Received December 24, 2020, accepted January 13, 2021, date of publication January 18, 2021, date of current version January 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3052500

Nonfrontal and Asymmetrical Facial Expression Recognition Through Half-Face Frontalization and Pyramid Fourier Frequency Conversion

TIANYANG CAO[1], CHANG LIU[1], JIAMIN CHEN[1], AND LI GAO[2]

[1]State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[2]Faculty of Education, Beijing Normal University, Beijing 100875, China

Corresponding authors: Tianyang Cao (cao_tian_yang@sina.cn) and Chang Liu (tuengineer@qq.com)

This work was supported in part by the Key Research Project of Frontier Science of the Chinese Academy of Sciences under Grant QYZDY-SSW-JSC037, in part by the National Natural Science Foundation of China under Grant 61901440, in part by the Beijing Municipal Natural Science Foundation under Grant 4202080, in part by the One Hundred Person Project of the Chinese Academy of Sciences, and in part by the Beijing Educational Science Planning Project under Grant AAA14003.

ABSTRACT Expression recognition in the wild is easily distorted by nonfrontal and asymmetry faces. In nonfrontal faces, some areas are compressed and distorted. Even after frontalization, these compressed areas may still be blurred and distort expression recognition. Additionally, asymmetrical expressions are common on half or local face areas and produce incorrect expression features. Therefore, this paper presents a half-face frontalization and pyramid Fourier frequency conversion method. Despite the location, range and intensity of incorrect expressions in nonfrontal faces being unknown, according to discrete Fourier transform, it can be proven that the frequency band of the correct expression is much larger than that of incorrect expression on the same face. This can be taken advantage of by pyramid frequency conversion, which is designed based on Fourier frequency conversion. It can adjust incorrect expression frequency in multiscales to take them out off the band-pass of the convolution operations of deep learning and be eliminated completely, whereas correct expression information is reserved. Thus, expressions can be recognized effectively.

INDEX TERMS Expression recognition, nonfrontal face, asymmetrical expression, discrete Fourier transform, frequency band, Fourier frequency conversion.

I. INTRODUCTION

Facial expression recognition (FER) is important for artificial intelligence (AI) [1]. When AI assists in our work, FER plays an important role for human state feedback, e.g., teaching effect analysis in network training. When AI runs independently, FER is more necessary to help intelligent devices to act as human beings, e.g., household robots.

The ideal method should recognize expression in the wild, which means the recognition method should adapt to multiview human faces [2], specifically, nonfrontal faces [3]; however, this has not been achieved [4]. Nonfrontal faces can seriously distort face images [5], [6]; the eyes, cheeks and mouth, which are very important for expression recognition, are all distorted. Not only their shapes but also their position relationship are changed and distorted [7], [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Shiqing Zhang.

There are usually three methods for solving expression recognition in the wild, including extracting facial key points, extracting whole or local facial appearance features, and nonfrontal face frontalization. All of these methods are devoted to extracting pose-invariant features for expression recognition; however, these methods have not accomplished nonfrontal face expression recognition effectively.

The facial keypoints are usually expression sensitive points [9], such as the corners of the eyes and mouth. Expression recognition can be realized according to the changes in their geometric position relationship. To extract feature points exactly, many self-study methods have been introduced to build models for automatically localizing feature points. For example, the hierarchical probabilistic model was used in reference [10] to extract keypoints on a distorted face. In addition, some traditional methods have also been improved to extract feature points from nonfrontal and distorted face images. Reference [11] upgraded the traditional active shape model (ASM) and active appearance model (AAM) to increase the accuracy in extracting feature points from faces in the wild. However, there are still some difficult problems to be solved, especially the distorted position relationship among feature points in photos. Despite exactly localizing these feature points, ordinary 2D face photos, which are only the projection of a real 3D face, cannot effectively show the geometric relationship among feature points. When feature point positions change with the head pose, the projection of their geometric relationship also changes and distorts and expression cannot be recognized exactly according to the distorted information.

The local or whole facial appearance expression features contain much more regional information than that of feature points; thus, this method can improve expression recognition accuracy. Powerful deep learning methods have been introduced to extract nonfrontal facial features [12]. Reference [13] introduced weighted mixture deep networks (VGG16 model) to reform traditional local binary patterns for expression. To improve model accuracy, Reference [14] proposed weighted center regression adaptive feature mapping (W-CR-AFM), which can automatically eliminate misclassified samples or add new samples to fine-tune and reformulate the expression recognition model. Reference [15] designed a multichannel pose-aware convolution neural network (MPCNN), which can make good use of facial features. It comprises three parts: multichannel feature extraction, joint multiscale feature fusion, and pose-aware recognition. However, expression recognition is also easily confused by the space information lost when a 3D face is projected on a 2D photo. For the same expression feature on the same face, its projection on a photo is changed with different head poses; even the deep learning expression recognition might also be disturbed.

By comparison, face frontalization is the most effective method for solving information loss caused by 3D face projection. The nonfrontal face can be directly transferred to the front face, and the facial image distortion can be eliminated. For the face frontalization, a common frontal face expression recognition method can be used directly. The typical frontalization method is generative adversarial nets (GAN) [16]. Reference [17] introduced GAN to frontalize multiview faces, and these frontalization faces can be directly used for expression recognition. 3D face rebuilding is another effective method for accomplishing frontalization, the typical method including the 3D morphable model (3DMM) [18], 3D dense face alignment (3DDFA) [19], and position map regression network (PRNet) [20]. Comparing with feature points and facial appearance information, expression can be recognized exactly through frontalization. Nevertheless, the frontalized face may be more blurred than a real front face, especially the compressed half-face areas which are turned away from the camera and contain much fewer pixels in the photo. The blurred face areas are unclear and easily disturbed expression recognition.

Additionally, even for real front face expression recognition, there is still some interference caused by asymmetry, which is natural for everyone's face, which easily causes an incorrect expression. According to the research of surgeons, local asymmetrical expressions produced by muscles and fatty deposits can cause incorrect expression features. Half-face asymmetrical expression, which is produced by the difference between brain hemispheres, can weaken the expression on half faces. Additionally, if the nonfrontal asymmetrical areas are also blurred after frontalization, incorrect expression information will be much intense and serious. In addition, incorrect expression information caused by these two problems may appear anywhere on the face, as nonfrontal distortion is changed with different head poses and asymmetry varies from person to person.

For solving these problems, this paper presents the half-face pyramid Fourier frequency conversion method. First, based on discrete Fourier transform, it can be proven that the frequency band of the correct expression is much larger than that of the incorrect expression on the same face. In addition, our method makes full use of this frequency difference to manipulate deep learning to filter out incorrect expressions. The key of deep learning for expression recognition is to extract face image features through convolution layers, which can be considered a group of two-dimensional filters. Thus, the best method for eliminating incorrect information is to adjust its frequency band out of the band-pass filter. Our pyramid method can accomplish this through Fourier frequency conversion, which proves the inverse relationship between image area size and image frequency. Local incorrect expression information produced by local face blur and asymmetry can be eliminated completely through this pyramid method by enlarging or shrinking face areas in multiscale to shift their frequency out of the band-pass filter in the convolution layers. Whereas correct expression information is reserved during this process because their frequency band is much wider than that of incorrect expressions. In addition, half-face incorrect expressions can be eliminated by this pyramid method by processing the left and right half-faces and finding out which half-face contains much more correct expression. After that, expression recognition accuracy can be improved effectively.

