Pilot facial feature recognition and fatigue classification method under high exposure environment

Huining Pei[1], **Jiali Sun**[1,2,3]

[1]College of Architecture and Art Design, Hebei University of Technology, Tianjin 300401, China

[2]sunjialihebut@163.com
[3]corresponding author

Abstract. To improve the recognition efficiency of pilot fatigue detection in high exposure environment, a facial feature recognition and fatigue classification method for pilots is proposed. Firstly, the high exposure images are pre-processed with frequency domain light normalization; secondly, a human face detector is trained based on YOLOv5s network model and detects the face, eyes and mouth; again, the fatigue feature values of eyes and mouth are obtained by fusing PERCLOS fatigue detection algorithm; finally, a fatigue classifier is built by RBF-ELM neural network and the fatigue parameters of eyes, mouth and yawning The fatigue parameters of the eye, mouth and yawn features were input to the classifier for fatigue classification. The experimental results show that the fatigue detection speed of the proposed method is up to 310 FPS when computed with GPU, which has higher recognition rate and better robustness compared with the classical PCA, CNN and DCT methods. The method can effectively improve the accuracy of pilot facial feature recognition and eye localization in high exposure environment, and has high practicality for non-contact pilot fatigue detection.

Keywords: pilot, facial feature recognition, fatigue detection, image pre-processing, YOLOv5 neural network.

1. Introduction

Lack of sleep during flight, long flight time and bad weather may cause brain fatigue in pilots due to high attention span for a long time, resulting in a decrease in decision making power, attention and reaction speed, which may lead to mishandling or misjudgment and become a potential safety hazard during flight.[1] This is a potential safety hazard during flight. Therefore, detects and evaluates the pilot's fatigue status as an important part of the mission.

Some classical convolutional neural networks, such as Lenet, Alexnet, Googlenet, VGGnet and Resnet, are widely used in the field of computer vision and have been applied to facial recognition technology by many researchers. GoogleNet, YOLO treats object detection as a regression problem rather than a classification problem, and uses a single neural network to perform all the basic steps to detect objects. It is a deep neural network-based target recognition and localization algorithm, which can be applied to pilot facial feature detection because of its fast operation, high detection accuracy, and ability to be applied to real-time systems.[2]

The tendency to sleep is one of the main signs of fatigue, and when a person is fatigued, the process of opening to closing and then opening the eyes takes a long time, and the longer the duration of eye closure, the more severe the fatigue.[3] PERCLOS (Percentage of Eyelid Closure Over the Pupil Over Time) measures the degree of fatigue by the proportion of the total time that the human eye is closed for a certain period of time. Compared with other methods, PERCLOS is a more visual and accurate reflection of driver fatigue. [4] In addition, the complex environment faced by pilots, in the case of the illumination of the cap is significantly affected by the angle between the sunlight and the flight path, which leads to a certain decrease of specificity in the detection of facial fatigue caused by the illumination. As a matter of fact, it is found that most of the flight accidents occur in clear daylight, which indicates the need for fatigue detection research under high exposure environment [5]. Furthermore, most of the existing studies only compare and analyze the two brain state characteristics of pilots' wakefulness and fatigue [6] , and only a few studies quantitatively assess the level of fatigue for pilots. Among them, Ahlstrom et al. [7] classified pilots' fatigue into three levels based on the Karolinska Sleepiness Scale (KSS): awake, fatigued, and severely fatigued; Larue et al. [8] used the pilot's alertness and the number of microsleeps as evaluation indicators to classify the pilot's fatigue status into four levels. However, the above studies seldom considered the influence of environmental factors on the fatigue state recognition results, and there is a problem that the classification of fatigue levels is relatively crude.

In summary, a method for facial feature recognition and fatigue classification of pilots in high-exposure environments is proposed. First, the high exposure facial images are pre-processed using image processing algorithms; second, the features of human face, eyes and mouth are recognized using the neural network YOLOv5; finally, the fatigue fusion results are judged and graded using the fatigue detection algorithm to assess the fatigue status of the pilot. To solve the problem of pilot fatigue judgment and grading under the interference of high intensity light outside the cockpit during aircraft takeoff and landing, and provide theoretical reference for the early warning of pilot fatigue.

2. Pilot facial feature recognition

2.1. Construction and processing of facial detection dataset

Since the pilot working images are confidential and not easily available, and the car driver facial dataset is similar to the airliner pilot driving scenario, which has negligible impact on the network training results, the public facial image dataset Celeba is used to train the network model. In addition, self-built datasets are used to expand the number of datasets and improve the generalization ability of the model.

2.1.1. Facial dataset construction

First, the high exposure images were uniformly processed by Photoshop to simulate the facial state of pilots under strong lighting environment; second, the high exposure images were normalized to normal lighting images by frequency domain lighting normalization and input to YOLOv5 network for feature recognition to improve the facial edge contour saliency and facial feature recognition rate. The dataset contains male and female facial images aged 18-70 years (12,190 images from Celeba and 3,100 images from the self-built database); 70% were randomly selected as the training set and 30% as the test set. Some of the dataset images are shown in Fig. 1.

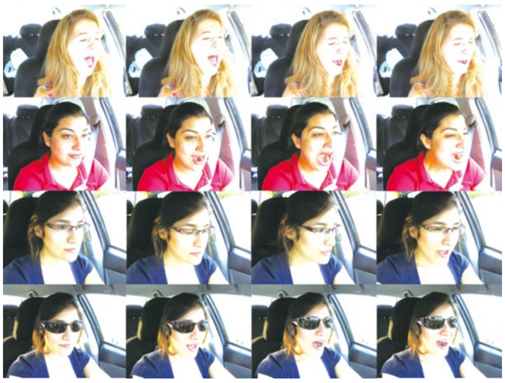


Figure 1. Images of some data sets

2.1.2. Frequency domain light normalization

In order to reduce the effect of illumination variation on pilot facial feature recognition, illumination preprocessing is required for facial image datasets, and there are two main methods for processing facial images under different illumination conditions. One normalizes the illumination conditions in the image and homogenizes the illumination in the image by weakening the differences. Such as histogram equalization [9] Gamma correction [10] DCT (Discrete Cosine Transform) [11] etc. Another method is the direct elimination of the illumination part of the image to remove the image caused by light recognition, such as Rstinex algorithm [12] The Rstinex algorithm, frequency-domain light normalization, and single-scale SSR and multi-scale MSR algorithms are generated according to the scale.

With reference to existing studies, it is known that frequency-domain light normalization can make the images acquired under any lighting conditions identical to those in the training library after normalization, while retaining the distinguishability of the face. Therefore, using the property that the phase frequency characteristics are independent of the illumination conditions, the frequency-domain light normalization method in the filtering processing method is selected to complete the pilot facial feature recognition. On the other hand, because the distinguishability between faces is small, the minimum non-zero feature vector is selected as the facial feature, and the results of the analysis of the overall illumination changes (brightening and darkening) and local area illumination changes of the face show that the algorithm is robust to illumination. The processing steps of illumination pre-processing are as follows.

First, the frequency-domain light normalization is applied to the sample images in the training set for the consistency of amplitude-frequency characteristics. Experimental simulations show that the frequency-domain light normalization facial recognition method is robust to light variations compared with the traditional method, as shown in the following procedure.

