IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 12, 2019, accepted March 21, 2019, date of publication March 29, 2019, date of current version April 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2908285

Accurate EEG-Based Emotion Recognition on Combined Features Using Deep Convolutional Neural Networks

J. X. CHEN[1,2], P. W. ZHANG[2], Z. J. MAO[3], Y. F. HUANG[3,4], D. M. JIANG[1], AND Y. N. ZHANG[1]

[1]School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China
[2]Department of Electronical and Information Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China
[3]Department of Electrical and Computer Engineering, The University of Texas at San Antonio, San Antonio, TX 78249, USA
[4]Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio, San Antonio, TX 78284, USA

Corresponding author: J. X. Chen (chenjx_sust@foxmail.com)

This work was supported by the Youth Science Foundation Projects of National Natural Science Foundation of China under Grant 61806118 (Project name: Research on EEG based emotion recognition with semi-supervised deep generative adversarial network, Project principal: J. X. Chen) and Grant 61806144 (Project name: Interpretation convolutional network based motor-cognitive joint brain network dynamically modeling and cognitive quantitative study, Project principal: Z. J. Mao).

ABSTRACT In order to improve the accuracy of emotional recognition by end-to-end automatic learning of emotional features in spatial and temporal dimensions of electroencephalogram (EEG), an EEG emotional feature learning and classification method using deep convolution neural network (CNN) was proposed based on temporal features, frequential features, and their combinations of EEG signals in DEAP dataset. The shallow machine learning models including bagging tree (BT), support vector machine (SVM), linear discriminant analysis (LDA), and Bayesian linear discriminant analysis (BLDA) models and deep CNN models were used to make emotional binary classification experiments on DEAP datasets in valence and arousal dimensions. The experimental results showed that the deep CNN models which require no feature engineering achieved the best recognition performance on temporal and frequency combined features in both valence and arousal dimensions, which is 3.58% higher than the performance of the best traditional BT classifier in valence dimension and 3.29% higher than that of BT classifier in arousal dimension.

INDEX TERMS EEG, emotion recognition, convolution neural network, combined features, deep learning.

I. INTRODUCTION

With the development of deep learning and artificial intelligence technology, emotion recognition has a broad application prospect in the field of human-computer interaction, which has been widely concerned by researchers [1]. Emotion recognition based on text, speech, facial expression and posture are appearing one after another, but these methods are subjective and cannot guarantee the authenticity of emotion. Physiological and psychological Studies have shown that changes in physiological signals tend to be much closer to people's real emotions than facial expressions, postures or voice [2]. However, the measured physiological signals such as EOG, ECG and EMG are still indirect reactions caused by emotions, which have the deficiency of lack of the reasonable evaluation criteria and low emotional recognition accuracy [3]. According to neurophysiology and psychology research, electroencephalogram (EEG) can not only reflect the various brain electrical activity and the functional state of the brain, but also can reflect the effective information of the human emotional state [4] and the generation or activity of emotion is closely related to the activity of the cerebral cortex [5]. In recent years, EEG signals have been gradually introduced into the field of emotion recognition because of their strong objectivity and high accuracy of classification [6].

Emotion recognition has achieved good classification results using the traditional machine learning classifiers. Kumar *et al.* [7] used linear kernel least squares support vector machine (LS-SVM) and back-propagation artificial neural network (ANN) to perform binary emotion recognition and achieved accuracy of 61.17% and 64.84% on valence and arousal emotional dimension respectively. Atkinson and Campos [8] combined the efficient feature selection method and the kernel-based SVM classifier to make emotion classification on the standard EEG Dataset, and gained the accuracy of 73.06% and 73.14% on valence and arousal dimensions. Chen *et al.* [9] proposed an EEG feature extraction algorithm based on the combination of Data

The associate editor coordinating the review of this manuscript and approving it for publication was Victor Hugo Albuquerque.

space adaptation (DSA) and Common spatial patterns (CSP) to alleviate the performance degradation of emotional classification caused by fluctuation and difference of day-to-day EEG signals from 12 subjects in 5 consecutive days.

The improvement of computer processing speed and computing ability makes it possible to design and implement deep learning networks. Deep Belief Network (DBN) [10], Convolutional Neural Network (CNN) [11], [12], Long/Short-Term Memory (LSTM) [13] Network and other deep network models have achieved excellent performance in computer vision, speech recognition and natural language processing. With the advancement of EEG signal acquisition technology, it becomes easier to obtain large-scale EEG data recordings and deep learning models are gradually used in the field of EEG-based emotion classification and recognition.

Suwicha *et al.* [14] applied a PCA-based principal component covariate adaptive transformation algorithm to extract the discriminant energy spectral density features of EEG signals and proposed a deep learning network based on Stacked Autoencoder to make emotion classification in valence and arousal dimensions, which classification accuracy was improved by 5.55% and 6.53% than that of traditional SVM and Naive Bayesian classifier respectively. Stober *et al.* [15] discussed the application of deep convolution automatic encoder in capturing the invariance of EEG data among different subjects in emotion classification task. Alhagry *et al.* [16] proposed a LSTM-based deep recursive neural network (RNN) which could automatically learn the features from the original EEG signals and then classify these features using a dense-layer classifier. The average subject-independent classification accuracy of this method in arousal, valence and likeness dimensions on DEAP dataset [17] was 85.65%, 85.45% and 87.99% respectively, which was much higher compared with the traditional method. Soleymani *et al.* [18] proposed a method to detect the emotional state of the subjects from their EEG signals and facial expressions in real time by using LSTM-RNN and continuous conditional random field (CCRF) algorithm and obtain better performance. Tripathi *et al.* [19] used deep convolutional neural network to classify DEAP data sets on valence and arousal emotional dimensions with accuracy of 75.58% and 73.28% respectively. Li *et al.* [20] used stacked auto encoder (SAE) and LSTM based recurrent neural network (RNN) to make emotion classification on mixed physiological signals including EEG with an accuracy of 79.26%. Salama *et al.* [21] proposed the application of 3D convolutional neural network (3D-CNN) in emotion recognition on multichannel EEG data and obtained the accuracy of 87.44% and 88.49% on valence and arousal dimension respectively.

Deep learning models can automatically end-to-end learn the abstract features from large scale raw samples, avoiding the engineering of feature extraction and feature selection. However, in the field of EEG signal recognition and classification, large-scale labeled EEG datasets are limited and many EEG-based BCI (Brain Computer Interface) applications often require high real-time performance. The application of deep learning models on EEG signals is just beginning, and the related performance still needs to be verified. Therefore, we selected to work on DEAP dataset [17] which includes large-scale EEG signals and emotion tags, extracted the time-domain, frequency-domain feature and their combined features associate to emotional dynamics, and made emotion classification on these features using deep convolutional neural network in the way like that in the field of computer vision. We took some state-of-the-art traditional machine learning classifiers as baselines which included Support Vector Machine (SVM), Bagging tree (BT), linear discriminative analysis (LDA) and Bayesian linear discriminant analysis (BLDA) models. The classification performance of our proposed deep CNN models was compared with those of baseline classifiers to verify the deep CNN models based on the combined features in time and frequency domain had higher performance on EEG emotion recognition.

The content of the paper was organized as follows, we first introduced the DEAP dataset, the data preprocessing method and details on how EEG features are extracted and combined. Next, we elaborated the EEG-based emotion classification models including the shallow machine learning classifiers, our proposed deep CNN models and the deconvolutional networks for hidden feature visualization. Then, we presented and analyzed the testing experiments and results, made comparison and discussion with the state-of-the-art shallow classifiers. Conclusion were provided in the end.

