IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 22, 2020, accepted October 2, 2020, date of publication October 13, 2020, date of current version October 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3030680

Ground Truth Dataset for EEG-Based Emotion Recognition With Visual Indication

GUOSHENG YANG, RUI JIAO, HUIPING JIANG, AND TING ZHANG
Brain Cognitive Computing Lab, School of Information Engineering, Minzu University of China (MUC), Beijing 100081, China

Corresponding author: Huiping Jiang (jianghp@muc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China Youth Project Funding under Grant 61503423, in part by the Leading Talent Program of State Ethnic Affairs Commission of China, and in part by the Double First-Class Special Funding of MUC.

ABSTRACT Along with extensive and successive applications, emotion recognition based on electroencephalogram has attracted more and more researchers. However, how to acquire sufficient and high-quality real emotion electroencephalogram to train emotion recognition model has always been a bottleneck issue in the electroencephalogram-based emotion recognition research field. Because the subject's emotion is easy to be affected by many factors and subjects do not always evoke emotion well, it's very hard to determine whether the real emotion electroencephalogram appears or not in the experiments. On the contrary to electroencephalogram, facial expression obtained without any intentional control is easy to be recognized by computers, and is also one of the primary and reliable cues for understanding emotions. Inspired by such a common sense, we proposed an approach to building up ground truth electroencephalogram dataset with visual indication. Firstly, the relationship between facial expression and electroencephalogram is analyzed in details from the viewpoints of biophysics and correlation. Secondly, based on the analysis result that when the subject's facial expression is evoked, it should be accompanied by the emotional electroencephalogram corresponding to it, the method of building up the ground truth electroencephalogram dataset with visual indication is put forward along with the computer automatic implementation. Thirdly, we have built up a ground truth electroencephalogram dataset which covers 3 kinds of emotions (i.e. joy, sadness, and neutral) of the subjects who are undergraduate and graduate students from Minzu University of China. Lastly, the validation of the established dataset is tested by the comparative experiments between the long short term memory emotion recognition models trained with two different datasets respectively. One dataset contains both truth and false electroencephalograms and another one only contains the truth data.

INDEX TERMS Ground truth, electroencephalogram, facial expression, electroencephalogram-based emotion recognition, visual indication, long short term memory.

I. INTRODUCTION

Emotion is the human affective response to the external environment or specific stimulation [1]. Emotional recognition is the process of identifying human emotional state [2], which is currently attractive to more and more researchers because of its extensive and successive applications, such as robotics [3], [4], rehabilitation [5], [6], marketing analysis [7], human-machine interface [8], [9], and adaptive e-learning system [10]. Up to now, various approaches are applied to deduce person's affective states from several aspects including features of behaviors, facial expression, voice intonation, etc. or from physiological signals including electroencephalogram (EEG), electrocardiograph (ECG), pulse rate, etc. Among these methods, EEG based method is extensively studied in the research field of emotion recognition because EEG detection instrument has some distinguished characteristics of mobility, non-expensive, high time resolution, and bearable space resolution. Besides of these highlights, EEG can provide more detailed and complex information with a non-invasive way for emotion recognition, and it cannot be hidden intentionally, which makes EEG based emotion recognition more effective and reliable [2].

Emotion recognition based on EEG mainly includes two modules: feature extraction and emotion classification. In the traditional methods, emotional EEG features are basically

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

extracted manually according to some rules, for example, the energy, power spectrum, and the energy ratio of different frequency bands; and these features are fed into a classifier to identify the emotional states. The classifiers are often traditional machine learning models, such as support vector machine [11], K-nearest neighbors [12], [13], etc.

As time-series signals, EEG signals reflecting brain activity are always dynamic. Although the manual feature extraction is simple in principle and is easy to be implemented, it cannot guarantee the real-time performance of the features. It may cause the loss of some feature information because traditional manual feature extraction is mostly based on the statistical properties of EEG in a sliding window with a certain size. The online performance of EEG signals is less considered.

In recent years, deep learning techniques, especially convolutional neural networks with two distinct advantages of being able to autonomously extract features from the signal, and as well as of high recognition accuracy if it is well trained with a vast quantity of labeled training data [2], have received extensive attention and achieved marvelous results in many fields, such as image recognition, natural language processing, speech recognition, etc. Activated by such attraction, many researchers have also tried to apply the deep convolutional neural network to the EEG-based emotional recognition [14], [15]. Although a lot of achievements of EEG-based emotion recognition with deep convolutional neural networks have been obtained in many application fields [16]–[19], [20], [21], such as movement imagery recognition, rehabilitation training, intelligent human-computer interaction (HCI), etc., one of the hard problems that should be addressed in EEG-based emotion recognition with deep convolutional neural networks is to acquire sufficient and high-quality ground truth training data. But it is very difficult to accurately collect the EEG ground-truth (or labeled) data corresponding to some kind of emotion induced by images or videos because of the following main reasons:

1) It is common sense that person's emotion is easy to be influenced by many factors, including social position, education background, workload, personal preference, etc. Therefore, we cannot ensure that subjects may always evoke emotion well in emotion recognition experiments. Furthermore, the subject's emotions presented by EEG signals are not always elicited when the subject is watching the stimulating images or videos.
2) Usually EEG signals are very weak with lower SNR (Signal to Noise Ratio), and have the non-linear, non-stationary, and time-varying characteristics [22]. It is difficult to collect the precisely ground truth emotion EEG data by just observation because we cannot know when exactly the real ground truth emotion is evoked, and how long it lasts even if subject's emotion is evoked.

But different from EEG, facial expression obtained without any intentional control is also one of the primary and reliable cues for human or computers to understand emotions and sentiments [23] through simply observation of the facial expressions of subjects. And it is widely believed that the appearance of facial expressions always accompanies with the EEG production synchronously.

Inspired by such a common sense, we have proposed an approach to building up "Ground Truth Dataset for EEG-based Emotion Recognition with Visual Indication" (DEVI) based on what we have done previously [24]–[26], in order to set up an accurate ground truth dataset for the training and test of the EEG-based emotion recognition models. This paper is organized as follows: we describe some works closely related with building up labeled dataset for EEG-based emotion recognition to highlight our proposal in Section II. Section III presents the relation analysis between facial expression and EEG from the viewpoints of biophysics and correlation. And then provides the method of building up the ground truth EEG dataset on the basis of above the relation analysis, and as well as the computer automatic implement method of the ground truth EEG data set. The design and implementation of the experiment are presented in Section IV. The test results of the effectiveness of established ground truth EEG data set and some discussions are presented in Section V. The summarization and conclusion are made in Section VI.

II. RELATED WORK

In this section, some works associated with building up ground truth EEG data set and emotion recognition based on the EEG and facial expression are presented to highlight our motivation.

Mohammad *et al.* [14], [27], [28] presented an approach to instantaneously detecting the emotions from EEG signals and facial expressions. The feature data of emotion is obtained according to the rule that power spectral densities (PSD) of EEG signals in different frequency bands are correlated with emotions [29]. But PSD is a statistical variable calculated during a period of time, which may largely weaken the on-line and time-varying characteristics of EEG signals. Yongrui Huang *et al.* [30] proposed a method of fusing both EEG and facial expression for multimodal emotion recognition. For EEG emotion detection, features are extracted based on PSD, and support vector machine model is employed for classification. But the EEG emotion data used to train the models is obtained by self-assessment of subjects after they watched movie clips. Thus, the on-line and time-varying characteristics of EEG signals cannot be guaranteed because of the absence of on-line evaluation.

