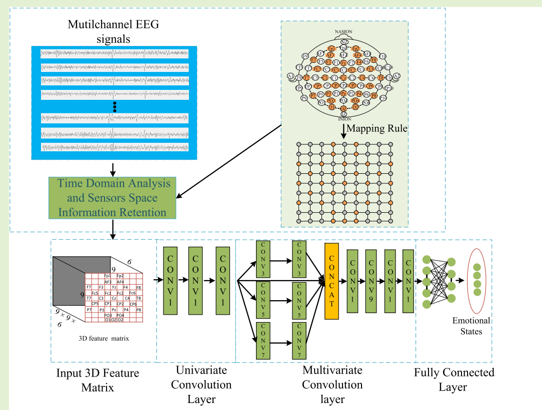
Emotion Recognition Using Three-Dimensional Feature and Convolutional Neural Network from Multichannel EEG Signals

Hao Chao, *Member, IEEE*, and Liang Dong

***Abstract*—Using electroencephalogram (EEG) signal to recognize emotional states has become a research hotspot of affective computing. Previous emotion recognition methods almost ignored the correlation and interaction among multichannel EEG signals, which may provide salient information related to emotional states. This article proposes a novel approach based on rearranged EEG features and deep learning algorithm. In particular, each channel EEG signal is first processed in time domain to get time-domain features. Then, features of all channels are treated as a three-dimensional (3D) feature matrix, according to positions of electrode sensors. This makes the features closer to the real response of the cerebral cortex. Subsequently, an advanced convolutional neural network (CNN) designed with univariate convolution layer and multivariate convolution layer is employed to deal with the 3D feature matrix for emotion recognition. A benchmark dataset for emotion analysis using physiological signal is employed to evaluate this method. The experimental results proved that the 3D feature matrix can effectively represent the emotion-related features in multichannel EEG signals and the proposed CNN can efficaciously mine the unique features of each channel and the correlation among channels for emotion recognition.**



Emotion Recognition Using Three-dimensional Feature and Convolutional Neural Network from Multichannel EEG Signals

***Index Terms*— Emotion recognition, multichannel EEG signal, three-dimensional feature, CNN, deep learning.**

I. Introduction

WITH the development of artificial intelligence, how to make computer own perception function, thinking function and behavior function similar to human has become a hot issue for researchers. Affective computing technology plays a key role when realizing human-computer interaction with intelligence and humanization. Most previous studies used facial [1] and speech [2] expressions to recognition emotions. However, these human signals are easy to be masked. Therefore, researchers try to use physiological signals, such as respiration, electrooculogram, galvanic skin response and electroencephalogram (EEG) [3]. Compared with other physiological signals, EEG can reflect dynamic changes of central neural system [4], and has been widely applied in various emotion recognition researches [5]–[7].

Extracting salient features related to emotional changes from EEG signals is the key to achieve satisfactory recognition performance. Various features extracted from time-domain, frequency-domain and time-frequency-domain have been used in many existing studies. The statistical features (mean, power, variance, etc.) of EEG are the most commonly used in previous researches [8], [9]. In addition, non-stationary index [10], Hjorth features (Activity, Mobility, and Complexity) [11], event-correlated potentials [12], and higher order crossing features [13] of EEG are also utilized for emotion recognition. Fast Fourier transform is the most frequently used analytical technique for EEG signals [14]. Power spectra density (PSD), power, and energy extracted from different EEG frequency bands are often employed in frequency-domain analysis [15]–[18]. Short-time Fourier transform is also used to acquire the time-varying characteristics reflected by EEG frequency data [19]. Koelstra *et al.* [20] presented a database named DEAP for the analysis of human emotional states,

Manuscript received July 28, 2020; accepted August 24, 2020. Date of publication September 1, 2020; date of current version December 16, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61502150, in part by the Fundamental Research Funds for the Universities of Henan Province under Grant NSFRF1616, in part by the Foundation for Scientific and Technological Project of Henan Province under Grant 172102210279, and in part by the Key Scientific Research Projects of Universities in Henan under Grant 19A520004. The associate editor coordinating the review of this article and approving it for publication was Prof. Kea-Tiong Tang. *(Corresponding author: Hao Chao.)*

The authors are with the Data Mining and Pattern Recognition Laboratory, School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454003, China (e-mail: chaohao1981@163.com; 211709020013@home.hpu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2020.3020828

and used the spectral power features of five frequency bands and Gaussian naive Bayes for emotion recognition. As EEG signal is time-varying, some researchers utilized discrete wavelet transforms [21] and Hilbert-Huang transform [22] methods to study EEG signals from both time and frequency domain. In addition, studies in neuroscience and neuropsychiatry have provided evidence for the connection between cognition and brain functional connectivity. Wu at al. [23] investigated the emotion-relevant brain functional connectivity patterns and evaluated the performance of the EEG connectivity feature for multimodal emotion recognition.

Many machine learning algorithms, such as k-nearest neighbor (KNN) [24], support vector machine (SVM) [25], and artificial neural network (ANN) [26] have been employed to identify emotional states. Transfer learning techniques have also been utilized to improve the performance of cross subject EEG classification [27]. Compared with traditional machine learning methods, deep learning technology has shown potential ability and achieved satisfactory performance [28]–[30]. Deep learning method can describe rich internal information because it can automatically obtain high-level features from data. Kwon *et al*. [31] extracted features from EEG signals and galvanic skin response (GRS) signals to construct fusion features, and then input into convolution neural network (CNN) for emotion recognition. Wang and Shang [32] introduced deep belief networks (DBNs) into emotion recognition. In our previous work, multiple domain EEG features were input into an improved deep belief network with glia chains to identify emotions [16].

Although emotion recognition researches based on EEG and deep learning technology have made some progress, there are still some challenges. First of all, most attentions were focused on how to obtain the frequency-domain, time-domain, and time-frequency domain features related to emotional state changes from each EEG channel. The interaction and correlation among multichannel EEG signals are ignored, which may also provide salient information. Furthermore, the traditional deep learning methods such as the DBN and the stacked autoencoder (SAE) cannot fully extract the correlation and interaction features among EEG channels in different brain regions.

For resolving these problems, this work introduces a special CNN model to process the 3D feature matrix for emotion recognition. Firstly, a three-dimensional (3D) feature matrix is constructed according to the arrangement of electrodes in the brain by using time-domain features. The 3D features matrix contains the unique features of each channel and the correlation features among channels, which is closer to the real brain response. Subsequently, a CNN is designed to extract high-level features related to emotions. The unique features of each single channel are fully extracted in the univariate convolution layer. In the multivariate convolution layer, the interaction and correlation among EEG signals are considered.

The paper is organized as follows. The emotion dataset, 3D feature matrix extraction method and the structure of proposed CNN are introduced in Section 2. The experimental results are described in Section 3. Section 4 provides a discussion. A brief conclusion of this work is presented in Section 5.

