IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Integrated neural network-based pupil tracking technology for wearable gaze tracking devices in flight training

Heming Zhang[1]. Changyuan Wang[1].
1 Optical Engineering, Xi'an Technological University, Xi'an, China

Corresponding author: Changyuan Wang (e-mail: 617222535@qq.com).
This work was supported in part by The National Natural Science Foundation of China] (No. 52072293).

ABSTRACT Pupil tracking technology is a tracking and detection method that uses eye image information to extract real-time position information of the pupil. Detecting the pilot's eye movement patterns and characteristics through pupil movement signals is an important part of monitoring the pilot's physiological characteristics. The current pupil tracking algorithm is prone to problems such as insufficient tracking accuracy and discontinuous pupil signals when faced with problems such as pupil occlusion caused by frequent blinking and loss of pupil information in dark light environments that occur during flight training for pilot students. To increase the tracking accuracy of pilots' pupils, this paper designs an integrated neural network-based pupil tracking technology for wearable gaze tracking devices in flight training. To solve the above problems, this paper builds a pupil positioning model based on the hybrid neural network by combining the feature pyramid and ViT network. On this basis, we built a hybrid neural network pupil tracking model for occluded pupil images based on the pilot eye data characteristics collected during flight training and designed a new loss function suitable for pupil detection. After verification, the pupil tracking algorithm we proposed has significantly improved the visual tracking accuracy with an error range of less than 5 pixels compared with existing methods, and the tracking accuracy can reach up to 85%. In pilot flight training, this algorithm has better pupil tracking stability, can effectively reduce pupil signal interference caused by pupil occlusion, and can achieve more accurate real-time tracking of pupils.

INDEX TERMS pupil-tracking; hybrid neural network; feature pyramid; ViT;

I. INTRODUCTION

Pupil tracking is the process of automatically positioning the pupil center and gaze point of the eyeball. Humans obtain external information mainly through visual information perceived by the human eye. Eye movement can intuitively reflect relevant data such as a person's gaze point and gaze time. It is of great significance for describing the process and characteristics of human visual perception and exploring the basic cognitive process of an individual. Obtaining data such as the pilot's operating response, flight status, and gaze habits by analyzing the pilot's eye movement patterns has become an important part of the current pilot's physiological data detection [1].With the rapid development of computer vision, the analysis and processing of visual information obtained from the external environment with the help of computer technology represented by artificial intelligence has been applied to people's real lives and has gradually become an important research direction of computer vision and artificial intelligence. In the field of flight training, people use eye movement detection equipment to obtain the pupil signals of student pilots and use computer vision and artificial intelligence technology to extract information such as the real-time position of the pupils. This method can help pilots complete flight training more efficiently. This detection method has been widely used in the field of flight training because the equipment is easy to wear and does not easily cause psychological discomfort to the subjects [2].

Because the camera distance of the head-mounted eye tracker is relatively close to the eyes and the relative position to the eyes is fixed, it avoids problems such as the restriction of the pilot's head angle by the desktop eye tracker. To collect complete pupil images during pilot training, this article uses a head-mounted eye tracker to detect pupil signals during pilot training. However, the pupil tracking algorithm used by currently commercially available eye trackers cannot meet the accuracy requirements of eye trackers for pilot flight training. When the subject makes actions such as closing eyes, blinking, or covering eyelashes, the detection accuracy of the pupil tracking algorithm is poor. In the case of large-area occlusion, the existing pupil tracking algorithm cannot

achieve continuous tracking and stable detection of pupil signals. When pilots perform multiple simulated flight missions, their eyes are in a state of high mental stress for a long time. Pilots' eyes will suffer from visual fatigue to varying degrees. During flight training, it is often difficult for pilots to keep their eyes open for a long time, and they often blink or squint frequently. It is difficult for existing algorithms to achieve long-term pupil tracking. In addition, the current pupil tracking algorithm is not adaptable to changes in the lighting environment. Since the pupil signal collection process requires detection in both light and dark environments, pupil positioning accuracy cannot be guaranteed in dark light. These result in a large amount of clutter and interference in the final pupil curve. Figure 1 shows the pupil tracking signal when looking at the flying target during flight training using eye tracker equipment in pilot training. The green signal line records the movement trajectory of the flying target over time, and the black line records the pupil signal recorded by the eye tracker when the pilot's eyes are looking at the flying target. As can be seen from Figure 1, due to the pilot's frequent squinting and blinking movements, there are many interferences and glitches in the pupil signal, which causes serious interference to pupil detection and reduces detection efficiency and detection accuracy. These problems can interfere with the analysis of pupil movements during flight training [3]. In addition, existing eye tracker equipment is often only suitable for a single data set for pupil tracking under pupil occlusion, and there are differences in recognition accuracy and work stability for images with different pupil characteristics.

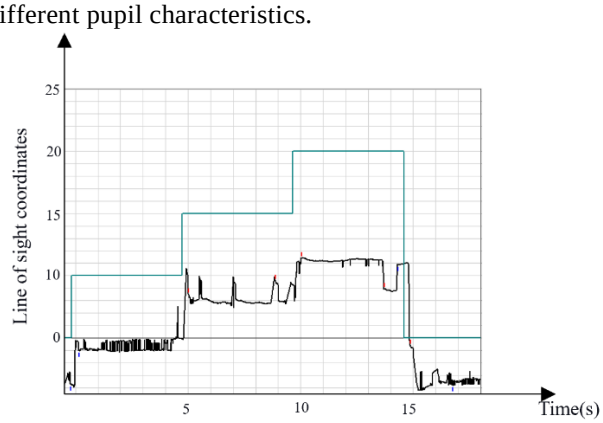


FIGURE 1. Pupil-tracking signal

In response to the above problems, we have carried out the following work: (1) To increase the accuracy of pupil detection, in this paper, we studied hybrid neural networks and built a hybrid neural network for pupil detection to achieve accurate positioning of pupils. (2) To improve the stability of the network in identifying different pupils, this article introduces an improved feature pyramid network and designs a new loss function based on the physiological characteristics of the eyes in pupil detection. (3) Aiming at the problem of pupil information loss caused by pupil occlusion, this paper builds an integrated pupil-tracking network by combining a hybrid neural network and a long short-term memory network, using the position of the pupil in past frames to calculate the current missing frame. Pupil position prediction.