Sander Koelstra *et al.* [31] built a multimodal data set for the analysis of human affective states, named as DEAP, which contained EEG and peripheral physiological signals, and as well as frontal face video. In the processing of constructing DEAP, participants performed a self-assessment of their levels of arousal, valence, liking, and dominance at the end of each trial. EEG signals for building EEG-based emotion recognition model are available from the DEAP, but the labeled emotion EEG data cannot be directly obtained from DEAP, and the on-line and time-varying characteristics of real EEG signals cannot be guaranteed.

Katsigiannis *et al.* [32] presented a multi-modal database containing subject EEG and electrocardiogram (ECG) signals recorded during affect elicitation by means of audio-visual stimuli, along with their self-assessment of their affective state in terms of valence, arousal, and dominance. All the signals were captured with portable, wearable, and wireless equipment, to allow the use of affective computing methods in everyday applications.

EEG signals for building EEG-based emotion recognition model are available from the proposed database, but it is not clear how to get the labeled emotion EEG from the proposed database.

To explore human emotions, Song *et al.* [33] designed and built a multi-modal physiological emotion database containing four modal physiological signals, namely EEG, galvanic skin response, respiration and electrocardiogram. Several feature extraction methods and classifiers were used to recognize the physiological responses of different emotions. The database has been made publicly available to encourage other researchers to use it to evaluate their own emotion estimation methods. But it is difficult to read the real ground truth EEG data from the data base as the label one to train the classifier.

As we state above, there are several public EEG datasets for EEG-based emotion recognition. However, only a limited amount of labeled data can be obtained to train the model. To address such issue, data augmentation method have been applied to generate EEG data recently. Some researchers have generated EEG data by applying a geometric transformation to the original data [34], [35]. Other researchers have focused on using deep generative models to generate artificial EEG data [2], [36]–[38]. Generating artificial data by applying a transformation from the original data is one of the conventional solutions to solving the data scarcity problem. EEG data augmentation can largely expand the database by adding the generated data, which is realistically like as the original data, rather than the original data.

EEG data augmentation can alleviate the overfitting of emotion recognition network parameters due to the increase of data volume. However, it does not necessarily improve the robustness and generalization of the emotion recognition network model because of two reasons. One is that the generated data is EEG pseudo-data, not real EEG data; another is that the variety of data is not increased. Therefore, collecting a large amount of real EEG emotion data is the most effective way to solving the data scarcity problem for training the generalized, and robust emotion recognition model.

All the members in our Brain Cognitive Computing Lab have been working on brain cognitive computing since 2009. Our research works are specializing in building up induced file database for generating different emotions, EEG database, and facial expression database, and in researching on emotion recognition, and intelligent HCI. Based on the study that we have done previously [24]–[26], we have proposed the approach to building up DEVI to construct ground truth dataset for EEG-based emotion recognition.

III. METHODOLOGY OF THE PROPOSED APPROACH

A. RELATION ANALYSIS BETWEEN FACIAL EXPRESSION AND EEG

Where does facial expression come from? When receiving an external (or internal) stimulus, human brain makes a judgment about the stimulus, and then generates emotions. At the same time, brain commands the facial muscles to move in proportion, and to produce facial expression which is the external manifestation of the brain neural activities.

Where does EEG come from? One standard view from a biophysics perspective is that an EEG is an extracellular current that reflects the sum of the postsynaptic potentials of thousands or even millions of identical pyramidal cells [39].

What relations are there between facial expression and EEG?

Functional imaging studies have shown that blood flow to the amygdala is accelerated and activated when recognizing fearful and sad faces [40], [41]. Phillips *et al.* [42] found that fearful faces activate the amygdala but not the insula. The amygdala locates in the temporal pole of the temporal lobe, distributes in BA38 area of Brodmann partition [54], and corresponds to T7 and T8 electrode positions of the International 10-20 System (10-20S). It is the part of the limbic system that generates, recognizes and regulates emotions. The insula locates in the depth of the lateral sulcus, distributes in BA13, BA14, BA15 and BA16 of Brodmann partition [54], and corresponds to FT7, FT8, FT9, FT10, TP7, TP8, TP9 and TP10 electrode positions of 10-20S. It is not only involved in the acquisition of basic emotions and perceptual information, but also in the cognitive process of higher emotions.

Killgore *et al.* [43] used fMRI to investigate neural activity in the amygdala and anterior cingulate gyrus when subjects unconsciously perceived happy or sad faces. The results suggest that the anterior cingulate gyrus and amygdala may play an important role in the subliminal awareness of emotional information. The cingulate gyrus distributing in BA23, BA24, BA26, BA29, BA30, BA31 and BA32 of Brodmann partition [55] and corresponding to Cz, Fpz, Fz, Cz and Pz electrode positions of 10-20S, is closely linked to emotions.

Mayberg [44] found that when healthy subjects tried to induce sadness, blood flow in the prefrontal cortex decreased. Ploghaus [45], [46] found that activation in this area increased when it induced painful anticipation. Spregelmeyer *et al.* [47] found that the dorsolateral prefrontal cortex was activated in all facial expressions of different negative emotions. Therefore, the prefrontal cortex may participate in general emotional processing, recognize the emotional meaning of stimuli, and regulate emotional reactions, emotional experiences and emotional behaviors dominated by the autonomous nervous system.

The prefrontal cortex, locating in the front of the frontal lobe, distributing in BA9, BA10, and BA11 of Brodmann partition [54], corresponding to F7, F8, F9, F10, AF7, and AF8 electrode positions of 10-20S, is an important brain tissue for information storage and processing, and is closely related to the production of emotions.

In addition, many research results have demonstrated the correlation between facial expression and EEG. For Example, Sun *et al.* [48] presented a framework to measure the correlation between spontaneous human facial affective expressions and relevant brain activity. The experimental results show strong correlation between the spontaneous facial affective expressions and the affective states related brain activity. Chakraborty *et al.* [49] simultaneously sampled EEG and facial expression of the subjects when they are watching the specific audio-visual stimulus. The nonlinear-correlation from EEG to the facial expression or from the facial expression to EEG is obtained by employing feed-forward neural network trained with back-propagation algorithm. A corpus creation of spontaneous facial expressions was obtained by Zatarain-Cabada Ramón *et al.* [50] through the recognition of EEG signals, and the facial expression classification in different categories such as boredom, excitement, etc.

From the brain function description, functional imaging research results, and correlation analysis between EEG and facial expression presented above, it can be seen that the production of facial expressions is always accompanied by physiological activities mainly executed in the relevant brain regions. This leads to the delayed continuous firing of large numbers of neurons in the related brain regions, which forms the local field potential (LFP) along with the generation of EEG [51]. Thus, we can make a conclusion that a subject's facial expression is evoked when the subject is shown emotional stimulated pictures or videos, the facial expression is always accompanied by the production of the emotional EEG. This is the basis for our proposed approach.

