electronics

MDPI

Article

Fatigue Driving Recognition Method Based on Multi-Scale Facial Landmark Detector

Weichu Xiao [1,2], Hongli Liu [1,*], Ziji Ma [1], Weihong Chen [3], Changliang Sun [1] and Bo Shi [1]

[1] College of Electrical and Information Engineering, Hunan University, Changsha 410082, China
[2] College of Information and Electronic Engineering, Hunan City University, Yiyang 413000, China
[3] College of Information Technology and Management, Hunan University of Finance and Economics, Changsha 410205, China
* Correspondence: hongliliu@hnu.edu.cn

Abstract: Fatigue driving behavior recognition in all-weather real driving environments is a challenging task. Accurate recognition of fatigue driving behavior is helpful to improve traffic safety. The facial landmark detector is crucial to fatigue driving recognition. However, existing facial landmark detectors are mainly aimed at stable front face color images instead of side face gray images, which is difficult to adapt to the fatigue driving behavior recognition in real dynamic scenes. To maximize the driver's facial feature information and temporal characteristics, a fatigue driving behavior recognition method based on a multi-scale facial landmark detector (MSFLD) is proposed. First, a spatial pyramid pooling and multi-scale feature output (SPP-MSFO) detection model is built to obtain a face region image. The MSFLD is a lightweight facial landmark detector, which is composed of convolution layers, inverted bottleneck blocks, and multi-scale full connection layers to achieve accurate detection of 23 key points on the face. Second, the aspect ratios of the left eye, right eye and mouth are calculated in accordance with the coordinates of the key points to form a fatigue parameter matrix. Finally, the combination of adaptive threshold and statistical threshold is used to avoid misjudgment of fatigue driving recognition. The adaptive threshold is dynamic, which solves the problem of the difference in the aspect ratio of the eyes and mouths of different drivers. The statistical threshold is a supplement to solve the problem of driver's low eye threshold and high mouth threshold. The proposed methods are evaluated on the Hunan University Fatigue Detection (HNUFDD) dataset. The proposed MSFLD achieves a normalized mean error value of 5.4518%, and the accuracy of the fatigue driving recognition method based on MSFLD achieves 99.1329%, which outperforms that of state-of-the-art methods.

Keywords: deep learning; facial landmark detector; fatigue driving recognition; multi-scale

Citation: Xiao, W.; Liu, H.; Ma, Z.; Chen, W.; Sun, C.; Shi, B. Fatigue Driving Recognition Method Based on Multi-Scale Facial Landmark Detector. *Electronics* **2022**, *11*, 4103. https://doi.org/10.3390/electronics11244103

Academic Editor: Felipe Jiménez

Received: 22 November 2022
Accepted: 7 December 2022
Published: 9 December 2022

1. Introduction

1.1. Background

Fatigue driving refers to the phenomenon of psychological and physiological dysfunction of drivers due to excessive mental and physical exertion during long-term driving. When a driver is fatigued, the physiological function, recognition, and control ability decline, and the driver cannot respond to the sudden accident in time, thus seriously affecting safe driving. According to a research by the World Health Organization, about 1.24 million people in the world die from road safety accidents every year, causing economic losses of tens of billions of dollars [1]. The Statistical Yearbook of Traffic Accidents estimates that about 17% of road traffic accidents are related to driver fatigue. A survey in Canada reported that 20% of fatal collisions involved fatigue [2]. Another survey in Pakistan reported 34% of road accidents were related to fatigue [3]. According to a US survey, 20% of fatal accidents are caused by drowsy drivers [4]. In the EU, fatigue driving leads to 20% of commercial transport crashes [5]. If drivers are reminded half a second in advance, nearly

60% of traffic accidents can be effectively avoided. Therefore, it is necessary to accurately recognize the fatigue driving state of drivers and issue warnings in time for traffic safety.

1.2. Motivation

Researchers around the world have carried out various studies on fatigue driving identification. From the dataset perspective, available datasets includes the yawning detection dataset (YawDD), the multi-facial action yawning dataset (MFAY), and NTHU drowsy driver detection dataset (NTHU-DDD) [6–8]. YawDD is a public yawning detection dataset, which includes normal, talking/singing, and yawning driving behaviors. Vehicles are stationary in the YawDD dataset. In terms of recognition methods, deep learning methods have been used to recognize fatigue driving behavior, and achieved high accuracy [9,10]. Given that fatigue driving behavior is a fine-grained activity, the focus should be on areas such as eyes and mouth. Savas et al. proposed a multi-tasking convolutional neural network model to detect driver fatigue. The Dlib algorithm is used to identify a driver's eye and mouth information [9]. Then, the system determines the fatigue parameters through a multi-task ConNN model. Finally, the duration of eye closure and the frequency of yawning are calculated, which can be used to determine the driver's fatigue level. Liu et al. proposed a multi task cascaded convolutional neural network (MTCNN) to detect the face and locate key points [10]. Then, the fatigue parameters of eyes and mouths are calculated through key points. Finally, two fatigue characteristic parameters are fused to judge the fatigue of drivers according to the PERCOS criterion and the fuzzy reasoning principle. However, enhancing the accuracy of fatigue driving behavior recognition is insufficient owing to the following reasons:

(1) The existing fatigue driving behavior recognition methods obtain the opening and closing state of eyes and mouths through the annotation frames or calculate the aspect ratio of eyes and mouth by annotating key points. If the detection accuracy of facial key points is low, then the recognition accuracy of fatigue driving behavior becomes low. Therefore, it is necessary to design a high-accuracy facial landmark detector.
(2) Most existing fatigue driving decision models use fixed thresholds to recognize drive fatigue. However, under the condition of fatigue driving, the aspect ratios of the eyes and mouths of different drivers vary. Therefore, the threshold value of the aspect ratio of the eyes and mouth of each driver in the fatigue driving state is different, requiring dynamic changes in the threshold value.
(3) At present, the public datasets of fatigue driving behavior mainly focuses on yawning behavior and rarely involves dozing behavior. The public dataset of facial key point detection rarely contains driver behavior images, and the task of manually marking 68 or 98 key points is arduous. Therefore, a dataset of driver behavior images needs be built. The dataset should consider reducing the number of key points manually marked in each image and various driving behavior types, including dozing, yawning, talking and normal.

1.3. Our Contributions

This study proposes a deep learning method to improve the detection accuracy of facial key points and applies it to the recognition of fatigue driving behavior. The main contributions of this study are outlined below.

(1) A novel deep learning framework called Multi-scale Facial Landmark Detector (MSFLD) is proposed to perform facial 23 key point detection. The MSFLD model replaces all the bottleneck layers in the traditional facial landmark detector with inverted residual blocks and increases the number of multi-scale fully connected layers, which reduces the number of model parameters. Thus, the proposed MSFLD can improve the detection accuracy of facial key points while keeping the detection speed constant.
(2) A MSFLD-based method is proposed for fatigue driving behavior recognition. The proposed method uses a spatial pyramid pooling and multi-scale output (SPP-MSFO) detection model to obtain the face region, detects 23 key points through MSFLD, calcu-

lates the fatigue parameter matrix according to the key points, and uses the combination of adaptive threshold and statistical threshold to determine the driver's fatigue status. This method not only improves the accuracy of fatigue driving behavior recognition, but also reduces the workload of labeling facial key points in dataset images.

(3) In the proposed fatigue driving recognition method, a driving behavior judgment strategy combining an adaptive threshold and statistical threshold is presented. The adaptive threshold addresses the problem of differences in the aspect ratio of the eyes and mouth of different drivers. The statistical threshold solves the problem that the adaptive threshold of the eyes is too low or the adaptive threshold of the mouth is too high. The combination of the two can avoid misjudgment and improve the recognition accuracy of driving behavior.

(4) The Hunan University Fatigue Driving Detection Dataset (HNUFDD) is built. The HNUFDD dataset includes yawning, dozing, talking, mouth closed and normal driving behavior types, and annotates 23 key points in the driver's face area. The proposed method is evaluated on the HNUFDD dataset, and the results show the superior performance of the proposed methods in comparison with state-of-the-art methods.

