IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 September 2022, accepted 8 October 2022, date of publication 17 October 2022, date of current version 1 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3215138

RESEARCH ARTICLE

Deep Learning-Based Auricular Point Localization for Auriculotherapy

XIAOYAN SUN[1,2], JIAGANG DONG[1], QINGFENG LI[2], DONGXIN LU[2], AND ZHENMING YUAN[1,2]

[1]School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China
[2]Engineering Research Center of Mobile Health Management System, Ministry of Education, Hangzhou 311121, China

Corresponding author: Zhenming Yuan (zmyuan@hznu.edu.cn)

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LGF20F020009, and in part by the Key Research and Development Program of Zhejiang Province under Grant 2020C03107.

ABSTRACT Auriculotherapy is one of the main forms of treatment in Traditional Chinese Medicine, whose potential as an alternative medicine for both health evaluation and disease treatment has been reported in many cases. However, its efficacy highly relies on the accurate localization of auricular points, which are not easy to be remembered due to their complexity. To explore an efficient way of locating auricular points, this study proposed a deep learning-based method of automatically locating auricular points from auricular images. A self-collected dataset named EID was created for TCM auriculotherapy research, with 91 auriculotherapy-related landmark points manually annotated according to the Chinese national standardization. A deep neural network structure was trained for landmark detection, and a direction normalization module was proposed to compensate for the detection error caused by the difference between the left and right ears. The trained model was validated on dataset EID. An average NME of 0.0514±0.0023 was achieved, which outperformed similar works. In addition, a certain auricular area corresponding to the digestive system was segmented based on the localized landmarks, and the results were tested in real-time video streaming. The proposed work for both auricular landmark and area identification can be widely used in auriculotherapy education and applications.

INDEX TERMS Auriculotherapy, deep learning, landmark detection.

I. INTRODUCTION

Auriculotherapy is a method of alternative therapy, whose efficacy has been reported in many publications, including the treatment for insomnia [1], [2], obesity [3], pain relief [4], and chronic fatigue syndrome of qi deficiency constitution [5]. In 1990, the World Health Organization (WHO) published the auricular acupuncture nomenclature and standardized 43 auricular points [6]. Auriculotherapy is part of Traditional Chinese Medicine (TCM), for both diagnosis and treatment. According to the latest National Standardization of Auricular Point of the People's Republic of China GB/T13734-2008 [7], 93 auricular points and 76 acupoint areas are defined. Correct identification and positioning of auricular points are essential for auricular point therapy. However, because the auricular elements are small and the number of acupoints is large, it is very difficult to remember them without years of practice. Therefore, auriculotherapy can only be performed by professional doctors based on their own experience. Several tools have been developed to assist the positioning procedure, with bioelectrical measurement, and skin dyeing [8], [9], [10], etc., which are time-consuming and not easy to be operated in general. This paper proposes a deep learning-based method to automatically identify the acupoints and auricular areas from ear images for auriculotherapy.

II. RELATED WORK

Currently, research for human auricle-related topics is mainly focused on ear detection from images [11], [12] or 3D point clouds [15], [16]. All these studies are focused on the external part of the auricle only, the detection of individual

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

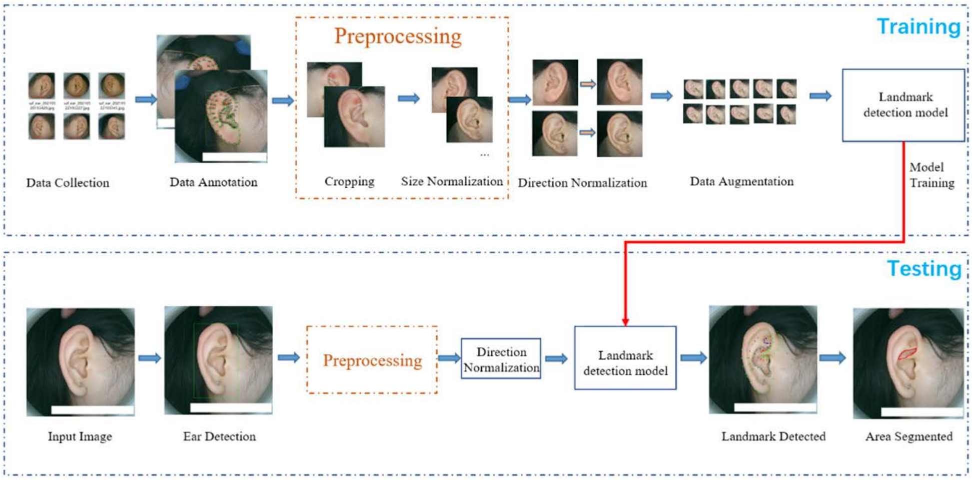


FIGURE 1. Overview of the proposed method.

auricular landmarks and elements has not yet been studied in depth. Little research studies the inner part of the auricle, especially for auriculotherapy-related tasks. Mussi et al. [18] used an image processing algorithm to segment the auricular areas based on depth map images. The study segmented seven major areas inside the auricle including the helix, the antihelix, and the concha cavity. Similar work is presented by Lei et al. [19] An ear tree-structured graph (ETG) was proposed and a 3-D flexible mixture model was trained to locate 18 landmarks of the auricle anatomy in 3D. However, no auriculotherapy-related landmarks are localized in both [18] and [19]. Wen et al. [20] used the ASM algorithm and 25 auricular subzones were divided according to the WFAM standard. While the deep learning-based methods have shown promising results in both landmark detection [29], [30] and area segmentation tasks [21], [42], one of the possible reasons for limited studies on auricle-related topics might be the missing of ear image datasets with annotated landmarks. One closely related research topic is facial landmark detection, and many studies localize landmark points by utilizing deep learning technology, including various of CNN-based networks [22], [24], [27], multi-task learning [23], [25], and transform learning [28] etc. For example, Sun et al. [22] proposed a three-level cascaded convolutional network to estimate the positions of facial keypoints. Zhang et al. [23] proposed a multi-task learning model to train facial landmark detection together with head pose estimation and facial attribute inference. Zhang et al. [25] proposed a deep cascading multi-task framework to enhance performance by making use of the intrinsic relationship between face detection and keypoints location. Kowalski et al. [26] proposed a deep alignment network (DAN), a robust face alignment method based on a deep neural network architecture. Zhang et al. [27] proposed a weakly supervised landmark-region-based convolutional neural network (LR-CNN) framework to detect facial components and landmarks simultaneously. However, these CNN-based networks usually have a large model size and may be not well applied to mobile devices, such as VGG16, and ResNet50. Zhao et al. [28] propose a lightweight model, namely Mobile Face Alignment Network (MobileFAN), using a simple backbone MobileNetV2 as the encoder and three deconvolutional layers as the decoder. Guo et al. [29] utilized a similar idea as Zhao's work, and applied MobileNetV2 as a backbone network but for facial keypoints localization. Saxen et al. [31] used two lightweight CNN, MobileNetV2 and Nasnet-Mobile for facial attributes detection, which performed faster than similar works. As these works indicated, MobileNetV2 has the advantages of being lightweight and high efficiency, therefore can be used on mobile devices.

