Theory and Practice of Science and Technology
2025, VOL. 6, NO. 8, 218-221
DOI: 10.47297/taposatWSP2633-456940.20250608

Research on Driver Fatigue Early Warning System Based on YOLO Algorithm

Jiaqi Zhang

School of Data Science, City University of Macau, Taipa, Macau, 999078, China

ABSTRACT
Driving fatigue is one of the key factors that cause major traffic accidents. In this paper, a real-time driver fatigue early warning system based on YOLO (You Only Look Once) algorithm is proposed and investigated to address the limitations of traditional fatigue detection methods. The core of the system adopts the improved YOLOv7 model to efficiently detect the key facial regions, combines with the dlib library to accurately locate 68 facial feature points, and extracts the eye aspect ratio and mouth aspect ratio as the key fatigue indicators. By introducing the Perclos model, the frequency of eye closure and yawning is analysed in 150 frames to quantify the fatigue state. Experiments on the NTHU-DDD dataset show that the system has an accuracy of 89.2% for distracting behaviour recognition and 91.5% for fatigue state determination, with good real-time performance. Meanwhile, this paper provides an in-depth analysis of the challenges faced by the system in complex scenarios such as extreme lighting and occlusion, and looks forward to future research directions such as lightweight deployment, multimodal fusion and cross-scene generalisation.

KEYWORDS
Yolo algorithm; Fatigue driving detection; Real-time warning; Computer vision; Multimodal fusion

1 Introduction

With the surge of car ownership, fatigue driving has become an important threat to global traffic safety. According to the World Health Organisation, about 30% of road fatalities are directly related to fatigue driving. Traditional detection methods, such as vehicle trajectory analysis and physiological signal monitoring, have defects such as high latency and intrusiveness, while real-time monitoring technology based on computer vision has become a research hotspot due to its non-contact and high-precision characteristics. The YOLO series algorithms, as a representative of single-phase target detection, have rapidly gained popularity in the field of fatigue detection by virtue of their end-to-end processing architecture and excellent real-time performance. This paper focuses on the technical evolution of YOLO algorithms in fatigue warning and analyses the future challenges and development trends.

2 Domestic and overseas research progress and current situation

2.1 Driver fatigue detection

Compared with traffic accidents in the awake state, fatigue driving accidents are often more serious in terms of lethality and damage. Currently, the identification technology for driver fatigue is mainly divided into two categories: active and passive. Active methods rely on subjective feedback from the driver, such as filling out questionnaires, scales or verbal answers, which has the advantage of being easy to operate, but has limitations such as interference with driving, response lag and strong subjectivity. Passive methods, on the other hand, have higher objectivity and timeliness through continuous collection of driver physiological or vehicle operation data and real-time analysis and judgement with the help of specific algorithms. At this stage, the research and practice of passive fatigue detection mainly focuses on three directions: fatigue analysis based on driver physiological signals, fatigue assessment based on vehicle handling performance, and fatigue state recognition based on computer vision technology.

2.2 Current status of domestic and overseas research

In the field of driver fatigue detection, biometric-based methods mainly rely on wearable sensing devices. Such methods assess the driver's fatigue state by capturing physiological signals such as brain waves and electromyographic activities during driving in real time. For example, a team of researchers from the University of Science and Technology of China (USTC) designed smart glasses to capture data on drivers' head movements and electroocular signals. They used Principal Component Analysis (PCA) to fuse multiple fatigue features and found that the fused features were highly correlated with the fatigue determination results based on EEG signals through Pearson correlation analysis. Other researchers explored non-contact respiratory monitoring techniques, and they confirmed the correlation between changes in respiratory rate and fatigue features using the intensive optical flow method. In terms of electroencephalogram (EEG) applications, foreign scholars have developed an intelligent detection system, the core of which is an innovative feature-generating network that uses texture descriptors to extract key information for fatigue identification. In order to improve the detection accuracy, domestic researchers also proposed a multidimensional fusion strategy that combines EMG signals with A-mode ultrasound signals for fatigue analysis. According to experiments, the recognition accuracy of this fusion method was improved by 8% to 13% compared with the single use of EMG or A-type ultrasound signals, respectively. In addition, a foreign research team has successfully achieved fatigue assessment by applying a neural network model to analyse drivers' heart rate variability (HRV) data, with a model accuracy of 90%.

