frontiers in Systems Neuroscience

EEG-based Emotion Classification Using Deep Neural Network and Sparse Autoencoder

Yuling Luo[1], Guopei Wu[1], Senhui Qiu[1*], Yuling Luo[1], Su Yang[2], Wei Li[3], Yifei Bi[4]

[1]Guangxi Normal University, China, [2]Xi'an Jiaotong-Liverpool University, China, [3]University of York, United Kingdom, [4]Leiden University, Netherlands

Submitted to Journal:
Frontiers in Systems Neuroscience

ISSN:
1662-5137

Article type:
Original Research Article

Received on:
29 Mar 2020

Accepted on:
12 Jun 2020

Provisional PDF published on:
12 Jun 2020

Frontiers website link:
www.frontiersin.org

Citation:
Luo Y, Wu G, Qiu S, Luo Y, Yang S, Li W and Bi Y(2020) EEG-based Emotion Classification Using Deep Neural Network and Sparse Autoencoder. *Front. Syst. Neurosci.* 14:43. doi:10.3389/fnsys.2020.00043

frontiers

EEG-based Emotion Classification Using Deep Neural Network and Sparse Autoencoder

Junxiu Liu [1,2], **Guopei Wu** [1,2], **Yuling Luo** [1,2], **Senhui Qiu** [1,2*], **Su Yang** [3], **Wei Li** [4,5], **and Yifei Bi** [6,7]

[1] *School of Electronic Engineering, Guangxi Normal University, Guilin, China*
[2] *Guangxi Key Lab of Multi-source Information Mining & Security, Guangxi Normal University, Guilin, China*
[3] *Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China*
[4] *Academy for Engineering & Technology, Fudan University, Shanghai, China*
[5] *Department of Electronic Engineering, The University of York, UK*
[6] *College of Foreign Languages, University of Shanghai for Science and Technology, Shanghai, China*
[7] *Department of Psychology, The University of York, UK*

Correspondence*:
Senhui Qiu
qiusenhui@mailbox.gxnu.edu.cn

ABSTRACT

Emotion classification based on brain-computer interface (BCI) systems is an appealing research topic. Recently, deep learning has been employed for the emotion classifications of BCI systems and compared to traditional classification methods improved results have been obtained. In this paper, a novel deep neural network is proposed for the emotion classification using EEG systems, which combines the Convolutional Neural Network (CNN), Sparse Autoencoder (SAE) and Deep Neural Network (DNN) together. In the proposed network, the features extracted by CNN are sent to SAE for encoding and decoding at first. Then the data with reduced redundancy are used as the input features of a DNN for classification task. The public datasets of DEAP and SEED are used for testing. Experimental results show that the proposed network is more effective than conventional CNN methods on the emotion recognitions. For DEAP dataset, the highest recognition accuracies of 89.49% and 92.86% are achieved for valence and arousal, respectively. While for SEED dataset, the best recognition accuracy reaches 96.77%. By combining CNN, SAE, DNN and training them separately, the proposed network is shown as an efficient method with a faster convergence than conventional CNN.

Keywords: EEG, Emotion Recognition, Convolutional Neural Network, Sparse Autoencoder, Deep Neural Network

1 INTRODUCTION

Brain Computer Interface (BCI) directly connects human (or animal) brain activity with artificial effectors (Kübler et al., 2009), which provides an interactive pathway between human brain and external devices

for various applications. The process of such interaction starts by recording the brain activities, through the signal processing and analysis to detect the users' intent (Tabar and Halici, 2016). Studies on BCI systems and its various implementations have been lasting for decades, and one of the most appealing research directions is the emotion recognition, due to its potential applications in numerous scenarios. Both non-physiological and physiological signals could be employed for emotion detections. Non-physiological signals include facial expression images (Lane et al., 1997), voice signal (Scherer, 1995) and body gesture (Cheng and Liu, 2008). Compared to the non-physiological signals, physiological signals can be detected by some wearable devices, such as electroencephalogram (EEG) (Zheng, 2017), electromyogram (Hiraiwa et al., 1989), electrocardiogram (Agrafioti et al., 2012), galvanic skin response, blood volume pressure and photoplethysmogram. Among these physiological signals, EEG signal has been widely used for the research on emotion recognitions (Chi et al., 2012; Liu et al., 2018c; Huang et al., 2015; Li et al., 2016). Captured from scalp by a number of EEG electrodes, emotion could be reflected immediately by EEG signal once a subject receives the stimulations.

There are two conventional rules to categorize human's emotions, namely the discrete basic emotion description and the dimension approaches. According to the discrete basic emotion description approach, emotions can be classified into six basic emotions including sadness, joy, surprise, anger, disgust and fear (van den Broek, 2013). For the dimension approach, the emotions can be classified into two dimensions (valence and arousal) or three dimensions (valence, arousal and dominance) (Zheng and Lu, 2015). Among these dimensions, valence describes the level of positive or negative of one person, arousal describes the level of excited or apathetic of emotion. The scale of dominance ranges from submissive (or without control) to dominance (or empowered). The emotion recognition is usually based on the dimension approach because of its simplicity, compared to the discrete basic emotion description (Zheng and Lu, 2015).

Early works on emotion recognition through analysing EEG signal could be traced back to more than fifty years ago (Fink, 1969). Many new methods on feature extraction and classification have been proposed for emotion detection recently (Petrantonakis and Hadjileontiadis, 2010). For the feature extraction, two types of feature are commonly used to analyse EEG signal: time-domain and frequency-domain features. Time-domain features capture the temporal information of signals such as fractal dimension feature (Hjorth, 1970), Hjorth feature and higher order crossing feature (Petrantonakis and Hadjileontiadis, 2010). The frequency-domain features can extract the useful information from the frequency perspective under different frequency bands. For instance, the EEG signal could be decomposed into δ (1-3 Hz), θ (4-7 Hz), α (8-13 Hz), β (14-30 Hz) and γ bands (31-50 Hz) (Petrantonakis and Hadjileontiadis, 2010; Hjorth, 1970; Li and Lu, 2009; Nie et al., 2011), where the features can be extracted from each of them. In addition, other features such as Deep Forest (Zhou and Feng, 2017), Statistical Characteristics (SC), Differential Entropy (DE) feature (Zheng et al., 2014), Pearson Correlation Coefficient (PCC) feature (Lewis et al., 2007) and Principal Component Analysis (PCA) feature (Subasi and Gursoy, 2010) are also used in emotion recognitions.

In the meantime, various classification methods have been used for emotion recognition, such as k-Nearest Neighbour (Bahari and Janghorbani, 2013), Multi-Layer Perceptron (Orhan et al., 2011). Support Vector Machine (SVM) and Linear Regression (LR) were used in (Wang et al., 2019), but recognition accuracy can be improved. In recent years, deep neural networks (DNN) (Tripathi et al., 2017) has been developed into one of the most effective and popular methods in many research fields (Liu et al., 2018a; Luo et al., 2018; Liu et al., 2019; Fu et al., 2017; Liu et al., 2018b). Convolutional Neural Networks (CNN) are widely used in computer vision, image classifications, visual tracking (Danelljan et al., 2016), segmentation and

object detections (Girshick et al., 2014). EEG emotion classification using CNN method is also explored in the approaches of (Tripathi et al., 2017). Cascade and parallel convolutional recurrent neural network is used for EEG human intended movements classification tasks (Zhang et al., 2018). Additionally, before applying the CNN, EEG data could be converted to image representation after feature extraction (Tabar and Halici, 2016). However, the accuracy of emotion recognition by using only CNN is not high. In the work of (Zhang et al., 2017), a deep learning framework consisting of the sparse autoencoder (SAE) and logistic regression was used to classify EEG emotion status. The sparse autoencoder was employed for feature extraction, and logistic regression was used to predict affective states. SAE is an unsupervised machine learning algorithm. By calculating the error between the output of SAE and original input, data could be reconstructed and useful features could be extracted for classification task. However, accuracy of that work is not high and there are no experiments for comparing to verify the work of SAE.

