Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 5675914, 11 pages
https://doi.org/10.1155/2020/5675914

Research Article

Enhanced Forest Microexpression Recognition Based on Optical Flow Direction Histogram and Deep Multiview Network

Huanmin Wang [1,2,3]

[1]*Mechatronics T&R Institute, Lanzhou Jiaotong University, Lanzhou 730070, China*
[2]*Engineering Technology Center for Informatization of Logistics & Transport Equipment, Lanzhou 730070, China*
[3]*Industry Technology Center of Logistics & Transport Equipment. Gansu, Lanzhou 730070, China*

Correspondence should be addressed to Huanmin Wang; wanghuanmin@mail.lzjtu.cn

Received 18 July 2020; Revised 13 August 2020; Accepted 17 August 2020; Published 31 August 2020

Guest Editor: Yi-Zhang Jiang

In order to recognize the instantaneous changes of facial microexpressions in natural environment, a method based on optical flow direction histogram and depth multiview network to enhance forest microexpression recognition was proposed. In the preprocessing stage, the histogram equalization of the acquired face image is performed, and then the dense key points of the face are detected. According to the coordinates of the key points and the face action coding system (FACS), the face region is divided into 15 regions of interest (ROI). In the feature extraction stage, the optical flow direction histogram feature between adjacent frames in ROI is extracted to detect the peak frame of microexpression sequence. Finally, the average optical flow direction histogram feature of the image sequence from the initial frame to the peak frame is extracted. In the classification stage, firstly, the head pose parameters under horizontal degrees of freedom are estimated to eliminate the influence of head pose motion, and a forest multiview conditional probability model based on deep multiview network is established. Conditional probability and neural connection function are introduced into the node splitting learning of random tree to improve the learning ability and distinguishing ability of the model on the limited training set. Finally, multiview-weighted voting is used to determine the categories of facial microexpressions. Experiments on CASME II microexpression dataset show that the proposed method can effectively describe the changes of microexpressions and improve the recognition accuracy compared with other new methods.

1. Introduction

Facial expression is a facial movement that reflects people's spirit and emotions. About 55% of information is transmitted through facial expression when people communicate. Therefore, it plays an important role in social communication. The efficiency of facial expression recognition is an important part of human-computer interaction, and the related research on facial expression recognition has become a hot spot [1]. However, most of the research on expression focuses on traditional expressions, i.e., macroexpressions or full-expressions, while less on microexpressions. Until recently, more and more scholars began to pay attention to the study of microexpressions [2]. Literature [3] focuses on such a meaningful topic and investigates how to make full advantage of the color information provided by the microexpression samples to deal with the microexpression recognition (MER) problem. Microexpressions are difficult to detect, and their duration is only 1/25 s to 1/5 s. Microexpressions may contain not only all muscular movements of ordinary expressions but also only part of them. It expresses the true feelings that human beings try to hide. It is a spontaneous expression. The above properties of microexpressions make it a window to understand human real feelings [4, 5]. Therefore, microexpressions have many potential applications, such as criminal investigation, national defense security, clinical diagnosis, and human-computer interaction. However, the characteristics of short duration, low intensity, and usually involving only local motion of microexpressions pose a great challenge to the recognition of microexpressions.

In order to recognize instantaneous microexpression changes in natural environment, a method based on optical

flow direction histogram and depth multiview network to enhance forest microexpression recognition is proposed, which is used to recognize microexpression in multipose view. The main innovations of this method are as follows:

(1) Accurately detect the dense key points of the face and divide the face area into ROI, which is different from the existing algorithms. The proposed algorithm extracts the average optical flow direction histogram features of the image sequence from the initial frame to the peak frame after extracting the adjacent two parts of the specific ROI and eliminates the interference of the unrelated region.

(2) Introducing visual angle network-enhanced forest neural network and conditional probability model into random forest learning can reduce the influence of attitude change and improve the classification accuracy, so as to obtain better classification results of microexpressions from multiple perspectives.

The organizational structure of this paper is as follows. Section 2 introduces the related research on microexpression recognition. Section 3 introduces the overall framework of the proposed methodology. Section 4 is the preprocessing and feature extraction of face images. Section 5 introduces enhanced forest microexpression classification based on multiview network. Section 6 is experiment. Section 7 summarizes the article and points out the future research directions.

2. Relevant Research

With the development of computer vision, some researchers began to try to extract various features for microexpression recognition. Li et al. [6] studied dynamic microexpression recognition under unconstrained conditions. A dynamic facial expression recognition based on multivisual and audio descriptor (DFER-MVAD) algorithm was proposed to extract dynamic microexpression features by using spatial-temporal local feature description of multivisual descriptors. However, this method is prone to overfitting when the amount of data is large. Zhao and Pietikäinen [7] present a local binary pattern from three orthogonal planes (LBP-TOP) for microexpression recognition. Pfister et al. [8] designed a two-stage system to realize the recognition of spontaneous microexpressions for the first time. In the first stage, the system uses Temporal Interpolation Model (TIM) to normalize the total number of frames of microexpression sequence, so as to solve the problem of short-term video. In the second stage, the system uses LBP-TOP to extract spatiotemporal local texture descriptor (SLTD) to recognize microexpressions. However, in uncertain environments, the robustness needs to be further improved. In [9], the RPCA method is used to extract minute motions of micro-expressions to replace differential images, and a Laplacian-based feature selection method is used to enhance the distinction between classes. A new differential spatiotemporal LBP (DiSTLBP-RIP) feature based on improved integral projection technique is proposed, which achieves good results in microexpression recognition. However, in uncertain environments, the recognition accuracy is good. The accuracy needs to be further improved. To the features based on LBP, a 3D gradient descriptor is proposed in literature [10], which combines K-means algorithm to identify the three stages of microexpression occurrence and emotional categories. In addition, with the wide application of optical flow in behavior detection and behavior recognition, a novel histogram of oriented optical flow (HOOF) feature based on deep learning is proposed in literature [11] for microexpression detection. Liu et al. [12] propose a main directional mean optical-flow (MDMO) feature to realize microexpression recognition and to achieve good results. On the contrary, with the successful application of deep learning in motion recognition, face authentication, expression recognition, etc., Patel et al. [13] apply deep learning to microexpression recognition, but because deep learning relies on large-scale datasets, the recognition rate of the proposed depth features does not exceed that of traditional features.