II. Related work

Previous pupil tracking and positioning technology determined pupil position information based on a priori information related to the human eye or face position. On this basis, researchers have further enriched pupil-tracking technology for specific environments and needs. For example, Ciesla and Koziol used a web camera combined with the relative position of the eyes on the face to build an artificial neural network and achieve real-time detection and positioning of pupils [3]. To solve the problem of interference from reflective objects such as glasses on pupil detection, Choi et al improved the neural network structure. Choi et al split the pupil image by combining the subject's facial image and achieved real-time position tracking and detection of the pupil [4]. The method of eye positioning based on facial area images improves the accuracy of the pupil recognition method to a certain extent. However, when the pupil area is severely occluded, the accuracy of pupil recognition needs to be further improved. Amine Kacete proposed a pupil detection method based on the Hough transform random regression tree to achieve high-precision pupil position estimation using a commercial eye tracker and to improve the tracking stability of the eye tracker in outdoor environments. This method optimizes the network structure based on the data distribution characteristics of the pupil image. This algorithm solves the problem of accurate positioning of pupils in scenes where the lighting environment changes frequently [5]. To improve the detection accuracy of the network for blurred pupil images and pupil images with background changes, Khan, W proposed a pupil positioning method combined with machine learning. This method identifies the subject's facial area through a pre-trained model, and on this basis achieves better detection results by improving the convolutional network structure [6]. To further improve the accuracy of pupil detection, Tian D. improved the decision tree algorithm and chose a moderate rate to increase the weight of training samples with larger errors. After experiments, the algorithm achieved more efficient pupil positioning. To solve the interference of artifacts and natural low-light conditions in eye movement videos on pupil recognition[7].Yuk-Hoi Yiu built a fully convolutional neural network for eye movement video data and finally achieved accurate and efficient segmentation of pupils [8]. Hongwei Ma initially positioned the pupil in the subject's eye image and then used a shallow CNN network to achieve accurate pupil position detection [9]. For eye images under infrared light, Sang Yoon Han used the U-Net model to segment the pupil area in the image and achieved an accuracy of 85.3% [10]. To eliminate the work of pupil labeling during the training of the pupil detection model, Pengxiang Xue proposed a fake data set generation

algorithm that integrates affine image features. This method has shown high stability in multiple validation data sets [11]. To improve the efficiency of artificial intelligence algorithms in the pupil recognition process, Kim S developed a lightweight pupil tracking algorithm. This algorithm uses a fast and accurate cascaded depth regression forest to meet high tracking accuracy while reducing the computational complexity of the algorithm [12].Nenad Markuš proposed a pupil positioning method based on a random regression tree set to solve the problem of accurate pupil positioning of low-resolution acquisition equipment. This method achieves accurate and stable tracking and positioning on mobile devices [13]. Lee, Y.W considered the impact of optical and motion blur, thick eyelashes, and light reflection from glasses on the quality of eye images, and studied the deep ResNet network to quickly identify pupil images. This algorithm has higher accuracy in eye-iris recognition. The previous algorithm has been significantly improved [14].

For fully automatic detection of nystagmus and eye movement, researchers currently mainly collect eye movement data for modeling and use artificial neural networks to track the position of the pupil. Wei K[15] designed a new nystagmus recorder using the OV4689 camera as an eye image tool for benign paroxysmal positional vertigo. This research extracts semantic information in images through the Yolov5 model and achieves real-time pupil segmentation through the improved Deplabv3+ network. The method proposed by Wei K achieves good results in pupil segmentation, but this method does acquisition not take into account the detection accuracy of pupils in dark light environments. To obtain the position information of the subject's eyes in different states, Newman J L [16] and others built a 2D convolutional neural network with the help of CAVA equipment. The researchers fused the head and eye movement data of patients when they experienced vertigo and realized vertigo detection based on the fusion of eye movement signals and head posture. Friedrich M U [17] built a ConVNG network for pupil detection with the help of smartphones .Shi, L[18] designed a pupil detection model based on the V_net network to obtain the position information of the subject's eyes in different states. The model employs a long and short memory network and is optimized for real-time detection of pupil coordinates even under occlusion. After testing, the VCF network proposed in this article was found to have a pupil detection accuracy of over 81% in a laboratory environment, within an error threshold of 5 pixels. Deng W et al.[19] addressed the issue of pupil center occlusion during the detection of benign paroxysmal vertigo by setting the pupil detection target as the lower pole of the pupil to determine the pupil position. This method was proven effective in diagnosing benign paroxysmal vertigo and has shown superiority [20].Wang L extracted local features and global features of eye images by combining the ResNest network and the ViT network. This method achieves a high pupil-tracking effect[21]. Experiments have confirmed that after importing some images from the TEyeD data set into the neural network, the algorithm's pupil detection accuracy can reach up to 85%.

III. Methods

To ensure the stable tracking of the pilot's pupils during the detection process, this paper proposes a pupil-tracking network for pilot training. The network consists of two parts: the pupil positioning network and the pupil tracking network, which respectively solve the poor pupil positioning stability and signal interference caused by pupil occlusion during the pupil signal collection process.

A. Pupil detection methods and data sources

To obtain eye-tracking data that is more in line with the pilot's flight training environment, we designed a head-mounted eye tracker for flight training. The acquisition end of the equipment uses a head-mounted eye tracker to collect the pupil images of the subject in real-time and upload them to the host computer.

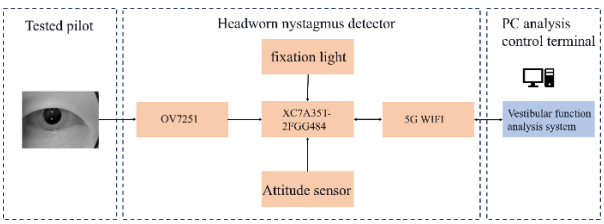


FIGURE 2. Structure diagram of head mounted nystagmus detector

The structure of the head-mounted eye tracker is shown in Figure 2. The head-mounted eye tracker integrates the OV7251 high frame rate camera. The eye tracker controls the OV7251 camera to collect the pilot's eye images through the built-in XILINX-XC7A35T-2FGG484 programmable logic controller. During the flight training process, the pilot under test looks at the display screen of the simulated flight platform through the observation glass. The camera obtains a clear image of the eye through the reflection of the viewing mirror. Obtaining eye images of the subject through this method can obtain a more comprehensive eye image without blocking the subject's line of sight.

