IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 December 2023, accepted 23 January 2024, date of publication 10 April 2024, date of current version 17 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3359244

RESEARCH ARTICLE

Facial Paralysis Symptom Detection Based on Facial Action Unit

HEQUN NIU[1,3], JIPENG LIU[2], XUHUI SUN[1,3], XIANGTAO ZHAO[1,3], AND YINHUA LIU[1,3,4]

[1]School of Automation, Qingdao University, Qingdao 266071, China
[2]College of Computer Science and Technology, Qingdao University, Qingdao 266071, China
[3]Institute for Future, Qingdao University, Qingdao 266071, China
[4]Shandong Key Laboratory of Industrial Control Technology, Qingdao 266071, China

Corresponding author: Yinhua Liu (liuyinhua@qdu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1313600.

ABSTRACT Facial paralysis refers to the abnormal behavior of facial muscles caused by a disorder of the facial nerve, mainly manifested as facial asymmetry. In recent years, deep learning has found extensive applications in facial paralysis detection research. However, most existing methods are constrained to assessing the severity of facial paralysis, thereby concealing crucial symptoms within black-box models. Compared to the severity of facial paralysis, the symptoms of facial paralysis are of greater significance to both physicians and patients. To address this issue, this paper proposes a facial paralysis symptom detection model based on facial action units (AUs). To enhance the accuracy of AU intensity prediction, a novel Difference Ensemble Method (DEM) is introduced. This method leverages differential information between frames within the same video to improve the accuracy of predictions for the current frame. Building upon the predicted AU intensity sequences for keyframes in a video, an interpretable model for detecting facial paralysis symptoms is designed. This model employs an active means to describe the asymmetry in facial muscle strength and utilizes co-occurrence matrices to detect synkinesis. It is noteworthy that DEM is exclusively trained on a dataset of normal faces but exhibits excellent performance when transferred to a facial paralysis dataset. Additionally, DEM exhibits higher accuracy in predicting AU intensity compared to existing methods. The F1 scores for detecting facial muscle function in the eyebrow, eye, and mouth regions with our proposed model are 80.0%, 79.23%, and 90.91%, respectively. To demonstrate the model's performance, a synkinesis detection experiment is conducted, further validating its applicability in facial paralysis detection.

INDEX TERMS Co-occurrence matrix, difference ensemble method, deep learning, facial paralysis symptom detection, facial action unit.

I. INTRODUCTION

Facial paralysis, also known as facial nerve paralysis, is a commonly encountered clinical condition [1]. Following the onset of the disease, motor dysfunction in the muscles responsible for facial expressions becomes apparent, resulting in asymmetrical mouth and eye movements. This condition significantly impacts patients' social activities, diminishing their quality of life [2]. Approximately 1.67% of people worldwide experience facial paralysis [3], and it affects individuals across various age groups. In the contemporary context, the prevalent pressures of survival and substantial workloads, particularly among the younger population, contribute to an alarming increase in the incidence of facial paralysis each year [4].

Detecting facial paralysis is crucial for assessing the degree of dysfunction in the facial nerve and muscles, as well as for monitoring changes in a patient's physical health during treatment and follow-up [5]. At present, the detection of facial paralysis is judged by clinicians according to their own clinical experience and relevant standards, due to the existing medical conditions and the limited number of relevant specialists, and the problem of doctor-patient imbalance is more serious [6]. In addition, it is difficult for paralyzed

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

patients to find their facial paralysis in the early stage, which makes it difficult for many patients to get timely and effective treatment, further aggravating the condition. On the other hand, doctors are influenced by subjective factors when making a diagnosis of facial paralysis. At present, the medical community utilizes various facial nerve grading scales to evaluate patients' faces. These including the Sunnybrook Facial Grading System (SFGS) [7], the House-Brackmann Scale [8], the Toronto facial grading system (TFGS) [9], [10], and Facial Nerve Grading System 2.0 (FNGS2.0) [11], among other. Unfortunately, there is no unified standard for evaluating facial nerve function, leading to inconsistent diagnostic conclusions and treatments by different doctors for the same patient. This discrepancy significantly impacts patients' timely consultations, doctors' choice of medical methods, and the prompt evaluation of treatment efficacy.

In recent years, numerous studies leveraging deep learning have achieved automatic detection of facial paralysis. For example, Hsu et al. introduced a deep hierarchical network (DHN) designed for the quantitative analysis of facial paralysis, utilizing a straight-line segmentation strategy based on face landmark localization [12]. Liu et al. proposed an objective method using facial videos and applying machine learning models to provide assessment results [13]. Parra-Dominguez et al. proposed a method for detecting facial paralysis in face photographs. Facial asymmetry was measured using facial landmarks, and a binary classifier based on a multilayer perceptron approach provided the output labels [5]. However, the method based on facial landmarks is not the underlying mechanism of facial paralysis and cannot further explain the symptoms and causes of facial paralysis.

Liu et al. proposed a parallel hierarchical convolutional neural network that combines the structure of Long Short-Term Memory (LSTM) networks to quantitatively assess the severity of facial paralysis through facial asymmetric features in regions and temporal changes in image sequences [14]. Xu et al. proposed a two-path LSTM network to extract global and local facial motor features, fuse the extracted advanced characterization information, and finally evaluate facial paralysis [15]. Zhang et al. proposed a deep learning-based method for the automatic prediction of facial paralysis grading [16]. These methods operate as end-to-end black box models, where the picture or video data of facial paralysis patients is input, and the facial paralysis grade is output. Although these methods demonstrate satisfactory predictive effects, they lack the capability to provide an explanation for the symptoms of facial paralysis. Both doctors and patients require an understanding of the underlying causes of facial paralysis, not solely its degree. Furthermore, it is important to note that deep learning methods heavily depend on a substantial amount of high-quality data. Individuals with facial paralysis often exhibit reluctance to share their photos or videos, posing a challenge in implementing deep learning methods with a limited dataset for facial paralysis.

Facial paralysis symptoms are interpreted based on the intensity of Facial Action Units (AUs) in the Facial Action Coding System (FACS), addressing the requirements of both physicians and patients. The FACS is a widely used protocol for recognizing and labeling facial expressions, providing a description of the movement of facial muscles [17]. AUs are defined as the minimum units of facial movement. They can appear individually or in combination, and each facial movement activates one or more AUs. The intensity of an action unit (AU) reflects the contraction state of the facial muscles. FACS uses the letters A to E to represent AU intensity changes from barely detected or tracked (A) to maximum intensity (E).

To accurately analyze facial paralysis symptoms, it is crucial to first accurately predict AU intensity. For example, Zhao et al. proposed a joint patch and multi-label learning (JPML) framework, leveraging group sparsity to identify important facial patches. JPML then employs a multi-label classifier, constrained by the likelihood of co-occurring AUs, to enhance prediction accuracy [18]. Li et al. designed a set of the adaptive region of interest (ROI) cropping networks, learning regional characteristics separately. They utilized multi-label learning to integrate the output of individual ROI cropping networks, explored interrelationships between these networks, and obtained global characteristics for sub-region AU detection [19]. Wu et al. introduced a contrasting feature learning method utilizing Convolutional Neural Network (CNN) learning. This method extracts feature differences between neutral faces and those displaying AUs, facilitating AU detection based on these distinctive features [20]. Yao Xia proposed a three regions-based attention network (TRA-Net) that divides the face into upper, middle, and lower regions. AUs are grouped according to their occurrence locations, and higher-level features are extracted using three consecutive soft attention modules for final AU detection [21]. It is evident that numerous studies enhance AU prediction accuracy by identifying differences between actions or by narrowing the model's focus. These insights serve as inspiration for the design of our model.

To address these problems and build upon previous experiences, a new method for facial paralysis symptom detection based on facial action units is proposed. The contributions are as follows:

1) We introduce a novel difference ensemble method to enhance the accuracy of AU intensity prediction for the current frame. This is achieved by leveraging differential information between the current frame and other frames within the same video.
2) We utilize the Active Mean (AM) to characterize AU intensity in video sequences for detecting abnormal facial functions and leverage a co-occurrence matrix for identifying synkinesis. This method not

only exhibits strong interpretability but also provides valuable insights into the essence of facial paralysis symptoms.

The remainder of this paper is organized as follows: Section II reviews relevant literature on facial action unit intensity prediction and facial paralysis detection. Section III provides a detailed description of the difference integration method and the facial paralysis symptom detection model. Section IV introduces the dataset and outlines the data preprocessing process. Section V describes the model training process, backbone network selection, comparative experiments, and the results of facial paralysis symptom detection. Finally, Sections VI and VII conclude with a discussion of the results.