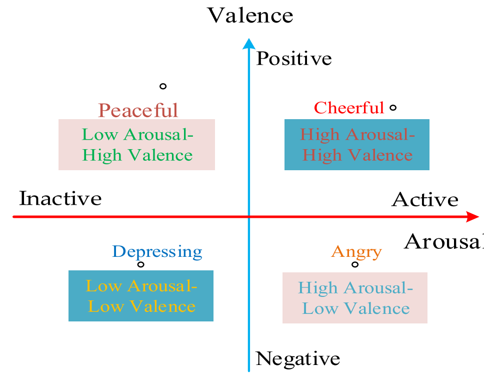


Fig. 1. Two-dimensional emotional space.

II. Materials and Methods

A. Dataset and Emotion Model

The most commonly used public dataset DEAP [20] is adopted to evaluate the proposed emotion recognition method. In this dataset, peripheral physiological and EEG signals of 32 subjects (including 16 males and 16 females, aged 19 to 37 years with an average of 26.9) were recorded while watching 40 clips of music videos as stimuli. Each video was selected to stimulate a related emotional state. For each subject, 40 trials were implemented, so there are 1,280 trials (40 trials $\times$ 32 subjects) in the dataset. The EEG signals were recorded in each trial included 63s (3s baseline signals and 60s stimulation signals) and 32 channels. After each trial, subject was asked to complete the self-assessment manikin (SAM) for valence, arousal, dominance, and liking. Each self-assessment dimension ranges from 1 to 9.

In this work, we used the pre-processed raw 32-channel signals and the corresponding two emotion self-assessment values (arousal and valence dimensions). The EEG signals (512 Hz) were downsampled to 128 Hz and processed to remove the electrooculogram artifacts. Then, the signals were filtered with cut-off frequencies of 4.0 and 45.0 by using band-pass filter.

The Russell's arousal-valence scale model is used for emotion analysis, due to its simplicity and extensive application. Fig. 1 shows the two-dimensional emotional space in which emotional states are described in a 2D space. Arousal varies from inactive (e.g., calm, uninterested) to active (e.g., alert, excited) and measures the activation of the sympathetic nervous system, represented by horizontal axis. Valence ranges from negative (e.g., stressed, sad) to positive (e.g., elated, happy) and measures subjective attitudes, represented by vertical axis.

Two types of emotion labeling schemes are used for arousal-valence scale. First, two emotion labels are defined in arousal and valence dimensions respectively. For each trial, if the associated self-assessment value of the valence is greater than five, then the trial is assigned to the high valence (HV) emotion class. Otherwise, the trial is assigned to the low valence (LV) emotion class. Similarly, there are low arousal (LA) and high arousal (HA) emotion classes in the arousal dimension. Secondly, four emotion labels are defined in the two-dimensional emotional space. If both the associated self-assessment values of the arousal and valence are greater than five, then the trial is assigned to the high arousal-high valence (HAHV) emotion class. Contrary, if both the associated self-assessment values of the arousal and valence are not greater than five, the trial is assigned to the low arousal-low valence (LALV) emotion class. Therefore, high arousal-low valence (HALV) and low arousal-high valence (LAHV) emotion classes can also be defined.

B. Multichannel EEG-Based 3D Feature Matrix

Due to the volume conductance effect, the reactions of two physically adjacent electrodes tend to be similar [33]. In the process of emotional state transition, the changes of EEG signals in different brain regions are not synchronous [34]. some researchers have found that retaining the position information of EEG electrodes when extracting features can provide gain information for emotion recognition [17]. Therefore, better recognition performance may be obtained if the above information can be preserved in the extracted emotion-related features. Considering the above reasons, a feature extraction method that can retain the above gain information is proposed.

The time-domain features are first extracted. Specifically, the time-domain features include mean μ_s, variance σ_s, standard deviation τ_s, the mean of absolute values of first difference δ_s, the mean of absolute values of second difference γ_s and the approximate entropy $ApEn$. Supposing that the EEG signal of each channel is represented by $s(t), t = 1, 2, 3, \ldots, T$, where T is the signal length. The calculation formulas of the first five features are expressed as:

$$\mu_s = \frac{1}{T}\sum_{t=1}^{T} s(t), \tag{1}$$

$$\sigma_s = \frac{1}{T}\sum\nolimits_{t=1}^{T}\left(s(t) - \mu_s\right)^2, \tag{2}$$

$$\tau_s = \sqrt{\frac{1}{T}\sum\nolimits_{t=1}^{T}\left(s(t) - \mu_s\right)^2}, \tag{3}$$

$$\delta_s = \frac{1}{T-1}\sum_{t=1}^{T-1} |s(t+1) - s(t)|, \tag{4}$$

$$\gamma_s = \frac{1}{T-2}\sum_{t=1}^{T-2} |s(t+2) - s(t)|. \tag{5}$$

Approximate entropy is often used to quantify the irregularity and complexity of time series. For the EEG signal $s(t)$, a threshold r for similarity comparison is defined, and then a measure m for length of the subsequence is determined. In this article, $m = 2$, $r = 0.2$. By reconstructing the original sequence, $(T - m + 1)$ subsequences $X(1), X(2), \ldots, (T - m + 1)$ are acquired. Each subsequence is represented by $X(i)$, where $X(i) = s(i), s(i+1), \ldots, s(i+m-1).d_m\,[X(i), X(j)]$ represents the distance between two arbitrary reconstruction vectors $X(i)$ and $X(j)$. After counting the number of vectors

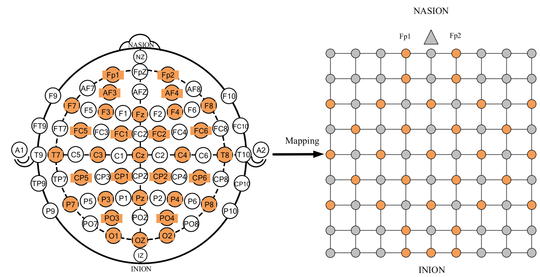


Fig. 2. International 10 / 20 system and 9 × 9 matrix.

that meet the following conditions, and the ratio to the total number of statistics is achieved, as in

$$C_i^m(r) = \frac{\mathrm{num}(d_m[X(i), X(j)] < r)}{T - m + 1}. \tag{6}$$

The average similarity rate $\Phi_m(r)$, when the number of subsequences is m, is defined as:

$$\Phi_m(r) = \frac{\sum_{i=1}^{T-m+1} \log(C_i^m(r))}{T - m + 1}. \tag{7}$$

According to above steps, the average similarity rate $\Phi_{m+1}(r)$ is calculated when the number of sequences is $m + 1$. The approximate entropy of EEG signal is expressed as:

$$ApEn = \Phi_m(r) - \Phi_{m+1}(r). \tag{8}$$

Six types of time-domain features are extracted from each EEG channel. The number of EEG channels used in this study is 32 (Fp1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, p2, Fz and Cz). Therefore, the number of features per sample is 192 (32 channels × 6 features).

