Scientific Reports

https://doi.org/10.1038/s41598-026-44994-4

Article in Press

A fatigue driving detection method based on driver posture and facial state analysis

Received: 26 December 2025

Accepted: 16 March 2026

Published online: 26 March 2026

Cite this article as: Hao Y., Sun X., Liu H. *et al.* A fatigue driving detection method based on driver posture and facial state analysis. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-44994-4

Yuting Hao, Xiuqian Sun, Hao Liu & Dapeng Wang

A Fatigue Driving Detection Method Based on Driver Posture and Facial State Analysis

Yuting Hao [1,2] , Xiuqian Sun [1, 2] , Hao Liu [1, 2] and Dapeng Wang [1,2,*]

1. Department of Automotive Engineering, Hebei Vocational University of Technology and Engineering, Xingtai 054000, China;

2. Hebei Special Vehicle Modification Technology Innovation Center, Xingtai 054000, China;

* Correspondence: wangdapeng@hevute.edu.cn

Abstract: Fatigue driving is one of the primary causes of traffic accidents, posing a serious threat not only to the driver's life and property but also to the safety of surrounding vehicles and pedestrians. Therefore, accurate and efficient detection of driver fatigue is essential for preventing traffic incidents. This study proposes a novel driver fatigue detection framework that integrates AlphaPose* with a Long Short-Term Memory (LSTM) network. To enhance the performance of AlphaPose, the original human detector is replaced with an optimized YOLOv11n model, which incorporates a Hybrid Pooling Fusion Block (HPFB) to improve feature representation and meet the requirements of keypoint estimation. Furthermore, a multi-view data processing strategy based on the Driving Monitoring Dataset (DMD) is introduced to capture driver behavior from frontal, lateral, and hand views. To effectively model the temporal dynamics of driver behavior, a fatigue behavior monitoring network is designed using LSTM. Comparative experiments demonstrate that the proposed AlphaPose*-LSTM-based system achieves superior performance in fatigue detection tasks compared to existing state-of-the-art approaches.

Keywords: AlphaPose; LSTM; YOLOv11n; Fatigue Driving; Driver Behavior Analysis; Deep Learning

1. Introduction

With the rapid growth of global vehicle ownership, road traffic safety issues have become increasingly severe, posing significant threats to human life and property. According to projections by the World Health Organization (WHO), by 2030, traffic accidents are expected to become the fifth leading cause of death worldwide [1], with a mortality rate surpassing that of many major diseases. A considerable proportion of traffic accidents are attributed to human operational factors, with driver-related errors being among the most critical [2]. One of the primary causes of such errors is driver fatigue—both neurological and visual—which impairs judgment and reduces the driver's ability to respond effectively to emergencies. Driver fatigue is widely recognized as a hazardous condition that compromises not only the driver’s safety but also endangers other road users [3]. Fatigue-related driving can generally be classified into two categories: (1) fatigue induced by prolonged driving and physical exertion, and (2) fatigue resulting from sleep deprivation, sustained psychological stress, or side effects of medication [4,5]. Both types significantly degrade a driver’s situational awareness, response time, decision-making capabilities, and visual acuity, thereby elevating the risk of accidents during critical moments. Fatigue can be further divided into physical and

mental components. Physical fatigue typically manifests as muscular discomfort, eye strain, or general exhaustion after extended periods of driving. In contrast, mental fatigue arises from psychological overload, stress, and insufficient rest, leading to attentional lapses and slower cognitive processing [6]. These dual aspects of fatigue—physical and mental—contribute to a range of impairments in driving behavior, including delayed reaction time, operational errors, diminished perception, frequent blinking, blurred vision, yawning, reduced brain activity, and repetitive head nodding. Collectively, these manifestations of fatigue-induced impairment severely compromise a driver's ability to maintain safe and attentive control of the vehicle.

Driver fatigue is recognized as a critical factor contributing to road traffic accidents, with fatigue-related crashes accounting for over 35% of traffic fatalities worldwide [7]. Therefore, identifying and mitigating driver fatigue is of paramount importance for ensuring road safety. Developing efficient and accurate fatigue detection systems is essential for reducing the incidence of fatigue-induced accidents. Broadly, existing fatigue detection methods can be categorized into three main approaches: (1) physiological signal-based detection, (2) vehicle dynamics-based detection, and (3) driver behavior-based detection. Physiological signal-based methods leverage sensors to monitor internal biometric parameters such as electroencephalo- grams (EEG) [8–12], electrocardiograms (ECG) [13–16], and electromyograms (EMG) [17–18] to infer the driver's fatigue level. While these techniques offer relatively high reliability and precision, their widespread deployment is limited by high equipment costs and intrusiveness, which can negatively impact driver comfort and acceptance. Another approach is based on analyzing vehicle dynamic parameters, such as speed fluctuations [19], steering wheel angle variations [20–21], acceleration profiles [22–23], and lane deviation patterns [24–25]. These systems infer fatigue by modeling the relationship between vehicle motion characteristics and driver input over time. Although such methods can achieve reasonable accuracy and real-time performance, they are susceptible to inter-individual variability, resulting in reduced robustness and poor generalizability across different drivers or vehicles. Due to the limitations of the above methods, achieving a widely applicable and non-intrusive fatigue detection system remains challenging. In recent years, rapid advancements in computing power and deep learning have enabled the development of data-driven fatigue detection frameworks. Deep learning-based models can effectively extract fatigue-related behavioral features from large-scale data without impeding driver experience. These models demonstrate high detection accuracy, excellent real-time performance, and strong generalization capabilities. Consequently, computer vision–based fatigue detection approaches—particularly those leveraging deep neural networks—are gaining increasing attention and are considered highly promising for practical deployment in intelligent driving systems.

Driver fatigue directly compromises the accuracy and safety of vehicle operation. As drivers transition from a state of alertness to fatigue, their posture and facial expressions undergo significant changes, including increased blink frequency [26–27], frequent yawning [28–29], and noticeable head drooping [30–32]. These representative fatigue-related behavioral features can be effectively identified and quantified using deep learning–based image analysis techniques. Liu et al. [28] proposed a dynamic facial fatigue detection algorithm tailored for small-object tracking. Their method integrates Coordinate Attention mechanisms and a

specialized small-object detection head into the YOLOv7 network to enhance fine-grained detection of the eyes and mouth. Additionally, image augmentation strategies were employed to further improve detection robustness. However, this approach heavily depends on accurate eye detection, rendering it susceptible to failure under occlusion conditions. Delwar et al. [33] introduced a lightweight fatigue detection approach based on facial analysis, utilizing models such as MobileNet, VGG16, and CNN to extract features from facial and ocular regions. Although suitable for edge computing deployments, its performance significantly degrades in complex driving scenarios involving low-light environments, motion blur, or driver head tilts. Yashaswini et al. [34] developed a "Journey Tracker" system for driver fatigue detection based on deep facial analysis. The system employs EfficientNet to monitor eye state and incorporates a glare detection filter and personalized pupil-tracking module to reduce false alarms. However, the method relies heavily on eye features, making it less effective when the driver's face is partially occluded or turned away. Furthermore, the system lacks head pose estimation, limiting its ability to differentiate between closed eyes and upward gaze, or between head tilts and fatigue. Zhou et al. [35] proposed a facial fatigue detection system based on the YOLOv8 object detection model, focusing on extracting and analyzing eye and mouth regions for behavioral assessment. While the system supports real-time monitoring, it predominantly relies on static image frames and lacks temporal modeling of fatigue behaviors such as blink duration or yawning frequency. Collectively, these studies highlight the potential and current limitations in fatigue detection research. They emphasize the need for more comprehensive feature modeling and robust behavioral interpretation. In response, the present study proposes a multi-feature fatigue detection framework that simultaneously considers ocular state, mouth movement, head posture, and hand position. This multi-modal approach aims to improve inference accuracy and model robustness. By integrating diverse fatigue-related indicators, the system can more reliably detect fatigue states and provide timely alerts to encourage rest, thereby reducing the risk of fatigue-induced traffic accidents.

To enhance the accuracy of driver fatigue detection systems, this study adopts convolutional neural network (CNN) techniques to achieve real-time and precise recognition of driver fatigue states. The system primarily analyzes driver-specific fatigue indicators such as ocular, oral, cranial, and manual features. In this work, we design and implement an integrated fatigue detection framework that combines AlphaPose [36] with a Long Short-Term Memory (LSTM) network [37]. To improve the performance of the pose estimation module in terms of both accuracy and efficiency, the original object detector within the AlphaPose framework is replaced with a lightweight and high-performance YOLOv11n [38]. Additionally, a Hybrid Pooling Fusion Block (HPFB) is introduced into the neck structure of YOLOv11n to enhance context representation and refine spatial detail sensitivity, thereby better supporting AlphaPose's keypoint estimation requirements. Furthermore, this study is the first to construct and utilize a tri-view behavior modeling strategy based on the Driving Monitoring Dataset (DMD) [39]. This strategy integrates front-view, side-view, and hand-view perspectives to enable comprehensive analysis of driver behavior patterns. To effectively capture the temporal dynamics of driving behaviors, an LSTM-based behavior recognition model is proposed. The LSTM module extracts sequential features from keypoint data, which are then passed through a fully connected layer with a ReLU activation function to enhance the

representational capacity. Subsequently, batch normalization (BN) is applied to stabilize training, and a dropout layer is employed to mitigate overfitting. A second fully connected layer performs feature dimensionality reduction, and the final classification output is generated through a Softmax layer, yielding the probability distribution over multiple fatigue levels. Experimental evaluations demonstrate that the proposed AlphaPose*-LSTM integrated framework consistently outperforms existing state-of-the-art methods across various driver fatigue detection benchmarks. It exhibits superior temporal modeling capacity and behavioral recognition accuracy. The system is capable of continuously monitoring driver fatigue status during vehicle operation and issuing early warnings when potential fatigue behaviors are detected, thereby contributing to improved road safety.