II. RELATED WORK

A. FACIAL ACTION UNIT INTENSITY PREDICTION

Facial AU intensity prediction, a pivotal task in facial behavior analysis, has garnered widespread attention in recent years. Zhang et al. proposed a weakly supervised block depth model based on two attentional mechanisms to predict AU intensity [46]. Chen et al. utilized a regional attentional AU strength estimation method via uncertainty-weighted multi-task learning with a multi-head self-attention mechanism to avoid redundancy and achieve attentional coding for each patch [47]. These innovative methods provide strong support for accurate estimation of facial action intensity. In the field of facial expression analysis and face recognition [57], [58], [59], facial action unit strength prediction has received much academic attention. Seuss et al. employed a hybrid approach to estimate AU strength for emotion assessment via linear regression [48]. Wang et al. introduced a multiple facial AU recognition and intensity estimation method, implemented by modeling the relationship between AUs in feature space and label space [22]. Hupont et al. proposed a real-time method for detecting the intensity of AUs based on the scale of a facial action coding system. Real-time processing is achieved by combining a histogram of gradient descriptors with a linear kernel support vector machine [23]. Wei et al. proposed a regression method capable of obtaining AU intensity robustly and accurately. The method extracts multi-scale spatial features and corresponding temporal features from faces in a sequence of video images and learns the local relationships of these spatiotemporal features [24]. Ge et al. designed the Adaptive local - global Relational Network to be flexibly adapted to facial tasks by adaptively mining explicitly defined muscle regions of the face to enhance the visual details of facial appearance and texture [44]. Ntinou et al. proposed a simple but effective method based on heatmap regression to solve the problem of localization and strength of AUs [49]. These methods not only contribute significantly to the basic AU intensity estimation but also have potential applications in sentiment analysis. To improve the efficiency of AU intensity estimation, Fan et al. introduced knowledge distillation (KD) for training models [50]. However, these methods do not fully consider the dynamic characteristics of the subject as an individual, resulting in the inability to comprehensively extract dynamic information. In this context, Ma et al. proposed a method to quickly construct an AU intensity prediction model and successfully constructed an automatic estimation model of AU intensity for face images [51]. Different from the above methods, DEM improves the accuracy of AU intensity prediction for the current frame with the assistance of other frames in the same video. Comparative experimental results demonstrate superior performance in AU intensity prediction compared to existing methods.

B. FACIAL PARALYSIS DETECTION

Detection of facial paralysis symptoms has been widely studied in recent years because of the obvious psychological and functional impact of this disorder on patients. In clinical practice, facial abnormalities are detected through a systematic visual examination of facial morphology and muscle movements [1]. During the assessment, the patient is asked to perform specific facial movements such as smiling, raising the eyebrows, closing the eyes, and bulging, which are then scored by the clinician on a facial grading scale [13]. In addition, there are many types of facial nerve grading scales available [7]. However, these traditional assessment methods are time-consuming and subjective, as the process must rely on medical professionals to perform them. In recent years, deep learning has been widely applied in the healthcare sector [60], [61], [62], particularly in the assessment of the severity of facial paralysis. The cascade encoder structure adopted by Wang et al. fully exploits the advantages of face semantic features in face spatial information extraction, which contributes to the accuracy of facial paralysis assessment [52]. While Parra-Dominguez et al. performed facial palsy detection on images by keypoint analysis [5]. Gogu et al. proposed an automatic facial paralysis recognition method for classifying facial paralysis and healthy subjects [53]. Ge et al. proposed a new adaptive local-global relational network (ALGRNet) for facial AU detection and used it for facial paralysis severity classification [45]. Literature [14], [15], and [16] also uses deep learning algorithms to classify the severity of facial paralysis. However, most current methods use end-to-end black-box models to classify the severity of facial paralysis but fail to adequately capture the problem of facial paralysis symptoms potentially in the black-box model. Boochoon et al. also pointed out several challenges faced by deep learning in facial paralysis detection, including the quality of the data and the interpretability of the model of the machine learning algorithm [63]. Compared to existing methods, this study uses facial AU intensity to detect facial muscle function and the AU co-occurrence matrix to detect synkinesis, which is more interpretable and can

TABLE 1. AUs from the Facial Action Coding System.

AU	Description	Muscle name
AU1	Inner Brow Raiser	Frontalis
AU2	Outer Brow Raiser	Frontalis
AU4	Brow Lowerer	Corrugator supercilii
AU5	Upper Lid Raiser	Eyelid muscles, Tarsal muscle
AU6	Cheek Raiser	Orbicularis oculi
AU9	Nose Wrinkler	Square muscle of upper lip
AU12	Lip Corner Puller	Zygomaticus major
AU15	Lip Corner Depressor	Triangularis
AU17	Chin Raiser	Mentalis
AU20	Lip Stretcher	Risorius muscle, Platysma muscle
AU25	Lips Part	Square muscle of lower lip, chin muscle, orbicularis oris muscle
AU26	Jaw Drop	Masseter, temporalis, pterygoid

meet the needs of doctors and patients for facial paralysis detection.

III. METHODOLOGY

A. FACIAL ACTION UNIT

The facial AU describes the changes in appearance caused by a set of facial muscle movements, and their combination can convey various facial expressions. Facial AUs are considered as the mapping from facial muscles to facial actions. They are chosen as the foundational elements for interpreting facial paralysis symptoms, distinguishing our method significantly from other facial paralysis detection methods. The 12 AUs used in our model and their corresponding muscles are shown in Table 1.

B. MODEL ARCHITECTURE

The overall architecture of the facial paralysis symptom detection model is shown in Figure 1. The model is divided into two parts based on the AUs: the prediction of AU intensity and the facial paralysis symptom detection based on the AU intensity. As facial paralysis symptoms are better revealed during dynamic movements, videos are selected as input for the AU intensity prediction model. Data processing is required before prediction can be performed. The processed data is input into the AU prediction model based on the difference ensemble method to predict AU intensity. To improve the accuracy of each frame's prediction, other frames from the same video are randomly selected to assist in the prediction process. Facial paralysis symptoms are detected using the predicted AU intensity. Detection of abnormal facial muscle function by calculating the active mean value of AU. Additionally, synkinesis between facial muscles is detected using the AU co-occurrence matrix.

C. DIFFERENCE ENSEMBLE METHOD

The difference ensemble method (DEM) is a strategy that can improve the accuracy of AU intensity prediction for the current frame by predicting the difference between frames. DEM requires the establishment of two models: the original value prediction model (OVPM) and the difference value prediction model (DVPM), with the data flow between the models illustrated in Figure 2.

To improve the accuracy of the ith frame original data X_i to predict Y_i, the original value prediction model needs another original data X_j to assist the prediction. In the original value prediction model, two sets of original data, X_i and X_j, serve as inputs, yielding predicted values Y_i and Y_j after model inference. Simultaneously, the model also generates intermediate features $F_i^{(k)}$ and $F_j^{(k)}$ corresponding to the input data X_i and X_j. The difference value prediction model takes the form of the difference $\Delta F_{i,j}^{(k)}$ between intermediate features $F_i^{(k)}$ and $F_j^{(k)}$ as its input, producing the difference $\Delta Y_{i,j}$ between the predicted values Y_i and Y_j. $\Delta F_{i,j}^{(k)}$ and $\Delta Y_{i,j}$ are defined as follows:

$$\Delta {F_{i,j}}^{(k)} = F_i^{(k)} - F_j^{(k)} \tag{1}$$

$$\Delta Y_{i,j} = Y_i - Y_j \tag{2}$$

where $F_i^{(k)}$ represents the kth intermediate feature generated when predicting input X_i. $\Delta Y_{i,j}$ is required as the label of the difference value prediction model during model training.

The predicted value of X_i is ${Y_i}^{(j)}$ with the assist of X_j. The ${Y_i}^{(j)}$ is calculated by

$$Y_i^{(j)} = Y_j + \Delta Y_{i,j} \tag{3}$$

where Y_j is original value predicted from X_j by original value prediction model. $\Delta Y_{i,j}$ is difference value predicted by difference value prediction model.

In order to "shop around", m original data $X_{j_1}, X_{j_2}, \ldots, X_{j_m}$ are used to assist in the prediction, and the prediction value of $Y_i^{(j_1)}, Y_i^{(j_2)}, \ldots, Y_i^{(j_m)}$ of Y_i are obtained respectively. The average of all prediction values is calculated as the final prediction values $\hat{Y}_i$. The overall structure of DEM is shown in Figure 3. The pseudo code is shown in Algorithm 1.