B. METHOD OF BUILDING UP EEG GROUND TRUTH DATA SET

Let a subject view a piece of stimulating movie clip lasting no more than 1 minutes [53], and synchronously sample the EEG signals and the facial expression signals. The sampling rates are f_1 for EEG, and f_2 for facial expression, where $f_1 = nf_2$, and n is a positive integer. The sampled raw EEG signal is expressed as $eeg\,(i, j, k, g)$ (where i represents the i^{th} subject, j for the j^{th} trial, k for the k^{th} kind of emotions $\in$ (joy, sadness, neutral), g for the g^{th} sampling), and the sampled raw facial expression is expressed as $fe\,(i, j, k, v)$ (v for the v^{th} sampling). All the signals of $eeg\,(i, j, k, g)$ and $fe\,(i, j, k, v)$ of the different subjects with different emotions in all trails are stored into EEG-A (the EEG data set including all ground truth EEG data and non-ground-truth EEG data) and FE-A data set (facial expression data including all ground truth facial expression data and non-ground truth facial expression data) respectively. Following the construction of data sets of EEG-A and FE-A, there are several steps to finish the construction of EEG-L data set (EEG data set only including ground truth EEG data).

Firstly, we develop a software tool with Python to segment the facial expression video $fe\,(i, j, k, v)$ from FE-A into separated sequence frames. The segment software tool is named as SVSF (segment video into separated frames), and the separated sequence frame is represented as $fesf\,(i, j, k, d)\,, d \in [1, 60 \times f_2]$.

Secondly, observe $fesf\,(i, j, k, d)$ to determine the special frame (named as frame-s, for example, given the m^{th} frame, $m \in [1, 60 \times f_2]$) from which the k^{th} emotion begin to appear, and another special frame (named as frame-e, for example, given q^{th}frame, $q > m \in [1, 60 \times f_2]$) from which the k^{th} emotion disappears. The frames $fesf\,(i, j, k, d)$between frame-s and frame-e are thought of the ground truth facial expression data for the k^{th} emotion, denoted as $fe_{se}\,(i, j, k, b)\,, b \in [m, m{+}1, \ldots, q]$, and stored into data set FE-L (facial expression data set only including ground truth facial expression data).

Lastly, the frame-s $fe_{se}\,(i, j, k, m)$ corresponding to the time point TSP $= m \times 1/f_2$ aligns to the EEG special sampling point from which the EEG signal of the k^{th} emotion real state starts to appear. The frame-e $fe_{se}\,(i, j, k, q)$ corresponding to the time point TEP $= q \times 1/f_2$ aligns to the EEG sampling point from which the EEG signal of the k^{th}emotion real state disappears. The EEG signals $eeg\,(i, j, k, g)$ between TSP and TEP are thought of the ground truth EEG data for the k^{th} emotion, denoted as $eeg_{se}\,(i, j, k, t)\,, t \in [TSP, TSP + 1/f_1, TSP + 2/f_1 \ldots, TEP]$, and stored into EEG-L.

Just as the procedure of building up the ground truth EEG data set of emotions stated above, all of such ground truth EEG data from different subjects with different emotions (joy, sadness, and neutral) in all trails are collected and stored into the ground truth EEG data set EEG-L, used to train the model for emotion recognition.

The whole procedure of building up the ground truth data set of emotions is shown as Fig.1 [26].

C. IMPLEMENTATION METHOD OF BUILDING UP EEG GROUND TRUTH DATA SET

Based on the method of building up ground truth EEG data set, we develop a software tool with Dlib (http://dlib.net/)to automatically generate EEG-L from EEG-A. We call it as BEEGLDS (Build EEG Labeled Data Set), which mainly consists of 5 components described as the following.

1) Image normalization

By using linear transformation, the original image is enlarged or reduced to be an image with uniform width and height.

2) Face alignment

By using affine transformation, the original image is rotated to put the head in the upright place, and the face is translated to the center of the image by using translation transformation.

3) Facial feature vector construction

For the i^{th} subject, the j^{th} trial, and the k^{th} emotion, extract 68 face key points of the frame $fesf\,(i, j, k, d)$,

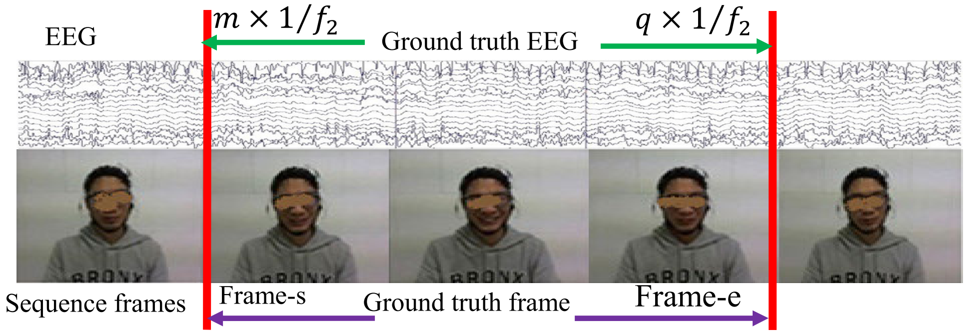

FIGURE 1. The whole procedure of setting up ground truth data set of emotion state.

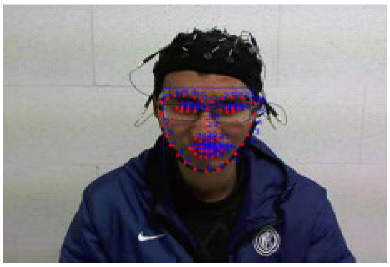

FIGURE 2. Extract 68 face key points by calling Dlib library (Derived from [26]).

$d \in [1, 60 \times f_2]$ (shown as Fig. 2) by calling Dlib library, and construct facial feature vector $\mathbf{V_{fe}}\ (i, j, k, d)$ by cascading the 68 face key points.

4) Emotion classifier and FE-L construction

Select the linear kernel function as the kernel function of SVM (supporting vector machine) to construct emotion classifier based on SVM. Apply the emotion classifier on $\mathbf{V_{fe}}\ (i, j, k, d)$ to recognize the special frame frame-s m^{th} and special frame frame-e $q^{th}$of the k^{th} emotion. Determine the frames $fe_{se}\ (i, j, k, b)$,$b\ \in\ [m, m+1, \ldots, q]$ and store them into FE-L.

5) EEG-L Construction

According to$fe_{se}\ (i, j, k, m)$ and $fe_{se}\ (i, j, k, q)$, calculate TSP $=m\times\frac{1}{f_2}$ and TEP $=q\times\frac{1}{f_2}$. Determine the ground truth EEG signals $eeg_{se}\ (i, j, k, t)\, t\ \in\ [TSP, TSP + 1/f_1, TSP + 2/f_1 \ldots, TEP]$ and store them into EEG-L.

D. METHOD of TESTING EEG-L

In order to test the effectiveness of EEG-L, we need to designate an EEG data set to be compared with EEG-L. To this end, we stochastically divide EEG-A data set into two subsets in a ratio of 7 to 3. The first subset is called as EEG-A-T (the training subset of EEG-A) used to train a model, the second subset is called as EEG-A-V (the verification subset of EEG-A) used to verify the model. Two EEG based emotion recognition models with same structure but different parameters are employed and trained. In the same training conditions (the learning rate, the loss function, batch size, and initial parameters are kept unchanged), we train the two models with EEG-L and EEG-A-T respectively. The model trained with EEG-L is called Model-L, while the model trained with the training data subset of EEG-A-T is called Model-A. When the two models are applied to the same test data set, we can determine which one is better between the two models by comparing the output performance of two models. If Model-L is better than model-A, there is no question that dataset EEG-L is more effective than dataset EEG-A-T in training EEG-based emotion recognition model because the only difference between the two models is that the parameters are trained with different data sets. Here, the output performance of the two models are described with recognition accuracy and stationarity.