The main contributions of this paper are as follows:

(1) A new dataset of auricle images is specifically collected for auricular acupoint localization using specially designed equipment, with 91 landmarks manually annotate;

(2) A deep learning-based landmark detection method is introduced with a direction normalization module to compensate for the asymmetry of the auricle shape;

(3) The special auricle area, i.e., cymba conchae is segmented according to the corresponding points for further application of auriculotherapy.

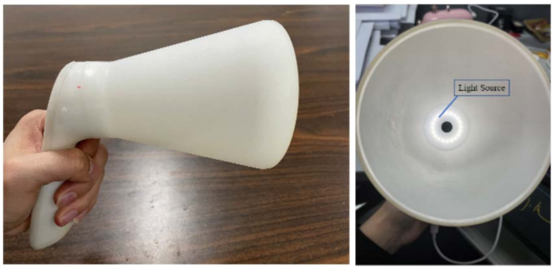
III. METHOD

An overview of our work was shown in Fig. 1. The main diagram of the proposed method included offline model training for landmark prediction and online application for both landmark detection and auricle area segmentation. The detailed steps are described in the following sections.

A. DATASET

1) IMAGE COLLECTION

To train the deep learning model for acupoints identification, the images showing the clear surface structure of the auricle are a prerequisite. However, most of the current datasets of auricular images are captured in wild without detailed



(a) (b)

FIGURE 2. Specially designed device for image collection. (a) device appearance, (b) embedded light source for lighting consistence.

TABLE 1. Details of the dataset EID.

	Number
Subjects	252
Sex	
Male	99
Female	153
Age	22±3
Images	
Left Ear Images	252
Right Ear Images	252

information. About annotations for landmarks that are needed for supervised training. Therefore, a handheld device (as shown in Fig. 2) was designed and used for data acquisition. The device is embedded with a stable light source covering the entire auricle so that the captured image contains a clear surface structure of the auricle. Both static images and dynamic videos could be captured with this device.

A self-collected image dataset consisting of a total of 252 participants' auricle images were collected with the device shown in Fig. 2. The statistical distribution of the participants is provided in Table 1. For each participant, a pair of images of the left and the right ear, and the size of each image with 500 × 500, were collected. The dataset is composed of 252 pairs resulting in a total of 504 images. In the following of this paper, Ear Image Dataset (EID) will be used as the short-term for this ear image dataset.

2) LANDMARK SPECIFICATION

All the landmarks are defined according to GB/T13734-2008 [7], under the instruction of a qualified TCM auriculotherapy practitioner. A total of 91 landmarks are defined as shown in Fig. 3, including 31 primary and 60 secondary landmarks, denoted with red and yellow dots, respectively. Primary landmarks are acupuncture points and some special points on the contour of the ear. To denote the auricular acupoint area better, 60 secondary landmarks are further introduced in between the primary landmarks. The numbering of landmarks is denoted in Fig. 3.

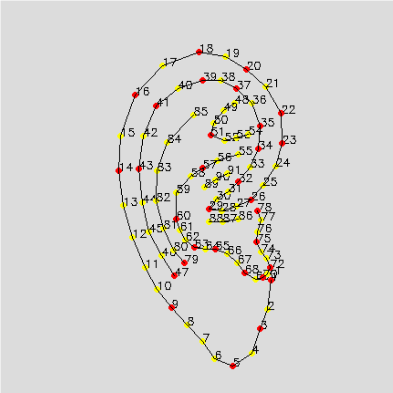


FIGURE 3. The landmark scheme was applied to the dataset. Red/yellow dots denote primary and secondary landmarks, respectively.

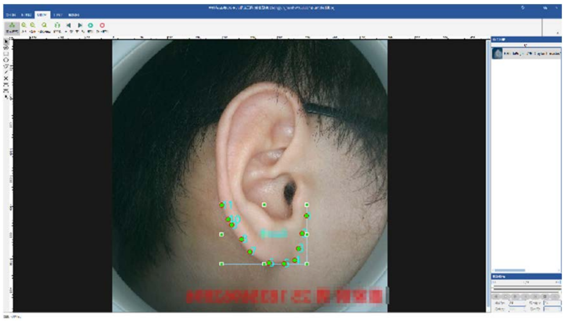
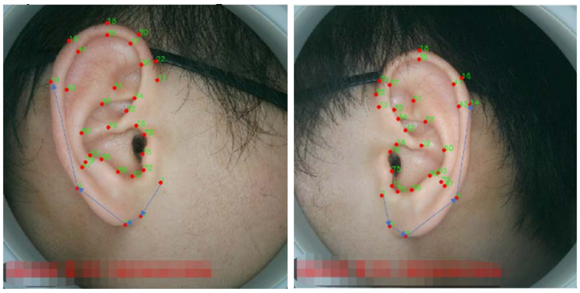


FIGURE 4. Annotating landmarks with a labeling tool.



(a) (b)

FIGURE 5. Landmarks annotation for (a) right and (b) left ear samples.

3) LANDMARK ANNOTATION

Labeling Tool - "CasiaLabeler" [13] is used for landmark annotation, which is shown in Fig. 4. All the landmarks are annotated in sequence from #1 to #91 as defined in Fig. 3 by a qualified TCM auriculotherapy practitioner. The coordinates of all the landmark points are recorded.

Fig. 5 illustrates the spatial distribution of the annotated landmarks for the right and left ear samples, respectively. Blue arrows indicated the directions when connecting the adjacent primary landmarks in sequence.