In Section V-B, an appropriate backbone network is selected for the original and difference value prediction model. In addition, the improvement effect of DEM was verified. Several key points should be considered when using the DEM:

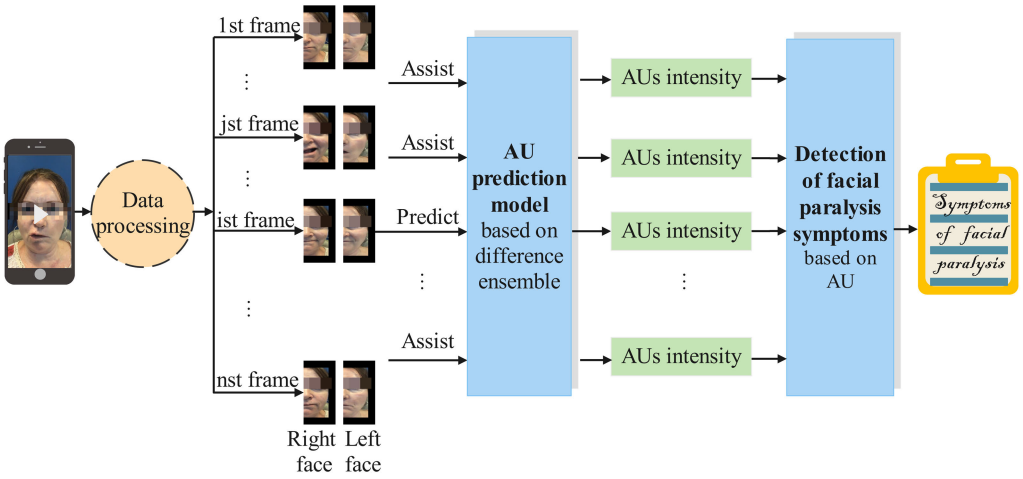


FIGURE 1. The overall architecture of facial paralysis symptom detection. Note that due to the principle of mirror symmetry, a flip operation was performed on the right face during data processing.

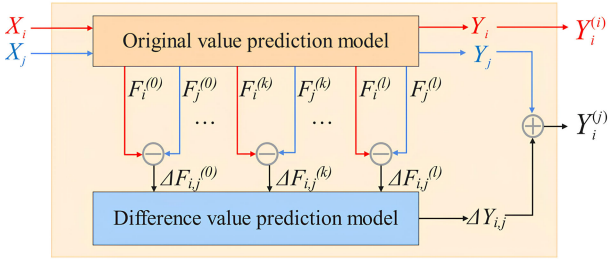


FIGURE 2. Data flow of the original value prediction model and the difference value prediction model.

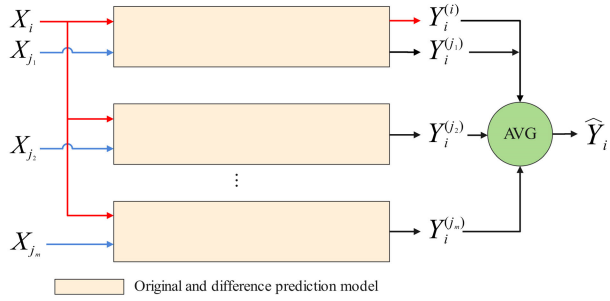


FIGURE 3. The overall structure of DEM.

1) The model parameters of the difference value prediction model depend on those of the original value prediction model. This dependency arises from the fact that the former requires the intermediate features of the latter as inputs, which are directly influenced by the model parameters of the original value prediction model. Consequently, any adjustments made to the model parameters of the original value prediction model require retraining the difference value prediction model. The trained original value prediction model and the difference value prediction model should be matched for use.

Algorithm 1 DEM-Based AU Intensity Prediction

Require: Facial movement video X

Ensure: AU intensity Y

1: $X_i, X_j \leftarrow$ Perform i_{th}, j_{th} frame extraction on X;
2: $Y_i, F_i^{(k)} \leftarrow OVPM\ (X_i)$ for each $k \in \{1, 2, \ldots, l\}$;
3: $Y_j, F_j^{(k)} \leftarrow OVPM\ \left(X_j\right)$ for each $k \in \{1, 2, \ldots, l\}$;
4: Compute $\Delta F_{i,j}^{(k)} = F_i^{(k)} - F_j^{(k)}$;
5: $\Delta Y_{i,j} \leftarrow DVPM\left(\Delta F_{i,j}^{(k)}\right)$;
6: Compute $Y_i^{(j)} = Y_j + \Delta Y_{i,j}$;
7: Calculate $Y_i = avg\left(Y_i^{(j)}\right)$;

2) The method proves more effective in enhancing results for problems involving the assessment of fuzzy degrees and may not be as effective in improving outcomes for classification problems, such as cat and dog recognition.
3) The method calculates the difference value by subtracting corresponding elements from two feature vectors. Thus, it is crucial to ensure that the feature meanings represented by the corresponding elements in the subtracted feature vectors are consistent.
4) Similar backbone networks are chosen for both the original value prediction model and the difference value prediction model to ensure seamless integration of shallow features from the former into the corresponding shallow locations of the latter, and to enable smooth integration of deep features from the former into the corresponding deep locations of the latter.

D. SYMPTOM DETECTION BASED ON AU

The AU intensity is predicted by DEM, as discussed in the previous section. Furthermore, the prediction accuracy can be gradually enhanced with the assistance of other frames.

TABLE 2. The 10 AUS used for abnormal facial function detection.

Facial Area	AUs
Eyebrows	AU1, AU2, AU4
Eyes	AU5, AU6
Mouth	AU12, AU15, AU17, AU20, AU25

In this section, AU intensity sequences are employed to detect two symptoms of facial paralysis: abnormal facial muscle function and synkinesis.

1) ABNORMAL FACIAL MUSCLE FUNCTION DETECTION MODEL

Abnormal facial muscle function is mainly reflected in the asymmetry of AU intensity values. Asymmetry is the main basis for our abnormal detection. The judgment of abnormal facial muscle function involves comparing the difference in AU intensity between the left and right faces. The active mean (AM) is defined to describe AU intensity in video sequences, and it can be represented as follows:

$$AM_{pos} = \frac{1}{|S|}\sum_{i \in S} AU_{i,p\ os} \tag{4}$$

$$S = \left\{ i \mid AU_{i,\text{ left}} \text{ is active or} AU_{i,\text{right}} \text{ is active} \right\} \tag{5}$$

where i represents the index of a video sequence, pos indicates the position of the left or right face, $AU_{i,pos}$ represents the AU intensity value in the ith frame at the pos position, and AM_{pos} represents the AM value of the pos position. The AU activities can be determined using a given threshold. If the value is greater than the threshold, the AU is considered active; otherwise, it is considered inactive. In this study, the threshold is set to 0.5.

The difference between the AM of left and right facial muscles is utilized to describe the asymmetry in their activation. It is defined as

$$\Delta AM = AM_{left} - AM_{right} \tag{6}$$

where ΔAM is the difference between the left and right facial muscles. If ΔAM is greater than 0, it means that the left facial muscle is stronger than the right facial muscle. On the contrary, if ΔAM is less than 0, it means that the right facial muscle is stronger than the left facial muscle.

In order to meet the clinical assessment requirements for facial paralysis, ten AUs were selected, corresponding to the eyebrow, eye, and mouth areas. Ten AUs in FACS were used for abnormal facial function detection as shown in Table 2.

The ΔAM sum of multiple AUs in the same area was taken as the difference value of left and right face activity in that area. If there is a large difference between the left and right faces, the muscle function of the weaker side is considered damaged. The degree of damage is quantified by the value of $|\Delta AM|$.

2) SYNKINESIS DETECTION MODEL

Facial synkinesis is a feature secondary to facial paralysis. Specifically, when a patient with facial paralysis attempts to voluntarily contract one facial area, abnormal and involuntary muscle contractions occur in the other facial area [26]. In other words, two uncorrelated AUs are activated simultaneously when synkinesis occurs in the face. The co-occurrence matrix (CM) was used to describe the probability of co-activation between AUs. Left and right face asymmetry is discussed separately. Four CMs are built, including $CM_{left|left}$, $CM_{left|right}$, $CM_{right|left}$, and $CM_{right|right}$. The probability of an element in row i and column j of $CM_{pos1|pos2}$ activation is defined as

$$P(AU_{pos1,i}\text{is active}|AU_{pos2,j}\text{is active}) = \frac{N_{AU_{pos1,i}+AU_{pos2,j}}}{N_{AU_{pos2,j}}} \tag{7}$$

where $pos1$ and $pos2$ represent the left or right face, i and j indicate the number of AU, $N_{AU_{pos1,i}+AU_{pos2,j}}$ is the total number of simultaneous occurrences of $AU_{pos1,i}$ and $AU_{pos2,j}$, and $N_{AU_{pos2,j}}$ is the number of occurrences of $AU_{pos2,j}$. A threshold is set to determine whether AU was activated or not. If it is greater than the threshold, the AU is considered active, otherwise, it is considered inactive. The threshold is set to 0.5.

The difference between the left and right faces is obtained by calculating the difference between the two co-occurrence matrices $CM_{right-left|left}$ and $CM_{left-right|right}$, defined as

$$CM_{right-left|left} = CM_{right|left} - CM_{left|left} \tag{8}$$

$$CM_{left-right|right} = CM_{left|right} - CM_{right|right}. \tag{9}$$

The co-occurrence difference matrix describes the difference in co-occurrence between the left and right sides. If the value is positive, it means that the co-occurrence of the opposite side is stronger than that of the same side. On the contrary, the same side is stronger than the opposite side. The degree of difference can be described by $|CM_{right-left|left}|$ and $|CM_{left-right|right}|$, which is the absolute value of the two matrices. If the difference exceeds this threshold, the two AUs are considered to be highly correlated. The threshold was taken to be 0.3.