In recent years, deep convolutional neural network (CNN) can automatically learn high-level features of images. Good progress has been made in the field of image recognition, and good recognition results have been achieved in the field of facial expression recognition based on CNN [14]. In [15], support vector machine (SVM) is used to replace the Softmax layer in traditional CNN, and 71.20% accurate recognition rate is achieved on FER2013 facial expression dataset. Yan et al. [16] proposed cascaded deep space-time network and apparent network models. Accurate recognition rates of 95.22% and 81.46% were obtained on enhanced CK+ and Oulu-CASIA expression sets, respectively. As can be seen, CNN has strong ability of feature learning and expression through multilayer neural network feedback learning, but it often relies on a large number of enhanced training datasets and powerful GPU computing power.

When facing the research of multiview facial expression recognition based on limited training data, deep learning methods are prone to problems such as overtraining and overfitting. In order to solve this problem, Zhou and Shi [17] try to pretrain CNN model on large-scale Image Net dataset and then fine-tune network parameters to achieve good expression recognition results. Zheng [18] tries to combine artificial feature extraction with CNN and proposes a facial expression recognition method based on local feature SIFT network, which achieves 78.9% accuracy on BU-3DFE multiview facial expression dataset. In addition, cascaded use of the two classifiers reduces the training parameters of the model, and some progress has been made in the field of object recognition. Although these methods improve the accuracy of facial expression recognition on limited datasets to a certain extent, they are still an open challenge in the natural environment with large gesture changes. To solve this problem, an enhanced forest microexpression recognition method based on optical flow direction histogram and depth

multiview network is proposed for automatic face microexpression recognition in attitude changing environment.

3. The Overall Framework of the Proposed Methodology

The overall structure of the proposed microexpression recognition is shown in Figure 1. Because microexpression usually involves only local motion, the proposed algorithm refines the range of feature extraction by dividing a specific ROI while eliminating the interference of irrelevant regions. Then, the peak frames of the microexpression sequence are detected by extracting the average optical flow direction histogram feature in the selected ROI. Finally, the image sequence from the initial frame to the peak frame is extracted. Based on the average optical flow direction histogram feature, a conditional probability model of forest multiview enhancement based on depth multiview network was established to recognize microexpressions.

4. Preprocessing and Feature Extraction of 4 Face Images

4.1. Face Image Acquisition and Preprocessing

4.1.1. Histogram Equalization. The basic idea of histogram equalization is that the histogram of the original image can be transformed into a uniformly distributed histogram [19, 20]. The dynamic range of the gray value of image pixel is increased, so the overall contrast of the image is enhanced. The operation procedures of histogram equalization are as follows.

(1) Give the gray value of the original image:
$r_0, r_1, \ldots, r_k$, $(k = 0, 1, \ldots, L-1)$,L is the total gray level of the image.

(2) Let $n(r_k)$be the probability of the occurrence of r_k gray levels so that the histogram can be obtained as follows:

$$p(r_k) = \frac{n(r_k)}{N}, \tag{1}$$

where $N$is the sum of all the pixels of image $f(x, y)$, and there are

$$\sum_{k=0}^{L-1} p(r_k) = 1. \tag{2}$$

(3) Calculate the cumulative probability of image pixel distribution:

$$p_f(r_k) = \sum_{j=0}^{k} p(r_j). \tag{3}$$

The mapping relationship between gray levels $r_k$and s_k is determined, and the cumulative probability $p_f(r_k)$is taken as the gray level transformation function $T(r_k)$.

(4) Calculate the values of each gray level, where $0 \le r_k \le 1$.

After output of the equalization result, the histogram distribution of the processed image becomes more uniform, and the pixels of the image are distributed to each gray level.

4.1.2. Face Key Point Detection and Region of Interest Partition. For ROI partition, the key points of dense face must be accurately detected first. In this paper, we use OpenFace, a facial behavior analysis tool proposed by Baltrugaitis and others, to extract the key points of face [21]. On the basis of constrained local neural model (CLNF) [22], this tool first detects the face frame using the face detector provided by Dlib Library and then learns a simple linear mapping from the face frame to the boundary of 43 key points to initialize the CLNF model. Later, more reliable mapping diagrams are calculated using the newly proposed local neural field (LNF), and nonuniform regularised mean shift is used as an optimization method to take the reliability of each region into account so as to make the results more accurate. In addition, OpenFace trains key point distribution models for eyes, lips, and eyebrows and fuses these key points. The detection effect of 43 key points of faces is shown in Figure 2.

After the dense key points are detected, the face area is divided into 15 ROIs according to the coordinates of the key points and FACS, as shown in Figure 2. FACS defines action unit (AU) according to the contraction or relaxation state of one or more facial muscles. Usually, an AU corresponds to a local movement of the face, so AU is often used to analyze facial movement [23]. Fifteen ROIs and corresponding AUs are classified in this paper, as shown in Table 1. There are many strategies for ROI partition, but the general principle is that ROI partition can neither be too sparse nor too dense. If the partition is too sparse, useful information may be omitted, and if the partition is too dense, redundant information may be introduced. Eyebrows and mouth are the most frequent areas in the process of microexpressions, so these areas are more carefully divided, while cheeks and foreheads are sparsely divided. In many previous segmentation strategies, the eye region is often regarded as a key object, but blinking can occur at any stage of expression, so this region may bring more misleading information, so this paper only divides the eyebrow region. At the same time, according to the situation of facial region movement, this paper chooses the area which is most closely related to microexpression and divides ROI into sparse areas as far as possible, which not only eliminates the interference of nonkey areas but also reduces the dimension of the average optical flow direction histogram feature. If the optical flow direction is divided into eight intervals, the dimension of the average optical flow direction histogram feature in a frame is