In this work, a novel network model combining CNN, SAE and DNN together is proposed to convert EEG time series into 2D images for a good emotion classification performance. The EEG signal is decomposed into several different bands. Based on frequency, time and location information, the 2D features are extracted from EEG data. Then convolutional layers of CNN are trained and used for further extracting features. SAE is used for reconstructing data obtained from convolutional layers, and in the step DNN is used for classification. Compared to other approaches, the proposed neural network model, which leverages the benefits of convolutional layers of CNN and sparsity of SAE, demonstrates a good classification accuracy and fast convergence. The procedure of the proposed method is summarized in Figure 1. Original EEG data is pre-processed, and features are extracted for deep learning model. After training and testing on the model, final classification results are obtained.

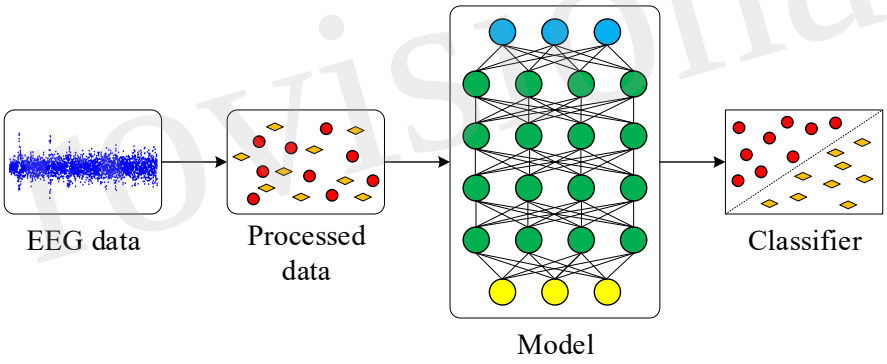

Figure 1. Emotion classification procedure in this work.

The rest of this paper is organized as follows: The proposed neural network model is presented in Section 2. Datasets and experimental results are provided in Section 3. Section 4 summaries the work and discusses the future work.

2 DEEP LEARNING FRAMEWORK

In this section, fundamental principles and essential network modules are presented. The novel model is also introduced in detail.

2.1 Convolution neural network (CNN)

The features extracted from original EEG data are sent to CNN first. CNN model includes several convolution-pooling layer pairs and one output layer. Before sending to CNN, features are concatenated

into image form which is then convolved with several one-dimensional filters in convolution layers. After the pooling layer, the data is further subsampled to images with smaller size. Network weights and filters in the convolution layers are learned through back-propagation algorithm.

In our experiments, data extracted from EEG signal are from four typical frequency bands, which include α (1-7 Hz), β (8-13 Hz), θ (14-30 Hz), and γ bands (30-45 Hz), using a Butterworth band-pass filter. After that, data is reformed into two-dimensional features such as PCC, which are the input for CNN. Detailed methods of this procedure is presented in Section 3.3 and Section 3.4. It is worth noting that the two-dimensional features contain not only frequency information, but also spatial location information of each electrode (Tabar and Halici, 2016). To preserve this information, one-dimensional filtering is applied in this work instead of two-dimensional filtering.

The CNN structure is relatively straightforward. Input vector is two-dimensional feature, which can be given by

$$x = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}, \tag{1}$$

where $m \times n$ is the shape of input vector x. The input two-dimensional feature is convolved with filters W_k at the convolution layer, which is given by

$$W_k = \begin{pmatrix} W_{11} \\ W_{21} \\ \dots \\ W_{i1} \end{pmatrix}, \tag{2}$$

where i is length of W_k and $i<m$ in equation 1. After the image convolution, output map is formed and the feature map at the given layer is obtained by

$$f(\alpha) = f(W_k \times x + b_k), \tag{3}$$

where $W_k \in R^{i \times 1}$ is the weight matrix and b_k is the bias value, k denotes the filter, for $k = 1, 2, \dots, n$ and n denotes the total number of filtering in convolutional layer. The activation function is f, which is a rectified linear unit ($ReLU$) function in this work. Compared with the traditional neural network activation functions such as $sigmoid$ and $tanh$, $ReLU$ is more efficient in avoiding gradient disappearance. $ReLU$ function is defined by

$$f(\alpha) = ReLU(\alpha) = \ln(1 + e^{\alpha}), \tag{4}$$

where α is defined in equation 3. At the max-pooling layer, the feature map is down sampled through the max-pooling function. Max-pooling is used because it is found that the maximum value from the selected values of a given feature map could be effectively extracted using this function.

After the last pooling layer, a fully connected layer is followed, in which output data from pooling layer is flattened. After that, fully connected layers named DNN are followed. In DNN, activation function of each layer is also $ReLU$. For the output layer, because there are two classification tasks including binary classification and multi-class classification, $sigmoid$ and $softmax$ are used, respectively. For binary-classification task, $Adadelta$ is used as optimizer, loss is calculated by binary crossentropy, which

is given by

$$loss = -\sum_{n=1}^{N} \hat{y}_i logy_i + (1-\hat{y}_i)log(1-\hat{y}_i), \tag{5}$$

where N is number of samples, y_i is label value which is the form of one-hot code, $\hat{y}_i$ is output from output layer where $sigmoid$ is used. For multi-class classification such as three-class classification, $Adam$ is used as optimizer, loss is calculated by categorical crossentropy, which is given by

$$loss = -\sum_{n=1}^{N} \hat{y}_{i1} logy_{i1} + \hat{y}_{i2} logy_{i2} + \hat{y}_{i3} logy_{i3}, \tag{6}$$

where N is number of samples, y_{i1}, y_{i2}, y_{i3} are values of label which is also the form of one-hot code, $\hat{y}_{i1}$, $\hat{y}_{i2}$, $\hat{y}_{i3}$ are three outputs from output layer where $softmax$ is used. Parameters in the model are updated by using back-propagation algorithm. The error between the desired output and the actual output is computed and the gradient descent method is applied to update parameters in order to minimize the error. Functions to update the weight and bias are shown by

$$W_k = W_k - \eta \frac{\partial E}{\partial W_k}, \tag{7}$$

$$b_k = b_k - \eta \frac{\partial E}{\partial b_k}, \tag{8}$$

where W_k is the weight matrix, b_k is the bias and η represents the learning rate, E is the error. E is equal to $loss$ in equation 5 and equation 6. The results obtained from this CNN will be used as a benchmark for the performance comparison in Section 3.

2.2 Sparse Autoencoder (SAE)

An autoencoder is a network including one input layer, one hidden layer and one output layer, which is used to preserve the essence of the input data as much as possible and remove the potential noise in an unsupervised manner. Therefore, the output data is simplified, and important information of input data is retained, which is beneficial for classification.

The structure of autoencoder is shown in Figure 2. The whole data processing is divided into encoding and decoding phases. In the encoding phase, the dimension of input data is reduced in one layer. When the decoded data arrives at the hidden layer, the dimension of input data reaches the same as the number of neurons predefined for this layer. The encoding function of hidden layer, h, is defined by

$$h = encoder(x) = f(W_k \times x + b_k), \tag{9}$$

where $W_k \in R^{m \times n}$ is the weight matrix between input layer and the next layer. As defined previously in CNN, b_k is also the bias vector and f represents the output function. Output function used in this part is $ReLU$, which is similar to the activation of CNN. Different from encoding phase, in decoding phase, the same number of neurons in output layers should be set as that of layers in encoding phase, in order to guarantee the output data has the same dimension as the input data. The decoding function is shown by

$$y = decoder(x) = g(W_k \times x + b_k), \tag{10}$$

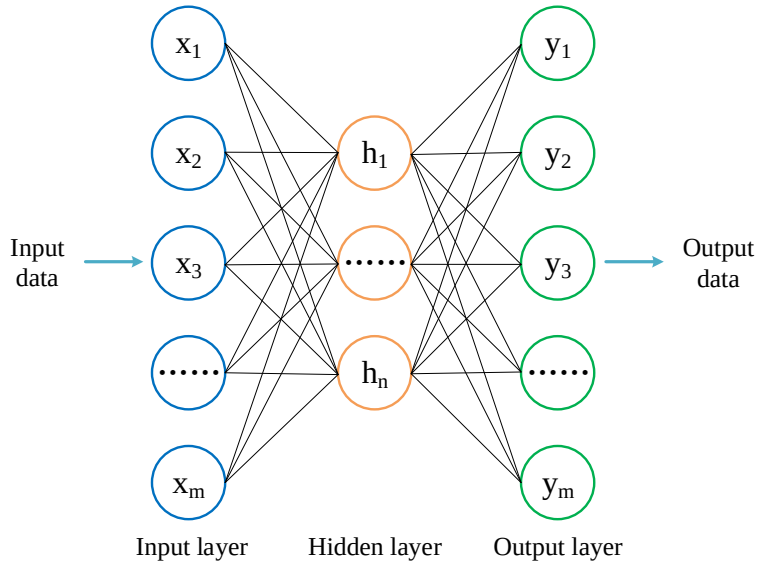

Figure 2. Autoencoder includes one input layer, one hidden layer and one output layer.