While synkinesis shares a resemblance with the co-occurrence matrix, there exist distinctions. Synkinesis necessitates the fulfillment of two conditions: firstly, a lower correlation among AU pairs in individuals without facial paralysis; secondly, a higher correlation among AU pairs in patients afflicted with facial paralysis. Consequently, when utilizing the co-occurrence matrix, it becomes imperative to selectively filter correlated AU pairs within the normal population.

Mavadati et al. statistically analyzed the AU co-occurrence matrix of 12 basic expressions of normal subjects' expressions and the results are shown in Fig. 4(a) [27]. To meet the needs of facial paralysis synkinesis detection, the AU co-occurrence matrix was filtered. In this matrix, elements

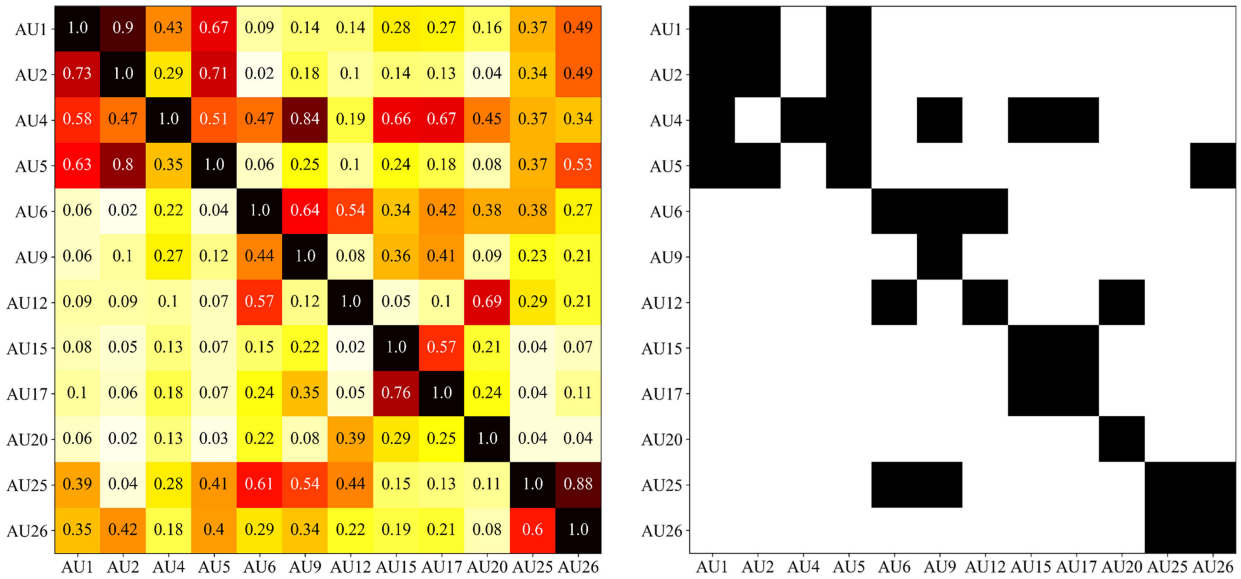


FIGURE 4. AU co-occurrence matrix and mask matrix to meet the needs of facial paralysis synkinesis detection.

above the threshold (0.5) were filtered out and set to black, while elements below the threshold were retained and set to white, resulting in the filtered result as shown in Fig. 4 (b). This matrix was used as a mask to filter the co-occurrence difference matrix. The pseudo code is shown in Algorithm 2.

Algorithm 2 Au-Based Facial Paralysis Detection

Require: AU intensity Y
Ensure: Facial Paralysis Symptoms Z
Compute AM by equation 4;
Compute $\Delta AM = AM_{left} - AM_{right}$;
if $\Delta AM > 0$ **then**
  $Z \leftarrow$ Abnormal right side of face;
else
  $Z \leftarrow$ Abnormal left side of face;
end if
Compute CM by equation 8 and equation 9;
Filter CM;
if $CM \neq 0$ **then**
  $Z \leftarrow$ synkinesis;
else
  $Z \leftarrow$ no synkinesis;
end if
return Z;

IV. DATA PROCESSING

A. DATASETS

The DISFA+ dataset is a public dataset with high-quality AU intensity labels [27]. It contains posed and spontaneous facial expression data for a set of individuals and provides manually labeled frame-based annotations of the intensity of twelve FACS facial actions. The intensity of each AU was labeled on a 6-point intensity scale [0-5]. DISFA+ selected nine subjects from DISFA and asked them to imitate 42 facial movements and record their postural facial movements [28]. These participants covered a wide range of ethnicities, including Asian, African American, and Caucasian. Compared to other public datasets (such as DISFA, BP4D [29]), the distinctive feature of the DISFA+ dataset lies in its unique collection process, which involves instructing participants to mimic 42 facial expressions. This approach is akin to the way doctors request patients to perform a series of facial movements during facial paralysis diagnoses. Furthermore, the DISFA+ dataset exhibits standardized and distinct AU features, facilitating enhanced model learning and performance. Consequently, we have chosen the DISFA+ dataset as the foundation for our research.

To test the performance of the facial paralysis symptom detection model, videos of the facial movements of 34 patients with facial paralysis were obtained from various channels. The demographic and facial behavior video information of the patients with facial paralysis is reported in Table 3. In these videos, specialists annotate the motor functions of the patient's left and right facial muscles (including the eyebrows, eyes, and mouth). Considering the privacy protection of patient information, this dataset could not be publicly published.

B. DATA PREPROCESSING

Process the input video data to transform it into half-face keyframe images. The data processing is divided into three main steps: face target detection, filtering to extract keyframes, and splitting the face into halves.

TABLE 3. Demographic and facial behavior video information of patients with facial paralysis.

Number of persons	34
Age range (years)	15-65
Mean age (years)	42
Gender	16 male, 18 female
Number of videos	34
Number of video frames	16700

1) FACE TARGET DETECTION

The Retinaface target detection model is used to select the face target box sequence in the video. 5 facial feature points (left eye, right eye, nose tip, left corner of mouth, right corner of mouth) are predicted by the Retinaface model.

2) FILTER TO GET KEYFRAMES

The method of filtering keyframes by subtracting two images requires the face's position in the image to remain relatively stable. To ensure facial stability, we select the midpoints of three feature points (left eye, right eye, and nose tip) that do not move with facial expressions as relative positioning points. Subsequently, the images are downsampled to 48 $\times$ 48 pixels using Bilinear interpolation. This downsampling helps eliminate the effects of small-scale shaking while retaining large-scale variations in facial behavior. The two downsampled images were subtracted and a pixel threshold was set to filter out similar frames.

3) SPLIT HALF OF THE FACE

Firstly, an affine transformation is applied using three feature points (left eye, right eye, and nose tip) to achieve face alignment. Secondly, the image is resized to 512 $\times$ 512 pixels by filling the surrounding areas with a grey bar. Thirdly, the face image is manually divided into two parts: the left and the right. Finally, the right half-face image is flipped to obtain the data format for the model input. The image size of the input model is 512 $\times$ 256.

V. EXPERIMENT AND RESULT ANALYSIS

In this section, model training and facial paralysis symptom detection experiments are conducted to select the best backbone network and validate the performance of the proposed model.

A. MODEL TRAINING

After data processing, a total of 10,010 frames were obtained from the DISFA+ dataset. These images were divided into an 8:2 ratio and used as a train set and test set for the original value prediction model, respectively. To train the difference value prediction model, a new dataset needs to be created. This model requires the difference between the feature maps of the two frames as well as the difference between the corresponding labels. To achieve this, we combined frames from the same video in the DISFA+ dataset in pairs, ensuring that the combined frames were not duplicated. After this process, 78,400 distinct frame pair combinations were obtained and subsequently divided into training and testing sets in an 8:2 ratio for the difference value prediction model.

The PyTorch is chosen as the deep learning framework and the NVIDIA GeForce RTX 3090 GPU is used for the experiments. After numerous tests, optimal training hyperparameters are determined: the initial learning rate is set to 0.0001, employing an exponential decay strategy for learning rate scheduling with a decay gamma coefficient of 0.98. The chosen loss function is cross-entropy loss. The batch size is adjusted based on the video memory size. In this study, the 3090 GPU has a video memory size of 32G. The batch size of the original value prediction model is set to 64. As the original value prediction model also utilizes video memory during the training of the difference value prediction model, the batch size for training the difference value prediction model is set to 16.

The input of the difference value prediction model depends on the output of the original value prediction model. Consequently, during model training, the original value prediction model is executed first, obtaining the feature map, which is then used as input for the difference value prediction model. The ΔY range outputted by the difference value prediction model is discrete and falls between -5 and 5. The final prediction results of the DEM are continuous numerical values, with a range limited to between 0 and 5.