The rest of this study is organized as follows: Section 2 reviews related work. Section 3 proposes a fatigue driving recognition method based on multi-scale facial landmark detector. Experimental verification is given in Section 4. Section 5 concludes this study.

2. Related Work

In this section, the methods of facial landmark detection and fatigue driving recognition are described.

2.1. Facial Landmarks Detection

Facial landmark detection is one of the key elements of fatigue driving recognition, and the object is to obtain key information about eyebrows, eyes, mouth and nose in fatigue driving recognition [11,12]. In recent years, many researches have been carried out on facial landmark detection [13–15]. With the superior performance of deep learning, Sun et al. first introduced a cascaded convolutional network for facial key point detection [16]. The proposed method only recognizes five facial key points, although its speed is fast. To improve the precision of facial key point recognition, Zhang et al. proposed a deep cascaded multitask convolutional neural network (MTCNN), which can realize face detection and key point detection from coarse to fine [17]. However, the operation of this algorithm is complicated. Deng et al. proposed a robust single-stage facial landmark detector, which adopts a multi-task learning strategy to predict face boxes, facial landmarks, and correspondence of each facial pixel simultaneously [18]. However, the size of the model increases when the ResNet50 network is used to train the retina face model. Guo et al. proposed a practical facial landmark detector, which consists of two subnets: a backbone network and an auxiliary network [19]. The proposed method is an end-to-end single-stage network capable of predicting 68 key points. However, 68 key points must be manually annotated for each image in the dataset. Liu et al. proposed a Densely U-Nets Refine Network (DURN), which is composed of DU-Net and Refine-Net, for facial landmark localization [20]. The proposed method can predict 106 key points, which contains more structural information than 68 key points; however, 106 key points must be manually annotated for each image in the dataset. Hassaballah et al. proposed a deep learning-based method using cascaded regression for coarse-to-fine detection of facial landmarks [21]. This proposed mothod is called coordinate regression with heatmap coupling (CR-HC). The method is composed of two-stage cascaded CNNs that are coupled with a heatmap module. The results show the performance of the method.

The above methods validate its performance on public datasets such as 300 W, AFLW, WIDER FACE, FDDB, and WFLW, but these datasets have few images of driver behaviors. Training network models based on these public datasets is not good for detecting the key points in driver behavior images. Therefore, a dataset including driver behavior images

need to be constructed. Such dataset should consider reducing the number of manually annotated key points in each image, and can calculate the fatigue driving parameters.

2.2. Fatigue Driving Recognition

Fatigue driving recognition can be divided into five categories based on input features: subjective report, biological, physical, vehicular, and hybrid [22]. The subjective report recognition method is based on a questionnaire survey of drivers, and a fatigue level is obtained after analysis and comparison according to the main symptoms of the respondents and the number of occurrences of various symptoms [23]. The advantage of this recognition method is that no invasive problems occur, and the disadvantage is that the process of measuring fatigue is not synchronized with the driving process and is easily affected by the subject's emotional and physical condition. The recognition method based on the driver's biological features usually has high accuracy, but the recognition process requires special equipment to measure biological signals such as electroencephalography [24], electrocardiogram [25], electro-oculography [26], and surface electromyogram [27]. The cost of the equipment is high, and the contact-type signal acquisition method is available. However, such equipment causes certain interference to drivers when driving, and most drivers hardly receive these devices. The recognition method based on vehicular features uses indicators such as steering wheel movement and lane offsets for recognition [28]. It only needs to obtain vehicle information and does not cause any interference to the driver. However, this recognition method is affected by road conditions and driver skills. The recognition method based on physical features uses image processing methods to detect the changes of the driver's individual features such as eyes, mouth, head, and facial expressions, to monitor the driver's fatigue states [29]. With the advantages of high accuracy and non-contact detection, this recognition method has become the mainstream of current research. However, the performance of this method is affected by objective factors, including light, shielding of glasses or masks, and vehicle movement. As a result, the recognition accuracy varies greatly in different environments, and its robustness is low. Moreover, user facial data needs be collected during the detection process, which involves user privacy issues. The recognition method based on hybrid features fuses various features for fatigue detection.

In the past few years, the deep learning methods on fatigue driving recognition have been developed with the success of computer vision [30–34]. Li et al. proposed a fatigue driving detection algorithm based on facial multi-feature fusion [30]. In this method, the improved YOLOv3-tiny convolution neural network is used to capture the face area, and the evaluation parameters of eye feature vector and oral feature vector are introduced. The driver's eye closure time and yawning frequency are calculated through the driver fatigue evaluation model to evaluate the driver's fatigue state. Du et al. put forward a multi-mode fusion recurrent neural network model by integrating the three characteristics of heart rate, mouth opening degree and eye opening degree, which can accurately recognize the fatigue driving state [31]. Raja et al. proposed a fatigue detection system based on multi-task cascaded convolutional neural networks [32]. The method uses a multi task cascaded convolutional neural network to predict the facial boundary box and five key points and infers the fatigue level of drivers by the percentage of closed eyes to pupils and the frequency of yawning and nodding. Jia et al. proposed a fatigue driving recognition algorithm based on deep learning and facial multi-index fusion [33]. The algorithm uses an improved MTCNN to detect the key points of the face, then determines the driver's eyes and mouth area according to the key points of the face, and finally judges the driver's fatigue state through E-MSR Net.

However, the existing methods for extracting fatigue features, such as eyes and mouth, are not accurate enough, and most of them ignore the time information of fatigue features and the relationship between features, thus reducing recognition accuracy. In contrast to simple static image classification and recognition, fatigue driving behavior is a continuous action. Simply using a single static image for classification and recognition loses impor-

tant temporal information. Therefore, the recognition of fatigue driving behavior should consider not only the opening and closing state of the eyes and the opening degree of the mouth, but also the duration or number of these actions.

3. Fatigue Driving Recognition Method Based on MSFLD

In this section, a fatigue driving recognition method based on the multi-scale facial landmark detector is proposed, including the overall architecture of the proposed method, the SPP-MSFO detection model, the MSFLD model and its learning algorithm, the parameter matrix of facial fatigue features, and driving behavior decision.

3.1. Overview of Architecture

The fatigue driving recognition method based on the MSFLD is mainly composed of four parts: the SPP-MSFO detection model, the MSFLD model, facial fatigue feature parameter matrix, and driving behavior decision module. The architecture overview diagram of the proposed method is shown in Figure 1. First, the input test video is framed, and the face region image is obtained through the SPP-MSFO detection model. Then, the MSFLD model is used to obtain the coordinates of 23 key points, and the aspect ratios of the left eye right eye and mouth are calculated in accordance with the coordinates of the key points to form a fatigue parameter matrix. Finally, the driving behavior decision module is used to judge whether the driver is fatigue driving, and the decision result is obtained as the output.

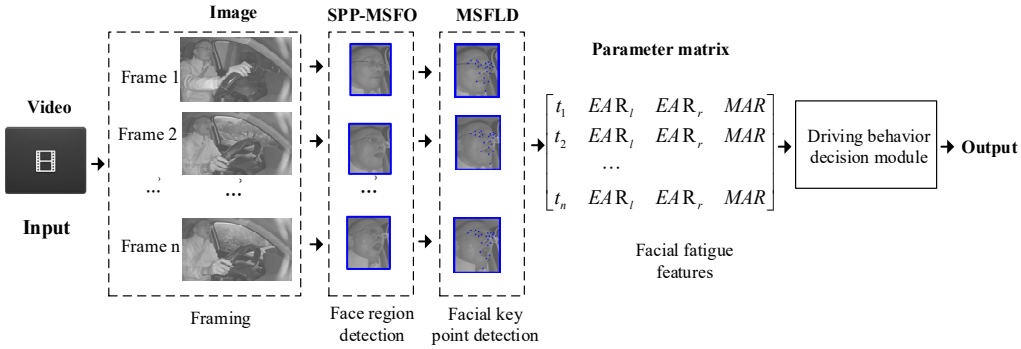


Figure 1. Architecture overview diagram of the proposed method.