Four individuals participated in the landmark annotation task for the 504 images in the dataset. A qualified TCM auriculotherapy practitioner is invited to train all the

TABLE 2. Results of intraclass correlation efficient.

Two-way mixed-effect/random-effect consistency	ICC	95%Confidence Intervals (CI)
X-axis Single Measurement ICC(C,1)	0.988±0.004	0.974 ~ 0.994
Y-axis Single Measurement ICC(C,1)	0.993±0.002	0.986 ~ 0.997

annotators beforehand and provide inspection during the annotating procedure.

The Intraclass Correlation Coefficient (ICC) [14] is used to evaluate the annotation reliability. ICC is one of the indicators used to measure and evaluate inter-observer reliability. The ICC value is between 0 and 1, and a high ICC close to 1 indicates the high reproducibility of numerical measurements made by different annotators. It is generally suggested that an ICC lower than 0.40 indicates poor reliability, and greater than 0.75 indicates high reliability [32]. The ICC is calculated as follows

$$ICC = \frac{MS_{observed} - MS_{error}}{MS_{observed} + (k-1)MS_{error}}. \tag{1}$$

n which, MS_{observed} is the mean square of the observed objects, MS_{error} is the mean square of the errors, MS_{observer} is the mean square of the observer, k is the number of observers.

The four annotators are asked to annotate a total of 8 images, respectively, which are composed of 4 images of the left ear and 4 images of the right ear. For each image, the X and the Y coordinates of each landmark point are recorded for analysis, and the mean and the standard deviation of the ICC calculated from the eight images are reported in Table 2. In this paper, the error between annotators is not considered, and the calculation is based on the original data. Therefore, the calculation mode ICC (C, 1) is selected, where C represents consistency and 1 represents a single measurement:

It can be seen that the ICC values are 0.988±0.004 (95%CI: 0.974∼0.994) and 0.993±0.002 (95%CI: 0.986∼0.997) for annotating in the X- or Y-axis, which indicates a high annotation consistency.

B. IMAGE PREPROCESSING

1) EAR DETECTION AND SIZE NORMALIZATION

A neural network is trained for ear detection with Dlib [33]. A total of 100 images are randomly chosen as the training data for the detector. The auricle area is manually annotated in the 100 images, and the model has trained accordingly.

With the trained ear detector, the bounding box of the entire auricle is achieved, which is expanded 1.1 times larger to form a new square-size bounding box. Images inside this square box are cropped and resized to 112 × 112 (Fig. 6) to speed up the landmark detection procedure.

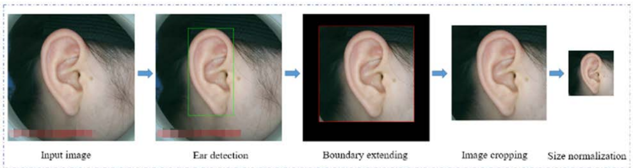


FIGURE 6. Ear detection and size normalization.

2) DIRECTION NORMALIZATION

There are certain differences between the left and right ears of humans [34]. As seen in Fig. 5, the annotated point sequences for left and right ears show obvious diversity differences in their directions. In our experiments, it is found that the efficacy of landmark detection is highly influenced by the ear direction captured in the sample image. To eliminate the influence caused by the difference in the directionality of the left and right ears, it is important to normalize the ear direction. The Haar [35] cascade detector in the OpenCV library is used to identify the ear direction as left or right. The image is flipped horizontally according to "Algorithm: Direction Normalization".

Algorithm 1 Direction Normalization

```
Input: original image ($I$)
Output: flipped image ($I_f$)
Detect ear direction (ed) from $I$
if ed is not standard
    Calculate the width ($W$) of $I$
    Set $i, j$ as the row and column indexes in $I$
    for $i$ 0 to $W$-1 do
      for $j$ 0 to $W$-1 do
          $I_f(i, j) = I(W\text{-}i, j)$
    end.
  end.
return $I_f$.
```

The coordinates of the annotated landmarks in the original image are flipped according to formula (2), where W represents the width of the image, (x_0, y_0) represents the coordinates of a landmark in the original image, and (x_f, y_f) represents the corresponding coordinates after the flipping.

$$\begin{bmatrix} x_f \\ y_f \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & W \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \tag{2}$$

3) DATA AUGMENTATION

Due to the relatively small amount of data in the data set EID, data augmentation techniques are applied to the training samples to expand the number of training data. To simulate different camera angles that possibly occur during the image capturing, rotation is performed for each image with a rotation angle between −30° and 30°, in every 5°. A linear

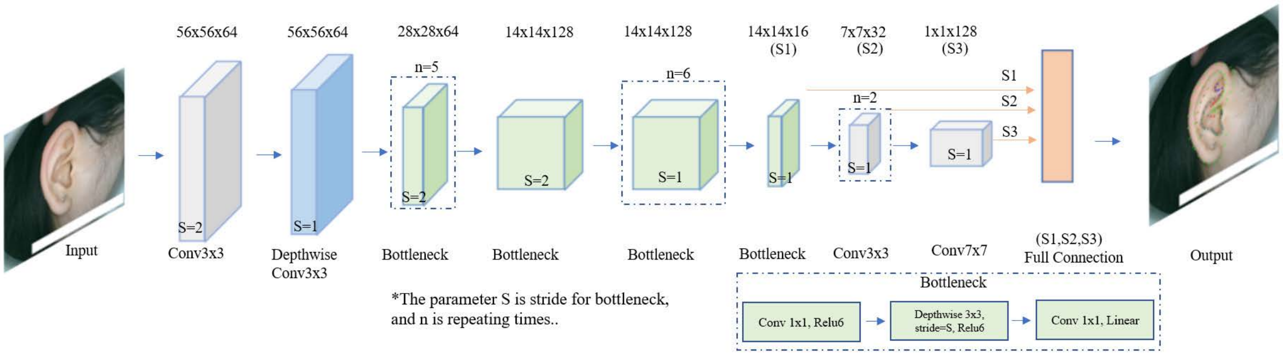


FIGURE 7. The network structure for landmark detection.

TABLE 3. Sample size before and after data augmentation.

Samples	Train Datasets		Test Datasets
	Before	After	
Left Ear	202	2424	50
Right Ear	202	2424	50

interpolation method as described in [36] is used to fill the rotated image. Therefore, 12 new samples are generated for each original training sample. The sample size before and after data augmentation is shown in Table 3.