The Fourier transform of the real signal 2D image *f*(*x*,*y*) is defined as

$$F(u,v)=\frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{n-1}f(x,y)exp\,[-j2pai\,(ux+vy)/N] \tag{1}$$

After the image Fourier transform, it is a complex signal in the frequency domain with its real part *R(u,v)* and imaginary part *I(u,v),* respectively.

$$R(u,v)\frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{n-1}f(x,y)cos\,[2pai\,(ux+vy)/N] \tag{2}$$

$$I(u,v)\frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{n-1}f(x,y)sin\,[-2pai\,(ux+vy)/N] \tag{3}$$

The complex signal of the Fourier transform of the image is expressed in polar coordinates to obtain the image amplitude-frequency characteristic |F(u,v)| and the phase-frequency characteristic *φ*(u,v).

$$|F(u,v)|=\sqrt{R^2(u,v)+I^2(u,v)} \tag{4}$$

$$\varphi(u,v)=arcta\,n\left[\frac{R(u,v)}{I(u,v)}\right] \tag{5}$$

The image recognition rate in the facial feature recognition system is greatly affected by the lighting conditions. With the change of lighting conditions, the relationship between the acquired facial image *f*(*x, y*) and the facial image *f*(*x, y*) under normal ambient lighting is as follows, and the relationship between the amplitude and phase frequency characteristics of the facial image *f*(*x, y*) and *f*(*x, y*) can be obtained from Eqs. (1) to (5) as Eq. (6)

$$f(x,y) = kf(x,y) \tag{6}$$

Where, when k>1, it means that the facial image *f*(*x, y) becomes stronger* overall relative to the lower image *f*(*x, y*) under normal lighting conditions; when 0<k<1, it means that the facial image *f*(x, y) becomes weaker overall compared to the lower image *f*(*x, y*) under normal lighting conditions.

From equation (7), it can be seen that when the illumination conditions of the facial image change, the amplitude-frequency characteristics also shift proportionally, while the phase-frequency characteristics remain unchanged. Based on the light-independent nature of the phase frequency characteristics of the image, the frequency-domain light normalization is used for facial feature recognition of pilots.

$$\{|F(u,v)| = k|(u,v)| \tag{7}$$

$$\phi(u,v) = \varphi(u,v) \tag{8}$$

Let the facial training set consisting of *M* facial samples bex_i , the solution of the mean value of the facial training set samples *m* is as in equation (9).

$$m = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{9}$$

The amplitude-frequency characteristics of the Fourier transform $|F_m(u,v)|$ as the amplitude-frequency characteristics of the face after light normalization, and the phase-frequency characteristics of the face as the phase-frequency characteristics after light normalization, constituting a complex signal in the frequency domain of the face image after light normalization of the face. The Fourier inverse transform of this signal is performed to obtain the light normalized face.

$$\widehat{x_i} = F^{-1}\{|F_m(u,v)|\angle\phi_i(u,v)\} \tag{10}$$

where the different facesx_i , theF_m is the amplitude frequency characteristic of the face, andφ_i is the phase frequency characteristic of the face. From equation (10), it can be seen that the amplitude-frequency characteristics of the face after normalization of illumination are consistent with the amplitude-frequency characteristics of the sample mean of the training set, so that the normalization of illumination of the face can be achieved while preserving the invariance of the phase frequency characteristics and thus ensuring the distinguishability of the face.

2.2. Pilot face detection based on YOLOv5 network

Accurate recognition of facial areas in the data set is the first step in non-contact fatigue detection, which requires recognition of the presence or absence of facial images in the image, and if so, determination of facial areas by judging facial features and contour information.[13] . YOLOv5s is more suitable for pilot fatigue detection because of its real-time, high accuracy and small size, and other versions of the network are widened and deepened on the basis of YOLOv5s, which can be easily modified through the configuration file, so the YOLOv5s network model is trained for pilot facial feature detection.

2.2.1. YOLOv5 network structure

The network structure of YOLOv5 is shown in Fig. 2. The network structure is mainly divided into 4 parts: Input, Base Backbone Network, Neck Network and Head Output.

(1) Inputs

The input side represents the image to be input, and at the input side the original image is often intervened, i.e., the input image is deflated to the input size of the grid, and a normalization operation is performed, and finally converted to $640 \times 640 \times 3$ tensor into the network.

(2) Benchmark Backbone Backbone Network

The CSPDarknet53 structure is used in YOLOv5, and the BottleneckCSP called Conv module in the previous version is removed, and the improved BottleneckCSP is called C3 module. C3 module concentrates the gradient changes in the feature map from beginning to end, which enables the YOLOv5 network to have a good learning capability, and to lightweight while maintaining good accuracy, while reducing computational bottlenecks and memory costs. There are two types of C3 modules in YOLOv5, which are classified as C3-False and C3-True, based on the presence or absence of residual edges.

YOLOv5 introduces the Focus module in the Backbone backbone network, which implements 32-fold downsampling of the input image with 4 Conv modules. the Focus module performs a slicing operation on the input image, taking a value every other pixel. Similar to the neighborhood downsampling, the original image data is sliced into 4 copies of data and stitched in the channel dimension, and finally the convolution operation is performed. This process concentrates the information in the image length and width dimensions into the channel dimension, so no information is lost. the Focus module reduces the cost of convolution and cleverly implements downsampling and increases the channel dimension by resetting the tensor dimension.

(3) Neck network

The Neck network is usually located in the middle of the benchmark Backbone network and Head network, using Neck network can further improve the diversity and robustness of the features, and better utilize the features extracted from the Backbone network, SPP (spatial pyramid pooling) module, FPN (feature pyramid network) module and PAN (path aggregation network) module are used in YOLOv5 to complete the information transfer fusion. feature pyramid network) module and PAN (path aggregation network) module are used in YOLOv5 to complete the information transfer and fusion.

(4) Outputs

The Head output layer mainly uses the previously extracted features to make predictions.