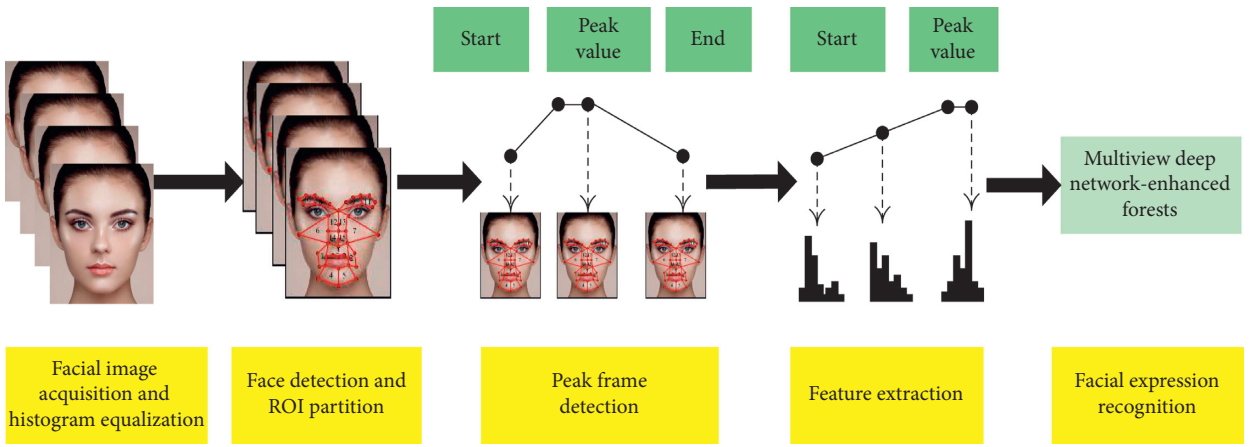

Figure 1: The overall structure block diagram of the proposed microexpression recognition method.

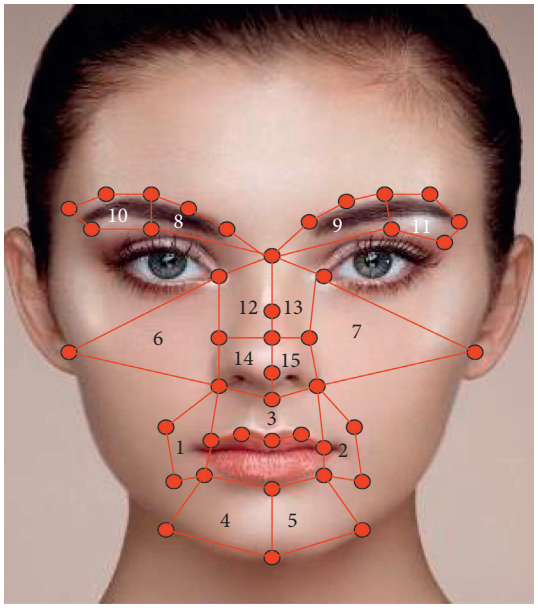

Figure 2: Key points and ROIs distribution.

Table 1: 15 ROIs and corresponding AUs and facial movements.

ROIs	AUs	Facial Movements
ROI^1,ROI^2	AU12, AU14, AU15, AU18	Lip corner
ROI^3,ROI^4	AU17, AU26	Chin
ROI^5	AU10	Upper lip
ROI^6,ROI^7	AU6, AU13	Cheek
ROI^8,ROI^9	AU1, AU4	Inner eyebrow
ROI^{10},ROI^{11}	AU2	Outer eyebrow
ROI^{12},ROI^{13}	AU9	Nose root
ROI^{14},ROI^{15}	AU16	Nose root

$15 \times 8 = 120$ dimension (in this paper, only the directional parameters of the optical flow are selected).

4.1.3. Lucas–Kanade (LK) Optical Flow Algorithms. Optical flow is used to estimate the relative motion between two frames occurring in time t and $t + \Delta t$. There are two kinds of optical flow computing techniques: dense optical flow and sparse optical flow, but the computational complexity of dense optical flow is higher than that of sparse optical flow. Therefore, LK optical flow algorithm [24, 25] is usually used. The algorithm is based on the following three assumptions:

(1) Constant brightness: that is, the brightness value of the pixels in the picture remains unchanged in a very short time.

(2) Micromotion: that is to say, the motion scale and frame of the pixel block in the picture are very small compared with the time change of the frame.

(3) Spatial consistency: that is, the pixels in the neighborhood of the picture have the same motion. Assuming that, in time, the brightness values of a pixel in a frame are $H(x, y, t)$ and $H(x + \Delta x, y + \Delta y, t + \Delta t)$; based on the first assumption of "constant brightness," it can be obtained by

$$H(x, y, t) = H(x + \Delta x, y + \Delta y, t + \Delta t). \tag{4}$$

At the same time, according to the hypothesis of the second "micromotion," Taylor expansion is carried out for equation (1):

$$H(x + \Delta x, y + \Delta y, t + \Delta t) = H(x, y, t) + \frac{\partial H}{\partial x}\Delta x + \frac{\partial H}{\partial y}\Delta y + \frac{\partial H}{\partial t}\Delta t + H.O.T. \tag{5}$$

By neglecting the higher order terms of order 2 or more and combining formula (1), it can be obtained that

$$\frac{\partial H}{\partial x}\Delta x + \frac{\partial H}{\partial y}\Delta y + \frac{\partial H}{\partial t}\Delta t = 0. \tag{6}$$

When $\Delta t \longrightarrow 0$,

$$\frac{\partial H}{\partial x}V_x + \frac{\partial H}{\partial y}V_y + \frac{\partial H}{\partial t} = 0, \tag{7}$$

where V_x and V_y are x and $y$components of optical flow velocity, respectively. The derivative forms are expressed by H_x, H_y, and H_t:

$$H_x V_x + H_y V_y = -H_t. \tag{8}$$

Formula (8) shows that there are two unknown parameters for a pixel. To solve this problem, LK optical flow algorithm still needs to rely on the third assumption, namely, spatial consistency. If a window of $5*5$ is selected around the current pixel, the following 25 equations can be obtained:

$$\begin{bmatrix} H_x(p_1) & H_y(p_1) \\ H_x(p_2) & H_y(p_2) \\ \vdots & \vdots \\ H_x(p_{25}) & H_y(p_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -\begin{bmatrix} H_t(p_1) \\ H_t(p_2) \\ \vdots \\ H_t(p_{25}) \end{bmatrix}. \tag{9}$$