In order to compare the recognition accuracy and stationarity of the two models, let two trained-well models of Model-A and Model-L be performed independently on the same testing data subset of EEG-A-V. It is easy to make a decision that who is better between Model-A and Model-L by comparing the recognition results of the two models. Thus, it is convinced to make a conclusion that EEG-L is more effective if the output performance of Model-L is better than that of the Model-A.

IV. EXPERIMENT

A. STIMULUS VIDEO SELECTION

In building EEG ground truth data set, video was selected as stimulus material. In the preliminary study, joy, sadness, and neutral emotions are considered. 30 video clips containing such emotions are manually selected from comedy, disaster, and documentary films. These film genres are mostly likely to elicit subject's emotions.

Selecting the most effective stimulus materials is crucial to eliciting emotional reactions from human subjects. It takes a certain amount of time to induce a subject's emotions. The induction time is too short to induce the subjects' emotions, but induction with too long time is easy to make the subjects tired. Psychologists recommended videos from 1 to 10 minutes long for elicitation of a single emotion [28], [56]. In general, the induction video should be as short as possible to avoid visual fatigue while keeping it long enough to ensure the effective eliciting emotional reactions. Here, the length of each movie clip is kept no more than 1 minutes [53].

Based on the preliminary study, 30 video clips were shown to more than 80 participants who are undergraduate and graduate students from Minzu University of China. The participants were asked to self-assess their emotion for each video clip by immediately rating the scale of arousal (ranging from calm to excited/activated) and valence (ranging from unpleasant to pleasant) in the Nine-point scale Form, shown as Table 1 and table 2. Prior to the questionnaire survey, all participants are given the meaning of different scales for self-assessment.

TABLE 1. Arousal nine-point scale form for video clip.

	Arousal / Video Name	Calm → Attention → Excited								
		1	2	3	4	5	6	7	8	9
Joy	Lost on journey									
	Crazy stone									
	A Chinese odyssey									
	Shaolin soccer(1)									
	Three Idiots									
	…									
Sadness	Hachiko: A Dog's Story									
	Titanic									
	Dad's lies									
	Beijing Love Story									
	Dead Poets Society									
	…									
Neutral	Tencent - Blue Planet II									
	Tencent - Great Wall									
	…									

According to Table 1 and Table 2, calculate the statistical average scales of the arousal and valence respectively for each video clip. 12 video clips were chosen based on the calculating results with the average scales of arousal and valence no less than 7 respectively, shown as Table 3.

B. PREPARATION of INDUCTION FILE

E-prime is a set of experimental generation system for computerized behavior research jointly developed by Carnegie Mellon University, the Center for Learning, Research and Development at the University of Pittsburgh, and Psychology software tools INC. It allows you to present a video stimulus file you've written, to provide specific time information and experimental details, and to keep the presentation of stimulus files in sync with the refreshment of the screen. Therefore, this experimental generation system is used to design experiments and play video stimulation files which are called as program of induction files.

At the very beginning of each stimulus video clip to be played, a 3-seconds red cross appears at the center of the screen to prompt the subject to pay attention, and then a 1-minute stimulus video clip (randomly selected from the 12 chosen stimulus video clips) is played on the screen.

TABLE 2. Valence nine-point scale form for video clip.

	Valence / Video Name	Unpleasant → little pleasant → Pleasant								
		1	2	3	4	5	6	7	8	9
Joy	Lost on journey									
	Crazy stone									
	A Chinese odyssey									
	Shaolin soccer(1)									
	Three Idiots									
	…									
Sadness	Hachiko: A Dog's Story									
	Titanic									
	Dad's lies									
	Beijing Love Story									
	Dead Poets Society									
	…									
Neutral	Tencent - Blue Planet II									
	Tencent - Great Wall									
	…									

TABLE 3. Information on selected video clips.

Emotion	Excerpt's source	Starting time	Ending time
Joy	Lost On Journey	0:44:07	0:45:07
Joy	Shaolin Soccer(1)	0:22:56	0:23:56
Joy	Crazy Stone	0:46:56	0:47:56
Joy	A Chinese Odyssey	0:41:33	0:42:33
Joy	Flirting Scholar	0:31:13	0:32:13
Joy	Shaolin Soccer(2)	0:44:43	0:45:43
Sadness	Hachiko: A Dog's Story	1:22:05	1:23:05
Sadness	Titanic	2:47:44	2:48:44
Sadness	Beijing Love Story	1:49:36	1:50:36
Sadness	Dad's lies	0:00:10	0:01:10
Sadness	Dearest	0:17:29	0:18:29
Sadness	Care for the elderly -- Package	0:00:16	0:01:16
Neutral	Tencent - Documentary blue Planet II	0:00:25	0:01:25
Neutral	Tencent - Documentary Great wall	0:00:09	0:01:09

After that, an instruction appears on the screen to prompt the subjects to press the button for self-assessment according to their emotional experience (positive, or negative, lasting about 27 seconds). Finally, the 30-seconds black screen was used to remove the influence of the last induction material. Such a process is called a trial. Followed on is the next trial till this group of experiment is finished (each group consists of 12 trials). The induction paradigm is shown in Fig.3. The whole process is programed into an induction file with the E-prime.

C. EXPERIMENT ENVIRONMENT

In our brain cognitive laboratory, the electrode Synamps2 amplifier connected with a 64-electrodes cap and a computer (named as computer 1) by USB interfaces is used to sample EEG signals with the sampling frequency 512 Hz. The

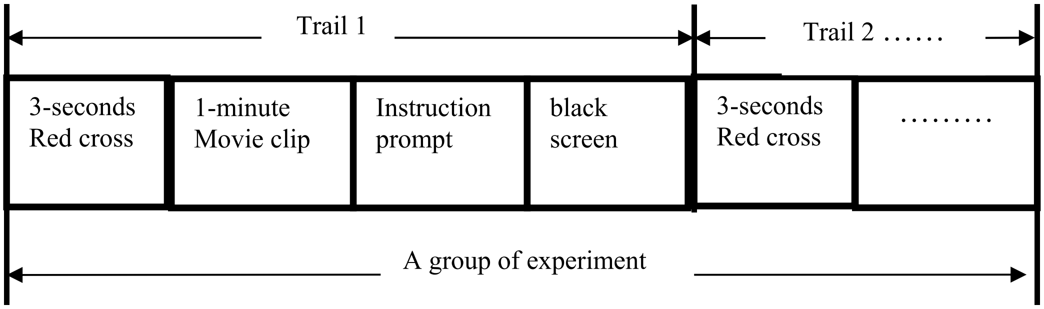

FIGURE 3. The induction paradigm.