TABLE 4. Accuracy of different networks in predicting raw values in the test set.

Model	Accuracy	Model	Accuracy
ResNet-50	88.81%	MobileNetV3	90.84%
ResNet-101	90.70%	EfficientNet-B6	90.56%
ResNet-152	90.95%	EfficientNet-B7	90.88%
DenseNet-169	88.95%	VGG-16	90.59%
DenseNet-201	89.56%	VGG-19	91.67%

B. SELECTION OF BACKBONE NETWORK

The most effective and suitable backbone network is selected through comparative experiments. Compare the performance of VGG, ResNet, DenseNet, MobileNet, and Efficient-Net in predicting original values [30], [31], [54], [55], [56]. The comparison results are shown in Table 4.

In Table 3, 5GG-19 demonstrates the highest performance in predicting the original values. In comparison to other networks, the VGG neural network demonstrates several advantages. Firstly, the VGG neural network has a relatively simple structure, facilitating the output of intermediate feature maps. Secondly, due to the significantly large fully connected part of the VGG neural network, the features in the convolutional part are relatively abstract. However, ResNet and DenseNet have deeper layer structures with a higher number of intermediate feature map outputs, making it challenging to determine the optimal solution. Moreover, they include only one fully connected layer, and

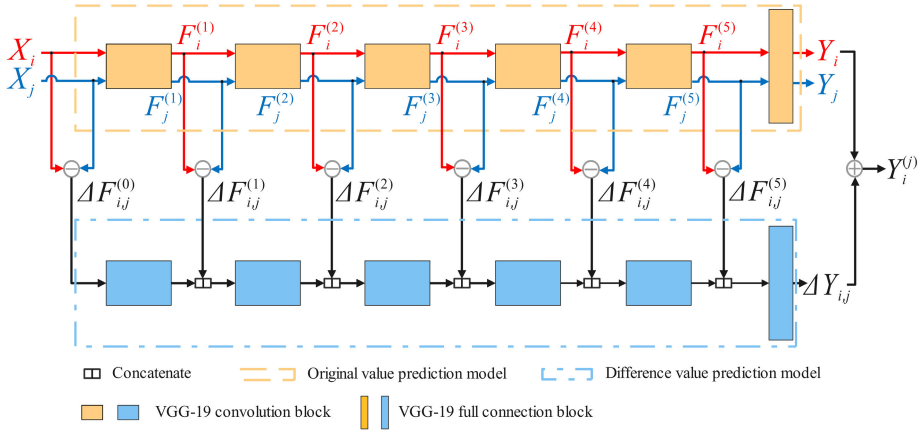


FIGURE 5. Network structure of DEM.

thus, in the later convolution stages, the original value prediction network is already close to the final prediction value. This approximation can interfere with the difference prediction model, affecting its ability to make more accurate predictions for differences. Experimental results indicate that the prediction accuracy of MobileNet and EfficientNet is above 90%, but still lower than VGG-19. Therefore, the VGG-19 was chosen as the backbone network for the original value prediction model, which is more suitable for DEM.

Combined with the discussion in Section III-C, the original value prediction model and the difference value prediction model should use similar backbone models. Therefore, VGG-19 was selected as the backbone network for the difference value prediction model. The network structure of the DEM is illustrated in Figure 5.

C. EFFECT IMPROVEMENT OF DEM

For each frame, m frames from the same set of keyframes are randomly selected as inputs to the DEM. The accuracy of predicting AU intensity and the processing time per image in the test set when m takes different values are shown in Figure 6.

From Figure 6, it can be seen that the accuracy of predicting AU intensity gradually increases from 91.07% to 96.66% with the increase of m value. The experimental data demonstrate that DEM is an effective method with a more noticeable improvement effect. In particular, the accuracy increased by 4.6% when m was between 0 and 2, after which the accuracy changed slowly. In addition, the processing time per image increases linearly with the growth of m. Therefore, in practical applications, the selection of m needs to find a balance between prediction accuracy and processing time. In this study, we consider setting the value of m to 2.

D. COMPARISON WITH RELATED WORKS ON AU INTENSITY PREDICTION

To demonstrate the superiority of our method, we compare it with other methods proposed in the literature. Table 5

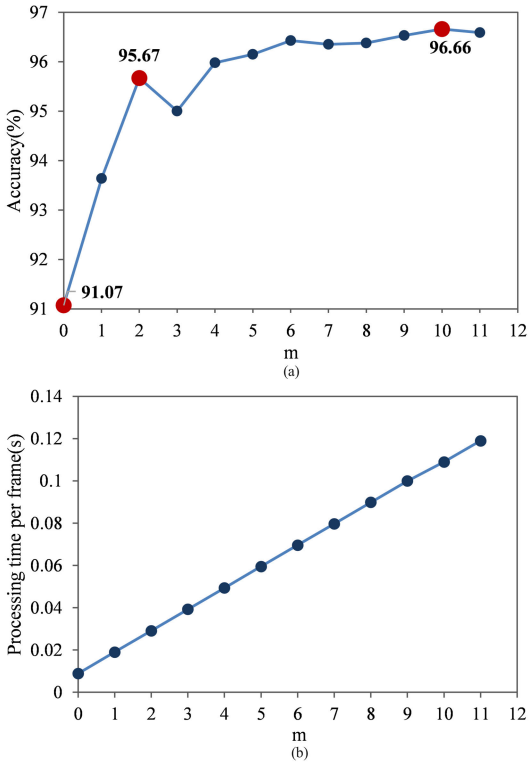


FIGURE 6. Effect improvement of the DEM. (a) represents the AU intensity prediction accuracy. (b) represents the processing time for each image.

presents a comparison of the Mean Absolute Error(MAE), Intraclass Correlation Coefficient(ICC), and Pearson Correlation Coefficient(PCC) results between our proposed method and other approaches on the DISFA+ dataset.

Walecki et al. [32] introduced the Copula Ordinal Regression (COR) framework to separate AU-dependent probabilistic modeling from edge modeling of AU intensity. Kaltwang et al. [33] proposed the Doubly Sparse Relevance Vector Machine (DSRVM) for the continuous estimation of facial behavior interpretation strength. Wang et al. [22]

TABLE 5. Performance comparison of AU intensity prediction on DISFA+.

Methods	MAE	ICC	PCC
COR [32]	1.15	0.19	0.22
DSRVM [33]	0.86	0.03	0.16
MTL+BN [22]	-	0.43	0.38
CCNN-IT [34]	0.61	0.45	-
BORMIR [35]	0.84	0.28	0.37
Wang et al. [36]	0.43	0.55	0.49
Ntinou et al. [49]	0.56	0.49	0.58
RA-UWML [47]	0.48	0.53	0.60
ALGRNet [45]	0.43	0.57	0.61
Ours	0.41	0.59	0.63

presented a multi-task feature learning technique for sharing features and a Bayesian network to capture AU label dependencies for AU intensity estimation. Walecki et al. [34] proposed a deep Convolutional Neural Networks-conditional random field (CNN-CRF) model to estimate multiple AU intensities. Zhang et al. [35] introduced a Bilateral Ordinal Relevance Multi-instance Regression model to learn a frame-level intensity estimator with weakly labeled sequences. Wang et al. [36] employed a deep framework to learn the basic attributes of each image. Support vector AU identification and support vector regression AU intensity estimation are trained by maximizing the log-likelihood AU mapping function. Ntinou et al. [49] utilized Heatmap Regression to estimate the AUs intensity. Chen et al. [47] adopted a regional attentional AU intensity estimation method with Uncertainty Weighted Multi-task Learning (RA-UWML) with a multi-head self-attention mechanism to avoid redundancy and achieve attentional coding for each patch. Ge et al. [45] proposed an adaptive local-global relational network (ALGRNet) model for facial AU detection to estimate the severity of facial paralysis. As can be seen from Table 5, our method achieves better performance compared to the others. This result fully validates the effectiveness and superiority of our model.

E. FACIAL PARALYSIS SYMPTOM DETECTION BASED ON AU

The AU intensity sequences for the left and right halves of the face in the video were obtained using the AU prediction model based on DEM. Facial paralysis symptoms were detected through an active-mean-based facial muscle function detection model and a co-occurrence matrix-based synkinesis detection model, both leveraging AU intensity sequence information.

1) ABNORMAL FACIAL MUSCLE FUNCTION DETECTION

The facial muscle function of the patients in the dataset was detected using an active-mean facial muscle function detection model. To ensure detection accuracy, a professional doctor labeled the facial muscle function strength of the left and right faces in the eyebrow, eyelid, and mouth areas in the videos of patients with facial paralysis. The labeling rules are as follows: if the left face is stronger than the right face, it is marked as 1; if the right face is stronger than the left face, it is marked as -1; if the difference between the two is not obvious, it is marked as 0. Subsequently, the facial muscle function of the dataset of patients with facial paralysis was evaluated, and the measure of the effect is shown in Figure 7.