C. LANDMARK DETECTION

1) NETWORK CONSTRUCTION

A convolutional neural network (CNN) can extract features of input layer by layer from low-dimensional to high-dimensional, which makes the feature extraction more accurate and achieves remarkable results. Backbone networks with stronger feature representation capabilities are introduced in recent years, such as VGG16 [37], ResNet50 [38], etc. However, the training of these networks requires a large amount of calculation, therefore it is hard to be run on mobile devices with limited computing power. The ultimate goal of our research is a mobile-based application for auriculotherapy, choosing a backbone network that can be used for mobile devices or embedded devices becomes the primary consideration. MobileNetV2 is used as the backbone network in this paper [29]. The network structure is shown in Fig. 7, the output contents of the latter three layers are fused to increase the performance.

2) EVALUATION METRICS FOR LANDMARK DETECTION

For the evaluation of landmark detection accuracy, the evaluation metric used in this paper is a normalized mean error (NME), and its definition is shown in formula (3):

$$\mathrm{NME} = \frac{1}{M}\sum\nolimits_{i=1}^{M}\frac{||pi - \hat{p}i||2}{d}. \tag{3}$$

where M is the number of landmark points in the image; p_i and $\hat{p}_i$ are the actual and predicted coordinates for the i^{th} point, respectively; d is the normalization factor, which is set as the ear width in our case.

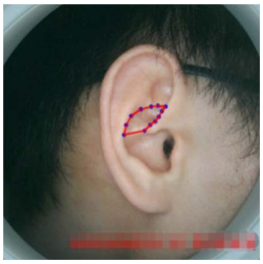


FIGURE 8. Illustration of the cymba conchae.

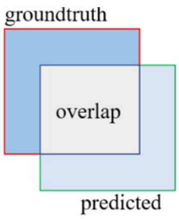


FIGURE 9. Illustration of IoU.

D. AURICULAR AREA SEGMENTATION

1) TARGET AREA CONSTRUCTION

After detecting all of 91 landmarks on the ear, the segmentation for some special auricular areas is also performed. In this paper, the cymba conchae is chosen as the target area for the segmentation, defined according to GB/T13734-2008. This area corresponds to the kidney, urinary system, and other organs, according to TCM theory. The health status of the digestive system inside the abdominal area could be reflexed by the appearance of the cymba conchae. The cymba conchae is illustrated in Fig. 8, as the area inside the red line. The segmentation of this area is done based on the identified landmarks #29∼34, #55∼60, as defined in Fig. 3.

2) EVALUATION METRICS FOR AREA SEGMENTATION

Intersection over Union (IoU) is used as the evaluation metric for the segmentation of the target auricular area. As shown in Fig. 9, ‘‘predicted’’ indicates the predicted

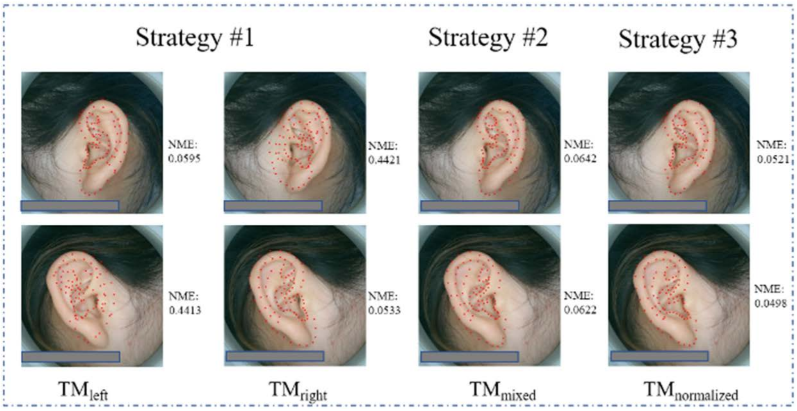


FIGURE 10. Comparison of results of different training strategies.

TABLE 4. Experimental environments and settings.

Parameter	Content
Deep Learning Framework	Pytorch1.7
Operation System	Ubuntu18.04
Hardware Configuration	Memory: 8GB
	Graphics Card: GTX1050-maxq
Hyper-parameters	Input shape: (112x112x3)
	Optimizer: Adam
	Epochs: 500
	Batch size: 16
	Learning rate: 0.0001

region, "groundtruth" indicates the ground-truth region, and "overlap" indicates the overlapping region of the two.

Its calculation is shown in the following formula (4):

$$IoU = \frac{overlap}{groundtruth + predicted - overlap}. \tag{4}$$

IV. AURICULAR AREA SEGMENTATION

Experiments are carried out to evaluate the effectiveness of our proposed method for landmark detection and area segmentation. The detailed information on experimental environments and settings for model training is shown in Table 4.

To obtain a more reliable and stable model, this paper uses the 5-fold cross-validation method. All the experimental results are reported by the 5-fold mean ± standard deviation.

A. LANDMARK DETECTION

Three strategies are designed and tested for model training, and the specific amount of experimental data is shown in Table 5:

Strategy #1: training two separate models with preprocessed data for the left and right ears, respectively;

Strategy #2: training a single model with left and right ear samples mixed;

Strategy #3: training a single model with direction normalization applied to preprocessed data.

Considering the asymmetry in the shape of one's left- and right-side ear, images captured for the left and right ear are trained separately for two models in Strategy #1. A total of 25 left samples and 25 right samples are tested with both two trained models, and the landmark detection results are listed in Table 6.

Results from Strategy #1 indicated that a model trained with single-side ear images only achieves low landmark detection accuracy for the opposite side ears. Therefore, samples with left and right ears are mixed in Strategy #2 for model training. The same testing data are used to evaluate the landmark detection accuracy, and the results are shown in Table 6.

To fully utilize the training samples for higher accuracy, in Strategy #3, the direction normalization module is introduced to normalize the direction of all the samples into the single one. The direction normalization operator as defined in session "DIRECTION NORMALIZATION" is applied to left ear samples before the model training. In this way, all the samples are treated in one direction, so that the best model performance can be expected. The experimental results are also shown in Table 6.