LK algorithm uses least squares estimation method to solve the overconstrained problem of formula (9) by minimizing $\|A\,d - b\|^2$. The standard form of LK algorithm is as follows:

$$\left(A^T A\right) d = A^T b. \tag{10}$$

Formula (9) is converted into the following form:

$$\begin{bmatrix} \sum H_x H_x & \sum H_x H_y \\ \sum H_x H_y & \sum H_y H_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -\begin{bmatrix} \sum H_x H_t \\ \sum H_y H_t \end{bmatrix}. \tag{11}$$

Finally, we can obtain

$$\begin{bmatrix} u \\ v \end{bmatrix} = \left(A^T A\right)^{-1} A^T b. \tag{12}$$

4.1.4. Peak Detection. Face alignment is a common preprocessing method in face recognition, but it will cause a certain degree of deformation of the face after face alignment, which will have a great impact on the very low intensity of microexpressions. In addition, considering that microexpressions occur in very short time and the degree of change between frames is very small, this paper does not carry out traditional face alignment, but first carries out peak frame detection. A complete microfacial sequence can be divided into three stages: initiation, initiation, and termination. The initiation and initiation stages can better reflect the microfacial category. For example, happy microfacial expressions are usually accompanied by mouth corner rising, while there is not much effective information for microfacial expression recognition in the fall stage. In addition, shorter microexpression sequences can reduce the noise caused by head posture changes, so microexpression sequences from the initial frame to the peak frame can be selected by peak detection. In this paper, we first extract the optical flow field between two adjacent frames in five regions around the eyebrow, mouth, and chin and then divide the optical flow direction into eight regions (as shown in Figure 3).

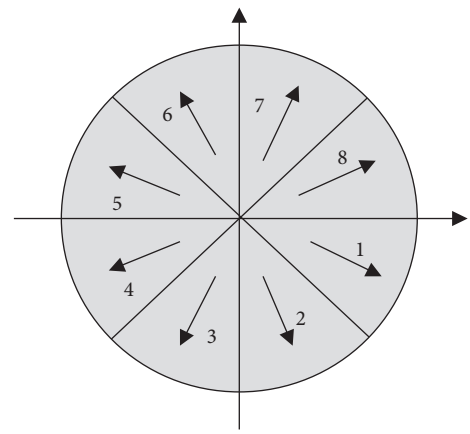

Figure 3: Diagram of 8 optical flow direction intervals.

Finally, the histogram features of optical flow direction in each area scale (shown in Figure 4(b) are calculated and the sum of downward and upward pixels is counted. The difference between them is δ as follows:

$$\delta_i^k = \sum_{p_i^k \in B_{\text{down}}} p_i^k - \sum_{p_i^k \in B_{\text{up}}} p_i^k. \tag{13}$$

Here, $i = 1, 2, \ldots, n$ is the index of frames of microexpression sequence. $k = 1, 2, \ldots, 5$ is the number of five regions R. p_i^k is a pixel belonging to $R_i^k.p$, B_{down} is the set of intervals 1, 2, 3, and 4 in the downward direction. B_{up} is a set of directions upward, i.e., intervals 5, 6, 7, and 8.The change of the positive and negative values of δ indicates the change of the direction of motion, so the peak frame is the number of frames when the positive and negative changes of δ. Photorheological changes of microexpression sequence are shown in Figure 4.

4.2. Microexpression Feature Extraction. Considering that the intensity of microexpressions is very low, even in peak frames, it is difficult to recognize them with traditional static features, so this paper extracts the dynamic optical flow characteristics of microexpressions. In addition, the microexpression sequence recorded by the high-speed camera changes very slightly between frames so that the main direction (including the optical flow direction interval with the largest number of pixels) in each ROI is not obvious compared with other directions and is easily interfered by other directions, so the optical flow $[V_x, V_y]$ between each frame and the first frame is extracted. Then, the Euclidean coordinates are transformed into polar coordinates (p, θ), and the direction of the optical flow is divided into eight directions, as shown in Figure 3. Finally, the histogram characteristics of optical flow direction in 15 ROIs are calculated. After calculating the optical flow direction histogram features in 15 ROIs, the optical flow direction histogram features ϑ_i of a frame can be obtained as follows:

