AFLFP: A Database with Annotated Facial Landmarks for Facial Palsy

Yifan Xia, Charles Nduka, Ruben Yap Kannan, Elena Pescarini, Juan Berner, and Hui Yu, *Senior Member*, *IEEE*

***Abstract*—Facial landmark detection is a crucial step for the task of computer-aided facial palsy diagnosis, which enables to focus on the affected facial regions for learning asymmetry, shape, and texture features of facial palsy. However, it is still very challenging to accurately detect salient landmarks on facial palsy images due to the unavailability of sufficient training databases providing annotated facial landmark images of facial palsy. To this end, we present a database in this paper named Annotated Facial Landmarks for Facial Palsy (AFLFP). AFLFP is a diverse, and reliable database that contains facial images with 16-class facial expressions of asymmetric facial expressions from 88 subjects. Each facial image is independently and manually annotated with 68 facial landmarks. This database is the first public manually annotated facial landmark database for facial palsy so far. Furthermore, to establish the benchmark results for the proposed database, we propose a deep neural network baseline with a two-stage cascaded fully convolutional network (FCN), which can detect facial landmarks in facial palsy from coarse to fine. The comprehensive experiments show that the proposed method performs better than the mainstream methods of machine learning and deep learning. And we have also compared the performance using normal faces and palsy faces respectively as the training data. The comparison results show that there are significant differences between them in terms of facial landmark detection, which further confirms the necessity to develop a facial landmark database specifically for facial palsy.**

***Index Terms*—Facial palsy, facial landmark detection, facial landmark database, deep learning.**

I. Introduction

FACIAL palsy, such as Bell's palsy, is a neuromuscular disorder disease which affects 11 to 40 persons per 100,000 worldwide and usually occurs between the ages of 15 and 50 [1]. The peak incidence of facial palsy more frequently appears in the spring and fall than other times of the year [2]. Almost a third of Bell's palsy victims have chronic symptoms of facial weakness, disability of facial expressions and movements [3], [4]. During the long-term process of rehabilitation, most facial palsy patients not only suffer physical pain but also have psychosocial problems, which seriously erodes the patient's health and the quality of daily life [5], [6]. Developing new treatments relies on having accurate, objective, and validated assessment tools. Thus, to alleviate the suffering of facial palsy patients, the first step is to make an accurate and effective assessment of facial palsy severity.

Traditional facial palsy diagnosis mainly relies on clinicians' subjective judgment. Clinicians make the visual inspection to get a comprehensive judgment after instructing the patient to perform certain facial expressions, which is based on various scoring standards and individual clinical experience [8]. Nevertheless, the conventional clinical method has the disadvantages of labor-intensive and subjective [7]. Hence, computer-aided diagnosis approaches for healthcare [4], [19], [49]-[54], [63] have been a hot research topic for various fields such as artificial intelligence, computer vision and human-computer interaction, that can potentially overcome the limitations of the conventional clinical method. Compared with conventional methods, computer-aided methods rely on computerized decision support to ensure that diagnosis is objective [9], [10]. During the process of rehabilitation, the objective diagnosis is beneficial for facial palsy patients to conduct assessments by themselves and enhance their confidence. Recent advances in computer-aided diagnosis methods of facial palsy have shown impressive performance and drawn prevalent attention [20]-[24], [31], [32], [43]. These approaches directly analyzing standardized 2D photography or videography of facial palsy captured from a camera through techniques of computer vision and imaging processing, which are highly efficient and cost effective. In particular, this makes it possible for many patients who have difficulties in visiting therapists to get real-time feedback whilst conducting therapy at home.

Currently, facial landmark detection plays a vital role in computer-assisted facial palsy diagnostic tasks since most existing methods depend on accurate facial landmarks to learn asymmetry, shape and texture features from facial images of facial palsy patients [14], [15]. In particular, accurate facial landmarks can enable the algorithm to focus on the affected facial regions and thus reduce the impact of the redundant facial

This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) (EP/N025849/1). (*Corresponding author: Hui Yu.*)

Yifan Xia and Hui Yu are with the School of Creative Technologies, University of Portsmouth, Portsmouth PO1 2DJ, U.K. (e-mail: yifan.xia@myport.ac.uk; hui.yu@port.ac.uk).

Charles Nduka and Ruben Yap Kannan are with the Queen Victoria Hospital, East Grinstead RH19 3DZ, U.K. (e-mail: charles.nduka@nhs.net; ruben.kannan@nhs.net). Elena Pescarini is with San Bortolo Hospital, 36100 Vicenza, Italy (e-mail: elena.pescarini@gmail.com). Juan Enrique Berner is with the Plastic Surgery Department, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Tyne NE7 7DN, U.K., and also with Kellogg College, University of Oxford, Oxford OX1 2JD, U.K. (e-mail: jeberner@gmail.com).

Fig. 1. Examples of facial images with annotated 68 facial landmarks from AFLFP.

image information to further improve the performance [7], [8], [9], [10]. The reason for this is because paralysis features usually are in specific facial regions such as eye and mouth rather than being evenly distributed across the whole face [17], [18]. For example, facial palsy patients usually have symptoms of eye narrowing and lip contour deformation in wrinkling nose expression. Thus, asymmetry symptoms of facial palsy can be easily inferred from different facial regions.

The emerging technologies in machine learning, computer graphics and deep learning which significantly improve the performance of various facial perception and computer vision tasks [42], [45],[48], [57]-[61], [64], have been advanced facial landmark detection to a novel state-of-the-art [44], [46]. Deep learning-based methods usually rely on a large dataset with elaborately annotated facial landmarks to enable better performance. However, there are still challenges for facial landmark detection on facial palsy faces. First, these existing facial landmark databases are mainly based on normal faces instead of palsy faces. Second, due to the nature of facial palsy and the complexity of annotation, it is not an easy task to collect a large annotated facial landmark database of facial palsy, which requires specific expertise and a long time to complete. Therefore, to facilitate accurate facial landmarks detection in facial palsy, facial landmark databases specifically for facial palsy with large valid annotated data are urgently needed.

Though 3D/4D technologies are available [43], [55], [57], images and videos are still the most widely used data. We thus have established a large-scale video-based facial landmark database for facial palsy, Annotated Facial Landmarks for Facial Palsy (AFLFP). AFLFP contains facial images with 16-class asymmetric facial expressions from 88 subjects. To ensure reliability, more than twenty well-trained annotators independently and manually annotated these facial images with 68 facial landmarks. And each annotated image is checked by different annotators. Fig. 1 shows examples of facial images with annotated facial landmarks from the AFLFP database. Considering the lack of facial landmark database for facial palsy, AFLFP as a publicly available database will lead to further advances in the development of automated methods for facial landmark detection and other analysis in facial palsy. In summary, the contributions of this paper are summarized as follows:

1) We present a manually annotated database AFLFP for facial landmark detection in facial palsy. This database is a large-scale, diverse, and reliable database which contains facial images with 16-class facial expressions from 88 subjects including facial palsy patients and facial palsy clinicians. To the best of our knowledge, AFLFP is the first public facial landmark database for facial palsy so far.

2) A two-stage cascaded fully convolutional network (FCN) as the deep neural network baseline is proposed, which can detect facial landmarks in facial palsy from coarse to fine. We have compared the proposed method with the mainstream methods of machine learning and deep learning to establish the baseline results of the proposed database.