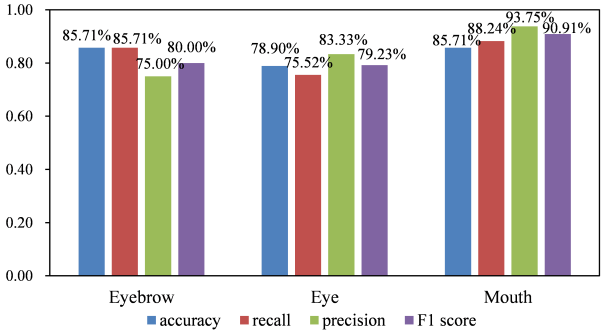


FIGURE 7. Effect of detection of abnormal facial muscle function.

As can be seen in Figure 7, the facial muscle function detection model has the highest detection effect for the mouth region and a slightly lower effect for the eyebrow and eye regions. Improvement in the detection results for the eyebrow and eye regions is planned for future work.

2) SYNKINESIS DETECTION

Only a small number of videos in the collected dataset show significant synkinesis. We demonstrate an experiment for synkinesis detection. Figure 8 illustrates the facial AU intensity curve of the patient. Figure 9 shows the co-occurrence difference matrix.

As seen in Figure 9, for AU5 and AU20 the co-occurrence value of the right face relative to the right face is stronger than that of the left face relative to the right face, as can also observed in Figure 8. The experimental results align with our expectations. Due to the absence of a gold standard, the outcomes of these experiments cannot be directly compared with those of other studies. In our forthcoming work, we plan to collect additional data for synkinesis detection experiments.

VI. DISCUSSION

Our research focuses on the use of computer vision techniques to detect facial paralysis, a common condition. By exploring the potential of extracting facial AU intensity from video data, we introduce a non-invasive method for detecting symptoms of facial paralysis, thereby advancing the use of artificial intelligence in healthcare. Our work addresses the need for low-stress, low-cost, and more accessible methods for facial paralysis detection, thereby enhancing the diagnostic and decision-making process in healthcare. In addition, the methods of our research can be integrated

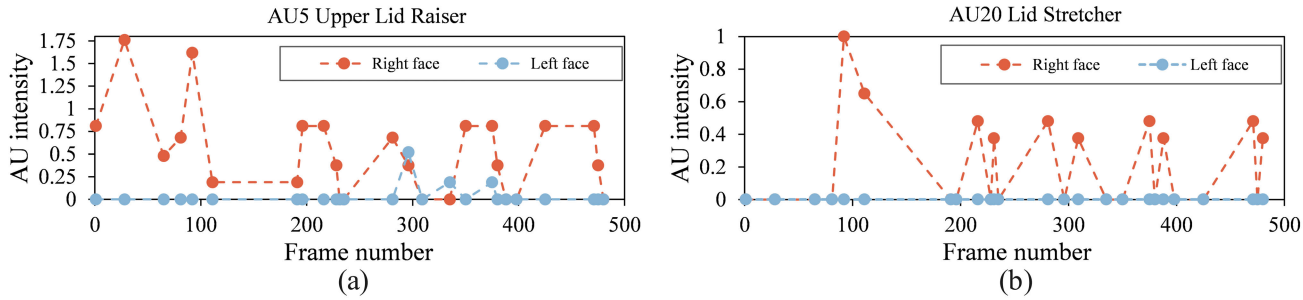


FIGURE 8. AU intensity curve for a certain patient. (a) represents the intensity variation of AU5 in the left and right half-faces. (b) represents the intensity variation of AU20 in the left and right half-faces.

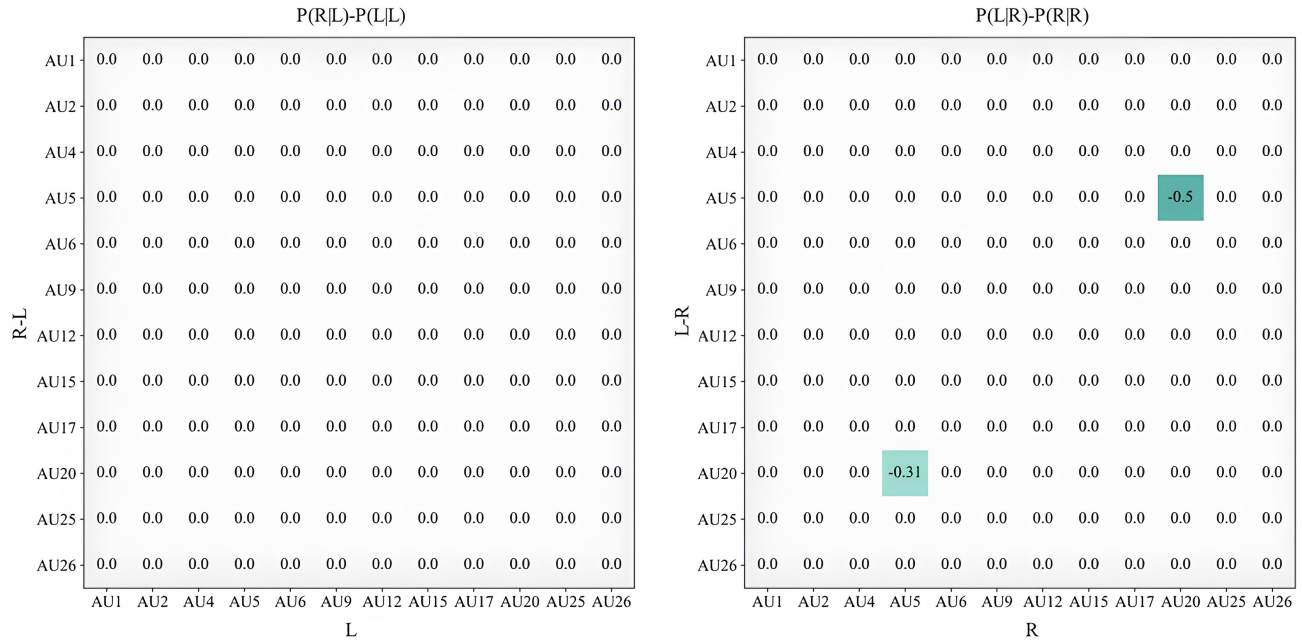


FIGURE 9. Co-occurrence difference matrix.

with medicine, enabling healthcare professionals to use advanced computational techniques to detect and analyze facial paralysis.

In this study, facial function was detected based on the asymmetry of AU intensity on both sides of the face. Facial asymmetry has also been highlighted in the literature [37], [38] as a key factor in detecting facial paralysis. Additionally, asymmetry analysis plays a crucial role in evaluating the success of surgical interventions following facial resuscitation therapy. Previous studies primarily concentrated on analyzing static asymmetrical facial features, neglecting adequate consideration of synkinesis features. To address this research gap, we opted to use video as an input for the model and propose methods for detecting synkinesis. Based on the experimental results, the method is promising and worthy of encouragement.

The facial paralysis symptom detection method proposed in this paper, based on facial AU, holds significant applicability and potential utility in diverse healthcare environments. Firstly, the method offers a non-invasive and privacy-preserving solution. In contrast to traditional detection methods like electro-neurography (ENoG) [39], surface electromyography (sEMG) [40], and electromyography (EMG) [41], the approach utilizes video data for facial feature extraction. These features can be easily captured, resulting in considerable time and resource savings, thereby enhancing efficiency in medical scenarios. Secondly, our method not only detects abnormal facial muscles through AU intensity but also identifies synkinesis using the AU co-occurrence matrix. To the best of our knowledge, there are no studies on synkinesis detection in facial paralysis, and we are the first to introduce it. Additionally, our method can be applied in telemedicine applications to facilitate remote detection and monitoring of individuals at risk of facial paralysis, especially in remote or resource-poor areas. Remote detection can further enable early identification

of potential facial paralysis symptoms, promoting timely intervention and treatment and thereby enhancing the overall quality of care for the individual.

In terms of computational complexity, our approach heavily relies on predicting and analyzing facial Action Unit (AU) intensity, which is a computationally relatively lightweight task. Nevertheless, our method necessitates processing a substantial amount of video data, potentially exerting pressure on computational resources, especially when dealing with high-resolution or high-frame-rate videos. However, given the continuous development of computing technology, we believe that this challenge can be effectively addressed.

This research currently has some limitations, primarily the lack of a large patient dataset. Because the sensitivity of patient privacy makes it challenging to collect enough data in healthcare, it is difficult to accumulate the large amount of data needed to train a robust and widely applicable machine learning model. This limitation has been acknowledged in other studies [42], [43]. To address this challenge efficiently, we utilized a normal face dataset for model training in the initial phase, and the transfer to facial paralysis symptom detection yielded positive results. Based on the available experimental results, the real-world application of the model may face some inherent constraints and limitations. Future work will concentrate on optimizing the model to enhance the accuracy of facial muscle function detection in the eyebrow and eye regions. In the meantime, we will gather more data from facial paralysis patients to validate the performance of the synkinesis detection model. Ultimately, we plan to integrate this model into an application to offer supportive assistance in clinical settings.