Fig. 10 shows the landmark detection results for two testing samples using different strategies, where the two sample captures left- and right-side ear, respectively. From the results, it is showed that model TM$_{mixed}$ and TM$_{normalized}$ performed well in predicting the mixed ear samples, and TM$_{normalized}$ achieved better results.

B. AURICULAR AREA SEGMENTATION

The cymba conchae is segmented based on landmarks identified using Strategy #3. The IoU results for the testing samples are listed in Table 7. For the left ears, the IoU is 0.6629 ± 0.0458. For the right ears, the IoU is 0.6826 ± 0.0103. The average IoU is 0.6731 ± 0.0233. Fig. 11 shows the results of landmark detection and auricular area segmentation for one of the testing images, where red and blue lines

TABLE 5. Statistics of experimental data.

Strategy	Number of training samples Left / Right / Total	Number of testing samples Left / Right / Total	Number of models	Experimental assessment	Evaluation metrics
#1	202 / 0 / 202	25/25/50	2	5-fold cross-validation	NME
#1	0 / 202 / 202	25/25/50			
#2	101 / 101 /202	25/25/50	1		
#3	101 / 101 /202	25/25/50	1		

TABLE 6. Results of different training strategies.

Strategy/Sample type		Left Ear	Right Ear	Average
Strategy #1	TM_{left}	0.0588±0.0024	0.4457±0.0026	0.2522±0.0025
	TM_{right}	0.4472±0.0019	0.0556±0.0038	0.2514±0.0026
Strategy #2	TM_{mixed}	0.0652±0.0036	0.0651±0.0034	0.0652±0.0034
Strategy #3	$TM_{normalized}$	**0.0531±0.0042**	**0.0497±0.0032**	**0.0514±0.0023**

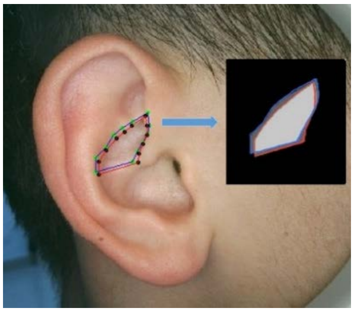


FIGURE 11. Illustration of the segmentation result. (Red / blue lines indicate the boundary of the annotated and segmented area, respectively).

TABLE 7. Segmentation results with TMnormalized.

Sample type	Left Ear	Right Ear	Average
IoU	0.6629±0.0458	0.6826±0.0103	0.6731±0.0233

indicate the boundary of the annotated and segmented area, respectively.

The segmentation is also tested in real-time with video streams captured by using the device described in session "IMAGE COLLECTION". Fig. 12 provides the segmentation results for the cymba conchae area in the video stream. It is seen that the segmentation is accurate and stable with a frame rate of 25fps.

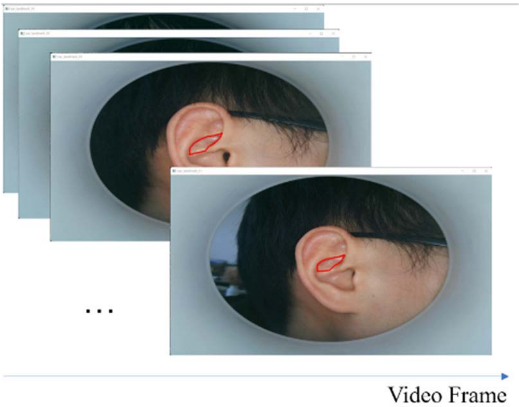


FIGURE 12. Segmentation results displayed in the real-time video streaming.

V. DISCUSSION

A self-collected dataset of ear images is introduced in this paper, namely EID. To the best of our knowledge, EID is the first dataset created for TCM auriculotherapy research, in which the acupoints related landmarks are annotated according to the Chinese national standardization. The comparison of EID with some of the popular ear image databases is given in Table 8, including UND-Collection E [39], WPUTEDB [40], iBUG-ears [17], and USTB Human Ear Image Library [41]. UND-Collection E includes 464 images from 114 human subjects, which were taken at different poses and lighting conditions, with only the right ears were captured. WPUTEDB contains 2071 images from 501 objects, which were collected as a testing tool for biometric algorithms. Images were acquired both outdoors and in a dark environment, with occlusions caused by hair, glasses, and earrings. All the above datasets contain ear images only, but no landmark annotation is provided. iBUG-ears database is an "in-the-wild" images dataset, which includes two sets of

TABLE 8. Comparison of our dataset (EID) with other datasets.

Dataset	Sample Characteristics	Image Characteristics	With Landmarks?
UND-Collection E (2002) [39]	464 images from 114 participants, captured in various pose and illumination conditions	Color images, right ears only, 640x480	none
USTB Human ear image library (2002) [41]	180 images from 60 participants, captured in a controlled environment	Grayscale image, right ear only, 80x150	none
WPUTEDB (2010) [40]	2071 images from 501 participants, captured in both outdoors and dark environments	Color images, varying sizes	none
iBUG-Collection A (2017) [17]	605 images, collected from Google Images	Color images, varying sizes	55 anatomical points
Ours (EID)	504 images from 252 participants, captured in a controlled environment	Color images, 500x500	91 auriculotherapy-related points

ear images: Collection A and Collection B, where 605 and 2058 images are collected from the Google Images and VGG databases, respectively. Because all the images are collected "in the wild", the surface structure of the ear in the images is usually not clear. iBUG-Collection A database includes 55 manually annotated landmark points for each image. However, these points are mainly based on the anatomical shape of the auricle only. USTB Human Ear Image Library uses a digital camera to take 180 images of the right ear from 60 participants, and the images are in 256 gray levels. Wen et al. [20] used some images in the USTB human ear dataset and manually labeled 65 landmarks, but only 30 images were annotated. Our EID dataset contains ear images collected with the specially designed device, therefore images with clear structure and color information are acquired. Compared with the 55 landmarks in iBUG-Collection A, EID annotates a total of 91 landmark points on the data. With the detailed images and annotated landmarks which identifies the acupoints in addition to the ear shape, EID can help researchers to identify the acupoints better.

Three training strategies are designed and tested. By comparing the results from Strategy #1 and #2, it can be seen that training with mixed samples improves the landmark detection results in general. Further experiments are performed to evaluate how the proportion of samples for different ear direction influence the final landmark detection accuracy, and the results are shown in Table 9. As shown in Fig. 13, for a certain ear sample, higher accuracy is achieved when more training samples of the same side ear are included, where the x-axis is the ratio of left ear samples to right ear samples, and the y-axis gives the resulted accuracy in average NME.