This article uses three data sets to train and verify the pilot pupil detection model. Data set I is the eye images collected when pilot students wear head-mounted eye trackers for flight training during simulated flight training. Dataset I is different from the existing pupil detection public beta data set. This data set more realistically records the pilot's eye activity status after completing the simulated flight. To verify the reliability of the network output results, Dataset II uses the open dataset TEyeD of human eye images and annotates them. TEyeD (also known as The Tel-Aviv University Eye Dataset) [22]is a dataset used for pupil detection and pupil tracking research. The dataset was created by researchers at Tel Aviv University in Israel. The TEyeD dataset contains pupil images from participants of different races, ages, and genders, covering a variety of lighting conditions, expressions, and postures. These

images cover pupil images from frontal and side angles, as well as images of different resolutions, enabling accurate pupil detection and tracking experiments in various scenarios. Dataset III is a data set collected by Świrski35 for pupil tracking. The Świrski35[23] pupil dataset is an important dataset for the field of pupil detection and pupil tracking. The dataset was created by researchers at the University of Warsaw, Poland, to promote research in fields such as computer vision and biometric identification. This dataset contains images of pupils from people of different races and ages, covering a variety of lighting conditions, expressions, and postures. The collection of pupil data sets includes frontal and side angle images, as well as pupil images of different resolutions. Data sets II and III are shown in Figures 3 and 4:

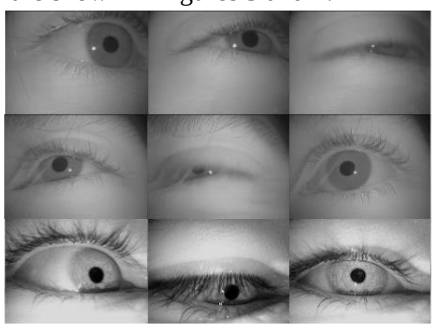


FIGURE 3. Dataset II eye images

B.Pupil positioning network based on hybrid neural network

This paper builds a pupil positioning network with the help of a serial network composed of optimized feature pyramid and ViT. The structure is shown in Figure 5. This model uses the feature pyramid network to extract the global features of the eye image, and then uses the ViT network to extract the local features of the pupil. Our network improves the anti-interference performance of pupil detection with the help of optimized feature pyramid, allowing the ViT network to obtain more accurate image features for the next step of spatial mapping.

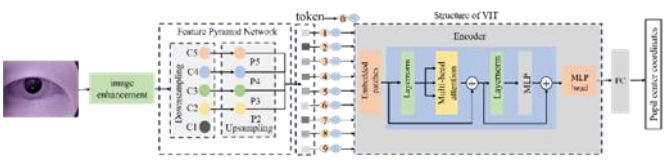


Figure 5. Pupil positioning network structure

1) IMAGE ENHANCEMENT

Since the camera frame rate of the head-mounted eye tracker can reach 100fps, this makes the exposure time of each frame of image short. Since some flight training sessions are conducted in a dark light environment, this results in the overall acquired eye images being darker. Commonly used methods such as low near-infrared fill light and short exposure time will cause the overall captured pupil image darker. Low image contrast will considerably affect the output accuracy of

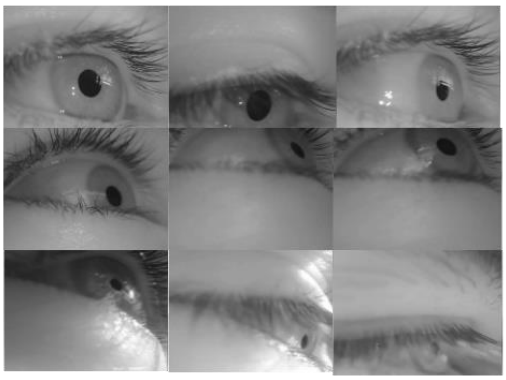


FIGURE 4. Dataset Ⅲ eye images

subsequent pupil detection algorithms, so it is necessary to improve the contrast ratio of the picture through image enhancement. We performed image enhancement algorithms such as gamma transformation, linear transformation, histogram normalization, and global histogram equalization on the eye images collected by the eye tracker, and drew the output image and grayscale distribution histogram. As shown in Figure 6, the image enhancement effect using the global histogram equalization algorithm is the best, and the pupil edges are more prominent.

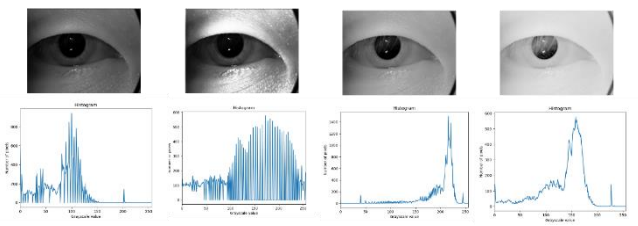


Figure 6. Image enhancement algorithm and grayscale distribution histogram

2) IMPROVED FEATURE PYRAMID NETWORK

In order to improve the generalization ability of eye feature extraction and accurate description of the overall features, the Feature Pyramid Network (FPN) in this article is improved. The improved feature pyramid network gives different weights to each feature layer by introducing an adaptive feature pyramid network. This method enables the feature pyramid network to more comprehensively extract pupil features and improve the network's adaptability to different eyes.

Feature Pyramid Network (FPN) [24] improves the accuracy of target detection tasks by combining the features of deep and shallow target detection images. The network of FPN consists of two structures: bottom-up path and top-down path. The bottom-up path consists of many convolutional modules, and each module contains many convolutional layers. In the bottom-up process, the spatial dimension is halved module by module (the step size is doubled). The output of each convolutional module will be used in a top-down path. The bottom-up path is that its dimensions change after the eye feature map is fed into

certain network layers. Layers that do not change the feature map size are grouped into a stage such that each extracted feature is the output of the last layer of each stage. Then, use the top-down path to achieve upsampling of the eye feature image. The feature maps generated using the pupil image through the FPN network through the top-down and bottom-up paths are C2, C3, C4, C5, and P2, P3, P4, and P5, respectively. The structure diagram of FPN is shown in Figure 7. This paper uses convolutions with convolution kernel sizes of 3×3, 5×5, and 7×7 to extract pilot pupil image information. The convolution kernel performs convolution operations on the features $V \in R^{H\times W\times C}$ respectively, and obtains the features $V1 \in R^{H\times W\times C/2}$, $V2 \in R^{H\times W\times C/4}$ 和 $V3 \in R^{H\times W\times C/4}$. The features of the first layer input image in the pilot's eye image contain noise and other information that interferes with pupil recognition, resulting in poor pupil detection. Deep features contain some information about pupils, but since they are mainly used for the prediction of large objects, to a certain extent, the detector may confuse the fuzzy semantics of pupils with other image information of the eye.