II. EEG DATASET AND FEATURE EXTRACTION

A. DEAP DATASET AND DATA PRE-PROCESSING

The experiment was carried on the DEAP dataset [17] developed by a team of researchers at Queen Mary University of London, which is a large open source dataset containing multiple physiological signals with emotional evaluations. In its data collection experiment, the induced EEG, ECG, EMG and other bioelectrical signals were detected and recorded while 32 subjects were watching 40 trials of music videos with different emotional tendencies for about 1 minute for each video. The subjects then rated the videos on a scale of 1-9 in terms of Arousal, Valence, Liking, Dominance and Familiarity. The rating value from small to large indicated that each index was from negative to positive, from weak to strong, respectively. The 40 stimulus videos included 20 high valence/arousal stimuli and 20 low valence/arousal stimuli. In this paper, we applied the first 32-channel EEG signals in DEAP dataset which had been pre-processed by down sampling to 128 Hz, band-pass filtering to 4-45 Hz, common average referencing and ocular artifacts removing by blind source separation algorithms. The duration of the denoised EEG signals in each trail is 63-s, including 60-s watching video and 3-s before watching, which have a total of 8,064 readings for each channel.

This paper extracted the 60-s EEG signals (7680 readings) induced by watching video in each trial and removed the 3s baseline signals before watching video to correct the changes of stimulus-independent signals and acquire the stimulus-related dynamics. Then in the time domain, the 60-s EEG signals was segmented into sixty 1-s epochs. The total number of EEG epochs from 40 trials of each subject was 40*60 = 2400, and the dimension of this dataset was 128 (time points) * 32 (channels)*1*2400 (epochs). Based on the emotional rating value of each video in the range of 1-9 in arousal and valence domain, the median 5 was used as the threshold to divide the rating value into two categories: more than 5 labeled with 1 meaning high arousal/valence, less than or equal to 5 labeled with 0 meaning low arousal/valence. Finally, the label data with dimension of 1*2400 corresponding to EEG signals was obtained. The dataset was then balanced to ensure that the number of two classes of labels and samples are equal, so as to reduce the impact of sample imbalance between classes on classification results and improve the normalization ability of classification models. For example, there were 1440 EEG samples with high arousal labels and 960 samples with low arousal labels for the fifth subject's dataset. We randomly extracted 960 samples from high labeled samples according to the number of low labeled samples, so that the number of two types of labels and corresponding EEG samples in each dataset was equal. Then we randomly extracted 30% of high arousal/ valence labels and 30% of low arousal/valence labels and their corresponding EEG samples as test data, the rest 70% labels and EEG samples were used as the training data, and it had been proved by experiments that such partitioning of cross-validation set could achieve better classification performance and better generalization. In this way, a 10-fold cross validation dataset of raw EEG data was constructed for each subject, which would be used to make feature extraction, train and test the shallow classifiers and our proposed deep CNN models as following methods.

B. EEG FEATURE EXTRACTION

In this paper, EEG features were extracted from time domain and frequency domain respectively. The amplitude at each time point of the pre-processed EEG data was taken as the original time domain feature which was called RAW feature. The RAW feature was then normalized by channels which was called NORM feature. Both RAW and NORM features had the same size of 128(time points)*32(channels)*1* epochs. In frequency domain, EEG signals always contain much information of rhythm frequency band. Many researches of neuroscience and psychology showed that five frequency bands of Delta (1Hz ∼ 4Hz), Theta (4Hz ∼ 8Hz), Alpha (8Hz ∼ 13Hz), Beta (13Hz ∼ 30Hz) and Gamma (above 30Hz ∼ 47Hz) in EEG signals were closely related to emotional and other psychological activities [22]. So, on these five frequency bands and the whole band (4-45Hz), we separately applied the fast Fourier algorithm to extract 64 power spectral density (PSD) features by sliding 0.5s Hamming window with 0.25s step along 1s EEG epoch on each channel. The experimental results showed that the PSD feature in the whole frequency band had better classification performance than those in the other sub-frequency bands, so we took the PSD feature with size of 64(features points) *32 (channels)*1* epochs in the whole frequency band as the typical frequency domain feature and called it FREQ feature for short. Because the combination of time domain and frequency domain features includes both global time and local frequency dynamics, it can better capture the emotional correlation in time and frequency domain synchronously. Therefore, we proposed to combine the RAW and NORM features in time domain with FREQ features separately and got the combined features which were called FREQRAW feature and FREQNORM feature. Before the combination, the RAW and NORM features were firstly down sampled to 64Hz to have the same size with the FREQ feature, and then was combined with FREQ feature separately by concatenating them in the third dimension, which made the size of the combined FREQRAW or FREQNORM feature was 64(feature points)*32(channels)*1* epochs. In the following experiments, we mainly dealt with these five types of EEG features including RAW, NORM, FREQ, FREQRAW and FREQNORM.

III. EEG-BASED EMOTION CLASSIFICATION MODELS

A. SHALLOW MACHINE LEARNING MODELS

In the DEAP Dataset, the participants' evaluation of each video was based on two-dimensional affective model of valence and arousal, and the rating values of these two indexes were all consecutive numbers from 1-9, representing the tendency from negative to positive, from weak to strong. Therefore, we chose to classify human emotion-related EEG data in valence and arousal dimensions. The valence index reflects the level of people's feeling of pleasure, the higher the valence, the more positive and happier, and the lower the valence, the more negative and sadder. The arousal index reflects the intensity of people's feeling, the higher the arousal, the more obvious and stronger the feeling, and the lower the arousal, the more implicit and weaker the feeling. We took four kinds of traditional shallow machine learning models including BT, LDA, BLDA and SVM which had shown good performance in our previous EEG-based emotion recognition study [23] as the benchmark to carry out binary emotion classification experiments.

Bagging Tree (BT) is a kind of supervised classification algorithm which combines a group of weak decision tree classifiers into a strong classifier through the iteration. Given a group of samples and each sample has a set of attributes and a predetermined category. In each iteration, 70% samples are randomly selected to form the sub training set D_t to train the t^{th} weak classifier and then put these samples back to repeat the same iteration 100 times. Finally, the category voted most by these 100 weak classifiers is chosen as the final classification result by voting. The purpose of linear discriminant analysis is to aggregate homogeneous samples and disperse heterogeneous samples. The Bayesian linear discriminant analysis (BLDA) algorithm calculates the posteriori probability and error probability of each sample and classifies the sample by estimating the maximum posteriori probability to minimize the expected loss. The core idea of the Support vector machine is to determine an optimal hyperplane, so that the samples fall in either side of the hyperplane and the distance between the sample and the hyperplane should be as large as possible. The SVM classifier can not only get the classification result with low error rate, but also make good classification decision on the test data out of the training set. In this paper, we applied the linear kernel function for SVM model for binary classification experiments.

B. DEEP CONVOLUTION NEURAL NETWORK MODELS

Deep learning aims to automatically learn and extract multi-level feature representation from raw data [24]. Convolutional neural network is a typical and widely used model for deep learning. The characteristics of convolutional neural network, such as local connection, weight sharing and down sampling operation, make it possible to effectively reduce the complexity of the network, reduce the number of training parameters and present the advantages of strong robustness and fault tolerance, as well as easy to train and optimize [25]. In this paper, we applied the convolutional neural network like that used in computer vision to make binary emotion classification on features from time domain, frequency domain and their combination in valence and arousal dimensions.

For deep CNN models, the essence of convolution kernel is to extract the deep and abstract information of input signals automatically. The convolution kernel is calculated as follows:

$$x_j^l = f(\sum_{i \in M_j} x_i^{l-1} * \boldsymbol{w}_{ij}^l + \boldsymbol{b}_j^l) \tag{1}$$

where x_j^l is the j^{th} feature of the layer l; w_{ij}^l is the connecting weight of the j^{th} feature of the layer l and the i^{th} feature of the layer l-1; b_j^l is the offset of the j^{th} feature of the layer l; * is the convolution operation; $f(\bullet)$ is the activation function.