Different from the existing facial key point detection techniques, this study proposes a method of using 23 key points. The facial key points are detected using the MSFLD model, and the number of key points is 23 instead of 68. The reasons for this are as follows: (1) Fatigue characteristic parameters, such as eye aspect ratio, and mouth aspect ratio can be calculated by using 23 key points. (2) Any model commits errors when detecting key points. The more key points, the greater the accumulated error when calculating fatigue characteristic parameters. That is, the number of facial key points has an effect on the accuracy of fatigue driving behavior recognition. (3) Marking facial key points is time-consuming. It is estimated that the workload of labeling 68 key points in the facial images of a dataset is nearly 3 times that of labeling 23 key points. Can fatigue driving detection be achieved without compromising accuracy using 23 key points? To our knowledge, this is the first time to use 23 key points for fatigue driving detection.

3.2. SPP-MSFO Detection

SPP-MSFO detection uses the SPP-MSFO model to detect facial regions from framed images. The SPP-MSFO model is a lightweight, single-stage object detection model. It consists of a backbone module, a spatial pyramid pooling (SPP) module and multi-scale feature output (MSFO) module, as shown in Figure 2.

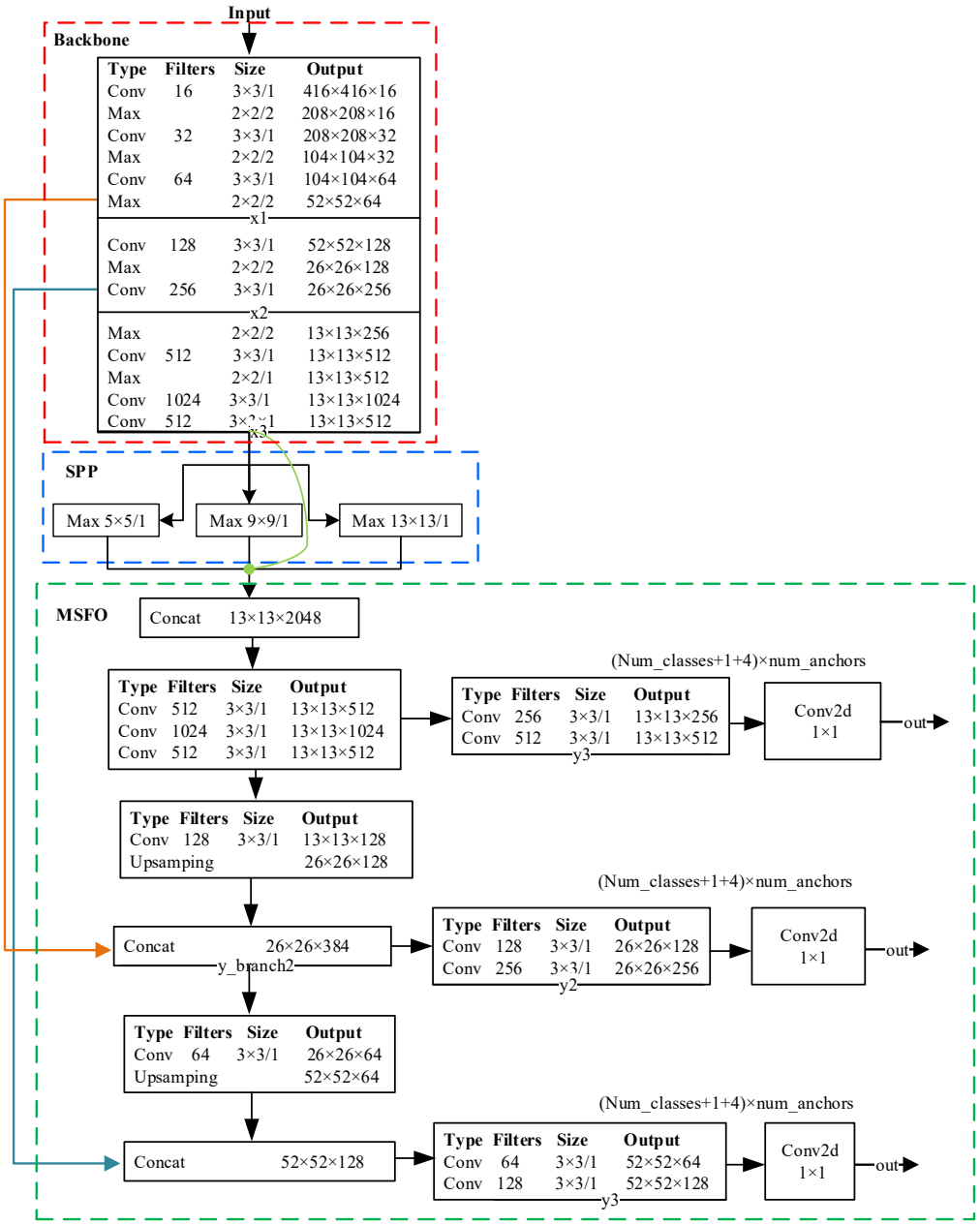


Figure 2. Structure of the SPP-MSFO model.

The SPP-MSFO model is based on the YOLOV3 tiny detection model with three innovations: (1) At the end of the backbone module, a convolutional operation with 512 channels and a convolution kernel of 3 × 3 is added. Branches are introduced from the maximum pooling layer of the sixth layer and the convolution layer of the ninth layer of the backbone module to facilitate subsequent feature fusion. (2) After the backbone module, the SPP module is added to realize the fusion of local features and global features, increase the size of receptive field, and enrich the expression ability of the feature map. (3) Adding the third scale (52 × 52) output of the characteristic map makes the receptive field smaller and improves the detection accuracy of the acquired face region image.

3.3. Model of The MSFLD and Its Learning Algorithm

The object of the multi-scale facial landmark detector is to detect facial key points for judging driving fatigue from the SPP-MSFO, including the key points of eyes and mouth. In this part, the MSFLD model and its learning algorithm are proposed.

3.3.1. The MSFLD Model

The MSFLD model is used to predict landmark coordinates, and the framework is shown in Figure 3. First, an image with a size of 112 × 112 × 3 is input into the model, and one pointwise convolution and one depth-wise convolution are performed. The results pass through four inverted bottleneck blocks and two convolution layers, greatly reducing the amount of computation and speeding up the operation of the model. Then, four average pooling layers and multi-scale fully connected layers are conducted. In this manner, the receptive field is increased. Thus, the global information of the face can be better obtained and the positioning accuracy of the model can be improved. Finally, the coordinates of 23 key points are obtained as the output of the MSFLD model.

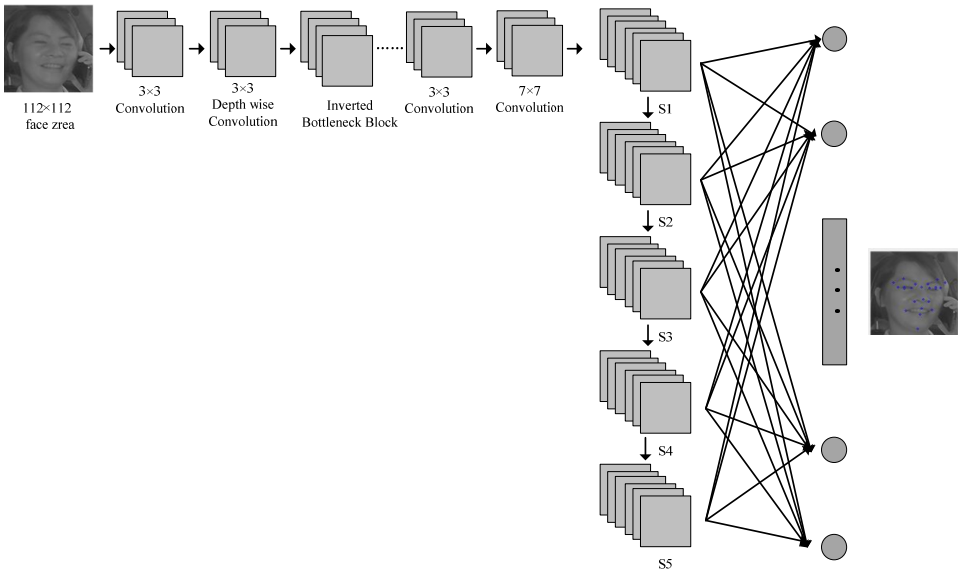


Figure 3. Architecture of the MSFLD model.