To enable the feature extraction network to obtain more pupil feature information and increase its adaptability to different human eye images. This article introduces an improved method of feature pyramid networks. Since the pupil information characteristics of each area in the human eye image differ, this paper introduces a bottom-up network structure with weight learning. This structure starts from the FPN output feature map P2 for convolution downsampling and calculates the weighted sum with the feature map P3. After the two are fused, a new feature map is obtained. The new feature map will continue to be fused in the same way. The final improved feature network output corresponds to the multi-layer output of the previous FPN. Compared with the ordinary FPN network structure, the features obtained in this way can extract more effective information about the eyes with the help of weights and reduce the impact of redundant information on detection accuracy. This network uses different weights to effectively reduce feature redundancy caused by the simple addition of features in human eye images and highlight pupil features.

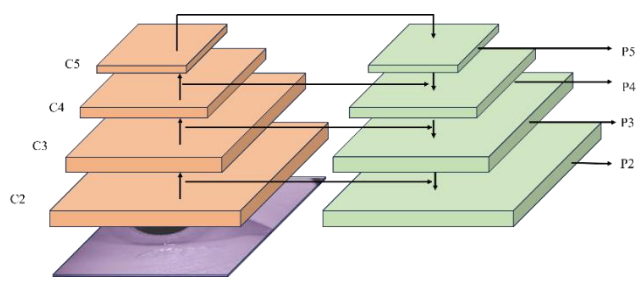


Figure 7. Feature pyramid structure

3) VISION TRANSFORMER

The network we proposed extracts global features of eye images through the ViT network. The ViT network consists of a multi-head self-attention block (Multi-head Self Attention, MSA) and a multi-layer perceptron block (Multi-Layer Perceptron, MLP). MSA is the core structure of the model. It performs weighted calculations on each position of the input sequence. and, realize the transmission and integration of information in the sequence. During the operation of MSA, a weight matrix is calculated, which further assigns a specific weight to each position of the input sequence, reflecting the importance of calculating the relationship between different positions of the curve. Before feeding image data into the model, the ViT network is trained with learnable class tokens. After the model training is completed, ViT uses this mark to perform the classification task; during this process, the ViT network stores the global position of each patch through superimposed position information encoding (Position Embedding).The ViT network divides the feature map into multiple patches of size H*W*C and calculates the mapping relationship of the patches to obtain vectors of corresponding lengths.

$$z_0 = [x_{\text{token}}; x_1E; x_2E; \dots; x_mE] + E_{\text{pos}}. \quad (1)$$

z_l represents the classification flag bits input to the multi-layer perceptron after L encoders; x_{token} represents the image block with the classification flag bits. In the formula: $\boldsymbol{E} \in R^{(S^2\times C)\times D}$ represents the embedding projection of Patch; $\boldsymbol{E}_{\text{pos}} \in R^{N\times D}$ represents the position embedding.

The encoder associates the divided image block information and uses the self-attention mechanism to assign position coding to the mapping vector to obtain the semantic representation of each eye feature image. The vector is then sent to the Transformer to avoid pixel-level attention operations. The location information calculation process is:

$$z_l' = \text{MSA}\left(LN(z_{l-1})\right) + z_{l-1},\ l = 1, \dots, L \quad (2)$$

$$z_l = \text{MLP}\left(LN(z_l')\right) + z_l',\ l = 1, \dots, L. \quad (3)$$

In formulas (2) and (3): z_l represents the encoded features. In the last layer of the encoder, the first Token in z_L^0 is taken as the global feature expression of the image and passed to the classifier for predicting the label.

4) LOSS FUNCTION

The complete intersection and union loss function (CIou) has shown good results in the target detection network. It uses the distance intersection and union loss function (DIoU) to combine the predicted frame's overlapped part, aspect ratio, and center point with the actual frame. An aspect ratio penalty term is added for the impact of the loss function, which has better effects in the multi-scale target detection process. To make the neural network converge faster and adapt to the shape changes of the pupil, we combined the CIoU function to redesign the overall loss function of the pupil.

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (4)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (5)$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \quad (6)$$

loU is the ratio of the intersection and union of the predicted box and the actual box; ρ is the distance between the center points of the two; b and b^g are the center points of the predicted box and the actual box respectively; c is the distance between the predicted box and the actual box The distance between the two non-adjacent corners of the circumscribed minimum rectangle formed by the box; α is the weight coefficient; v is the aspect ratio consistency parameter; w, w^{gt} are the widths of the predicted box and the actual box respectively; h, h^{gt} are the heights of the predicted box and the actual box.

Since the eye image collected by the nystagmus detector is information from the front of the eye, the eye image information is relatively fixed. In order to better detect the pupil position information, we describe the position-related information of the pupil by setting the pupil center, rotation angle, and eye axis, as shown in Figure 8. The overall loss function of the hybrid neural network for pupil detection can be expressed as:

$$\begin{aligned}\mathcal{L}(y, \hat{y}) = & \frac{1}{m}\sum_{k=1}^{m} \alpha \mathcal{L}_C(y, \hat{y}) \\ & + \beta\left[(a^k - \hat{a}^k)^2 + \left(b^k - \hat{b}^k\right)^2\right] \\ & + \gamma\left(\theta^k - \hat{\theta}^k\right)^2\end{aligned} \quad (7)$$

Here $y = (c_x, c_y, a, b, \theta)$, and c_x, c_y represents the pupil center, a, b represents the axis distance, θ represents the rotation angle, $\mathcal{L}_C$ is the (DCIoU) loss function, α, β, γ represent the weight.