Fig. 2 shows the international 10 / 20 system and its mapping matrix. The left part describes the distribution of electrodes in the cerebral cortex. Among them, those marked with orange are the electrodes used for EEG data in this article. The matrix of the mapping is described on the right. The size of the mapping matrix is 9 × 9, but only 32 channels are used. In order to keep the spatial information intact without affecting the features, 0 is used to complete the unused channels. The orange position in the matrix is filled with the feature value of the corresponding electrode. The position in the matrix, where the electrode is not used, is set to 0.

In order to represent the features reasonably, each same type of features of all channels are arranged according to the mapping rules in Fig. 2. For instance, for each type of features of all channels, a 9 × 9 × 1 feature matrix will be constructed. For all six types of features of each sample, a 3D feature matrix of size 9 × 9 × 6 can be obtained, as shown in Fig. 3. The characters in the red grid are the names of electrodes.

The 3D feature matrix not only contains unique features of each EEG channel, but also retains the interaction and correlation information among channels. In addition, the spatial information of brain and synchronous change between brain regions can also be well characterized [34]. This feature representation method can show the changes of EEG signals on the cerebral cortex more directly and accurately.

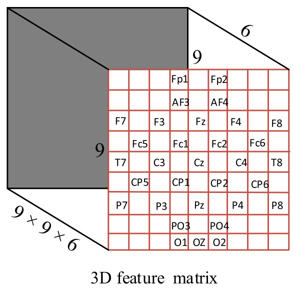


Fig. 3. 3D feature matrix.

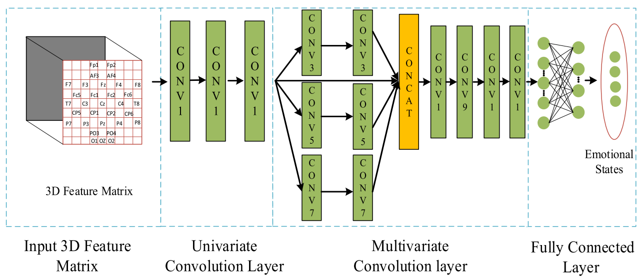


Fig. 4. Architecture of the proposed emotion recognition CNN model.

C. CNN With Specific Structures for Emotion Recognition

In the deep learning methods, CNN, which uses convolution layers for feature extraction has made a breakthrough in image-related tasks. The success of CNN attracted us to introduce it into EEG-based emotion recognition.

Lin *et al.* [35] introduced 1 × 1 convolution for the first time in their proposed network model. Szegedy *et al.* [36] used parallel multiscale convolution filters to obtain different information of the input image in their proposed Inception network model to obtain better image representation. Liu *et al.* [37] proposed a CNN model with 1 × 1 convolution and multiscale convolution structure for time series classification. Inspired by their researches, this work tried to use the above two structures in CNN to extract the unique emotion-related features of each EEG signal and the correlation among EEG signals in different brain regions from 3D feature matrix for emotion recognition.

Fig. 4 describes the structure of the proposed CNN model. The first part is feature input layer, and the input feature is a 3D feature matrix. Univariate convolution layer is the second part, which consists of 1 × 1 convolution layers. The third part is multivariate convolution layer, which mainly includes a four-way convolution. The last part is fully connected layer.

In the univariate convolution layer, univariate convolutions are applied to the 3D feature matrix. The purpose of the univariate convolution layer is to extract features from each EEG channel separately. To enhance the abstract ability of local model, filters with plane size of 1 × 1 are used in univariate convolution layer to scan the channels to increase nonlinearity. In addition, the relu activation function (ReLU) is used to get the nonlinear results after each 1 × 1 convolution.

Inspired by Iception network structure [36], a four-way convolution is used to find the optimal local brain construction. Filters of different sizes can extract the symmetric property of the large area and the particular characteristics of the small area in the brain. Three scales of filters are used in multivariate convolution layer, including 3 × 3, 5 × 5, and 7 × 7. This setting depends on the form of the 3D feature matrix. Each different convolutional layer consists of two layers. The first convolutional layer is considered to find the local correlation of EEG channels by grouping the local channels together. The useful information between groups can be taken by the second convolutional layer. The same padding is used on each convolutional layer. In addition, ReLU activation function is applied after each convolutional operation. Moreover, the output of the univariate convolutional layer is also directly used as another way of convolution.

After the above process is completed, all output from the previous sections is parallelly merged. This preserves the unique features from each group of convolutional layers. Next, a convolution operation is performed at the output of concatenation of the inception units, where the size of the filter is the same as that of the input data. As a result, the filter compresses each tensor into a vector from different ways.

Finally, the output of the multivariate convolution layer is fed to the fully connected layer. Besides, in order to prevent over fitting, two dropout layers are employed before the final full connected layer. Softmax function outputs the emotion recognition result. The loss function of this model is cross-entropy function, as in

$$H(p, q) = -\sum_{i} p(i) \log q(i), \tag{9}$$

where $q(i)$ is the estimated distribution and $p(i)$ is the true distribution. In addition, the adaptive moment estimation optimization with initial learning rate 0.0001 is used to minimize the loss function.

Different from traditional CNN, the proposed CNN model dedicates to deal with multichannel EEG signals, in which the network considers the correlation and interaction among channels. The structural characteristics of CNN used in this work enable it to make full use of the emotion-related features in the 3D feature matrix, so as to improve the emotion recognition performance. Full details about the network architecture used in this work can be found in Table I. The network parameters which achieve the relatively good result of our experiments are adopted.

III. RESULTS

DEAP data set is used to evaluate the emotion recognition method proposed in this work. Sufficient data are necessary for a deep learning model to obtain meaningful results. Therefore, a temporal segmentation method is used to augment the number of samples. First, a 3-second pretrial baseline in each EEG signal is removed. Then, the raw EEG signals of each trial are divided through sliding window technology. In order to ensure the validity of extracted features and the sufficient number of samples, the duration of the sliding window is set to 6s. The sliding windows are nonoverlapping to avoid the intra-trial redundant information. The segments are regarded as independent samples which are divided in the same period for all channels in one trial, and the new samples inherited the emotion labels of the original. The total numbers of samples used in this study is 12800, with different emotion labels, as shown in Table II.

TABLE I
FULL DETAILS OF THE PROPOSED EMOTION RECOGNITION CNN MODEL

Layer Type	Layer Parameters	Output Shape
Input	9 × 9 ×6	9 × 9 × 6
Univariate convolution layer		
1 × 1 convolution	32 1 × 1 filters	9 × 9 × 32
1 × 1 convolution	64 1 × 1 filters	9 × 9 × 64
1 × 1 convolution	128 1 × 1 filters	9 × 9 × 128
Multivariate convolution layer		
Multi-size filter layer 1	128 3 × 3 filters 128 5 × 5 filters 128 7 × 7 filters	9 × 9 × 128 9 × 9 × 128 9 × 9 × 128
Multi-size filter layer 2	128 3 × 3 filters 128 5 × 5 filters 128 7 × 7 filters	9 × 9 × 128 9 × 9 × 128 9 × 9 × 128
Concatenation	x-axis	36 × 9 × 128
1 × 1 convolution	128 1 × 1 filters	36 × 9 × 128
9 × 9 convolution	176 9 × 9 filters	4 × 1 × 176
1 × 1 convolution	194 1 × 1 filters	4 × 1 × 194
1 × 1 convolution	256 1 × 1 filters	4 × 1 × 256
Fully connected layer		
Dense layer 1	1024 units	1024
Dropout layer 1	Rate: 0.6, training: False	
Dense layer 2	512 units	512
Dropout layer 2	Rate: 0.3, training: False	
SoftMax		4 or 2