2. Materials and Methods

In monotonous driving scenarios, drivers are particularly prone to fatigue, which can significantly increase the risk of traffic accidents. During prolonged driving sessions, drivers are often unaware that they have entered a fatigued state. Prior to the occurrence of fatigue-related accidents, drivers typically experience a transitional phase characterized by a struggle against fatigue. An effective fatigue detection and alerting system can provide timely warnings, thereby reducing the likelihood of accidents. Fatigue-induced crashes are more likely to result in severe injuries or fatalities and can lead to substantial economic losses. According to a survey conducted by Maycock in the United Kingdom, 29% of 4,600 respondents admitted to having experienced near-sleep episodes while driving. Sleep laboratory experiments further revealed that individuals who fall asleep for one to two minutes often deny having done so, whereas approximately 50% of participants acknowledged falling asleep after two to four minutes of sleep. Subjective sleep onset is typically a gradual process, during which the driver's perception and awareness become impaired. Fatigue compromises a driver’s ability to assess the driving environment and increases the probability of operational errors. Data from the National Highway Traffic Safety Administration (NHTSA) in the United States indicates that approximately 25% of annual traffic accidents are related to driver fatigue. Therefore, early prediction of the driver’s state and timely intervention upon detecting signs of fatigue are crucial to mitigating the risk of traffic accidents.

Driver behavior prediction poses a significant challenge, as it requires the rapid identification of potentially hazardous events within a limited time window. During driving, drivers exhibit a wide range of operational behaviors, many of which must be recognized to determine whether they are in a normal or fatigued state. Effectively detecting these behavioral states necessitates the fusion of multi-modal sensor data. In this study, we utilize tri-view driver monitoring data, which includes facial feature detection, body posture estimation, and hand state analysis, as illustrated in Fig. 1. The facial region provides critical indicators of driver fatigue. Among facial cues, the eyes have been extensively studied — for instance, datasets such as DR(eye)VE [41], DADA [42], BDD-A [43], DrivFace [44], and AttenD [45] explore the relationship between eye behavior and safe driving from various perspectives. Fatigue is a major factor impairing driving safety. It can be detected by monitoring eye and mouth activity, including parameters such as blink duration, blink frequency, and PERCLOS [46], a common metric for fatigue estimation. Head posture is also a reliable indicator of fatigue-related behavior. Studies like DriveAHead [47] and DD-Pose [48] leverage yaw, pitch, and roll angles to assess

driver state. In addition, body posture, particularly observed from side-view cameras, plays a crucial role in evaluating driver state. The StateFarm dataset [49] contains nine categories of distracted driving behaviors. While still images often fail to capture temporal information, video sequences provide valuable inputs for spatiotemporal modeling, improving classification accuracy when temporal dynamics are incorporated. Moreover, the driver's hand position and movement are strongly correlated with attention and fatigue. Datasets such as CVRRHANDS 3D [50] and VIVA-Hands [51] annotate hand positions with bounding boxes and enable hand detection through multi-scale convolutional networks. In our experiments, we incorporate Fault Tree Analysis (FTA) to analyze potential anomalies in facial, body, and hand features under fatigued driving conditions, as illustrated in Fig. 2.

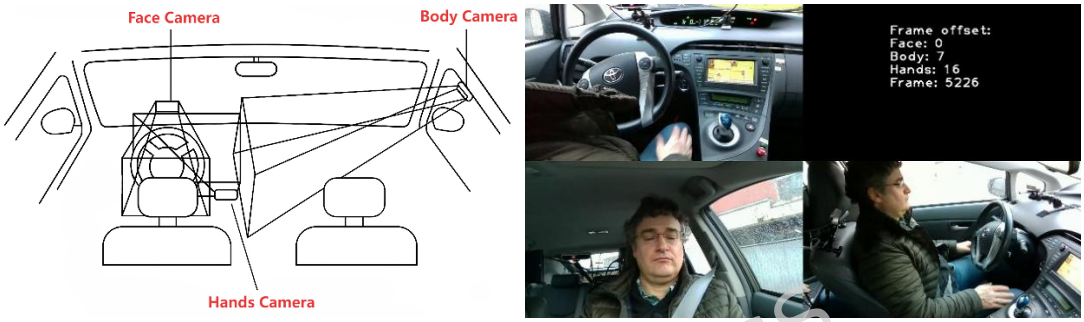


Figure 1. Data Acquisition Method of the Driving Monitoring Dataset (DMD).

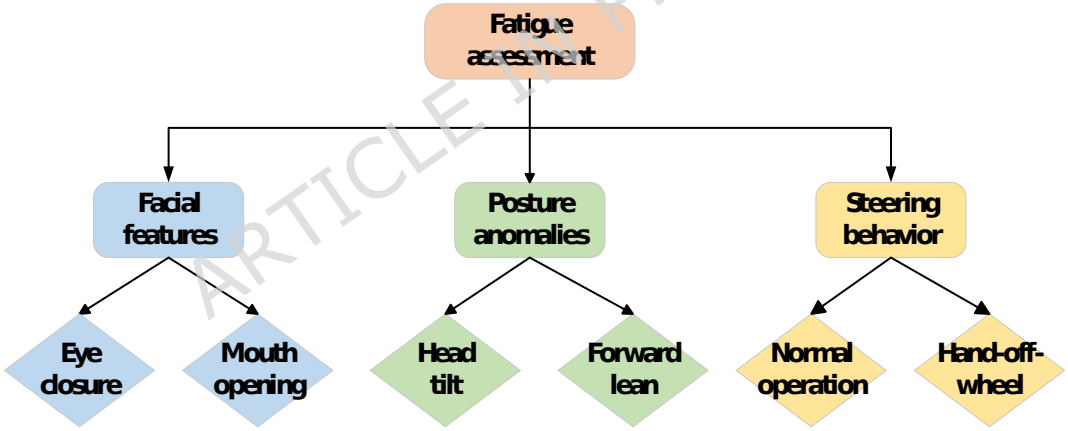


Figure 2. Potential Abnormal Behaviors Exhibited by Drivers in a Fatigued State.

As illustrated in Figure 2, various abnormal behaviors can be categorized according to different levels of driver fatigue. Experimental observations indicate that fatigue often leads to partial or complete eyelid closure, with a noticeable reduction in eye opening. Additionally, fatigued drivers tend to exhibit increased mouth opening beyond normal levels. Fatigue-induced postural abnormalities are also evident, such as forward head tilting and lateral head inclination (either left or right). It is worth noting that due to proper seatbelt use during experiments, torso tilting was not considered a significant factor in this study. Moreover, fatigue impacts hand behavior on the steering wheel—when fatigue becomes severe, the driver may release one or both hands from the wheel. Based on these findings, a fatigue detection framework was developed using Fault Tree Analysis (FTA), as illustrated in Figure 3.

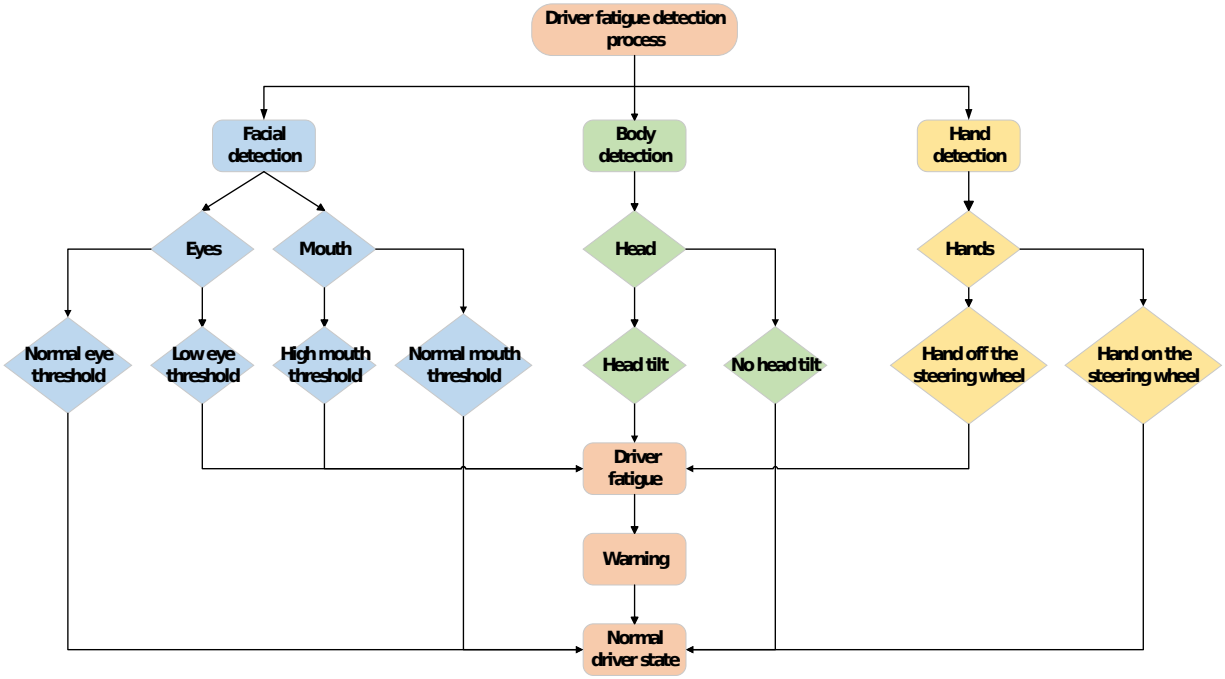


Figure 3. Hierarchical Structure of Driver Fatigue Detection Based on Behavioral Indicators.