3) We have explored the differences in facial landmark detection between normal faces and palsy faces. The experimental results demonstrate that creating a facial landmark database specifically for facial palsy is very necessary for further performance improvement of facial landmark detection for facial palsy.

Here is the remainder structure of this paper. In Section II, we provide a brief description of the related work about facial landmark detection in facial palsy. Section III introduces the details about the proposed AFLFP database. Section IV describes the proposed two-stage cascaded FCN in detail. In Section V, we show the results of the experimental evaluation on the proposed AFLFP database. Finally, we conclude this paper and give future works in Section VI.

II. Related Work

The research on computer-aided facial palsy diagnosis has witnessed promising progress in recent years, which makes it possible to assist or replace some work of clinicians with a machine to make a diagnosis [20]-[24], [31], [32]. The

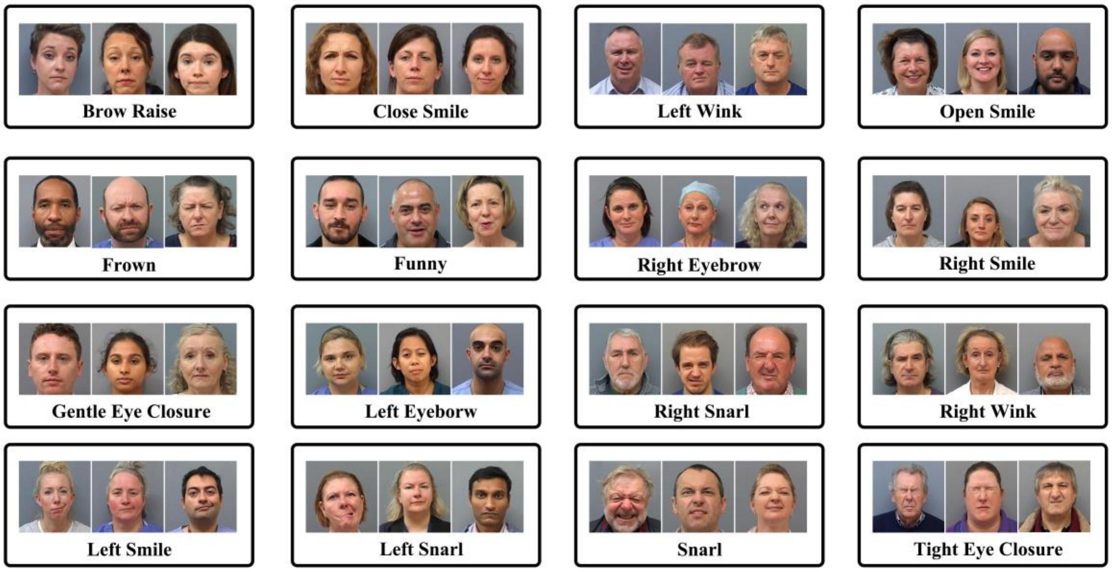

Fig. 2. Examples of 16-class facial expressions from AFLFP.

computer-aided method can avoid subjective bias from clinicians in diagnosis, which opens a promising direction for the development of facial palsy diagnosis. The development of hardware technologies provides opportunities for accessing devices with embedded cameras to capture images/videos. This has made existing methods of facial palsy diagnosis mainly focus on analyzing images/videos captured from patients, which is cost effective and convenient to use.

Facial landmark detection is essential to the computer-aided facial palsy diagnosis since the severity of facial palsy is usually quantified based on detected facial landmarks [12], [13], [16], [25]-[30]. Most existing works quantifying facial palsy are built on top of facial landmarks to learn asymmetry, shape and texture features from facial images of patients captured from an ordinary camera. Burres [11] evaluated the facial motor function of the patient's face at rest or specified facial expressions utilizing 10 facial landmarks to calculate 13 distances, which was a pioneer work in this field. However, the distance was determined by utilizing a hand-held caliper to gauge manually marked facial landmarks, which is an inefficient process. The development of computer technology allows landmark detection and distance calculation to be carried out in a fully automated way, which has promising applicability in this field. Wang *et al.* [14] utilized facial asymmetry features to evaluate the degree of facial palsy. They first trained an ASM model to detect 68 facial landmarks of the facial palsy patient. Then, they divided the face into eight regions according to the detected landmarks and extracted facial asymmetry features from these regions. Kim *et al.* [15] proposed a diagnosis system for facial palsy based on the smartphone that utilized the iPAR-CLR method to detect facial landmarks. Based on the detected facial landmarks, they utilized the displacement of the eyebrows and mouth regions to compute the level of facial asymmetric.

Unlike the methods mentioned above that learn asymmetry features based on facial landmarks for facial palsy diagnosis, facial landmarks can also be useful when trying to focus on the affected facial regions and then reduce the impact of redundant information to improve the performance. Deformity symptoms usually appear on some crucial regions of the facial palsy patient's face, such as inability to close the eye fully and the corner of the mouth pulls down [17], [18]. Guo *et al.* [9] proved that it had obvious correlates between features extracted from the mouth region and facial palsy. Barbosa *et al.* [10] utilized the region-based feature for the classification of facial palsy. Thus, the discriminative features of facial palsy can be extracted from the crucial regions. Hsu *et al.* [8] designed a deep hierarchical network to quantitatively analyze facial palsy. This network first detected facial landmarks and located the palsy regions. Then, it quantitatively analyzed the level of the facial palsy according to detected facial landmarks. Recently, Liu *et al.* [7] utilized the facial landmarks detected from IntraFace [33] to divide the whole face into two palsy regions and designed a hierarchy network to evaluate facial palsy. This network could extract facial paralysis features from these palsy regions and significantly reduce redundant information from unrepresentative regions on feature learning.

Although current facial landmark detection methods trained on normal facial databases can work somehow on the face of facial palsy patients, the detected landmarks are not accurate. The main reason is that the performance is mitigated by the training data and the quality and diversity of databases directly determines the performance of facial landmark detection to a certain extent. Currently, most existing facial landmark databases are mainly collected from healthy subjects instead of facial palsy patients, such as LFPW [34], HELEN [35], AFW [36] and 300-W [37]. Although there are some facial palsy databases such as YFP [8] and MEEI [65], they don't provide the annotations of facial landmarks. Unlike the normal facial databases mentioned above, it is difficult to collect a large