TABLE II
THE NUMBER OF SAMPLES PER EMOTION CLASS FOR EACH LABELING SCHEME

Emotion recognition task	Emotion Description Scale	Emotion Class	Data quantity
Binary classification	Arousal	HA	7370
		LA	5430
	Valence	HV	7080
		LV	5720
Four-class classification	2D arousal-valence plane	HAHV	4390
		LAHV	2690
		LALV	2740
		HALV	2980

A. Performance Analysis

Samples of all subjects are used to evaluate the model, which can be used to illustrate the generalization ability of the proposed method. The models are implemented by using TensorFlow and are deployed on a GeForce GTX 1060 GPU. The 10-fold cross validation technique is used to analyze the emotion recognition performance under two types of emotion labeling schemes. Specifically, all samples are randomly divided into ten subsets. One subset is regarded as a test set, and another subset as a validation set. The remaining subsets are regarded as a training set. The above process is repeated ten times until all subsets are tested. Batch size is set to 40. In order to find the best training point, the results of one experiment in four-class classification task is exploited, as shown in Fig. 5.

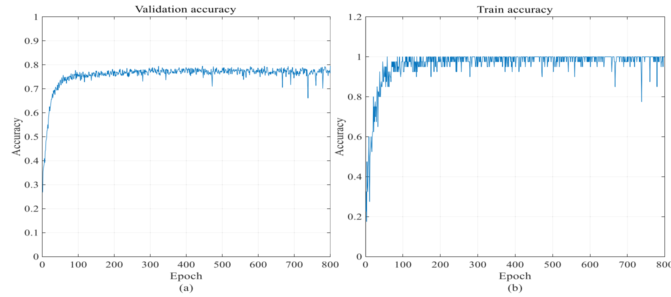


Fig. 5. A four-class classification result to find the optimal training point. (a) Validation accuracy of the proposed CNN model. (b) Training accuracy of the proposed CNN model.

TABLE III
EMOTION RECOGNITION RESULTS OF BINARY CLASSIFICATION USING THE PROPOSED CNN AND 3D FEATURE MATRIX

Predict	Label: Arousal High	Arousal Low	Valence High	Valence Low
High	6505	943	6215	987
Low	865	4487	865	4733
F1 score	0.8780		0.8703	
Precision	0.8734		0.8630	
Recall	0.8826		0.8778	
Kappa	0.7103		0.7067	
Accuracy	0.8588 ± 0.0162		0.8553 ± 0.0206	

As shown in Fig. 5(a), when the epoch is less than 400, the validation accuracy increases with the increase of epoch, but when it is greater than 400, the validation accuracy does not increase, and has a downward trend. As shown in Fig. 5(b), when the epoch is greater than 200, the training accuracy is close to 1. And when the epoch is over 400, its fluctuation is larger with the increase of epoch. Therefore, in order to obtain convincing emotion recognition results, the epoch of CNN used in this work is set to 400.

Classification accuracy (Acc), F1 score (F1), Precision (Pre) Recall and Kappa are used to evaluate the performance of the CNN model. The parameters of F1 score, Precision and Recall are set to macro-average for multiclass task. The results of emotion recognition using two types of emotion labeling schemes are shown in Table III and Table IV, respectively.

In the binary classification task, the proposed method achieves accuracies of 0.8588 in the arousal dimension and 0.8553 in the valence dimension. Meanwhile, it achieves F1 scores of 0.8780 in the arousal dimension and 0.8703 in the valence dimension. In the four-class classification task, the proposed method achieves accuracy of 0.7677 and F1 score of 0.7631. Satisfactory results have also been obtained on other evaluation methods. The recognition results based on two types of emotion labeling schemes show the effectiveness of the proposed method.

TABLE IV
EMOTION RECOGNITION RESULTS OF FOUR-CLASS CLASSIFICATION USING THE PROPOSED CNN AND 3D FEATURE MATRIX

Predict	Label			
	HAHV	LAHV	LALV	HALV
HAHV	3529	333	309	335
LAHV	267	1968	219	150
LALV	257	219	2004	172
HALV	337	150	172	2323
F1 score	0.7631			
Precision	0.7649			
Recall	0.7616			
Kappa	0.6845			
Accuracy	0.7677 ± 0.0226			

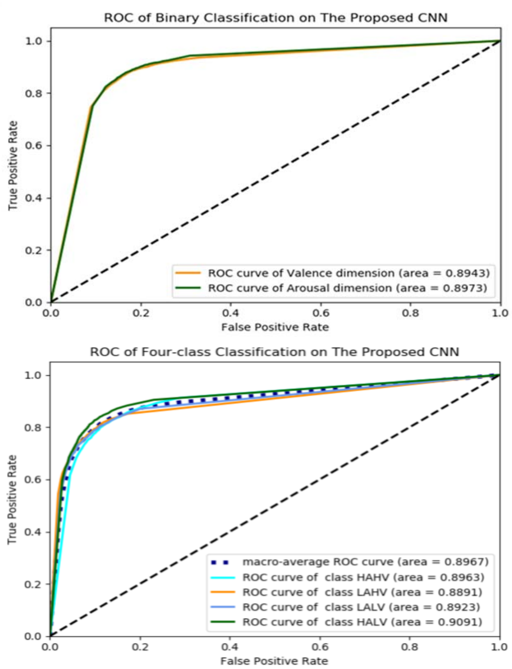


Fig. 6. ROC Curve on the proposed CNN. Top: ROC Curves of arousal and valence dimensions. Bottom: ROC Curves of four emotion classes and macro-average ROC curve.

The Receiver Operating Characteristic (ROC) curve is also used to evaluate the performance of the proposed CNN. The area under the ROC curve measures the performance of supervised classification rules. A satisfactory classification rule is reflected by a ROC curve which lies in the upper left triangle of the square. And the higher the area value, the better the ranking performance. Fig. 6 shows the ROC curves of the proposed CNN model on two tasks. The convex ROC curves and the high area value (Four-class classification: 0.8967, Arousal: 0.8973, Valence: 0.8943) exhibit the excellent classification performance of the proposed CNN.

B. Comparison Between 3D Feature Matrix and 2D Feature Matrix

To prove the advantages of 3D feature matrix, the proposed CNN model is also used to deal with the same time domain features without permutation. Six types of features of 32 channels can construct a two-dimensional (2D) feature matrix of size 32×6 (32 channels $\times$ 6 features). According to the size of 2D feature matrix, the filter sizes of the multivariable convolution layer are set to 2×2, 3×3 and 5×5 respectively. In the second convolution layer after concatenation, the filter size is set to 32×6. Samples of all subjects and the 10-fold cross validation technique are also used here. When epoch is more than 300, the CNN model is stable. Therefore, batch size and epoch are also set to 40 and 400 respectively.