The ReLU [26] function is often used as the activation function because it is simple to implement, can accelerate the calculation and convergence, has no saturation problem, and greatly alleviates the phenomenon of gradient dissipation. The formula of the ReLU function is as follows:

$$f(x) = \max(0, x) \tag{2}$$

The pooling layer is a structure for down sampling the features obtained from the convolutional layers, which can reduce the amount of computation and the degree of overfitting of the network to some extent, thus improving the performance of the CNN model [27]. The pooling method includes average pooling and maximum pooling. In this paper, we applied the maximum pooling layer to divide the convolutional features into several disjoint regions of n*n and used the maximum value of each region to represent the down sampled convolutional feature. After the output of the last convolutional layer, two fully connected layers were appended to combine all the extracted features together. A dropout layer was added after the second full connection layer which could significantly reduce the overfitting of the model by ignoring a part of the feature detectors (making a part of node value in hidden layer 0) in each training batch. Finally, a Soft-max classification layer was connected. It was an extended supervised learning algorithm based on logistic regression, which was often used in combination with deep learning or unsupervised learning algorithm.

Due to the temporal and spatial correlation of EEG signals, we considered to use eight convolutional filters on local/global and space/time combinations and one convolutional filter the same as that of the CNN model used for computer vision (CVCNN). These eight combined convolutional filters were separately labeled as LS, LT GS, GT, LSLT, LSGT, GSLT and GSGT. LS stood for local spatial filter, LT stood for local time filter, GS stood for global spatial filter, GT stood for global time filter, LSLT stood for local time local spatial filter, LSGT stood for local space global time filter, GSLT stood for global space local time filter and GSGT stood for global space global time filter. Our previous experimental results [28], [29] had shown that CVCNN, GSCNN and GSLTCNN models presented better performance than other filter models in EEG-based rapid sequence visual presentation (RSVP) event classification. So, we mainly applied these three deep CNN models to make EEG-based emotion classification.

In EEG-based emotion classification experiment, the selection of parameters for CNN models is very important, including the number and setting of convolutional layers, pooling layers, feature maps and the full connection layers and the selection of optimization algorithms. In order to balance the depth of the model with the limited training samples, we defined a searching space for selecting model parameters: the number of convolutional layers and pooling layers is ranging from 1 to 3, the number of full connection layers is ranging from 1 to 2, and the number of feature maps of each convolutional layer is ranging from $2^3 to 2^7$. By training and testing the above three CNN models configurated with each combination of parameters from the searching space, the model parameters with the best classification performance was selected. During the searching process, to simplify the calculation, we set the batch size as 20, the size of convolution kernel in pooling layer was 2*2, the size of convolution kernel in CVCNN model was 5*5, the size of convolution kernel in GSLTCNN model was 20*5, and the size of convolution kernel in GSCNN model was 20*1. Four convolution kernels with step 1 were used in each convolutional lay of three CNN models. The full connection layer was put after the last convolutional layer and its pooling layer, a Soft-Max classifier which had two classes of output was appended at the end of the network. ReLU function was used as the activation function and cross entropy was used as the loss function of the network. We set the rate of dropout as 85% to avoid overfitting. All CNN models were optimized by stochastic gradient descent algorithm.

Through experiments and comparisons of various models in search space, the best mode parameters were determined which meant the CVCNN, GSCNN and GSLTCNN models with 2 convolution layers, 2 pooling layers, 2 fully connected layers and 1 Soft-Max layer were selected as the optimal models for emotion classification on DEAP dataset which first convolution layer had 128 feature maps and second convolution layer had 64 feature maps. The structure of our deep CNN models was shown as Fig.1.

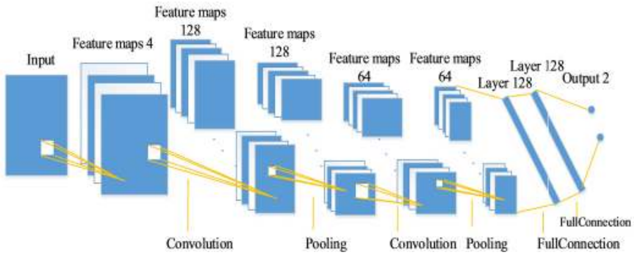


FIGURE 1. **Structure of our deep CNN models.**

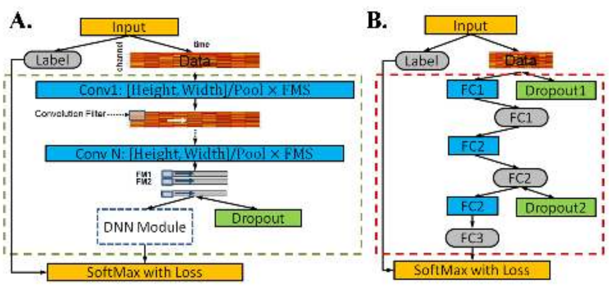


FIGURE 2. **The architecture of the proposed deep CNN model. A. The detailed structure of deep CNN model. There are N convolutional layers, where each blue box represents a convolutional operation and the texts inside represent [kernel shape] / MP width × Feature Map Size. "FM" denotes feature map. B. The detailed architecture of the DNN Module in A. It contains fully connected modules (blue boxes) and hidden units (gray ovals).**

The architecture of the proposed deep CNN was shown as Fig.2. Multiple filters or kernels were convolved with the input data in terms of vectorized EEG epochs in each convolutional layer, which were designed to capture different local temporal and spatial EEG features. The output of a convolution layer from one kernel was called a feature map (FM). All the output feature maps were combined by the fully connected layers at the end of the last convolution layer. The pseudocode of our proposed deep CNN procedure was as Table 1.

C. DECONVOLUTIONAL NEURAL NETWORK FOR VISUALIZING FEATURES OF DEEP CNN MODELS

To make our proposed deep CNN models easy to be understood, we applied the deconvolutional network [30] on our trained CNN models to visualize the features extracted by them. During the implementation of a deconvolutional network, the most activated hidden unit on the top convolution layer after pooling was firstly selected. A forward propagation was used to produce a set of hidden units in the top convolutional layer to determine the most activated unit for each epoch. Then, we calculated the max L1 norm for each hidden unit and the unit with the largest L1 norm was fed into the deconvolutional network. The deconvolution algorithm shown as Fig. 3 was like that proposed in [31], which consisted of 3 steps to deconvolve the activated hidden units and map them back to the input EEG.

TABLE 1. **Pseudo code of GSLT-CNN for subject identification.**

Line	pseudo code
1:	Collect data from 4 experiments as $\boldsymbol{v} \in \mathcal{R}^{M\times T}$, M as channel size, T as time samples
2:	Initialize CNN filter weight $\boldsymbol{W}_{ct}^{ik}$ for $\boldsymbol{c}^{th}$ ($\boldsymbol{c} = \boldsymbol{1,2,\ldots,M}$ as channel size) channel, $\boldsymbol{t}^{th}$ ($\boldsymbol{t} = \boldsymbol{1,2,\ldots,T'}$ as time sample size) time sample, $\boldsymbol{k}^{th}$ ($\boldsymbol{k} = \boldsymbol{1,\ldots K}$ as feature map size) kernel, and $\boldsymbol{i}^{th}$ ($\boldsymbol{i} = \boldsymbol{1,\ldots N}$ as convolutional layer size) layer.
3:	Initialize CNN fully connected weight $\boldsymbol{U}_{p,q}^{j}$, where $\boldsymbol{p}$ is the feature size of the top convolutional layer, $\boldsymbol{q}$ is the size of hidden units, and j is the size of fully-connected layers.
4:	For $\boldsymbol{i} < \boldsymbol{N}$ ($\boldsymbol{N}$ is the convolutional layer)
5:	$\boldsymbol{convolution(v)_{kt} = ReLU\left(\sum_{c,t}^{M,T} W_{ct}^{ik} v\right)}$
6:	$\boldsymbol{v \leftarrow convolution(v)_{kt}}$
7:	$\boldsymbol{v \leftarrow pooling(v)}$
8:	Iteration level: $\boldsymbol{i \leftarrow i + 1}$
9:	For $\boldsymbol{j} < \boldsymbol{F}$ ($\boldsymbol{F}$ is the fully-connected layer)
10:	$\boldsymbol{fc(v) = ReLU\left(\sum_{p,q}^{j} U_{p,q}^{j} v\right)}$
11:	$\boldsymbol{v \leftarrow fc(v)}$
12:	Iteration level: $\boldsymbol{j \leftarrow j + 1}$
13:	Back-propagate Loss ($\boldsymbol{v}$, $\boldsymbol{L}$), $\boldsymbol{L}$ as the true label