Driving performance-based methods constitute the second category of major research paths. The core of this type of approach lies in inferring the driver's fatigue state characteristics by parsing the key dynamics parameters during vehicle driving, such as longitudinal velocity, longitudinal acceleration, throttle pedal opening, and steering wheel angle. Aiming at the effects of fatigue in different driving environments, some scholars have systematically investigated the effects of fatigue on handling behaviours under a variety of road conditions by using a driving simulator platform, and their analyses have focused on typical performance indicators such as longitudinal speed change, steering wheel operation dynamics, and vehicle lateral position deviation. Another research team constructed structural equation modelling with a sizeable sample of 307 drivers to explore in-depth the intrinsic correlation between driver fatigue and a variety of driving performance. In order to assess the effect of vehicle equipment on fatigue, a group of subjects with rich driving experience were tested under real road conditions, and the study focused on analysing their vehicle dynamics to reveal the actual effect of the sports seat system on relieving driver fatigue in the driving process. In addition, some researchers in China have proposed a detection framework that integrates multiple sources of information, including longitudinal speed, steering wheel angle, and lateral vehicle position, with the aim of achieving a more comprehensive and accurate identification of driver fatigue.

3 Basic theory of relevant models

3.1 YOLOv5 network structure

The YOLOv5 backbone network is based on the CBS module (Conv-BN-SiLU), the C3 module and the SPPFF module, which extracts multi-scale features while compressing the image size and expanding the number of channels. The neck network combines the top-down feature pyramid network FPN with the bottom-up path aggregation network PANet to form a dual-path structure, which deeply fuses shallow, medium and deep feature information. The head network is responsible for predicting the location, category and confidence of the target. Three different scales of detection heads are included to process the feature maps from the neck network, which correspond to output 80x80, 40x40, and 20x20 feature maps when inputting a 640x640 image.

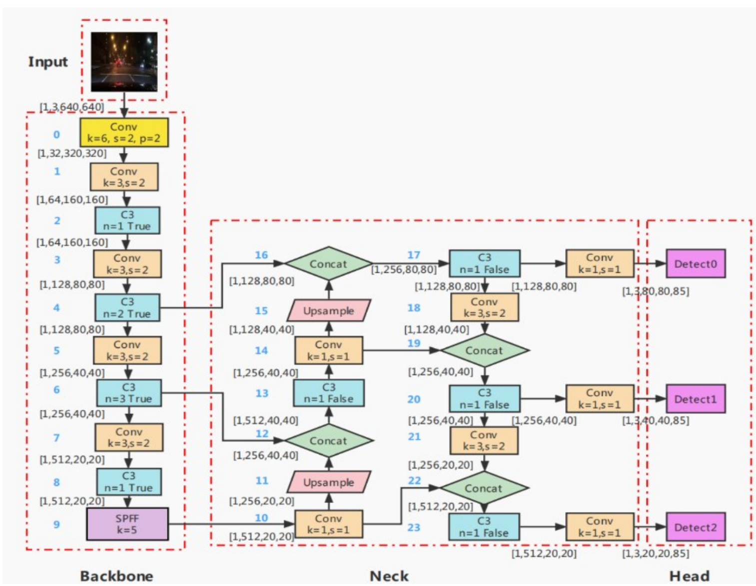

Figure 1

3.2 YOLOv7 network structure

YOLOv7 adopts a multi-stage backbone network from Stage 2 to Stage 5, which mainly consists of CBS, ELAN, and MP modules. residual structure is incorporated into ELAN and MP modules to alleviate the problem of deep gradient disappearance. The input 640×640×3 image is processed in four stages, and the original feature map resolution is halved and the number of channels is multiplied step by step, and the final output is a 20×20×1024 feature map. The neck network uses FPN combined with PANet structure for multi-scale feature fusion. The Stage3, Stage4 and Stage5 features processed by SPPCSPC are fused to generate P3-P5 and N3-N5 through top-down (up-sampling) and bottom-up (down-sampling) paths, which combines shallow detailed features such as contours and colours with deeper semantic features to significantly improve the detection capability of small targets. The head network, on the other hand, adopts a decoupled head design to separate the two tasks of classification and bounding box regression, and follows the anchor box-based detection method. Three different scales of fused feature layers N3, N4, N5 are inputted into the detection head, and each feature point predicts three anchor frames, and a dynamic label assignment strategy is used to judge the category and location of the targets in the anchor frames.