Table 1 lists the configuration of each module of the MSFLD model. Each line represents a sequence of identical layers, repeating n times. All layers in the same sequence have the same number c of output channels. The first layer of each sequence has a stride s. The expansion factor t is always applied to the input size. p is for padding. As shown in the table, the convolution kernel size is 3 × 3 in the first two convolutions. In the inverted bottleneck blocks, convolution kernels of size 3 × 3 and 7 × 7 are used. The scales of the four pooling layers are 56 × 56, 28 × 28, 14 × 14 and 7 × 7, respectively.

Table 1. MSFLD network configuration.

Input	Operator	*t*	*c*	*n*	*s*	*p*
112 × 112 × 3	Conv 3 × 3	–	64	1	2	1
56 × 56 × 64	Depth wise Conv 3 × 3	–	64	1	1	1
56 × 56 × 64	Inverted bottleneck	2	64	3	2	1
28 × 28 × 64	Inverted bottleneck	3	96	3	2	1
14 × 14 × 96	Inverted bottleneck	4	144	4	2	1
7 × 7 × 144	Inverted bottleneck	2	16	1	1	1
7 × 7 × 16	Conv 3 × 3	–	32	1	1	1
7 ×7 × 32	Conv 7 × 7	–	128	1	1	0
(S1) 56 × 56 × 64	Avg pool	–	64	1	2	1
(S2) 28 × 28 × 64	Avg pool	–	96	1	2	1
(S3) 14 × 14 ×96	Avg pool	–	144	1	2	1
(S4) 7 × 7 × 144	Avg pool	–	128	1	1	0
(S5) 1 × 1 × 128	–	–	128	–	–	–
In_feature = 496	Full connection	–	46	1	–	–

3.3.2. Learning Algorithm of the MSFLD

The 23 key points of the driver's face are obtained by the trained MSFLD. The training strategy of the MSFLD is proposed on the basis of the above analysis, as shown in Algorithm 1.

Algorithm 1: Training strategy of MSFLD

Input: Given training samples 17441 face region images of size 112 × 112, $X = \{X_1, X_2, \ldots, X_M\}$ and their 23 key point annotations, $Y = \{Y_1, Y_2, \ldots, Y_M\}$.
Output: The well trained model MSFLD
1: Construct the MSFLD model shown in Figure 3;
2: Initialize the parameters $\theta(w, b, \alpha)$, set the batch size (i.e., 96);
3: **Repeat**
4: Randomly select a batch instances X_b from X;
5: Forward learn training samples through the MSFLD model;
6: Compute the loss function L_2 by Equation (1);
7: Propagate the error back-through MSFLD and update the parameters of MSFLD;
8: Find L_2 by minimizing L_2 with X_b;
9: **Until** end condition is satisfied.

The key details are illustrated as follows:

(1) In Line 1, the structure of the MSFLD model is constructed. This model consists of convolutions, inverted bottleneck blocks, average pooling, and multi-scale fully connected layers. The overview of the MSFLD architecture is illustrated in Figure 3.

(2) In Line 2, parameters of the MSFLD model, including the weight value w, bias b, learning rate α, and batch size, are initialized. The initialization scheme for these parameters is described in detail in Section 4.

(3) In Lines 3–9, the strategies of forward learning and backward propagation are used to train the MSFLD model. In the backward propagation, the model uses Adam to optimize parameters. In Line 6, the loss function L_2 is defined as shown in Equation (1).

$$L_2 = \frac{1}{M} \sum_{m=1}^{M} \sum_{n=1}^{N} w_n \|d_n^m\|_2^2 \quad (1)$$

where M denotes the number of training images in each process, N is the pre-defined number of landmarks to be detected for each face, w_n is the weight, $\|d_n^m\|$ designates a certain metric to measure the distance of the n-th landmarks of the m-th input [14].

(4) In Line 9, model training is completed until the end condition is met. The iteration limit and early stop policy are used as the end conditions. At the end of the training, the MSFLD model with optimal parameters for 23 key point detection is obtained.

3.4. Facial Fatigue Feature Parameter Matrix

The parameter matrix of facial fatigue feature is the basis for judging driver fatigue, and its value is calculated according to facial key points obtained from the MSFLD model. This part presents the feature extraction of eye fatigue, the feature extraction of mouth fatigue, and the calculation of fatigue parameter matrix.

3.4.1. Feature Extraction of Eye Fatigue

The eyes are an important fatigue characterization parameter and can indicate whether the driver is dozing off according to the degree of eye opening and closing. Based on 23 key points obtained by the MSFLD, 8 points are selected to extract eye closure features, including 4 in the left eye and 4 in the right eye. The coordinates of the four key points of the left eye are (x_6, y_6), (x_7, y_7), (x_8, y_8), (x_{13}, y_{13}). The coordinates of the four key points of the right eye are (x_9, y_9), (x_{10}, y_{10}), (x_{11}, y_{11}), (x_{12}, y_{12}). It is noted that x represents the abscissa and y represents the ordinate. The left eye aspect ratio (EAR) EAR_l and the right

eye aspect ratio EAR_r are used to judge the driver's eye opening and closing state. EAR_l and EAR_r are calculated as Equations (2) and (3), respectively.

$$EAR_l = \frac{y_{13} - y_7}{x_8 - x_6} \tag{2}$$

$$EAR_r = \frac{y_{12} - y_{10}}{x_{11} - x_9} \tag{3}$$

where x_6, x_8, x_9, and x_{11} are the abscissas of the key points of the left eye and the right eye, respectively. y_7, y_{13}, y_{10}, and y_{12} are the vertical coordinates of the key points on the face of the left eye and the right eye, respectively.

3.4.2. Feature Extraction of Mouth Fatigue

The mouth feature is also an important fatigue characterization parameter. By judging the opening degree of the mouth, it can be judged whether the driver is in a yawning state. The mouth usually has three states: closing, talking and yawning. The driver yawns with his mouth wide open. That is, when yawning, the height of the mouth increases and the width decreases. The mouth aspect ratio MAR is used to judge the driver's mouth opening and closing state, and is computed as Equation (4).

$$MAR = \frac{y_{21} - y_{19}}{x_{20} - x_{18}}, \tag{4}$$

where x_{18}, x_{20} are the abscissas of the two key points on the left and right of the mouth, y_{19}, y_{21} are the ordinates of the two key points above and below the mouth.

3.4.3. Calculate the Fatigue Parameter Matrix

On the basis of the coordinates of facial 23 key points in each frame, the aspect ratios of the left eye, the right eye and the mouth can be calculated in accordance with Equations (2)–(4). The fatigue parameter matrix is constructed by using EAR_l, EAR_r, and MAR of each frame, as shown in Equation (5). In Equation (5), the first column represents the number of frames, the second, third and fourth columns represent the aspect ratio of the left eye, right eye and mouth in the corresponding frame, respectively.

$$\begin{bmatrix} t_1 & EAR_{l1} & EAR_{r1} & MAR_1 \\ t_2 & EAR_{l2} & EAR_{r2} & MAR_2 \\ \vdots & \vdots & \vdots & \vdots \\ t_n & EAR_{ln} & EAR_{rn} & MAR_n \end{bmatrix}, \tag{5}$$

3.5. Driving Behavior Decision Module

Based on the above parameter matrix, the method of combining adaptive threshold and statistical threshold is proposed for fatigue driving decision.