The possible explanation for this might be that the networks learned both the structural and directional features of the auricles. Because the 91 landmarks are annotated in sequential order, the shapes constructed from the annotated points of the left and right ear are different. With Strategy #3, the directional features provided with all the training samples are the same, therefore, the structural features i.e., the localizations of landmarks are extracted in the biggest content.

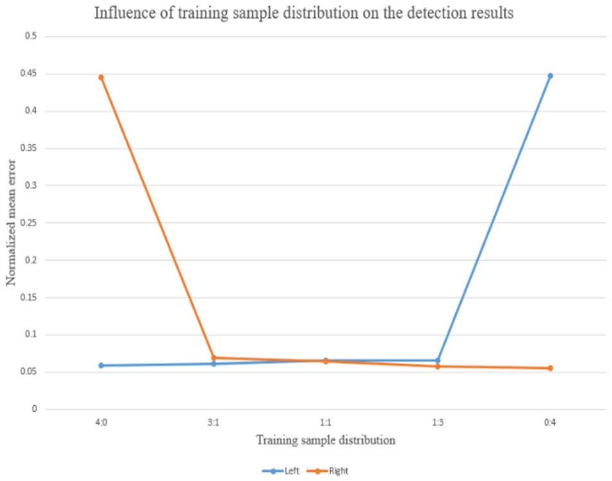


FIGURE 13. Influence of training sample distribution on the detection results.

TABLE 9. Experimental results of different sample proportions.

Proportion (Left: Right)/Sample type	Left	Right
4:0	0.0588±0.0024	0.4457±0.0026
3:1	0.0615±0.0035	0.0693±0.0035
1:1	0.0652±0.0036	0.0651±0.0034
1:3	0.0658±0.0027	0.0573±0.0025
0:4	0.4472±0.0019	0.0556±0.0026

The training Strategy #1 could be considered as the special case of Strategy #2, where the split for the ear samples with different directions is all to none. Therefore, for single-sided ears, Strategy #1 results in higher accuracy than Strategy #2, but it requires the training of two models separately. The results from Strategy #1 also indicate that the accuracy for right ears is higher than for left ears (0.0588±0.0024 vs. 0.0556±0.0038). Therefore, the normalization direction is set to right-sided in Strategy #3 for better performance. Whether such a difference exists for all cases or only for our dataset requires further exploration.

TABLE 10. Comparison of our results with others.

Method	Task	Datasets	Number of Landmarks	Results for Landmark Detection (NME)	Results for Area Segmentation (IoU)
Wen[20]	D	30 images from USTB	65	6.67±0.59 (Euclid distance)	none
Zhou[17]	D	iBUG-Collection A	55	0.0522±0.0246	none
Mussi [18]	S	self-generated 3D models	none	none	0.9961±0.10 (Similarity)
Ours	D	iBUG-Collection A	55	**0.0493±0.0012**	none
MTCNN[25]	D&S	EID	91	0.0765±0.0143	0.6421±0.0341
TCDCN[23]	D&S	EID	91	0.0715±0.0046	0.6375±0.0532
Ours	D&S	EID	91	**0.0514±0.0023**	**0.6731±0.0233**

D: Landmark Detection S: Auricular Area Segmentation

As indicated in Table 6, the highest accuracy for landmark detection is achieved when Strategy #3 is applied, where all the training samples are processed in one single direction as well as the testing samples. The direction normalization module not only maximizes the size of training samples with the same side ear but also provides randomness to the training samples.

Further experiments are performed to find how sufficiently different strategies utilize the samples in the dataset. In the experiments reported in Table 5, the size of the training sample for Strategy #2 and Strategy #3 are set unchanged with Strategy #1 (202 training samples in total), so only the impact of the sample direction can be revealed. When all the available samples (202 left and 202 right samples) are used for training, the average NME can be improved from 0.0652±0.0034 to 0.0554±0.0011 with Strategy #2, and from 0.0514±0.0023 to 0.0500±0.0011 with Strategy #3. Among all these results, training with the direction normalization module achieved the highest accuracy.

Both tasks of landmark detection and auricular area segmentation for auriculotherapy are completed in this paper. Experiments show that the proposed method achieved an average NME of 0.0514±0.0023 for landmark detection of a total of 91 points, and the IoU for left and right ears are 0.6629±0.0458 and 0.6826±0.0103, respectively. For each testing sample, its NME and IoU' are shown in Fig. 14, where IoU'=1-IoU. Since the segmentation accuracy is negatively correlated with the localization accuracy, the area segmentation accuracy is slightly higher with right ears than with left ears.

Research reported for similar tasks are very limited. The results of our work are compared with some other research, and the results are listed in Table 10. Among these, Wen et al. [20] proposed an ASM-based method for landmark detection and auricular division. 65 points were selected by the authors according to WFAS STANDARD-002:2013 [43], and 32 divisions were divided by interpolating adjacent points on the arcs. However, only 30 images from USTB were annotated and used for model training. 10 images were tested in their paper, which was even selected in the training sets, and no quantitative results were reported for area division. As seen in their paper, only images with right-sided ears were selected, which is partly because the ASM algorithm is sensitive to the initialization. Such a problem could also benefit from the direction normalization module we introduced in the paper. Mussi et al. [18] identified the contours of anatomical regions of ears. Although the segmentation accuracy was higher than ours, however, such results were based on depth map images, which were not as convenient as color images for data collection. In [17], Zhou et al. explored the ear landmark localization results with iBug-Collection A using a different method. The best result reported was 0.0522±0.0246, which was slightly better than the result we achieved. To better evaluated the proposed landmark detection method, it was also tested on iBug-Collection A. The final result (0.0493±0.0012) showed that our method was superior to that proposed in [17]. We also tested MTCNN [25] and TCDCN [23] networks on our dataset, and the average NME were 0.0765±0.0143, and 0.0715±0.0046.