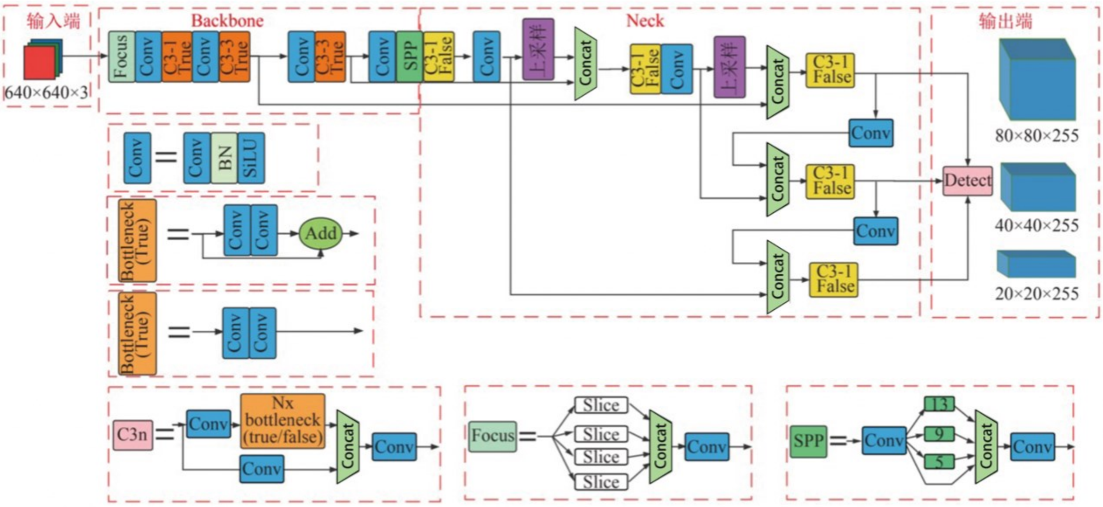


Figure 2. YOLO network structure diagram

2.2.2. *Face detection*

The YOLOv5 network model is used for face detection, and the size of the input image is 416*416. After convolution and batch normalization operations on the input image, the input image is sampled 3 times, 32 times, 16 times and 8 times to obtain the multi-scale feature map. After 32 down-samplings, the feature maps are too small, so YOLOv5 uses up-sampling with a step size of 2 to double the size of the resulting feature maps, i.e., 16 times the down-sampling. Similarly, a feature map with 16 samples is sampled in steps of 2 and a feature map with a sample size of 8 times is obtained for detection using depth features. Detecting targets at different scales is a challenge, especially for small targets. Feature Pyramid Network (FPN)[8] is a feature extractor designed to improve the accuracy and extraction speed feature extractor, which replaces the feature extractor in the detector (e.g., faster R-CNN) and generates higher quality feature map pyramids. The training process for facial detection in the YOLOv5 network is divided into three main basic steps.

Step 1: Determine the training sample set

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \tag{11}$$

where *x is the training sample*, *y is the* training sample label (-1 is the non-facial sample and 1 is the facial sample), and *n* is the number of samples.

Step 2: Initialization of weights, each sample is initialized to the same weight, which is calculated by equation (12).

$$\omega_{1,i} = \frac{1}{n}, i = 1,2, \dots, n \tag{12}$$

where$\omega_{1,i}$ is the weight value at the first iteration.

Step 3: Let the number of iterations be$m = 1,2, \dots, M$, the formula for calculating the weights of the weights at the *mth iteration* is equation (13).

$$e_{m,i} = \frac{w_{m,i}}{\sum_{j=1}^{n} w_{m,j}} \tag{13}$$

where, *m* is the number of iterations and *e* is the weight value at the time of iteration. An example of facial recognition training results is shown in Fig. 3.

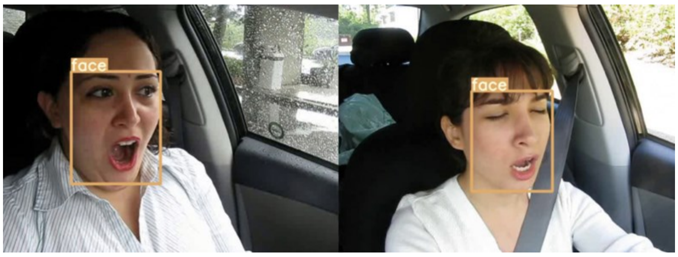
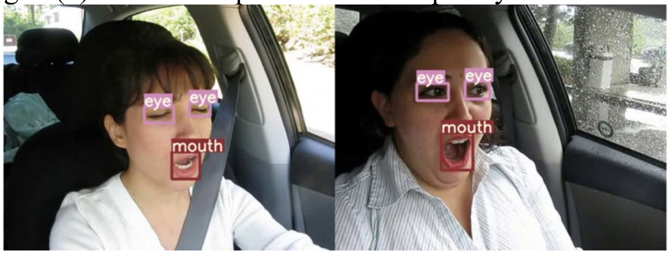


Figure 3. Training results of face detection test

2.2.3. *Eye and mouth detection*

When a person is fatigued, the magnitude of the change of the eyes and mouth is more obvious.[14] The driver's mental state is judged by the state of the eyes and mouth in consecutive frames, which can be tracked while detecting the state of the eyes and mouth features in each frame. If the eyes are not opened for several frames after closing, the state of fatigue is judged. Therefore, when the degree of eye closure of a person is less than a certain threshold and the person's mouth feature is greater than a certain threshold, the person is judged to be in a fatigue state.

According to the a priori knowledge of "three chambers and five eyes" of human face, we can deduce that the eye area is located in the upper 3/5 of the face and the mouth area is located in the lower 1/5 of the face. Based on this facial geometry, we can quickly locate the eye and mouth regions and further reduce the computational effort of the detection algorithm.[15] This further reduces the computational effort of the detection algorithm. Fast facial detection using YOLOv5[16] The Haar wavelet algorithm [17] Based on the Haar wavelet algorithm, the driver's face is pre-processed to detect and locate the position of the human eyes and mouth. Based on the location of facial feature points to determine the eye and mouth feature frame is shown in the boxed area in Fig. 4, Fig. 4 (a) shows the open mouth and closed eye state, and Fig. 4 (b) shows the open mouth and open eye state.



(a) Open-mouth, closed-eye state (b) Open-mouth, open-eye state

Figure 4. Eye and mouth detection tracking results

The most obvious and important aspect of facial fatigue information is the degree of eye opening and closing.[18] The driver's eyes are naturally open when awake. When the driver is awake, the eyes are naturally open; if the driver is moderately fatigued, the eyes are closed for a longer period of time with frequent blinking; if the driver is fatigued, the eyes are closed. PERCLOS) is currently the most effective fatigue detection algorithm recognized internationally, which indicates the proportion of human eye closure time to total time within a certain period of time. It has three judgment criteria, EM, P70 and P80, of which P80 is the best [19] , that is, 80% of eye closure means in a fatigue state, and the calculation formula is shown in equation (14). Therefore, the fatigue detection algorithm based on PERCLOS fatigue detection uses the P80 judgment criterion, and the eye opening and closing state is usually judged by the eye aspect ratio *r. The* eye aspect ratio is shown in Fig. 5, and the calculation formula is shown in Equation (15).

$$P = \frac{m}{n} \times 100\% \tag{14}$$

where$P$ is the proportion of time spent with the eyes closed, and$m$ is the number of eye closure frames, and$n$ is the total number of frames during the detection time.

$$r = \frac{h}{\omega} \tag{15}$$

where h andω are the height and width of the eye, respectively.

Although the PERCLOS eye fatigue value can reflect the degree of fatigue, the accuracy of determining fatigue is greatly improved by fitting it with other fatigue judgment characteristics. Since the mouth opening and closing state is also one of the evaluation factors of fatigue [20] , so on the other hand, the generalized use of PERCLOS calculation rules selects the mouth yawning feature as another feature for judging the fatigue status of pilots. As shown in Fig. 6, the mouth aspect ratio k *is used* as the basis for determining the mouth opening and closing, and the calculation formula is shown in equation (16).

$$k = \frac{a}{b} \tag{16}$$

where *k* is the mouth height to width ratio, and *a* and *b* are the height and width of the mouth, respectively.

