a multi-task cascade architecture, simultaneously performs comprehensive facial posture analysis and facial detection. Additionally, a lightweight network cascade based on SqueezeNet is employed to detect facial keypoints, identifying 72-point coordinates crucial for assessing fatigue. By extracting fatigue features from these keypoints, a multimodal fusion approach to determine fatigue is proposed. The paper concludes with the development of an online fatigue detection and early warning system, showcasing practical applications of these research findings

Keywords:

Deep Learning; Fatigue Detection; Multimodal Fusion.

1. Introduction

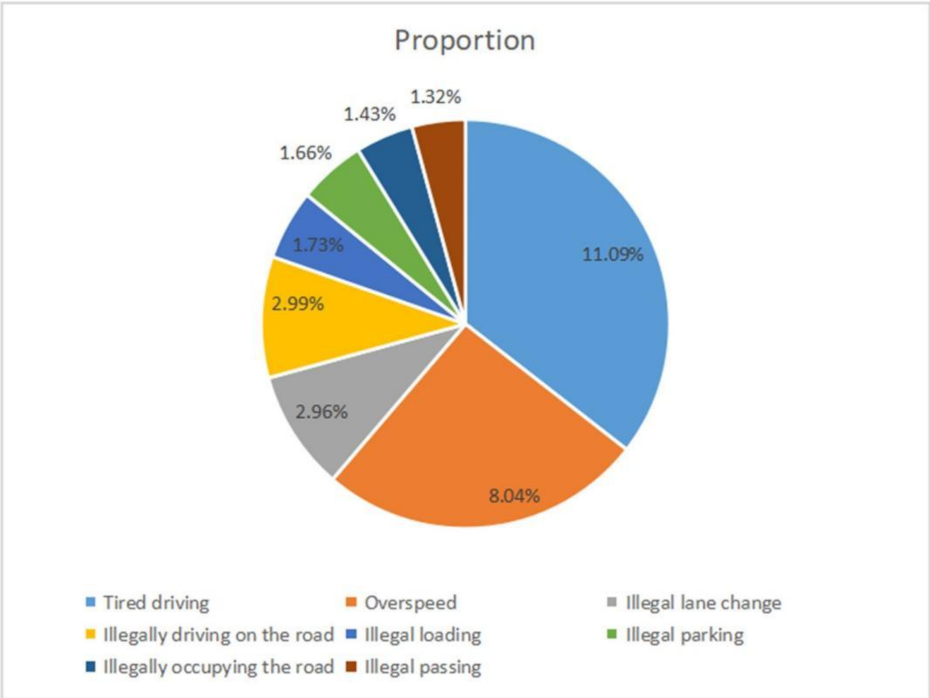
1.1. The Research Background and Significance

Fatigue driving always threatens the road traffic safety. With the rapid development of unmanned technology and 5G technology commercial landing, the car driving auxiliary technology has also been many manufacturers layout to their product line, many models have begun to carry driving assistance system (ADAS), ADAS including adaptive cruise control system, parking assist system, vehicle deviation warning system, lane change auxiliary system,collision warning system, etc. Although these driver assistance systems have brought unprecedented convenience and driving experience to the drivers, the traffic accidents caused by fatigue driving every year still always threaten the road traffic safety, and they need to be solved urgently. Therefore, the driver fatigue detection system should be added to the future driver assistance system to solve the problem of fatigue driving.Since there is no clear standard for the fatigue state, the proportion of traffic accidents caused by actual fatigue driving may be higher than the statistics. The dangers of fatigue driving threaten global road traffic safety.According to the National Highway Traffic Safety Administration, the traffic accident rate caused by fatigue driving in the United States is equally bad, with 1,500 direct deaths and 71 0,000 personal injuries. According to the statistics of relevant departments in Germany, France, Japan and other countries, the casualties caused by fatigue driving account for about 15% to 20% of the total number of personal injury accidents.

Figure 1. Analysis of the main causes of highway accidents

Fatigue driving causes high accident rate and serious consequences, but effectively reminding the driver of the fatigue state can greatly reduce the probability of traffic accidents caused by fatigue.