The recognition results under the two types of emotion labeling schemes are shown in Table V. Firstly, the comparison is carried out on the emotion recognition accuracy. Compared with 2D feature matrix, 3D feature matrix improves the emotion recognition accuracies by 2.54% in the arousal dimension and 2.89% in the valence dimension. In the four-class classification task, 3D feature matrix improves the emotion recognition accuracy by 4.21%. Secondly, the performance of 3D feature matrix and 2D feature matrix is analyzed by using a Wilcoxon signed-rank test ($\alpha < 0.05$). Here, the recognition performance is similar as a null hypothesis. The null hypothesis is accepted if α is less than p-value. The p-values of arousal, valence and four-class classification are 0.023, 0.002 and 0.002, respectively. The performance of 3D feature matrix on the other four evaluation methods is also better than that of 2D feature matrix. The results show that the performance of 3D feature matrix is superior to 2D feature matrix. The reason may be that the 3D feature matrix, which is closer to the real cerebral cortex response, contains the spatial domain features and correlation information among channels, which can improve emotion recognition performance. Satisfactory results are also obtained by using 2D feature matrix, which proves the validity of the proposed CNN model. Moreover, according to the differences of recognition accuracies between 2D and 3D feature matrix under two types of emotion labeling schemes, it can be seen that the gain information in 3D feature matrix improves the performance of multiclass emotion recognition task more obviously.

C. Advantages of Time Domain Features

In order to illustrate the advantages of using time-domain features in this work, frequency-domain features, fusion features, and raw EEG data are all used for comparison. Specifically, the frequency-domain features include the average PSD values (4 power $\times$ 32 EEG channels) of all EEG channels in theta (4–8 Hz), alpha (8–12Hz), beta (12–30 Hz), and gamma (30–45 Hz) bands, extracted from each sample in Table II. The fusion features (10 features $\times$ 32 EEG channels) combine frequency-domain features and time-domain features. To reduce computational complexity, the first 20 seconds signal of each sample in the DEAP dataset is used. Then, the same sample augment method as in Table II is used to increase the number of samples to 12800, except that the duration of the sliding window is 2s. In this way, samples using the raw EEG data (256 points $\times$ 32 EEG channels) for emotion recognition are acquired. These three types of features

TABLE V
THE RESULTS OF 3D FEATURE MATRIX AND 2D FEATURE MATRIX USING THE PROPOSED CNN

Emotion dimension	Recognition results										Wilcoxon signed-rank test
	2D feature matrix					3D feature matrix					
	Acc	F1	Pre	Recall	Kappa	Acc	F1	Pre	Recall	Kappa	
Arousal	0.8375	0.8594	0.8559	0.8629	0.6668	0.8588	0.8780	0.8734	0.8826	0.7103	$P = 0.0230$
Valence	0.8313	0.8493	0.8388	0.8601	0.6576	0.8553	0.8703	0.8630	0.8778	0.7067	$P = 0.0020$
Four-class Classification	0.7367	0.7332	0.7321	0.7357	0.6443	0.7677	0.7631	0.7649	0.7616	0.6845	$P = 0.0020$

TABLE VI
THE RESULTS OF DIFFERENT TYPES OF FEATURES USING THE PROPOSED CNN

Feature Type	Recognition results														
	Arousal					Valence					Four-class Classification				
	Acc	F1	Pre	Recall	Kappa	Acc	F1	Pre	Recall	Kappa	Acc	F1	Pre	Recall	Kappa
Time-domain features	0.8588	0.8780	0.8734	0.8826	0.7103	0.8553	0.8703	0.8630	0.8778	0.7067	0.7677	0.7631	0.7649	0.7616	0.6845
Frequency-domain features	0.6789	0.7276	0.7111	0.7449	0.3371	0.6632	0.6883	0.7051	0.6723	0.3226	0.4609	0.4499	0.4523	0.4518	0.2729
Fusion features	0.8335	0.8553	0.8559	0.8548	0.6594	0.8328	0.8492	0.8469	0.8516	0.6615	0.7523	0.7474	0.7520	0.7440	0.6631
Raw EEG data	0.5921	0.6482	0.6439	0.6526	0.1631	0.5429	0.5945	0.5836	0.6059	0.0712	0.3296	0.3179	0.3214	0.3179	0.0883

also construct the 3D feature matrix according to the method proposed in this work. The experimental settings and network parameters remain unchanged.

The performance of CNN using time-domain features, frequency-domain features, fusion features and raw EEG data are shown in Table VI for both binary classification task and four-class classification task. In the arousal dimension, compared with frequency-domain features, fusion features and raw EEG data, time-domain features improve the emotion recognition accuracies by 26.50%, 3.04% and 45.04% respectively. In the valence dimension, time-domain features improve the emotion recognition accuracies by 28.97%, 2.70% and 57.54% respectively. In the four-class classification task, time-domain features improve the emotion recognition accuracy by 66.57%, 2.05% and 132.91% respectively. In other evaluation methods, time-domain features have also obtained the best performance, which proves its advantages in characterizing emotion-related features in EEG signals. Moreover, it is observed that the performance of emotion recognition using raw EEG data is the worst. The reason may be that, the raw EEG data contains too much information and the network cannot learn effective emotional discrimination features from it. Moreover, using raw EEG data for emotion recognition requires huge memory and computing resources.

D. Advantages of The Proposed CNN's Structures on Emotion Recognition

To illustrate the advantages of the proposed model structures, three comparison CNNs (CNN-1, CNN-2 and CNN-3) are used. CNN-1 is used to show the advantages of the proposed CNN's 1×1 convolution layer. Instead of using 1×1 convolution filters, CNN-1 uses 3×3 convolution filters in univariate convolution layer. CNN-2 is used to show the advantages of the proposed CNN's multivariable convolution layer which only uses one-way convolution with 5×5 filters in the multivariable convolution layer. In addition, in order to demonstrate the advantages of the two structures mentioned above, CNN-3 is used. CNN-3 has 3×3 convolution filters in univariate convolution layer and one-way convolution with 5×5 filters in the multivariable convolution layer. Other parameters in three comparison CNNs remain the same. Samples of all subjects and the 10-fold cross validation technique are also used here. For a fair comparison, epoch and batch size are also set to 400 and 40, respectively. The recognition results under the two types of emotion labeling schemes using CNNs with different structure and 3D feature matrix are shown in Table VII.