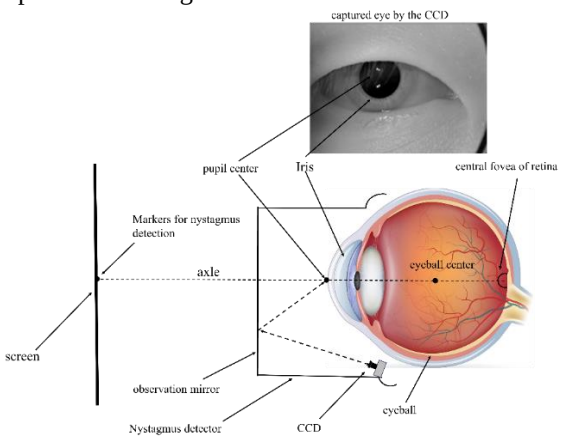


Figure 8. Principle of eye image collection by head-mounted vestibular detector

B. Pupil tracking network for missing pupil frames

When the subject performs flight training, the movement signal of the pupil is periodic, and during the detection process, the period, frequency and other parameters of the pupil movement signal are relatively fixed. On this basis, to reduce the computing time of the system, we use the long and short memory network to track the pupil, which is used to predict the pupil coordinates of the occluded pupil and the missing frame of the image. To increase the output accuracy of the long and short memory network, we used the long and short memory network to build a pupil prediction network for frames with missing pupil information.

As shown in Figure 9, the eye images at t_1 and t_2 are input into the pupil tracking network, and the pupil information is detected through the pupil positioning network. The network uses the long short-term memory network to calculate the current pupil position information based on the pupil center of past frames. The expression of the pupil tracking network is:

$$D(t) = \begin{cases} I(D_t \mid D_t - 1, D_t - 2), t \in T \\ L(Z_t), t \in T \end{cases} \quad (8)$$

In Formula 8, D_t and Z_t represent the output and input images of the network at time t, respectively, and T represents the pupil tracking process time. Within T time, I (pupil prediction network for pupil missing frames) predicts the pupil information of the current frame based on the pupil information of the previous two frames. During the non-tracking period, L (pupil positioning network based on the hybrid neural network) detects the corresponding pupil information based on the current frame image.

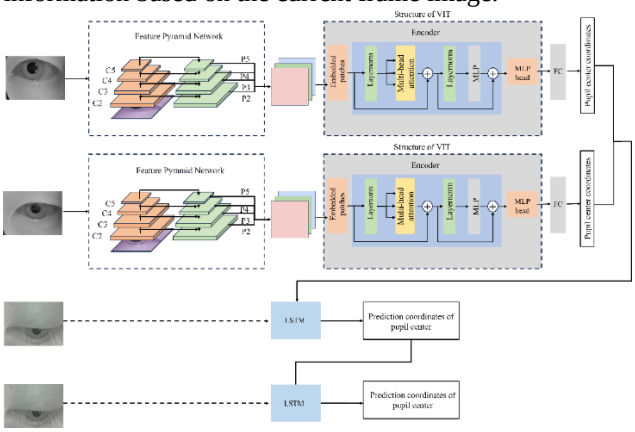


Figure 9. Pupil tracking network structure

LSTM neural network is a particular type of recurrent neural network (RNN)[25]. It proposes a solution to the problem of long-term dependence on input information of RNN by introducing three gates: forgetting, input, and output, which enables LSTM to learn long-term dependence on input information compared to RNN, and simultaneously solves the unreasonable gradient existing in RNN. This problem enables LSTM to maintain a stable error range during backpropagation, improving the prediction accuracy of short-term predictions. Figure 10 shows the LSTM neural network structure.

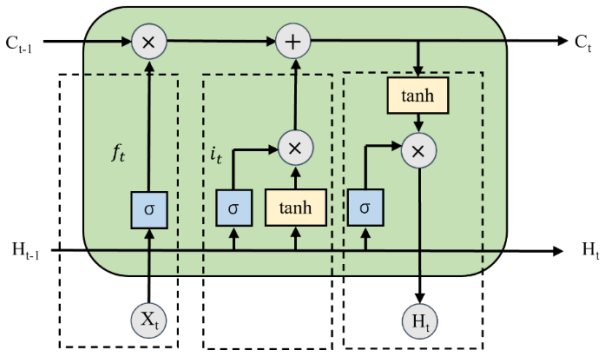


Figure 10. LSTM network structure

The LSTM network is responsible for obtaining the pupil coordinates at this time using the pupil coordinate information of the previous n frames. The long and short memory network first reads the pupil coordinate data and establishes the correlation characteristics between the pupil coordinates and time through the input gate, forgetting gate, and output gate of the long and short memory network. Finally, according to the output accuracy, different weights are set for various hidden layer units of the long and short memory network, and the prediction results are finally output after training.

Assume that the long short-term memory network has a total of m neurons, where the output value of the x-th neuron is c_x, the neuron's weight is v_x, and the output value is y_t.

$$v_x = \frac{\exp(c_x)}{\sum_{i=1}^{m} c_i} \tag{9}$$

$$y_t = \sum_{i=1}^{t} v_i c_i \tag{10}$$

We have verified through experiments that when the number of long and short memory network neurons is 20, the batch processing volume is 1024, and the training period is 10, the output accuracy of the long short memory network is the most accurate.

III. Experiments and results discussion

To test the accuracy of the pupil positioning and tracking model in the pilot pupil detection process, we recorded the eye movement images and pupil data of 25 student pilots wearing head-mounted eye trackers and completing flight training in simulated flight training. The test process is shown in Figure 11. During this process, we collected a total of 150,286 eye images. We use 80% of them as the training set, and the remaining images as the test set as the training set I, and finally import the data set I, data set II, and data set III respectively into the network for training.

Figure 11. Student pilot flight training process chart

To determine the optimal training methods and parameters for tracking accuracy, this paper designed six pupil prediction tracking network models. The network model parameter settings are shown in Table 1. The activation functions of each layer are linear. The network uses the Euclidean distance between the predicted value and the actual value output by the model as an evaluation index for pupil center detection. The absolute distance between the model's predicted and actual values is used as a measure of pupil radius detection. Network output accuracy is detected by calculating the Euclidean distance between the predicted value and the actual position.