This pyramid Fourier frequency conversion can effectively improve the emotion recognition rate. In addition, before processing half faces, this method takes advantage of the PRNet for face frontalization [20]. The deep learning used in this method is a pretrained model designed by reference [21]. In the test, two expression in the wild databases, SFEW and FER 2013 are used; both of them consist of many nonfrontal facial expression images collected from films or the Internet. For our method, the recognition rates of pleasantness and unpleasantness are 83.5% for 88.1% for SFEW. The recognition rates of pleasantness and unpleasantness are 85.2% and 85.9% for FER 2013.

II. METHOD DESIGN

In this section, the incorrect expression feature is analyzed. First, its frequency band is much different from that of the correct expression on the same face. In addition, then, according to this difference, the half-face pyramid Fourier frequency conversion method is designed for expression recognition.

A. FREQUENCY FEATURE ANALYSIS FOR INCORRECT EXPRESSION

The most useful expression recognition method is deep learning. Its core is convolution layers, and image convolution operations in these layers can be taken as two-dimensional filters for extracting image features whose frequency is consistent with their band-pass. In addition, the image frequency of expression corresponds to the contraction or relaxation extent of facial skin. For every kind of expression, deep learning can build a group of filters in its convolution layers. Therefore, the essence of face image expression recognition for a face image is to determine the expression with the most features in the image.

However, the convolution layers are easily disturbed by the incorrect expression information caused by nonfrontal and asymmetry. The features of this incorrect expression are analyzed in this section.

1) INCORRECT EXPRESSION CAUSED BY ASYMMETRY

Asymmetry can cause half-face incorrect expression and local face incorrect expression.

a: HALF-FACE INCORRECT EXPRESSION

Half-face incorrect expression caused by asymmetry is shown in figure 1. Scientists found that different expressions, such as pleasantness/unpleasantness, and approach/withdrawal, are controlled by different brain hemispheres. There are many studies on this issue. One theory proposes that the right hemisphere is more dominant for negative/unpleasantness human expression, whereas the left hemisphere is more active for positive/pleasantness expressions [22], [23]. However, this is not certain. Another theory proposes that the right hemisphere is dominant for the processing of all expressions [24], [25].

(a) (b) (c)

FIGURE 1. Asymmetrical face. (a) Original face, (b) "Neutral" right half face, (c) "Angry" left half face.

Despite the relationship between hemisphere and expression might vary from person to person and is still being researched, this asymmetrical expression can seriously disturb expression recognition. If the asymmetrical expression is inconsistent with that of the training sample, it is easy to produce an incorrect recognition result. For example, in the simplest case, the left half of the test face image is more sensitive to expression, while the sensitivity on the training sample face is the opposite, which will disturb the recognition result.

Additionally, the asymmetry may also disturb the recognition model established by the training process. For some people, the left side of their faces might be sensitive to pleasantness and the right half to unpleasantness. While this is uncertain, others might be on the opposite. Using these expression samples to train deep learning, the recognition model cannot be trained accurately.

b: LOCAL INCORRECT EXPRESSIONS ON FACE AREAS

Reference [26], [27] proved the movement of facial muscles and fatty deposits, which are basic to producing expressions, are often asymmetrical. It is easy to make facial skin around them to distort and become much different from the other areas [28]. This might produce local incorrect expression information and decrease correct expression information.

Expression recognition is a classification issue. For each kind of expression, deep learning pretrains a group of filters. The recognition result is the expression whose filters can extract most of the features from this facial image.

Ideally, the frequency band of facial correct expression should be consistent with its band-pass filter, and most of its information can be extracted. This filtering process is shown in figure 2 (a), (c) and (d). The spectrum of the original correct expression is shown in figure 2 (a). The spectrum of its band-pass filter is shown in figure 2 (c). The filter result is shown in figure 2 (d), and most frequency information can be extracted.

However, in reality, the frequency band of distorted facial areas and the band-pass filter is often misplaced. The distortion caused by local asymmetrical areas including enlarged and shrunken areas according to Fourier transform can both cause image frequency conversion. So the correct expression frequency band is easily disturbed and shifted. After filtering, only a little correct information might be reserved. If the reserved information is less than that of the incorrect expression, error recognition is easily caused. This affect caused by misplacement can be shown in figure 2 (b) (e), compared with the spectrum of ideal original correct expression (figure 2 (a)), the frequency of the correct expression is shifted (figure 2 (b)). Its filter result (figure 2 (e)) is much less than the original filter result (figure 2 (d)), and this easily causes an incorrect expression recognition result.

However, since incorrect frequency shift can weaken the correct expression information, the correct frequency shift can also eliminate incorrect expression. Although the asymmetrical samples may interfere with the trained deep learning expression recognition model, this problem can be solved by eliminating as much incorrect expression information as possible. To achieve this, the pyramid Fourier frequency conversion method is presented in section 2.1. It can shift the

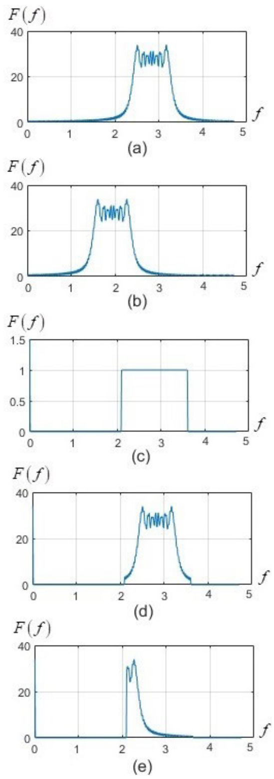


FIGURE 2. Expression information extraction through filters in convolution layers. (a) The spectrum of ideal correct expression, (b) The spectrum of real correct expression, (c) The spectrum of filter pass-band for correct expression, (d) The spectrum of filtering result of ideal correct expression, (e) The spectrum of filtering result of real correct expression. ($F(f)$ is the information amplitude in frequency domain, f is the frequency axis in frequency domain). (In order to simplify analysis, this figure only shows filtering process on one-dimensional, the two-dimensional image spectrum also follows the same principle.)

frequency of expression information in multiscale to eliminate local incorrect expression information and make filters reserve correct expression information. In addition, half-face incorrect expression can also be solved by comparing the processed results on two half faces through this pyramid method and finding which half face contains much more correct expression information.

2) INCORRECT EXPRESSION CAUSED BY A NONFRONTAL FACE

If the asymmetry is combined with image distortion caused by a nonfrontal face, incorrect expression will be much more serious.