$$\Omega = \frac{1}{n_f} \sum_{i=1}^{n_f} \vartheta_i, \tag{14}$$

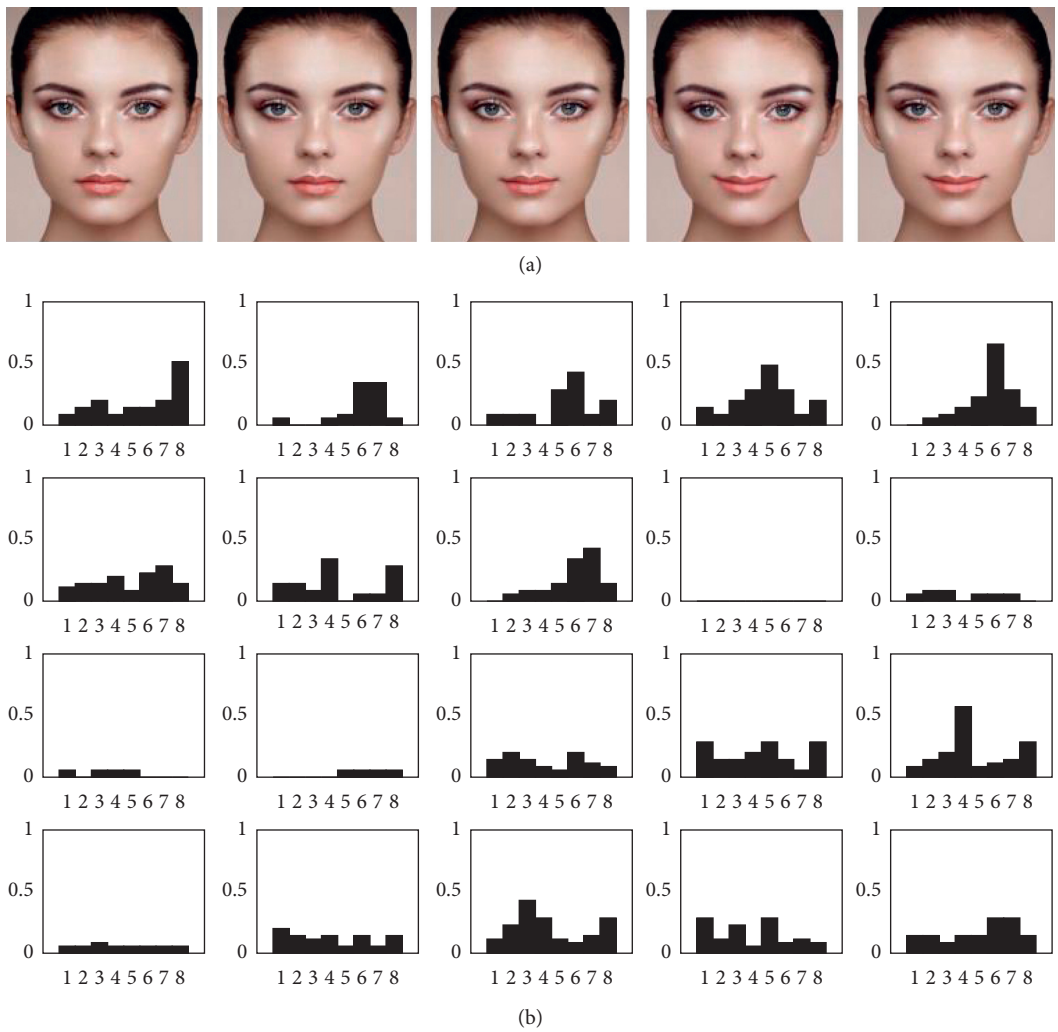

Figure 4: Photorheological chart of microexpression Sequence. (a) The original sequence of pictures in three stages of microexpression. (b) Optical flow direction histogram between adjacent frames in the whole process.

where n_f is the total number of frames from the start frame to the peak frame of the current microexpression sequence and ϑ_i is the optical flow direction histogram feature of frame f_i.

5. Enhanced Forest Microexpression Classification Based on Multiview Network

5.1. Multiview Network-Enhanced Forest Model. The training block diagram of multiview network enhanced forest is shown in Figure 5. It is composed of multiple network enhancers under different attitude perspectives. Each subforest is made up of a deep migration facial feature training depth network enhancement tree from a special perspective. Each deep network reinforcement tree learning includes three processes: reinforcement joint layer, node learning layer, and multiperspective decision-making voting layer [26]. Under the condition of head posture, the enhanced joint layer uses the conditional depth network from different perspectives to enhance the expression of the average optical flow direction histogram features of the microexpression sequence, so as to obtain more discriminative expression of enhanced depth features. In the node learning layer, NCSF splitting function is used to rank the depth enhancement features at each splitting node. The learning condition depth network enhances the forest splitting nodes and grows iteratively to the leaf nodes. The multiview decision-making voting layer chooses the probability model of leaf nodes from different perspectives by weights and classifies the multiview expressions.

Under different head postures, based on the above pretraining VGG-Face full connection layer, the learning condition feature expression set P: $\{P = (\vartheta_i, \theta_i), \pi\}$ of deep migration feature is obtained. Of which, ϑ_i is the average optical flow direction histogram feature of microexpression sequence, θ_i is the current head posture, and π is the expression category.

5.1.1. Enhanced Joint Layer. Connection function f_n based on the hidden layer in CNN enhances conditional feature expression P of face sub-block and uses enhanced feature

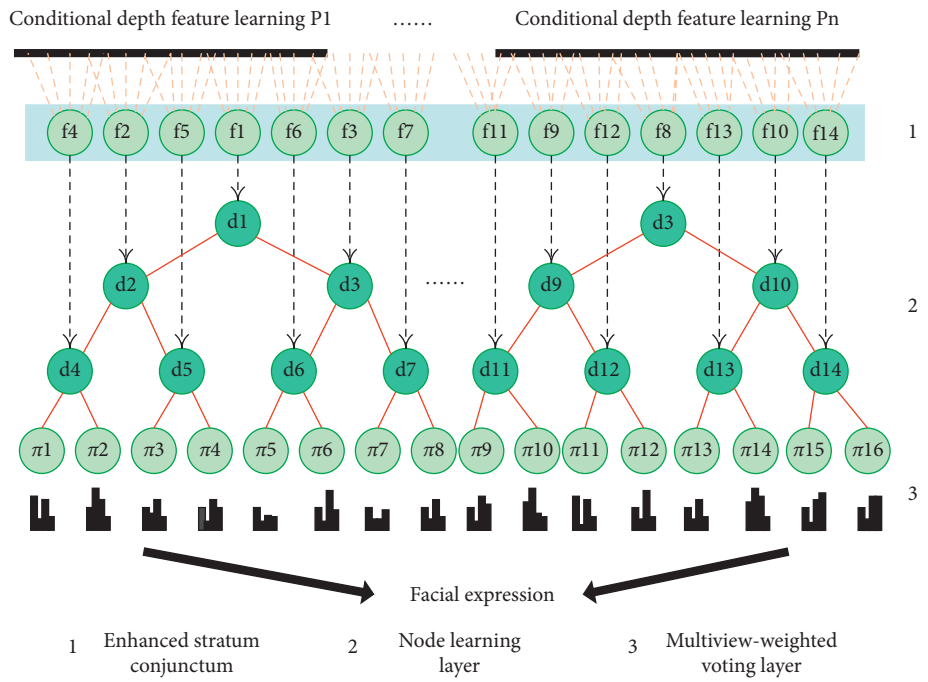

Figure 5: Training block diagram of multiview network enhanced forest.

expression as node feature selection of network-enhanced forest:

$$d_n\left(P, Y \mid \Omega_\theta\right) = \sigma\left(f_n\left(P, Y \mid \Omega_\theta\right)\right), \tag{15}$$

where σ is sigmoid function; Ω_θ is expression subforest under different attitude angles; n is a splitting node of deep network enhanced forest; and$Y$ is network model parameters. Deep network enhances the number of nodes of a training tree in the forest, i.e., enhances the feature dimension for the output of the enhanced joint layer.