where $W_k \in R^{n \times m}$. After encoding and decoding phases, the model is trained and the parameters could be obtained by minimising the cost function which is defined by

$$min \sum \left| E(x^i, y^i) \right|, \tag{11}$$

where y^i is output data and x^i is original input data. When the network is trained, output values are reconstructed and the shape of which is equal to that of input data. Parameters of the model could be updated according to

$$W_k = W_k - \eta \frac{\partial E(x^i, y^i)}{\partial W_k}, \tag{12}$$

$$b_k = b_k - \eta \frac{\partial E(x^i, y^i)}{\partial b_k}, \tag{13}$$

where η denotes the learning rate of the network. E is error in SAE. For details of optimizer and E, they are the same as that in Section 2.1 in binary-classification task.

In order to increase the generalization of the network and improve the training efficiency of the proposed network, a sparse constraint on the activity of the hidden representations is added in this work. Sparse constraint helps suppress activation of neurons in hidden layer and useful features can be extracted by autoencoder. Thus, the cost function in sparse autoencoder is described by

$$J_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{m} KL(\rho || \hat{\rho_j}), \tag{14}$$

where $\hat{\rho_j}$ is the average activation of hidden unit j, ρ is the sparsity level, β is the weight of the sparsity penalty term. KL is Kullback–Leibler divergence, which ensures the sparsity of neurons in hidden layer. KL is defined by

$$KL(\rho||\hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho)\frac{1-\rho}{1-\hat{\rho}_j}, \tag{15}$$

$$\hat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m} f_j(x^i), \tag{16}$$

where m denotes the number of samples at unite j in hidden layer, f_j denotes the activation of hidden neuron j.

2.3 Combined CNN-SAE-DNN

EEG signal is quite sensitive to a variety of factors during acquisition, such as environmental interference and human's own emotional fluctuations. Therefore, EEG signals may be mixed with a variety of noise, which would undoubtedly influence the required brain patterns and the experimental results. In addition, in some experiments, subjects were unable to perform the emotion collection task successfully and the experimental results were deviated greatly. In order to overcome these problems, a deep learning network structure is proposed in this work. The structure of the proposed network is shown in Figure 3.

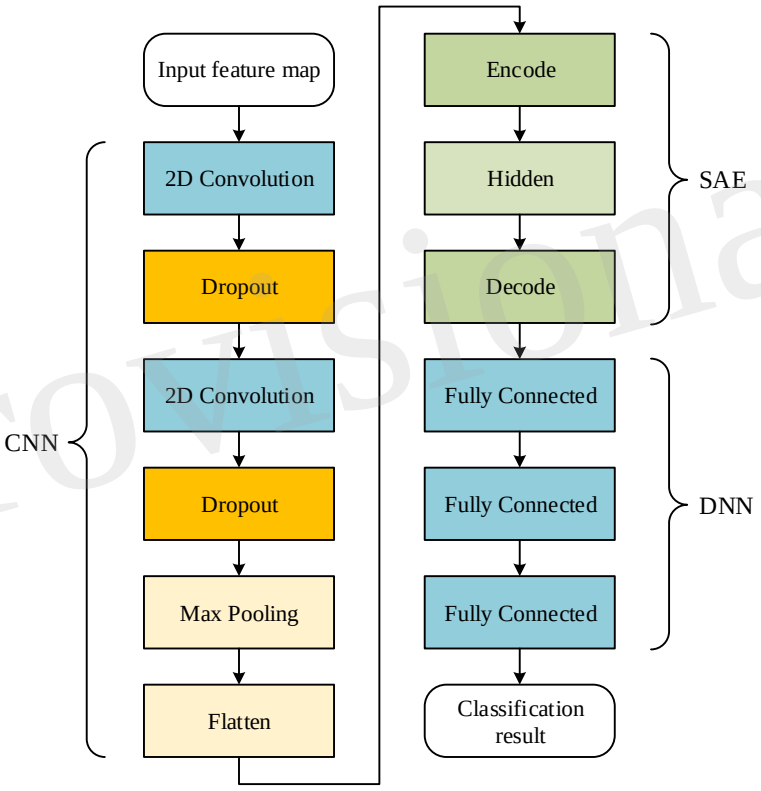

Figure 3. The proposed network includes CNN, SAE and DNN, in which CNN and SAE are used for feature extraction and DNN is used for classification.

As shown by Figure 3, in the proposed network, CNN structure consists of two convolutional layers and one max pooling layer. Dropout connects to each convolutional layer. SAE consists of one encode layer, one hidden layer and one decode layer. In DNN, there are three fully connected layers used for classification. Given features such as PCC for input of the proposed network, the output of max-pooling layer is used as the input of SAE. Finally, the output of SAE is used as the input of DNN for classification.

The training procedure is that CNN with one fully-connected output layer are trained for some epochs using all samples using all features, and the output layer is abandoned after training. Then, by sending features to input of the trained CNN, output of max pooling layer can be obtained. The output is flattened to one-dimension data and it is set as the input of SAE. After unsupervised learning of SAE, data is reconstructed. The reconstructed data is divided for training and testing in DNN, i.e. CNN and SAE are trained separately. Thus, before data is classified in DNN, training in CNN and SAE can be seen as a part of feature extraction. It should be noticed that DNN used for finally classification is not the fully-connected output layer abandoned from CNN in the first step. DNN is never trained before output of SAE is obtained as input data for DNN.

Another CNN with the same parameters and structure as the whole proposed network is set as comparison in order to test the performance of the proposed network fairly. If adding more layers in this CNN, If adding more layers in this CNN, accuracy doesn't improve and leads to overfitting problem. For experiments on this CNN, features are splited directly 80% for training and the rest for testing.

3 DATASETS AND EXPERIMENTS

In this section, two datasets of DEAP (Koelstra et al., 2012) and SEED (Zheng and Lu, 2015) are used to evaluate the proposed network model. Data processing methods and experiment results are presented.

3.1 Emotional EEG Datasets

The DEAP dataset was collected from 32 subjects when they were watching 40 one-minute music and video clips. The age of the subjects ranges between 19 and 37 years old, and half of them are males. During the 40 trials for each subject, various signals were recorded as 40-channel data, which include EEG, electromyogram, breathing zone, plethysmograph, temperature and so on (Koelstra et al., 2012). EEG signal was recorded at 512 Hz. The data was segmented into trials of 60 seconds and a bandpass frequency filter was applied after that. After each trial, the participants were asked to do a self-assessment about their emotional levels including four different scales such as valence, arousal, dominance and liking.

The EEG signal is down-sampled into 128 Hz for the experiments in this work, where the frequency of EEG data is from 4.0 Hz to 45.0 Hz. Valence and arousal are the two scales chosen for this work. Each of them is ranging from one (low) to nine (high) and scales are divided into two parts to construct our binary-classification tasks. Similar to the work in (Koelstra et al., 2012), valence is divided into high (ranging from five to nine) and low valence (range from one to five) according to the valence scale, and according to the arousal scale, arousal is divided into high (ranging from five to nine) and low arousal (ranging from one to five).