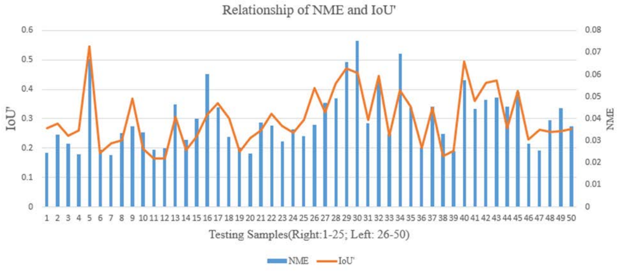


FIGURE 14. Relationship of NME and IoU'.

VI. CONCLUSION

In this paper, an end-to-end auricular acupoint localization method based on the deep model MobileNetV2 was proposed. A direction normalization module was introduced to compensate for the differences between left and right ears. A total of 91 auriculotherapy-related landmark points were detected from the image automatically and an average NME of 0.0514±0.0023 was achieved with the 5-fold cross-validation, which outperforms similar works. An annotated ear image dataset EID was collected to help studies on acupoints localization. Auricular area segmentation was also performed in this paper, and the average IoU for the target area (cymba conchae) was 0.6731±0.0233. This method can also be extended to the detailed division of the auricular areas including the triangular fossa, the helix, and other areas. Experiments indicated that by using the proposed method, acupoints could be detected in both static images and video streams in real-time, which provided the potential to identify the localization of auriculotherapy practice.

Limitation remains in this study. Firstly, the dataset contains the ear images of healthy and young students only. According to TCM experience, the auricular appearance could be different between unhealthy and healthy people. In our future work, more data will be collected with various age and health status distribution, so that an in-depth study could be performed. Secondly, the reported results for auricular area segmentation are not high. The area was segmented based on the detected landmarks. In future work, methods to improve the accuracy of landmark detection with various deep-models will be studied, and the deep learning-based segmentation algorithm will also be explored.

REFERENCES

[1] L. K. P. Suen, A. Molassiotis, C. H. Yeh, and S. K. W. Yeung, "Auriculotherapy for insomnia in elderly people: A 6 week, double-blinded, randomised pilot study," *Lancet*, vol. 390, p. S58, Dec. 2017.

[2] L. K. P. Suen, A. Molassiotis, S. K. W. Yueng, and C. H. Yeh, "Comparison of magnetic auriculotherapy, laser auriculotherapy and their combination for treatment of insomnia in the elderly: A double-blinded randomised trial," *Evidence-Based Complementary Alternative Med.*, vol. 2019, pp. 1–19, May 2019.

[3] M. C. Santos, J. R. Rothstein, and C. D. Tesser, "Auriculotherapy in obesity care in primary health care: A systematic review," *Adv. Integrative Med.*, vol. 9, no. 1, pp. 9–16, Mar. 2022.

[4] M. Murakami, L. Fox, and M. P. Dijkers, "Ear acupuncture for immediate pain relief-a systematic review and meta-analysis of randomized controlled trials," *Pain Med.*, vol. 18, no. 3, pp. 551–564, 2017.

[5] Y. Y. Xu, J. H. Liu, H. Ding, H. Tang, S. Y. Song, W. Q. Zhong, and Z. B. Pan, "Clinical research of auricular gold-needle therapy in treatment of chronic fatigue syndrome of Qi deficiency constitution," *Chin. Acupuncture Moxibustion*, vol. 39, no. 2, pp. 128–132, 2019.

[6] *Report on the Working Group on Auricular Acupuncture Nomenclature, Lyon, France, 28–30 November 1990*, World Health Organization, Geneva, Switzerland, 1991.

[7] L. Zhou and B. Zhao, *Nomenclature and Location of Auricular Points*, 1st ed. Beijing, China: Standard Publishing House, 2008.

[8] J. H. Liu, Y. Y. Xu, and G. Z. Xu, "Auricular medicine is the 'bridge' of the integrated traditional Chinese medicine and western medicine," *Chin. J. Integr. Traditional Western Med.*, vol. 39, no. 6, pp. 750–752, 2019.

[9] A. Wirz-Ridolfi, "The history of ear acupuncture and ear cartography: Why precise mapping of auricular points is important," *Med. Acupuncture*, vol. 31, no. 3, pp. 145–156, Jun. 2019.

[10] Y.-T. Yang, Q.-F. Huang, Y.-F. Jia, J. Liu, and X.-P. Ma, "Current situation and analysis of clinical application of auricular acupoint sticking therapy," *J. Acupuncture Tuina Sci.*, vol. 14, no. 2, pp. 141–148, Apr. 2016.

[11] Z. Emersic, D. Susanj, B. Meden, P. Peer, and V. Struc, "ContexedNet: Context–aware ear detection in unconstrained settings," *IEEE Access*, vol. 9, pp. 145175–145190, 2021.

[12] I. I. Ganapathi, S. Prakash, I. R. Dave, and S. Bakshi, "Unconstrained ear detection using ensemble-based convolutional neural network model," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 1, Jan. 2020, Art. no. e5197.

[13] *CasiaLabeler*. Accessed: Jun. 8, 2020. [Online]. Available: https://github.com/msnh2012/CasiaLabeler

[14] J. J. Bartko, "The intraclass correlation coefficient as a measure of reliability," *Psychol. Rep.*, vol. 19, no. 1, pp. 3–11, Aug. 1966.

[15] Q. Zhu, Z. Mu, and L. Yuan, "Corresponding keypoint constrained sparse representation three-dimensional ear recognition via one sample per person," *IET Biometrics*, vol. 11, no. 3, pp. 225–248, May 2022.

[16] S. Prakash and P. Gupta, "A rotation and scale invariant technique for ear detection in 3D," *Pattern Recognit. Lett.*, vol. 33, no. 14, pp. 1924–1931, Oct. 2012.

[17] Y. Zhou and S. Zaferiou, "Deformable models of ears in-the-wild for alignment and recognition," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 626–633.

[18] E. Mussi, M. Servi, F. Facchini, R. Furferi, L. Governi, and Y. Volpe, "A novel ear elements segmentation algorithm on depth map images," *Comput. Biol. Med.*, vol. 129, pp. 104–157, Feb. 2021.

[19] J. Lei, X. You, and M. Abdel-Mottaleb, "Automatic ear landmark localization, segmentation, and pose classification in range images," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 46, no. 2, pp. 165–176, Feb. 2016.

[20] J. Wen, M. Jiang, Y. Wang, N. Huang, and M. Gao, "An auricular division method based on ASM algorithm," *Technol. Health Care*, vol. 29, pp. 487–495, Jan. 2021.