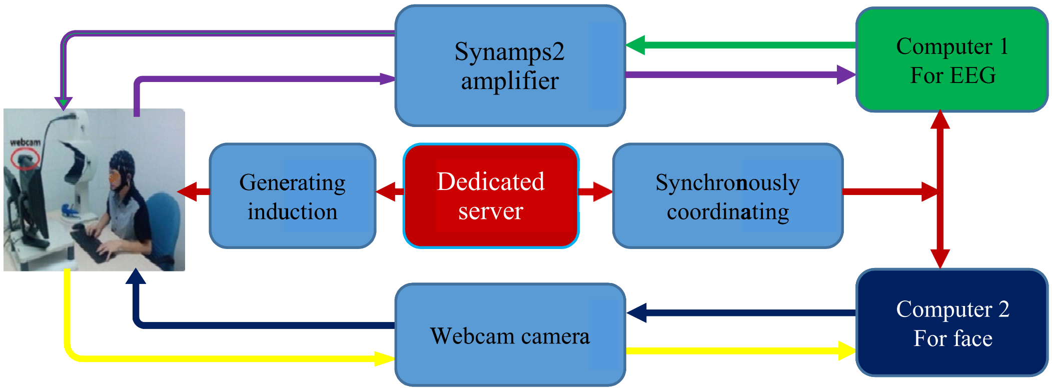

FIGURE 4. Experiment environment.

electrode distribution of the electrode cap is based on the 10/20 system electrode placement method commonly used at present. Scan4.5 software developed by Neuroscan Company is employed to sample and collect EEG signals.

Web camera (Logitech Carl Zeiss Tessar with 1920∗1080 resolution) and another computer (named as computer 2) along with the sampling program developed by us are employed to sample facial expression with the sampling frequency 32 fps (frame per second). A dedicated server is applied to run the induction files developed with E-prime according to the description in part B of Section IV, and to concurrently run another program (also developed by us) of coordinating computer1 and computer2 in order to get the records of EEG and facial expression synchronously.

A screen is fixed in front of a subject, 0.5 meters far away from the subject. The induction file is played on the screen, and the video sound is played simultaneously by the dual-channel audio system. The camera is placed directly above the screen, about 50 centimeters away from the subject. The whole structure of the experiment environment is shown in Fig.4.

D. EXPERIMENTAL PROTOCOL

This study protocol was approved by the institutional review boards (ECMUC2019008CO) at Minzu University of China. 24 undergraduate and graduate students from Minzu University of China are invited to participate in the experiment. They are half male and half female, aged between 21 and 25 years old, with normal vision or corrected vision, right-handedness, and no history of neurological injury or mental illness. They are seated in the font of the screen, approximately away 0.5 meter from the screen [50], [52].

Prior to the experiment, firstly, all subjects provide IRB-approved written informed consent after they are given an explanation about the experimental procedure. Next, they are given a set of instructions about the experimental protocol and the meaning of the feedback buttons. Following that, one subject is brought into the experimental room to start experiments when all instructions are clear to him.

In the experimental room, when the 64-electrodes cap is put on the subject's head and all electrode signals are well checked, the subject is taught with the teaching induction file (similar to the formal induction file developed by us in advance) to familiarize themselves with the system. The formal experiment starts after all these preparation works are done, and subject operates following the instruction of the formal induction file.

E. BUILD UP EEG GROUND TRUTH DATA SET EEG-L

For each trial, $eeg\,(i, j, k, g)$ and $fe\,(i, j, k, v)$ are synchronously collected with the given sampling frequency ($f_1 = 512Hz$ for EEG, $f_2 = 32Hz$ for facial expression) under the coordination of the dedicated server. According to the method and the implementation of building up EEG ground

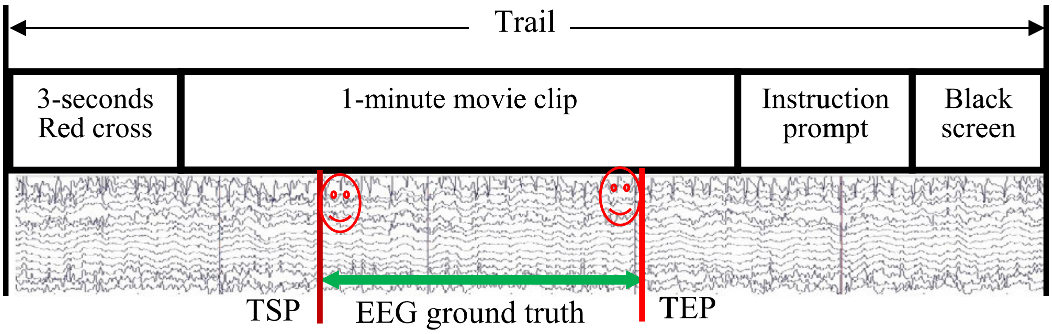

FIGURE 5. The procedure of getting EEG ground truth from one trail.

truth data set presented in Section 6, computer software tool SVSF is performed on $fe\,(i, j, k, v)$ to segment the facial expression images into $60 \times 32 = 1920$ separated sequence frames $fesf\,(i, j, k, d)\,, d \in [1, 1920]$. And then the computer software tool BEEGLDS is performed on the $fesf\,(i, j, k, d)\,, d \in [1, 1920]$ to recognize the special frame frame-s m^{th} and special frame frame-e q^{th} of the k^{th}emotion. The frames $fe_{se}\,(i, j, k, b)\,, b \in [m, m+1, \ldots, q]$ are taken as the facial expression ground truth data, and store them into the facial expression ground truth data set FE-L.

In terms of $fe_{se}\,(i, j, k, m)$ and $fe_{se}\,(i, j, k, q)$, BEEGLDS calculates TSP $= m \times 1/f_2$ and TEP $= q \times 1/f_2$, and take the EEG signals $eeg_{se}\,(i, j, k, t)\, t \in [TSP, TSP + 1/f_1, TSP + 2/f_1 \ldots, TEP]$ as the ground truth EEG data and store them into EEG-L. The whole procedure in detail is shown as Fig.5.

V. EEG-L VALIDATION TESTING AND DISCUSSION

A. TRAIN MODEL-L AND MODEL-A

According to the description presented in the part D of Section III, Both Model-L and Model-A are long short term memory (LSTM) Networks. They are employed as the emotion recognition models to test the effectiveness of the data set EEG-L. Each of Model-L and Model-A has four layers, and each layer has 32 hidden nodes. On the Tensorflow development platform with the Python language, Model-L is trained with the EEG-L data set, and Model-A is trained with the training data subset of EEG-A-T.

For training Model-L, read EEG training data indexed by subjects in EEG-L. For each subject, take 64-electrodes EEG data as a one-step read, and continuously read the EEG data at the 10 consecutive sampling points. All these EEG data form a 64×10 matrix which is taken as the input to be fed into Model-L once a time. In the training process, the learning rate is set to 0.05, the loss function is the cross entropy function, and dropout rate is set to 0.5 to avoid overfitting.

In the experiment, split the EEG ground truth training data of one subject into a series of batches (each batch with size of 64) in term of the input dimension of the model, and input them in turn to train the Model-L. And then, calculate the accuracy rate of the trained Model-L on the current training data set when the training is finished. The training results are shown as Fig.6 [26], where the vertical axis represents the accuracy rate, and the horizontal axis represents the training number (epoch).