At present, most studies are fatigue by detecting changes in the biological state of the human body during fatigue.



Determination of L. L status. In the traditional driving fatigue detection methods, the disadvantages such as poor reliability, non-real-time and unfriendly interaction are increasingly prominent, which cannot meet the needs of application scenarios. A driver fatigue detection and early warning device that can solve the above pain points is crucial. This paper studies a real-time, reliable and non-contact driving fatigue detection method. Facial area detection and facial multi-attribute analysis are realized, and the detection results are used to extract the fatigue characteristics. Finally, the fatigue state of drivers is evaluated according to the fusion decision algorithm, and an online fatigue driving detection system is set up to prompt and early warning.

1.2. Research Status

Early studies on fatigue detection were mostly carried out in the medical field, by monitoring heart rate, brain waves, and blood pressure.

And other changes in the fatigue status. With the deepening of substantive research, diversified detection methods are constantly being proposed.

1.2.1. Classification of Fatigue Detection Methods

The specific detection methods are mainly divided into subjective detection methods and objective detection methods.

The subjective detection method is to judge the existence of fatigue through self-evaluation and others' evaluation. In the self-evaluation, the Karolinska Sleepiness Scale (KSS) is most commonly used as the evaluation standard. This method is simple, universal and widely used, but it cannot play its value for the real-time and auxiliary requirement scenarios. In particular, the subjective evaluation method is highly subjective, which leads to the failure to effectively evaluate its accuracy.

The objective detection method eliminates the interference of subjective factors, and monitors the

ISSN:2377-0430
Vol. 4, No. 4, 2024

specific fatigue characteristics with the help of the testing equipment.

Measurement, such as physiological information, vehicle driving information, etc. A way to judge the fatigue state. It is mainly divided into three categories, as shown in Table 1:

Table 1. Objective method comparison of fatigue test

objective ethods	Method description	accuracy	practicability
The physiological characteristics of the driver	Detection of physiological signals such as EEG and blood pressure	fine	same as
Characteristics of vehicle behavior	Test the driving state of the motor vehicle	same as	fine
Driver's facial features	Driver to for changes in the facial features	fine	fine

1.3. This Paper Studies the Content

The main content of this paper is based on the fatigue driving detection technology of deep learning. At the same time, this research focuses more on the analysis of face attributes under the road traffic environment, and the main research contents are as follows:

(1) Study the deep network model of facial detection and facial attribute analysis fusion

Based on the study and improvement of existing networks, we propose a convolutional neural network model with multi-task cascade structure.

Complete the task of face detection, the facial attributes are analyzed to reduce the task process of face analysis. Aiming at the complex light, partial occlusion, and low resolution in practical application scenarios, the speed and accuracy of multi-task model processing are improved by optimizing the network structure and training methods.

(2) Study the network model of face keypoint detection using the cascade network strategy For the traditional face key points, the network network layer is too deep, the model parameters are various, the operation efficiency is not high, and check.

For the problem of low measurement accuracy, a face key point detection model based on cascade strategy is proposed. This method uses lightweight network, has reduced model size, improves the running time of the detection algorithm and the practicability of the algorithm.

(3) Study the application of facial multi-attribute recognition in the state analysis of road traffic drivers The driver's facial attribute information reflects the actual driving state. Through the multi- attribute fatigue feature of the driver's face.

To determine the fatigue driving status, in this study, we analyze multiple attributes of the face, extract multiple groups of fatigue characteristics, and use the method of information fusion, so as to realize the practical application function of fatigue driving detection.

2. Fusion Network for Face Detection and Facial Pose Analysis

Some excellent target detection networks such as Faster RCNN, SSD, YOLO use multi-task learning strategy. These excellent target detection networks all detect and classify targets at the final output, using a network similar to the traditional classification or regression network. The fusion network of face detection and head posture analysis is inspired by the simultaneous detection and analysis task under the single model structure, and adds the regression task of head posture analysis to the traditional detection task.