In the binary classification task, compared with CNN-1, CNN-2 and CNN-3, the proposed CNN improves the recognition accuracies of 20.26%, 4.795% and 22.97% respectively in the arousal dimension. In the valence dimension, the proposed CNN improves the recognition accuracies of 20.84%, 5.475% and 24.99% respectively. In the four-class classification task, the proposed CNN improves the recognition accuracies of 35.73%, 6.462% and 55.47% respectively. Wilcoxon signed-rank test is also used here to compare the performance between the proposed CNN and comparison CNNs. The p-values of CNN-1, CNN- 2 and CNN-3 in two

TABLE VII
THE RESULTS OF THE PROPOSED CNN, CNN-1, CNN-2 AND CNN-3 USING 3D FEATURE MATRIX

Model	Recognition results														
	Arousal					Valence					Four-class Classification				
	Acc	F1	Pre	Recall	Kappa	Acc	F1	Pre	Recall	Kappa	Acc	F1	Pre	Recall	Kappa
The proposed CNN	0.8588	0.8780	0.8734	0.8826	0.7103	0.8553	0.8703	0.8630	0.8778	0.7067	0.7677	0.7631	0.7649	0.7616	0.6845
CNN-1	0.7141	0.7559	0.7454	0.7666	0.4133	0.7078	0.7278	0.7508	0.7062	0.4131	0.5656	0.5515	0.5559	0.5491	0.4066
CNN-2	0.8195	0.8425	0.8465	0.8385	0.6312	0.8109	0.8234	0.8554	0.7937	0.6223	0.7211	0.7165	0.7165	0.7174	0.6226
CNN-3	0.6984	0.7520	0.7156	0.7924	0.3722	0.6843	0.7053	0.7253	0.6864	0.3624	0.4938	0.4813	0.4864	0.4824	0.3133

TABLE VIII
RECOGNITION RESULTS OF DIFFERENT CLASSIFIERS

Classifier	Recognition results														
	Arousal					Valence					Four-class Classification				
	Acc	F1	Pre	Recall	Kappa	Acc	F1	Pre	Recall	Kappa	Acc	F1	Pre	Recall	Kappa
The proposed CNN	0.8588	0.8780	0.8734	0.8826	0.7103	0.8553	0.8703	0.8630	0.8778	0.7067	0.7677	0.7631	0.7649	0.7616	0.6845
RDF	0.7743	0.7976	0.8169	0.7791	0.5376	0.7684	0.7843	0.8077	0.7623	0.5341	0.6449	0.6348	0.6506	0.6269	0.5131
DT	0.7278	0.7647	0.7657	0.7636	0.4463	0.7212	0.7477	0.7531	0.7423	0.4404	0.5854	0.5803	0.5800	0.5806	0.4409
KNN	0.6416	0.6998	0.6759	0.7254	0.2567	0.6055	0.6478	0.6398	0.6561	0.1995	0.4209	0.3977	0.4106	0.3959	0.2008
SVM-linear	0.5922	0.7274	0.5913	0.6449	0.0653	0.5650	0.7082	0.5630	0.5542	0.0408	0.3595	0.1951	0.4028	0.2755	0.0396
SVM-rbf	0.5787	0.7294	0.5787	0.5058	0.0134	0.5565	0.7090	0.5564	0.4770	0.0143	0.3427	0.1305	0.2544	0.2503	0.0003
ANN	0.6497	0.7151	0.6497	0.7636	0.2651	0.6090	0.6738	0.6255	0.7302	0.1930	0.4168	0.3856	0.4093	0.3851	0.1874

tasks are all 0.002. The performance of the proposed CNN on the other four evaluation methods is also better than that of the comparison CNNs. The comparison results show that the performance of the proposed CNN is much better than the comparison CNNs. This proves the superiority of the proposed CNN model structure.

From the results of CNN-1 and CNN-2, it can be seen that the influence of univariate convolution layer on recognition results is more obvious than that of multivariable convolution layer. The reason is that the time domain features of each channel provide the main information related to emotion state. It also shows that the correlation among channels and spatial information of brain can bring benefits to emotion recognition. Among all the comparison results, the performance of the four-class classification task changed more obviously using difference CNN models, which indicated that it would be more beneficial for multiclass emotion recognition task to make full use of each channel features and correlation information among EEG channels.

E. Comparison Between The Proposed CNN and Some Common Classifiers

In this section, some machine learning methods using the same time-domain features for emotion recognition analysis are also constructed. Machine learning methods including SVM with linear kernel (SVM-linear), SVM with rbf kernel (SVM-rbf), Decision Tree (DT), Random Decision Forest (RDF), ANN (activation = relu, batch_size = auto, max_iter = 200, learning_rate_init = 0.001, alpha = 1e-05, beta_1 = 0.9, beta_2 = 0.999) and KNN (neighbors = 5). Specially, the ANN has two 100 nodes hidden layers. All parameters not mentioned in each comparison classifier use default values. And all comparison classifiers are implemented by the Scikit-learn toolkit [38]. All experiments are conducted in two types of emotion labeling schemes by using samples of all subjects, and a 10-fold cross-validation technique is also adopted here. In order to obtain the best emotion recognition results of different classifiers, the features are normalized into 0 to 1 before they are input to the SVM-linear, SVM-rbf and ANN classifiers, and are not normalized for other classifiers.

The recognition results of classifiers using the same features are shown in Table VIII. In the binary classification task, the proposed CNN achieves the highest average recognition accuracies in two dimensions. In the arousal dimension, compared with the SVM-rbf with the worst performance and RDF with the best performance of the comparison classifiers, the proposed CNN improves the recognition accuracies by 48.4% and 10.91%. In the valence dimension, compared with RDF, which performs best in the comparison classifiers, the proposed CNN improves the recognition accuracy by 11.31%. Compared with the worst performing SVM-rbf in the comparison classifiers, the proposed CNN improves the recognition accuracy by 53.69%. The proposed CNN also achieves the highest recognition accuracy in the task of four-class classification. Compared with RDF, DT, KNN, SVM-linear, SVM-rbf and ANN, the proposed

TABLE IX
RECOGNITION RESULTS OF DIFFERENT CLASSIFIERS USING HIGH-LEVEL FEATURES

Method	Recognition results														
	Arousal					Valence					Four-class Classification				
	Acc	F1	Pre	Recall	Kappa	Acc	F1	Pre	Recall	Kappa	Acc	F1	Pre	Recall	Kappa
CNN+RDF	0.9582	0.9637	0.9662	0.9612	0.9147	0.9510	0.9559	0.9519	0.9598	0.9008	0.9398	0.9389	0.9404	0.9374	0.9184
CNN+DT	0.9542	0.9602	0.9615	0.9589	0.9063	0.9436	0.9491	0.9481	0.9501	0.8860	0.9099	0.9086	0.9085	0.9086	0.8779
CNN+KNN	0.9708	0.9746	0.9761	0.9731	0.9403	0.9615	0.9653	0.9627	0.9679	0.9221	0.9417	0.9409	0.9426	0.9394	0.9210
CNN +SVM-linear	0.9734	0.9743	0.9733	0.9754	0.9395	0.9645	0.9680	0.9645	0.9716	0.9281	0.9430	0.9423	0.9435	0.9411	0.9227
CNN +SVM-rbf	0.9552	0.9605	0.9748	0.9466	0.9088	0.9505	0.9560	0.9407	0.9717	0.8995	0.8914	0.8924	0.9028	0.8848	0.8518
CNN+ANN	0.9725	0.9761	0.9767	0.9755	0.9438	0.9633	0.9672	0.9566	0.9781	0.9256	0.9413	0.9403	0.9412	0.9394	0.9203

CNN improves the recognition accuracy by 19.04%, 31.14%, 82.39%, 113.54%, 124.01% and 84.19%, respectively. The performance of the proposed CNN on the other four evaluation methods is also better than that of the comparison machine learning methods. Apparently, the CNN proposed in this work can achieve much better performance in two emotion classification tasks. In addition, compared with other classifiers, the proposed CNN has more significant performance in the task of four-class classification than in binary classification task. This shows that this method has more advantages in the relatively difficult multiclass emotion classification tasks.