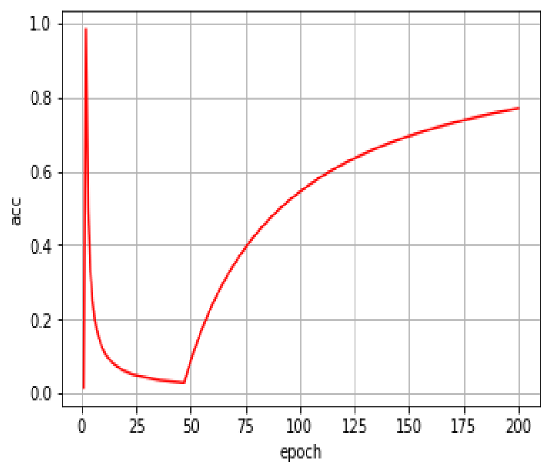

FIGURE 6. The trained result of Model-L.

The same process is performed on the training data subset EEG-A-T of EEG-A to train the Model-A. The training results are shown as Fig.7 [26], where the vertical axis represents the accuracy rate, and the horizontal axis represents the training number (epoch).

B. TEST THE DATA SET EEG-L

To test the EEG-L effectiveness, the EEG data of 8 subjects were randomly selected from the subset EEG-A-V of EEG-A. They are employed as the testing data. The trained-well models of Model-A and Model-L are performed independently on the same testing EEG data of a subject. This process is called as one experiment. One subject's EEG data is only employed for one experiment. Therefore, there are 8 experiments in total. After each experiment, the emotional classification accuracy of each model is calculated for each subject, and the computing results of the two models on the same testing data set are shown as Table 4 and Table 5 [26].

According to Table 4 and Table 5, the average of classification accuracy and the variance of accuracy for each model

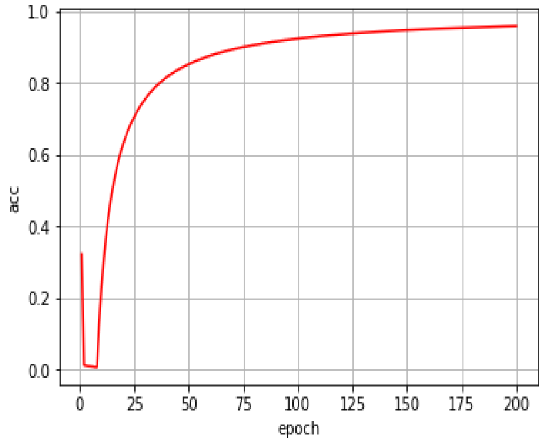

FIGURE 7. The trained result of Model-A.

TABLE 4. The classification results of trained–well Model-A.

Subject	Accuracy
Subject 1	84.7189%
Subject 2	84.5686%
Subject 3	96.0394%
Subject 4	84.5686%
Subject 5	90.3630%
Subject 6	93.0190%
Subject 7	81.9126%
Subject 8	76.6449%

TABLE 5. The classification results of trained–well Model-L.

Subject	Accuracy
Subject 1	85.0061%
Subject 2	76.9473%
Subject 3	94.3200%
Subject 4	90.9286%
Subject 5	94.3863%
Subject 6	94.3485%
Subject 7	88.4661%
Subject 8	90.9286%

are calculated, and the computing results of the two models are shown as Table 6 [26].

As can be seen from the Table 6, the accuracy of emotion recognition of Model-L is 89.4164% higher than 86.4794% of Model-A. The variance of the accuracy of emotion recognition of Model-L is 0.0036 lower than 0.0039 variance of the accuracy of emotion recognition of Model-A. This experimental results demonstrate that the output performance of Model-L is better than that of Model-A. In other words, dataset EEG-L is more effective than dataset EEG-A-T in training EEG-based emotion recognition model. Therefore, the proposal of building up EEG ground truth dataset for EEG-based emotion recognition with visual indication is effective.

TABLE 6. Results of the two models for emotion recognition.

Model	Average Accuracy	Variance of Accuracy
Model-A	86.4794%	0.0039
Model-L	89.4164%	0.0036

C. DISCUSSION

1) The accuracy and the stationarity of emotion recognition have strong relations with the EEG-based emotion recognition model. The higher the quality of the model is, the higher the emotion recognition accuracy will be. When the structure of the model is fixed, the quality of the model mainly depends on the model parameters. In general, the parameters of the model are obtained through training with a large amount of data. Therefore, the EEG based emotion recognition model has strong relations with the EEG training data set. A large amount of ground truth EEG training data can guarantee the high quality of the EEG-based model. While EEG signals that do not contain emotional information or do not show significant emotional characteristics degraded the performance of emotion recognition model, which leads to lower recognition accuracy and the poor stationarity. By using of the method proposed in this paper, the non-ground truth EEG data is filtered out, the real and reliable EEG signals are collected to form the EEG ground truth data set used to train the deep model for emotion recognition based on EEG. The accuracy of the recognition model is obviously enhanced as a result. At the same time, the stationarity of the model is also obviously improved because of the variance decrease of emotion recognition accuracy.

2) In building up the ground truth dataset for EEG-based emotion recognition with facial expression indication, facial expression recognition is a very important link, which affects the effectiveness of EEG-L. The higher the accuracy of facial expression recognition is, the more efficiency of EEG-L will be. The accuracy of facial expression is affected by many factors, among them occlusion (resulted from some shelfs, for

example, wearing spectacles) is one of the most prominent, and it is a hot research topic. Many methods are put forward to deal with such a problem, and many progresses have been achieved [59-61].

In this paper, we are mainly focus on verifying the effectiveness of our proposed method. The refinement of the proposed approach, such as considering facial expression recognition with spectacles, is our future research work to be presented in another paper. Based on this idea, Dlib+SVM, one of the many emotional recognition algorithms, is selected to verify our idea because of its simplicity and ease of implementation.

As for the occlusion, we try to invite subjects with normal vision or with the contact lenses to participate experiments, which can avoid defeating emotional recognition software by wearing spectacles.

3) It is possible that the face of a subject can be weakly deformed because of wearing the electrode cap. The weakly facial deformation has some negative influence on the accuracy of facial expression recognition, which leads to the non-ground-truth EEG data arising in the dataset EEG-L. In this case, the facial expression with weak deformation is taken as noised training data, which is used to train the kernel function of SVM classifier with end-to-end learning so that it can tolerate the micro-deformation to some extent, and can also recognize the real expression from the micro-deformation facial expression.

4) Up to now, there are only 3 types of emotion data of joy, sadness, and neutral in EEG-L. This is not enough for scientific research in terms of quantity and variety. At present, EEG-L is improving by adding other emotions such as amusement, fear, horror, surprising, boredom, etc. If you are interesting in EEG-L, or need EEG-L for your academic research rather than commerce, you can obtain the relevant data from EEG-L by contacting the corresponding author of this paper.

VI. CONCLUSION AND FUTURE WORK

In this paper, based on the result derived from the biophysics and correlation analysis that when the subject's facial expression is evoked, it should be accompanied by the emotional EEG corresponding to it, we proposed an approach to building up ground truth dataset for EEG-based emotion recognition with visual indication. We set up a ground truth EEG dataset EEG-L with the computer tools VSF and BEEGLDS developed by us according to the proposed approach. This preliminary EEG-L covers joy, sadness, and neutral emotions. The experimental results clearly show that the emotion recognition accuracy and stationarity of the LSTM emotion recognition model trained with EEG-L can be enhanced, comparing to that of the same model trained with EEG data subset EEG-A-T. In other words, the ground truth EEG data set EEG-L is effective to enhance the performance of the EEG-based emotion recognition model. Thus, there is no doubt to say collecting EEG signals with visual indication is one of the effective ways to acquiring sufficient and high-quality ground truth EEG training data.