2.1. First Level Network

FCN has the ability to handle pictures of any size, and its convolution calculation method is basically similar to the method of "sliding" window to get the candidate window. Since in the highly convolution optimized cuDNN library, the convolution operation can classify all the candidate Windows by one calculation, this operation method is more efficient than the CPU serial processing with "sliding" operation. You can modify the same step size in the convolution operation by modifying the step size in the slide window.

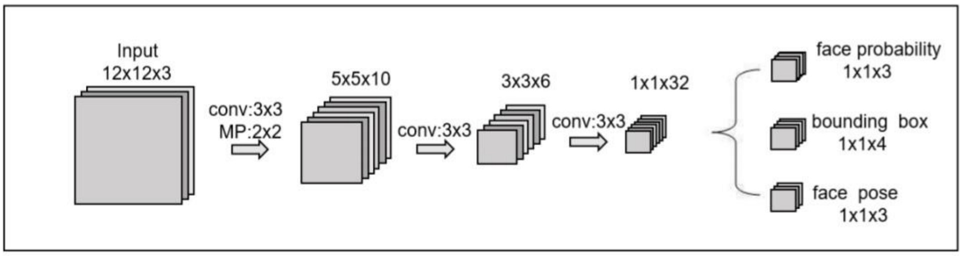


Figure 2. The first level of the network structure

The first convolution layer uses a filter of 3310 and the pooling layer uses a maximum pooling of 22. The second convolutional layer uses a 3316 size filter. The third convolutional layer used a 3332 size filter. Finally, 11-size filters were used for face classification, bounding box regression, and facial pose prediction.

To improve the ability of the model to fit to the null attachments, the PReLU activation function replaces the longer-used ReLU activation function by default. The ReLU activation function lacks the ability to respond to the negative value input, while PReLU adds a negative value input linear term on the basis of the ReLU activation function, and still has a certain response ability when the input is less than 0.

The judgment of facial area draws on the processing method of the literature. In the traditional MTCNN, P-Net uses 1x1x2 vector in the network structure to make facial judgment. Traditional model training only comprehensive data and facial data training, all facial area and the comprehensive area defined from cross and ratio (Intersection over Union, IoU) calculation, IoU as a commonly used evaluation of face detection, in the imitation real southwest university engineering master's degree thesis 20 test and analysis section is introduced in detail. When IoU is greater than 0.7 is identified as comprehensive area, less than 0.3 is identified as non- facial area. However, a large number of IoU images between 0.3 and 0.7 often contain the data of the face area to be trained. Direct abandonment is undoubtedly a waste of a large amount of valuable data, in order to avoid the occurrence of such phenomena. The candidate window is divided into three categories, namely: comprehensive part area, some facial area and non-facial area, mining the potential value of IoU images between 0.3 and 0.7.

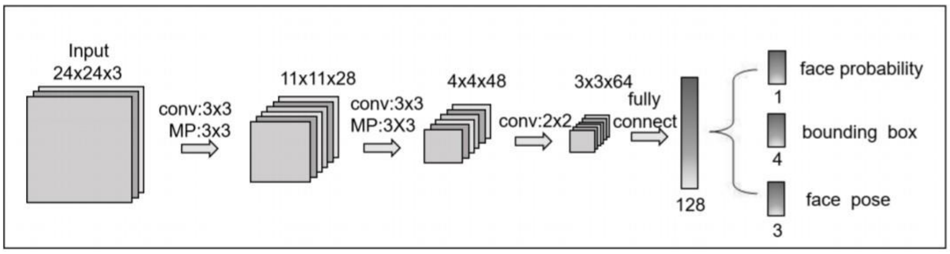
Through the generation of face area confidence map, we can see that the three classification face classifier can get better face classification effect.