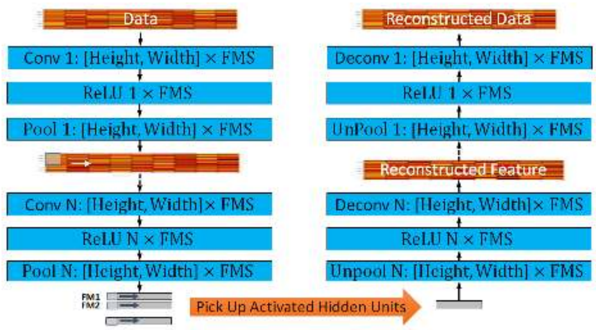


FIGURE 3. **Illustration of deconvolutional network for reconstructing the activated hidden units.**

The first step was to make a reverse operation of pooling that was called unpooling. For our CVCNN, GSCNN and GSLTCNN, the max pooling operation was non-invertible. But we could get an approximate inverse value by marking the locations of the maxima within each pooling region with a set of switch variables. Thus, these switches were used to place the reconstructions from the layer above into the appropriate locations to preserve the structure of the stimulus [32], while other locations were filled with zeros instead. The second step was rectification, where the function of ReLU was applied to the unpooling output. The third step was unfiltering. As the convolution filters in our CVCNN, GSCNN and GSLTCNN were all linear operations, the reverse operation of filtering convolved the feature maps from upper layer using the transposed versions (flipping each filter vertically and horizontally) of the same filters applied on the rectified maps after step 2. After repeating these deconvolution operations for each CNN module, we could visualize the reconstructed EEG patterns underlying the most activated hidden units.

IV. EXPERIMENTS AND RESULTS DISCUSSION

For each type of RAW, NORM, FREQ, FREQRAW and FREQNORM features extracted and balanced from each subject, 30% of high-class labels and 30% of low-class labels and their corresponding EEG data were randomly extracted as the test set, the rest 70% of the labels and EEG data were used as the training set, and 10% of the training data was then randomly selected as the verification set while training began. In this way, a 10-fold cross-validation set was constructed for each type of feature of each subject, which was used as input to train and test the shallow baseline classifiers and our proposed deep GSCNN, GSLTCNN and CVCNN models.

The area under the ROC (Receiver Operating Characteristic) curve called as AUC (Area Under roc Curve) was used as a measurement to evaluate the performance of the above classification models. The value of AUC usually ranges from 0.5 to 1 and the larger the AUC value is, the better the model is. If AUC $=$ 1, it indicates when using this model for prediction, the perfect prediction result can be obtained no matter what threshold is set. While if AUC $= 0.5$, it indicates the model cannot complete the classification or prediction work. The average AUC of each classifier on 10-fold cross-validation set was taken as the performance metric of this classifier on the corresponding feature of each subject. The AUC values of 32 subjects was then averaged as the final within-subject classification performance of the corresponding model on corresponding type of feature.

The aim of our experiments was to evaluate the binary classification performance of four shallow baseline classifiers and three CNN models on valence and arousal dimensions to find out the optimal classification model and the feature with the best performance. we also reconstructed and visualized the hidden units learned from CVCNN model by deconvolution network to show the CVCNN model could learn the discriminative emotional features.

A. EMOTION CLASSIFICATION IN VALENCE

In valence dimension, we applied four shallow classifiers including BT, LDA, BLDA, and SVM and three deep CNN models including CVCNN, GSCNN, and GSLTCNN to make within-subject binary (high/low) emotion classification. The experimental results were shown in Table 2. From that we could see in valence dimension, BT, BLDA and SVM models all showed the best performances on the FREQ features in frequency domain, which were much higher than those on RAW and NORM features in time domain, and little higher than those on FREQRAW and FREQNORM combination features except SVM showed better performance on FREQRAW feature. The performances of three CNN models on FREQ features were also much higher than those on the RAW and NORM features, and the recognition performances on the FREQRAW and FREQNORM combination features were still much higher than those on all single features including FREQ feature.

TABLE 2. The average auc of 7 models on 5 features of 32 subjects in valence.

Models	RAW	NORM	FREQ	FREQRAW	FREQNORM
BT	0.5889	0.5822	0.9642	0.9250	0.9254
BLDA	0.5120	0.5134	0.8115	0.6786	0.8093
SVM	0.5590	0.5591	0.7596	0.9234	0.7460
LDA	0.5147	0.6453	0.5147	0.6361	0.6361
CVCNN	0.6221	0.6551	0.9307	0.9933	0.9997
GSCNN	0.6242	0.6394	0.8875	0.9933	1
GSLTCNN	0.6716	0.6350	0.8523	0.9946	1

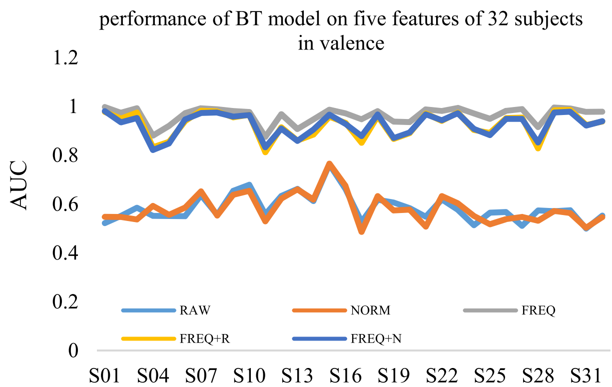


FIGURE 4. Classification performance of BT model on five features of 32 subjects in valence dimension.

For traditional shallow classifiers, the BT model showed the best recognition performance, with the best average AUC of 96.42% on the FREQ feature and the best average AUC of 92.54% on the combined FREQNORM feature. The binary valence classification results of BT model on five features of 32 subjects were shown in Fig.4, from which we found out the EEG features of the 15th subject presented the best recognition performance. The LDA model has the best performance in identifying time domain features with an average AUC of 64.53%. The SVM model has the best recognition performance on the combined FREQRAW features with an average AUC of 92.34% and an average ACC of 87.07%. Compared with the binary valence classification result of literature [15], the accuracy of SVM model was increased by 19.86%.

For three types of deep CNN models, their average AUC performances on FREQNORM features were all approximately 1, which were significantly higher than those on single frequency or time features, which showed that our proposed deep CNN models could obtain nearly perfect performance on the combined FREQNORM feature for predicting high or low valence of emotion. Among them, the deep

CVCNN model with convolution kernel of 5*5 had the best average AUC performance on total five features, reaching 84.02%, which was 4.31% higher than the average AUC of 79.71% of the best shallow BT model on total five features. Additionally, the average ACC of deep CVCNN model was up to 88.76%, which was 9.5% higher than the best accuracy of 79.26% of the LSTM RNN model proposed in literation [15] for binary valence classification on the same DEAP dataset. The performance of the CVCNN model on five features of 32 subjects was shown in Fig.5, from which we also found out the 15th subject presented the best recognition performance and the AUC values of all subject on FREQNORM and FREQRAW features were significantly higher than those on single time or frequency features.