Based on current research, further research works should be addressed as the following:

1) The refinement of the proposed approach, such as considering facial expression recognition with spectacles, and facial micro-deformation resulted from wearing electrode cap, is one of our future research works.
2) EEG-L should be improving by adding other emotions such as horror, surprising, boredom, etc. for various academic research. In addition, more subjects should be invited to participate the experiment, to enhance EEG-L in terms of quantity and variety, which can deal with the influence of individual difference on the accuracy of emotion recognition.

ACKNOWLEDGMENT

The authors would like to thank all subjects who participated in this research and the technical support from FISTAR Technology Inc. s

REFERENCES

[1] G. Chen, X. Zhang, Y. Sun, and J. Zhang, "Emotion feature analysis and recognition based on reconstructed EEG sources," *IEEE Access*, vol. 8, pp. 11907–11916, 2020, doi: 10.1109/ACCESS.2020.2966144.

[2] F. Wang, S.-H. Zhong, J. Peng, J. Jiang, and Y. Liu, "Data augmentation for EEG-based emotion recognition with deep convolutional neural networks," in *Proc. Int. Conf. Multimedia Modeling*, Bangkok, Thailand, 2018, pp. 82–93.

[3] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks," *Sci. Rep.*, vol. 6, no. 1, p. 38565, Dec. 2016, doi: 10.1038/srep38565.

[4] T. Carlson and R. J. D. Millán, "Brain-controlled wheelchairs: A robotic architecture," *IEEE Robot. Autom. Mag.*, vol. 20, no. 1, pp. 65–73, Mar. 2013.

[5] C. Pandarinath, P. Nuyujukian, C. H. Blabe, B. L. Sorice, J. Saab, F. R. Willett, L. R. Hochberg, K. V. Shenoy, and J. M. Henderson, "High performance communication by people with paralysis using an intracortical brain-computer interface," *eLife*, vol. 6, Feb. 2017, Art. no. e18554, doi: 10.7554/eLife.18554.

[6] K.-S. Hong and M. J. Khan, "Hybrid brain–computer interface techniques for improved classification accuracy and increased number of commands: A review," *Frontiers Neurorobotics*, vol. 11, p. 35, Jul. 2017, doi: 10.3389/fnbot.2017.00035.

[7] D. McDuff, R. E. Kaliouby, J. F. Cohn, and R. W. Picard, "Predicting ad liking and purchase intent: Large-scale analysis of facial responses to ads," *IEEE Trans. Affect. Comput.*, vol. 6, no. 3, pp. 223–235, Jul. 2015.

[8] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain–computer interface," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 44, pp. E6058–E6067, 2015.

[9] Y. Li, J. Pan, F. Wang, and Z. Yu, "A hybrid BCI system combining P300 and SSVEP and its application to wheelchair control," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3156–3166, Nov. 2013, doi: 10.1109/TBME.2013.2270283.

[10] A. Sun, Y.-J. Li, Y.-M. Huang, and Q.Li, "Using facial expression to detect emotion in e-learning system: A deep learning method," in *Proc. Int. Symp. Emerg. Technol. Educ.*, 2017, pp. 446–455.

[11] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, Apr. 2014.

[12] M. Murugappan, N. Ramachandran, and Y. Sazali, "Classification of human emotion from EEG using discrete wavelet transform," *J. Biomed. Sci. Eng.*, vol. 3, no. 4, pp. 390–396, 2010.

[13] M. S. Özerdem and H. Polat, "Emotion recognition based on EEG features in movie clips with channel selection," *Brain Informat.*, vol. 4, no. 4, pp. 241–252, Dec. 2017.

[14] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Trans. Affect. Comput.*, vol. 7, no. 1, pp. 17–28, Jan. 2016.

[15] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.

[16] S. Alhagry, A. Aly, and R. A.Elkhoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 355–358, 2017.

[17] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul./Sep. 2020.

[18] S.-E. Moon, S. Jang, and J.-S. Lee, "Convolutional neural network approach for EEG-based emotion recognition using brain connectivity and its spatial information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 2556–2560.

[19] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, Jul. 2017.

[20] S. B. Wankhade and D. D. Doye, "IKKN predictor: An EEG signal based emotion recognition for HCI," *Wireless Pers. Commun.*, vol. 107, no. 2, pp. 1135–1153, Jul. 2019.

[21] F. Cavallo, F. Semeraro, L. Fiorini, G. Magyar, P. Sinčák, and P. Dario, "Emotion modelling for social robotics applications: A review," *J. Bionic Eng.*, vol. 15, no. 2, pp. 185–203, Mar. 2018.

[22] J. Fedjaev, "Decoding EEG brain signals using recurrent neural networks," M.S. thesis, Dept. Elect. Eng., Technische Universitat München, Munich, Germany, 2017.

[23] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Inf. Fusion*, vol. 37, pp. 98–125, Sep. 2017.

[24] Y. Lu, H. Jiang, and W. Liu, "Classification of EEG signal by STFT-CNN framework: Identification of right-/left-hand motor imagination in BCI systems," in *Proc. 7th Int. Conf. Comput. Eng. Netw. PoS(CENet)*, Shanghai, China, Jul. 2017, pp. 22–23.

[25] Z. Zhou, "Research on EEG signal characteristic representation in emotion recognition," M.S. thesis, School Inf. Eng., Minzu Univ. China, Beijing, China, 2015.

[26] Y. Lu, "Research on several problems of affective computing based on deep Learning," M.S. thesis, School Inf. Eng., Minzu Univ. China, Beijing, China, 2019.

[27] M. Soleymani, S. Asghari-Esfeden, M. Pantic, and Y. Fu, "Continuous emotion detection using EEG signals and facial expressions," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.

[28] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.

[29] R. J. Davidson, "Affective neuroscience and psychophysiology: Toward a synthesis," *Psychophysiology*, vol. 40, no. 5, pp. 655–665, Sep. 2003.

[30] Y. Huang, J. Yang, P. Liao, and J. Pan, "Fusion of facial expressions and EEG for multimodal emotion recognition," *Comput. Intell. Neurosci.*, vol. 2017, pp. 1–8, Sep. 2017.

[31] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan./Mar. 2012.

[32] S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 98–107, Jan. 2018.

[33] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "MPED: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12177–12191, 2019.

[34] M. M. Krell and S. K. Kim, "Rotational data augmentation for electroencephalographic data," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 471–474.

[35] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proc. IEEE*, vol. 103, no. 6, pp. 871–890, Jun. 2015.

[36] K. Gregor Hartmann, R. Tibor Schirrmeister, and T. Ball, "EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals," 2018, *arXiv:1806.01875*. [Online]. Available: http://arxiv.org/abs/1806.01875

[37] Y. Luo and B.-L. Lu, "EEG data augmentation for emotion recognition using a conditional Wasserstein GAN," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 2535–2538.