The driver fatigue detection process is divided into three main components: facial status analysis, head posture evaluation, and hand position monitoring. Facial status analysis primarily involves the detection of eye and mouth states. When the eye-opening ratio falls below a predefined minimum threshold, the driver is considered to be in a fatigued state. Conversely, if the eye-opening ratio exceeds this threshold, the driver is deemed to be in a normal driving condition. Similarly, when the mouth-opening ratio exceeds a designated maximum threshold, it indicates a potential fatigue state; otherwise, it corresponds to a normal state. Head posture evaluation focuses on detecting head inclination. A non-tilted head posture is indicative of normal driving behavior, whereas noticeable head tilting suggests the onset of fatigue. Hand position monitoring assesses whether the hands are appropriately placed on the steering wheel. When both hands are operating the wheel in a standard manner, the driver is considered alert. In contrast, if either or both hands disengage from the steering wheel, this is interpreted as a sign of fatigue. Maintaining consistent hand placement on the wheel is associated with normal driving status.

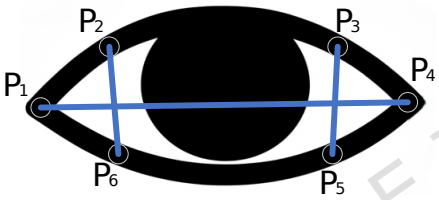
Fatigue detection based on facial, body, and hand behavior during driving presents a significant challenge, as it requires accurately determining the driver's fatigue state under limited contextual information. Normal driving behavior can be decomposed into a spatiotemporal process, wherein the driver continuously interacts with various environmental variables over time and space. Key components of this interaction include the coordinated movement of the driver's arms, eyes, and torso. The integration of spatial-temporal structure and sequential behavior modeling is essential for effectively identifying fatigue-related patterns in driver behavior.

As driver fatigue accumulates progressively, a range of characteristic behaviors begin to emerge, including yawning, frequent blinking, and the downward movement of the head and hands. These fatigue-related indicators

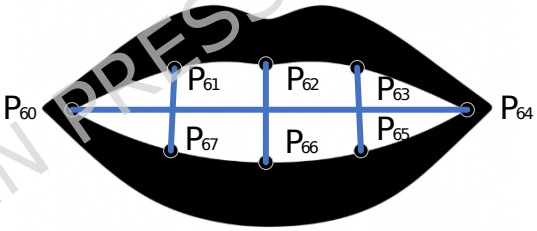
cannot be reliably identified from a single frame of visual data, as temporally isolated observations lack contextual continuity. Effective recognition of fatigue onset and progression requires temporal modeling to capture behavioral trends over time. For instance, in single-frame analysis, momentary actions such as glancing down at the dashboard or navigation system may be incorrectly interpreted as fatigue. Similarly, natural blinking may be misclassified as a fatigue symptom. Continuous multi-frame monitoring mitigates such misjudgments by capturing the temporal dynamics of driver behavior.

Table 1. outlines the classification hierarchy of driver fatigue monitoring systems.

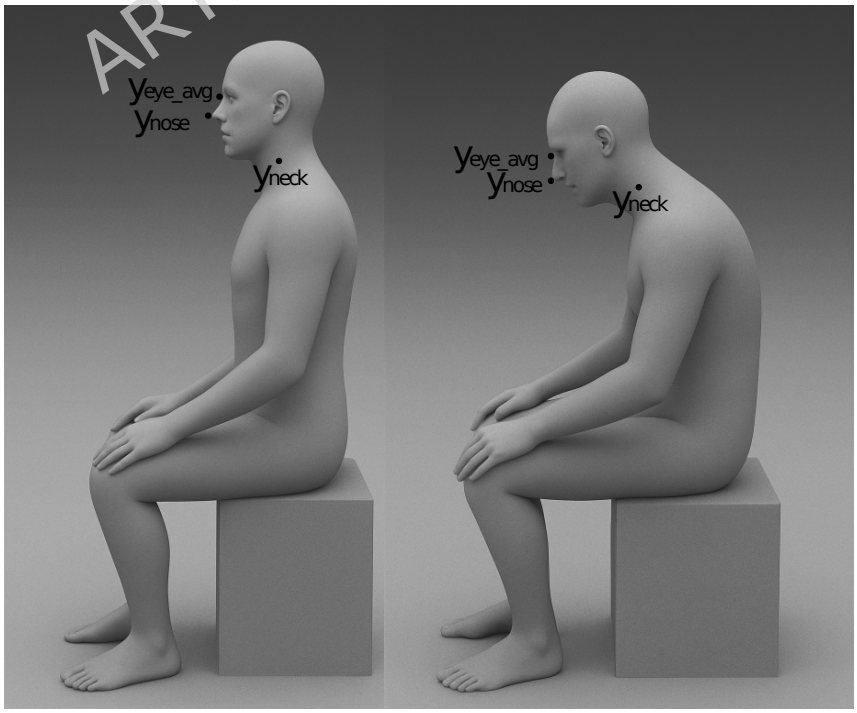
Temporal Frame	Eye	Mouth	Head	Hand	Result
All Five Regions in Normal State for 20 Consecutive Frames	Eyes Open	Mouth Normally	Normal Posture	Proper Hand Grip	Normal
Any Fatigue Indicator Sustained Over 20 Consecutive Frames	Eye Closure	Mouth Wide Open	Head Tilt /Forward Lean	Hand Off Steering Wheel	Fatigued
Multiple Fatigue Indicators Sustained Over 20 Consecutive Frames					Severely Fatigued



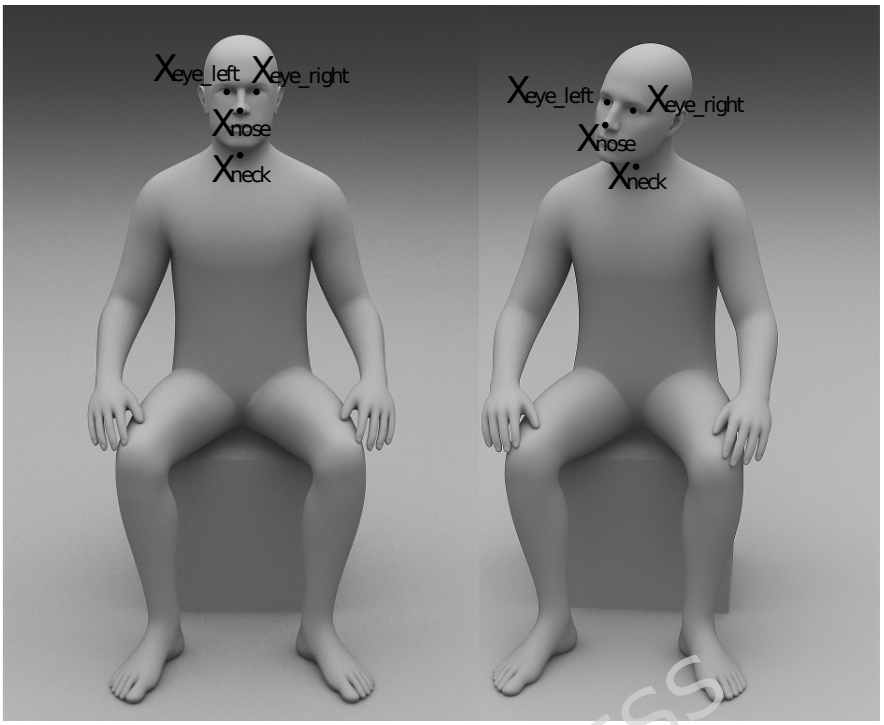
(a) Eye keypoint information.



(b) Mouth keypoint information.



(c) Data obtained from the driver side monitoring system.



(d) Data obtained from the driver front-view monitoring system.

Figure 4. (a) Distribution of key landmark points around the eyes; (b) distribution of key landmark points around the mouth; (c) information obtained from the driver's side-view monitoring system, including the average eye coordinates, nasal tip coordinate, and neck coordinate, where the y-axis follows the conventional image coordinate system, increasing from top to bottom; (d) information obtained from the driver's front-view monitoring system, including eye coordinates, nasal tip coordinate, and neck coordinate, where the x-axis follows the conventional image coordinate system, increasing from left to right.

Eye opening and closure are determined based on the Eye Aspect Ratio (EAR), which is mathematically defined by Equation (1):

$$EAR = \frac{\|P_2 - P_6\| + \|P_3 - P_5\|}{2 \cdot \|P_1 - P_4\|} \quad (1)$$

In Equation (1), P_1 to P_6 represent six key facial landmarks around the eye, as illustrated in Figure 4 (a).The Eye Aspect Ratio (EAR) value is inversely correlated with eye openness—i.e., the smaller the EAR, the higher the likelihood that the eyes are closed. When the EAR exceeds 0.20, it indicates that the eyes are open and the driver is in a normal, alert state. If the EAR falls within the range of 0.18 to 0.20 for 20 consecutive frames or more, it suggests a partially closed state of the eyes, indicating the onset of driver fatigue. When the EAR lies between 0.12 and 0.18 for at least 20 consecutive frames, the eyes are considered half-closed, corresponding to a moderate level of fatigue where the upper and lower eyelids are approaching each other. If the EAR drops below 0.12 for 20 or more consecutive frames, it signifies tightly closed eyes, implying that the driver is likely experiencing significant fatigue or possibly falling asleep. Since this study is focused on fatigue warning, an EAR value below 0.20 sustained over 20 consecutive frames is

interpreted as an eye-closure fatigue state, indicating that the driver may require an alert or intervention.