SEED dataset was collected from 15 subjects (7 males) when they were asked to watch 15 film clips. The duration of each film clip is about 4 minutes and each film is easily understood in order to elicit emotion of 15 subjects participating in the experiments effectively. There are 15 trials for each subject and each trial lasts for 305 seconds consisting of a hint of start for 5 seconds, a movie clip for 4 minutes, a self-assessment for 45 seconds and a rest for 15 seconds. EEG data in SEED dataset was collected from 62 electrodes, which includes more information than the DEAP dataset. After collection, EEG data was downsampled to 200 Hz and applied with a bandpass filter from 0-75 Hz.

Similar to the DEAP dataset, in this dataset, the data is applied with a frequency filter from 4.0-45.0 Hz in order to equitably evaluate the proposed network. Negative, positive and neutral are emotion labels in this dataset, which represent the subjects' emotion states during each experiment. Label value of negative, positive and neutral is -1, 1, 0, respectively. Thus, labels in SEED dataset include three categories.

3.2 Experiment Setting

In order to test the efficacy of the proposed network, CNN model and the proposed network are trained by using data obtained from two time windows with different length, thus, in total four groups of experiments are conducted. For experiments in CNN used for comparison, after feature extraction of EEG data, 80% samples are used as training data and the rest samples are used as test data among all of the data. Average accuracy is calculted from accuracies of the last ten epochs in each experiment. For the proposed network, before training data and testing data are divided, CNN and SAE in the proposed network are trained using features. After that, features are sent to the input of CNN and the output data of SAE is obtained. The output data after feature extraction is divided 80% for training and 20% for testing in DNN. In this work, Keras and Tensorflow (Abadi et al., 2016) are used for the proposed network implementation. For detailed free parameters in the proposed network, they are described in Section 3.3 and Section 3.4, respectively.

3.3 Experiments on DEAP dataset

Length of data in DEAP dataset is 63 seconds and the first 3 seconds is removed in experiments. Then band pass filtering is applied. Among 40 channels, EEG data is contained in 32 channels, which is chosen for experiments. After that, EEG signals are decomposed into α (1-7 Hz), β (8-13 Hz), θ (14-30 Hz), and γ bands (30-45 Hz). After band pass filtering, signal windowing on four frequency bands is applied. EEG signals are divided into short time frames in order to facilitate signal processing, thus time windows with different overlaps are applied to EEG data in order to increase samples for training. Two window sizes, 8 seconds and 12 seconds, are used for evaluating the proposed network. From the start of each recorded EEG signal, data is segmented by a sliding time window with an overlap for each frequency band. For each trial of 60 seconds, 14 segments are obtained using an 8-second time window moving every 4 seconds, 7 segments are obtained using a 12-second time window moving every 8 seconds. Finally, in total of 32 participants, 17,920 (14 segments$\times$40 trials$\times$32 participants) and 8,960 (7 segments$\times$40 trials$\times$32 participants) samples are obtained using time windows of 8 seconds and 12 seconds, respectively. Segment labels are the same as the label of the original sample.

After that, three different features, namely PCC, PCA and SC, are extracted to evaluate the proposed network. For PCC-based features, PCC of data in every two channels is calculated and a 32$\times$32 PCC matrix is constructed for one sample. For PCA-based features, dimension of data from each channel is reduced into 32 and features with the shape of 32$\times$32 are obtained. For SC-based features, four different characteristics are extracted including variance, mean, kurtosis and skewness. These statistical characteristics of data are calculated together and finally a 32$\times$4 matrix is obtained. In the proposed work, the features are separately extracted in each of the frequency bands (α, β, θ and γ bands). According to the work in (Wang et al., 2018) and other similar researches, data of four frequency bands are used together in order to get the best results. After data is processed, for the data obtained using time window of 8 seconds, the shapes of the above three different feature matrixes are 17,920$\times$4$\times$32$\times$32, 17,920$\times$4$\times$32$\times$32 and 17,920$\times$4$\times$32$\times$4, respectively. For data obtained using time window of 12 seconds, they are 8,960$\times$4$\times$32$\times$32, 8,960$\times$4$\times$32$\times$32 and 8,960$\times$4$\times$32$\times$4, respectively. Detailed configuration of the proposed network for DEAP dataset is shown by Figure 4. For SC, input shape is 32$\times$4. These features are two-dimensional, which are suitable inputs for CNN and the proposed network.

As shown in Figure 4, for DEAP dataset, two convolutional layers and one max-pooling layer are applied for the proposed network. Kernel size is set to 3$\times$1 and pooling size is set to 3$\times$3. Input data shape is 32$\times$32. The numbers of kernels in convolutional layer are set to 32 and 64, respectively. In SAE, the numbers of neurons in encode, hidden and decode layers are set to 512, 128 and 512, respectively. In DNN, the numbers of three fully connected layers are set to 512, 256 and 2, respectively. In the proposed network,

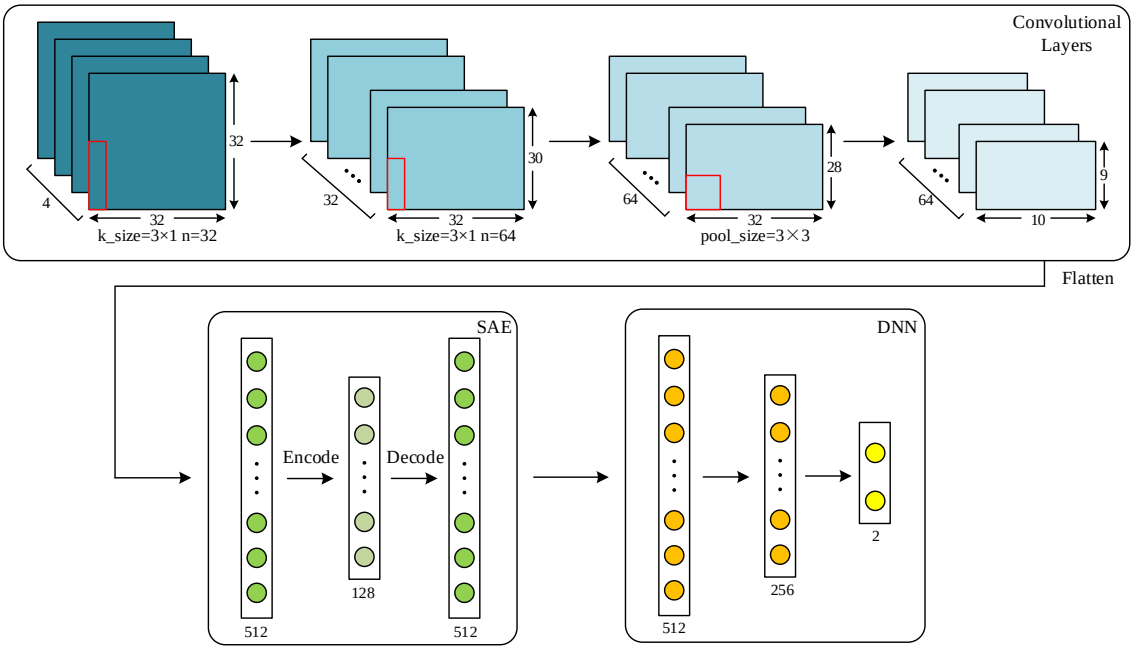

Figure 4. Configuration of the proposed network for the DEAP dataset.

the training epochs, batch size, learning rate in CNN are set to 50, 128 and 0.01. Epoch, batch size, learning rate in SAE are set to 100, 64 and 0.01, respectively. For those of DNN, they are set to 100, 128 and 0.01.

In the proposed network, the training epochs are carried out in convolutional layers and SAE for features extraction, training and testing epochs are carried out in DNN for classification. Another CNN with the same parameters and structure as the proposed network is served as a baseline method to evaluate the performance of the proposed network. Epoch, batchsize and learning rate in this CNN are set to 100, 128 and 0.01. Parameters in this CNN are the same as that of the proposed network . The experiments results of data obtained from time window of 8 seconds are shown by Table 1.