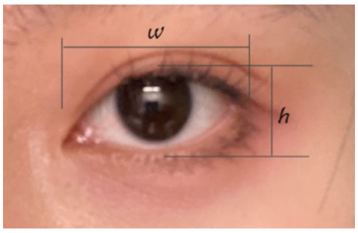


Figure 5. Eye height to width ratio

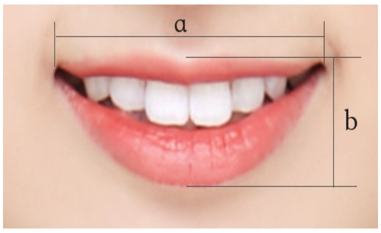


Figure 6. Mouth height to width ratio

3. Pilot fatigue classification

3.1. Pilot fatigue detection

Based on the PERCLOS judgment criteria, the fatigue judgment formula for the percentage of pilot's eyes closed and mouth open time can be expressed as follows.

$$\frac{\omega}{t} \geq \vartheta \tag{17}$$

style.ω the time for which the eye is closed more than 80% of the time, and$t$ is the detection unit time, and ϑ is the threshold.

During the training, whenϑ 0.48 is taken, when the detection value$\geq$ 0.48, it is judged that the fatigue state appears in the image of the frame, and for the mouth action feature, when the yawning state appears in several consecutive frames (3-4s), it is judged as pilot fatigue.

The experiments segmented the long video in the database into video clips of 10 consecutive frames and used the ratio of the number of frames in which the fatigue state appeared to determine the factual fatigue state of the pilots, the input variables used PERCLOS value and the number of yawns, and the parameter fuzzification of the fatigue level is shown in Table 1. Where, when the PERCLOS value< 0.15, or the number of yawning frames< 0.5, the personnel is considered to be in the awake state; when the PERCLOS is between 0.15 and 0.3, or the number of yawn frames is between 0.5 and 3, it is judged to be mildly fatigued; when the PERCLOS value> 0.3, or the number of yawning frames> 3, it is judged as fatigue state.

Table 1. Parameter fuzzification of fatigue level

	Sobriety	**Mild fatigue**	**Fatigue**
Peclos	< 0.15	[0.15,0.3]	> 0.3
Yawn/Frame	< 0.5	[0.5,3]	> 3

3.2. Fatigue classification based on RBF-ELM neural network

The RBF-ELM (Radial Basis Function-Extreme Learning Machine) neural network extends the extreme learning machine (ELM) to the radial basis function (RBF) network proposed by Huang et al.[21] The RBF-ELM uses the RBF network structure (i.e., three-layer network structure) and the ELM learning model (i.e., randomly generating weighted parameters from the input to the hidden layer, and then solving the weighted parameters from the hidden layer to the output layer by the analytical solution).[19] . Therefore, it has not only the general nonlinear approximation ability of RBF network, but also the fast learning ability of ELM. the learning algorithm of RBF-ELM is to randomly select the center and width of RBF kernel in the hidden layer, and solve the weights between the hidden layer and the output layer by the analytical solution, therefore, using RBF-ELM for the establishment of pilot fatigue classifier can output more accurate fatigue classification results. For the facial features of pilots in fatigue state, including blinking, yawning and other fatigue features, the parameter changes can be detected to determine whether the pilot is fatigued or not. The system uses PERCLOS value [19] and blink frequency as the main parameters for detecting fatigue, and the mouth yawning feature as a secondary parameter to achieve fatigue grading, and the neural network tool in MATLAB is used to build a 3-layer neural network structure [21] , as shown in Fig. 7.

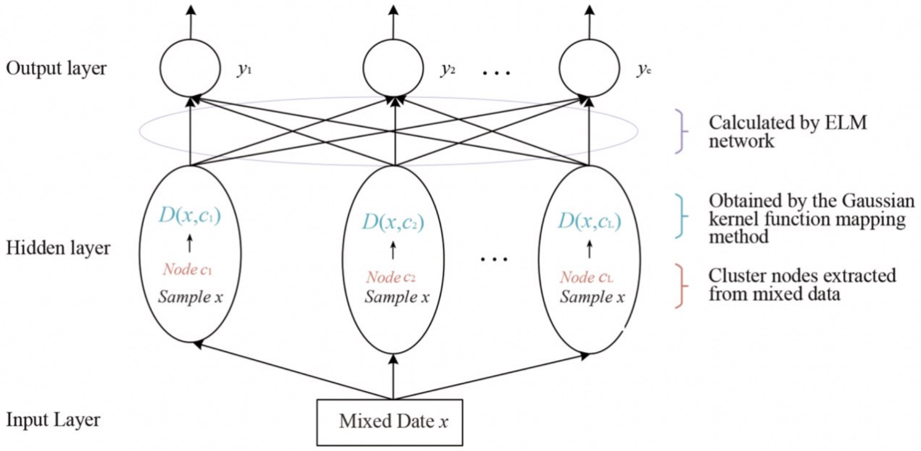


Figure 7. Model structure of RBF-ELM

(1) Input layer. Receives mixed (numeric or categorical) data with attributes as input and distributes the input data to the hidden layer. Where.M_c nodes receive categorical attributes and the remaining nodes receive numerical attributes.

(2) Hidden layer. consists of RFB centralized nodes$C_1 = (l = 1,..., L)$ which firstly completes the calculation of the center distance between x andC_1 center distance between the nodes, followed by mapping the output of this node by Gaussian kernel functionD_x as shown in equation (18).

$$\phi(x, C_1, \sigma_1) = exp(-D(-D(x, C_1)^2/2\sigma_1^2) \tag{18}$$

whereC_1 is the cluster center extracted from the training dataset, andσ_1 is the randomly generated width.

(3) Output layer. This layer uses the ELM method to calculate the weights between the hidden layer and the output layer by CNODES, and the formulae are shown in (19) to (20).

$$W = \left(H^T + \frac{I}{C}\right)^{-1} H^T \begin{bmatrix} T_1 \\ \vdots \\ T_N \end{bmatrix}, \tag{19}$$

$$where H = \begin{bmatrix} \phi(x_1, C_1, \sigma_1) & \cdots & \phi(x_1, C_L, \sigma_L) \\ \vdots & \ddots & \vdots \\ \phi(x_N, C_1, \sigma_1) & \cdots & \phi(x_N, C_L, \sigma_L) \end{bmatrix} \in R^{N \times L} \tag{20}$$

where$C$ is the regulation parameter, andT_i is the 0-1 vector, wherey_i term is 1 and the other terms are 0. In this layer, the first$i$ neuron output in this layer is$\widehat{y_l} = \sum_{l=1}^{L} \phi(x_i, c_i, \sigma_l) w_{li}$ The output of the first neuron in this layer is . According to equation (17), for any sample$x$ The predicted value for any sample is shown in equation (21).

$$label(x) = \arg max_i \, \widehat{y_l} \tag{21}$$

The principle of parameter learning is constructed by using error direction propagation and random search. Assuming that the actual output of the system is y, the desired output is denoted as d, and p *is* set as the sample natural number, the error is denoted as$e_p = y_p - d_p$, the error function is denoted as$E_p = \frac{1}{2} e_p^2 = \frac{1}{2}(y_p - d_p)^2$, then the center of the affiliation function and the width update mechanism are shown in (22-23).

$$\mu_{ij}(k+1) = \mu_{ij}(k) + \frac{\partial E_p}{\partial \mu_{ij}} \tag{22}$$

$$\sigma_{ij}(k+1) = \sigma_{ij}(k) + \frac{\partial E_p}{\partial \mu_{ij}} \tag{23}$$

In the video clip, the research idea of judging the fatigue state of human body by mouth features is to first extract single-frame image features, so as to obtain sequence features, and complete the classification of yawning process by sequence features. When yawn frequency≥10% (for one deep yawn, or multiple shallow yawns), the appearance of fatigue features can be fully confirmed.