The degree of mouth opening is quantified using the Mouth Aspect Ratio (MAR), as formally defined in Equation (2):

$$MAR = \frac{\|P_{61} - P_{67}\| + \|P_{62} - P_{66}\| + \|P_{63} - P_{65}\|}{3 \cdot \|P_{60} - P_{64}\|} \quad (2)$$

In Equation (2), the points P_{60} to P_{67} correspond to eight key facial landmarks located on the lips, as illustrated in Figure 4 (b).The Mouth Aspect Ratio (MAR) increases with the likelihood of yawning—i.e., a larger MAR indicates a higher probability of a yawning event. When the MAR is below 0.20, it indicates that the mouth is in a normally closed state, suggesting the driver is not speaking or yawning. If the MAR falls within the range of 0.20 to 0.35 for 20 or more consecutive frames, it denotes a slightly open mouth, typically associated with mild speech behavior. In this study, a MAR value between 0.35 and 0.50 sustained over 20 or more consecutive frames is defined as a clearly open mouth, indicating the driver may be yawning or inhaling—both potential signs of fatigue. When the MAR exceeds 0.50 for 20 or more consecutive frames, it suggests a significantly open mouth, reflecting a pronounced yawn or a high degree of fatigue.

Fatigue detection based on head posture mainly relies on two types of head tilt: horizontal (lateral) tilt and forward/backward tilt. A lateral head tilt between 0° and 5° suggests a normal posture, indicating that the driver is maintaining a forward-facing head position. If the tilt falls between 5° and 10° for 20 or more consecutive frames, it is regarded as a slight tilt, possibly indicating a mildly fatigued state with minor head leaning. In this study, a lateral tilt between 10° and 20° sustained over 20 frames is considered a noticeable tilt, signifying distraction or declining alertness. A lateral tilt greater than 20° maintained for 20 or more consecutive frames is interpreted as a severe tilt, strongly suggesting fatigue or fatigue, such as nodding off while driving.

The forward tilt of the driver's head is quantified through the normalized characteristic parameter Δy, as mathematically defined in Equations (3) through (5):

$$\Delta y = y_{nose} - y_{eye\_avg} \quad (3)$$

$$face\_height = y_{neck} - y_{eye\_avg} \quad (4)$$

$$\Delta y_{norm} = \frac{\Delta y}{face\_height} \quad (5)$$

In Equation (3), y_{nose} denotes the y-coordinate of the nasal tip while y_{neck} represents the y-coordinate of the neck position. In Equation (4), y_{eye_avg} indicates the average y-coordinate value of both eyes. The normalized vertical displacement, Δy_{norm}, is used to evaluate the vertical posture of the driver's head. The information obtained from the driver-side monitoring system is shown in Figure 4 (c), When $\Delta y_{norm} < 0.15$, it indicates that the driver is keeping their head up or looking straight ahead. When Δy_{norm} is between 0.15 and 0.25 for more than 20 consecutive frames, the driver is considered to have a slight downward head tilt, suggesting early signs of fatigue. In this study, a value of Δy_{norm} between 0.25 and 0.4 sustained for over 20 consecutive frames indicates a clear downward head posture, associated with fatigue or the onset of microsleep. When $\Delta y_{norm} \geq 0.4$ for more than 20 frames, the driver’s head is considered to be significantly drooping, which is a strong indicator of fatigue or imminent sleep.

In the experiment, the inclination of the driver's head was determined by calculating the tilt angle of the line connecting both eyes. Under normal conditions, the head remains upright, and this angle is close to zero. A significant deviation in either the positive or negative direction indicates that the head is tilted to the left or right. In addition, the angle between the line from the nasal tip to the neck and the vertical direction was used to further assess head tilt. To eliminate the influence of inter-individual differences and varying camera distances, a normalization process was applied, as defined in Equation (6) :

$$\Delta\square_{norm} = \frac{|x_{nose} - x_{neck}|}{d_{eye}} \quad (6)$$

In Equation (6), x_{nose} represents the x-coordinate of the nasal tip, x_{neck} denotes the x-coordinate of the neck position, and d_{eye} refers to the interocular distance. Based on information obtained from the driver's frontal monitoring system, as illustrated in Figure 4 (d). when $\Delta\square_{norm} < 0.05$, the driver's head is considered to be in a normal forward-facing posture. When $\Delta\square_{norm}$ falls within the range of 0.05 to 0.15 and persists for more than 20 consecutive frames, it indicates a slight lateral head tilt, which suggests early signs of driver fatigue. If $\Delta\square_{norm}$ remains between 0.15 and 0.3 for over 20 frames, the head tilt is more pronounced, reflecting drowsiness or inattentiveness. When $\Delta\square_{norm} > 0.3$ and this condition continues for more than 20 frames, the driver's head is considered to be significantly tilted to the left or right, which serves as a strong indicator of fatigue or imminent sleep. As for the driver's hand position: When both the left and right hands are firmly holding the steering wheel, the hand status is considered normal. If the driver is driving with one hand relaxed or temporarily releasing one hand from the steering wheel for more than 20 consecutive frames, it is classified as relaxed or single-handed driving. If the distance between either wrist and the center of the steering wheel exceeds the wheel's radius for more than 20 consecutive frames, this is interpreted as a sign of severe fatigue. The driver behavior analysis employs the AlphaPose algorithm, which implements detection of 136 human keypoints using a top-down approach that is particularly suitable for driver fatigue detection during vehicle operation. AlphaPose combines a Single Person Pose Estimator (SPPE) network with multi-stage regression refinement to achieve high-precision detection of Halpe-136 full-body keypoints, including facial, head, torso and hand components. The Halpe-136 keypoint dataset consists of 26 body keypoints (indices 0-25), 68 facial keypoints (indices 26-93), and 42 hand keypoints (indices 94-135), totaling 136 keypoints. For driver fatigue detection applications, keypoints located below the left and right hips are excluded from analysis, and duplicate keypoints maintain identical coordinate values and confidence scores in the output. Consequently, the actual number of utilized keypoints is reduced to 119.

The AlphaPose algorithm requires an input image with human detection results provided by a detector for driver pose estimation. The system then obtains the driver's full-body posture through tracking networks and pose estimation. The acquisition of the driver's behavioral feature information constitutes the most critical component in behavior recognition. To enhance driver behavior prediction, Long Short-Term Memory (LSTM) networks are incorporated to analyze temporal sequences in video data for predicting behavioral characteristics. The 119 keypoints from the AlphaPose algorithm model are converted into vector representations for subsequent processing.

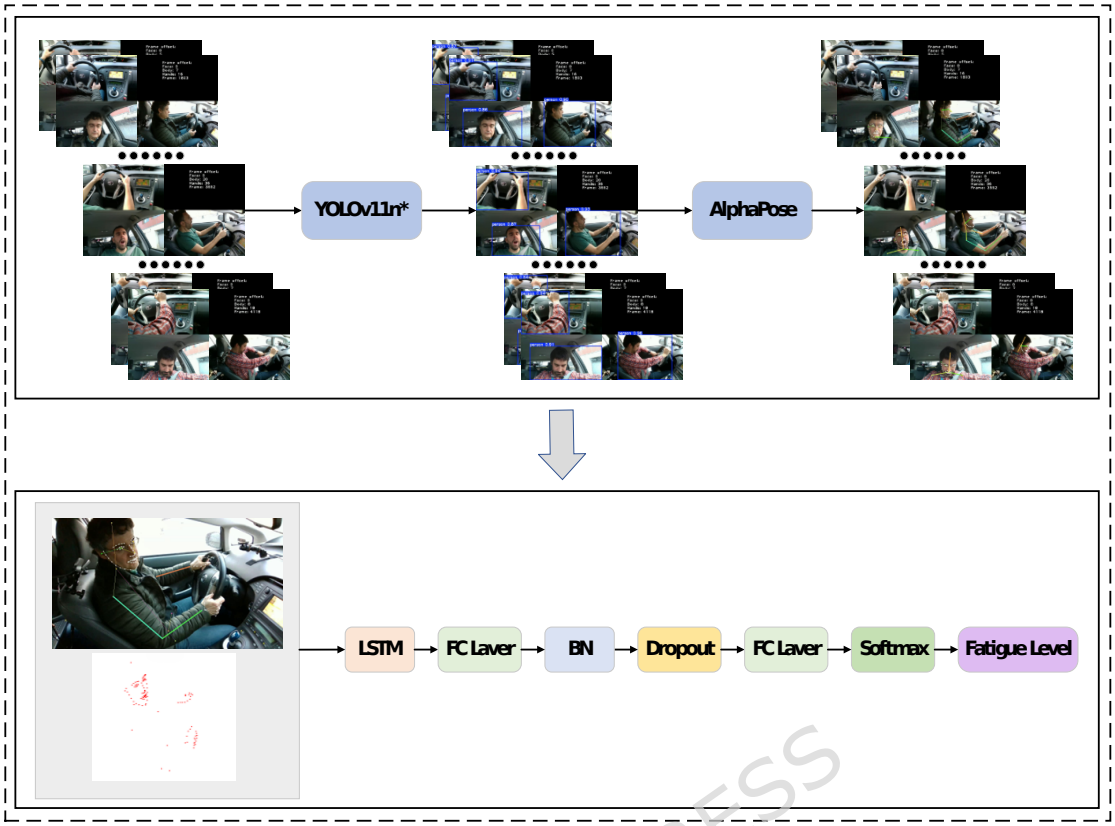


Figure 5. Driver Fatigue Behavior Monitoring Network.

The objective of a driver fatigue detection system is to accurately and efficiently identify the driver's fatigue state during vehicle operation, thereby enabling timely intervention or corrective actions. As illustrated in Figure 5, the proposed fatigue detection pipeline integrates a high-speed human detection module based on the YOLOv11n object detection algorithm, known for its superior real-time performance. YOLOv11n facilitates the AlphaPose framework by rapidly and precisely localizing human regions within the input image, providing accurate cropping for subsequent behavior analysis. Precise and real-time detection of driver regions is crucial for ensuring reliable behavior inference. During experimentation, it was observed that YOLOv11n occasionally failed to correctly detect the driver in blurred images. To address this issue and enhance detection robustness without compromising efficiency, we introduced architectural optimizations to the original YOLOv11n model, as depicted in Figure 6.