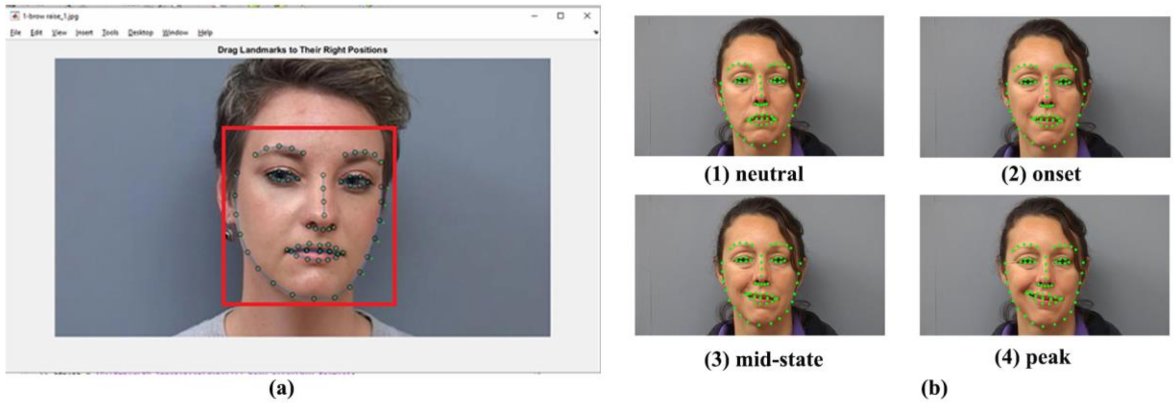

Fig. 3. Database annotations. (a): A screenshot of the software used to annotate facial landmarks. Only one detected face in each image is detected (shown in the red bounding box) and annotated (green circles) by the software. If some landmarks need to be adjusted, annotators drag those landmarks (black circles) directly to their right positions manually. (b): Sample face of four keyframes (1) neutral, 2) onset, 3) a mid-state between onset and peak, 4) peak) with 68 landmarks automatically annotated using the Supervised Descent Method (SDM) [39] method and manually corrected by our annotators.

annotated facial landmark database of facial palsy due to the nature of facial palsy. And there are no publicly available databases providing facial palsy images annotated with accurate landmarks. To fill this research gap, we present a facial landmark database specifically for facial palsy, called Annotated Facial Landmarks for Facial Palsy (AFLFP). AFLFP is a diverse, and reliable database that contains annotated facial landmarks of facial images with 16-class facial expressions of asymmetric facial expressions. To the best of our knowledge, the proposed database is the first public facial landmark database specially for facial palsy so far. In addition, to address the detection of facial landmarks in facial palsy, we propose a deep neural network baseline with a two-stage cascaded fully convolutional network (FCN), which can detect facial landmarks in facial palsy from coarse to fine. And we compared the proposed deep neural network baseline with the mainstream methods of machine learning and deep learning to establish the baseline results of the proposed database. We have also explored the differences in facial landmark detection between normal faces and palsy faces.

III. AFLFP Database

AFLFP database is a large-scale, diverse, and reliable facial landmark database for facial palsy. This database is created and manually annotated by volunteers and facial palsy clinicians, which includes facial images of facial palsy patients. In this section, we introduce the details of creating this database.

A. *Experiment Setup*

In the experiment, each participant was asked to perform 16-class facial expressions including Brow Raise, Close Smile, Frown, Funny, Gentle Eye Closure, Left Eyebrow, Left Smile, Left Snarl, Left Wink, Open Smile, Right Eyebrow, Right Smile, Right Snarl, Right Wink, Snarl and Tight Eye Closure. Fig. 2 exhibits specific samples of 16-class facial expressions. The experimental environment is an illumination-controlled room. A fixed camera at the front view of participant was used to continuously record the participant faces during the process of performing specified expression. The fixed camera with a resolution of 1920×1080 pixels collected data of participant faces at a frame rate of 30 fps.

B. *Participants and Ethics Statement*

We recruited 88 facial palsy participants (38 males and 50 females) in total to participate in the study. All participants in this experiment have signed the written consent form (local research ethics committee approval number 139558). Before the experiment began, each participant was instructed with an experimental tutorial. And the whole experimental procedure was also agreed by all the participants. The participants were allowed to withdraw from the experiment once they felt any uncomfortable. We only collected video capture from participants who agreed to use their video images for research purposes. Some participants also gave consent to displaying their facial images in research articles and reports.

C. *Creating AFLFP*

1) Data selection: Two research experts were recruited in this experiment to review the video clips collected from 88 participants. To ensure the availability of video clips, they were required to conduct the preliminary screening. They selected 16 video clips for each participant and each video clip lasted 10-30 seconds. The selection criteria for video clips are as follows: 1) the participant's face must be clearly visible and without occlusions in each video clip; 2) the complete process of performing specific expression segments should be included in each video clip, and 3) each video clip contains the content of only one expression out of 16-class facial expressions. Another two research experts with extensive experience in facial palsy analysis evaluated each of the selected video clips. The consensus of the two experts determines the final choice of video clips.

2) Database Annotation: After data selection, we have obtained 16 expression videos for each participant. To reduce the complexity of annotation, we only selected four key frames in each video representing the key states within the progression of one facial expression, which are neutral, onset, a mid-state

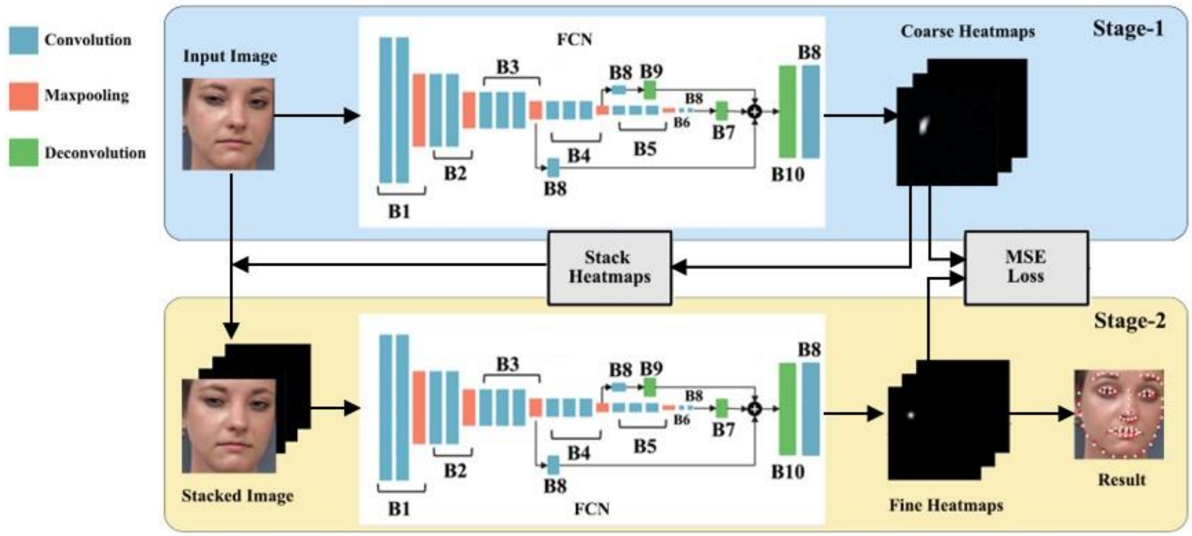

Fig. 4. The flowchart of the proposed two-stage cascaded FCN used for facial landmark detection. The blocks B1-b10 are defined in Table 1. Both stages have the similar FCN structure, but they are assigned with different learning tasks. During the training, these two stages are supervised by MSE loss at the same time.

between onset and peak, peak for each of the subject's expression video. This database totally includes about 5,632 facial images, 1,408 samples for each key state and 64 samples for each subject. Fig. 3(b) shows samples of four key frames selected from expression videos. We employed 22 annotators including research students and university researchers with expertise in image processing. Each labeler was randomly assigned the same number of images. They were given a tutorial on facial landmark detection before starting annotating to ensure the quality of the annotation work. The annotators were also given some demonstration facial images with the perfect landmark locations as examples. They were instructed to annotate facial images using the landmark configuration of 300-W [37] (68-point markup). As it is very time-consuming and hard to annotate these facial images, we developed a software for facial landmark annotation based on Matlab to facilitate the process. The software first utilizes Viola-Jones face detector [38] to locate the face region and automatically detects the rough facial landmarks using the SDM method [39]. Then, annotators independently and manually corrected imprecise or missed facial landmarks. And each annotated image is checked by different annotators, who were given enough time to ensure reliability. Fig. 3(a) shows a screenshot of the user interface of the annotation software, and some annotation results of sample faces are shown in Fig. 3(b).