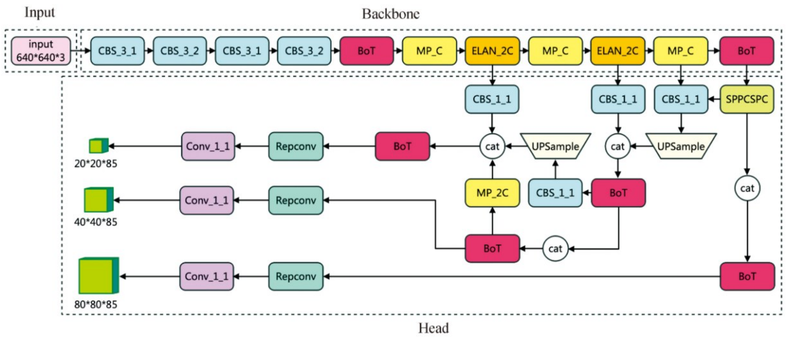

Figure 2

3.3 model comparison

The YOLO family is widely known for its excellent balance of speed and accuracy, and YOLOv5 and YOLOv7 are two important representative models in the family, which, despite the continuity of their names, were developed by different teams and have significant differences in design concepts and implementation details, each showing unique advantages and limitations. YOLOv5's main advantages are its high engineering efficiency and deployment friendliness. YOLOv5's main advantage is its high engineering efficiency and deployment friendliness. Its codebase is well structured and documented, and it provides a wide range of pre-trained models from the very small YOLOv5n to the large YOLOv5x, which makes it easy for developers to quickly get started, train, and deploy it to a variety of platforms. SPPF (Spatial Pyramid Pooling Fast) module, PANet Path Aggregation Network, and other mature technologies, YOLOv5 achieves a good level of accuracy while maintaining a high inference speed. However, YOLOv5 has relatively limited innovation in model architecture, and its core components are mostly borrowed or optimised from its predecessors such as YOLOv3, YOLOv4. It relies heavily on training techniques, and the potential of the model's own architecture has been gradually exploited in the subsequent development, while YOLOv7's core strength lies in its innovation in model architecture and the higher accuracy it achieves. YOLOv7 is not a direct iteration of YOLOv5, but presents a series of original designs, including: Extended Efficient Layer Aggregation Network (E-ELAN) as the backbone, which significantly improves the learning ability of the network without compromising the gradient path. Innovative model scaling strategies to more efficiently adjust model depth and width. Dramatically improve the final model accuracy through more sophisticated training strategies. This enables YOLOv7 to significantly outperform YOLOv5 and other competitors of the same period on benchmark datasets such as MS COCO, especially on larger model sizes. The main drawbacks are that the ease of engineering and deployment is slightly less than YOLOv5, and the ease of use of the codebase and the scale of community support are not as mature as the latter in the early days. Some of the complex training strategies increase the training cost and the requirement of computational resources, while some innovative structures may bring a slight speed compensation, which requires careful selection of model variants in the pursuit of the ultimate inference speed.

4 Driver fatigue warning system

4.1 Dataset collection

The NTHU-DDD (Drowsy Driver Dataset) dataset was used for model training and evaluation. This dataset is one of the benchmark datasets widely used in the field of fatigue driving detection, and its core value is to provide synchronised eye and mouth state information. The dataset is a collection of facial videos recorded by 18 subjects in a simulated driving environment, containing more than 650,000 frames of annotated images in total, covering a wide range of states such as awake, mildly fatigued and severely fatigued.

4.2 Experimental procedure

The experiment adopts a modular architecture design, integrating computer vision and deep learning technologies to achieve multimodal analysis of fatigue state and behaviour recognition. The system core consists of three key modules. Module 1, YOLOv7 target detection module, implements real-time target detection based on YOLOv7, specifically trained to identify four distracting behaviours: using a mobile phone (phone), smoking (smoke), drinking water (drink), and fatigue (drowsy), accelerated by FP16 semi-precision inference, which improves the speed of inference by more than 40% in GPU environment, and optimizes the detection results by using the Non-Maximum We use non-maximum suppression (NMS) to optimise the detection results, and the confidence threshold is set to 0.7 to effectively filter the output of standardised detection results from false positives. Second, fatigue feature analysis module, based on dlib's 68-point facial keypoint detection to achieve fatigue feature extraction, the eye aspect ratio (EAR) and mouth aspect ratio (MAR) algorithms are set to EAR_THRESH=0.3, MAR_THRESH=0.65 to determine the state of the eye closure and yawning. Third, for fatigue determination, the system introduces the Perclos model to calculate the fatigue score with 150 frames as a cycle, integrating the frequency of eye closure and yawning, and when the score exceeds 0.38, it is judged as a fatigue state and an alert is issued.