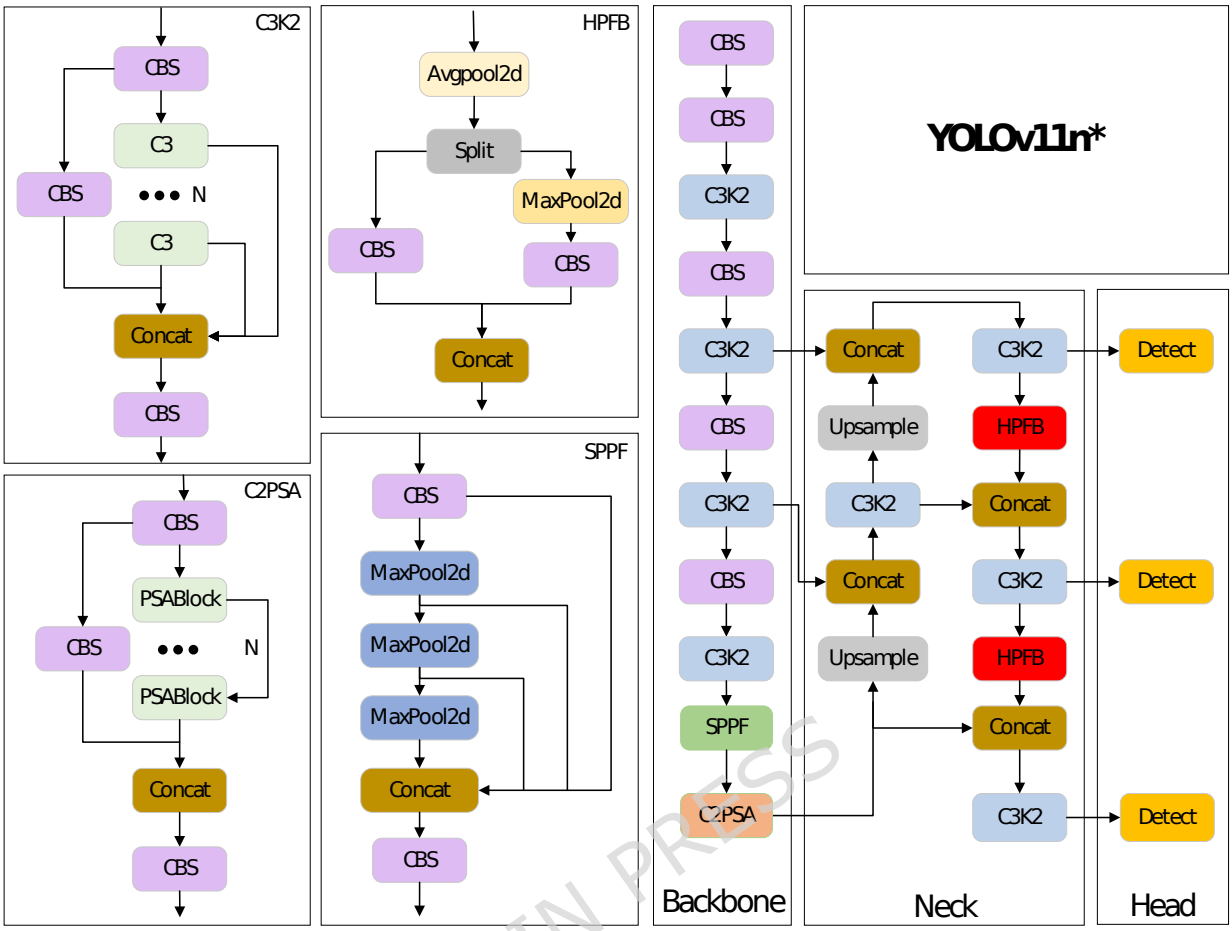


Figure 6. Network Architecture of YOLOv11n*

To enhance the model's adaptability for driver detection tasks, the training process utilized the DMD (Driving Monitoring Dataset), from which a subset of 10,000 images was selected for model training. In order to improve the compatibility of the YOLOv11n architecture with the characteristics of the DMD dataset, this study introduces an optimization strategy based on a novel Hybrid Pooling Fusion Block (HPFB). The HPFB module first employs an average pooling operation (AvgPool2d) to enlarge the global receptive field, thereby enhancing the contextual modeling capability. Subsequently, a max pooling operation (MaxPool2d) is applied to reinforce edge and detail representations, which contributes to the improved detection of low-resolution targets. This mechanism strengthens the expressiveness of subtle fatigue-related features such as eye closure, head drooping, and mouth opening. The formulation of the proposed HPFB mechanism is shown as follows:

$$X_{avg} = \text{AvgPool2D}(X) \quad (7)$$

$$X_{max} = \text{MaxPool2D}(X) \quad (8)$$

$$F_{avg} = \text{CBS}(X_{avg}) \quad (9)$$

$$F_{max} = \text{CBS}(X_{max}) \quad (10)$$

$$F = \text{Concat}(F_{avg}, F_{max}) \quad (11)$$

Equation (7) represents the application of average pooling to the input feature map X, while Equation (8) denotes the application of max pooling to the same feature map. In Equation (9), the average-pooled feature map is further processed by a CBS module composed of Convolution, Batch Normalization, and the SiLU

activation function. Similarly, Equation (10) describes the application of the CBS module to the max-pooled feature map. Finally, Equation (11) illustrates the concatenation of the features obtained from both average and max pooling branches, thereby achieving a fusion that simultaneously enhances the network's capacity for global context modeling and edge feature representation. This architectural optimization is designed to improve the model's effectiveness in fatigue-related cues such as eye closure, mouth opening, and head drooping.

YOLOv11n* is employed to detect the driver within video frames, after which behavior analysis is conducted using the AlphaPose algorithm based on the Halpe-136 skeletal model for the driver identified by YOLOv11n*. The pose estimation network utilizes a ResNet-50 backbone, and the input image resolution is set to 256 × 192. The overall structure of the backbone network is based on ResNet-50, as illustrated in Equation (12).

$$Y = F(x,\{W_i\}) + x \tag{12}$$

Feature representations are extracted through the network's convolutional layers. In Equation (12), the parameter set $\{W_i\}$ denotes the weights of the i-th layer, while x represents the input feature map of the residual block. The residual connection structure enhances the stability of gradient propagation.

The Heatmap Head generates a 2D Gaussian heatmap for each keypoint, as described in Equation (13).

$$H_k(x,y) = \exp\left(-\frac{(x-x_k)^2+(y-y_k)^2}{2\sigma^2}\right) \tag{13}$$

In Equation (13), H_k denotes the heatmap corresponding to the k-th keypoint, with a spatial resolution matching the network's output size. The value $H_k(x,y)$ represents the probability that the pixel at location (x,y) corresponds to keypoint k. The parameter σ refers to the standard deviation of the Gaussian function, which controls the spatial spread of the keypoint heatmap. The coordinates (x_k,y_k) indicate the ground-truth position of the k-th keypoint within the heatmap.

The coordinate regression module identifies the keypoint location by selecting the pixel with the maximum response in the heatmap, as defined in Equation (14).

$$(x_k^{pred},y_k^{pred}) = \arg\max_{x,y} H_K(x,y) \tag{14}$$

The variables $\left(x_k^{pred},y_k^{pred}\right)$ in Equation (14) represent the predicted coordinates of the k-th keypoint. These are obtained by identifying the location of the maximum response value in the corresponding heatmap, with the position of this peak indicating the most probable location of the keypoint.

Confidence Estimation: The confidence score for a keypoint is defined as the maximum value within its corresponding heatmap, as shown in Equation (15).

$$c_k = \max_{x,y} H_K(x,y) \tag{15}$$

In Equation (15), the confidence score c_k for keypoint k is obtained by taking the maximum value from its corresponding heatmap.

The Long Short-Term Memory (LSTM) network is employed to extract temporal features from the sequence of keypoints, enabling the analysis of fatigue-related behavioral patterns over time. Following the LSTM layer, a fully connected layer with a ReLU activation function is applied to project the output into a new feature space for enhanced nonlinear feature representation. A Batch Normalization (BN) layer is subsequently introduced to normalize the outputs, accelerate convergence, mitigate overfitting, and stabilize the training process. To further reduce the risk of

overfitting, a Dropout layer is added after BN, which randomly deactivates a portion of neurons during training to improve model generalization.

A second fully connected layer is then used to further refine the feature space, preparing the representation for the final classification task. A Softmax layer at the output converts the final feature representation into a probability distribution over fatigue levels, thereby estimating the likelihood that the driver is in a particular fatigue state. The input sequence of keypoints and behavioral features is represented by Equation (16)：

$$X = [X_1, X_2, \ldots, X_T] \tag{16}$$

In Equation (16), T denotes the total number of time steps (i.e., the number of consecutive frames, such as 30 frames), and $X_T \in R^d$ represents the feature vector at the T-th frame, composed of all keypoint coordinates and their corresponding confidence scores.

The LSTM module receives the time series of keypoints and behavioral features and extracts the temporal dependencies, as shown in Equation (17)：

$$H = \text{LSTM}(X) = [h_1, h_2, \ldots, h_T] \tag{17}$$

In Equation (17), X denotes the input temporal feature sequence, and h_T represents the LSTM hidden state at the T-th frame.

A fully connected layer is then applied to perform a linear transformation for feature mapping, as shown in Equation (18)：

$$Z_1 = W_1 h_{\text{last}} + b_1 \tag{18}$$

In Equation (18), W_1 represents the weight matrix of the fully connected layer that projects the hidden state to a new feature space, b_1 denotes the bias vector of the fully connected layer, and Z_1 corresponds to the output feature vector of this layer.