VII. CONCLUSION

In this study, we propose a highly interpretable model for detecting facial paralysis symptoms. The prediction accuracy of AU intensity is significantly improved by introducing a novel DEM. The method fully leverages information from other frames in the same video to effectively support the prediction of the current frame. We utilize the AU intensity of video keyframes to detect facial paralysis symptoms. Abnormal facial muscle function is detected by analyzing the mean AU intensity values. The AU co-occurrence matrix is used to detect facial paralysis synkinesis. Following comparative experiments, the backbone network that is most suitable for the DEM is selected, providing a solid foundation for accurate prediction of AU intensity. The experimental results demonstrate that the method not only outperforms existing methods in terms of performance but also exhibits the ability to effectively transition from normal facial AU detection to detecting facial paralysis patients. Future research will focus on optimizing the model to further improve the accuracy of facial muscle function detection in the eyebrow and eye regions. It is also planned to increase the dataset size to validate the synkinesis detection model to ensure its robustness in a wider range of scenarios and populations.

REFERENCES

[1] M. A. Alagha, A. Ayoub, S. Morley, and X. Ju, "Objective grading facial paralysis severity using a dynamic 3D stereo photogrammetry imaging system," *Opt. Lasers Eng.*, vol. 150, Mar. 2022, Art. no. 106876.

[2] L. Ishii, A. Godoy, C. O. Encarnacion, P. J. Byrne, K. D. O. Boahene, and M. Ishii, "Not just another face in the crowd: Society's perceptions of facial paralysis," *Laryngoscope*, vol. 122, no. 3, pp. 533–538, Mar. 2012.

[3] R. F. Baugh, G. J. Basura, L. E. Ishii, S. R. Schwartz, C. M. Drumheller, R. Burkholder, N. A. Deckard, C. Dawson, C. Driscoll, M. B. Gillespie, R. K. Gurgel, J. Halperin, A. N. Khalid, K. A. Kumar, A. Micco, D. Munsell, S. Rosenbaum, and W. Vaughan, "Clinical practice guideline," *Otolaryngol.-Head Neck Surg.*, vol. 149, pp. 1–27, Nov. 2013.

[4] I. Mavrikakis, "Facial nerve palsy: Anatomy, etiology, evaluation, and management," *Orbit*, vol. 27, no. 6, pp. 466–474, Jan. 2008.

[5] G. S. Parra-Dominguez, R. E. Sanchez-Yanez, and C. H. Garcia-Capulin, "Facial paralysis detection on images using key point analysis," *Appl. Sci.*, vol. 11, no. 5, p. 2435, Mar. 2021.

[6] G. F. Volk, R. A. Schaede, J. Thielker, L. Modersohn, O. Mothes, C. C. Nduka, J. M. Barth, J. Denzler, and O. G. Lichius, "Reliability of grading of facial palsy using a video tutorial with synchronous video recording," *Laryngoscope*, vol. 129, no. 10, pp. 2274–2279, Oct. 2019.

[7] J. G. Neely, N. G. Cherian, C. B. Dickerson, and J. M. Nedzelski, "Sunnybrook facial grading system: Reliability and criteria for grading," *Laryngoscope*, vol. 120, no. 5, pp. 1038–1045, May 2010.

[8] S. D. Reitzen, J. S. Babb, and A. K. Lalwani, "Significance and reliability of the House–Brackmann grading system for regional facial nerve function," *Otolaryngol.-Head Neck Surgery*, vol. 140, no. 2, pp. 154–158, Feb. 2009.

[9] F. T. Kayhan, D. Zurakowski, and S. D. Rauch, "Toronto facial grading system: Interobserver reliability," *Otolaryngol.-Head Neck Surgery*, vol. 122, no. 2, pp. 212–215, Feb. 2000.

[10] T. Li, J. Ren, and X. Peng, "Therapeutic observation on superficial needling with different frequencies for intractable facial paralysis," *J. Acupuncture Tuina Sci.*, vol. 17, no. 6, pp. 432–437, Dec. 2019.

[11] J. T. Vrabec, D. D. Backous, H. R. Djalilian, P. W. Gidley, J. P. Leonetti, S. J. Marzo, D. Morrison, M. Ng, M. J. Ramsey, B. M. Schaitkin, E. Smouha, E. H. Toh, M. K. Wax, R. A. Williamson, and E. O. Smith, "Facial nerve grading system 2.0," *Otolaryngol. Head Neck Surg.*, vol. 140, no. 4, pp. 445–450, Apr. 2009.

[12] G. J. Hsu, J.-H. Kang, and W.-F. Huang, "Deep hierarchical network with line segment learning for quantitative analysis of facial palsy," *IEEE Access*, vol. 7, pp. 4833–4842, 2019.

[13] Y. Liu, Z. Xu, L. Ding, J. Jia, and X. Wu, "Automatic assessment of facial paralysis based on facial landmarks," in *Proc. IEEE 2nd Int. Conf. Pattern Recognit. Mach. Learn. (PRML)*, Chengdu, China, Jul. 2021, pp. 162–167.

[14] X. Liu, Y. Xia, H. Yu, J. Dong, M. Jian, and T. D. Pham, "Region based parallel hierarchy convolutional neural network for automatic facial nerve paralysis evaluation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2325–2332, Oct. 2020.

[15] P. Xu, F. Xie, T. Su, Z. Wan, Z. Zhou, X. Xin, and Z. Guan, "Automatic evaluation of facial nerve paralysis by dual-path LSTM with deep differentiated network," *Neurocomputing*, vol. 388, pp. 70–77, May 2020.

[16] T. Wang, S. Zhang, L. Liu, G. Wu, and J. Dong, "Automatic facial paralysis evaluation augmented by a cascaded encoder network structure," *IEEE Access*, vol. 7, pp. 135621–135631, 2019.

[17] E. B. Prince, K. B. Martin, and D. Messinger, "Facial action coding system," in *Psychology, Computer Science*, 2015.

[18] K. Zhao, W.-S. Chu, F. De la Torre, J. F. Cohn, and H. Zhang, "Joint patch and multi-label learning for facial action unit and holistic expression recognition," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3931–3946, Aug. 2016.

[19] W. Li, F. Abtahi, and Z. Zhu, "Action unit detection with region adaptation, multi-labeling learning and optimal temporal fusing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6766–6775.

[20] B.-F. Wu, Y.-T. Wei, B.-J. Wu, and C.-H. Lin, "Contrastive feature learning and class-weighted loss for facial action unit detection," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Bari, Italy, Oct. 2019, pp. 2478–2483.