Table 1. Average accuracies comparisons on DEAP dataset using different features extracted from data with length of 8 seconds between two networks.

Network	Labels	PCC	PCA	SC
CNN	Valence	78.80%	73.32%	71.10%
	Arousal	82.25%	72.76%	73.04%
Proposed Network	Valence	89.49%	75.59%	81.93%
	Arousal	92.86%	85.87%	82.94%

From Table 1, among all features extracted from EEG data, the PCC feature is demonstrated to be better than most of the other features on both CNN and the proposed network. The proposed network can reach the recognition accuracy of 92.86% on arousal by using PCC. Moreover, recognition accuracies of most experiments on the proposed network are better than CNN (3.27%∼13.11% improvement). As described previously, this is due to the inclusion of SAE, which can not only reconstruct data from convolutional layers and pooling layer, but also extract features further and make the data easier to be recognized than CNN.

Training loss of SAE is shown by Figure 5, data reconstruction is achieved when the loss does not change sharply and data reconstruction is fast during training process in SAE. For other extracted features (except PCC), the recognition accuracy of each method is better compared to the work in (Zhang et al., 2016a) (81.21% for valence and 81.26% for arousal).

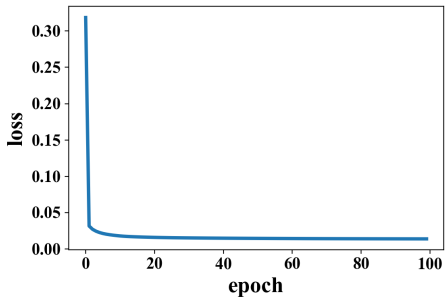

Figure 5. Change of loss when data is reconstructed in SAE on DEAP dataset.

Figure 6 and Figure 7 show the accuracies of CNN and the proposed network. Red lines in figures denote the average accuracy in the last 10 epochs. It can be seen that the accuracy of CNN is gradually converged. For the the proposed network, the accuracy is fast converging at the beginning of epoch after less than 10 epochs. This is because features are easy to be recognized using output data obtained from SAE, before being classified by DNN. For features extracted by the PCC and other methods, accuracy of proposed network has a faster convergence than CNN.

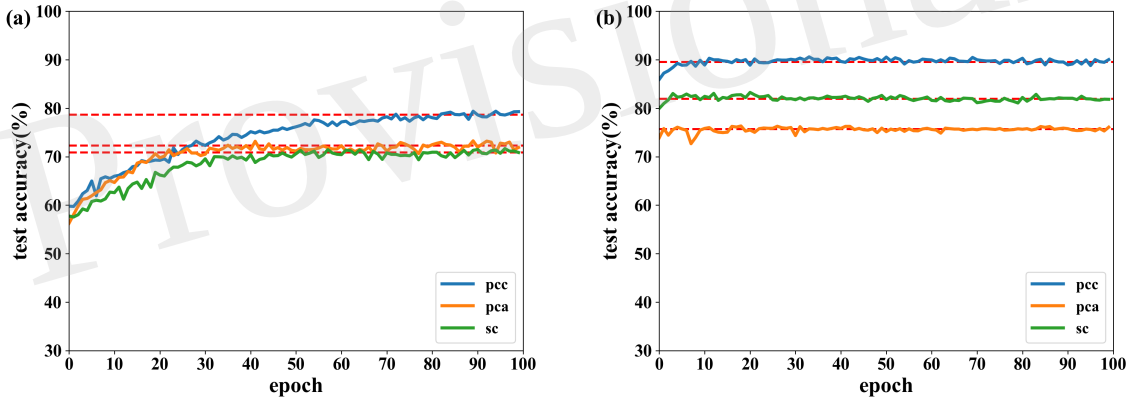

Figure 6. Accuracy comparison of two networks on valence using data with length of 8 seconds on DEAP dataset, in which (a) is result of CNN and (b) is result of the proposed network.

Similarly, results using data obtained from time window of 12 seconds are shown in Table 2. From Table 2, accuracy obtained using data with length of 12 seconds is lower than that of 8 seconds. It is more difficult to collect emotion information when the stimulation time is increasing. Most studies related with the classification of EEG data focused on short length of time. In this experiment, higher classification accuracy is achieved on data of 12 seconds than that of shorter length in other studies such as the work in (Zhang et al., 2017), which shows the good effectiveness of the proposed network.

Accuracies for classification on data of 12 seconds on both CNN and the proposed network are shown by Figure 8 and Figure 9. It can be found that higher recognition accuracy is obtained by the proposed

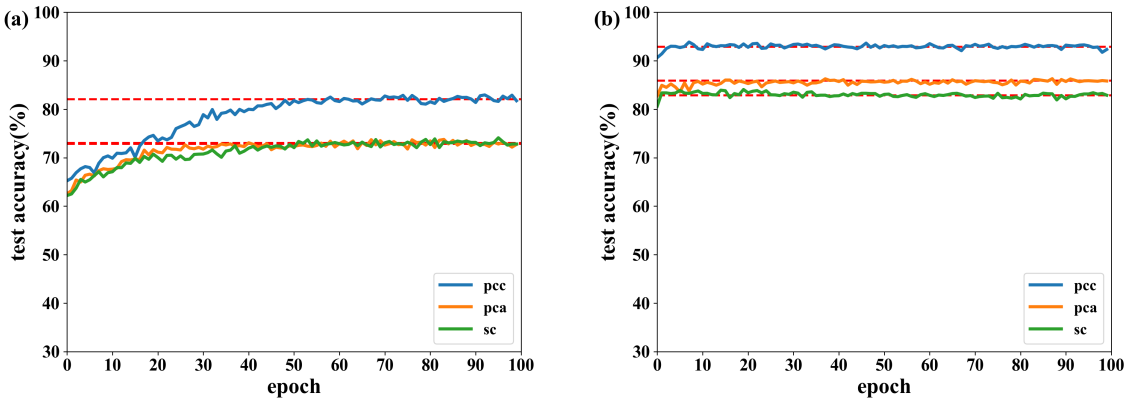

Figure 7. Accuracy comparison of two networks on arousal using data with length of 8 seconds on DEAP dataset, in which (a) is result of CNN and (b) is result of the proposed network.

Table 2. Average accuracy comparisons on DEAP dataset using different features extracted from data with length of 12 seconds between two networks.

Network	Labels	PCC	PCA	SC
CNN	Valence	75.13%	67.23%	66.09%
	Arousal	76.12%	69.20%	69.48%
Proposed Network	Valence	82.16%	76.34%	73.41%
	Arousal	85.47%	79.11%	75.44%

network. Moreover, the classification accuracy of the proposed network has a faster convergence in each experiment.

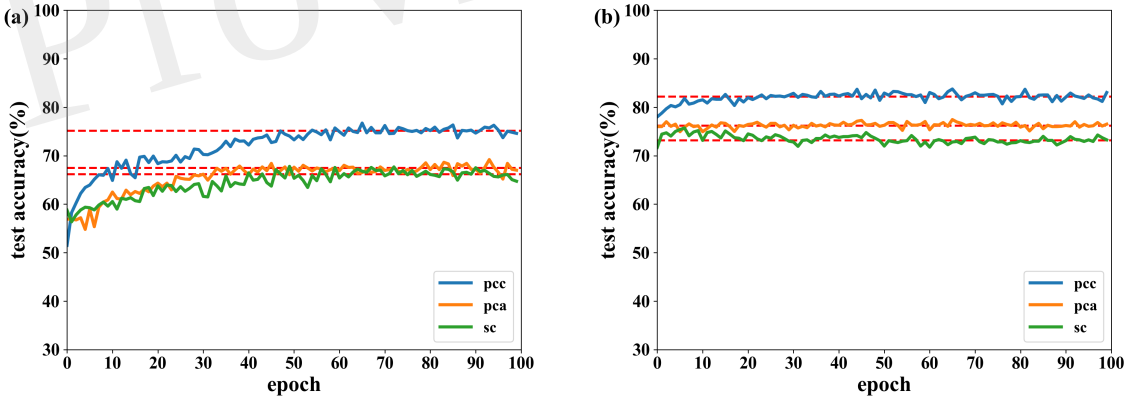

Figure 8. Accuracy comparison of two networks on valence using data with length of 12 seconds on DEAP dataset, in which (a) is result of CNN and (b) is result of the proposed network.