The batch normalization layer enhances training stability through normalization followed by weighted restoration, as mathematically formulated in Equation (19):

$$\hat{Z}_1 = \text{BN}(Z_1) = \gamma \cdot \frac{Z_1 - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \tag{19}$$

In Equation (19), Z_1 denotes the output of the preceding fully connected layer, μ and σ^2 represent the mean and variance of the current batch, respectively, and ϵ is a small constant added to prevent division by zero. γ and β are learnable scale and shift parameters, respectively. The normalized and scaled output $\hat{Z}_1$helps maintain numerical stability, accelerates convergence, and enhances the model's generalization ability.

The Dropout layer, which is used to mitigate overfitting, randomly sets a portion of activations to zero with a predefined probability, as expressed in Equation (20):

$$Z_2 = \text{Dropout}(\hat{Z}_1) \tag{20}$$

In Equation (20), $\hat{Z}_1$ represents the feature output after batch normalization. The Dropout operation randomly sets a portion of the elements in $\hat{Z}_1$ to zero based on a predefined probability, effectively reducing overfitting by preventing co-adaptation of neurons. The resulting feature vector after dropout is denoted as Z_2, which serves as the input to the subsequent fully connected layer.

The final classification is performed by a fully connected (FC) layer followed by a Softmax activation function, as defined in Equations (21) and (22), producing a probability distribution over the predicted fatigue levels.

$$Z_3 = W_2 Z_2 + b_2 \tag{21}$$

$$y = \text{Softmax}(Z_3) \tag{22}$$

In Equation (21), W_2 denotes the weight matrix of the second fully connected (FC) layer, Z_2 represents the feature vector obtained after the Dropout operation, and b_2 is the corresponding bias vector. The output Z_3 constitutes the final set of logits prior to classification. In Equation (22), Z_3 refers to the output of the last FC layer, and $\mathrm{Softmax}(Z_3)$ converts each logit into a class probability, yielding □, which represents the final predicted probability distribution across fatigue levels.

3. Experimental results and analysis

The experimental platform used in this study consists of an Intel® Xeon® Bronze 3106 CPU @ 1.70GHz × 8 and an NVIDIA GeForce RTX 3090 GPU. The operating system is Ubuntu 20.04, and the development environment includes PyTorch 1.12.1, Python 3.8.20, and CUDA 11.3. To evaluate the performance of the optimized model, a comprehensive set of metrics is adopted, including Precision, Recall, mean Average Precision (mAP), Frames Per Second (FPS), Accuracy, and F1-score. We utilized a publicly available dataset named "DMD" [39], which contains driver fatigue video recordings. The video data were sampled at intervals of 5 frames and subsequently annotated in TXT format. A total of 10,000 images were selected for the training set, 1,000 for the test set, and 1,000 for the validation set. To ensure consistency in both ablation experiments and comparative evaluations, a unified set of hyperparameters was employed during model training. Specifically, the batch size was set to 16, and the maximum number of epochs was 300. In addition, the loss function was kept identical to that of the baseline model. For the LSTM-based fatigue detection model, the training configuration includes the Adam optimizer, an initial learning rate of 1e-3, 100 training epochs, and an L2 regularization coefficient of 1e-4. The YOLOv11n-related experimental parameters are tuned based on the original network configuration as described in [38].

Precision refers to the proportion of correctly predicted positive samples among all samples predicted as positive, as shown in Equation (23):

$$\mathrm{P\square ecision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{23}$$

In Equation (23)□In this equation, TP (True Positive) denotes the number of samples correctly predicted as positive, while FP (False Positive) denotes the number of samples incorrectly predicted as positive.

Recall refers to the proportion of actual positive instances that are correctly identified by the model, as defined in Equation (24):

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{24}$$

In Equation (24)□In this equation, FN (False Negative) indicates the number of actual positive samples that were missed by the model.

Mean Average Precision (mAP) represents the average area under the Precision-Recall (P-R) curve across all object classes, serving as a comprehensive metric for evaluating the detection performance of a model, as shown in Equation (25):

$$\mathrm{mAP} = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} \mathrm{AP_i} \tag{25}$$

In Equation (25)□In this equation, mAP (mean Average Precision) is computed as the average of AP (Average Precision) values across all object classes, where AP measures the area under the P-R curve for a specific class.

Frames Per Second (FPS) indicates the number of image frames that the model can process per second, as defined in Equation (26):

$$\mathrm{FPS} = \frac{\text{Total number of frames}}{\text{Total consumption time}} \tag{26}$$

Accuracy measures the proportion of correctly classified samples out of the total number of samples, as defined in Equation (27):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (27)$$

In Equation (27), In this equation, TN (True Negative) refers to the number of correctly predicted negative samples.

F1-score is the harmonic mean of precision and recall, providing a balanced evaluation when there is an uneven class distribution. It is defined in Equation (28):

$$F1 = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (28)$$

In Equation (28), If there is a large discrepancy between Precision and Recall, the F1-score tends to be low. The F1-score ranges from 0 (worst) to 1 (best).

Parameter (M) represents the total number of all learnable weights and biases in the model. FLOPs (G) represents the number of floating-point operations required for a single forward inference of the model.

To further validate the effectiveness of the HPFB module in optimizing YOLOv11n, we conducted ablation experiments. The experimental results are shown in Table 2.

Table 2. Ablation Study on the Optimization of YOLOv11n.

Model	Precision/%	Recall/%	mAP/%	FPS	Parameter (M)	FLOPs (G)
YOLOv11n	83.9	81.6	82.5	80.6	2.6	6.5
YOLOv11n+HPFB	84.5	81.9	83.7	75.6	2.66	6.6
YOLOv11n+2HPFB	85.8	82.7	85.1	71.7	2.7	6.7

YOLOv11n in Table 2 denotes the baseline detection model. The term YOLOv11n+HPFB refers to the incorporation of one HPFB module between the second and third detection heads of YOLOv11n, as illustrated between the second and third detection heads in Figure 6. YOLOv11n+2HPFB indicates the addition of two HPFB modules—one between each pair of detection heads in the three-head structure of YOLOv11n, as shown by the two HPFB insertions in Figure 6. As shown in Table 2, with the integration of HPFB modules ahead of the detection heads in YOLOv11n, the FPS begins to decline, while the number of Parameters and FLOPs experiences a slight increase. However, the introduction of the HPFB modules effectively enhances the model's Precision, Recall, and mAP metrics. The real-time performance can fully meet the experimental requirements of the detection model. Therefore, YOLOv11n+2HPFB is selected as our optimized network model, denoted as YOLOv11n*.

To validate the effectiveness of the optimized model YOLOv11n*, a comparative analysis was conducted against YOLOv8n, YOLOv10n, and the baseline YOLOv11n, as presented in Table 3. The baseline model selected for this study was YOLOv11n, which demonstrates strong real-time performance along with relatively high detection accuracy. However, during experimentation, it was observed that the model exhibited issues such as missed detections and duplicate detections in frames of lower quality. Consequently, further optimization was performed, resulting in the YOLOv11n* model shown in Table 3. Although YOLOv11n* exhibits a reduction in detection speed compared to YOLOv8n [52],

YOLOv10n [53], and the original YOLOv11n, along with increases in both model parameters and FLOPs, it achieves a significant improvement in detection precision and mean average precision (mAP), demonstrating the effectiveness of the proposed enhancements.

Table 3. Performance Comparison of Network Models.

Model	Precision/%	Recall/ %	mAP/%	FPS	Parameter (M)	FLOPs (G)
YOLOv8n	80.1	79.8	79.3	102.1	3.2	8.7
YOLOv10n	81.1	80.9	81.7	72.1	8.1	24.8
YOLOv11n	83.9	81.6	82.5	80.6	2.6	6.5
YOLOv11n*	85.8	82.7	85.1	71.7	2.7	6.7

A visual performance comparison was performed to verify the improvements of YOLOv11n*, with the original and optimized model outputs shown in Figures 7 and 8, respectively.

Figure 7. Visual Detection Results of YOLOv11n for Driver Monitoring.

Figure 8. Visualized Detection Results of YOLOv11n* for Driver Monitoring.

As evidenced by the driver detection visualization results of YOLOv11n in Figure 7, the original YOLOv11n object detection model exhibited three critical limitations: (1) overflow of hand keypoints beyond detection boundaries, (2) duplicate detections, and (3) missed detections - all of which failed to meet the requirements for accurate driver fatigue monitoring. The optimized YOLOv11n* model, demonstrated in Figure 8's visualization results, successfully addresses these deficiencies and achieves compliance with fatigue detection specifications.

YOLOv11n* provides AlphaPose with driver detection bounding boxes that exhibit superior real-time performance and detection accuracy, thereby enabling more effective keypoint estimation within the bounded regions. The precisely estimated keypoints are subsequently extracted and fed into the LSTM-based driver fatigue detection network. This study conducts comprehensive multi-perspective driver fatigue detection using the DMD – Driving Monitoring Dataset. To further validate the efficacy of the proposed LSTM network, comparative analyses were performed against other state-of-the-art algorithms. Table 4 presents a detailed performance comparison between: CNN+Stacked LSTM [54], CNN+LSTM [55], 3DCNN+LSTM [3], LSTM-CNN [56], Multi-Feature LSTM [57], Residual Swin Transformer [58], Vision transformers [59], AlphaPose+LSTM and the proposed AlphaPose*+LSTM framework. Table 4 reports the experimental results of the integrated framework. Specifically, "AlphaPose+LSTM" refers to the combination of YOLOv11n with AlphaPose+LSTM, whereas "AlphaPose*+LSTM" indicates the integration of the optimized YOLOv11n* with AlphaPose+LSTM. As evidenced by the metrics in Table 4, the AlphaPose*+LSTM algorithm demonstrates exceptional performance across all evaluated parameters compared to alternative approaches.