4.3 Results and problems

The following is the result display, when checking the behaviour of using mobile phone, drinking water, smoking and fatigue will be carried out instant alert. The accuracy rate of identifying distracting behaviours is 89.2%. The accuracy rate of determining the state of fatigue is 91.5%.

Figure 3

At the same time, however, there are some problems with the system that need to be improved. Firstly, the detection accuracy of the YOLO algorithm for critical areas such as the eyes and mouth drops dramatically when encountering extreme lighting or weather conditions such as strong backlighting, low light at night, alternating light and darkness in tunnels, and reflections from rainy and foggy windshields. Existing methods such as image enhancement and multi-spectral fusion suffer from high cost and insufficient generalisation ability, and lack a universal low-cost solution. Secondly, the occlusion problem, masks, sunglasses, hats, scarves, and hand occlusion such as face rubbing can lead to complete loss of facial features. Existing models are weak in inferring fatigue features in heavy occlusion, and lack effective "occlusion inference" mechanisms such as indirect fatigue judgement using head posture or body movements.

5 Summary and outlook

The future development direction and challenges of the fatigue driving detection system based on YOLO algorithm are as follows. Firstly, the adaptability to complex situations: for example, when entering and exiting tunnels or meeting at night, the lighting conditions such as glare of opposing headlights change drastically, and there are also drivers who are blocked by glasses, sunglasses, moustache and so on, and secondly, some body postures such as turning the head sharply, tilting the body forward and other extreme cases can easily lead to the system's misdetection and omission of the detection. Secondly, the problem of individual differences: the physiological characteristics and fatigue patterns of different age groups, genders and races are very different, making it difficult for a single model to detect them all. For example, the natural blinking frequency of the elderly is low, and some people have unusual fatigue behaviour. The last is the deep fusion of multi-modalities: combining near-infrared imaging, millimetre-wave radar, steering wheel operation signals with visual features and other technologies to construct a fusion model with stronger robustness.

References

[1] Zhu Guiwei. Research on defect detection algorithm of strip steel based on improved YOLOv7 [D]. Dalian Jiaotong University, 2025.
[2] Han T. Research on bird target detection and recognition method based on YOLOv8 [D]. Dalian Jiaotong University, 2025.
[3] B. Wang. Research on underwater small target detection algorithm based on YOLOv7 [D]. Dalian Jiaotong University, 2025.
[4] Gao Peixin, Lei Xingshan. Design and implementation of intelligent automobile fatigue driving detection system based on Android[J]. Automotive Maintenance Technician, 2025(10):9-11.
[5] Sheng Jingwen. Research on lightweight traffic sign recognition algorithm based on YOLOv11[D]. Anhui University of Architecture, 2025.
[6] Yang Song, Peng Suzhen. Train fatigue driving detection algorithm based on YOLOv8n[J]. Information Technology and Informatisation, 2025 (04):70-73.
[7] Junli Xu. Driving fatigue detection based on functional brain network and graph convolutional network[J]. Journal of Automotive Safety and Energy Conservation, 2025,16(02):226-233.
[8] DU Wei, NING Wu, MENG Linan et al. Improved YOLO-based driver fatigue detection algorithm for mining cards[J]. Modern Electronic Technology, 2025,48(07):126-131.
[9] Qichao Jiang, Chengbo Yu, Yiguo Xuan et al. YOLOv5 driver fatigue detection algorithm based on lightweight backbone[J]. Science, Technology and Engineering, 2024,24(16):6766-6774.
[10] Zhang Renbao. Research on vehicle driver monitoring safety early warning system based on deep learning [D]. Shanghai University, 2023.
[11] Fan Q. H., Jiang X. C., Yang G. J. et al. Research on test method and trend of fatigue risk evaluation of driving[J]. Journal of Taiyuan University of Technology, 2021,52(04):645-653.
[12] Zhang H. Research on fatigue driving detection and its early warning method based on driver physiological signals [D]. Jilin University, 2021.
[13] Jiao Yingying. Research on driver drowsiness detection based on EEG and EEG signals [D]. Shanghai Jiao Tong University, 2019.