Table 1. Different network parameters

	Model 1 a	Model b	Model c	Model e	Model f	structure 6
Input_size	(3,3)	(3,3)	(2,3)	(2,3)	(1,3)	(1,3)
Hidden_size	100	100	50	50	25	25
Batch_size	1024	5120	1024	5120	6500	5120
Epoch	10	40	10	40	30	10

This article verifies the optimal neural network model parameters by importing Data Set I into the pupil tracking network. During the verification process, the interference due to the pupil positioning output results is eliminated to better verify the pupil prediction network model. We always use the label values provided by the data set for the pupil image parameters and import them into different network models for training. We verified different model parameters by testing the network's pupil position prediction accuracy at the time interval of multiple frames of pupil images and then found the optimal network parameters. It can be seen from the figure that the parameters of Model 1 have the optimal output results. After multiple tests and verifications, this article determines that the parameters of Model 1 are used as the training parameters of the pupil tracking network.

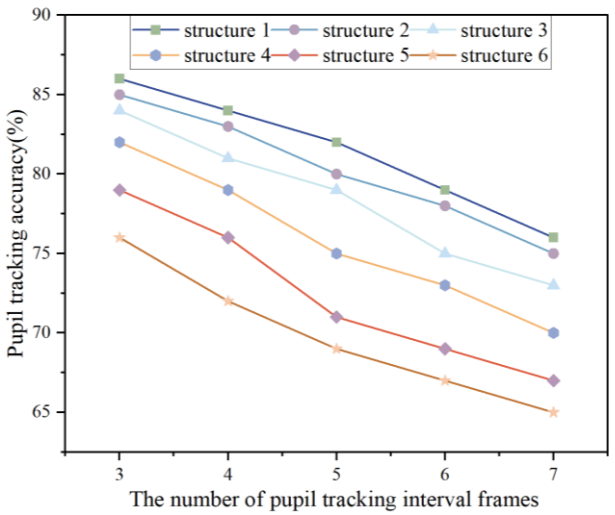


Figure 12. **Output results of different network structures**

We imported the data sets into the FPN, ViT, DeepVOG model [20], and HVit model [21] and the network we designed respectively. The accuracy and stability of the network output results are verified by comparing them with the baseline network FPN, ViT, and the improved network DeepVOG model and HVit. Figures 14, 15, and 16 show the pupil positioning accuracy of different networks after importing different data sets into FPN, ViT, DeepVOG network, HVit network, and our network respectively. We compare and verify the pupil positioning accuracy of each network by counting the error between the predicted value of the pupil center output by different networks and the actual value when it is less than that of different pixels.

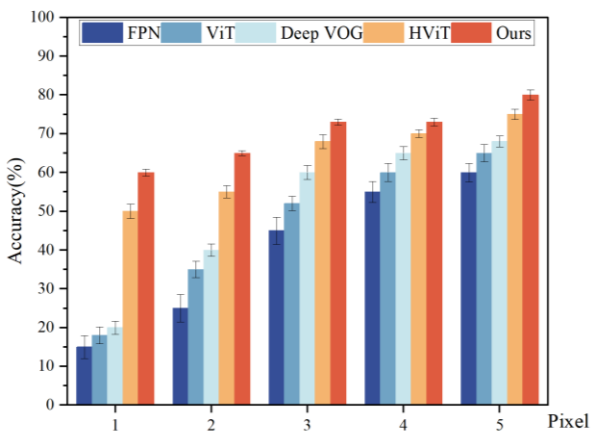


Figure 13. **Comparison of training output results of Data Set I**

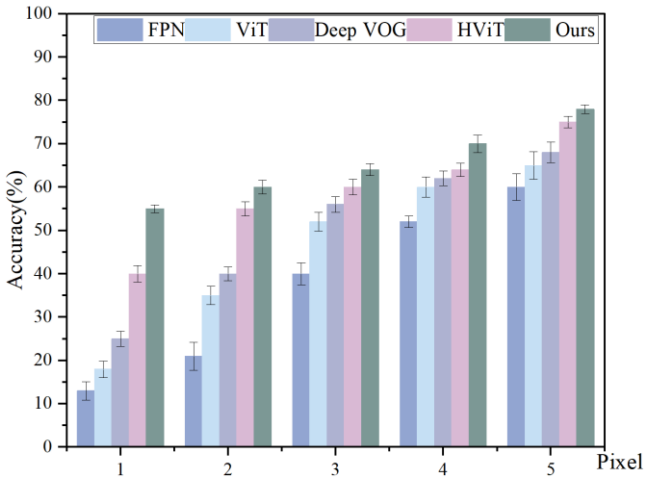


Figure 14. **Comparison of training output results of Data Set Ⅱ**

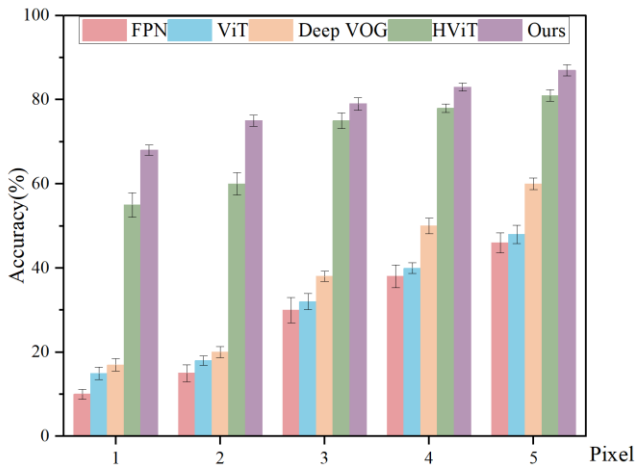


Figure 15. **Comparison of training output results of Data Set Ⅲ**

The comparison of training output results of data sets I, II, and III is shown in Figures 13, 14, and 15. According to the output accuracy comparison chart after importing Data Set I into each network, we can see that the pilots under test had frequent blinking, squinting movements, and changes in lighting conditions during the simulated flight experiment. The tracking accuracy of FPN and ViT networks is greatly affected by interference, and the detection accuracy is both lower than 50%. The visual tracking accuracy of the FPN network is 10% when the error is less than 1 pixel. The detection accuracy of the DeepVOG network is below 60% when the error is 5 pixels. Because the HViT network uses a hybrid neural network structure, its pupil detection accuracy is improved compared to the DeepVOG model. The HViT network's pupil detection accuracy for each error radius is improved by an average of 21% compared with the DeepVOG model. Our proposed model improves visual tracking targets by more than 10% compared to the HViT model, with errors within the 1 pixel and 2 pixels range. Our model achieves the highest accuracy of 85% for visual tracking target detection within an error range of 5 pixels. The visual tracking accuracy for errors less than 1 pixel also reaches 60%. This shows that our model has improved output accuracy for pupil detection data with frequent interference

compared with previous models. For Dataset II, we want to test the stability and accuracy of the network in detecting different eye images. We import the entire dataset into each model for training. Compared with the previous output accuracy of the HViT model when 5,000 eye data were imported, the network we proposed still maintained good training accuracy in the training results of Data Set II. Our network achieves a maximum detection accuracy of 82% for pupil detection, regarding the output results of Dataset III, the proposed model increased by 6% under the same pixel error compared with the HViT model, which has the highest accuracy among all models.