The contraction or relaxation extent of skin and facial features on face images, which is equivalent to expression image frequency, is easily distorted when the head pose changes. According to the theory of Fourier frequency conversion, when a head turns, the half nonfrontal face image toward the camera is widened, it can decrease the frequency of facial expression information. In addition, the half nonfrontal face image away from camera is compressed, which can enlarge the frequency of facial expression information. This can be analyzed as follows:

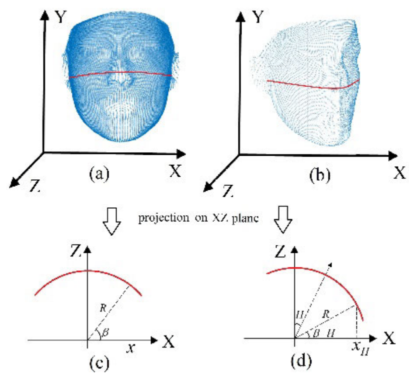


FIGURE 3. Face curve deform caused by head turning. (a) Face curve (red) on 3D frontal face. (b) Face curve (red) on 3D nonfrontal face. (c) Frontal face curve (red) projection on XZ plane (horizontal). (d) Nonfrontal face curve (red) projection on XZ plane (horizontal).

Taking the facial image distortion caused by head turning, which is a common head movement, as an example:

The changing pixel coordinates are the embodiment of facial image distortion. This change can be decomposed as X-direction (image horizon) and Y-direction (image vertical). The interference caused by head turning angle H is mainly concentrated in the X-direction. This can lead to a distorted expression image.

As shown in figure 3, the red curves in figure 3 (c) (d) is the projection of 3D frontal and nonfrontal faces curves (red curves in figure 3 (a) (b)) on XZ plane (horizontal). Comparing with the projection curves in figure 3 (c), the projection curve in figure 3 (d) has turned angle H.

For the point original coordinate x, it is calculated through function (1).

$$x = R \cdot \cos\beta \tag{1}$$

where β is the face point radius angle. When head turning angle H, the point coordinate is changed to x_H:

$$x_H = R \cdot \cos(\beta - H) \tag{2}$$

The interference can be shown more clearly after decomposition by a trigonometric formula:

$$x_H = R\cos\beta\cos H + R\sin\beta\sin H = x\cos H - \Delta x \tag{3}$$

where, $\Delta x = -R\sin H\cos H$. This distortion caused by H can be analyzed according to the spectrum analysis of Fourier transform [29]. The frontal face spectrum is as follows:

$$F(\omega) = \int_{-\infty}^{\infty} f(x) \cdot e^{-j\omega x} dx \tag{4}$$

The nonfrontal face spectrum on the X-direction is as follows:

$$\begin{aligned} F_H(\omega) &= \int_{-\infty}^{\infty} f(x_H) \cdot e^{-j\omega x} dx \\ &= \int_{-\infty}^{\infty} f(x\cos H - \Delta x) \cdot e^{-j\omega x} dx = F\left(\frac{\omega}{\cos H}\right) e^{-j\omega\Delta x} \end{aligned} \tag{5}$$

Compared with the spectrum $F(\omega)$ of the original information, the nonfrontal face spectrum is changed as $F\left(\frac{\omega}{\cos H}\right)e^{-j\omega\Delta x}$. The face distortion causes the frequency to be compressed $\frac{1}{\cos H}$ times, and phase added $-\Delta x$. This frequency change can disturb the filtering results in convolution layers directly. The larger the head turn, the more serious the expression features are distorted. For the same reason, the facial feature distortion on the image vertical direction caused by head pitching is similar to that of head turning.

Nonfrontal face frontalization is the best method for solving this problem; thus, this paper introduces the PRNet network, which is a deep learning method and can accomplish 2D face image frontalization. However, there are still some problems. The resolution of these facial features and skin in compressed half face is very low in the original image and still blurred in the frontalization image. If the asymmetrical areas are also the blurred area, the incorrect expression will be much serious.

To solve this problem, the essential difference between incorrect and correct expression on the same face image is analyzed according to their spectrum in the next section. The pyramid Fourier frequency conversion method is designed based on these spectrum differences.

3) FREQUENCY BANDS DIFFERENCE BETWEEN INCORRECT AND CORRECT EXPRESSIONS

It is more difficult to eliminate local incorrect expressions than half-face incorrect expressions. Half-face incorrect expressions can be eliminated by recognizing left and right half-face expressions separately and identifying which half face contains much more correct expression. However, the local incorrect expression caused by local asymmetrical and blurred face areas is much more difficult to eliminate because it may mix with the correct expression on the face image.

To make matters worse, these local incorrect expressions are uncertain, as their location and range and intensity are different in different faces. The only known is that the face area showing the correct expression is much larger than that of the incorrect expression; otherwise, it cannot be called the correct expression. However, because of the unknown position and range of incorrect expression in every face image, this area difference between correct and incorrect expressions is difficult to apply on the face image space domain.

This paper takes advantage of this difference in the face image frequency domain. In addition, this difference is clearer in the spectrum of Fourier transform. As the face image detected by computer vision is a digital image, the spectrum of correct and incorrect expression on the same face can be analyzed according to the theory of discrete Fourier transform (DFT) [30]:

$$F_c(n) = \sum_{k=0}^{N_c-1} f_c(k)e^{-\frac{2\pi}{N_c}nk} \tag{6}$$

$$F_w(n) = \sum_{k=0}^{N_w-1} f_w(k)e^{-\frac{2\pi}{N_w}nk} \tag{7}$$

where, $f_c()$ and $f_w()$ are the correct and incorrect expression functions, their area sizes are N_c and N_w, $N_c \gg N_w$, n is the frequency in the discrete spectrum and equivalent to the frequency symbol f in the continuous domain, $n = f$. To simplify the analysis, this paper only calculates DFT on the X-axis, and the transform on the Y-axis is the same.

The highest frequency $f_{\max,c}, f_{\max,w}$ of the DFT of the correct and incorrect expression can be detected relative to the number of face points in their areas. Nc is the number of correct expression pixels, and N_w is the number of incorrect expression pixels. As $N_c \gg N_w$ in the same face, the correct expression frequency that can be detected is much larger than that of the incorrect expression, $f_{\max,c} \gg f_{\max,w}$. So the frequency band of the correct expression is larger than that of the incorrect expression on the same face. This can be shown in figure 4.

The frequency band of the correct expression information is much wider than that of the incorrect information. This can be taken advantage of to design an incorrect expression elimination method.

The filtering ability relies on its band-pass. Only the information whose frequency band overlaps (or partially overlaps) band-pass filter can be reserved by the deep learning convolutional layers; the others are filtered out. For every expression, the convolutional layers pretrain a group of filters. The band-pass of each filter group is approximately equal to the frequency band of its corresponding expression information. The different filter groups can extract the expression information and filter out the other expression information. This process can be shown more clearly in a spectrum, as shown in figure 5. Figure 5 (a) is the original spectrum of the correct expression (front) and incorrect expression (back). Figure 5 (d) is the band-pass filter for the correct expression. The filtering result is shown in figure 5 (e). Figure 5 (b) is the band-pass filter for the other expression, and the filtering result is shown in figure 5 (c). To simplify the analysis, figure 5 only takes one filter for each expression as an example.

Although much more of the correct expression information can be reserved than the incorrect expression after filtering, the remaining incorrect expressions might still mislead the recognition result. The best method for improving recognition accuracy is to eliminate the incorrect expression completely through the convolution layer filters.