[21] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, "RIC-UNet: An improved neural network based on UNet for nuclei segmentation in histology images," *IEEE Access*, vol. 7, p. 21 420-21 428, 2019.

[22] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3476–3483.

[23] Z. Zhang, "Facial landmark detection by deep multi-task learning," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 94–108.

[24] Y. Wu, T. Hassner, K. Kim, G. Medioni, and P. Natarajan, "Facial landmark detection with tweaked convolutional neural networks," in *Proc. IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 3067–3074, Dec. 2018.

[25] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[26] M. Kowalski, J. Naruniec, and T. Trzcinski, "Deep alignment network: A convolutional neural network for robust face alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 88–97.

[27] R. Zhang, C. Mu, M. Xu, L. Xu, and X. Xu, "Facial component-landmark detection with weakly-supervised LR-CNN," *IEEE Access*, vol. 7, pp. 10263–10277, 2019.

[28] Y. Zhao, Y. Liu, C. Shen, Y. Gao, and S. Xiong, "MobileFAN: Transferring deep hidden representation for face alignment," *Pattern Recognit.*, vol. 100, Apr. 2020, Art. no. 107114.

[29] X. Guo, S. Li, J. Yu, J. Zhang, J. Ma, L. Ma, W. Liu, and H. Ling, "PFLD: A practical facial landmark detector," 2019, *arXiv:1902.10859*.

[30] R. Chen, Y. Ma, L. Liu, N. Chen, Z. Cui, G. Wei, and W. Wang, "Semi-supervised anatomical landmark detection via shape-regulated self-training," *Neurocomputing*, vol. 471, pp. 335–345, Jan. 2022.

[31] F. Saxen, P. Werner, S. Handrich, E. Othman, L. Dinges, and A. Al-Hamadi, "Face attribute detection with MobileNetV2 and NasNet-mobile," in *Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2019, pp. 176–180.

[32] G. Perinetti, "StaTips—Part IV: Selection, interpretation and reporting of the intraclass correlation coefficient," *South Eur. J. Orthodontics Dentofacial Res.*, vol. 5, no. 1, pp. 3–5, May 2018.

[33] Z. Liu, "Application of dlib in face recognition technology," *Pract. Electron*, vol. 2020, no. 21, pp. 39–41, 2020.

[34] P. Lu, L. Tsao, C. Yu, and L. Ma, "Survey of ear anthropometry for young college students in China and its implications for ear-related product design," *Hum. Factors Ergonom. Manuf. Service Industries*, vol. 31, no. 1, pp. 86–97, Jan. 2021.

[35] F. J. M. Shamrat, A. Majumder, P. R. Antu, S. K. Barmon, I. Nowrin, and R. Ranjan, "Human face recognition applying Haar cascade classifier," in *Pervasive Computing and Social Networking*. Cham, Switzerland: Springer, 2022, pp. 143–157.

[36] S. Elaw, "Face detection in crowded human images by bi-linear interpolation and adaptive histogram equalization enhancement," *Amer. J. Comput. Sci. Technol.*, vol. 3, no. 4, pp. 68–75, 2020.

[37] G. Lou and H. Shi, "Face image recognition based on convolutional neural network," *China Commun.*, vol. 17, no. 2, pp. 117–124, Feb. 2020.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[39] *Und Biometric Dataset Collection E*. Accessed: 2002. [Online]. Available: https://cvrl.nd.edu/projects/data/

[40] D. Frejlichowski and N. Tyszkiewicz, "The west pomeranian university of technology ear database-a tool for testing biometric algorithms," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2010, pp. 227–234.

[41] *Ear Recoginition Laboratory at USTB*. Accessed: Jun. 2018. [Online]. Available: http://www1.ustb.edu.cn/resb/en/index.htm

[42] A. Amer, T. Lambrou, and X. Ye, "MDA-UNet: A multi-scale dilated attention U-Net for medical image segmentation," *Appl. Sci.*, vol. 12, no. 7, p. 3676, Apr. 2022.

[43] World Federation of Acupuncture-Moxibustion Societies(WFAS), "Auricular acupuncture point (WFAS standard-002: 2012)," *World J. Acupuncture-Moxibustion*, vol. 23, no. 3, pp. 12–21, 2013.

XIAOYAN SUN was born in Hangzhou, China, in 1980. She received the B.S. degree in biomedical engineering from Zhejiang University, in 2004, the M.S. degree in biomedical engineering from Peking University, in 2007, and the Ph.D. degree in electrical and computer engineering from Old Dominion University, Norfolk, VA, USA, in 2012. She was at Hangzhou Normal University, Hangzhou, in 2013. She is currently a Lecturer with the School of Information Science and Engineering. Her research interests include computer-aided surgery and unobtrusive health monitoring.

JIAGANG DONG was born in Wuhu, Anhui, China, in 1995. He received the B.S. degree in software engineering from the Hefei Normal University of Hefei, China, in 2018. He is currently pursuing the M.S. degree in technology of computer application with Hangzhou Normal University, Hangzhou, Zhejiang, China. His research interests include medical image processing and deep learning.

QINGFENG LI received the M.B.A. degree from Zhejiang University. Inheritor of JIAQI LI's ear acupoint diagnosis and treatment technology and the Distinguished Expert of the Engineering Research Center of Mobile Health Management System, Ministry of Education, Hangzhou Normal University. He has published many books and undertaken a number of provincial projects in ear acupoint research.

DONGXIN LU was born in 1971. He received the master's degree in computer application technology from Harbin Engineering University, in 1996, and the doctor's degree in computer application technology from the Harbin Institute of Technology, in 2000. He is the Deputy Secretary General of CCF Human–Computer Interaction Committee, an Honorary Member of CCF YOCSEF, the Former Vice Chairman of CCF Shenzhen Branch, the Vice Chairperson of CCF YOCSEF Headquarters, and the Chairperson of CCF YOCSEF Shenzhen sub forum.

ZHENMING YUAN received the Ph.D. degree from the College of Computer Science and Technology, Zhejiang University. He is currently a Professor with the College of Information Science and Engineering, Hangzhou Normal University, and the Vice Dean of the Engineering Research Center of Intelligent Healthcare, Ministry of Education, China. His main research interests include artificial intelligent in medical, machine learning, and big data mining.