3.5.1. Adaptive Threshold Calculation

The adaptive threshold is obtained by calculating the eye aspect ratio and mouth aspect ratio of the first p frames of each test video and then taking the average value, where p is an integer. The adaptive threshold of eyes EAR_{at} and the adaptive threshold of mouth MAR_{at} are calculated as shown in Equations (6) and (7). The value of adaptive threshold is dynamical, which effectively deals with the problem of different aspect ratio of eyes and mouth of different drivers.

$$EAR_{at} = \frac{\frac{1}{p}\sum_1^p EAR_l + \frac{1}{p}\sum_1^p EAR_r}{2} \tag{6}$$

$$\mathrm{MAR_{at}} = \frac{1}{\mathrm{p}} \sum_{1}^{\mathrm{p}} \mathrm{MAR} \tag{7}$$

3.5.2. Statistical Threshold Calculation

The statistical threshold is an average value based on different driving behavior types and different drivers. It mainly solves the problem wherein the adaptive threshold may have a low eye threshold and a high mouth threshold when testing fatigue driving behavior videos.

The steps of statistical threshold are as follows: First, the average of the eye aspect ratio and mouth aspect ratio for normal, closed mouth, and talking driver behavior are calculated. Normal, closed mouth, and talking are non-fatigue driving behavior, and 23 key points have been manually marked in the dataset. Then, the eye aspect ratio of doze driving behavior and the mouth aspect ratio of yawn driving behavior are computed. Finally, the eye aspect ratio of non-fatigue driving behavior and the eye aspect ratio of doze driving behavior are averaged. The result obtained is the statistical threshold of eyes, denoted as $\mathrm{EAR_{st}}$. $\mathrm{EAR_{st}}$ is calculated as shown in Equation (8).

$$\mathrm{EAR_{st}} = \frac{\frac{1}{\mathrm{m}} \sum_{1}^{\mathrm{m}} \mathrm{EAR_{normal}} + \frac{1}{\mathrm{n}} \sum_{1}^{\mathrm{n}} \mathrm{EAR_{cm}} + \frac{1}{\mathrm{u}} \sum_{1}^{\mathrm{u}} \mathrm{EAR_{talk}} + \frac{1}{\mathrm{k}} \sum_{1}^{\mathrm{k}} \mathrm{EAR_{doze}}}{4} \tag{8}$$

where $\mathrm{EAR_{normal}}$, $\mathrm{EAR_{cm}}$, $\mathrm{EAR_{talk}}$, and $\mathrm{EAR_{doze}}$ are the eye aspect ratio for normal, closed mouth, talking, and doze driving behavior, respectively. m, n, u and k are the number of samples of normal, closed mouth, talking and dozing in the dataset marked with 23 key points, respectively.

Similarly, the statistical threshold of the mouth $\mathrm{MAR_{st}}$ is obtained by averaging the mouth length width ratio of non-fatigue driving behavior and the mouth width ratio of yawn driving behavior, as shown in Equation (9).

$$\mathrm{MAR_{st}} = \frac{\sum_{1}^{\mathrm{m}} \mathrm{MAR_{normal}} + \frac{1}{\mathrm{n}} \sum_{1}^{\mathrm{n}} \mathrm{MAR_{cm}} + \frac{1}{\mathrm{u}} \sum_{1}^{\mathrm{u}} \mathrm{MAR_{talk}} + \frac{1}{\mathrm{v}} \sum_{1}^{\mathrm{v}} \mathrm{MAR_{yawn}}}{4} \tag{9}$$

where $\mathrm{MAR_{normal}}$, $\mathrm{MAR_{cm}}$, $\mathrm{MAR_{talk}}$, and $\mathrm{MAR_{yawn}}$ are the mouth aspect ratio for normal, closed mouth, talking, and yawn driving behavior, respectively. m, n, u and v are the number of samples of normal, closed mouth, talking and yawning in the dataset marked with 23 key points, respectively.

3.5.3. Fusion Strategy of Adaptive and Statistical Thresholds

When the test video is input to the model, the driving behavior type is unknown. If the input test video is a dozing driving behavior, the adaptive threshold of the eyes is too low. In this case, it will lead to errors in the judgment of driving behavior. Therefore, the threshold of the eye aspect ratio of fatigue driving behavior is obtained by taking the maximum value of the adaptive threshold and the statistical threshold, which can avoid misjudgment. The combined threshold of eyes $\mathrm{EAR_{ct}}$ was shown in Equation (10).

$$\mathrm{EAR_{ct}} = \max\{\mathrm{EAR_{at}}\ ,\ \mathrm{EAR_{st}}\}. \tag{10}$$

Similarly, if the input test video input is a yawning driving behavior, the adaptive threshold of the mouth is too high. This will also lead to misjudgment of driving behavior. Thus, the threshold of the mouth aspect ratio of fatigue driving behavior is obtained by taking the minimum value of the adaptive threshold and the statistical threshold, which can avoid misjudgment. The combined threshold of mouth $\mathrm{MAR_{ct}}$ is calculated as shown in Equation (11).

$$\mathrm{MAR_{ct}} = \min\{\mathrm{MAR_{at}}\ ,\ \mathrm{MAR_{st}}\}. \tag{11}$$

4. Experiments

In this section, the effectiveness of the proposed approach is evaluated on the HNUFDD dataset, including the convergence and parameter sensitivity of the models and the accuracy in comparison with existing methods.

4.1. Settings

This subsection describes the datasets, experimental conditions, metrics and baselines.

4.1.1. Dataset Description

The HNUFDD dataset was used in our experiments. To collect the HNUFDD dataset, our group carefully configured the following components: camera, environment, participants, and videos. The driving behavior video acquisition system, which was provided by Kunshan Stellate Ship Intelligent Technology Co., Ltd. (Kunshan, China) was installed in the front right side of the driver in the car. The videos were collected using an infrared camera with fill light at a resolution of 1920 × 1080 pixels, 24-bit depth, and 25 frames per second. To reflect varying illumination conditions, the videos were recorded from early morning until sunset and sometimes into the evening. Moreover, the weather varies from sunny to rainy. To reflect a real driving environment, we recorded some driving behavior videos while the car was moving. The participants were asked to sit in the driver's seat and wear their seat belts. The dataset contained videos of 34 male and 16 female volunteers with different ages and facial characteristics. People with and without glasses, different hairstyles and different clothing participated. In the dataset, most participants were filmed with videos of five categories of driving behavior, i.e., normal, dozing, yawning, mouth closed, and talking. The videos lasted about 15 s. This dataset has a total of 346 videos.

4.1.2. Dataset Preprocessing

The data obtained from the real scene is collected into the dataset through the process of framing, selection, marking. First, each video was converted into images by framing processing. Then, images that correspond to each type of driving behavior were manually selected and retained. The dataset contained five categories: normal, dozing, yawning, closed-mouth, and talking driving behavior, with a total of 22,007 images, including 17,441 images in the training set and 4566 images in the test set. Finally, the eyebrows, eyes, nose, and mouth of the driver image were labeled with 23 key points by "labelme" software (version 5.0.1), as shown in Figure 4, to create a labeled dataset.

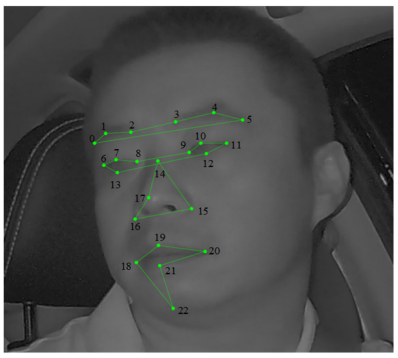


Figure 4. Annotation map of 23 key points on the face.