Therefore, in this paper, this is accomplished by adjusting the face image frequency, and image is proposed as follows:

Because the frequency band of the correct expression is much larger than that of incorrect expression, if the image frequency can be shifted $\Delta\omega$, which is larger than the incorrect information frequency band and smaller than the correct information frequency band, the narrow-band incorrect information moves away from its band-pass filter and

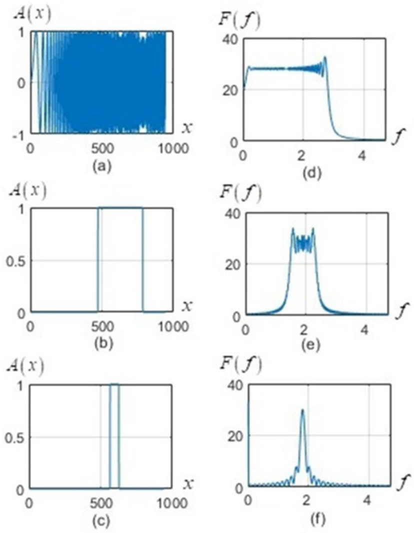


FIGURE 4. The difference between correct and incorrect expression. (a) Original expression information. (b) Face area of correct expression information. (c) Face area of incorrect expression information. (d) The spectrum of original whole face expression. (e) The large frequency band of correct expression on face area. (f) The small frequency band of incorrect expression on face area. ($A(x)$ is the information amplitude in space domain, x is the space axis in image space domain; $F(f)$ is the information amplitude in frequency domain, f is the frequency axis in frequency domain). (In order to simplify analysis, this figure only shows expression characteristic on the one-dimensional, and two-dimensional spectrum also follows the same principle.)

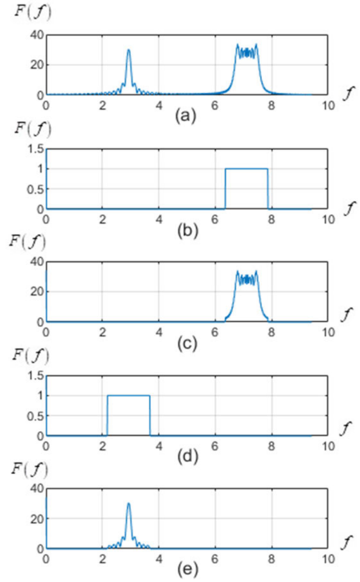


FIGURE 5. Expression information processing displayed in frequency spectrum. (a) Incorrect expression (front) and correct expression (back), (b) The convolution layer filter for processing correct expression, (c) Filtering result of correct expression, (d) The convolution layer filter for processing incorrect expression, (e) Filtering result of incorrect expression, ($F(f)$ is the information amplitude in frequency domain, f is the frequency axis in frequency domain). (In order to simplify analysis, this figure only shows one filter for each expression.)

is eliminated, whereas the broadband correct information can still be reserved after it is processed by convolution layers.

This frequency shift is shown in figure 6. Figures 6(a) (d) (g), shows the result of shifting the face image frequency to the left (shift $-\Delta\omega$). The narrow-band incorrect information moves away from its band-pass filter and is eliminated by the filters in the convolutional layers (figure 6(d)), whereas the broadband correct information can still be reserved by the convolutional layers (figure 6(g)). The expression recognition can be more exact according to the filtering result of frequency left shift information.

Figures 6(b) (e) (h) show the result of shifting the image frequency to the right (shift $\Delta\omega$). The narrow-band incorrect information also moves away from its band-pass filters and is eliminated by the filters in the convolutional layers (figure 6(e)), whereas the broadband correct information can still be partially reserved by the convolutional layers (figure 6(h)). The expression recognition can also be more exact according to the filtering result of the frequency right shift information.

Considering that incorrect expressions might distribute in many frequency points, only shifting the image frequency once is not enough to filter out all the incorrect expression information. To solve this problem, the pyramid Fourier frequency conversion method is designed in this paper. It can shift face image frequency in multirange, from $-\Delta\omega$ (left shift) to $\Delta\omega$ (right shift), and incorrect expression information can be eliminated completely.

B. EXPRESSION RECOGNITION BY HALF-FACE PYRAMID FOURIER FREQUENCY CONVERSION

This half-face pyramid Fourier frequency conversion method can eliminate both local incorrect expression and half-face incorrect expression, and improve recognition accuracy effectively.

For a face with any pose, as shown in figure 7, the structure of expression recognition is shown in figure 8. It is composed of face frontalization and pyramid Fourier frequency conversion for two half faces and an expression recognition model. This pyramid structure is the core of expression recognition, and the face expression can be recognized according to the results of pyramid Fourier frequency conversion in its two half faces. The expression recognition model is a pretrained deep learning model downloaded from the GitHub free open source library and was designed by the authors of reference [21]. The expressions which appear most frequently in recognition results of these multiscale face areas can be taken as the correct expression of this face image.

1) PYRAMID FOURIER FREQUENCY CONVERSION METHOD

The pyramid Fourier frequency conversion method is designed based on the theory of Fourier frequency conversion. As shown in figure 9, the left and right half faces are processed by two pyramid structures separately, and face image frequency multishift can be accomplished by changing the image size in multiscale.

Face areas input into this pyramid can be divided into two classes: the small areas cut from the face image and the larger

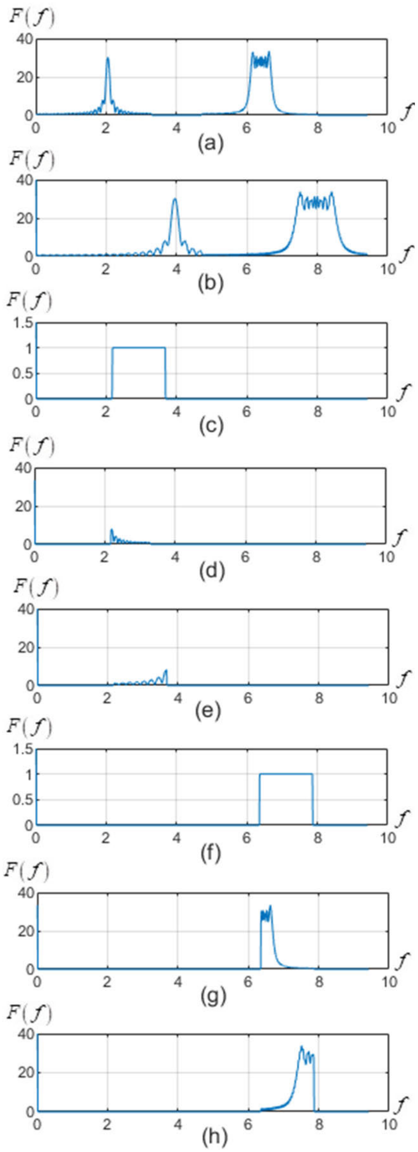


FIGURE 6. **The spectrum of face frequency shift and processing result through convolution layers. (a) The frequency left shifted incorrect expression (front) and correct expression (back), (b) The frequency right shifted incorrect expression (front) and correct expression (back), (c) The convolution layer filter for processing incorrect expression, (d) Filtering result of frequency left shifted incorrect expression, (e) Filtering result of frequency right shifted incorrect expression, (f) The convolution layer filter for processing correct expression, (g) Filtering result of frequency left shifted correct expression, (h) Filtering result of frequency right shifted correct expression.**