The specific steps are as follows.

Step 1: Extraction of faces from a single frame image.

Step 2: Extract the mouth feature points and calculate the mouth aspect ratio using the height and width of the mouth, in equation (16)ϑ take 0.48, when the eye opening and closing degree≥0.48 is recognized as yawning state, if the time exceeds 3 seconds go to step 3, no return to step 1.

Step 3: Count yawning characteristics, Count=Count+1, when the frequency of yawning characteristics in 30 seconds exceeds the locking criterion, it is recognized as the current fatigue state.

4. Experimental validation and analysis

4.1. Facial detection dataset

Since both visual feature-based pilot and car driver detection require a video capture device above the driver's seat, their fatigue detection processes and algorithms are relatively similar in principle. However, the training parameters of the models may differ due to the differences in the acquisition device, acquisition environment, and the collected data set. Since real pilot flight data are more confidential and not easily accessible, and network training using processed video datasets can approximate a real-time facial acquisition scenario, public car driving video datasets are used instead and the effectiveness of the algorithm is verified.

In order to focus on the effectiveness of the algorithm in the case of high exposure, while taking into account the actual application of the camera capture image resolution is not good in most cases, the images used in the experiments are processed by Gaussian blurring and then exposure processing [22] The images used in the experiments are Gaussian blurred and then exposed. The experiment uses the public video dataset YawDD to simulate the real time scenario.[23] The experiment uses the public video dataset YawDD, which has 351 videos of car drivers driving, each video is labeled with four states: normal, talking, singing and fatigue yawning, and the states of normal, talking and singing are labeled as non-fatigue states, and the fatigue yawning is labeled as fatigue states. All video clips in the dataset were segmented to obtain frame sequences, and two frames were selected in each video clip of 7 seconds in length to create a subset of data as a regular component of the driver, and a total of 849 video clips were obtained for the experiment. This method of filtering videos from the dataset was used to better extract sequences that indicate that the driver is in a fatigue state and to exclude sequences that may affect the output of the neural network.

19-女人（没有眼镜）-哈欠.avi
19-女人（没有眼镜）-说话.avi
20-女人（没有眼镜）-哈欠.avi
20-女人（没有眼镜）-说话.avi
20-女人（没有眼镜）-正常.avi
21-女人（没有眼镜）-哈欠.avi
21-女人（没有眼镜）-说话.avi
22-女人（没有眼镜）-哈欠.avi
22-女人（没有眼镜）-说话.avi
22-女人（没有眼镜）-正常.avi
22-女人（太阳镜）-哈欠.avi
22-女人（太阳镜）-说话.avi
22-女人（太阳镜）-正常.avi
23-女人（没有眼镜）-哈欠.avi
23-女人（没有眼镜）-说话.avi
23-女人（没有眼镜）-正常.avi
24-女人（没有眼镜）-哈欠.avi
24-女人（没有眼镜）-说话.avi
24-女人（没有眼镜）-正常.avi
25-女人（没有眼镜）-哈欠.avi
25-女人（没有眼镜）-正常.avi

Figure 8. YawDD dataset

4.2. Experimental results and analysis

(1) Pilot face detection based on YOLOv5 network

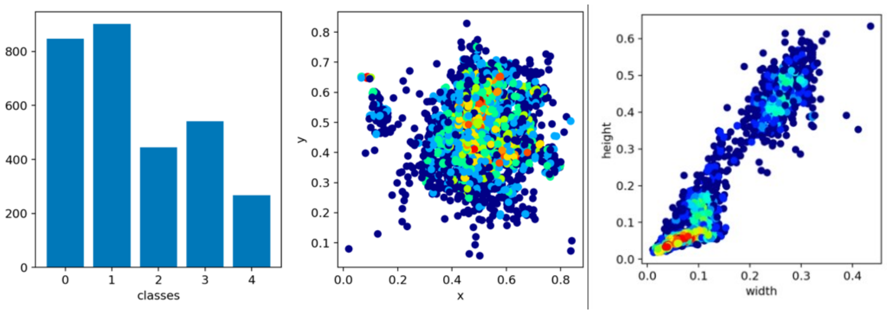
The training data was introduced into the YOLOv5 network for training, and the trained model was validated with the validation set data for model overfitting, and then tested with the test dataset. Each video file of the YawDD dataset was decomposed into 100 images according to key frames, and a total of 35,100 photos were introduced into the algorithm already written in PyCharm for facial region recognition, and the test results are shown in Table 2: the number of facial regions recognized by the network was 34832, and the facial recognition rate was 99.24%.

Table 2. Facial region recognition results in YawDD dataset

Number of face photos/photo	Number of facial areas recognized	Recognition rate/%
35100	34832	99.24

(2) Eye and mouth detection based on YOLOv5 network

The fatigue features were classified into 5 categories: 0 normal, 1 open eyes, 2 closed eyes, 3 open mouth, and 4 closed mouth. The visualized distribution of the dataset is shown in Fig. 9, where Fig. 9(a) shows the number of 5 categories of fatigue features distributed in the dataset after the experiment, Fig. 9(b) shows the distribution of the coordinates of the center points of the feature boxes labeled in the labels, and Fig. 9(c) shows the distribution of the length and width of the feature boxes.



(a) Distribution of the number of 5 types of fatigue features (b) Distribution of the coordinates of the center points of the marked feature frames (c) Distribution of the length and width of the feature frames

Figure 9. Visualization of data set distribution

When loU is set to 0.5 in the network, the AP of all images in each category is calculated, and then all categories are averaged, i.e., mAP (Mean Average Precision), which indicates the average mAP on different loU thresholds (from 0.5 to 0.95 in steps of 0.05), as shown in Fig. 10. The larger area under the curve means the larger the AP value, the higher the detection accuracy of the category.