Table 4. Comparative Analysis Between the AlphaPose+LSTM Algorithm and State-of-the-art Methods.

Model	Accuracy	Precision	Recall	F1
CNN+Stacked+LSTM	88.60%	-	87.10%	-
CNN+LSTM	94.90%	-	-	-
3DCNN+LSTM	96.00%	93.00%	100.00 %	96.00%
LSTM-CNN	85.60%	-	-	94.00%
Multi-Feature LSTM	95.23%	-	-	-
Residual Swin Transformer	96.40%	-	-	-
Vision transformers	96.56%	-	-	-
AlphaPose+LSTM	96.20%	92.50%	99.00%	96.00%

AlphaPose*+LSTM	97.60%	94.30%	100.00 %	97.30%

The fatigue driving detection system is designed to promptly issue warnings when drivers enter a fatigued state, thereby preventing potential traffic accidents. System real-time performance serves as a critical evaluation metric. During the experiments, it was found that selecting a 20-frame time window not only ensures real-time detection but also allows the driver fatigue detection system to effectively identify instances of fatigue exhibited by drivers during operation. Therefore, the system analyzes driver fatigue state based on 20-frame image sequences, with Table 5 documenting the time consumption for processing each frame at different inference stages. The complete fatigue detection cycle for 20 frames requires 958 milliseconds. Both detection accuracy and temporal efficiency demonstrate that the AlphaPose*+LSTM framework represents an advanced solution for fatigue driving detection. Although the full processing time of a 20-frame sequence is 958 ms, this value represents the total computational cost rather than the response latency. In practical deployment, the system operates using a sliding-window streaming mechanism. After an initial warm-up stage to accumulate the first 20 frames, fatigue state predictions are updated continuously for each newly incoming frame. Therefore, the effective online response latency is approximately equal to the per-frame inference time (47.9 ms), enabling near real-time fatigue alerts. Compared with block-wise sequence processing that requires waiting for an entire frame batch, the sliding-window strategy significantly reduces response delay and improves system safety in high-speed driving scenarios.

Table 5. Documents the per-frame computational latency of the fatigue driving detection system during inference.

Model	YOLOv11n*	Res50 Halpe-136	LSTM	Total
Inference Time	13.9ms	32.2ms	1.8ms	47.9ms

To validate the effectiveness of the fatigue driving detection system, we conducted visual detection of drivers under different fatigue states. As illustrated in Figure 9, the visualization results of driver fatigue state detection are presented. Furthermore, to further verify the robustness of the proposed system under occlusion and motion blur conditions, visual detection was performed in such scenarios. As shown in Figure 10, the corresponding visualization results under partial occlusion and motion blur are demonstrated.

Figure 9. Visualization Results of Driver Fatigue State Detection.

Figure 10. Robustness Detection of Drivers under Partial Occlusion and Motion Blur Conditions.

The visualization results presented in Figure 9 and 10 demonstrate that the incorporation of multi-view information significantly enhances the robustness of the detection model, particularly under challenging conditions such as strong illumination, motion blur, and partial occlusion. The visualization results in Figure 9 and Figure 10 illustrate that the integrated AlphaPose* and LSTM approach can reliably determine driver fatigue severity through postural state analysis.

4. Conclusions

Fatigued driving behavior has been widely recognized as one of the key factors contributing to traffic accidents. To address this issue, this study proposes an integrated driver fatigue monitoring system that combines pose estimation with temporal sequence modeling, incorporating an improved AlphaPose* algorithm and Long Short-Term Memory (LSTM) networks. In the detection module, we enhance

both accuracy and processing efficiency of keypoint detection by replacing the original pedestrian detector in AlphaPose with YOLOv11n and introducing a Hybrid Pooling Fusion Block (HPFB) in its neck structure. The HPFB first incorporates average pooling (AvgPool2d) to expand the global receptive field of feature maps, thereby improving contextual information modeling capability. Subsequently, max pooling (MaxPool2d) is applied to emphasize edge and texture details, enhancing the representation of targets in low-resolution regions. This fusion strategy significantly improves recognition accuracy for subtle facial dynamics critical to fatigue detection, such as eye closure, head drooping, and mouth opening. For behavioral modeling, we propose a multi-perspective analysis framework based on the Driving Monitoring Dataset (DMD), which structurally models driver behavior through three distinct pathways: frontal facial images, side posture, and hand operations. This approach substantially enriches the characterization of fatigued driving features. Furthermore, to effectively capture the temporal evolution of driver behavior, we construct an LSTM-based sequential behavior recognition network that extracts temporal dependency information from keypoint sequences, enabling continuous monitoring and identification of fatigue state progression. Experimental validation demonstrates that the proposed system achieves superior performance across multiple driver fatigue recognition tasks, showing promising potential for practical applications.

Author Contributions: Conceptualization, Y.H. and D.W.; methodology, Y.H.; software, X.S.; validation, Y.H., X.S. and H.L.; formal analysis, Y.H.; investigation, Y.H.; resources, X.S.; data curation, X.S.; writing—original draft preparation, Y.H.; writing—review and editing, H.L.; visualization, X.S.; supervision, Y.H.; project administration, H.L.; funding acquisition, D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article. The experimental data employed in this study are derived from the open-source DMD video dataset [39].

Ethics and Deployment: The experimental data employed in this study are derived from the open-source DMD video dataset [39]. Due to the limitations of this dataset, the proposed model is not directly applicable to nighttime or low-illumination environments. In practical applications, the parameters for evaluating fatigue should be adjusted based on specific data characteristics to ensure compliance with the requirements of the model. The primary application of this model lies in driver fatigue warning tasks. In real-world deployment, repeated testing of the model ' s outcomes is necessary to ensure its robustness and reliability. This study provides reference technical indicators; however, actual applications should carefully consider the influence of environmental conditions and various external factors.

Acknowledgments: We are especially grateful to our team members for their excellent cooperation and patient support during the literature analysis process.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Global status report on road safety 2018. Geneva: World Health Organization. License: CC BY- NC-SA 3.0 IGO,2018.
2. AAA Foundation for Traffic Safety. 2019 traffic safety culture index. AAA Foundation for Traffic Safety, Technical report, 2020.
3. Sara A. Alameen, Areej M. Alhothali. A Lightweight Driver Drowsiness Detection System Using 3DCNN With LSTM [J]Computer Systems Science & Engineering,2023,44(1):895-912
4. National Highway Traffic Safety Administration, Traffic safety facts 2015. https://crashstats.nhtsa. dot.gov/Api/Public/Publication/812384, 2015.
5. Vanlaar Ward, Herb Simpson, Dan Mayhew, et al., Fatigued and drowsy driving: a survey of attitudes, opinions and behaviors, J. Saf. Res. 39 (2008) 303–309, https://doi.org/10.1016/j.jsr.200 7.12.007.
6. Chen, C.S.; Lu, J.; Ma, K.K. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface. Lect. Notes Comput. Sci. 2017, 10116, V–VI.
7. Gaëtan Merlhiot, Mercedes Bueno, How drowsiness and distraction can interfere with take-over performance: A systematic and meta-analysis review, Accident Analysis & Prevention, Volume 170, 2022, 106536, ISSN 0001-4575, https://doi.org/10.1016/j.aap.2021.106536.
8. Z. Lian, T. Xu, Z. Yuan, J. Li, N. Thakor and H. Wang, "Driving Fatigue Detection Based on Hybrid Electroencephalography and Eye Tracking," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 11, pp. 6568-6580, Nov. 2024, doi: 10.1109/JBHI.2024.3446952.
9. Ardabili, S.Z.; Bahmani, S.; Lahijan, L.Z.; Khaleghi, N.; Sheykhivand, S.; Danishvar, S. A Novel Approach for Automatic Detection of Driver Fatigue Using EEG Signals Based on Graph Convolutional Networks. Sensors 2024, 24, 364. https://doi.org/10.3390/s24020364
10. Siddhad G, Dey S, Roy P P, et al. Awake at the Wheel: Enhancing Automotive Safety Through EEG-Based Fatigue Detection[C]//International Conference on Pattern Recognition. Springer, Cham, 2025: 340-353.
11. Zhou X, Lin D, Jia Z, et al. An EEG channel selection framework for driver drowsiness detection via interpretability guidance[C]//2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2023: 1-5.
12. Rezaee Q, Delrobaei M, Giveki A, et al. Driver drowsiness detection with commercial EEG headsets[C]//2022 10th RSI International Conference on Robotics and Mechatronics (ICRoM). IEEE, 2022: 546-550.
13. Halomoan, J.; Ramli, K.; Sudiana, D.; Gunawan, T.S.; Salman, M. ECG-Based Driving Fatigue Detection Using Heart Rate Variability Analysis with Mutual Information. Information 2023, 14, 539. https://doi.org/10.3390/info14100539
14. Zeng, C.; Zhang, J.; Su, Y.; Li, S.; Wang, Z.; Li, Q.; Wang, W. Driver Fatigue Detection Using Heart Rate Variability Features from 2-Minute Electrocardiogram Signals While Accounting for Sex Differences. Sensors 2024, 24, 4316. https://doi.org/10.3390/s24134316
15. Xiong H, Yan Y, Sun L, Liu J, Han Y, Xu Y. Detection of driver drowsiness level using a hybrid learning model based on ECG signals. Biomed Tech (Berl). 2023 Oct 13;69(2):151-165. doi: 10.1515/bmt-2023-0193. PMID: 37823389.