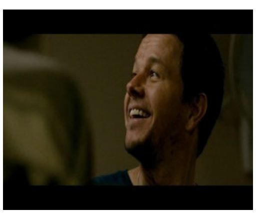


FIGURE 7. **Original "pleasantness" face.**

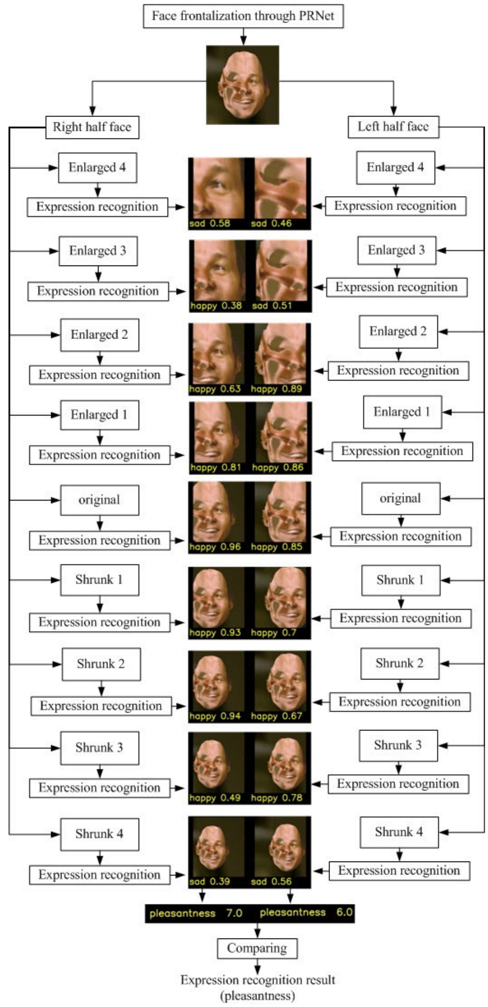


FIGURE 8. **The image process of half-face pyramid Fourier frequency conversion and expression recognition result.**

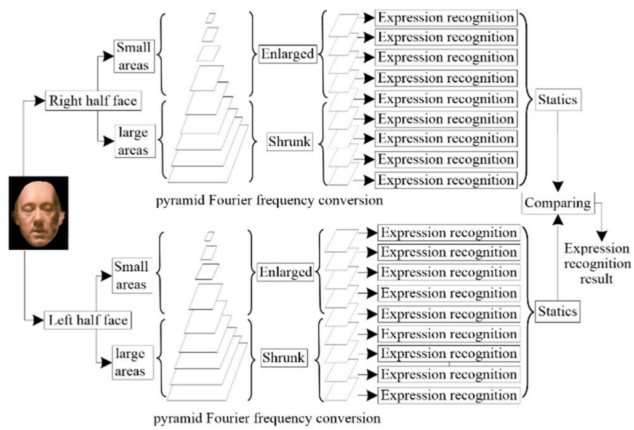


FIGURE 9. **The structure of half-face pyramid Fourier frequency conversion.**

areas composed of the face and its background. In addition, the left and right eyes, which are the most sensitive points on the face, are taken as the center of these areas. These cut face areas are adjusted to a standard face image size and input into deep learning for expression recognition. According to Fourier frequency conversion, the expression frequency in the image area is inverse proportional to the change in this area size. Small areas enlarged can decrease expression frequency, shrunken large areas can enlarge expression frequency. After these processes, the narrow-band incorrect expression information frequency is adjusted to multiscale and can be easily eliminated by the convolutional layers of the expression recognition model (deep learning). Whereas the broadband correct expression information can be reserved and taken advantage of by deep learning for recognizing expression.

a: SMALL AREAS ENLARGED FOR DECREASING EXPRESSION FREQUENCY

Small face areas are cut from left or right half faces with a group of gradually reduced square masks, which are smaller than the face size. These cut face areas are enlarged to the standard image size. This is the frequency band shrinking process.

For example, the original image Fourier transform from the space domain to the frequency domain is $z(x, y) \leftrightarrow Z(u, v)$, given that the frequency band of correct expression information is $\Delta u = \omega_{u,2} - \omega_{u,1}$, $\Delta v = \omega_{v,2} - \omega_{v,1}$ and $u \in [\omega_{u,1}, \omega_{u,2}]$, $v \in [\omega_{v,1}, \omega_{v,2}]$. If the image size is increased a times, the image frequency band of the correct expression is decreased as $u_w \in \left[\frac{\omega_{u,1}}{a}, \frac{\omega_{u,2}}{a}\right]$, $v_w \in \left[\frac{\omega_{v,1}}{a}, \frac{\omega_{v,2}}{a}\right]$. This can be proven through the theory of Fourier transform:

$$z(ax, ay) \leftrightarrow \frac{1}{a^2} Z\left(\frac{u}{a}, \frac{v}{a}\right) = \frac{1}{a^2} Z(u_w, v_w) \tag{8}$$

Ideally, for the filter in the convolution layer, which is used to extract correct expression information, its band-pass should be nearly unified with that of correct expression, $u \in [\omega_{u,1}, \omega_{u,2}]$, $v \in [\omega_{v,1}, \omega_{v,2}]$. To extract correct expression information and effectively recognize expression, as shown in functions (9) (10), the decreased expression band and its band-pass filter should be partially overlapped:

$$\left[\frac{\omega_{u,1}}{a}, \frac{\omega_{u,2}}{a}\right] \cap [\omega_{u,1}, \omega_{u,2}] = \left[\omega_{u,1}, \frac{\omega_{u,2}}{a}\right] \neq \emptyset \tag{9}$$

$$\left[\frac{\omega_{v,1}}{a}, \frac{\omega_{v,2}}{a}\right] \cap [\omega_{v,1}, \omega_{v,2}] = \left[\omega_{v,1}, \frac{\omega_{v,2}}{a}\right] \neq \emptyset \tag{10}$$

The maximum increased time a that can make functions (9) and (10) tenable is the minimum between $\frac{\omega_{u,2}}{\omega_{u,1}}$ and $\frac{\omega_{v,2}}{\omega_{v,1}}$.

b: LARGE AREAS DECREASED FOR ENLARGING EXPRESSION FREQUENCY

These areas are cut from the left and right half faces and their background with a group of gradually increased square masks, which are larger than the face size and smaller than the image size. In addition, the cut face areas are decreased to the standard image size. This is the frequency band enlarging process.

If the area size is decreased to $\frac{1}{b}$, the frequency band of the correct expression is increased to $u_s \in [b\omega_{u,1}, b\omega_{u,2}]$, $v_s \in [b\omega_{v,1}, b\omega_{v,2}]$:

$$z\left(\frac{x}{b}, \frac{y}{b}\right) \leftrightarrow b^2 Z(bu, bv) = \frac{1}{b^2} Z(u_s, v_s) \tag{11}$$

For the filter, if the increased expression band and its band-pass filter are still partially overlapped, as shown in functions (12) (13), the correct expression feature can be extracted and the expression can be recognized effectively.

$$[b\omega_{u,1}, b\omega_{u,2}] \cap [\omega_{u,1}, \omega_{u,2}] = [b\omega_{u,1}, \omega_{u,2}] \approx \emptyset \tag{12}$$

$$[b\omega_{v,1}, b\omega_{v,2}] \cap [\omega_{v,1}, \omega_{v,2}] = [b\omega_{v,1}, \omega_{v,2}] \approx \emptyset \tag{13}$$

The minimum shrink time $\frac{1}{b}$ that can make functions (12) (13) tenable is the maximum between $\frac{\omega_{u,1}}{\omega_{u,2}}$ and $\frac{\omega_{v,1}}{\omega_{v,2}}$.