IV. Proposed Method

In this section, we provide the details of the proposed deep neural network baseline for landmark detection of facial palsy images. The main purpose of the proposed deep network baseline is to establish baseline results for the proposed database, which will provide a benchmark for researchers to validate their algorithms for landmark detection of facial palsy images. Our approach for multiple landmark detection regresses heatmaps directly from input images during training and predicts the heatmap for each facial landmark. The location with the maximum response value on the predicted heatmap is regarded as the landmark prediction result. Aiming for a low landmark detection error, our proposed architecture has a cascade structure as shown in Fig. 4. Our proposed cascade structure includes two connected stages, which can regress a set of heatmaps of facial landmarks from coarse to fine.

In the facial landmark detection task, landmark detection methods based on deep neural networks (DNNs) have shown impressive performance recently, which is to predict the location of facial landmarks via regression. Existing works for deep learning-based landmark detection methods can be roughly divided into two categories: direct regression and heatmap regression. The direct regression method takes facial images as input and treats landmark coordinates as the optimization target, which uses the network to directly regress landmark coordinates and predicts a set of facial landmark coordinate values. However, the direct regression method usually requires a fully connected layer with many network parameters to model the nonlinearity and it is difficult to learn the image to coordinate mapping. Moreover, the important spatial information may be lost to some extent due to the fully connected layer used in the networks. Thus, under the influence of the interference factors such as large poses and occlusion, the prediction of some landmarks by direct regression methods might lead to low detection accuracy or even failure.

The heatmap regression method outperforms the direct regression method, which has good spatial perception ability and makes the network more accurate and robust. This kind of method regresses the heatmap of each landmark and locates the landmark coordinate according to the maximum value position on the heatmap. By enabling an image-to-image mapping, facial landmark detection can be can be treated as a subproblem of image segmentation [40], [41], [47], [62]. Therefore, existing heatmap regression methods usually utilize the structure of the fully convolutional network (FCN) which can preserve spatial

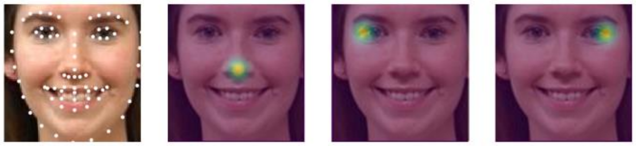

Fig. 5. The sample heatmaps generated using Gaussian kernel. The first left image shows the ground truth facial landmarks. The others show some sample generated heatmaps of facial landmarks (with the original image behind). From left to right: nose ridge tip, right eye corner outer and left eye corner outer.

TABLE 1
BLOCK SPECIFICATION FOR THE PROPOSED TWO-STAGE CASCADED FCN. NOTATIONS (CHANNELS, KERNEL, STRIDE) AND (KERNEL, STRIDE) ARE USED TO DEFINE THE CONV AND POOLING LAYERS.

1	B2	B3	B4	B5
x conv layer 4,3x3,1x1) laxpooling x2, 2x2)	2x conv layer (128,3x3,1x1) Maxpooling (2x2, 2x2)	3x conv layer (256,3x3,1x1) Maxpooling (2x2, 2x2)	3x conv layer (512,3x3,1x1) Maxpooling (2x2, 2x2)	3x conv laye (512,3x3,1x Maxpooling (2x2, 2x2)
6	B7	B8	B9	B10
onv layer 4096,3x3,1x1) onv layer 4096,1x1,1x1)	deconv layer (68, 4x4, 4x4)	conv layer (68, 1x1,1x1)	deconv layer (68, 2x2, 2x2)	deconv layer (68,8x8, 8x8

details and capture the precise location of each landmark from feature maps. In this paper, we propose a coarse-to-fine cascade regression method to regress a set of heatmaps in which the location of each pixel stores a value representing the probability of that location is the correct landmark. The overall network framework is shown in Fig. 4. To establish an image-to-image mapping, we transform the coordinates of 68 facial landmarks to the 68-channel heatmaps. As shown in Fig. 5, we use a channel heatmap to represent each facial landmark. Each heatmap is generated from a 2D Gaussian distribution which is normalized to the range from 0 to 1, and the center of the distribution is the facial landmark. Given a target landmark with a coordinate (x_l, y_l), the heatmap H is thus given by,

$$H(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(x-x_l)^2+(y-y_l)^2}{2\sigma^2}\right) \qquad (1)$$

where $H(x,y)$ is the corresponding pixel value of the heatmap at position (x,y) of the image, which indicates the probability of this location being the landmark; σ represents the standard deviation which is the peak width of the distribution in the heatmap image for the landmark. Thus, heatmap pixels that closer to the target coordinate have higher values, but the values smoothly and rapidly decrease for pixels further from the target coordinate.

According to the structure of the VGG-FCN network [40], we design a two-stage cascade network architecture which uses a coarse-to-fine strategy to regress 68-channel heatmaps of landmarks directly from input images (see Fig. 4). Both stages have a similar FCN structure with skip connections, but they are assigned different learning tasks. We set the parameters of layers in each stage such as filter size, kernel size and stride following those in [40]. The parameter details of the network architecture are illustrated in Table 1.

In our design, the structure of FCN in Stage 1 takes the facial images as input to produce coarse 68-channel heatmaps H_S1 of the landmarks. After that, an additional FCN in Stage 2 is used to refine the coarse 68-channel heatmaps from Stage 1 for further improving accuracy and robustness. As shown in Fig. 4, the 68-channel heatmaps from Stage 1 and the input image are stacked as the input of Stage 2. Given a set of input images and corresponding target heatmaps, these two stages are trained at the same time during the training process. We take generated target heatmaps from a 2D Gaussian distribution as ground truth to supervise the predicted heatmaps from both Stage 1 and Stage 2. Therefore, the final loss function of the proposed network is the sum of the MSE loss functions of the two stages:

$$loss = \frac{1}{N}\sum_{n=1}^{N}\|h_n^{s1} - g_n\|^2 + \frac{1}{N}\sum_{n=1}^{N}\|h_n^{s2} - g_n\|^2 \qquad (2)$$

where h_n^{s1} and h_n^{s2} represent the heatmap of n-th landmark predicted by Stage 1 and Stage 2 respectively; g_n refers to the ground truth heatmap of n-th landmark; N is the total number of target facial landmarks. To evaluate the performance of the facial landmark detection, the fine heatmaps from Stage 2 must be converted to normal landmark coordinates. Generally, existing methods usually treat the pixel coordinates with the largest responses in the heatmap as the target coordinates. To get more accurate results, we use the average result of several largest responses of the predicted heatmaps to replace the largest one as the final facial landmark coordinates.