The results in this subsection demonstrate that accuracies can reach 92.86% on data of 8 seconds and 85.47% on data of 12 seconds. When the same feature is used for comparison, the proposed network is more powerful in classifying EEG emotion data than CNN. Finally, the proposed network has a quicker convergence speed.

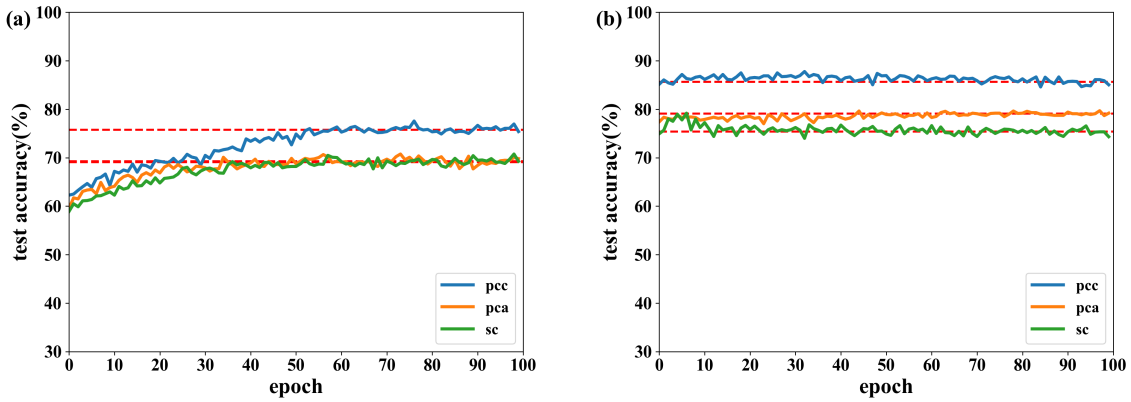

Figure 9. Accuracy comparison of two networks on arousal using data with length of 12 seconds on DEAP dataset, in which (a) is result of CNN and (b) is result of the proposed network.

3.4 Experiments on SEED dataset

There are totally 675 trials in SEED dataset. According to the work in (Zheng and Lu, 2015), the first sample of each subject is chosen and finally 225 samples are obtained. Because of different data length in each channel, the data segment of 80 seconds is chosen to reduce influence of unstable signal at beginning and end of the whole signal and finally data with the shape of 16,000×225×62 is obtained. Moreover, the data is processed as the same way in DEAP dataset, each sample is divided into different frames with different time windows. Two time windows, 8 and 12 seconds, are also used in SEED dataset. 19 and 9 segments are obtained separately from data using time window of 8 seconds moving every 4 seconds and time window of 12 seconds moving every 8 seconds for each sample. Thus, in total of 225 trials, 4,275 (19 segments×15 trials×15 participants) and 2,025 (9 segments×15 trials×15 participants) samples are obtained, respectively.

The detailed configuration of the proposed network for SEED dataset is shown by Figure 10. Amount of data extracted from this dataset is much less than DEAP dataset, thus the two classifiers used on this dataset are a little different. For PCC, input data is 62×62, and the numbers of kernels are separately set to 8 and 16 in two convolutional layers. In SAE and DNN, number of each layer is set the same as that on DEAP dataset, except that the number of output layer is set to 3 because this is a three-classification task on SEED dataset. After training in CNN and SAE for features extraction, DNN is used for final classification. Similar to the DEAP dataset, the same features are extracted for SEED dataset. For PCA and SC, input shape is 62×62 and 62×4, respectively. For CNN used for comparison, parameters are also set the same as the proposed network.

The experiment results of data obtained from time windows of 8 seconds and 12 seconds are shown by Table 3. Accuracy under this dataset is higher than DEAP dataset. The highest average accuracy could reach 96.77%, which is better than the work in (Wang et al., 2018), 90.2%. For the data obtained from time window of 12 seconds, the best accuracy could reach 94.62%, which shows that PCC-based features have better performance than others. When the data is reconstructed by the SAE, the change of loss on SEED is shown by Figure 11. Loss drops immediately after several epochs, i.e. the data reconstruction can be achieved fast when SAE is being trained.

Accuracies under different features extracted from data of 8 seconds and 12 seconds are shown by Figure 12 and Figure 13. It is shown that recognition accuracies of the proposed network are better than CNN for almost all features, especially the PCC-based features. The proposed network can achieve faster

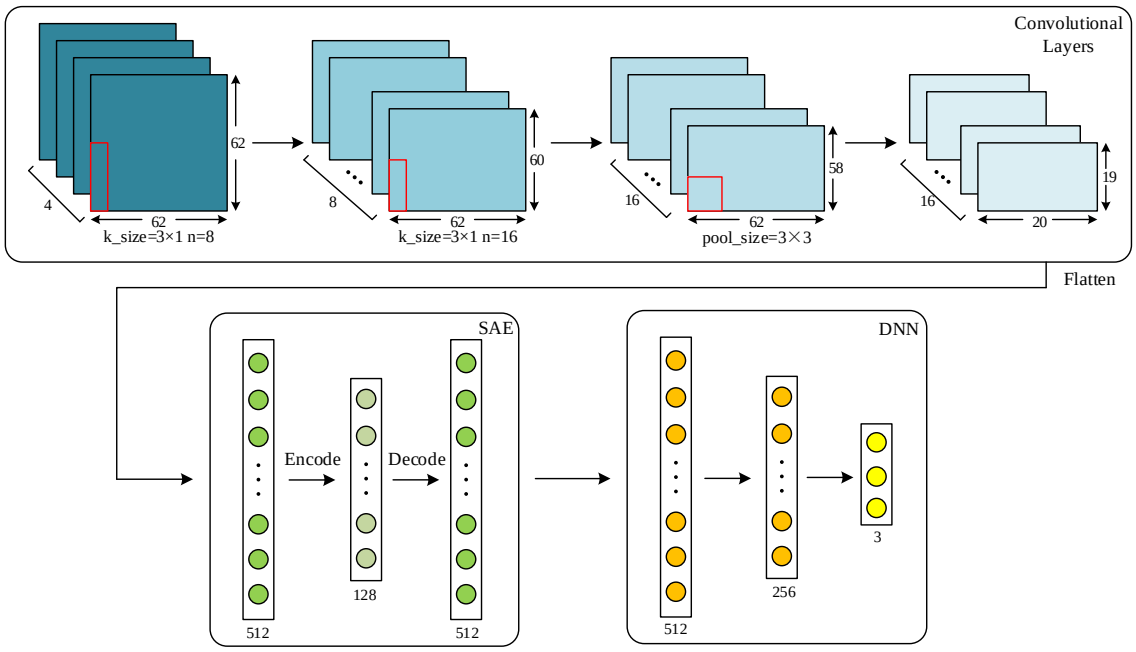

Figure 10. Configuration of the proposed network for the SEED dataset.

Table 3. Average accuracy comparisons on SEED dataset using different features extracted from data with length of 8 seconds between two networks.

Network	Data Length	PCC	PCA	SC
CNN	8 seconds	93.71%	77.66%	83.32%
	12 seconds	91.53%	70.59%	75.48%
Proposed Network	8 seconds	96.77%	88.90%	87.73%
	12 seconds	94.62%	70.62%	79.09%

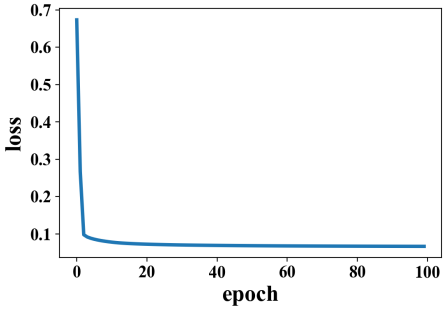

Figure 11. Change of loss when data is reconstructed by SAE on SEED dataset.

convergence on classification accuracy than CNN on SEED dataset. Experiments on these two datasets shows that the proposed network performs better than original CNN in emotion recognition.