However, for these incorrect features, despite their distribution in many frequency points, their frequency bands $[\omega_{u,1,w}, \omega_{u,2,w}]$, $[\omega_{v,1,w}, \omega_{v,2,w}]$ are much smaller than that of correct expressions and easy to move off their band-pass filter $[\omega_{u,1f}, \omega_{u,2f}]$, $[\omega_{v,1f}, \omega_{v,2f}]$. Taking the Fourier transform on the X-axis as an example:

$$\left[\frac{\omega_{u,1,w}}{a}, \frac{\omega_{u,2,w}}{a}\right] \cap [\omega_{u,1,f}, \omega_{u,2,f}] = \left[\omega_{u,1,f}, \frac{\omega_{u,2,w}}{a}\right] \approx \emptyset \tag{14}$$

$$[b\omega_{u,1,w}, b\omega_{u,2,w}] \cap [\omega_{u,1,f}, \omega_{u,2,f}] = [b\omega_{u,1,w}, \omega_{u,2,f}] \approx \emptyset \tag{15}$$

where $[b\omega_{u,1,w}, b\omega_{u,2,w}]$ is increased incorrect expression frequency band, $\left[\frac{\omega_{u,1,w}}{a}, \frac{\omega_{u,2,w}}{a}\right]$ is its decreased frequency band.

Assume that the band-pass of all expression filters are the same and equal to Δu on the u-axis of the spectrum. For the enlarged face area, the largest increased time for correct expression that can still be detected is $\frac{\omega_{u,2}}{\omega_{u,1}} = \frac{\omega_{u,1}+\Delta u}{\omega_{u,1}}$. While for the incorrect expression, as its frequency band is very narrow $\frac{\Delta u}{n}$ ($n \gg 1$), its area can only be increased to $\frac{\omega_{u,2,w}}{\omega_{u,1,f}} = \frac{\omega_{u,1,f} + \frac{\Delta u}{2}\left(1+\frac{1}{n}\right)}{\omega_{u,1,f}}$, much smaller than that of the correct expression. So, if the face area is increased from $\frac{\omega_{u,2,w}}{\omega_{u,1,f}}$ to $\frac{\omega_{u,2}}{\omega_{u,1}}$ and input into deep learning for expression recognition, the broadband correct expression can be reserved by its filters, whereas the narrow-band incorrect expression is moved off its band-pass filter and eliminated.

For the decreased face area, the correct expression can be decreased to $\frac{\omega_{u,1}}{\omega_{u,2}} = \frac{\omega_{u,1}}{\omega_{u,1}+\Delta u}$. The incorrect expression can only be decreased to $\frac{\omega_{u,1,w}}{\omega_{u,2,f}} = \frac{\omega_{u,1,f}}{\omega_{u,1,f} + \frac{\Delta u}{2}\left(1+\frac{1}{n}\right)}$, much larger than that of the correct expression. So, if the decreased face area is decreased from $\frac{\omega_{u,1,f}}{\omega_{u,2,w}}$ to $\frac{\omega_{u,1}}{\omega_{u,2}}$ and input into deep learning for expression recognition, the broadband correct expression can be reserved by its filters, whereas the narrow-band incorrect expression is moved off its band-pass filter and eliminated.

Therefore, if the suitable enlarging and shrinking range can be chosen, the correct expression information can be reserved whereas the incorrect expression is filtered out, and the expression can be recognized effectively. Through the test, the suitably increased and decreased time is 2∼0.6, and 9 layers of multiscale for this pyramid structure can effectively accomplish expression recognition.

2) PYRAMID PROCESSING FOR HALF FRONTALIZATION FACES AND EXPRESSION RECOGNITION

Before the face image is processed by the pyramid Fourier frequency conversion, the face need to be frontalized and divided as left and right half faces.

Nonfrontal face frontalization can be accomplished through PRNet, which is a deep learning method. PRNet can transfer the 2D face photo into a 3D face [20]. The core of PRNet is composed of 10 residual blocks and 17 transposed deconvolution layers. The self-study ability of residual blocks is very strong and can extract many 3D features directly from a 2D image. According to these features, deconvolution layers build the 3D face. The frontalization face can be obtained by rotating the 3D face to the front through an Euler matrix $A(heading, pitch, roll)$ whose heading, pitch and roll angles are also calculated through PRNet.

The two frontalization half faces are processed by this pyramid structure separately. The expression result, which appears most in these recognition results of multiscales of two half faces, is the correct expression.

III. EXPERIMENTAL RESULTS

To test this expression recognition method effectively, two expression in the wild databases, SFEW and FER2013, are introduced. SFEW is composed of the face images captured from many famous films. FER2013 is composed of face images downloaded from the Internet. Both of them are very useful for testing expression recognition.

Because the features in a nonfrontal face, such as eyes, mouth and checks, are often distorted and blurred, nonfrontal faces cannot show the expression detail as clearly as that of a front face in the most cases. Thus, this paper only divided the expressions as ''pleasantness'', ''unpleasantness'', ''surprise'', and ''neutral''. Unpleasantness is composed of negative expressions, including ''angry'', ''disgust'', ''fear'', and ''sad''. The emotion ''surprise'', as it often combines with other expressions, can be divided as three kinds: ''pleasant surprise'', ''unpleasant surprise'', and ''neutral surprise''. Its experimental result is analyzed at the end of this section.

Through our method, the recognition result of figure.7 is shown in figure.10. The expression is correctly recognized.

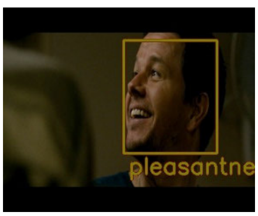


FIGURE 10. **''Pleasantness'' is correct recognized by our method.**

In the test, for the recognition rate of database FER2013, the present methods, such as VGG [31] is 71.29%, reference [4] is 71.14%, eXnet [31] is 73.54 %. For the recognition of ''unpleasantness'' and ''pleasantness'', as ''angry'', ''disgust'', ''fear'', ''sad'' are combined as ''unpleasantness'', the recognition rate can be a little higher, and the reported light-CNN [4] are 81.5% and 88%, pre-trained CNN [4] are 84.25% and 91% [4]. Our method recognition rates are 85.9% and 85.2%, and much exactly for recognizing ''unpleasantness''. While as FER2013 may still contain some expression images which are made by following the instruction of the photographers, and a little different from that of expression in-the-wild.

For the database SFEW, its expression images are much more complex than those of FER 2013; thus, our recognition experiment is mainly focused on this database.