To verify the network's pupil-tracking effect during the pupil detection process, we used the hybrid network model of pupil tracking we designed and the previous pupil-tracking algorithm to collect pupil signals from the pilot. In the pupil-tracking verification experiment, we added a moving simulated sight target to the flight training and asked the pilots in flight training to look at the moving sight target to test the network's pupil-tracking effect. During the verification process, the moving sight targets moved in different modes to simulate different pupil movements of the pilot.

Figure 16. Image of flight training sight target during simulated flight training

Figure 17 shows the pilot's pupil signals collected when the simulated visual target moves discretely in five positions. The pilot's pupil movement tracking trajectory is achieved using the pupil tracking algorithm. The black line represents the pilot's eye movement trajectory, and the green line represents the target's movement trajectory. Comparing Figure 17 with Figure 1, we can find that the pupil-tracking trajectory using the pupil-tracking algorithm is more stable than that without the tracking algorithm. The pupil-tracking algorithm reduces glitches and noise interference caused by blinking and squinting. Figure 18 shows the pilot's pupil signal collected when the simulated visual target is continuously moving. Figure 18(a) and Figure 18(b) show the eye-tracking trajectories collected using other pupil-tracking algorithms and using the pupil-tracking algorithm we designed respectively. It can be seen from Figure 18(a) that the subject's pupil occlusion occurred at 0.5 seconds and 8.4 seconds respectively. When the pupil tracking algorithm was not used, the pupil detection device was unable to obtain the pupil center coordinates, and obvious disturbances occurred. According to Figure 18(b), after using the pupil tracking algorithm, the pilot's pupil tracking was relatively stable. The pilot blinked at 0.53 seconds, and the tracking network predicted the pupil position of the missing pupil frame. Figures 17 and 18 illustrate that the pupil tracking model can reduce the interference of pupil information caused by pupil occlusion of pilots, and improve the accuracy of pupil information collection.

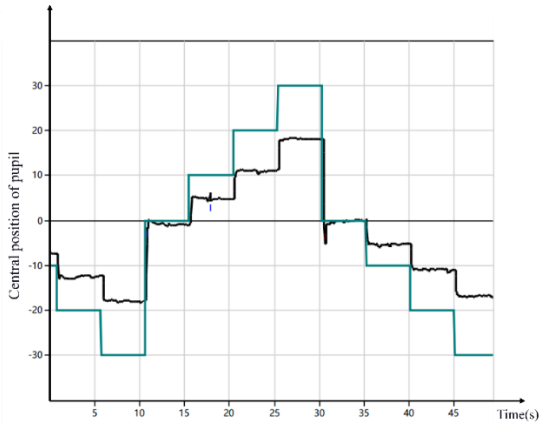
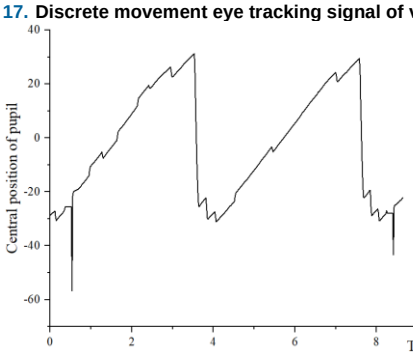
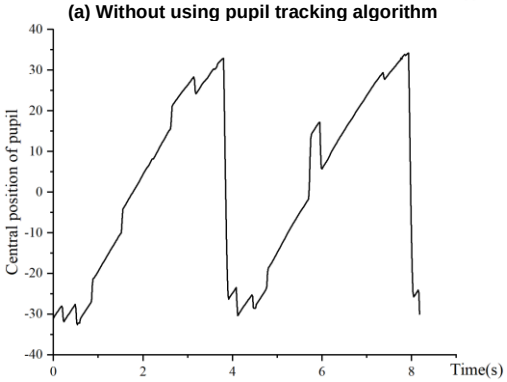


Figure 17. Discrete movement eye tracking signal of visual target



(a) Without using pupil tracking algorithm



(b) Using pupil tracking algorithm

Figure 18. Continuous movement of visual target eye tracking signal

To further test the role of the pupil-tracking network in actual simulated flight training operations, we collaborated with a pilot training center to test the pupil-tracking network algorithm in a real-world pilot simulation environment. Since the relative distance between the pilot's head and the screen of the simulated flight platform remains fixed during flight training, we can further obtain the pilot's gaze direction by calculating the pupil rotation angle. We asked pilots to look at instruments and targets outside the cabin during the training process and used the pilot's gaze direction obtained by the pupil tracking network to evaluate the performance of the neural network in flight training operations. The pilot gaze detection results during flight training are shown in Figure 19. Based on the pilot's gaze detection results, we use the pupil tracking algorithm to accurately distinguish the

pilot's gaze behavior on different targets during actual training flights. In actual flight training, the instrument area that the pilot is looking at is large (radius larger than 40 pixels), and our proposed pupil tracking method can help determine the direction of pupil movement. The output of this algorithm ensures that it is not confused with the gaze behavior of looking at other objects. In addition, we determine and track the pilot's pupil position through multiple frames of pupil images. The erroneous tracking results of the pupil tracking network can be eliminated through consecutive multiple frames of images. Therefore, the pupil tracking algorithm proposed in this article can accurately track the pupil movement of pilot training operations.