V. EXPERIMENTAL EVALUATION

The baseline results on the AFLFP database are established in this section. Since the technologies of deep neural networks (DNNs) have achieved impressive performance in various computer vision tasks, we present a deep neural network baseline with a cascaded FCN that detects facial landmarks for facial palsy. In the following subsections, we first describe the experimental settings including some implementation details and evaluation metrics. We then conduct baseline experiments on the AFLFP database and compare the proposed baseline with the mainstream methods of machine learning and deep learning. Finally, the differences in facial landmark detection between normal faces and palsy faces are compared to demonstrate the necessity of creating a facial landmark database specifically for facial palsy.

A. *Implementation Details*

AFLFP contains 5,632 facial images with annotated facial landmarks collected from 88 facial palsy patients and facial palsy clinicians. To evaluate the performance of the facial landmark detection task, 10-fold cross-validation method was adopted in our experiment. We divided the 88 subjects of the dataset into 10 subsets. 90% of the 88 subjects (80 subjects) were used as the training set, and 10% of the 88 subjects (8 subjects) were selected as the test set. Facial images of different subjects were not mixed to keep the experiment as subject

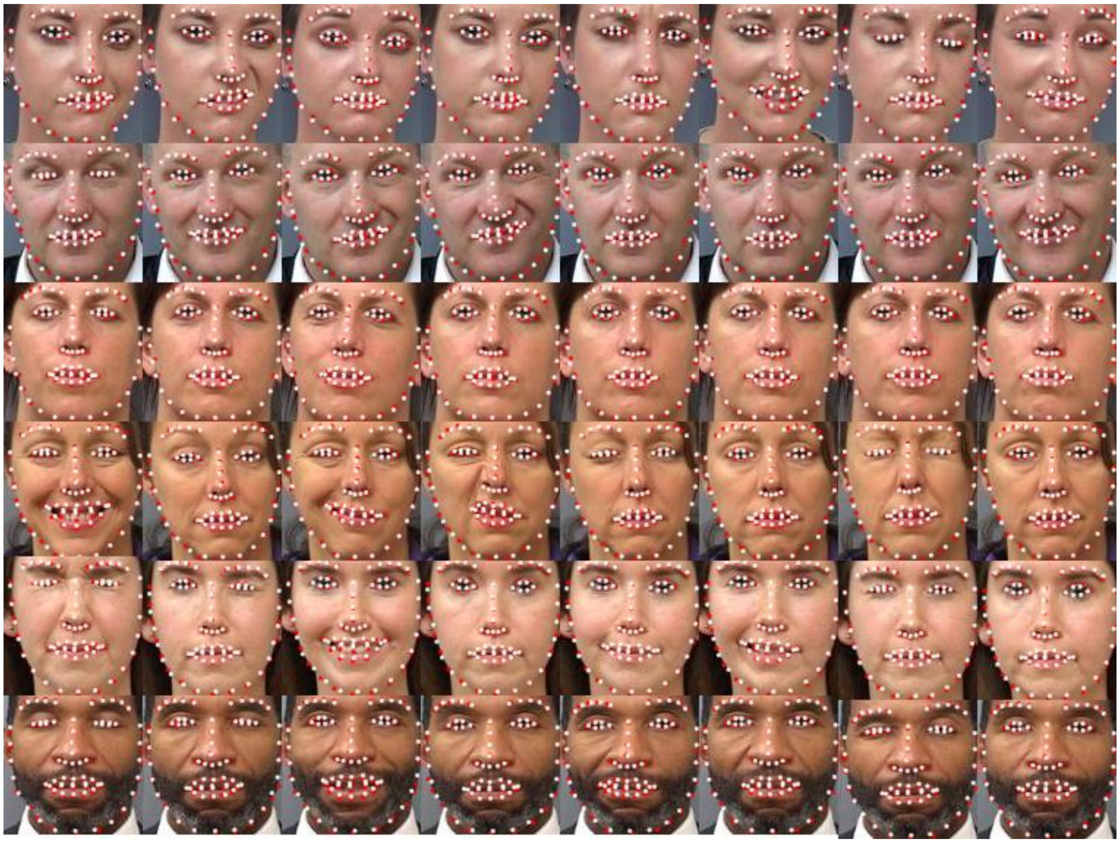

Fig. 6. Example results in each row are detection results of facial landmarks for the samples in the AFLFP database, where 68 landmarks were employed for evaluation. Red: ground truth from. White: predictions of the proposed method.

independent. During the stage of preprocessing, the face area of each image was first detected by the Viola-Jones face detector [38] and cropped according to the detected bounding box. Then, all the cropped images were resized to 96×96px and transformed to grayscale, which can maintain stable performance and improve efficiency during training. The ground truth heatmap of each facial landmark is generated by using the 2D Gaussian distribution (the standard deviation is set to 5) with a size of 96×96 px.

We implemented the proposed cascaded FCN illustrated in Fig. 4 using the TensorFlow framework. The proposed network was trained through an 8 GB NVIDIA GeForce GTX 1080 GPU on a desktop PC with a 4.20GHz Intel Core i7 processor, 16 GB of RAM memory. Adam optimizer with β_1 = 0.9, β_2 = 0.999 was utilized to optimize the whole network. The learning rate was initialized as 0.0001 and then was reduced by the step decay policy which divided the learning rate by 10 after each 30 epochs. We set the batch size as 32. To evaluate the performance of the proposed cascaded FCN architecture, final facial landmark coordinates are obtained as the average result of the largest 25 responses of the predicted heatmaps instead of the largest one.

B. *Evaluation Metric*

During the evaluation stage, the metric of the normalized mean error (NME) which is commonly used in the evaluation of the facial landmark detection task was used to evaluate the proposed method. The NME can be formulated as:

$$NME = \frac{1}{N}\sum_{n=1}^{N}\frac{\|y_n - \hat{y}_n\|_2}{\sqrt{w_{bbox} * h_{bbox}}} \qquad (3)$$

where N represents the number of images in the database; y_n is the ground truth facial landmark and $\hat{y}_n$ denotes the predicted facial landmark; w_{bbox} and h_{bbox} are the width and height of the ground truth bounding box respectively; $\|\cdot\|_2$ represents the L_2 norm of the matrix.

C. *Baseline Experiment*

To evaluate our proposed cascaded FCN on the AFLFP database, 16-class facial expressions were used including Brow Raise (BR), Close Smile (CS), Frown (FR), Funny (FN), Gentle Eye Closure (GEC), Left Eyebrow (LE), Left Smile (LSM), Left Snarl (LSN), Left Wink (LW), Open Smile (OS), Right Eyebrow (RE), Right Smile (RSM), Right Snarl (RSN), Right Wink (RW), Snarl (SN) and Tight Eye Closure (TEC). Fig. 6 shows the qualitative results of the proposed method on the AFLFP database. The white points represent the detected facial landmarks by the proposed method and the red points denote the ground truth provided by the AFLFP database. From Fig. 6, we can find that our predictions are very close to the ground truth.

For quantitative results, we analyzed both Mean and standard deviation (StD) values of NME to evaluate the performance of the proposed method on the AFLFP database (see Table 2). The

TABLE 2
COMPARISON OF FACIAL LANDMARK DETECTION ON 16-CLASS FACIAL EXPRESSIONS FROM AFLFP USING MEAN AND STD OF NME (%). THE HIGHEST RESULTS ARE HIGHLIGHTED IN BOLD.