For pleasantness, the expression recognition process and test result are shown in figures 11, 12, and 13. The woman in figure 11 is smiling and only showing the side of her face to the camera. Through the pyramid Fourier frequency conversion on the two side faces, the incorrect expression information is eliminated effectively. The number of correct expression recognition results is much larger than that of incorrect recognition results on the same half frontalization face and larger than the number of incorrect recognition results on the other half frontalization face. The smile expression can be recognized exactly as ''pleasantness'', as shown in figure 13.

FIGURE 11. **Original ''pleasantness'' face.**

For unpleasantness, the expression recognition process and test results are shown in figures 14, 15, and 16:

The person in figure 14 is sad. Her face is distorted with her head turned, and some face areas are also shaded. Through the pyramid Fourier frequency conversion on the two half frontalization faces, the incorrect expression information is eliminated effectively. The number of correct expression recognition results is much larger than that of incorrect recognition on the same half face and larger than the number of incorrect recognition results on the other half face.

The sad expression can be recognized exactly as ''unpleasantness'', as shown in figure 16.

For the recognition rate of database SFEW, the latest reported methods, such as the deep learning method Multiple deep CNNs is 55.96% [32], RAN(VGG16+ResNet18) is 56.4% [33], and gACNN is 54.47% [33], and ensemble IL-CNN [34] is 59.41%. For recognizing unpleasantness and pleasantness, the recognition rates can be a little higher, and the reported recognition rates of ensemble IL-CNN [34] increasing to 64.82% and 73.7%. Whereas recognition rates for our method could be 88.1% and 83.5%.

For the emotion ''surprise'', as it often combines with ''happy'' or ''angry'', the recognition result is divided into

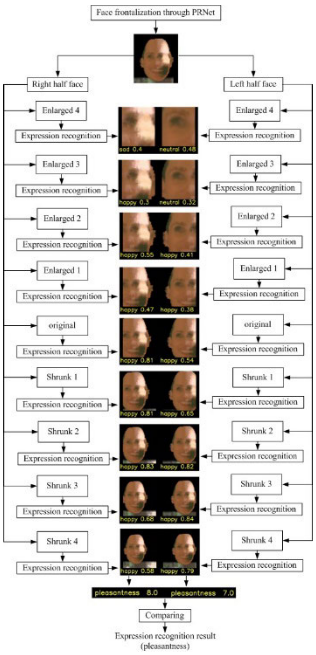


FIGURE 12. The image process of half faces pyramid Fourier frequency conversion for the "Pleasantness" expression.

FIGURE 13. "Pleasantness" is correctly recognized by our method.

FIGURE 14. Original "unpleasantness" face.

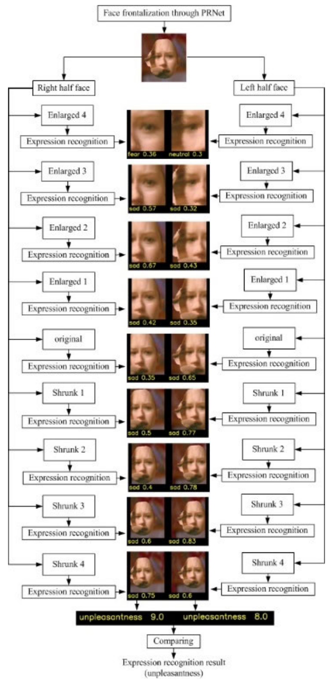


FIGURE 15. The image process of half faces pyramid Fourier frequency conversion for the "Unpleasantness" expression.

FIGURE 16. "Pleasantness" is correctly recognized by our method.

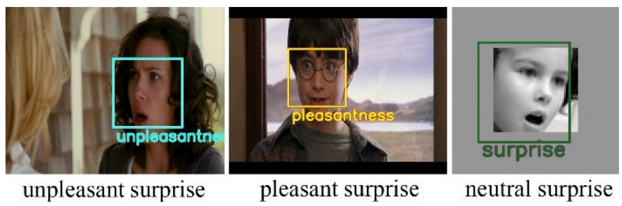


FIGURE 17. Recognition results of three kinds of "surprise".

three kinds: "pleasant surprise", "unpleasant surprise", and "neutral surprise" in our experiment. The recognition results are shown in figure 17. For the database SFEW, "unpleasant surprise" is recognized as "unpleasantness" (recognition result of the left face in figure 17) and the recognition rate is 79.3%, "pleasant surprise" is recognized as "pleasantness" (recognition result of the middle face in figure 17) and the recognition rate is 10.3%, and "neutral surprise" is recognized as "surprise" and the recognition rate is 6.1%. In addition, only 4.3% of images are incorrectly recognized as "neutral". For the database FER 2013, the "unpleasant surprise" recognition rate is 64.6%, the "pleasantness surprise" recognition rate is 14.9%, and the "neutral surprise" recognition rate is 15.3% (recognition result of the right face in figure 17). In addition, only 5.2% of the images are incorrectly recognized as "neutral". "Surprise" can be recognized effectively through our method.

IV. CONCLUSION

Facial expression recognition in the wild is an important issue for artificial intelligence (AI). While recognition is disturbed seriously by nonfrontal and asymmetrical faces, every face has asymmetrical expressions. The intensities between two half-face expressions are usually different. The asymmetrical movement of muscles and fatty deposits easily produces local incorrect expression. Although frontalization can solve the nonfrontal face problem, the distorted face may be blurred in some areas after frontalization, which aggravates the incorrect expressions caused by asymmetry. Thus, after proving the frequency band differences between correct and incorrect expressions on the same face through the theory of discrete Fourier transform, this paper presents the half-face pyramid frequency conversion method. This method can take advantage of the filters in deep learning convolutional layers for expression recognition, and decreasing and increasing faces in multiscale to eliminate incorrect expression information. According to the theory of Fourier frequency conversion theory, the incorrect expressions are filtered out by this pyramid structure because they are narrow-band, whereas the correct expressions can be reserved because they are broadband. To eliminate the asymmetrical half-faces, this paper processes the two half faces separately, and the final recognition result can be obtained by comparing the two half-face expression results in different frequency conversion ranges. In the test, the expression in the wild databases SFEW and FER 2013 were used. The average recognition rates of pleasantness and unpleasantness is larger than 80%.

ACKNOWLEDGMENT

The authors of this work are especially grateful to the authors of reference [21] for their expression recognition open source software on GitHub, the PRNet authors for their open source software on GitHub, and the Face++ expression recognition platform provided by Megvii Technology Ltd. No conflict of interest was declared. All the face images in this manuscript are taken from the open dataset SFEW or down free from the Internet, and they can be used directly.

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, ''Emotion recognition in human-computer interaction,'' *IEEE Signal Process. Mag.*, vol. 18, no. 1, pp. 32–80, Jan. 2001.

[2] S. Moore and R. Bowden, ''Local binary patterns for multi-view facial expression recognition,'' *Comput. Vis. Image Understand.*, vol. 115, no. 4, pp. 541–558, Apr. 2011.

[3] K. Yan, W. Zheng, T. Zhang, Y. Zong, and Z. Cui, ''Cross-database non-frontal facial expression recognition based on transductive deep transfer learning,'' *IEEE Access*, vol. 7, pp. 1–8, 2019.

[4] J. Shao and Y. Qian, ''Three convolutional neural network models for facial expression recognition in the wild,'' *Neurocomputing*, vol. 355, pp. 82–92, Aug. 2019.