Figure 19. Pilot visual gaze test results

IV. Discussions and Conclusion

This paper designs a hybrid neural network pupil-tracking algorithm for pilot pupil detection. First of all, to be more consistent with the pilot's eye movement behavior during flight, we integrate pilot pupil detection into simulated flight training. This paper solves the problem of inaccurate pupil detection accuracy in flight training by designing and improving the network structure. To solve the problem that the eye images are dark and the network is not adaptable to different eye images during the experiment, we connected the improved feature pyramid network and the ViT network, built a pupil detection model, and designed a new loss function. to obtain more accurate pupil coordinates of the pilot. In response to the problem of pupil area occlusion in pilots such as blinking, we predict the short-term coordinates of the pupil by integrating an adaptive LSTM network to solve the pupil occlusion problem. Experiments show that the accuracy of the pupil-tracking hybrid neural network we designed is improved compared to previous pupil-tracking and positioning algorithms. Our method can achieve a maximum tracking accuracy of 87%, which is higher than previous pupil-tracking networks. When the error between the predicted value and the real value of the pupil tracking network we built is less than 5 pixels, its accuracy remains above 65%. We performed pupil signal tracking on pilots using a pupil tracking network. After testing, the pupil tracking network we designed effectively reduced the pupil signal interference caused by the pilot's pupil occlusion.

REFERENCES

[1] West P D B, Sheppard Z A, King E V. Comparison of techniques for identification of peripheral vestibular nystagmus[J]. The Journal of Laryngology & Otology, 2012, 126(12): 1209-1215.
[2] Jiayu Hang,Yuan Gao,Chengfei Li. Research progress on vestibular function and vestibular illusion training for pilots [J]. Medical Journal of National Defending Forces in Northwest China,2019,40(07):452-456.
[3] Gwon O H, Kong T H, Key J, et al. Auto-pattern recognition for diagnosis in benign paroxysmal positional vertigo using principal component analysis: a preliminary study[J]. Research in Vestibular Science, 2022, 21(1): 6-18.
[4] Ciesla M, Koziol P. Eye pupil location using webcam[J]. arXiv preprint arXiv:1202.6517, 2012.
[5] Choi J H, Lee K I, Song B C. Eye pupil localization algorithm using convolutional neural networks[J]. Multimedia Tools and Applications, 2020, 79: 32563-32574.
[6] Kacete A, Royan J, Seguier R, et al. Real-time eye pupil localization using Hough regression forest[C]//2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016: 1-8.
[7] Khan W, Hussain A, Kuru K, et al. Pupil localisation and eye centre estimation using machine learning and computer vision[J]. Sensors, 2020, 20(13): 3785.
[8] Tian D, He G, Wu J, et al. An accurate eye pupil localization approach based on adaptive gradient boosting decision tree[C]//2016 Visual Communications and Image Processing (VCIP). IEEE, 2016: 1-4.
[9] Yiu Y H, Aboulatta M, Raiser T, et al. DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning[J]. Journal of neuroscience methods, 2019, 324: 108307.
[10] Ma H, Shen R, Ye J, et al. High-Automatical and High-Accurate Pupil Location Neural Network via FRST FPL[C]//2023 7th International Conference on Machine Vision and Information Technology (CMVIT). IEEE, 2023: 45-51.
[11] S. Y. Han, H. J. Kwon, Y. Kim and N. I. Cho, "Noise-Robust Pupil Center Detection Through CNN-Based Segmentation With Shape-Prior Loss," in *IEEE Access*, vol. 8, pp. 64739-64749.
[12] Xue P, Wang C, Huang W, et al. Pupil centre's localization with transformer without real pupil[J]. Multimedia Tools and Applications, 2023: 1-18.
[13] Kim S, Jeong M, Ko B C. Energy efficient pupil tracking based on rule distillation of cascade regression forest[J]. Sensors, 2020, 20(18): 5141.
[14] Nenad Markuš, Miroslav Frljak, Igor S. Pandžić, Jörgen Ahlberg, Robert Forchheimer,Eye pupil localization with an ensemble of randomized trees,Pattern Recognition,Volume 47, Issue 2,2014, 578-587.
[15] Lee, Y.W.; Kim, K.W.; Hoang, T.M.; Arsalan, M.; Park, K.R. Deep Residual CNN-Based Ocular Recognition Based on Rough Pupil Detection in the Images by NIR Camera Sensor. *Sensors* **2019**, *19*, 842.
[16] Wei K, Yang Q, Yang X, et al. Application of a pupil tracking method based on Yolov5-Deeplabv3+ fusion network on a new BPPV nystagmus recorder[C]//International Conference on Biomedical and Intelligent Systems (IC-BIS 2022). SPIE, 2022, 12458: 948-955.
[17] Newman J L, Phillips J S, Cox S J. Detecting positional vertigo using an ensemble of 2D convolutional neural networks[J]. Biomedical Signal Processing and Control, 2021, 68: 102708.
[18] Friedrich M U, Schneider E, Buerklein M, et al. Smartphone video nystagmography using convolutional neural networks: ConVNG[J]. Journal of Neurology, 2023, 270(5): 2518-2530.
[19] Shi, L., Wang, C., Tian, F. et al. An integrated neural network model for pupil detection and tracking. Soft Comput 25, (2021).10117–10127
[20] Deng W, Huang J, Kong S, et al. Pupil trajectory tracing from video-oculography with a new definition of pupil location[J]. Biomedical Signal Processing and Control, 2023, 79: 104196.
[21] Yiu Y H, Aboulatta M, Raiser T, et al. DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning[J]. Journal of neuroscience methods, 2019, 324: 108307.
[22] W. Fuhl, G. Kasneci and E. Kasneci, TEyeD: Over 20 million real-world eye images withpupil, eyelid, and iris 2D and 3D segmentations, 2D and 3D landmarks, 3D eyeball, gazevector, and eye movement types, 2021 IEEE Int. Symp. Mixed and Augmented Reality(ISMAR) (4–8 October 2021, Bari, Italy, 2021), pp. 367–375.
[23] L. Świrski, A. Bulling and N. Dodgson, Robust real-time pupil tracking in highly o®-axisimages, Proc. Symp. Eye Tracking Research and Applications (28–30 March 2012, Santa Barbara California, United States, 2012), pp. 173–176.
[24] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
[25] Memory L S T. Long short-term memory[J]. Neural computation, 2010, 9(8): 1735-1780.

Heming Zhang received a master's degree in electronic information engineering from Xi'an Technological University in 2022. His research areas include artificial intelligence, computer vision and human-machine hybrid direction.

Changyuan Wang received Master's degree in Applied Mathematics from the Northwest University of Technology in April 1988. In 2011, he graduated from the Xi'an University of Technology majoring in Mechanical Engineering and received a Doctor's Degree. At present, he is Professor and Doctoral Supervisor of the Xi'an Technological University.