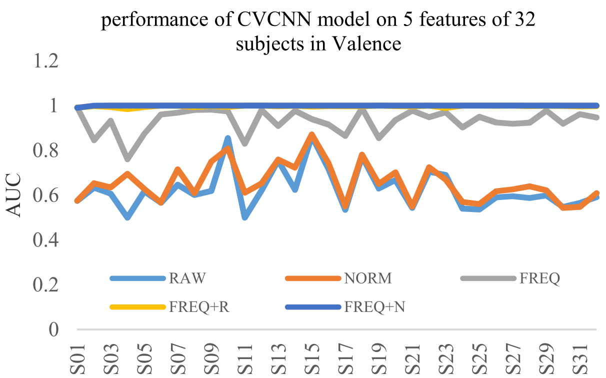
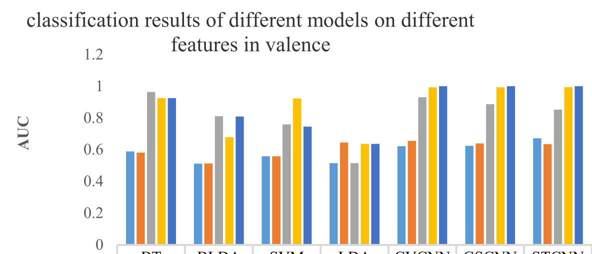


FIGURE 5. Classification performance of CVCNN model on 5 features of 32 subjects in valence dimension.



	BT	BLDA	SVM	LDA	CVCNN	GSCNN	STCNN
RAW	0.5889	0.512	0.559	0.5147	0.6221	0.6242	0.6716
NORM	0.5822	0.5134	0.559	0.6453	0.6551	0.6394	0.635
FREQ	0.9642	0.8115	0.7596	0.5147	0.9307	0.8875	0.8523
FREQRAW	0.925	0.67868	0.9234	0.6361	0.9933	0.9933	0.9946
FREQNORM	0.9254	0.8093	0.746	0.6361	0.9997	1	1

FIGURE 6. Comparison of classification results of different models on different features in valence.

The comprehensive comparison Histogram of the classification performances of shallow models and deep models in valence dimension were shown in Fig.6. On the RAW feature in the time domain, the deep GSLTCNN model showed the highest classification performance, with an average AUC of 67.16%, which was 8.27% higher than the average AUC of 58.89% of the best performing BT model in the shallow classifiers. On the NORM feature in the time domain, the deep CVCNN model showed the highest classification performance, with the average AUC of 65.51%, which was 0.98% higher than the average AUC of 64.53% of the best performing LDA model in the shallow classifiers. On the FREQ feature in the frequency domain, BT model showed the best performance, with an average AUC of 96.42%, which was 3.35% higher than the average AUC (93.07%) of the best performing deep CVCNN model. On the FREQRAW combination feature, the deep GSLTCNN model showed the best performance, with an average AUC of 99.46%, which was 6.96% higher than the average AUC of 92.5% of the best performing BT model in the shallow classifiers. On the FREQNORM combination feature, the deep GSLTCNN model showed the best performance, with an average AUC of 99.99%, which was 7.45% higher than the average AUC of 92.54% of the best performing BT model in the shallow classifiers. Thus, it could be seen that the EEG-base emotion classification performances of deep CNN models were almost better than those of shallow classifiers on four types of features, especially on combination features of time domain and frequency domain.

B. EMOTION CLASSIFICATION IN AROUSAL

In arousal dimension, we also applied four shallow classifiers including BT, LDA, BLDA, and SVM and the deep CVCNN, GSCNN, and GSLTCNN modes to make within-subject binary emotion classification. The experimental results were shown in Table 3.

TABLE 3. The average AUC of 7 classification models on 5 features of 32 subjects on arousal.

	RAW	NORM	FREQ	FREQRAW	FREQNORM
BT	0.9671	0.5517	0.5067	0.9441	0.9442
BLDA	0.7963	0.5032	0.5036	0.6743	0.7951
SVM	0.7525	0.5590	0.5531	0.9462	0.7353
LDA	0.6395	0.5486	0.5486	0.6327	0.6327
CVCNN	0.6012	0.6176	0.8851	0.9988	1.0000
GSCNN	0.5902	0.5987	0.8802	0.9930	1.0000
GSLTCNN	0.6175	0.5670	0.8390	0.9958	1.0000

From Table 3 we found out in arousal dimension, different from the classification results in valence dimension, the recognition performances of BT, BLDA and LDA models on RAW features in time domain were almost much higher than those on NORM and FREQ features and little higher than those on FREQRAW and FREQNORM combination features. While the performances of three CNN models on FREQ features were still much higher than those on the RAW and NORM features, and the recognition performances on the FREQRAW and FREQNORM combination features were still much higher than those on all single features including FREQ feature.

For the shallow classifiers, the BT model still had the best recognition performance, with the average AUC of 96.71% on the RAW feature in time domain and the average AUC of 94.42% on the FREQNORM combination feature. The result of binary arousal classification of 32 subjects on five

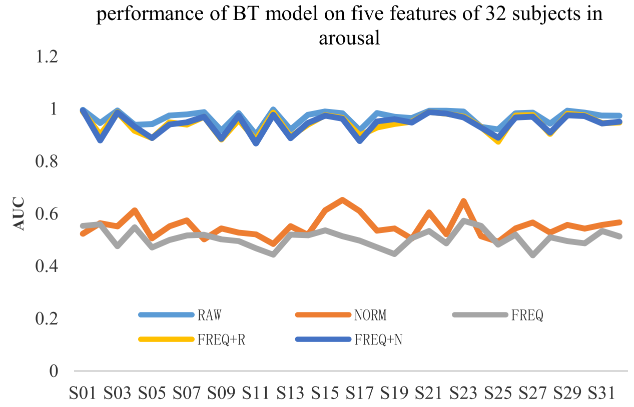


FIGURE 7. Classification performance of BT model on five features of 32 subjects in arousal dimension.

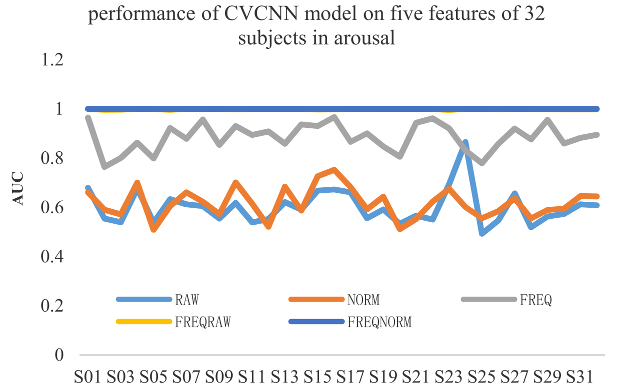
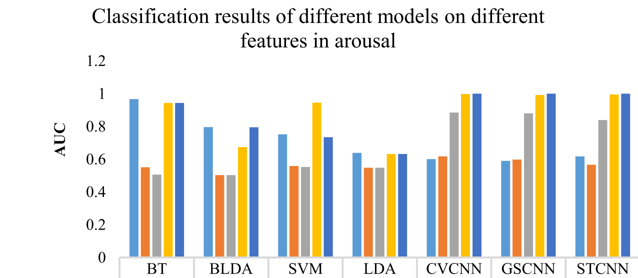


FIGURE 8. Classification performance of CVCNN model on 5 features of 32 subjects in arousal dimension.



	BT	BLDA	SVM	LDA	CVCNN	GSCNN	STCNN
RAW	0.9671	0.7963	0.7525	0.6395	0.6012	0.5902	0.6175
NORM	0.5517	0.5032	0.559	0.5486	0.6176	0.5987	0.567
FREQ	0.5067	0.5036	0.5531	0.5486	0.8851	0.8802	0.839
FREQRAW	0.9441	0.6743	0.9462	0.6327	0.9988	0.993	0.9958
FREQNORM	0.9442	0.7951	0.7353	0.6327	1	1	1

FIGURE 9. Performance comparison histogram of different models on different features in arousal.

features by BT model was shown in Fig.7, from which we found out the EEG features of the 23th subject presented the best recognition performance. While the SVM model showed the best recognition performance on combined FREQRAW feature in time-frequency domain, with an average AUC of 94.62% and average ACC of 86.98%. Compared with the binary arousal classification result of literature [15], the accuracy of EEG-based emotion classification by SVM was increased by 18.43%.