5.1.2. Node Learning Layer. In order to learn and grow split nodes, an NCSF splitting model is proposed in this paper, which combines the information gain (IG) of decision tree with the loss function of the deep network model, and is used to enhance the node growth of multiview deep network.

Random stochastic gradient descent (SGD) to minimize model risk:

$$Y^{(t+1)} = Y^{(t)} - \frac{\eta}{|B|} \sum_{(P,\pi)\in B} \frac{\partial L(Y, \pi; P)}{\partial Y}, \tag{16}$$

where $\eta > 0$ is the learning rate, π is the expression tag, and B is the randomly extracted feature subset, and the loss item of the training feature set is

$$L(Y, \pi; P) = -\sum_n p\left(\pi \mid d_n, Y, P\right) \log\left(p\left(\pi \mid d_n, Y, P\right)\right), \tag{17}$$

where $p(\pi \mid d_n, Y, P)$ is the probability of learning expressions. According to the derivative chain rule,

$$\frac{\partial L(Y, \pi; P)}{\partial Y} = \sum_{n\in N} \frac{\partial L(Y, \pi; P)}{\partial f_n(P, Y \mid \theta)} \cdot \frac{\partial L(Y, \pi; P)}{\partial Y}, \tag{18}$$

where the second derivative can be obtained by optimizing the network parameters and the first derivative depends on the selection of the left and right subnodes of the tree. Available:

$$\frac{\partial L(Y, \pi; P)}{\partial f_n(P, Y \mid \theta)} = -\sum_n \left(d_n^R(P, Y \mid \theta) + d_n^L(P, Y \mid \theta)\right), \tag{19}$$

where $d_n^R(P, Y \mid \theta)$ and $d_n^L(P, Y \mid \theta)$ represent the right and left nodes of the tree, respectively. When the information gain IG is maximum, the left and right subnodes of the spanning tree are split. When the depth of the tree reaches the maximum or the loss function converges iteratively, the leaf nodes are generated; otherwise, the iterative node learning is continued.

5.1.3. Multiperspective-Weighted Voting Layer. After generating leaf nodes, $p(\pi \mid \theta, l)$ is defined as a probability expression model from the perspective of head pose on leaf nodes, which can be expressed as a multiparameter Gauss mixture model:

$$p(\pi \mid \theta, l) = N\left(\pi \mid \theta; \overline{\pi \mid \theta}, \sum_l^{\pi\mid\theta}\right), \tag{20}$$

where $\overline{\pi|\theta}$ and $\sum_l^{\pi|\theta}$ are expression probability mean and covariance matrices on leaf nodes.

In order to eliminate the influence of different perspectives on facial expression recognition, a multiview weight voting algorithm was used to vote the expression

probability of face block on leaf node l from different perspectives, and the expression category probability of view subforest Ω_θ was obtained:

$$p\left(\pi \mid \Omega_\theta\right) = \frac{1}{k}\sum_{t=1}^{k} C_\theta p_{a_t}(\pi \mid \theta, l), \tag{21}$$

where $a_t$are trees in subforests, C_θ are weights of subforests in perspective, and k are trees of training trees in subforests.

5.2. Head Pose Parameter Estimation. In order to eliminate and correct the influence of head posture on expression recognition, accurate head posture estimation is the prerequisite for successful multiview expression recognition. Among the three dimensions of head posture movement, because the horizontal perspective in the natural environment has the greatest influence on expression, this paper divides head posture into nine disjoint subsets in the horizontal direction: {90°, 60°, 45°, 30°, 0°, −30°, −60°, −90°}, that is, nine different perspectives, and takes the probabilistic model of training head posture from each perspective as the prior probability of expression recognition. The multiparameter Gauss probability model for head attitude:

$$p(\theta \mid l) = N\left(\theta; \overline{\theta}, \sum_{l}^{\theta}\right), \tag{22}$$

where $\overline{\theta}$ and $\sum_l^\theta$ are the mean and covariance matrices of head pose probability on leaf nodes.

5.3. Multiview Conditional Probability Model and Microexpression Recognition. A priori conditional probability model of facial expression based on head posture parameters is established to simulate multiview facial expression probability:

$$p(\pi \mid P) = \int p(\pi \mid \theta, P)p(\theta \mid P)\mathrm{d}\theta. \tag{23}$$

In order to obtain $p(\pi \mid \theta, P)$, the training set is first divided into training subsets from different perspectives based on θ. The θ-parameter space can be discretized into disjoint subsets 33, and then formula (23) can be transformed into disjoint subsets Ω_c:

$$p(\pi \mid P) = \sum_{c}\left(p\left(\pi \mid \Omega_c, P\right)\right)\int p(\theta \mid P)\mathrm{d}\theta, \tag{24}$$

where $p(\theta \mid P)$ is obtained by head pose estimation and conditional probability $p(\pi \mid \Omega_c, P)$ can be obtained by training based on disjoint subset Ω_c. Finally, the probability $p(\pi \mid P)$ of expression categories is calculated by conditional multiview weight voting, i.e.,

$$p(\pi \mid P) = \frac{1}{C}\sum_{c=1}^{C} p\left(\pi \mid \Omega_c\right)p\left(\theta \mid \Omega_c\right), \tag{25}$$

where C is the number of subforests Ω_c, $p(\pi \mid \Omega_c)$is derived from equation (7), and $p(\theta \mid \Omega_c)$ is the conditional probability of head posture from this perspective.

6. Experiments

In order to verify the validity of the average optical flow direction histogram feature, experiments were carried out on CASME II microexpression dataset and compared with MDMO [12], DFER-MVAD [6], and DiSTLBP-RIP [9]. The hardware environment is MacBook Pro (13″, 2017), CPU_Intel Core i5/graphics card_Intel Iris Plus Graphics 640, 2.3 GHz, and hard disk_256 G/memory_8 G/.