4.1.3. Experimental Conditions

The experiments were conducted on a 64-bit Ubuntu 20.04 platform with Intel x299 Core i9-10900X CPU @ 3.7 GH, NVIDIA GeForce RTX 3090 and 48 GB memory. Python language and PyTorch framework were used.

The size of the input face region images is 112 × 112 × 3, where 3 is the number of channels of images, and the height and width of the image are 112. Parameter initialization

in forward pass and backward fine-tuning is important for model training. In this study, the weights between layers were initialized randomly and obeyed uniform distribution. All biases were initialized as zero. Model optimization used stochastic gradient a with momentum of 0.9, learning rate of 0.0001, and batch size of 96. All the deep learning models were trained with the same optimization scheme.

4.1.4. Evaluation Metrics

To evaluate the performance of the proposed methods, normalized average error (NME) and accuracy are used. NME is used to measure the performance of the proposed MSFLD model, which is a common evaluation index for facial key point detection. NME is the average of the normalized error of all annotation landmarks, and is defined:

$$\mathrm{NME} = \frac{1}{\mathrm{N}} \sum_{\mathrm{k}=1}^{\mathrm{N}} \frac{\|\mathrm{p_k} - \mathrm{g_k}\|_2^2}{\mathrm{d}}, \tag{12}$$

where $\mathrm{p_k}$ and $\mathrm{g_k}$ are the coordinates of the *k*th predicted key point and the real key point, respectively. d is the Euclidean distance between the key points of the two outer eye corners. The smaller the NME value, the better the detection performance of the model.

Accuracy is an important index to measure fatigue driving recognition performance, and its definition is shown in Equation (13).

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{13}$$

where TP is the number of true positives, TN is the number of true negatives, FN is the number of false negatives, and FP is the number of false positives. In performing the evaluation experiments, if the input model is a "doze" or "yawning" video, and the test result shows "fatigue", MSFLD successfully detected fatigue driving behavior, i.e., the detected result is a true positive; otherwise, the test result is classified as a false negative. If the input model is "closed" or "normal" or "talking" video and the test result does not display "fatigue", the test result is classified as a true negative; otherwise, the test result is classified as a false positive.

4.1.5. Baselines

The proposed method is compared with MTCNN [17], Retina Face [18], PFLD [19], and DURN [20]. The detailed descriptions of the state-of-the-art methods are as follows.

MTCNN [17]: MTCNN is a multitask cascaded convolutional network, which is composed of a proposal network (P-Net), refine network (R-Net), and output network (O-Net), which can realize face detection and key point detection from coarse to fine. The image pyramid of MTCNN can transform the size of the initial image. The P-Net model is used to generate numerous candidate target regions, and the R-Net model is used to select and regress the target regions, excluding most of the negative examples, The O-Net model discriminates and regresses the remaining target region boxes to achieve face region detection and key point detection.

Retina Face [18]: Retina Face is a robust single-stage facial landmark detector, which uses a multi-task learning strategy to predict face scores, face boxes, facial landmarks, and 3D position and correspondence of each facial pixel simultaneously. Retina Face is designed on the basis of the feature pyramids with independent context modules. The Mobile0.25 network or the Resnet50 network is used to train the Retina Face model.

PFLD [19]: PFLD is a practical facial landmark detector, which consists of two sub subnets, i.e., the backbone network and the auxiliary network. The backbone network is composed of two convolutional layers, four bottleneck layers, and three fully connected layers, which are used to predict the location of feature points. The auxiliary network is composed of four convolutional layers and two fully connected layers, which are used to

predict face pose. The model is trained for face key point detection on the 300 W and AFLW datasets. The size of the PFLD model is 6.6MiB.

DURN [20]: DURN is a densely U-nets refine network for facial landmark localization, which is composed of DU-Net and Refine-Net. The DU-Net model consists of three DU-Net cascades, each of which contains four multi-scale intermediate supervisions. The DURN model uses MobileNetv3 as the backbone to extract features, performs feature fusion through PAN and SSH, performs multi-scale prediction on three scales (1/8, 1/16, 1/32), and adds integral regression to obtain the face box and key point coordinates.

4.2. Experimental Results

In this subsection, the effectiveness of the proposed method is evaluated in terms of model convergence and compared with baselines.

4.2.1. Convergence Analysis of the MSFLD

To observe the convergence of MSFLD, the training process was analyzed. In the experiment, we set parameter initialization. The task is to detect 23 key points of faces from the HNUFDD dataset. During the training process, the Adam optimizer was used to update the parameters, the batch size was set to 96, the initial learning rate was 0.0001, and a total of 300 epochs were trained. Figure 5 illustrates the curve of the training loss with respect to the number of epochs. The curve becomes flat as the number of training epochs increases. Starting from the 64th epoch, the train loss of MSFLD stabilized. This result indicates the convergence of MSFLD.

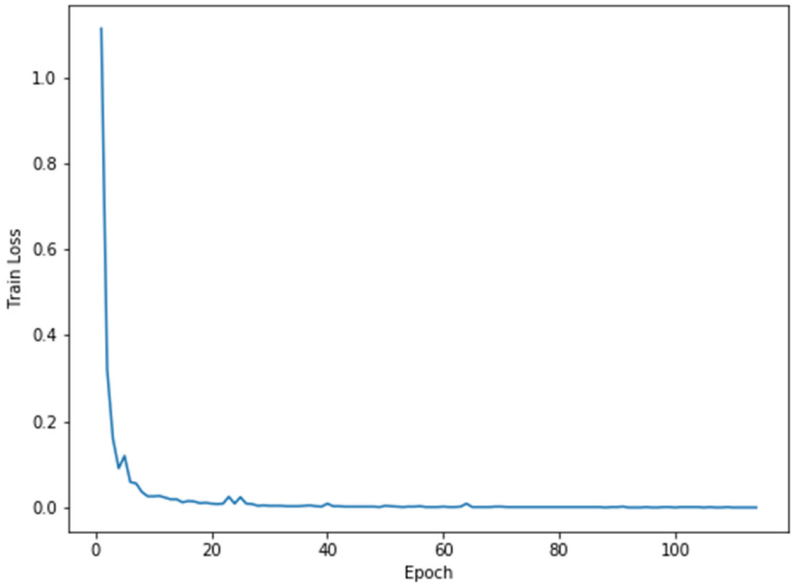


Figure 5. Training process of MSFLD on the HNUFDD dataset.

4.2.2. Ablation Study of the MSFLD

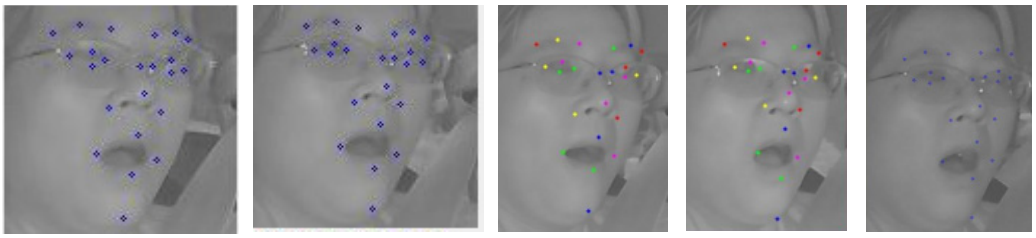
We conducted experiments on the HNUFDD dataset using different configurations of Bottleneck, Inverted bottleneck and fully connected layers of different scales. The experimental results are shown in Table 2. As can be seen from Table 2, the size of the model is reduced by 0.5MiB by replacing Bottleneck with Inverted bottleneck, indicating that the model reduces the amount of parameters; on the basis of Inverted bottleneck, a 56×56 and 28×28 multi-scale fully connected layer is added, that is, increasing from 3 scales to 5 scales, the NME value is reduced by 0.9327%, indicating that the localization accuracy of the facial landmark detector is improved. This result demonstrates the validity of the MSFLD model design.

Table 2. Effectiveness of using different configuration in the MSFLD model on the HNUFDD dataset.