F. Superior High-Level Features Extraction Performance of The Proposed CNN

The CNN designed in this work extracts the unique features of each EEG channel and the correlation features among channels from the 3D feature matrix. In order to demonstrate the effectiveness of this method, the high-level features extracted by CNN are input into some machine learning algorithms. The 3D feature matrix of each sample is input into the trained CNN model in Section III. A. Then, the outputs of 512 neurons in the last layer of the fully connected layer are used as the high-level features. The category and parameters of classic machine learning algorithms used in this section are the same as those in TABLE VIII. Samples of all subjects and the 10-fold cross validation technique are also used here.

The performance of classic machine learning algorithms using high-level features is shown in Table IX. In the binary classification task, by using SVM-linear to classify high-level features, the highest average recognition accuracies of 0.9734 in the arousal dimension and 0.9645 in the valence dimension are achieved. SVM-linear also achieves the highest average recognition accuracy of 0.9430 in the task of four-class classification. Satisfactory performance has also been obtained in other evaluation methods. Compared with using softmax for classification, using CNN for high-level features extraction and classic machine learning algorithms for classification can significantly improve emotion recognition performance. In addition, the recognition performance of classic machine learning algorithms using high-level features is also significantly improved compared to using time-domain features. The experimental results prove that making full use of the unique features of each EEG channel and the correlation among channels can improve the performance of emotion recognition, and the CNN can effectively extract these salient features.

IV. DISCUSSIONS

A. Advantages of The Proposed Method

First of all, although 3D feature matrix needs feature construction process, this work shows that the 3D feature matrix can improve the representation ability of multichannel EEG emotion-related features. It demonstrates that reasonable feature extraction and representation methods are very important to improve the performance of emotion recognition.

Secondly, 3D feature matrix is very similar to multichannel image. In view of this, we used CNN with special structures to process the 3D feature matrix. In the univariate convolution layer, 1×1 convolution is used to extract the unique features of each EEG channel and deepen the network. Multivariate convolution layer uses parallel filters of different sizes to extract the regional features of different areas of the cerebral cortex. The unique combination method of filters also preserves the regional features of different areas. Compared with the classical machine learning methods, CNN has higher computational complexity, but it can extract effective features from data. According to the EEG data form, constructing the deep network with corresponding structure is the research focus of improving the performance of emotion recognition.

B. Comparison With Existing Methods

The recognition performance of our approach under two types of emotion labeling schemes is compared with some existing researches. For all reported researches, the dataset (DEAP) and the schemes of labeling are the same. Table X details the features and classifiers used in the comparison researches. Recognition accuracy is adopted for comparison because it is most commonly used.

As shown in Table X, the performance of our approach is better than the comparison methods, no matter whether

TABLE X
DETAILS FROM VARIOUS PREVIOUS RESEARCHES. H-ATT-BGRU, HIERARCHICAL BIDIRECTIONAL GATED RECURRENT UNIT NETWORK; LSTM, LONG SHORT-TERM MEMORY NETWORKS

Emotion recognition task	Method	Features and Classifier	Accuracy	
			Arousal	Valence
Binary classification	Koelstra *et al.* [20]	Power spectral features, Gaussian naïve Bayes classifier	0.6200	0.5760
	Xu *et al.* [39]	Narrow-band spectral features, DBN	0.6988	0.6688
	Chen *et al.* [40]	Raw EEG features, H-ATT-BGRU	0.6650	0.6790
	Kwon *et al.* [31]	Wavelet transformed spectrogram of EEG, Short time zero Crossing rate of GRS, CNN	0.7656	0.8046
	Xing *et al.* [41]	PSD, SAE+LSTM	0.8110	0.7438
	Our approach	Statistical features, CNN	**0.8588**	**0.8553**
		Statistical features, CNN + SVM-linear	**0.9734**	**0.9645**
Four-class Classification	Zubair and Yoon [42]	Statistical and wavelet-based features, SVM	0.4970	
	Mei *et al.* [43]	Connection matrix of the brain structure, CNN	0.7310	
	Kwon *et al.* [31]	Wavelet transformed spectrogram of EEG, Short time zero Crossing rate of GRS, CNN	0.7343	
	Our approach	Statistical features, CNN	**0.7677**	
		Statistical features, CNN + SVM-linear	**0.9430**	

it is a binary classification task or a four-class classification task. Previous studies have suggested a variety of emotion related features. However, the reasonable representation of features was often ignored. Deep network CNN, DBN, SAE, LSTM and H-ATT-BGRU have been verified on DEAP dataset by comparison researches in Table X. However, the proposed CNN gets the best performance. So, constructing a suitable network structure according to the data form is the key to obtaining distinguishing features and satisfactory emotion recognition performance. The comparison results show that our approach is excellent in multichannel EEG emotion recognition.

C. Potential in Multichannel EEG Applications

The complex patterns in EEG data can be effectively represented by using the feature representation method similar to the arrangement of signal sensors. And the psychological and physiological information represented by the patterns, such as emotion and disease, can be further analyzed. This method also shows CNN's potential in other fields, not just image processing. The overall work enables an effective method for the emotion classification of EEG and holds great potential for other classification tasks related to EEG, such as epileptic dementia and Alzheimer's disease detection. Moreover, the proposed method also provides a new way for multisensory signals recognition and prediction task.

V. CONCLUSION

A novel multichannel EEG-based emotion recognition method based on 3D feature representation and CNN is proposed in this work. A deep CNN, which contains univariate convolution layer and multivariable convolution layer, is used to extract the unique information of each EEG channel and the correlation information among channels of the brain from the 3D feature matrix for emotion recognition. The results show that 3D feature matrix can effectively represent emotion-related features of multichannel EEG signals, and the proposed CNN can make full use of these features for emotion recognition. Moreover, the emotion recognition performance of the proposed method is compared with some existing methods and shows advantages, which proves the feasibility and effectiveness of the proposed emotion recognition method.

REFERENCES

[1] Y. Huang, J. Yang, P. Liao, and J. Pan, "Fusion of facial expressions and EEG for multimodal emotion recognition," *Comput. Intell. Neurosci.*, vol. 2017, no. 2, pp. 1–8, 2017.

[2] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognit.*, vol. 44, no. 3, pp. 572–587, Mar. 2011.

[3] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," *Comput. Methods Programs Biomed.*, vol. 140, pp. 93–110, Mar. 2017.

[4] W. Heller, "Neuropsychological mechanisms of individual differences in emotion, personality, and arousal," *Neuropsychology*, vol. 7, no. 4, pp. 476–489, Oct. 1993.