Recall is the recall rate, and the formula is shown in equation (24). precision is the training model checking accuracy, and the formula is shown in equation (25). The two are contradictory metrics, and if the two can be better balanced, a better detection effect will be obtained under different conditions, i.e., the curve area in Fig. 10.

$$Precision = \frac{TP}{TP+FP} \tag{24}$$

$$Recall = \frac{TP}{TP+FP} \tag{25}$$

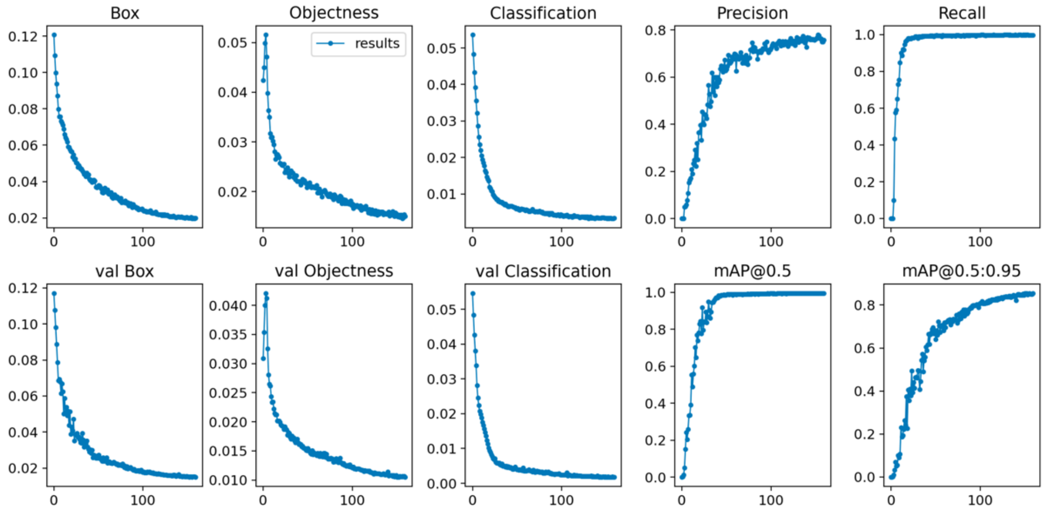
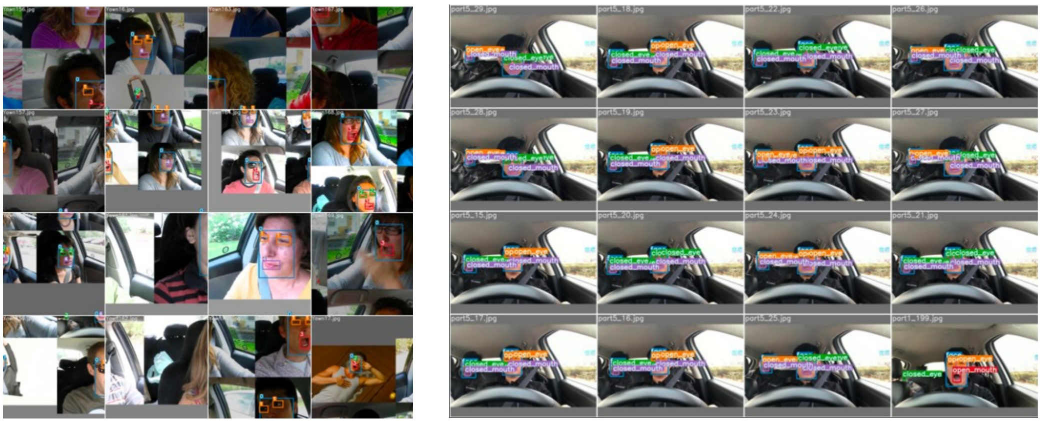


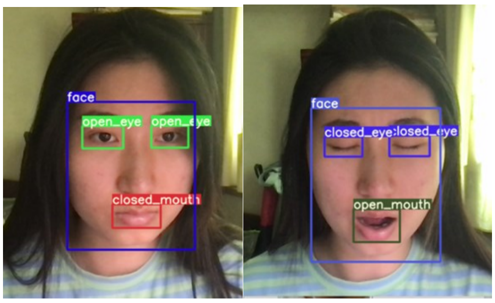
Figure 10. Model visualization results

Fig. 11 shows part of the fatigue labeling process in the training set and test set during the experiment. The trained algorithm can complete real-time facial feature recognition and eye and mouth fatigue feature detection for the video database. Fig. 12 shows the experimental diagram to test the real-time detection effect of the target detection model, which proves that the algorithm has the real-time detection effect and can accurately identify the real-time fatigue state of the subject.



(a) Training set fatigue feature labeling (b) Test set fatigue feature labeling

Figure 11. Experimental fatigue feature labeling



(a) Normal state (b) Fatigue driving state

Figure 12. Actual detection results of the target detection model

(3) Fusion of eye and mouth features

Since the yawning frequency can effectively reflect the pilot fatigue state [24][25] so the PERCLOS parameter is also applied to the detection of yawning frequency, and the two parameters are calculated as shown in Equation (26).

$$\begin{cases} P_e = \frac{\Sigma T_e}{T} \times 100\% = \frac{\Sigma N_e}{N} \times 100\% \\ P_m = \frac{\Sigma T_m}{T} \times 100\% = \frac{\Sigma N_m}{N} \times 100\% \end{cases} \tag{26}$$

Where P_{e} and P_m are the PERCLOS parameters for the eye and mouth, respectively.T_e andT_m denote the cumulative time that the eyes are closed and the mouth is open at an angle above the normal state per unit time T, respectively; N is the total number of video frames captured per unit time T, the$\mathrm{N_e}$ and$\mathrm{N_m}$ are the cumulative number of frames per unit time T where the eyes are closed and the mouth is open beyond the normal state, respectively. *The* PERCLOS parameters for the eyes and mouth are set to appropriate thresholds, and when the calculated P_{e} or P_m exceeds the set threshold, the driver is considered to be in a fatigue state.

To determine the appropriate PERCLOS threshold, 100 video sequences were randomly intercepted from the YawDD video dataset publicly used to design and test the fatigue detection algorithm, of which 50 were in the fatigue state and 50 were in the normal state, each for 10 s. After testing, the results were known as shown in Fig. 13, and the PERCLOS parameters of the eyes in the fatigue stateP_{e} between 0.30 and 0.70, and the PERCLOS parameters of the mouthP_m in the range of 0.35-0.80[14] . For this reason, comparison experiments were done for different thresholds of PERCLOS parameters at an interval of 0.05 whenP_{e} = 0.35, the P_m = 0.4, the least false and missed detections and the highest detection accuracy were achieved.

(4) RBF-ELM network-based classification of pilots' facial fatigue levels

The results of calculating PERCLOS values are shown in Fig. 14, from which we can find that the PERCLOS fatigue values are r

elatively low when the tester is awake in the video, and when the fatigue characteristics of the tester increase gradually with time, the PERCLOS fatigue values also increase, indicating that the PERCLOS fatigue values are able to characterize the degree of fatigue.

According to the analysis of the method of detecting fatigue of the mouth feature in this paper, the experiment obtained the conclusion in Fig. 15 that a deep yawn appeared between 40-150 frames, a shallow yawn appeared between 250-400 frames and a shallow yawn also appeared between 450-550 frames. According to Eq. (16), the fatigue feature frequency is 10% for 800 frames (within 30 seconds), indicating that the fatigue state of the mouth can be successfully detected.