Moreover, EEG data divided by a fixed time window with different overlaps on SEED dataset is tested. Besides time window of 8 seconds with overlap of 4 seconds, overlaps of 6 and 8 seconds are also tested. Due to the highest accuracy, PCC-based features are used in these experiments and classification results are shown by Figure 14.

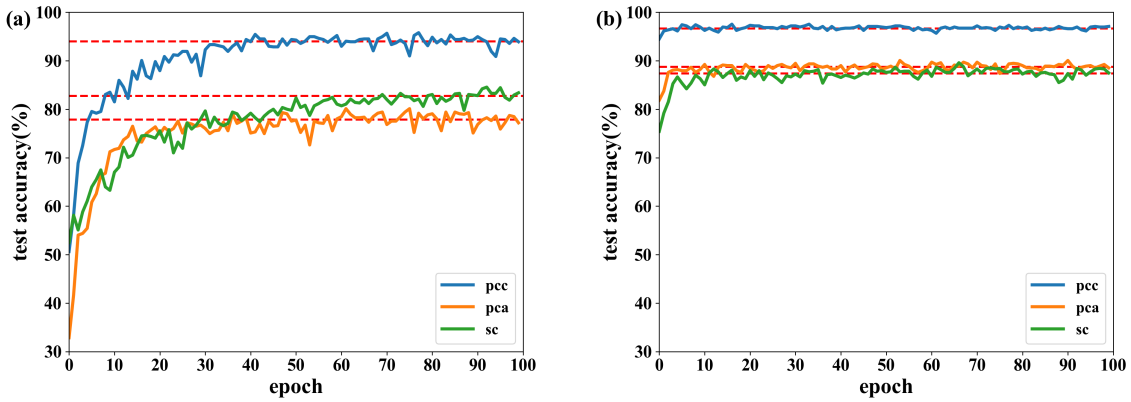

Figure 12. Accuracy comparison of two networks using data with length of 8 seconds on SEED dataset, in which (a) is result of CNN and (b) is result of the proposed network.

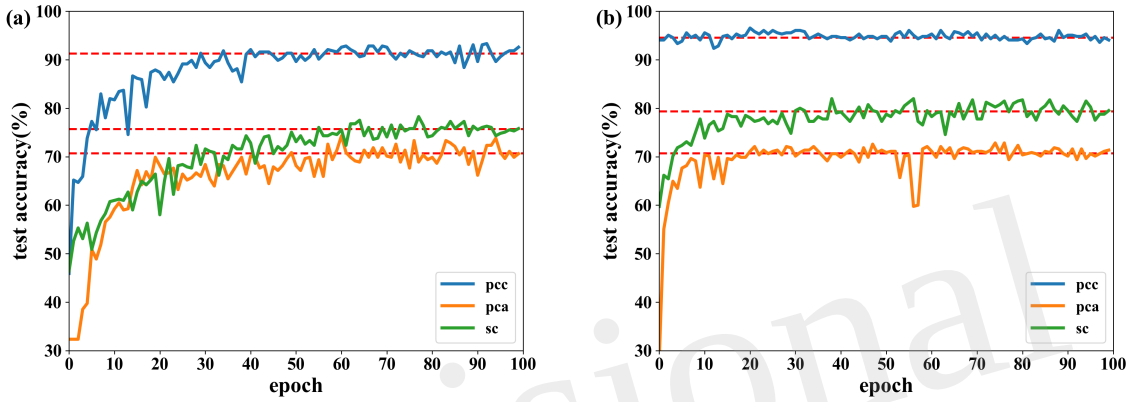

Figure 13. Accuracy comparison of two networks using data with length of 12 seconds on SEED dataset, in which (a) is result of CNN and (b) is result of the proposed network.

As shown by Figure 14, recognition accuracy could reach the highest value while the overlap is 4 seconds. The shorter the overlap is, the more similar neighbouring data segments, i.e. features could be learned better when similar information is included in each trial. However, when the overlap is too short, the number of data segments is larger which requires a longer time for training. In this experiment, data with length of 8 seconds and overlap of 4 seconds could reach the best result.

In a short summary, the best recognition could reach 96.77% on the three-class classification. The proposed network demonstrates to be more powerful in classifying EEG emotion data than CNN on SEED dataset. For the data with same length, length of overlap has an impact on recognition accuracy where 4-second overlap obtained the best performance.In addition, the proposed network is also compared with other research works using DEAP and SEED datasets and the results are shown by Table 4. For complexity analysis, the number of parameters is 7.55×10^5 and 7.50×10^5 for the networks used for DEAP and SEED, respectively.

Table 4 shows the results of (García et al., 2016) achieved 88.3% on valence and 90.6% on arousal. However, data used for experiments is limited and classification model is more fit in classifying small amount of data with high dimensions. The approach of (Koelstra et al., 2012) used a Gaussian Bayes classifier and experiment results proved that EEG signals are effective in emotion recognition by DEAP

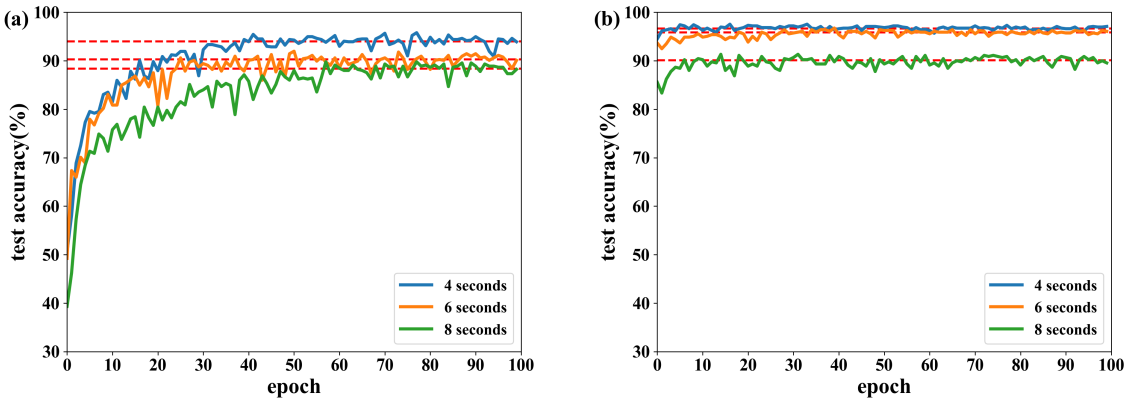

Figure 14. Accuracy comparison of two networks using features extracted from different length of overlap on SEED dataset, in which (a) is result of CNN and (b) is result of the proposed network.

Table 4. Performance comparisons with other approaches.

Classification methods	DEAP Dataset		SEED Dataset
	valence	arousal	
CNN+statistical methods (Tripathi et al., 2017)	81.4%	73.4%	/
Gaussian Bayes (Koelstra et al., 2012)	57.6%	62.0%	/
Deep SAE+RSP (Zhang et al., 2017)	73.1%	80.8%	/
BDGLS+DE (Wang et al., 2018)	/	/	93.7%
DGCNN+DE (Song et al., 2018)	/	/	90.4%
GP+LVM (García et al., 2016)	88.3%	90.6%	/
BLSTM+DE (Wang et al., 2019)	/	/	94.96%
Physe Space Dynamics (Soroush et al., 2019)	84.6%	87.4%	/
SDEL+PCA (Ullah et al., 2019)	82.8%	74.5%	/
This work [PCC]	**89.49%**	**92.86%**	**96.77%**

dataset. Recent study (Tripathi et al., 2017) used extracted data for classification and better accuracy results are obtained by using CNN, where the classification accuracy of valence and arousal is 81.4% and 73.4%, respectively. The approaches of (Wang et al., 2018) and (García et al., 2016) used DE-based features and dynamical graph convolutional neural networks and the accuracy achieved 93.7%.In the approach of (Wang et al., 2019), BLSTM and other machine learning classifiers such as SVM and LR are used for emotion recognition. BLSTM achieve the best accuracy of 94.96% on SEED dataset, which is better than SVM and LR. In the approach of (Soroush et al., 2019), phase space dynamics is introduced to classify emotions, which achieves 87.42% on arousal and 84.59% on valence, respectively. Sparse discriminative ensemble is used for feature extraction in (Ullah et al., 2019) and achieves 82.81% on valence, 74.53% on arousal, respectively. In this work, both of the DEAP and SEED datasets are used for experiments, where accuracies achieve 89.49%, 92.86% in valence and arousal on DEAP dataset, and 96.77% on SEED dataset. Results demonstrate that the proposed network is more powerful than CNN and other approaches.