For three types of deep CNN models, their average AUC performances on the FREQNORM feature were all approximately 1, which were significantly higher than those on single frequency or time domain features. It showed our proposed deep CNN models could obtain nearly perfect prediction results when they were used to predict high or low arousal of emotion on the FREQNORM combination features. Among them, the deep CVCNN model with convolution kernel of 5*5 still had the best average AUC performance on total five features, reaching 82.05%, which was 3.77% higher than the average AUC of 78.28% of the best shallow BT model on total five features. Additionally, the average ACC of deep CVCNN model was up to 85.57%, which was 10.22% higher than the best accuracy of 75.35% of the LSTM RNN model proposed in literation [15] for binary arousal classification on the same DEAP dataset. The performance of the deep CVCNN model on five features of 32 subjects was shown in Fig.8, from which we found out the 16th subject presented the best recognition performance.

The comprehensive comparison Histogram of the classification performances of shallow models and deep models in arousal dimension were shown in Fig.9. On the RAW feature in the time domain, the BT model showed the highest classification performance, with an average AUC of 96.71%, which was significantly higher than those of the three depth CNN models and other shallow models. On the NORM feature in the time domain, the deep CVCNN model showed the highest classification performance, with the average AUC of 61.76%, which was 5.86% higher than the average AUC of 55.9% of the best performing SVM model in the shallow classifiers. On the FREQ feature in the frequency domain, the CVCNN model still showed the best performance, with an average AUC of 88.51%, which was 33.2% higher than the average AUC of 55.31% of the best performing SVM model in the shallow classifiers. On the FREQRAW combination feature, the deep CVCNN model showed the best performance, with an average AUC of 99.88%, which was 6.26% higher than the average AUC of 94.62% of the best performing SVM model in the shallow classifiers. On the FREQNORM combination feature, the average AUC values of the three deep CNN models were all approximately 1, which were 5.58% higher than the average AUC value of 94.42% of the best performing BT model in the shallow classifiers. Thus, it could be seen that the EEG-base emotion classification performances of deep CNN models were almost better and robust than those of shallow classifiers on four types of features, especially on combination features of time domain and frequency domain.

C. PERFORMANCE EVALUATION OF CNN MODELS

As shown in Fig.6 and Fig.9, the CVCNN model had the general best performance in EEG-based emotion classification in valence and arousal dimensions, and its classification performance on the single FREQ feature was always higher than that on RAW or NORM feature in time domain. In our opinions, its best performance close to 1 on the FREQNORM and FREQRAW combination features was largely due to its

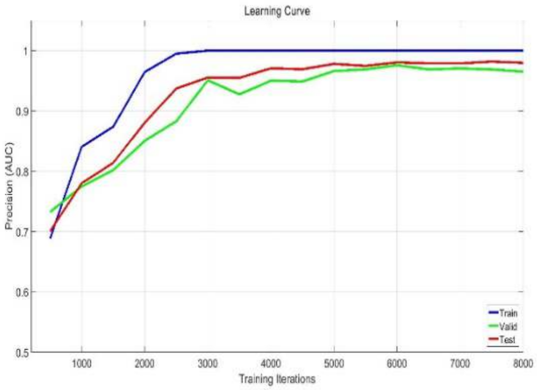


FIGURE 10. Learning curve of CVCNN model on FREQ feature in valence classification.

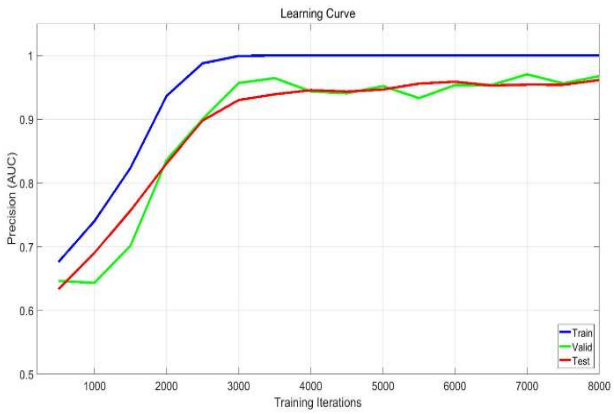


FIGURE 11. Learning curve of CVCNN model on FREQ feature in arousal classification.

higher performance on the FREQ feature. Therefore, we plotted the learning rate curve of the CVCNN model on the FREQ feature in valence and arousal dimension separately shown in Fig.10 and Fig.11 to evaluate the learning performance of the model.

As seen from Fig.10 and Fig.11, in both valence and arousal dimensions, the classification accuracy on the training set increased rapidly with the increase of iteration times and the overfitting on the training set was achieved after about 3K iterations. But during the training process, the classification accuracy on test set and validation set both increased slowly, the learning curve on verification set fluctuated slightly while increasing, and the learning curve on test set fluctuated little and stops growing after about 4K iterations, which indicated that the deep CVCNN model had good and stable classification performance on FREQ feature in frequency domain. The learning rate curves of GSCNN and GSLTCNN in valence and arousal dimensions were almost the same as that of CVCNN, so there was no more repetition.

D. FEATURE RECONSTRUCTURE AND VISUALIZATION

WE also reconstructed and visualized the hidden units learned from CVCNN model by deconvolution network to show the CVCNN model could learn the discriminative

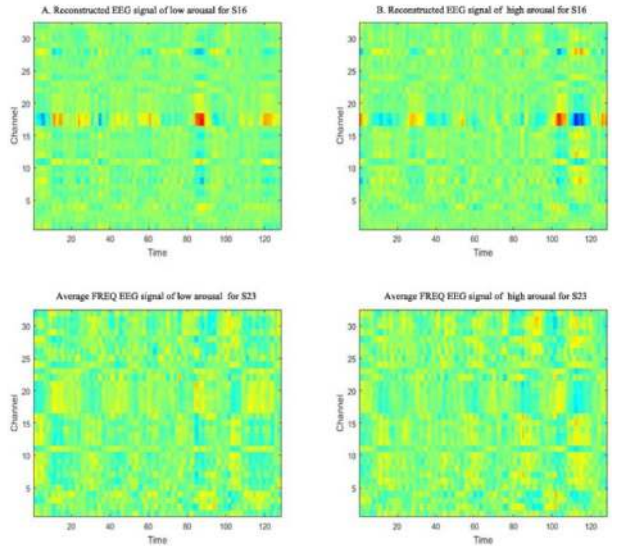


FIGURE 12. The RAW average EEG signal and reconstructed by the most activated CVCNN hidden units of low/high arousal from subject 23. A) The RAW average EEG signal (bottom) and reconstructed EEG signal (top) of low arousal from subject 23. B) The RAW average EEG signal of high arousal (bottom) and reconstructed EEG signal (top) of high arousal from subject 23. Horizontal axis is the time point of 1 second epoch, and the vertical axis is the channel. Deconvolutional network was used for reconstruction. The color axis is scaled from −3 to 3.

emotional features. Fig.12 showed the Average FREQ EEG signal and the reconstructed FREQ EEG signal by the most activated CVCNN hidden units from subject 16 who had the best performance by CVCNN model in arousal dimension. As we could see, the reconstructed signal of this subject was different from the average EEG signal and it showed some task specific features. For example, for the low arousal task, the highlighted reconstructed features existed between 600-700ms, while these kinds of features did not appear in the high arousal task. For the high arousal task, the highlighted reconstructed features existed between 800-850ms, while these kinds of features do not appear in the low arousal task. The same conditions also existed in the original average FREQ EEG signals in low/high arousal tasks, but the different pattern was not so obvious as that in the constructed signals, which showed that the CVCNN model could learn more discriminant hidden features for emotion classification.