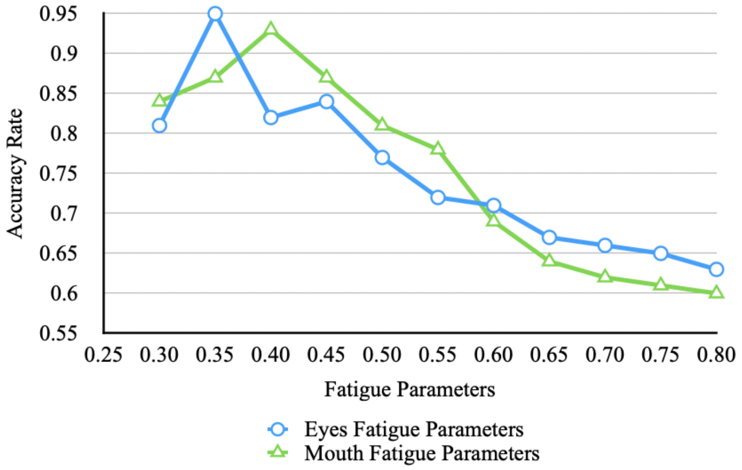


Figure 13. Comparison of experimental results

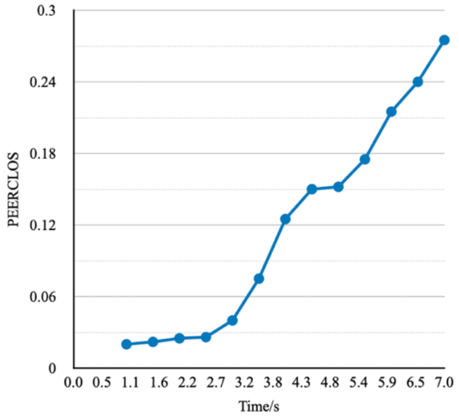


Figure 14. PERCLOS growth curve

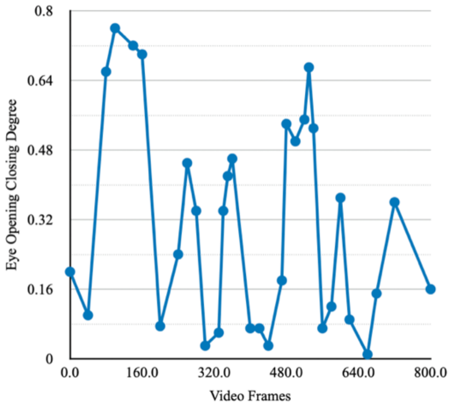


Figure 15. Mouth feature curve

4.3. *Performance comparison of different algorithms*

To verify the advantages of the model in fatigue detection, the model is compared with CNN (Convolutional Neural Network)[26] , PCA (Principal Component Analysis)[27] and DCT (Discrete Cosine Transform)[28] The model is compared with classical facial fatigue detection algorithms such as CNN (Convolutional Neural Network), PCA (Principal Component Analysis), and DCT (Discrete Cosine Transform), and the training and evaluation of facial fatigue detection are performed on the same database with a test set of1.32×10^4 test set of 5749 images containing faces from 5749 individuals. The experiments show that CNN can process images taken under high exposure conditions, but the performance degrades under poor lighting conditions. PCA and DCT are able to recognize images taken under average lighting conditions, but perform poorly in the dataset after high exposure processing. The evaluation shows that YOLOv5 in facial recognition outperforms other techniques in recognition regardless of the conditions of lighting conditions, background environment and facial features, and better meets the fatigue detection of pilots in real situations. Moreover, the RFB-ELM network model utilizes the nonlinear approximation capability of the classical RFB neural network and the fast learning capability in the ELM model to show high accuracy and processing efficiency in the fatigue classification process.

In the comparison experiments, the YOLO-RBF-ELM algorithm outperforms other algorithms of the same category in terms of comprehensive computational performance and shows a more obvious superiority in computational speed. Fig. 16 and Table 3 show the comparison data of different algorithms, and the experiments show that the mAP during the training of YOLOv5-RBF network can reach 246% of CNN, AP_50 is 136% of PCA, and AP_Medium is 343% of CNN.

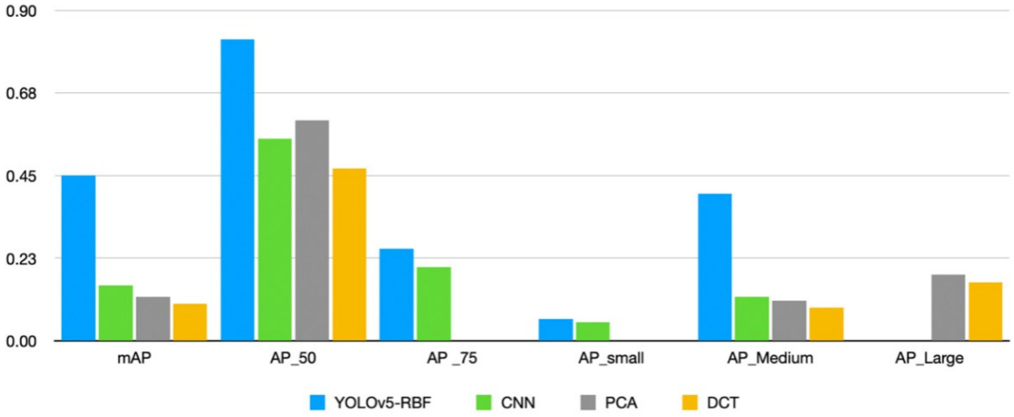


Figure 16. Performance comparison of different algorithms

Table 3. Computational performance of different algorithms

	YOLOv5-RBF-ELM	CNN	PCA	DCT
mAP	0.45	0.15	0.12	0.10
AP_50	0.82	0.55	0.60	0.47
AP _75	0.25	0.20	0.00	0.00
AP_small	0.06	0.05	0.00	0.00
AP_Medium	0.40	0.12	0.11	0.09
AP_Large	0.48	0.20	0.18	0.16

Each method was run 10,000 times on the dataset images, then 10 iterations were performed and the average time spent was taken. Fig. 17 shows the results of comparing the computational speed of different algorithms. Since the RFB network is able to approximate any nonlinear function with arbitrary accuracy as BP neural network, and has a fast convergence speed and strong generalization and parallel processing information capability, which makes it a fast learning capability, the computational speed of YOLO-RBF-ELM network model reaches 310 FPS when computed on GPU, which is a significant improvement over CNN, PCA, and DCT network frameworks. significant improvement over CNN, PCA, and DCT network frameworks.

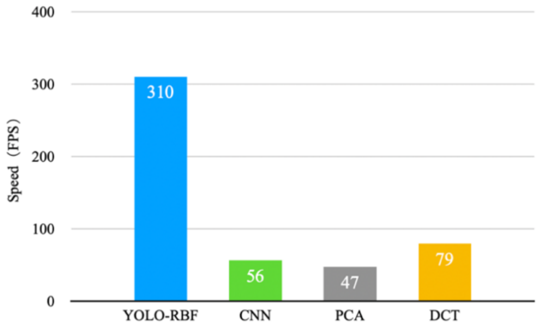


Figure 17. Comparison of computational speed of different algorithms

5. Conclusion

A non-contact pilot fatigue detection algorithm is proposed for high exposure of pilot's facial images due to the strong illumination environment that may be faced during the flight. The system uses frequency-domain light normalization to pre-process the data set images to make the face distinguishable under high exposure conditions; in facial detection, YOLOv5 with strong target detection capability is used; the fatigue grading is accomplished using an improved RBF-ELM network model to determine the grading of eye and mouth fatigue feature values based on PERCLOS method. The improved network has faster learning ability and more accurate fatigue grading results. In order to verify the effectiveness of the system, three model structures, CNN, PCA and DCT, were selected for comparison and validation. The experimental results show that the system can effectively identify the fatigue flight status of pilots and has a more obvious speed improvement in GPU calculation, which lays a foundation for non-contact pilot fatigue detection.