Since the head pose analysis task is added along with the face detection task, the third type of output of the model requires a predictive output of the head poses. Head posture is one of the important facial attributes, and face detection is more difficult under multiple posture and abnormal posture. The conventional face pose estimation is to set a fixed angle value at intervals, and then determine the closest angle value to the detection result, and the output is the discrete angle value. While the facial pose is in a three-dimensional space, the pose is estimated as continuous angular values. Especially in specific application scenarios, the facial posture changes frequently and varies greatly. Due to the strong correlation between head posture and face detection itself, and the head posture analysis task is added to face detection, and the multi-task learning strategy improves the performance of the detector. Generally, the human head can be modeled as a rigid object without entity, with the degree of freedom in three directions, described by the attitude Angle, including in-plane rotation Angle (yaw), up and down flip Angle (pitch), left and right flip Angle (roll), with the positive face as the starting point, the right hand is the positive direction, and the radian unit is used when output.

2.2. The Second Level Network

By filtering with the set threshold, the second level network can refine the output results of the first level network, and the output of the boundary regression box can get the candidate box corresponding to the face region. NMS is then used to pick the candidate boxes with the highest scores in the neighborhood and to suppress the windows with low scores. The candidate box after the non-maximum suppression treatment is further processed by the second level network.

Figure 3. The second level of the network structure



2.3. Third Level Network

Compared with the first two layers, the third level network structure is the most complex. In order to ensure the final model accuracy, after the first two layers have been screened for strict candidate regions, the deeper convolution layer is used to extract the input image features, and the full connected layer is used to ensure the output accuracy. To ensure that the network has a wider input, the third level network uses an input of 48x48. Due to the obvious reduction of candidate regions, the actual computational cost is not significantly improved, and the accuracyof the model maximizes the third level network.

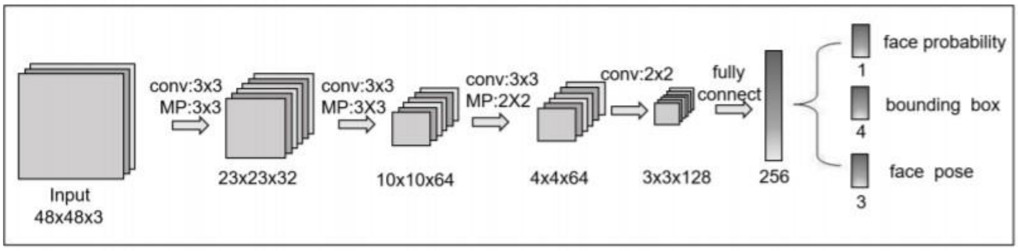


Figure 4. Third level network structure

2.4. Experimental Results and Analysis

The algorithm performance is evaluated to compare the fairness of the experiment, and the unified test code provided by FDDB is used. The algorithm proposed in this study is named as the face detection and head pose analysis fusion network (Face Detection And Pose Analysis United CNN, FPU-CNN), The test results were compared with the research algorithms PyraminBox, Conv3d, FastCNN, CCF, npdface, MultiresHPM, DDFD and CasCNN in the last 5 years, Algorithmic data was obtained from the FDDB official website, Based on the algorithm evaluation results on FDDB and the accuracy comparison results under low FP (False Positive), The proposed fusion model of face detection and head pose analysis is better than most face detection algorithms in face detection tasks, At the same time, there is still a gap with some excellent face detection algorithms such as PyraminBox, FastCNN, etc.

3. Facial Attribute Analysis Network

Based on the facial attribute analysis of facial key points, this study is a further extension of the facial attribute characteristics based on the facial recognition and facial pose analysis fusion network (FPU-CNN). After the face detection and head attitude analysis of the face detection and head attitude analysis, the facial attribute analysis network based on face key points is used to conduct more in-depth facial attribute analysis. For the analysis of face attributes, the face attributes are analyzed according to the specified regional key points.

Recently, the main research direction of deep convolutional network is on the accuracy of model detection. For the same detection accuracy, a streamlined model architecture has the following advantages: reducing the communication pressure with the server during training, less parameters, less data downloaded from the cloud, and more suitable for deployment on memory-limited devices such as FPGA. Based on these advantages, this study proposes that SqueezeNet network completes the detection task of face keypoints. SuqeezeNet achieves the same accuracy rate as AlexNet on ImageNet, but only uses 1 / 50 parameters compared with AlexNet. Using model compression techniques, it is possible to compress the SqueezeNet to 0.5MB, which is 1 / 510 of the AlexNet model size.