Formulation	NME (%)	Model Size (MiB)
Bottleneck + S1 + S2 + S3	6.4803	6.6
Inverted bottleneck + S1 + S2 + S3	6.3845	6.1
Inverted bottleneck + S1 + S2 + S3 + S4 + S5 (MSFLD)	5.4518	6.2

4.2.3. Facial Key Point Detection

To verify the performance of the proposed MSFLD, experiments on facial key point detection were carried out on the HNUFDD dataset. The detection results with 23 key points are shown in Figure 6 and Table 3 shows the comparison results on the HNUFDD dataset for facial key point detection.



MSFLD PFLD Retina Face DURN MTCNN

Figure 6. Detection results of 23 key points using various methods on the HNUFDD dataset.

Table 3. Comparison of different methods on the HNUFDD dataset for facial key point detection.

Method	NME (%)	Model Size (MiB)	Excution Time (s)
MTCNN [17]	9.0951	1.97	10.72
Retina Face _Resnet50 [18]	5.7063	104.7	8.18
PFLD [19]	6.4803	6.6	4.29
DURN_ Mobilenetv3 [20]	9.1648	3.6	6.02
MSFLD	5.4518	6.2	4.82

As can be seen from Table 3, the NME value using the MSFLD is 5.4518%, which is lower than that of other methods. This result may benefit from using the inverse residual blocks instead of the traditional convolution operations and increasing the number of multi-scale fully connected layers. Such results indicate that the MSFLD is suitable for face key point detection. For example, the NME value using the PFLD model is 6.4803%, and that using the MTCNN model is 9.0951%. The NME of the proposed method is 1.0285% lower than the former and 3.6433% lower than the latter. The NME value using the Retina_Resnet50 model is 5.7063%, and the model size is 104.7MiB. The NME value and model size value of our proposed MSFLD model are 0.2545% lower, and the model size value 98.5MiB lower than those of the Retina_Resnet50 model, respectively. To measure the feasibility of the proposed MSFLD method, the execution time is calculated by detecting images in the test set. As can be seen from Table 3, the MSFLD method has low execution time for facial key points detection. Such results indicate that the proposed MSFLD is more effective than state-of-the-art methods in real scenes for face key point detection.

4.2.4. Fatigue Driving Recognition

To evaluate the performance of the proposed fatigue driving recognition method, this part carried on the experiments from aspects of parameter sensitivity and threshold combination.

This experiment is to explore the influence of adaptive threshold with changing p value on fatigue driving decision performance. The adaptive threshold of the video is obtained by calculating the eye aspect ratio and mouth aspect ratio of the first p frames of each test video and then taking the average value. Table 4 shows the results with different adaptive

thresholds on the HNUFDD dataset for fatigue driving behavior recognition. When p is 30, the accuracy rate is up to 90.4624%, which was higher than that of other p values.

Table 4. Comparison with different adaptive thresholds on the HNUFDD dataset for fatigue driving behavior recognition.

p	Accuracy (%)
30	90.4624
35	89.5954
40	89.0173
50	87.8613

This experiment investigates the impact of different statistical thresholds on fatigue driving decision performance. The statistical thresholds are formed as follows: in the dataset with 23 key points manually marked, the average values of the eye aspect ratio and mouth aspect ratio of 4616 images under the three driving behaviors of normal, mouth closed and talking were 0.3729 and 0.4559, respectively. Then, the average eye aspect ratio of the 1833 images under the dozing driving behavior was 0.2452; the average mouth aspect ratio of the 1539 images under the yawning driving behavior was 0.8299. Lastly, the statistical thresholds of the eyes and the mouth were 0.3091 and 0.6429, respectively. Table 5 shows the recognition accuracy of fatigue driving behavior when EAR and MAR take different values. When EAR = 0.3091 and MAR = 0.6429, the accuracy rate is 94.5087%, which is higher than that of other EAR and MAR values.

Table 5. Comparison with different statistical thresholds on the HNUFDD dataset for fatigue driving behavior recognition.

EAR	MAR	Accuracy (%)
0.2452	0.8299	83.5260
0.3729	0.4559	44.5087
0.3091	0.6429	94.5087

This experiment discusses the performance of the proposed method with the different combination thresholds. The driver's eye state can be judged in accordance with Equation (10). EAR_{ct} is set to the maximum value of adaptive threshold and the statistical threshold. When the aspect ratio of the driver's eye is lower than EAR_{ct}, the driver's eyes is considered closed at this moment. If it is higher than EAR_{ct}, it is considered that the driver's eyes are open at this moment. Similarly, the driver's mouth state is judged in accordance with Equation (11). The combined threshold of mouth MAR_{ct} was set to the minimum of the adaptive threshold and the statistical threshold. When the aspect ratio of the mouth is higher than MAR_{ct}, driver's mouth is considered as yawning at this moment. When it is lower than MAR_{ct}, it is considered that the driver's mouth is closing or talking at this moment. In more than 55% of the images in the test video, if the eye aspect ratio is less than the set threshold EAR_{ct}, or the mouth aspect ratio is greater than the set threshold MAR_{ct}, the video is judged to have fatigue driving. Different combination thresholds are obtained when p, EAR, and MAR take different values. The recognition accuracy of fatigue driving behavior on the HNUFDD dataset is shown in Table 6. When p = 30, EAR = 0.3091 and MAR = 0.6429, the recognition accuracy of the proposed method with the combined threshold is 99.1329, which is higher than that of other combined thresholds.

Table 6. Comparison with different combination thresholds on the HNUFDD dataset for fatigue driving behavior recognition.

p	EAR	MAR	Accuracy (%)
30	0.3091	0.6429	**99.1329**
35	0.3091	0.6429	97.9769
40	0.3091	0.6429	97.1098
50	0.3091	0.6429	95.9538

This experiment is to study the impact of the proposed method with different threshold strategies on the performance of fatigue driving recognition. As can be seen from Table 6, the best result is obtained when $p = 30$, EAR $= 0.3091$ and MAR $= 0.6429$. In this experiment, the methods of fatigue driving recognition with adaptive threshold, statistical threshold and combined threshold are conducted according to this set of values. Table 7 shows the comparison of the fatigue driving recognition methods with different thresholds on the HNUFDD dataset. The result shows that the accuracy of the method using combined threshold is higher than the other two methods.

Table 7. Comparison with different thresholds on the HNUFDD dataset for fatigue driving behavior recognition.

Threshold Strategy	Accuracy (%)
Adaptive threshold	90.4624
Statistical threshold	94.5087
Combination threshold	**99.1329**

This experiment is conducted to compare the performance of the proposed method with existing methods for fatigue driving recognition. Table 8 shows the comparison results on the HNUFDD dataset for fatigue driving behavior recognition. The MSFLD achieves the accuracy of 99.1329%, which is higher than that of other four methods. Such results indicate that the proposed MSFLD is more effective than the state-of-the-art method in real scenes for fatigue driving behavior recognition.

Table 8. Comparison with existing methods on the HNUFDD dataset for fatigue driving behavior recognition.

Research	Methodology	Accuracy (%)
Liu et al. [10]	MTCNN	68.2081
Deng et al. [18]	Retina Face _Resnet50	82.3699
Guo et al. [19]	PFLD	73.4104
Liu et al. [20]	DURN_ Mobilenetv3	67.9191
Proposed method	MSFLD	99.1329

5. Conclusions

In this study, we have presented a fatigue driving recognition method based on the MSFLD. The proposed method is composed of face region detection, facial key points detection, parameter matrix construction, and fatigue driving decision. The MSFLD method based on deep learning are proposed to adaptively detect facial key points. For fatigue driving decision, the method of combining with adaptive threshold and statistical threshold is proposed to avoid misjudgment of fatigue driving in the real scenario. The propsed MSFLD method achieves NME of 5.4518 for facial 23 key points detection, and the proposed fatigue driving recognition method obtains an accuracy of 99.1329% on the HNUFDD dataset. Thus, the proposed method based on the MSFLD improves the performance of fatigue driving recognition while reducing the workload of labeling facial key points.