[38] Y. Luo, L.-Z. Zhu, Z.-Y. Wan, and B.-L. Lu, "Data augmentation for enhancing EEG-based emotion recognition with deep generative models," 2020, *arXiv:2006.05331*. [Online]. Available: http://arxiv.org/abs/2006.05331

[39] M. X. Cohen, "Where does EEG come from and what does it mean?" *Trends Neurosci.*, vol. 40, no. 4, pp. 208–218, Apr. 2017.

[40] H. C. Breiter, N. L. Etcoff, P. J. Whalen, W. A. Kennedy, S. L. Rauch, R. L. Buckner, "Response and habituation of the human amygdala during visual processing of facial expressions," *Neuron*, vol. 17, no. 5, pp. 875–887, 1996, doi: 10.1016/S0896-6273(00)80219-6.

[41] J. S. Morris, C. D. Frith, D. I. Perrett, D. Rowland, A. W. Young, A. J. Calder, and R. J. Dolan, "A differential neural response in the human amygdala to fearful and happy facial expressions," *Nature*, vol. 383, no. 6603, pp. 812–815, Oct. 1996.

[42] M. L. Phillips, A. W. Young, C. Senior, M. Brammer, C. Andrew, A. J. Calder, E. T. Bullmore, D. I. Perrett, D. Rowland, S. C. R. Williams, J. A. Gray, and A. S. David, "A specific neural substrate for perceiving facial expressions of disgust," *Nature*, vol. 389, no. 6650, pp. 495–498, Oct. 1997.

[43] W. D. S. Killgore and D. A. Yurgelun-Todd, "Activation of the amygdala and anterior cingulate during nonconscious processing of sad versus happy faces," *NeuroImage*, vol. 21, no. 4, pp. 1215–1223, Apr. 2004.

[44] H. S. Mayberg, M. Liotti, S. K. Brannan, S. Mcginnis, R. K. Mahurin, P. A. Jerabek, J. A. Silva, J. L. Tekell, C. C. Martin, J. L. Lancaster, and P. T. Fox, "Reciprocal limbic-cortical function and negative mood: Converging PET findings in depression and normal sadness," *Amer. J. Psychiatry*, vol. 156, no. 5, pp. 675–682, 1999.

[45] A. Ploghaus, I. Tracey, J. Gati, S. Clare, R. Menon, P. Matthews, and J. Rawlins, "Dissociating pain from its anticipation in the human brain," *Science*, vol. 284, no. 5422, pp. 1979–1981, Jun. 1999.

[46] A. Ploghaus, L. Becerra, C. Borras, and D. Borsook, "Neural circuitry underlying pain modulation: Expectation, hypnosis, placebo," *Trends Cognit. Sci.*, vol. 7, no. 5, pp. 197–200, May 2003.

[47] R. Sprengelmeyer, M. Rausch, U. T. Eysel, and H. Przuntek, "Neural structures associated with recognition of facial expressions of basic emotions," *Proc. Roy. Soc. London. Ser. B, Biol. Sci.*, vol. 265, no. 1409, pp. 1927–1931, Oct. 1998.

[48] Y. Sun, H. Ayaz, and A. N. Akansu, "Neural correlates of affective context in facial expression analysis: A simultaneous EEG-fNIRS study," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 820–824, doi: 10.1109/GlobalSIP.2015.7418311.

[49] A. Chakraborty, P. Bhowmik, S. Das, A. Halder, A. Konar, and A. K. Nagar, "Correlation between stimulated emotion extracted from EEG and its manifestation on facial expression," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2009, pp. 3132–3137, doi: 10.1109/ICSMC.2009.5346157.

[50] Z.-C. Ramon, B.-E.-M. Lucia, L.-H. Daniel, and R.-F.-J. Mario, "Creation of a facial expression corpus from EEG signals for learning centered emotions," in *Proc. IEEE 17th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2017, pp. 387–390.

[51] G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes," *Nature Rev. Neurosci.*, vol. 13, pp. 407–420, May 2012.

[52] V. Zotev, R. Phillips, K. D. Young, W. C. Drevets, and J. Bodurka, "Prefrontal control of the amygdala during real-time fMRI neurofeedback training of emotion regulation," *PLoS ONE*, vol. 8, no. 11, Nov. 2013, Art. no. e79184.

[53] A. Basu and A. Halder, "Facial expression and EEG signal based classification of emotion," in *Proc. Int. Conf. Electron., Commun. Instrum. (ICECI)*, kolkata, India, Jan. 2014, pp. 1–4.

[54] *The Anatomical Structure and Function of the Brain—Brodmann Partition*. Accessed: Jul. 20, 2020. [Online]. Available: https://wenku.baidu.com/view/b36b00616e175f0e7cd184254b35eefdc9d31507.html

[55] *Cingulate Cortex*. Accessed: Aug. 9, 2018. [Online]. Available: https://baike.baidu.com/item/%E6%89%A3%E5%B8%A6%E7%9A%AE%E5%B1%82/6027041?fr=aladdin

[56] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers," *Cognition Emotion*, vol. 24, no. 7, pp. 1153–1172, Nov. 2010.

[57] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Trans. Affect. Comput.*, early access, Mar. 17, 2020, doi: 10.1109/TAFFC.2020.2981446.

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

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

GUOSHENG YANG received the M.S. and Ph.D. degrees in control theory and control engineering from the Harbin Institute of Technology, China, and Beijing Institute of Technology, China, in 1988 and 2002, respectively. He is currently a Professor with the School of Information Engineering, Minzu University of China, Beijing, China, where he directs the Pattern Recognition and Intelligent Systems Research Laboratory. He was a Postdoctoral Fellow with the Institute of Automation, Chinese Academy of Sciences, from May 2002 to May 2004, engaged in the study of multi-sensor fusion and image processing. From September 2005 to May 2007, he was a Postdoctoral Fellow with the University of Saskachewan, Canada, engaged in the research on the image processing and emotion recognition. His research interests mainly include computer vision, brain inspired computing, multimodality information fusion, pattern recognition, and artificial intelligence. His researches have been funded by the National Natural Science Foundation of China (NSFC), National Science and Technology Support Program (China), and National High-tech R&D Program (China). He has published several books and more than one hundred technical papers in refereed journals and conference proceedings. In addition, he has been a reviewer for many professional journals and conferences.

RUI JIAO was born in Nanyang, Henan, China, in 1996. She received the bachelor's degree in Internet of Things engineering from Pingdingshan University, Henan, in 2018. She is currently pursuing the master's degree in computer science and technology with the Minzu University of China, Beijing. Her research interests include artificial intelligence, construction of brain networks, and application in emotion analysis.

HUIPING JIANG was born in Nanyang, Henan, China, in 1975. She received the Ph.D. degree in microelectronics and solid state electronics from the Harbin Institute of Technology (HIT), in 2005. She is currently an Associate Professor of computer sciences with the Information Engineering Department, Minzu University of China. Her current research interests include different aspects of affective computing, artificial intelligence, and intelligent learning.

TING ZHANG was born in Nanyang, Henan, China, in 1980. He received the bachelor's degree in computer science and technology from Capital Normal University, in 2008. He is currently a Lecturer with the Computer Department, Minzu University of China. His current research interests include different aspects of deep learning, image processing, and natural language processing.