There is still room for improvement in the selection of the database in the article, and the subsequent production of a database using autonomously filmed videos can be used to train the network model; in addition, there are limitations for the possible noise and occlusion in the image sequences, so this important factor is planned to be included in the subsequent thinking and research. The future goal is to develop a real-time monitoring system that incorporates more features, such as mouth features, eye features, and image sequence features, to monitor the factual fatigue state of pilots.

References

[1] Feng Chuanyan, Wanyan Xiaoru, Chen Hao et al. Context-awareness model and application based on multi-resource load theory[J]. Journal of Beijing Aviation University, 2018, 44(7): 1438-1446.

[2] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, “A Review of Yolo Algorithm Developments,” Procedia Comput. Sci., vol. 199, pp. 1066–1073, 2021, doi: 10.1016/j.procs.2022.01.135.

[3] C. Jie and F. Xiaoyi, “Method of Facial Expression Recognition Based on Coarse-to-fine Classification,” 2007.

[4] T. Pradhan, A. N. Bagaria, and A. Routray, “Measurement of PERCLOS using eigen-eyes,” 4th Int. Conf. Intell. Hum. Comput. Interact. Adv. Technol. Humanit. IHCI 2012, pp. 0–3, 2012, doi: 10.1109/IHCI.2012.6481864.

[5] C. D. Wickens, J. Helleberg, and X. Xu, “Pilot maneuver choice and workload in free flight,” Hum. Factors, vol. 44, no. 2, pp. 171–188, 2002, doi: 10.1518/0018720024497943.

[6] DENG Zhenghong, HUANG Yijie, LI Xiang et al. Research on video-based driving fatigue detection technology[J]. Journal of Northwestern Polytechnic University, 2015, 33(6): 1101-1106.

[7] C. Zheng, B. Xiaojuan, and W. Yu, “Fatigue driving detection based on Haar feature and extreme learning machine,” J. China Univ. Posts Telecommun., vol. 23, no. 4, pp. 91–100, Aug. 2016, doi: 10.1016/S1005-8885(16)60050-X.

[8] T. Shimizu, T. Nanbu, and T. Sunda, “An exploratory study of the driver workload assessment by brain functional imaging using onboard fNIRS,” 2011. doi: 10.4271/2011-01-0592.

[9] S. C. F. Lin et al., “Image enhancement using the averaging histogram equalization (AVHEQ) approach for contrast improvement and brightness preservation,” Comput. Electr. Eng., vol. 46, pp. 356–370, 2015, doi: 10.1016/j.compeleceng.2015.06.001.

[10] Luo Y.X., Jia B., Qiu X.Y. et al. Brain fatigue state identification for pilots based on Gamma deep belief network[J]. Journal of Electronics, 2020, 48(6): 1062-1070.

[11] W. Chen, M. J. Er, and S. Wu, “Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain,” IEEE Trans. Man Cybern., vol. 36, no. 2, pp. 458–466, 2006.

[12] Xia Linlin, Yue Meng, Czer Wang et al. Sparse direct method VSLAM 3D reconstruction based on Retinex theory[J]. Chinese Journal of Inertial Technology, 2021, 29(2).

[13] R. Neha and S. Nithin, “Comparative Analysis of Image Processing Algorithms for Face Recognition,” Proc. Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2018, no. Icirca, pp. 683–688, 2018, doi: 10.1109/ICIRCA.2018.8597309.

[14] X. Fan, B. C. Yin, and Y. F. Sun, “Yawning detection for monitoring driver fatigue,” Proc. Sixth Int. Conf. Mach. Learn. Cybern. ICMLC 2007, vol. 2, no. August, pp. 664–668, 2007, doi: 10.1109/ICMLC.2007.4370228.

[15] S. Singh and N. P. Papanikolopoulos, “Monitoring driver fatigue using facial analysis techniques,” IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC, pp. 314–318, 1999, doi: 10.1109/itsc.1999.821073.

[16] B. Mandal, L. Li, G. S. Wang, and J. Lin, “Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 3, pp. 545–557, 2017, doi: 10.1109/TITS.2016.2582900.

[17] Ma Boyu, Yinwei Yu. Research and implementation of face recognition system based on AdaBoost algorithm[J]. Journal of Instrumentation, 2016, 37: 163-167.

[18] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “SphereFace: Deep hypersphere embedding for face recognition,” Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 6738–6746, 2017, doi: 10.1109/CVPR.2017.713.

[19] O. Of and M. Carriers, “PERCLOS : A Valid Psychophysiological Measure of Alertness As Assessed by Psychomotor Vigilance,” October, vol. 31, no. 5, pp. 1237–1252, 1998, [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:PERCLOS+:+A+Valid+Psychophysiological+Measure+of+Alertness+As+Assessed+by+Psychomotor+Vigilance#0

[20] C. Jacobé de Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J. L. Vercher, “Detection and prediction of driver drowsiness using artificial neural network models,” Accid. Anal. Prev., vol. 126, no. November, pp. 95–104, 2019, doi: 10.1016/j.aap.2017.11.038.

[21] S. Ding, L. Xu, C. Su, and F. Jin, “An optimizing method of RBF neural network based on genetic algorithm,” Neural Comput. Appl., vol. 21, no. 2, pp. 333–336, Mar. 2012, doi: 10.1007/s00521-011-0702-7.

[22] Wang Y-L, Chen L, Lai M et al. Fuzzy face recognition based on trace transform and rotation incremental modulation features[J]. Journal of Electronics, 2021, 12(12): 2439-2448.

[23] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, “YawDD: A yawning detection dataset,” Proc. 5th ACM Multimed. Syst. Conf. MMSys 2014, pp. 24–28, 2014, doi: 10.1145/2557642.2563678.

[24] Y. C. Lee, J. D. Lee, and L. N. Boyle, “Visual attention in driving: The effects of cognitive load and visual disruption,” Hum. Factors, vol. 49, no. 4, pp. 721–733, Aug. 2007, doi: 10.1518/001872007X215791.

[25] Song, Cambridge, Wang, Feng, Niu, Jin et al. A spatio-temporal neural network-oriented method for potential emotion recognition[J]. Journal of Xi'an University of Electronic Science and Technology, 2021, 48(4): 160-167.

[26] W. H. Gu, Y. Zhu, X. D. Chen, L. F. He, and B. B. Zheng, “Hierarchical CNN-based real-time fatigue detection system by visual-based technologies using MSP model,” IET Image Process., vol. 12, no. 12, pp. 2319–2329, 2018, doi: 10.1049/iet-ipr.2018.5245.

[27] B. Fatima, A. R. Shahid, S. Ziauddin, A. A. Safi, and H. Ramzan, “Driver Fatigue Detection Using Viola Jones and Principal Component Analysis,” Appl. Artif. Intell., vol. 34, no. 6, pp. 456–483, 2020, doi: 10.1080/08839514.2020.1723875.

[28] T. Tuncer, S. Dogan, F. Ertam, and A. Subasi, “A dynamic center and multi threshold point based stable feature extraction network for driver fatigue detection utilizing EEG signals,” Cogn. Neurodyn., vol. 15, no. 2, pp. 223–237, 2021, doi: 10.1007/s11571-020-09601-w.