V. CONCLUSION

In this paper, the traditional machine learning and deep learning models for EEG-based emotion classification were established, and their classification performances were verified on the EEG signals of DEAP dataset. We firstly extracted the preprocessed and denoised 60-s 32-channel EEG data induced by watching video from 40 trials in DEAP dataset for each of 32 subjects, then removed the 3-s baseline signals before watching video to correct the changes of stimulus-independent signals, then segmented the 60-s EEG signals into sixty 1-s epochs to make the dimension of each EEG sample similar to that of the image, such as 128 (time points)*32 (channels)*1*2400 (epochs), and finally the dataset was balanced to ensure that the number of two classes of labels and samples are equal to improve the normalization ability of classification models. We obtained five types of features including the RAW and NORM features in time domain, FREQ feature in frequency domain, and the combination features of FREQRAW and FREQNORM in both time and frequency domain through the way introduced above. Then, 10-fold cross validation dataset was constructed for each type of feature of each subject, and was input into 4 shallow classifiers and 3 deep CNN models to make binary within-subject emotion classification experiments in valence and arousal dimensions.

The experimental results showed that the deep CNN models always performed better and more stably on FREQ feature and the FREQRAW and FREQNOR combination features, and the best classification performance of deep CNN models on combination features was obviously better than the best performance of shallow classifiers. Although the shallow BT model showed the optimal performance on single FREQ or RAW EEG feature, its performance was not stable and consistent, for example, it showed the best performance on FREQ feature in valence dimension, but in arousal dimension its prediction performance on FREQ feature was very poor, on the contrary, it showed the best performance on the RAW feature.

So, it was concluded that based on the time and frequency combination EEG features, our proposed deep CNN model similar to that used for image classification in computer vision could automatically learned the discriminant stimulus-related EEG dynamics end-to-end and achieve the optimal and robust performance of binary emotion classification in valence and arousal dimensions, which not only avoided the large engineering of manual feature extraction and feature selection before traditional machine learning classification, but also improve effectively improves the accuracy and stability of EEG emotion recognition. It also provided a valuable method for developing high performance brain-computer interface for EEG-based emotion recognition and regulation.

Although good experimental results have been obtained in our present study, further research is still needed on how to extract more discriminative EEG features to make cross-subject emotion classification, how to select, construct and optimize deep learning models with higher accuracy, robustness and generalization for EEG-based emotion recognition, and how to incorporate some emotional-related brain neurogenic analysis into the analysis of experimental results. All these are the main contents of our next research work.

REFERENCES

[1] G. M. Lu, L. Yuan, W. J. Yang, J. J. Yan, and H. B. Li, "Speech emotion recognition based on long-term memory and convolutional neural network," *J. Nanjing Univ. posts Telecommun.*, vol. 38, no. 5, pp. 1–7, Nov. 2018. doi: 10.14132/j.cnki.1673-5439.2018.05.009.

[2] Y. F. Chen, Y. L. Cui, and S. X. Wang, "Review of emotion recognition technology based on physiological signals," *System Simul. Technol.*, vol. 13, no. 1, pp. 1–5, Sep. 2017.

[3] J. F. Jiang, Y. Zeng, Z. M. Lin, T. Li, and B. Yan, "Research review of emotion assessment based on brain electrical signals," *J. Inf. Eng. Univ.*, vol. 17, no. 6, pp. 693–866, May 2016.

[4] Y. Sun, J. J. Ma, and X. Y. Zhang, "EEG emotion recognition based on nonlinear global features and spectral features," *Comput. Eng. Appl.*, vol. 54, no. 17, pp. 116–121, Jul. 2008.

[5] J. Li, G. Z. Liu, and J. Gao, "Emotion classification based on EEG signal," *J. Beijing Univ. Inf. Sci. Technol.*, vol. 32, no. 2, pp. 34–39, May 2017.

[6] J. R. Zhang and G. Wang, "Emotion recognition based on EEG research," *Comput. Appl. Res.*, vol. 36, no. 11, pp. 1–6 Nov. 2018. [Online]. Available: http://kns.cnki.net/kcms/detail/51.1196.TP.20180811.1329. 062. html

[7] N. Kumar, K. Khaund and S. M. Hazarika, "Bispectral analysis of EEG for emotion recognition," *Procedia Comput. Sci.*, vol. 84, pp. 31–35, May 2016, doi: 10.1016/j.procs.2016.04.062.

[8] J. Atkinson and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," *Expert Syst. Appl.*, vol. 47, pp. 35–41, Apr. 2015. doi: 47.10.1016/j.eswa.2015.10.049.

[9] J. X. Chen, R. Zheng, P. W. Zhang, and X. Y. Jia, "EEG emotion classification based on data space adaptation and Common Spatial Pattern," *Comput. Eng.*, to be published. doi: 10.19678/j.issn.1000-3428.0050948.

[10] G. E. Hinton, S. Osindero, and Y. W. The, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[11] P. Y. Simard, D. Teinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. 7th Int. Conf. Document Anal. Recognit.*, Aug. 2003, pp. 958–963.

[12] A. Krizhevsky, S. Ilya, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 26th Annu. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] 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, pp. 1–10, Sep. 2014.

[15] S. Stober, A. Sternin, and A. M. Owen, "Deep feature learning for EEG recordings," *Comput. Sci.*, vol. 165, pp. 23–31, Jun. 2015.

[16] S. Alhagry, A. Aly, and A. Reda, "Emotion recognition based on EEG using LSTM recurrent neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, p. 10, Aug. 2017, doi: 10.14569/IJACSA.2017.081046.

[17] 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./Mar. 2012.

[18] 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./Mar. 2016.

[19] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on DEAP Dataset," in *Proc. 19th IAAI Conf.*, Feb. 2017, pp. 4746–4752, Feb. 2017.

[20] Y. J. Li, J. J. Huang, H. Y. Wang, and N. Zhong, "Research on multi-mode physiological signal fusion and emotion recognition based on SAE and LSTM RNN," *J. Commun.*, vol. 38, no. 12, pp. 109–120, Jun. 2017.

[21] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Shalaby, "EEG-based emotion recognition using 3D convolutional neural networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 329–337, Jan. 2018.

[22] Balconi, Michela and G. Mazza, "Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues. ERS/ERD and coherence measures of alpha band," *Int. J. psychophysiol.*, vol. 74, no. 2, pp. 58–65, Jan. 2009.

[23] Y. Liu, O. Sourina, and M. K. Nguyen, "Real-time EEG-based emotion recognition and its applications," *Trans. Comput. Sci. XII*, vol. 6670, pp. 256–277, Apr. 2011, doi: 10.1007/978-3-642-22336-5_13.

[24] J. Y. Zhang, H. L. Wang, Y. Guo, and X. Hu, "Research review on deep learning," *Comput. Appl. Res.*, vol. 35, no. 07, pp. 1921–1928, Jun. 2018.

[25] J. Liu, Q. Yuan, G. Wu, and X. Yu, "Review of convolutional neural networks," *Comput. Era*, vol. 11, pp. 19–23, Mar. 2018.

[26] K. O. , "Shea, and R. Nash. (2015). "An introduction to convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1511.08458

[27] J. Zhang, S. Li, and Z. Yin, "Pattern classification of instantaneous mental workload using ensemble of convolutional neural networks," *IFAC-Papers OnLine*, vol. 50, no. 1, pp. 14896–14901, Jul. 2017.

[28] S. Ahmed, L. M. Merino, Z. Mao, J. Meng, K. Robbins, and Y. Huang, "A deep learning method for classification of images RSVP events with EEG data," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 33–36.

[29] M. Hajinoroozi, Z. Mao, and Y. Huang, "Deep transfer learning for cross-experiment prediction of rapid serial visual presentation events," in *Proc. IEEE Brain Comput. Interface Meeting*, Jun. 2013, pp. 55–60.

[30] Z. Mao, W. Yao, and Y. Huang, "EEG based biometric identification with deep learning,"in *Proc. 8th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, May 2017, pp. 1–10.

[31] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 8689. 2014, pp. 818–833.