6.1. Dataset and Parameter Settings. CASME II database is maintained by Institute of Psychology, Chinese Academy of Sciences. Its temporal resolution is 200 fps. At the same time, the area of the face area in CASME II is about 280 ∗ 340 pixels, which can extract more details from the short duration and low emphasis microexpressions. In the process of CASME II acquisition, some participants were required to maintain a neutral state throughout the video viewing process, while the other participants inhibited facial movement only when they realized that microexpressions were about to appear. Based on the above characteristics, CASME II is selected as the experimental dataset, which contains 247 microfacial expression samples selected from nearly 3000 facial movements of 26 participants. Each microfacial expression sample is marked with start frame, peak frame, and end frame. In addition, action units and emotional categories are also marked. Because the number of samples in some categories is too small, the 247 samples are usually divided into six categories: angry, disgust, fear, happy, sadness, and surprise.

The training set includes 117 microfacial images selected from human face motion, and the test set includes 130 microfacial images selected from human face motion. In addition, the experiment is based on the CAFE framework to train depth network to enhance forest. Some important parameters in training are as follows: learning rate is 0.01, epochs are 5000, number of split iterations is 1000, and depth of tree is 15.

6.2. Recognition Results on CASME II Dataset. In order to reduce the influence of posture on facial expression recognition, nine types of head posture estimation are performed on CASME II dataset using the proposed method. As shown in Table 2, the experimental parameters are consistent with those of facial expression recognition. It can be seen that the average accuracy of the proposed method is 98.40%, which shows that it can well register head pose changes.

Figure 6 illustrates the proposed method for recognizing confusion matrix in CASME II microexpression dataset. It can be seen that the average recognition accuracy under different perspectives is 86.63%.

6.3. Comparison and Analysis with Other Methods. Three methods, MDMO, DFER-MVAD, and DiSTLBP-RIP, were used to estimate nine kinds of head posture in CASME II dataset. Figure 7 shows the average recognition

Table 2: Accuracy of recognition of different attitude perspectives in CASME II microexpression dataset.

Attitude Perspective (°)	Accuracy rate (%)	Attitude Perspective (°)	Accuracy rate (%)
−90	99.69	30	98.89
−60	97.31	45	97.71
−45	97.47	60	97.21
−30	99.12	90	99.49
0	99.51	—	—

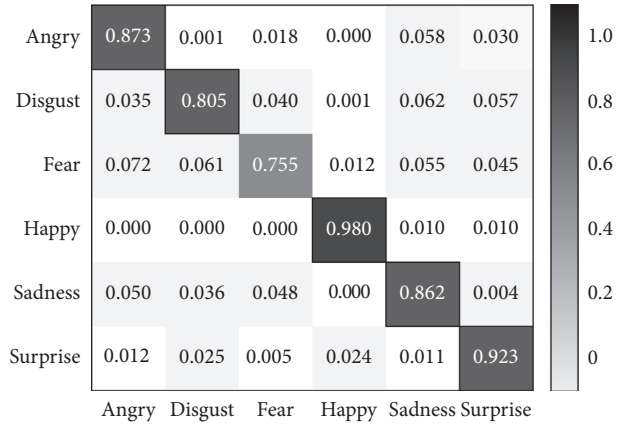

Figure 6: Recognition confusion matrix of CASME II microexpression dataset.

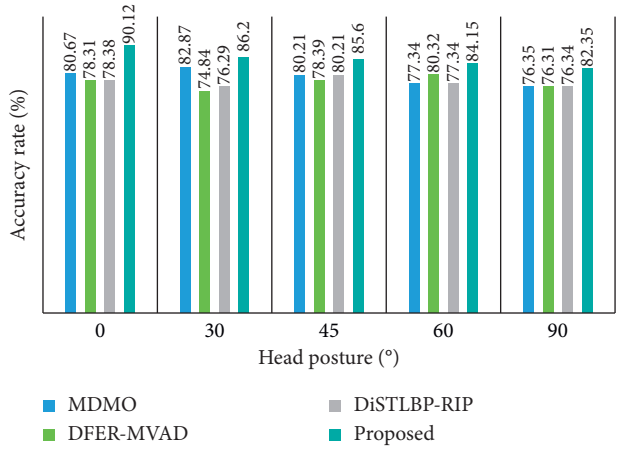

Figure 7: Comparison of recognition accuracy of microexpression from different perspectives.

rate of the proposed method from different perspectives compared with MDMO, DFER-MVAD, and DiSTLBP-RIP. It can be seen that the recognition rate of the proposed method is higher than that of the other three methods at five different angles in the horizontal right direction. The accuracy of the proposed method is the highest when the head posture is 0 degrees, which is 90.12%, while the accuracy of the 90 degrees view of the head posture is reduced to 82.35%.

Table 3 shows the overall recognition results of the proposed method and three other facial expression recognition methods on CASME II microexpression dataset. In the previous section, experiments were carried

Table 3: Performance comparison of several methods on CASME II microexpression dataset.

Method	Pose category	Accuracy rate (%)	STD
MDMO	9	79.49	1.4
DFER-MVAD	9	77.63	0.9
DiSTLBP-RIP	9	77.71	1.3
The proposed method	9	86.63	0.7

out under nine different perspectives. The average accuracy of the proposed method is 86.63%, while the average accuracy of the other three methods is not 80%. At the same time, 0.7% STD shows the robustness of the method.

According to the analysis of the experimental results, the ROI is divided into sparse areas as far as possible, which eliminates the interference of nonkey areas. Therefore, the proposed method can also obtain better recognition results under the change of large perspective, and the recognition rate is higher than that of MDMO, DFER-MVAD, and DiSTLBP-RIP methods, which are more than 3.00%. In other methods, RPCA is used to extract the micromotion of microexpression to replace the differential image, and the key points cannot be detected correctly in some samples. Experiments on CASME II microexpression dataset show that compared with other classical methods, the proposed method extracts the average optical flow direction histogram features from the initial frame to the peak frame, which can effectively describe the changes of microexpressions and improve the recognition accuracy.