In the facial attribute analysis task, the state of the eye and mouth was judged, and the network cascade was used to output the facial key points, eye key points and mouth key points respectively. First in the first level of the network, output 72 key points of the face. After the first level network initially outputs the 72 key points of the face, the approximate positions of the eye (left and right eye) and the mouth are roughly obtained. Then, the second level network still uses the cascade network structure based on the SqueezeNet network to further accurately determine the key point coordinates of the left eye, right eye and mouth.

For the face key point detection task, this study uses the network cascade method to further improve the Squeeze-Net network, forming the same type of network with different complexity. In addition to the advantages of not losing the model scale, the improved network still has a good detection accuracy in the face key point detection task.

4. Multimodal Fusion Fatigue Driving Detection Method

Based on the above results of the fusion network of face detection and face posture analysis, face key point detection and face attribute analysis network, this study proposes a multi-mode fusion fatigue determination criterion to integrates various fatigue features to determine the fatigue state.

Compare the proposed multi-index fusion detection algorithm with a single index detection algorithm:

(1) Compared with the single index detection algorithm that only uses the PERCLOS index, it can be seen that the detection results of the multi-index fusion detection algorithm in this studyhave been improved in the awake state, fatigue state and severe fatigue state. After analyzing the samples with only PERCLOS identification indicators, it is found that the samples with wrong identification generally have large changes in external light, and the side often appear in normal driving, and the interference of positive strong light greatly reduces the detection accuracy of the algorithm. In the multi-index strategy, although the analysis of mouth state and eye state is greatly influenced by light factors, the head posture analysis task is not sensitive to the complex illumination of the environment. The fatigue index extracted by the head posture analysis ensures the accuracy of the fusion decision method and improves the robustness of thealgorithm.

(2) Compared with the fatigue identification results of only using the nodding frequency condition attribute, the fatigue determination method of multimodal fusion proposed in this study is also significantly improved. To nod frequency alone this fatigue characteristic identification error sample analysis, in the process of driving drivers communicate with others or humming songs, nodding behavior under the fatigue state, according to the single condition attribute appear misjudgment, according to the fusion decision method can be corrected to such misjudgment, ensure accurate identification results.

(3) Compared with the fatigue identification results of only using the yawning frequency condition attribute, the fatigue determination method of multi-feature index fusion proposed in this paper also has obvious advantages. The analysis of the identification error samples that used the fatigue feature of yawning frequency, which is unique, shows that, similar to the error sample that only used the PERCLOS index as the fatigue determination standard, the eye and mouth properties of the face were greatly affected by light. Multimodal strategy will reduce the misjudgment of a single index strategy.

References

[1] Tong Ji. The number of motor vehicles in China continues to grow rapidly [J]. Education Vision, 2018(4): p.75-75.

[2] Li Duhou, Liu Qun, Yuan Wei, etal. Relationship between fatigue driving and traffic accidents [J]. Journal of Transportation Engineering, 010 (2): p.104-109.

[3] Saroj K.L.Lal, Ashley Craig.A critical review of the psychophysiology of driver fatigue[J].Biological Psychology, 55(3):p.0-194.

[4] Li Zengyong, Jiao Kun. Correlation analysis of simulated mental load and heart rate variability [J]. Beijing Biomedical Engineering, 2002,21 (3): p.190-193.

[5] Bao Sarina, Zhu Shlin, Qi Chunhua, et al. Experimental study on drivers' neck muscle fatigue [J]. Chinese Journal of Safety Science, 2014,24 (5): p.68-72.

[6] Schmidhuber, Jürgen.Deep learning in neural networks: An overview[J].Neural Netw, 61:p.85-117.

[7] Zhu Rongxin:Detection and evaluation of combined harvester driving fatigue based on physiological signals(Ph.D,Northeast Agricultural University, China 2016),p.1