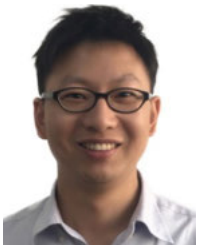
[21] Y. Xia, "Upper middle and lower region learning for facial action unit detection," 2020, *arXiv:2002.04023*.
[22] S. Wang, J. Yang, Z. Gao, and Q. Ji, "Feature and label relation modeling for multiple-facial action unit classification and intensity estimation," *Pattern Recognit.*, vol. 65, pp. 71–81, May 2017.
[23] I. Hupont and M. Chetouani, "Region-based facial representation for real-time action units intensity detection across datasets," *Pattern Anal. Appl.*, vol. 22, no. 2, pp. 477–489, May 2019.
[24] C. Wei, K. Lu, W. Gan, and J. Xue, "Spatiotemporal features and local relationship learning for facial action unit intensity regression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Anchorage, AK, USA, Sep. 2021, pp. 1109–1113.
[25] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System*, 2nd ed. Salt Lake City, UT, USA: Research Nexus eBook, 2002.
[26] T. Hadlock and N. Jowett, "Contemporary management of bell palsy," *Facial Plastic Surg.*, vol. 31, no. 2, pp. 093–102, May 2015.
[27] M. Mavadati, P. Sanger, and M. H. Mahoor, "Extended DISFA dataset: Investigating posed and spontaneous facial expressions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Las Vegas, NV, USA, Jun. 2016, pp. 1452–1459.
[28] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "DISFA: A spontaneous facial action intensity database," *IEEE Trans. Affect. Comput.*, vol. 4, no. 2, pp. 151–160, Apr. 2013.
[29] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard, "BP4D-spontaneous: A high-resolution spontaneous 3D dynamic facial expression database," *Image Vis. Comput.*, vol. 32, no. 10, pp. 692–706, Oct. 2014.
[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[32] R. Walecki, O. Rudovic, V. Pavlovic, and M. Pantic, "Copula ordinal regression for joint estimation of facial action unit intensity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 4902–4910.
[33] S. Kaltwang, S. Todorovic, and M. Pantic, "Doubly sparse relevance vector machine for continuous facial behavior estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1748–1761, Sep. 2016.
[34] R. Walecki, O. Rudovic, V. Pavlovic, B. Schuller, and M. Pantic, "Deep structured learning for facial expression intensity estimation," *Comput. Sci.*, pp. 3405–3414, Apr. 2017.
[35] Y. Zhang, R. Zhao, W. Dong, B.-G. Hu, and Q. Ji, "Bilateral ordinal relevance multi-instance regression for facial action unit intensity estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7034–7043.
[36] S. Wang, B. Pan, S. Wu, and Q. Ji, "Deep facial action unit recognition and intensity estimation from partially labelled data," *IEEE Trans. Affect. Comput.*, vol. 12, no. 4, pp. 1018–1030, Oct. 2021.
[37] G. M. Guanoluisa, J. A. Pilatasig, and V. H. Andaluz, "GY MEDIC: Analysis and rehabilitation system for patients with facial paralysis," in *Proc. Int. Symp. Integr. Uncertainty Knowl. Modelling Decision Making*. Cham, Switzerland: Springer, Cham, 2019, pp. 63–75.
[38] G. M. Guanoluisa, J. A. Pilatasig, L. A. Flores, and V. H. Andaluz, "GY MEDIC v2: Quantification of facial asymmetry in patients with automated Bell's palsy by AI," in *Proc. Int. Conf. Augmented Reality, Virtual Reality Comput. Graph.*, 2019, pp. 351–361.
[39] K. J. Kwon, J. H. Bang, S. H. Kim, S. G. Yeo, and J. Y. Byun, "Prognosis prediction changes based on the timing of electroneurography after facial paralysis," *Acta Oto-Laryngologica*, vol. 142, no. 2, pp. 213–219, Feb. 2022.
[40] P. Zhu, H. Wang, L. Zhang, and X. Jiang, "Deep learning-based surface nerve electromyography data of e-health electroacupuncture in treatment of peripheral facial paralysis," *Comput. Math. Methods Med.*, vol. 2022, May 2022, Art. no. 8436741.
[41] K. Machetanz, F. Grimm, R. Schäfer, L. Trakolis, H. Hurth, P. Haas, A. Gharabaghi, M. Tatagiba, and G. Naros, "Design and evaluation of a custom-made electromyographic biofeedback system for facial rehabilitation," *Frontiers Neurosci.*, vol. 16, Mar. 2022, Art. no. 66173.
[42] A. Raj, O. Mothes, S. Sickert, and G. F. Volk, "Automatic and objective facial palsy grading index prediction using deep feature regression," in *Medical Image Understanding and Analysis*, vol. 1248. Cham, Switzerland: Springer, pp. 253–266.
[43] C. Sforza, E. Ulaj, D. M. Gibelli, F. Allevi, V. Pucciarelli, F. Tarabbia, D. Ciprandi, G. Dell'Aversana Orabona, C. Dolci, and F. Biglioli, "Three-dimensional superimposition for patients with facial palsy: An innovative method for assessing the success of facial reanimation procedures," *Brit. J. Oral Maxillofacial Surg.*, vol. 56, no. 1, pp. 3–7, Jan. 2018.
[44] X. Ge, J. M. Jose, P. Wang, A. Iyer, X. Liu, and H. Han, "ALGRNet: Multi-relational adaptive facial action unit modelling for face representation and relevant recognitions," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 5, no. 4, pp. 566–578, Oct. 2023.
[45] X. Ge, J. M. Jose, P. Wang, A. Iyer, X. Liu, and H. Han, "Automatic facial paralysis estimation with facial action units," 2022, *arXiv:2203.01800*.
[46] Y. Zhang, H. Jiang, B. Wu, Y. Fan, and Q. Ji, "Context-aware feature and label fusion for facial action unit intensity estimation with partially labeled data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 733–742.
[47] H. Chen, D. Jiang, Y. Zhao, X. Wei, K. Lu, and H. Sahli, "Region attentive action unit intensity estimation with uncertainty weighted multi-task learning," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 2033–2047, Jul./Sep. 2023.
[48] D. Seuss, T. Hassan, A. Dieckmann, M. Unfried, K. R. R. Scherer, M. Mortillaro, and J.-U. Garbas, "Automatic estimation of action unit intensities and inference of emotional appraisals," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1188–1200, Apr./Jun. 2023.
[49] I. Ntinou, E. Sanchez, A. Bulat, M. Valstar, and G. Tzimiropoulos, "A transfer learning approach to heatmap regression for action unit intensity estimation," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 436–450, Jan. 2023.
[50] Y. Fan, J. C. K. Lam, and V. O. Li, "Distilling region-wise and channel-wise deep structural facial relationships for FAU (DSR-FAU) intensity estimation," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 986–997, Apr./Jun. 2023.
[51] J. Ma, X. Li, Y. Ren, R. Yang, and Q. Zhao, "Landmark-based facial feature construction and action unit intensity prediction," *Math. Problems Eng.*, vol. 2021, pp. 1–12, Mar. 2021.
[52] Z. Shao, Z. Liu, J. Cai, and L. Ma, "JÂA-Net: Joint facial action unit detection and face alignment via adaptive attention," *Int. J. Comput. Vis.*, vol. 129, no. 2, pp. 321–340, Feb. 2021.
[53] S. R. Gogu and S. R. Sathe, "AutoFPR: An efficient automatic approach for facial paralysis recognition using facial features," *Int. J. Artif. Intell. Tools*, vol. 32, no. 2, Mar. 2023, Art. no. 2340005.
[54] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.
[55] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 1314–1324.
[56] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
[57] F. Boutros, V. Struc, J. Fierrez, and N. Damer, "Synthetic data for face recognition: Current state and future prospects," *Image Vis. Comput.*, vol. 135, Jul. 2023, Art. no. 104688.
[58] S. H. Abdulhussain, B. M. Mahmmod, A. AlGhadhban, and J. Flusser, "Face recognition algorithm based on fast computation of orthogonal moments," *Mathematics*, vol. 10, no. 15, p. 2721, Aug. 2022.
[59] H. Ma, K. Xu, X. Jiang, Z. Zhao, and T. Sun, "Transferable black-box attack against face recognition with spatial mutable adversarial patch," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 5636–5650, 2023.
[60] V. Ravi, H. Narasimhan, C. Chakraborty, and T. D. Pham, "Deep learning-based meta-classifier approach for COVID-19 classification using CT scan and chest X-ray images," *Multimedia Syst.*, vol. 28, no. 4, pp. 1401–1415, Aug. 2022.
[61] Y. Wu, H. Guo, C. Chakraborty, M. Khosravi, S. Berretti, and S. Wan, "Edge computing driven low-light image dynamic enhancement for object detection," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 5, pp. 3086–3098, Sep./Oct. 2023.
[62] A. Kishor, C. Chakraborty, and W. Jeberson, "Reinforcement learning for medical information processing over heterogeneous networks," *Multimedia Tools Appl.*, vol. 80, pp. 23983–24004, Mar. 2021.
[63] K. Boochoon, A. Mottaghi, A. Aziz, and J.-P. Pepper, "Deep learning for the assessment of facial nerve palsy: Opportunities and challenges," *Facial Plastic Surg.*, vol. 39, no. 5, pp. 508–511, Oct. 2023.

HEQUN NIU is currently pursuing the master's degree in electronic information with the College of Automation, Qingdao University. Her main research interests include image/video processing, face analysis, pattern recognition, computer vision, intelligent healthcare, and affective computing.

JIPENG LIU is currently pursuing the bachelor's degree in software engineering with the College of Computer Science and Technology, Qingdao University. His main research interests include behavior detection and recognition in videos, abnormal facial behavior detection, and intelligent healthcare.

XUHUI SUN is currently pursuing the master's degree in electronic information with the College of Automation, Qingdao University. His research interests include biomedical signal decoding, human–computer interaction, and the applications of rehabilitation robotic systems.

XIANGTAO ZHAO received the bachelor's degree in 2021. He is currently pursuing the master's degree with the Institution for Future, Qingdao University, researching image processing and behavior recognition related work.



YINHUA LIU received the doctor's degree in mechanical engineering from Sophia University, Japan, in March 2013. He joined Qingdao University, in August 2017. He has participated in a number of key research and development projects of the Ministry of Science and Technology. He has published more than ten papers in international and domestic journals and conferences. He received or accepted more than 30 patents in the State Intellectual Property Office. In the field of offshore photovoltaic, intelligent parking, smart city, and other fields, we have developed a number of industry-university-research innovation projects, and won a number of innovation and entrepreneurship awards. We have accumulated rich project resources and technical teams, among which more than 30 team research and development personnel are specialized in software, machinery, electromechanical, embedded, design, and other fields, with interdisciplinary ability and experience. He has a number of product research and development experience and project management experience. His research interests include intelligent wearables, offshore photovoltaic, deep learning, and artificial intelligence. He won the Outstanding Mentor of RoboMaster University Single Competition, the First Prize of Shandong New Generation Mobile Internet Innovation Application Skills Competition, and many other awards.