7. Conclusions

In order to recognize instantaneous microexpression changes in natural environment, a method based on optical flow direction histogram and depth multiview network to enhance forest microexpression recognition is proposed, which is used to recognize microexpression in multipose view. Considering the low intensity of microexpressions, even in peak frames, it is difficult to recognize them with traditional static features, so the dynamic optical flow features of microexpressions are extracted in this paper. From the summary of the experimental results, it can be seen that the learning ability of the model can be effectively improved by establishing the conditional probability model of forest multiview enhancement based on depth multiview network and introducing conditional probability and neural link function into node splitting learning of random tree. In addition, using multiview-weighted voting decision to classify facial microexpressions can effectively improve the

discriminant ability of the model on the limited training set. Because the average optical flow direction histogram features depend on the first frame as the starting frame of expression occurrence and also depend on the accuracy of peak frame detection, these limitations will be further solved and the robustness to light and noise will be improved in the future work.

Data Availability

The data included in this paper are available without any restriction.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Scientific Research Projects of Colleges and Universities in Gansu Province (no. 2018C-10).

References

[1] B. Yang, J. Cheng, Y. Yang et al., "MERTA: micro-expression recognition with ternary attentions," *Multimed Tools and Applications*, 2019.

[2] Q. Li, S. Zhan, L. Xu, and C. Wu, "Facial micro-expression recognition based on the fusion of deep learning and enhanced optical flow," *Multimedia Tools and Applications*, vol. 78, no. 20, pp. 29307–29322, 2019.

[3] H. Sadeghi and A.-A. Raie, "Human vision inspired feature extraction for facial expression recognition," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 30335–30353, 2019.

[4] M. Mandal, M. Verma, S. Mathur, S. K. Vipparthi, S. Murala, and D. Kranthi Kumar, "Regional adaptive affinitive patterns (RADAP) with logical operators for facial expression recognition," *IET Image Processing*, vol. 13, no. 5, pp. 850–861, 2019.

[5] H. Li and G. Wen, "Sample awareness-based personalized facial expression recognition," *Applied Intelligence*, vol. 49, no. 9, pp. 2956–2969, 2019.

[6] H. F. Li, Q. Li, and L. Zhou, "Dynamic facial expression recognition based on multi-visual and audio descriptors," *Acta Electronica Sinica*, vol. 47, no. 8, pp. 1643–1653, 2019, (in Chinese).

[7] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[8] T. Pfister, X. Li, G. Zhao et al., "Recognising spontaneous facial micro-expressions," in *Proceedings of the IEEE International Conference on Computer Vision, ICCV 2011*, pp. 1449–1456, IEEE, Barcelona, Spain, November 2011.

[9] H. Xiaohua, S. J. Wang, X. Liu et al., "Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 32–47, 2017.

[10] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor," in *Proceedings of the International Conference on Crime Detection & Prevention*, pp. 1–6, Dubai, UAE, December 2010.

[11] X. Li, J. Yu, and S. Zhan, "Spontaneous facial micro-expression detection based on deep learning," in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pp. 1130–1134, IEEE, Chengdu, China, March 2017.

[12] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, and X. Fu, "A main directional mean optical flow feature for spontaneous micro-expression recognition," *IEEE Transactions on Affective Computing*, vol. 7, no. 4, pp. 299–310, 2016.

[13] D. Patel, X. Hong, and G. Zhao, "Selective deep features for micro-expression recognition," in *Proceedings of the 2016 23rd international conference on pattern recognition (ICPR)*, pp. 2258–2263, IEEE, Cancun, Mexico, December 2016.

[14] J. Li, D. Zhang, J. Zhang et al., "Facial expression recognition with faster R-CNN," *Procedia Computer Science*, vol. 107, pp. 135–140, 2017.

[15] T. Zhang, Z. W. Zheng, and J. Y. K. YanZong, "A deep neural network-driven feature learning method for multi-view facial expression recognition," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2528–2536, 2016.

[16] A. Yan, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pp. 1–10, Lake Placid, NY, USA, March 2016.

[17] Y. Zhou and B. E. Shi, "Action unit selective feature maps in deep networks for facial expression recognition," in *Proceedings of the International Joint Conference on Neural Networks*, pp. 2031–2038, IEEE, Anchorage, AK, USA, May 2017.

[18] W. Zheng, "Multi-view facial expression recognition based on group sparse reduced-rank regression," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 71–85, 2014.

[19] K. Xia, H. Yin, P. Qian, Y. Jiang, and S. Wang, "Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images," *IEEE Access*, vol. 7, pp. 96349–96358, 2019.

[20] S. Kansal, S. Purwar, and R. K. Tripathi, "Image contrast enhancement using unsharp masking and histogram equalization," *Multimedia Tools and Applications*, vol. 77, no. 20, pp. 26919–26938, 2018.

[21] T. Baltrusaitis, P. Robinson, and L. P. Morency, "OpenFace: an open source facial behavior analysis toolkit," in *Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10, IEEE, Lake Placid, NY, USA, March 2016.

[22] T. Baltrusaitis, P. Robinson, and L. P. Morency, "Constrained Local Neural Fields for robust facial landmark detection in the wild//," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 354–361, Sydney, Australia, December 2013.

[23] S. Wang, G. Peng, S. Chen, and Q. Ji, "Weakly supervised facial action unit recognition with domain knowledge," *IEEE Transactions on Cybernetics*, vol. 48, no. 11, pp. 3265–3276, 2018.

[24] E. Antonakos, J. Alabort-I-Medina, G. Tzimiropoulos, and S. P. Zafeiriou, "Feature-based lucas-kanade and active appearance models," *IEEE Transactions on Image Processing*, vol. 24, no. 9, pp. 2617–2632, 2015.

[25] A. Taimori and A. Behrad, "A new deformable mesh model for face tracking using edge based features and novel sets of energy functions," *Multimedia Tools and Applications*, vol. 74, no. 23, pp. 10735–10759, 2015.

[26] Z. Yue, F. Yanyan, Z. Shangyou, and P. Bing, "Facial expression recognition based on convolutional neural network," in *Proceedings of the 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, pp. 410–413, Beijing, China, 2019.