16. Wang, L.; Song, F.; Zhou, T.H.; Hao, J.; Ryu, K.H. EEG and ECG-Based Multi-Sensor Fusion Computing for Real-Time Fatigue Driving Recognition Based on Feedback Mechanism. Sensors 2023, 23, 8386. https://doi.org/10.3390/s23208386
17. Satti AT, Kim J, Yi E, Cho HY, Cho S. Microneedle Array Electrode-Based Wearable EMG System for Detection of Driver Drowsiness through Steering Wheel Grip. Sensors (Basel). 2021 Jul 27;21(15):5091. doi: 10.3390/s21155091.
18. Mahmoodi M, Nahvi A. Investigation of Sleep Deprivation Effect on Driver's Electromyography Signal Features in a Driving Simulator. J Sleep Sci. 2019;3(3-4):53-62.
19. Karthikeyan V, Kumar B P, Babu S S, et al. A narrative vehicle protection representation for vehicle speed regulator under driver exhaustion--a study[J]. arXiv preprint arXiv:1402.3657, 2014.
20. Li, Z.; Chen, L.; Peng, J.; Wu, Y. Automatic Detection of Driver Fatigue Using Driving Operation Information for Transportation Safety. Sensors 2017, 17, 1212. https://doi.org/10.3390/s17061212
21. Li, Z.; Li, S.E.; Li, R.; Cheng, B.; Shi, J. Online Detection of Driver Fatigue Using Steering Wheel Angles for Real Driving Conditions. Sensors 2017, 17, 495. https://doi.org/10.3390/s17030495
22. Zhendong Lan, Jian Zhao, Pengbo Liu, Chi Zhang, Nana Lyu, Lie Guo, Driving fatigue detection based on fusion of EEG and vehicle motion information, Biomedical Signal Processing and Control, Volume 92,2024,106031,ISSN 1746-8094, https://doi.org /10.1016/j.bspc.2024.106031.
23. Md. Ebrahim Shaik, A systematic review on detection and prediction of driver drowsiness, Transportation Research Interdisciplinary Perspectives, Volume 21,2023,100864, ISSN 2590-1982, https://doi.org/10.1016/j.trip.2023.100864.
24. Shanmeng Zhao, Yaxue Peng, Yaqing Wang, Gang Li, Mohammed Al-Mahbashi1 Lightweight YOLOM-Net for Automatic Identification and Real-Time Detection of Fatigue Driving. Computers, Materials & Continua 2025, 82(3), 4995-5017. https://doi.org/10.32604/cmc.2025.059972
25. Li, A., Ma, X., Guo, J. et al. Driver fatigue detection and human-machine cooperative decision-making for road scenarios. Multimed Tools Appl 83, 12487–12518 (2024). https://doi.org/10.1007/s11 042-023-15994-7
26. Akin A, Kalkan H. Detecting Driver Fatigue With Eye Blink Behavior[J]. arXiv preprint arXiv:2407.02222, 2024.
27. Dreißig M, Baccour M H, Schäck T, et al. Driver drowsiness classification based on eye blink and head movement features using the k-NN algorithm[C]//2020 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2020: 889-896.
28. Shugang Liu, Yujie Wang, Qiangguo Yu, Jie Zhan, Hongli Liu, Jiangtao Liu. A Driver Fatigue Detection Algorithm Based on Dynamic Tracking of Small Facial Targets Using YOLOv7. IEICE Transactions on Information and Systems, Vol. E106-D, No. 11, pp. 1881–1890, 2023.
29. Nguyen, P.V.; Smith, A.; Lee, B. Driver Yawning Detection System Using Key Facial Landmarks and Mouth Aspect Ratio. In: Advances in Computer Vision. Lecture Notes in Computer Vision. Springer, 2022.
30. Dkhil M B, Wali A, Alimi A M. Towards a new system for drowsiness detection based on eye blinking and head posture estimation[J]. arXiv preprint arXiv:1806.00360, 2018.
31. Lyu J, Yuan Z, Chen D. Long-term multi-granularity deep framework for driver drowsiness detection[J]. arXiv preprint arXiv:1801.02325, 2018.

32. Albadawi, Y.; AlRedhaei, A.; Takruri, M. Real-Time Machine Learning-Based Driver Drowsiness Detection Using Visual Features. J. Imaging 2023, 9, 91. https://doi.org/10.3390/jimaging9050091
33. Delwar, T.S.; Singh, M.; Mukhopadhyay, S.; Kumar, A.; Parashar, D.; Lee, Y.; Rahman, M.H.; Sejan, M.A.S.; Ryu, J.Y. AI- and Deep Learning-Powered Driver Drowsiness Detection Method Using Facial Analysis. Appl. Sci. 2025, 15, 1102. https://doi.org/10.3390/app15031102
34. Yashaswini NL, Arun V, Shashikala BM, Raj S, Vani HY and Flammini F (2024) Journey tracker: driver alerting system with a deep learning approach. Front. Robot. AI 11:1433795. doi: 10.3389/frobt.2024.1433795
35. Zhou C, Zhao Y, Liu S, et al. Research on driver facial fatigue detection based on Yolov8 model[C]//2024 5th International Conference on Information Science, Parallel and Distributed Systems (ISPDS). IEEE, 2024: 282-285.
36. H. S. Fang et al., "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 6, pp. 7157-7173, 1 June 2023, doi: 10.1109/TPAMI.2022.3222784.
37. Sepp Hochreiter, Jürgen Schmidhuber; Long Short-Term Memory. Neural Comput 1997; 9 (8): 1735–1780. doi: https://doi.org/ 10.1162/neco.1997.9.8.1735
38. Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[J]. arXiv preprint arXiv:2410.17725, 2024.
39. Ortega J D, Kose N, Cañas P, et al. Dmd: A large-scale multi-modal driver monitoring dataset for attention and alertness analysis[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 387-405.
40. Maycock G. Sleepiness and driving: the experience of U.K. car drivers. Accid Anal Prev. 1997 Jul;29(4):453-62. doi: 10.1016/s0001-4575(97)00024-9. PMID: 9248503.
41. Palazzi A, Abati D, Solera F, et al. Predicting the Driver's Focus of Attention: the DR (eye) VE Project[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(7): 1720-1733.
42. Fang J, Yan D, Qiao J, et al. DADA: Driver attention prediction in driving accident scenarios[J]. IEEE transactions on intelligent transportation systems, 2021, 23(6): 4959-4971.
43. Xia Y, Zhang D, Kim J, et al. Predicting driver attention in critical situations[C]//Asian conference on computer vision. Cham: Springer International Publishing, 2018: 658-674.
44. Vineetha Vijayan, Pushpalatha KP, A Comparative Analysis of RootSIFT and SIFT Methods for Drowsy Features Extraction, Procedia Computer Science, Volume 171, 2020, Pages 436-445, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.04.046.
45. Bobbie Seppelt, Sean Seaman, Linda Angell, Bruce Mehler, and Bryan Reimer. 2017. Differentiating Cognitive Load Using a Modified Version of AttenD. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17). Association for Computing Machinery, New York, NY, USA, 114–122. https://doi.org/10.1145/ 3122986.3123019
46. B.T. Zhang, W.W. Chang, X.L. Li, Fatigue Driving Detection Based on Spatial-Temporal Electroencephalogram Features and Parallel, vol. 23, 2023, pp. 315–325, https://doi.org/10.16097/j.cn ki.1009-6744.2023.02.033.
47. Schwarz, A.; Haurilet, M.; Martinez, M.; Stiefelhagen, R. DriveAHead—A Large-Scale Driver Head Pose Dataset. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HA, USA, 21–26 July 2017; Volume 2017, pp. 1165–1174.

48. Roth, M.; Gavrila, D.M. DD-pose—A large-scale driver head pose benchmark. In Proceedings of the IEEE Intelligent Vehicles Symposium, Paris, France, 9–12 June 2019; Volume 2019, pp. 927–934.
49. Liu D, Yamasaki T, Wang Y, et al. Toward extremely lightweight distracted driver recognition with distillation-based neural architecture search and knowledge transfer[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 24(1): 764-777.
50. E. Ohn-Bar and M. M. Trivedi, "The Power Is in Your Hands: 3D Analysis of Hand Gestures in Naturalistic Video," 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 2013, pp. 912-917, doi: 10.1109/CVPRW.2013.134.
51. N. Das, E. Ohn-Bar and M. M. Trivedi, "On Performance Evaluation of Driver Hand Detection Algorithms: Challenges, Dataset, and Metrics," 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, 2015, pp. 2953-2958, doi: 10.1109/ITSC.20 15.473.
52. Ultralytics. (2023). Ultralytics YOLOv8. GitHub repository. https://github.com/ultralytics/ultralytics
53. Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
54. W. Zhang and J. Su, “Driver yawning detection based on long short term memory networks,” in Proc. of IEEE Sym. Series on Computational Intelligence (SSCI), Honolulu, HI, USA, pp. 1–5, 2017.
55. Y. Fei, B. Li and H. Wang, “Long short-term memory network based fatigue detection with sequential mouth feature,” in Proc. of Int. Sym. on Autonomous Systems (ISAS), Guangzhou, China, pp. 17–22, 2020.
56. Chungho Lee, Jinung An, LSTM-CNN model of drowsiness detection from multiple consciousness states acquired by EEG, Expert Systems with Applications, Volume 213, Part B,2023,119032, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.119032.
57. Moredo, M.J.R.; Celino, J.D.S.; Ibarra, J.B.G. Multi-Feature Long Short-Term Memory Facial Recognition for Real-Time Automated Drowsiness Observation of Automobile Drivers with Raspberry Pi 4. Eng. Proc. 2025, 92, 52. https://doi.org/10.3390/engproc2025092052
58. Xiao, W.; Liu, H.; Ma, Z.; Chen, W.; Hou, J. FPIRST: Fatigue Driving Recognition Method Based on Feature Parameter Images and a Residual Swin Transformer. Sensors 2024, 24, 636. https://doi.org/10.3390/s24020636
59. Krisna, G. S.; Supriya, K.; Vardhan, J. Vision transformers and YoloV5 based driver drowsiness detection framework[J]. arXiv preprint arXiv:2209.01401, 2022.