[5] 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.

[6] S. Raghu and N. Sriraam, "Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms," *Expert Syst. Appl.*, vol. 113, pp. 18–32, Dec. 2018.

[7] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, no. 4, pp. 94–106, Apr. 2014.

[8] T. Y. Chai *et al.*, "Classification of human emotions from EEG signals using statistical features and neural network," *Int. J. Eng. Sci.*, vol. 1, pp. 1–6, Jan. 2009.

[9] Y. Liu and O. Sourina, "Real-time fractal-based valence level recognition from EEG," in *Transactions on Computational Science XVIII*. Berlin, Germany: Springer, 2013, pp. 101–120.

[10] E. Kroupi, A. Yazdani, and T. Ebrahimi, "EEG correlates of different emotional states elicited during watching music videos," in *Proc. 4th Int. Conf. Affect. Comput. Intell. Interact.*, Oct. 2011, pp. 457–466.

[11] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, Sep. 1970.

[12] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis, "Toward emotion aware computing: An integrated approach using multichannel neurophysiological recordings and affective visual stimuli," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 589–597, May 2010.

[13] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, Mar. 2010.

[14] Z. Yin, Y. Wang, L. Liu, W. Zhang, and J. Zhang, "Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination," *Frontiers Neurorobotics*, vol. 11, no. 19, Apr. 2017.

[15] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.

[16] H. Chao, H. Zhi, L. Dong, and Y. Liu, "Recognition of emotions using multichannel EEG data and DBN-GC-Based ensemble deep learning framework," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–11, Dec. 2018.

[17] H. Chao, L. Dong, Y. L. Liu, and B. Y. Lu, "Emotion recognition from multiband EEG signals using CapsNet," *Sensors*, vol. 19, no. 9, pp. 1–16, 2019.

[18] D. Nath, M. Singh, D. Sethia, D. Kalra, and S. Indu, "A comparative study of subject-dependent and subject-independent strategies for EEG-based emotion recognition using LSTM network," in *Proc. the 4th Int. Conf. Compute Data Anal.*, Mar. 2020, pp. 142–147.

[19] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, "Real-time movie-induced discrete emotion recognition from EEG signals," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 550–562, Oct. 2018.

[20] S. Koelstra *et al.*, "DEAP: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[21] R. E. J. Yohanes, W. Ser, and G.-B. Huang, "Discrete wavelet transform coefficients for emotion recognition from EEG signals," in *Proc. 34th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 2251–2254.

[22] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "Toward an EEG-based recognition of music liking using time-frequency analysis," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3498–3510, Dec. 2012.

[23] X. Wu, W.-L. Zheng, and B.-L. Lu, "Investigating EEG-based functional connectivity patterns for multimodal emotion recognition," 2020, *arXiv:2004.01973*. [Online]. Available: http://arxiv.org/abs/2004.01973

[24] M. Li, H. Xu, X. Liu, and S. Lu, "Emotion recognition from multichannel EEG signals using K-nearest neighbor classification," *Technol. Health Care*, vol. 26, no. 4, pp. 509–519, Jul. 2018.

[25] I. Wichakam and P. Vateekul, "An evaluation of feature extraction in EEG-based emotion prediction with support vector machines," in *Proc. 11th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, May 2014, pp. 106–110.

[26] A. T. Sohaib, S. Qureshi, J. Hagelbäck, O. Hilborn, and P. Jerica, "Evaluating classifiers for emotion recognition using EEG," in *Proc. Inter Conf Augment. Cognit.*, 2013, pp. 492–501.

[27] T. Duan, M. Chauhan, M. Abuzar Shaikh, and S. Srihari, "Ultra efficient transfer learning with meta update for cross subject EEG classification," 2020, *arXiv:2003.06113*. [Online]. Available: http://arxiv.org/abs/2003.06113

[28] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *Sci. World J.*, vol. 2014, Sep. 2014, Art. no. 627892.

[29] W. Liu, W. L. Zheng, and B. L. Lu, "Emotion recognition using multimodal deep learning," in *Proc. 23rd Int. Conf. Neural Inform. Process.*, 2016, pp. 521–529.

[30] B. Yang, X. Han, and J. Tang, "Three class emotions recognition based on deep learning using staked autoencoder," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017, pp. 1–5.

[31] Y.-H. Kwon, S.-B. Shin, and S.-D. Kim, "Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system," *Sensors*, vol. 18, no. 5, p. 1383, Apr. 2018.

[32] D. Wang and Y. Zhang, "Modeling physiological data with deep belief networks," *Int. J. Inf. Edu. Technol.*, vol. 3, no. 5, pp. 505–511, 2013.

[33] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Hum. Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, Nov. 2007.

[34] T. Iwaki and M. Noshiro, "EEG activity over frontal regions during positive and negative emotional experience," in *Proc. ICME Int. Conf. Complex Med. Eng. (CME)*, Jul. 2012, pp. 418–422.

[35] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*. [Online]. Available: http://arxiv.org/abs/1312.4400

[36] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[37] C.-L. Liu, W.-H. Hsaio, and Y.-C. Tu, "Time series classification with multivariate convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4788–4797, Jun. 2019.

[38] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12 no. 10, pp. 2825–2830, Jan. 2012.

[39] H. Xu and K. N. Plataniotis, "EEG-based affect states classification using deep belief networks," in *Proc. Digit. Media Ind. Acad. Forum (DMIAF)*, Jul. 2016, pp. 148–153.

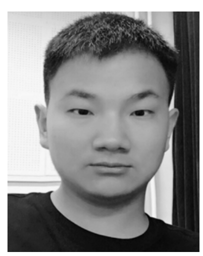
[40] J. X. Chen, D. M. Jiang, and Y. N. Zhang, "A hierarchical bidirectional GRU model with attention for EEG-based emotion classification," *IEEE Access*, vol. 7, pp. 118530–118540, 2019.

[41] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "SAE+LSTM: A new framework for emotion recognition from multi-channel EEG," *Frontiers Neurorobotics*, vol. 13, p. 37, Jun. 2019.

[42] M. Zubair and C. Yoon, "EEG based classification of human emotions using discrete wavelet transform," in *Proc. IT Con. Sec.*, 2018, pp. 21–28.

[43] H. Mei and X. Xu, "EEG-based emotion classification using convolutional neural network," in *Proc. Int. Conf. Secur., Pattern Anal., Cybern. (SPAC)*, Dec. 2017, pp. 130–135.

Hao Chao (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in June 2012. He is currently a Lecturer with the Data Mining and Pattern Recognition Laboratory, School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo. His current research interests include pattern recognition, speech recognition, and EEG signal analysis and processing.



Liang Dong received the B.S. degree in computer science and technology from Henan Polytechnic University, Jiaozuo, China, in 2017. He is currently pursuing the Master of Engineering degree with the Data Mining and Pattern Recognition Laboratory, School of Computer Science and Technology, Henan Polytechnic University. He has published six articles in international journals and conferences. His research interests include signal processing, deep learning, pattern recognition, and machine learning.