Method	BR		CS		FR		FN		GEC		LE		LSM		LSN	
	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD
Emotrics	14.3	2.8	14.7	3.7	14.9	3.1	15.7	7.5	15.1	3.4	15.5	2.9	15.1	2.9	15.2	3.1
SDM	13.1	3.1	14.1	3.5	13.5	3.2	16.3	7.3	13.2	2.4	13.8	2.8	14.4	3.3	14.2	3.8
VGGNet	11.9	2.7	12.5	3.7	11.8	3.2	14.9	7.6	11.7	2.7	11.5	3.0	12.3	2.4	12.2	2.7
FCN	11.7	2.4	11.8	2.6	11.7	2.4	13.4	3.4	11.9	2.5	11.3	2.0	12.0	2.5	11.9	2.2
Cascaded FCN	**11.5**	**1.8**	**11.1**	**2.0**	**11.1**	**2.1**	**12.6**	**3.0**	**11.4**	**2.0**	**10.8**	**1.9**	**11.9**	**2.4**	**11.5**	**2.1**

LW		OS		RE		RSM		RSN		RW		SN		TEC		Avg	
Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD
14.7	3.9	15.4	3.3	15.1	3.5	15.4	3.8	15.4	2.8	14.6	3.4	15.7	4.4	17.0	5.1	15.2	3.7
14.1	3.2	14.5	3.0	13.4	3.2	14.5	3.0	14.2	2.8	14.1	3.7	14.7	4.0	14.3	4.3	14.1	3.8
11.8	2.5	12.4	2.7	12.2	3.6	13.4	3.7	12.2	3.0	12.0	2.5	12.8	4.2	13.1	4.0	12.4	3.7
11.9	2.2	12.2	2.0	11.9	2.1	12.9	2.7	12.1	2.6	11.9	2.3	12.8	3.7	12.2	2.6	12.1	2.5
11.4	**2.1**	**12.0**	**1.7**	**11.2**	**2.0**	**12.4**	**2.6**	**11.6**	**2.0**	**11.5**	**2.2**	**12.0**	**2.9**	**11.9**	**2.5**	**11.6**	**2.3**

better performance of facial landmark detection is, the smaller Mean and StD of NME are. For each category of facial expressions by using our method, the facial expression of Left Eyebrow has the lowest Mean of NME which is 10.8%. The Mean of NME of Funny can reach 12.6% which is the hardest one among all facial expressions. Moreover, the StD of NME of Open Smile is about 1.7% which is the lowest among the 16-class facial expressions. The expression of Funny has the highest StD of NME, which is about 3.0%. The average Mean and StD values of the proposed method on the AFLFP database is about 11.6% and 2.3% respectively.

To establish baseline results for the proposed database, we also compared the performance of the proposed method with the mainstream methods of machine learning and deep learning. In this experiment, we used the Emotrics [29] and SDM [39] as the machine learning baseline. Emotrics is an open-source software based on machine learning to automatically detect facial landmarks for patients with facial palsy. And SDM is one of the most widely used algorithms for the task of facial landmark detection. For deep learning-based methods, VGGNet [56] is the baseline of the direct regression method. And we also used a single FCN structure as the baseline of the heatmap regression method. Table 2 illustrates the comparison results on the AFLFP database. From Table 2, we can conclude observations as follows. First, deep learning-based methods are overall better than machine learning methods. Second, the performance of the heatmap regression method utilizing a single FCN structure achieves better performance than the baseline of the direct regression method. Finally, it is obvious that the performance of the proposed cascaded FCN outperforms the method using a single FCN structure, indicating that the proposed method has robust performance on the AFLFP database.

D. Comparisons between Normal Faces and Palsy Faces

We also conducted a comparison experiment to explore the differences in facial landmark detection when using normal faces and palsy faces as the training data. By this experiment, we aimed to identify the necessity of creating a facial landmark database specifically for facial palsy. For fair comparison, we utilized the proposed cascaded FCN with the same parameters for facial landmark detection. During the experiment, we

TABLE 3
EVALUATION OF EFFECT OF DIFFERENT TRAINING DATA IN FACIAL LANDMARK DETECTION ON THE AFLFP DATABASE. THE HIGHEST RESULTS ARE HIGHLIGHTED IN BOLD.

Training data	NME (%)	
	Mean	Std
Normal face	16.3	3.0
Palsy face	**12.8**	**2.5**

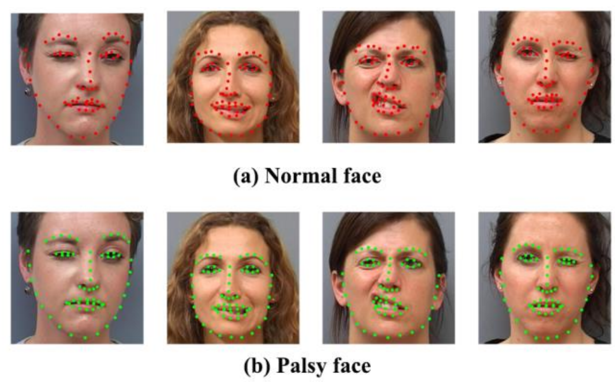

Fig. 7. Qualitative comparison with the different training data.

obtained two training models with two different types of training data which are normal faces from general facial landmark databases and palsy faces from the AFLFP database respectively. Both training models were tested on the palsy faces from AFLFP database. For normal faces, all the training samples from LFPW [34], HELEN [35] and AFW [36] form the training set which has 3148 images in total. To avoid the bias due to different training sizes, we balanced the size of these two databases selecting the same amount of data from the AFLFP database. The metrics of Mean and StD of NME were also used to evaluate the performance. The comparison results on the AFLFP database are shown in Table 3. From Table 3, the model using palsy face data achieves better performance with a Mean of NME of 12.8% and an StD of NME of 2.5%, which are 3.5% and 0.5% lower than the model using normal face data, respectively. Due to limited space, we didn't list the specific performance of 16-class facial expressions in Table 3. Since there is a very large gap between the performance of the model using palsy face data and that of the model using normal face data from Table 3, it can be inferred that the model trained with facial palsy data has better performance on all 16-class facial expressions. One reason for this phenomenon is that palsy faces are much more complex and difficult than normal faces in terms of facial landmark detection because of the large and arbitrary facial deformation. Moreover, we also displayed the results of the qualitative comparison in Fig. 7. From Fig. 7, we can find that the prediction results of facial landmark detection from the model trained with facial palsy data are more accurate. Therefore, it is useful to develop a specific facial landmark database for facial palsy application, which can further improve the performance of existing methods of facial landmark detection in facial palsy.

VI. Conclusion and Future Work

In this paper, a facial asymmetry database for facial landmark detection is presented, which is called Annotated Facial Landmarks for Facial Palsy (AFLFP). AFLFP is a large-scale, diverse, and reliable database that contains facial images with 16-class facial expressions from 88 facial palsy patients, volunteers, and clinicians. To the best of our knowledge, this database is the first public facial landmark database specifically for facial palsy so far. To establish the baseline results of the proposed database, a deep neural network baseline with a two-stage cascaded fully convolutional network (FCN) is proposed to detect facial landmarks. Baseline experiments show that the proposed deep neural network baseline performs better than the mainstream methods of machine learning and deep learning on the AFLFP database. Comparison experiments in facial landmark detection between normal faces and palsy faces show that creating a facial landmark database specifically for facial palsy is very necessary for further performance improvement.