In our future work, multimodal data (i.e., head posture, vehicle driving speed, acceleration) will be considered to expand the fatigue parameter features to enhance the robustness of the system. Based on these data features, the driving behavior decision method can be improved to meet the fatigue driving detection accuracy in complex scenes.

Author Contributions: Data curation, B.S.; Funding acquisition, H.L.; Investigation, Z.M. and B.S.; Methodology, Z.M.; Project administration, H.L.; Software, W.X. and C.S.; Writing—original draft, W.X.; Writing—review & editing, W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 61971182 and 62173133) and by Natural Science Foundation of Hunan Province (Grant No. 2021JJ30145).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. Global status report on road safety 2013: Supporting a decade of action. *Inj. Prev.* **2013**, *15*, 286.
2. Road Safety in Canada. Available online: https://www.tc.gc.ca/ (accessed on 24 March 2022).
3. Azam, K.; Shakoor, A.; Shah, R.A.; Khan, A.; Shah, S.A.; Khalil, M.S. Comparison of fatigue related road traffic crashes on the national highways and motorways in Pakistan. *J. Eng. Appl. Sci.* **2014**, *33*, 47–54.
4. AAA Foundation for Traffic Safety. Available online: https://www.aaafoundation.org (accessed on 10 January 2022).
5. Fatigue. Available online: https://ec.europa.eu/transport/roadsafety/ (accessed on 21 January 2022).
6. Abtahi, S.; Omidyeganeh, M.; Shirmohammadi, S.; Hariri, B. YawDD: A Yawning Detection Dataset. In Proceedings of the ACM Multimedia Systems, Singapore, 19 March 2014; pp. 24–28. [CrossRef]
7. Yang, H.; Liu, L.; Min, W.; Yang, X.; Xiong, X. Driver Yawning Detection Based on Subtle Facial Action Recognition. *IEEE Trans. Multimed.* **2021**, *23*, 572–583. [CrossRef]
8. Köstinger, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Annotated Facial Landmarks in the Wild: A large-scale, real-world database for facial landmark localization. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Barcelona, Spain, 6–13 November 2011; pp. 2144–2151. [CrossRef]
9. Savaş, B.K.; Becerikli, Y. Real Time Driver Fatigue Detection System Based on Multi-Task ConNN. *IEEE Access* **2020**, *8*, 12491–12498. [CrossRef]
10. Liu, W.; Tang, M.; Wang, C.; Zhang, K.; Wang, Q.; Xu, X. Attention-guided Dual Enhancement Train Driver Fatigue Detection Based on MTCNN. In Proceedings of the International Academic Exchange Conference on Science and Technology Innovation (IAECST), Guangzhou, China, 10–12 December 2021; pp. 1324–1329. [CrossRef]
11. Salem, E.; Hassaballah, M.; Mahmoud, M.M.; Ali, A.M.M. Facial Features Detection: A Comparative Study. In Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2021), Settat, Morocco, 28–30 June 2021; pp. 402–412.
12. Khabarlak, K.; Koriashkina, L. Fast facial landmark detection and applications: A survey. *J. Comput. Sci. Technol.* **2022**, *22*, 12–41. [CrossRef]
13. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 1867–1874. [CrossRef]
14. Hassaballah, M.; Bekhet, S.; Rashed, A.A.M.; Zhang, G. Facial Features Detection and Localization. *Recent Adv. Comput. Vis.* **2019**, *804*, 33–59. [CrossRef]
15. Kansizoglou, I.; Misirlis, E.; Tsintotas, K.; Gasteratos, A. Continuous Emotion Recognition for Long-Term Behavior Modeling through Recurrent Neural Networks. *Technologies* **2022**, *10*, 59. [CrossRef]
16. Sun, Y.; Wang, X.; Tang, X. Deep Convolutional Network Cascade for Facial Point Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3476–3483. [CrossRef]
17. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Proc. Let.* **2016**, *23*, 1499–1503. [CrossRef]
18. Deng, J.; Guo, J.; Zhou, Y. Retinaface: Single-Stage Dense Face Localisation in the Wild. Available online: https://arxiv.org/abs/1905.00641 (accessed on 20 May 2022).
19. Guo, X.J.; Li, S.Y.; Yu, J.K. PFLD: A Practical Facial Landmark Detector. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 1–11. [CrossRef]
20. Liu, Y. Grand Challenge of 106-Point Facial Landmark Localization. In Proceedings of the IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China, 8–12 July 2019; pp. 613–616. [CrossRef]

21. Hassaballah, M.; Salem, E.; Ali, A.M.M.; Mahmoud, M.M. Deep recurrent regression with a heatmap coupling module for facial landmarks detection. *Cogn. Comput.* **2022**, 1–15. [CrossRef]
22. Sikander, G.; Anwar, S. Driver Fatigue Detection Systems: A Review. *IEEE Trans. Intell. Transp.* **2019**, *20*, 2339–2352. [CrossRef]
23. Portouli, E.; Bekiaris, E.; Papakostopoulos, V.; Maglaveras, N. On-road experiment for collecting driving behavioural data of sleepy drivers. *Somnology* **2007**, *11*, 259–267. [CrossRef]
24. Yang, Z.; Ren, H. Feature Extraction and Simulation of EEG Signals During Exercise-Induced Fatigue. *IEEE Access* **2019**, *7*, 46389–46398. [CrossRef]
25. Chui, K.T.; Tsang, K.F.; Chi, H.R.; Ling, B.W.; Wu, C.K. An Accurate ECG-Based Transportation Safety Drowsiness Detection Scheme. *IEEE Trans. Ind. Inform.* **2016**, *12*, 1438–1452. [CrossRef]
26. Tsuchida, A.; Bhuiyan, M.S.; Oguri, K. Estimation of drowsiness level based on eyelid closure and heart rate variability. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009; pp. 2543–2546. [CrossRef]
27. Balasubramanian, V.; Adalarasu, K. EMG-based analysis of change in muscle activity during simulated driving. *J. Bodyw. Mov. Ther.* **2007**, *11*, 151–158. [CrossRef]
28. Yang, J.H.; Mao, Z.H.; Tijerina, L.; Pilutti, T.; Coughlin, J.F.; Feron, E. Detection of Driver Fatigue Caused by Sleep Deprivation. *IEEE Trans. Syst. Man Cybern.* **2009**, *39*, 694–705. [CrossRef]
29. Lee, B.G.; Chung, W.Y. Driver Alertness Monitoring Using Fusion of Facial Features and Bio-Signals. *IEEE Sens. J.* **2012**, *12*, 2416–2422. [CrossRef]
30. Li, K.; Gong, Y.; Ren, Z. A Fatigue Driving Detection Algorithm Based on Facial Multi-Feature Fusion. *IEEE Access* **2020**, *8*, 101244–101259. [CrossRef]
31. Du, G.; Li, T.; Li, C.; Liu, P.X.; Li, D. Vision-Based Fatigue Driving Recognition Method Integrating Heart Rate and Facial Features. *IEEE Trans. Intell. Trans.* **2021**, *22*, 3089–3100. [CrossRef]
32. Raja, M.S.; Manu, V.S.; Reshma, D. A Real-time Fatigue Detection System using Multi-Task Cascaded CNN Model. In Proceedings of the IEEE International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 24–25 April 2021; pp. 674–679. [CrossRef]
33. Jia, H.; Xiao, Z.; Ji, P. Fatigue Driving Detection Based on Deep Learning and Multi-Index Fusion. *IEEE Access* **2021**, *9*, 147054–147062. [CrossRef]
34. Hao, Z.; Li, Z.; Dang, X.; Ma, Z.; Liu, G. MM-LMF: A Low-Rank Multimodal Fusion Dangerous Driving Behavior Recognition Method Based on FMCW Signals. *Electrcnics* **2022**, *11*, 3800. [CrossRef]