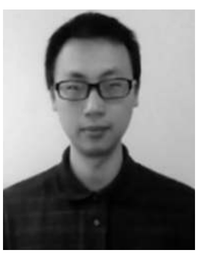
[32] Z. Mao, W. Yao, and Y. Huang, "Design of deep convolutional networks for prediction of image rapid serial visual presentation events," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, May 2017, pp. 2035–2038.

J. X. CHEN was born in Shihezi, Xinjiang, China, in 1979. She received the B.S. and M.S. degrees from the Department of Electrical and Information Engineering, Shaanxi University of Science and Technology, China, in 2002 and 2005, respectively. She is currently pursuing the Ph.D. degree in computer science and technology with Northwestern Polytechnical University, Xi'an, China.

From 2006 to 2013, she was a Lecturer with the Shaanxi University of Science and Technology. From 2016 to 2017, she was a Visiting Scholar with The University of Texas at San Antonio, TX, USA, where she researched on deep learning and electroencephalogram (EEG)-based event detection. Since 2013, she has been an Associate Professor with the Department of Electrical and Information Engineering, Shaanxi University of Science and Technology. She has authored one book and more than 20 articles. She holds two inventions. Her research interests include machine learning and pattern recognition, deep learning, EEG signal processing, and emotion recognition. Since 2016, she has been a member of the Chinese Computer Society.

P. W. ZHANG was born in Qianxian, Shaanxi, China, in 1977. He received the B.S. and M.S. degrees from the Department of Electrical and Information Engineering, Shaanxi University of Science and Technology, China, in 1996 and 2005, respectively.

From 2006 to 2013, he was a Lecturer with the Shaanxi University of Science and Technology, where he has been an Associate Professor with the Department of Electrical and Information Engineering, since 2013. He has authored one book and more than ten articles. He holds two inventions. His research interests include machine learning, pattern recognition, electroencephalogram (EEG) signal processing, and emotion recognition.



Z. J. MAO was born in Shangrao, Jiangxi, China, in 1991. He received the B.S. degree in applied mathematics from Jilin University, China, in 2011, and the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, The University of Texas at San Antonio, TX, USA, in 2013 and 2016, respectively. Since 2017, he has been a Postdoctoral Researcher with the BCI Laboratory, The University of Texas at San Antonio. His research interests include deep learning, brain–computer interface applications, biocomputing, electroencephalogram (EEG) signal processing, and event detection.

Y. F. HUANG was born in Xi'an, China, in 1973. He received the B.S. degree from Northwestern Polytechnical University, China, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from the State University of New York at Stony Brook, USA, in 1998 and 2001, respectively.

Since 2002, he has been with the Department of Electrical and Computer Engineering, The University of Texas at San Antonio (UTSA), where he is currently a Professor. He is also an Adjunct Professor with the Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio. He is also a Visiting Professor with the Center of Bioinformatics, Harvard Center for Neurodegeneration and Repair. His research interests include the areas of computational biology, brain–computer interface, statistical modeling, and Bayesian methods. He currently focuses on developing passive electroencephalogram (EEG)-based brain–machine interaction and deep learning algorithms for EEG data analysis.

Dr. Huang was a recipient of the U.S. National Science Foundation (NSF) Early Career Award, in 2005, the Best Paper Award of 2006 Artificial Neural Networks in Engineering Conference, and the 2007 Best Paper Award of *IEEE Signal Processing Magazine*. His research has been supported by NSF, National Institute of Health, Air Force Office of Scientific Research, Army Research Lab, Department of Defense, and Qatar National Research Fund. He is an Organizer of workshops and several special sessions including the IEEE Workshop on Genomic Signal Processing and Statistics, the Workshop on Systems Biology and Medicine, and the IEEE Bioinformatics and Biomedicine Conference. He is an Associate Editor of the IEEE Transactions on Signal Processing, *BMC Systems Biology*, the *EURASIP Journal on Bioinformatics and Computational Biology*, and the *International Journal of Machine Learning and Cybernetics*.

D. M. JIANG was born in Zhengzhou, Henan, China, in 1973. She received the B.S., M.S., and Ph.D. degrees in computer Science and technology from Northwestern Polytechnical University, Xi'an, China, in 1996, 1998, and 2000, respectively.

From 2000 to 2003, she was a Postdoctoral Researcher of mobile station of control science and technology with Northwestern Polytechnical University (NPU). In 2010, she was promoted to be a Professor with the School of Computer Science, NPU. From 2001 to 2002 and from 2006 to 2007, she was involved in research on the International Cooperation Project between the Chinese Ministry of Science and Technology and the Belgian Government, Vrije Universiteit Brussel. Since 2003, she has been with the School of Computer Science, NPU. Since 2005, she has been serving as the Chinese Contact Person for the Joint Laboratory of Audiovisual Signal Processing of Northwestern Polytechnical University and Brussels Free University. Since 2011, she has been a Supervisor of Ph.D. student with the School of Computer Science, NPU. She is the Host of three national projects and three key provincial projects of Shaanxi and more than 50 articles. Her research interests include speech processing of audio–visual fusion, audio–visual emotion analysis and recognition, facial animation synthesis of expressive speakers, and emotion recognition of physiological signals.

Dr. Jiang was a recipient of the Second Prize for outstanding academic papers of Natural Science of Shaanxi and the First Prize for science and technology in Shaanxi University. Since 2011, she has been serving as the Contact Person for NPU–Europe Liama Research Union. She is the Director of the Shaanxi Image and Graphics Society and the Shaanxi Signal Processing Society.

Y. N. ZHANG was born in Liquan, Shaanxi, China, in 1967. She received the B.S. degree in electronic engineering from the Dalian University of Technology, Dalian, China, in 1988, the M.S. and Ph.D. degrees in electronic engineering from Northwestern Polytechnical University, Xi'an, China, in 1992 and 1996, respectively, the Ph.D. degree from the Key Laboratory of Radar Defense Science and Technology, Xi'an University of Electronic Science and Technology, Xi'an, in 1999, and the Ph.D. degree from the School of Computer Engineering, Northwestern Polytechnical University, in 2002.

From 2000 to 2010, she was the Director of the Shaanxi Key Laboratory of Speech and Image Information Processing, Northwestern Polytechnical University, where she was promoted to a Professor with the School of Computer Science, in 2003. From 2012 to 2018, she was the Dean of the School of Computer Science, Northwestern Polytechnic University. Since 2005, she has been a Supervisor of Ph.D. student with the School of Computer Science, NPU, where she has been an Assistant President, since 2015. She has authored three books and more than 100 papers which have been published in authoritative journals and important international conferences at home and abroad, such as IEEE TPAMI, IEEE TIP, PR, IEEE TSMC-B, *Information Fusion*, CVPR, and ICCV. She holds more than 50 patents for invention. Her research interests include image processing, pattern recognition, computer vision, and remote sensing. She has undertaken more than 40 national-level projects such as 973 national defense projects, key projects of the National Natural Science Foundation of China, 863 national defense projects, and preliminary research of assembly projects.

Dr. Zhang was a recipient of one First Prize for scientific and technological progress in Shaanxi and one First Prize for National Defense Technological Invention. She was rewarded as one of the Advanced Individuals of 863 Science and Technology Tackling of General Assembly, in 2011. She was selected as the Young and Middle-Aged Science and Technology Innovation Leading Talent in Talents Promotion Plan of the Ministry of Science and Technology of the CPC, in 2013. She was selected as one of the first batch of scientific and technological innovation leaders in the ''ten thousand people plan'' of the Organization Department of the Central Committee of the CPC, in 2014. She serves as the Standing Director for the Chinese Stereology Society, the Deputy Director and the Secretary-General for the Image Analysis Branch, the Director for the Chinese Image Graphics Society and the Chinese Artificial Intelligence Society, a member for the Signal Processing Branch of the Chinese Electronics Society, and a Standing Vice-Chairman and the Secretary-General for the Shaanxi Signal Processing Society.