After the publication of this work, the AFLFP database will be made public to researchers to evaluate the methods of facial landmark detection on a database specifically for facial palsy applications. We believe that the release of this database will provide a benchmark for researchers to validate the approaches of facial landmark detection in facial palsy analysis and bring more interested investigators into this study. Moreover, we hope that the combination of this database and the recent progress in deep learning will improve the performance of computer-aided facial palsy diagnosis systems in detecting facial landmarks. In the future, we will keep expanding the database including increasing the subject number and the diversity of the severity degree of facial palsy.

Acknowledgement

The authors would like to thank Peter Anderson (Queen Victoria Hospital, East Grinstead, UK) for the assistance.

References

[1] J. A. McCaul, L. Cascarini, D. Godden, D. Coombes, P. A. Brennan, and C. J. Kerawala, “Evidence based management of Bell’s palsy,” Brit. J. Oral Maxillofacial Surg., vol. 52, no. 5, pp. 387–391, 2014.

[2] J. Finsterer, “Management of peripheral facial nerve palsy,” European Archives of Oto-Rhino-Laryngology, vol. 265, no. 7, pp. 743–752, 2008.

[3] A. Y. Fattah et al., “Facial nerve grading instruments: Systematic review of the literature and suggestion for uniformity,” Plastic Reconstructive Surg., vol. 135, no. 2, pp. 569–579, 2015.

[4] H. Zhu, G. Han, L. Shu, and H. Zhao, “Arvanet: Deep recurrent architecture for ppg-based negative mental-state monitoring,” IEEE Transactions on Computational Social Systems, vol. 8, no. 1, pp. 179–190, 2020.

[5] J. Lou, H. Yu, and F.-Y. Wang, “A review on automated facial nerve function assessment from visual face capture,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 488–497, Feb. 2020.

[6] J. Tavares-Brito, M. M. van Veen, J. R. Dusseldorp, F. Bahmad, and T. A. Hadlock, “Facial palsy-specific quality of life in 920 patients: Correlation with clinician-graded severity and predicting factors,” Laryngoscope, vol. 129, no. 1, pp. 100–104, Jan. 2019.

[7] 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 Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 10, pp. 2325–2332, 2020.

[8] G.-S.-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, Dec. 2019.

[9] Z. Guo et al., “An unobtrusive computerized assessment framework for unilateral peripheral facial paralysis,” IEEE J. Biomed. Health Informat., vol. 22, no. 3, pp. 835–841, May 2018.

[10] J. Barbosa, W.-K. Seo, and J. Kang, “ParaFaceTest: An ensemble of regression tree-based facial features extraction for efficient facial paralysis classification,” BMC Med. Imag., vol. 19, no. 1, pp. 1–4, Dec. 2019.

[11] S. A. Burres, “Facial biomechanics: The standards of normal,” Laryngoscope, vol. 95, no. 6, pp. 708–714, 1985.

[12] L. Modersohn and J. Denzler, “Facial paresis index prediction by exploiting active appearance models for compact discriminative features,” in Proc. 11th Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl., 2016, pp. 271–278.

[13] T. Wang, J. Dong, X. Sun, S. Zhang, and S. Wang, “Automatic recognition of facial movement for paralyzed face,” Bio-medical materials and engineering, vol. 24, no. 6, pp. 2751–2760, 2014.

[14] T. Wang, S. Zhang, J. Dong, L. Liu, and H. Yu, “Automatic evaluation of the degree of facial nerve paralysis,” Multimedia Tools Appl., vol. 75, no. 19, pp. 11893–11908, Oct. 2016.

[15] H. S. Kim, S. Y. Kim, Y. H. Kim, and K. S. Park, “A smartphone-based automatic diagnosis system for facial nerve palsy,” Sensors, vol. 15, no. 10, pp. 26 756–26 768, 2015.

[16] H. Yoshihara, M. Seo, T. H. Ngo, N. Matsushiro, and Y.-W. Chen, “Automatic feature point detection using deep convolutional networks for quantitative evaluation of facial paralysis,” in Proc. 9th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI), Oct. 2016, pp. 811–814.

[17] W. S. W. Samsudin, R. Samad, M. Z. Ahmad, and K. Sundaraj, “Forehead lesion score for facial nerve paralysis evaluation,” in Proc. IEEE Int. Conf. Autom. Control Intell. Syst. (I2CACIS), Jun. 2019, pp. 102–107.

[18] M. P. Lafer, “Management of long-standing flaccid facial palsy static approaches to the brow, midface, and lower lip,” Otolaryngologic Clinics North Amer., vol. 51, no. 6, pp. 1141–1150, 2018.

[19] N. Bui, J. Yen, and V. Honavar, “Temporal causality analysis of sentiment change in a cancer survivor network,” IEEE transactions on computational social systems, vol. 3, no. 2, pp. 75–87, 2016.

[20] 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, Oct. 2019.

[21] A. Song, Z. Wu, X. Ding, Q. Hu, and X. Di, “Neurologist standard classification of facial nerve paralysis with deep neural networks,” Future Internet, vol. 10, no. 11, p. 111, Nov. 2018, doi: 10.3390/fi10110111.

[22] M. Sajid, T. Shafique, M. Baig, I. Riaz, S. Amin, and S. Manzoor, “Automatic grading of palsy using asymmetrical facial features: A study complemented by new solutions,” Symmetry, vol. 10, no. 7, p. 242, Jun. 2018, doi: 10.3390/sym10070242.

[23] L. Liu, G. Cheng, J. Dong, S. Wang, and H. Qu, “Evaluation of facial paralysis degree based on regions,” in Proc. 3rd Int. Conf. Knowl. Discovery Data Mining, Jan. 2010, pp. 514–517.

[24] K. Anguraj and S. Padma, “Evaluation and severity classification of facial paralysis using salient point selection algorithm,” Int. J. Comput. Appl., vol. 123, no. 7, pp. 23–29, Aug. 2015.

[25] J. Barbosa et al., “Efficient quantitative assessment of facial paralysis using iris segmentation and active contour-based key points detection with hybrid classifier,” BMC Med. Imag., vol. 16, no. 1, p. 23, Mar. 2016.

[26] C. H. J. Tzou et al., “Evolution of the 3-dimensional video system for facial motion analysis: Ten years’ experiences recent developments,” Ann. Plastic Surg., vol. 69, no. 2, pp. 173–185, 2012.

[27] T. H. Ngo, Y.-W. Chen, M. Seo, N. Matsushiro, and W. Xiong, “Quantitative analysis of facial paralysis based on three-dimensional features,” in Proc. Int. Conf. Image Process. (ICIP), Sep. 2016, pp. 1319–1323.

[28] G.-S. J. Hsu, W.-F. Huang, and J.-H. Kang, “Hierarchical network for facial palsy detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2018, pp. 580–586.

[29] D. L. Guarin, J. Dusseldorp, T. A. Hadlock, and N. Jowett, “A machine learning approach for automated facial measurements in facial palsy,” JAMA Facial Plastic Surg., vol. 20, no. 4, pp. 335–337, 2018.