[5] P. Werner, F. Saxen, A. Al-Hamadi, and H. Yu, ''Generalizing to unseen head poses in facial expression recognition and action unit intensity estimation,'' in *Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Lille, France, May 2019, pp. 1–8.

[6] A. Samanta and T. Guha, ''On the role of head motion in affective expression,'' in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 2886–2890.

[7] S. Li, W. Deng, and J. Du, ''Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2584–2593.

[8] J. Hu, B. Yu, Y. Yang, and B. Feng, ''Towards facial de-expression and expression recognition in the wild,'' in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Cambridge, U.K., Sep. 2019, pp. 157–163.

[9] S. P. Khandait, R. C. Thool, and P. D. Khandait, ''Automatic facial feature extraction and expression recognition based on neural network,'' *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 1, pp. 113–118, 2011.

[10] Y. Wu, Z. Wang, and Q. Ji, ''A hierarchical probabilistic model for facial feature detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 1–4.

[11] M. Iqtait, F. S. Mohamad, and M. Mamat, ''Feature extraction for face recognition via active shape model (ASM) and active appearance model (AAM),'' in *Proc. IOP Conf., Mater. Sci. Eng.*, Mar. 2018, vol. 332, pp. 1–9.

[12] C. J. Lin, C. H. Lin, and S. H. Wang, ''Multiple convolutional neural networks fusion using improved fuzzy integral for facial emotion recognition,'' *Appl. Sci.*, vol. 9, no. 13, pp. 1–14, 2019.

[13] B. Yang, J. Cao, R. Ni, and Y. Zhang, ''Facial expression recognition using weighted mixture deep neural network based on double-channel facial images,'' *IEEE Access*, vol. 6, pp. 4630–4640, 2018.

[14] B.-F. Wu and C.-H. Lin, ''Adaptive feature mapping for customizing deep learning based facial expression recognition model,'' *IEEE Access*, vol. 6, pp. 12451–12461, 2018.

[15] Y. Liu, J. Zeng, S. Shan, and Z. Zheng, ''Multi-channel pose-aware convolution neural networks for multi-view facial expression recognition,'' in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Xi'an, China, May 2018, pp. 458–465.

[16] R. Huang, S. Zhang, T. Li, and R. He, ''Beyond face rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Moscow, Russia, Oct. 2017, pp. 2458–2467.

[17] Y.-H. Lai and S.-H. Lai, ''Emotion-preserving representation learning via generative adversarial network for multi-view facial expression recognition,'' in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Xi'an, China, May 2018, pp. 263–270.

[18] V. Blanz, T. Vetter, and A. Rockwood, ''A morphable model for the synthesis of 3D faces,'' in *Proc. ACM SIGGRAPH*, 2002, pp. 187–194.

[19] X. Zhu, X. Liu, Z. Lei, and S. Z. Li, ''Face alignment in full pose range: A 3D total solution,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 78–92, Jan. 2019.

[20] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, ''Joint 3D face reconstruction and dense alignment with position map regression network,'' in *Proc. 15th Eur. Conf. Comput. Vis. (ECCV)*, vol. 11218, 2018, pp. 557–574.

[21] O. Arriaga, M. Valdenegro-Toro, and P. Plöger, ''Real-time convolutional neural networks for emotion and gender classification,'' 2017, *arXiv:1710.07557*. [Online]. Available: http://arxiv.org/abs/1710.07557

[22] J. C. Borod, C. S. Haywood, and E. Koff, ''Neuropsychological aspects of facial asymmetry during emotional expression: A review of the normal adult literature,'' *Neuropsychol. Rev.*, vol. 7, no. 1, pp. 41–60, Mar. 1997.

[23] R. Campbell, ''The lateralization of expression: A critical review,'' *Int. J. Psychol.*, vol. 17, nos. 2–3, pp. 211–229, 1982.

[24] R. J. Davidson, ''Emotion and affective style: Hemispheric substrates,'' *Psychol. Sci.*, vol. 3, no. 1, pp. 39–43, Jan. 1992.

[25] R. J. Davidson, ''Cerebral asymmetry, emotion and affective style,'' *J. Clin. Exp. Neuropsychol.*, vol. 12, no. 1, pp. 361–387, 1995.

[26] H. Sackeim and R. Gur, ''Asymmetry in facial expression,'' *Science*, vol. 209, no. 4458, p. 834, 1980.

[27] H. Sackeim and A. Hemispheric, ''Asymmetry in the expression of positive and negative emotions,'' *Arch. Neurol.*, vol. 39, pp. 210–218, Apr. 1982.

[28] H. A. Sackeim and R. C. Gur, ''Lateral asymmetry in intensity of emotional expression,'' *Neuropsychologia*, vol. 16, no. 4, pp. 473–481, Jan. 1978.

[29] R. Bracewell, *The Fourier Transform and its Applications*. New York, NY, USA: McGraw-Hill, 2002.

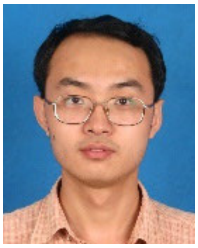
[30] F. J. Harris, ''On the use of windows for harmonic analysis with the discrete Fourier transform,'' *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.

[31] M. N. Riaz, Y. Shen, M. Sohail, and M. Guo, ''EXnet: An efficient approach for emotion recognition in the wild,'' *Sensors*, vol. 20, no. 4, p. 1087, Feb. 2020.

[32] K. Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, ''Region attention networks for pose and occlusion robust facial expression recognition,'' *IEEE Trans. Image Process.*, vol. 29, pp. 4057–4069, 2020.

[33] Y. Li, J. B. Zeng, S. G. Shan, and X. L. Chen, ''Occlusion aware facial expression recognition using cnn with attention mechanism,'' *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2439–2450, May 2019.

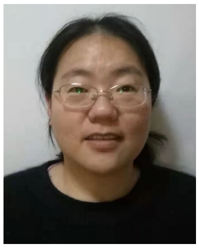
[34] J. Cai, Z. Meng, A. S. Khan, Z. Li, J. O'Reilly, and Y. Tong, ''Island loss for learning discriminative features in facial expression recognition,'' in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Xi'an, China, May 2018, pp. 302–309.



TIANYANG CAO received the Ph.D. degree in control science and engineering, in 2012. From 2015 to 2017, he was a Postdoctoral Research Fellow with the Institute of Electronics, Chinese Academy of Sciences. He is currently an Assistant Research Fellow with the State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include computer vision, image processing, signal processing, and multi-sensor information fusion.

CHANG LIU received the Ph.D. degree in electronic science and technology engineering, in 1996. He is currently a Research Fellow with the State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include electronics, MEMS sensor, novel sensing materials, image processing, and signal processing.

JIAMIN CHEN received the Ph.D. degree in materials science and engineering, in 2017. He is currently a Research Fellow with the State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include magnetic sensor, novel sensing materials, MEMS sensor, and image processing.



LI GAO received the Ph.D. degree in educational economy and management, in 2012. She is currently an Assistant Research Fellow with the Faculty of Education, Beijing Normal University. Her research interests include educational psychology, and teaching effect analysis in networks teaching and distance learning.