[30] O. Azoulay, Y. Ater, L. Gersi, Y. Glassner, O. Bryt, and D. Halperin, “Mobile application for diagnosis of facial palsy,” in Proc. 2nd Int. Conf. Mobile Inf. Technol. Med., Prague, Czech Republic, Nov. 2014.

[31] C. A. Banks, P. K. Bhama, J. Park, C. R. Hadlock, and T. A. Hadlock, “Clinician-graded electronic facial paralysis assessment: The eFACE,” Plastic Reconstructive Surg., vol. 136, no. 2, pp. 223–230, 2015.

[32] T. H. Ngo, M. Seo, N. Matsushiro, and Y.-W. Chen, “Evaluation of facial paralysis based on spatial features of filtered images,” Int. J. Biosci., Biochem. Bioinf., vol. 6, no. 1, p. 1, 2016.

[33] F. De la Torre, W.-S. Chu, X. Xiong, F. Vicente, X. Ding, and J. Cohn, “IntraFace,” in Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG), May 2015, pp. 1–30.

[34] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, “Localizing parts of faces using a consensus of exemplars,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 12, pp. 2930–2940, 2013.

[35] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, “Interactive facial feature localization,” in European conference on computer vision. Springer, 2012, pp. 679–692.

[36] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012, pp. 2879–2886.

[37] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, “300 faces in-the-wild challenge: The first facial landmark localization challenge,” in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 397–403.

[38] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, Dec. 2001, p. I-511.

[39] X. Xiong and F. De la Torre, “Supervised descent method and its applications to face alignment,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2013, pp. 532–539.

[40] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[41] Y. Xia, H. Yu, and F.-Y. Wang, “Accurate and robust eye center localization via fully convolutional networks,” IEEE/CAA Journal of Automatica Sinica, vol. 6, no. 5, pp. 1127–1138, 2019.

[42] Zhang, Wenjin, and Jiacun Wang. "Dynamic hand gesture recognition based on 3D convolutional neural network models." In 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC) (pp. 224-229). IEEE, 2019.

[43] C. Timen, S. Vinayahalingam, K. J. Ingels, S. J. Bergé, T. J. Maal, and C. M. Speksnijder, “Reliability and agreement of 3d anthropometric measurements in facial palsy patients using a low-cost 4d imaging system,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 8, pp. 1817–1824, 2020.

[44] W. Deng, Y. Fang, Z. Xu, and J. Hu, “Facial landmark localization by enhanced convolutional neural network,” Neurocomputing, vol. 273, pp. 222–229, 2018.

[45] C. C. dos Santos, J. L. A. Samatelo, and R. F. Vassallo,“Dynamic gesture recognition by using cnns and star rgb: A temporal information condensation,” Neurocomputing, vol. 400, pp. 238–254, 2020.

[46] X. Tang, F. Guo, J. Shen, and T. Du, “Facial landmark detection by semi-supervised deep learning,” Neurocomputing, vol. 297, pp. 22–32, 2018.

[47] N. Zeng, H. Li, Z. Wang, W. Liu, S. Liu, F. E. Alsaadi, and X. Liu, “Deep-reinforcement-learning-based images segmentation for quantitative analysis of gold immunochromatographic strip,” Neurocomputing, 2020.

[48] Z. Niu, G. Zhong, H. Yu, A Review on the Attention Mechanism of Deep Learning, Neurocomputing, 2021, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.03.091.

[49] Z. Wang, J. Xin, Z. Wang, H. Gu, Y. Zhao, and W. Qian, “Computer-aided dementia diagnosis based on hierarchical extreme learning machine,” Cognitive Computation, pp. 1–15, 2020.

[50] S. Zhang, H. Yu, T. Wang, J. Dong, T.. Pham,Linearly augmented real-time 4D expressional face capture,Information Sciences, Volume 545, 2021, Pages 331-343, ISSN 0020-0255.

[51] X. Fei, L. Shen, S. Ying, Y. Cai, Q. Zhang, W. Kong, W. Zhou, and J. Shi, “Parameter transfer deep neural network for single-modal b-mode ultrasound-based computer-aided diagnosis,” Cognitive Computation, vol. 12, no. 6, pp. 1252–1264, 2020.

[52] D. Haputhanthri, G. Brihadiswaran, S. Gunathilaka, D. Meedeniya, S. Jayarathna, M. Jaime, and C. Harshaw, “Integration of facial thermography in eeg-based classification of ASD,” International Journal of Automation and Computing, pp. 1–18, 2020.

[53] W. Liu, Z. Wang, X. Liu, N. Zeng, and D. Bell, “A novel particle swarm optimization approach for patient clustering from emergency departments,” IEEE Transactions on Evolutionary Computation, vol. 23, no. 4, pp. 632–644, 2018.

[54] J. Li, N. Li, X. Shao, J. Chen, Y. Hao, X. Li, and B. Hu, “Altered Brain Dynamics and Their Ability for Major Depression Detection using EEG Microstates Analysis,” IEEE Transactions on Affective Computing, 2021.

[55] X. Zhang, L. Yin, J. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. Girard, “BP4D-Spontaneous: A high resolution spontaneous 3D dynamic facial expression database”, Image and Vision Computing, 32 (2014), pp. 692-706.

[56] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[57] H. Yu, O. Garrod, and P. Schyns, “Perception-driven facial expression synthesis.”, Computers & Graphics, pp.152-162. 2012.

[58] Q. Xiao, C. Li, Y. Tang, and X. Chen, “Energy efficiency modeling for configuration-dependent machining via machine learning: A comparative study,” IEEE Transactions on Automation Science and Engineering, 18(2), 717-730, 2020.

[59] Y. Guo, Y. Xia, J. Wang, H. Yu, and R.-C. Chen, “Real-time facial affective computing on mobile devices,” Sensors, vol. 20, no. 3, p. 870, 2020.

[60] Hu, Bin, and Jiacun Wang. "Deep learning based hand gesture recognition and UAV flight controls." International Journal of Automation and Computing, 17(1), 17-29, 2020.

[61] W. Zhang, J. Wang, and F. Lan. "Dynamic hand gesture recognition based on short-term sampling neural networks." IEEE/CAA Journal of Automatica Sinica, 8(1), 110-120, 2020.

[62] Z. Tu, A. Abel, L. Zhang, B. Luo, and A. Hussain, “A new spatio-temporal saliency-based video object segmentation,” Cognitive Computation, vol. 8, no. 4, pp. 629–647, 2016.

[63] S. Ghosh and T. Anwar, “Depression intensity estimation via social media: A deep learning approach,” IEEE Transactions on Computational Social Systems, vol. 8, no. 6, pp. 1465–1474, 2021.

[64] Y. Xia, W. Zheng, Y. Wang, H. Yu, J. Dong, and F.-Y. Wang, “Local and global perception generative adversarial network for facial expression synthesis,” IEEE Transactions on Circuits and Systems for Video Technology, 2021.

[65] J. J. Greene, D. L. Guarin, J. Tavares, E. Fortier, M. Robinson, J. Dusseldorp, O. Quatela, N. Jowett, and T. Hadlock,“The spectrum of facial palsy: The meei facial palsy photo and video standard set,” The Laryngoscope, vol. 130, no. 1, pp. 32–37, 2020.