4 DISCUSSION AND CONCLUSION

There are some points worth discussing. First, the proposed model can be trained through an end-to-end way which is different from this work. The end-to-end training method was tested and it obtained a similar performance. However, training model can be further investigated and optimized in the future work. Second, labels are used in feature extraction. It should be noticed that many feature extraction algorithms use

labels such Relief and ReliefF (Kira and Rendell, 1992), where feature weights are calculated according to samples in the same and different classes. Label information has been used in the feature extraction process (Bohgaki et al., 2014; Zhang et al., 2016b). Third, constructing an autoencoder-like structure is another way for emotion recognition, which can be investigated in the future work.

In this work, a new deep network is proposed to classify EEG signals for emotion recognition. The CNN and the proposed network are applied for two different datasets, i.e. the DEAP and SEED datasets. In the proposed network, CNN and SAE are trained for features extraction, in which by combining supervised learning of CNN and unsupervised learning of SAE, more useful features are extracted. Experimental results show that the the proposed network achieves a better performance than CNN and other approaches. It also shows that when embedding a SAE structure into a CNN, accuracy is better comparing to CNN with the same parameters and structure as the proposed network. In the proposed network, three different features are extracted for classifications. Results showed that by using PCC-based features, the average recognition accuracy of the proposed network can reach 89.49% on valence, 92.86% on arousal for DEAP and 96.77% for SEED, where the the proposed network has a faster convergence speed. In addition, overlap length also affects the performance, and results under SEED dataset showed that data of 8s with overlap of 4s can achieve the best result. It is also found that the data processed by SAE is easily classified in the proposed network, which indicates that the SAE is effective in extracting features from EEG data. Future work will consider using SAE and other classifiers to further improve the classification performance.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China under Grants 61976063 and 61603104, the Guangxi Natural Science Foundation under Grant 2017GXNSFAA198180, the funding of Overseas 100 Talents Program of Guangxi Higher Education, research Fund of Guangxi Key Lab of Multi-source Information Mining & Security (19-A-03-02), and the Innovation Project of Guangxi Graduate Education under Grant YCSW2020102.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation* (USENIX Association), 265–283

Agrafioti, F., Hatzinakos, D., and Anderson, A. K. (2012). Ecg pattern analysis for emotion detection. *IEEE Transactions on Affective Computing* 3, 102–115. doi:10.1109/T-AFFC.2011.28

Bahari, F. and Janghorbani, A. (2013). Eeg-based emotion recognition using recurrence plot analysis and k nearest neighbor classifier. In *2013 20th Iranian Conference on Biomedical Engineering (ICBME)* (IEEE), 228–233

Bohgaki, T., Katagiri, Y., and Usami, M. (2014). Pain-relief effects of aroma touch therapy with citrus junos oil evaluated by quantitative eeg occipital alpha-2 rhythm powers. *Journal of Behavioral and Brain Science* 4, 11–22. doi:10.4236/jbbs.2014.41002

Cheng, B. and Liu, G. (2008). Emotion recognition from surface emg signal using wavelet transform and neural network. In *2008 2nd International Conference on Bioinformatics and Biomedical Engineering* (IEEE), 1363–1366

Chi, Y. M., Wang, Y.-T., Wang, Y., Maier, C., Jung, T.-P., and Cauwenbe, G. (2012). Dry and noncontact eeg sensors for mobile brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 20, 228–235. doi:10.1109/TNSRE.2011.2174652

Danelljan, M., Robinson, A., Khan, F. S., and Felsberg, M. (2016). Beyond correlation filters: Learning continuous convolution operators for visual tracking. In *European Conference on Computer Vision*, eds. B. Leibe, J. Matas, N. Sebe, and M. Welling (Springer), 472–488

Fink, M. (1969). Eeg and human psychopharmacology. *IEEE Transactions on Information Technology in Biomedicine* 9, 241–258. doi:10.1146/annurev.pa.09.040169.001325

Fu, Q., Luo, Y., Liu, J., Bi, J., Qiu, S., Cao, Y., et al. (2017). Improving learning algorithm performance for spiking neural networks. In *2017 IEEE 17th International Conference on Communication Technology (ICCT)* (IEEE), 1916–1919

García, H. F., Álvarez, M. A., and Álvaro A. Orozco (2016). Gaussian process dynamical models for multimodal affect recognition. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE), 850–853

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE), 580–587

Hiraiwa, A., Shimohara, K., and Tokunaga, Y. (1989). Emg pattern analysis and classification by neural network. In *Conference Proceedings., IEEE International Conference on Systems, Man and Cybernetics* (IEEE), 1113–1115

Hjorth, B. (1970). Eeg analysis based on time domain properties. *Electroencephalography and clinical neurophysiology* 29, 306–310. doi:10.1016/0013-4694(70)90143-4

Huang, Y.-J., Wu, C.-Y., Wong, A. M.-K., and Lin, B.-S. (2015). Novel active comb-shaped dry electrode for eeg measurement in hairy site. *IEEE Transactions on Biomedical Engineering* 62, 256–263. doi:10.1109/TBME.2014.2347318

Kira, K. and Rendell, L. A. (1992). A practical approach to feature selection. In *Machine Learning Proceedings*. 249–256

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdan, A., Ebrahimi, T., et al. (2012). Deap: A database for emotion analysis ;using physiological signals. *IEEE Transactions on Affective Computing* 3, 18–31. doi:10.1109/T-AFFC.2011.15

Kübler, A., Furdea, A., Halder, S., Hammer, E. M., Nijboer, F., and Kotchoubey, B. (2009). A brain – computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. *Annals of the New York Academy of Sciences* 1157, 90–100. doi:10.1111/j.1749-6632.2008.04122.x

Lane, R. D., Reiman, E. M., Bradley, M. M., Lang, P. J., Ahern, G. L., Davidson, R. J., et al. (1997). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia* 35, 1437–1444. doi:10.1016/S0028-3932(97)00070-5

Lewis, R. S., Weekes, N. Y., and Wang, T. H. (2007). The effect of a naturalistic stressor on frontal eeg asymmetry, stress, and health. *Biological Psychology* 75, 239–247. doi:10.1016/j.biopsycho.2007.03.004

Li, M. and Lu, B.-L. (2009). Emotion classification based on gamma-band eeg. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE), 1223–1226

Li, X., Hu, B., Sun, S., and Cai, H. (2016). Eeg-based mild depressive detection using feature selection methods and classifiers. *Computer Methods and Programs in Biomedicine* 136, 151–161. doi:10.1016/j.cmpb.2016.08.010

Liu, J., Huang, Y., Luo, Y., Harkin, J., and McDaid, L. (2019). Bio-inspired fault detection circuits based on synapse and spiking neuron models. *Nerocomputing* 331, 473–482. doi:10.1016/j.neucom.2018.11.078

Liu, J., Mcdaid, L. J., Harkin, J., Karim, S., Johnson, A. P., Millard, A. G., et al. (2018a). Exploring self-repair in a coupled spiking astrocyte neural network. *IEEE Transactions on Neural Networks and Learning Systems* 30, 865–875. doi:10.1109/TNNLS.2018.2854291

Liu, J., Sun, T., Luo, Y., Fu, Q., Cao, Y., Zhai, J., et al. (2018b). Financial data forecasting using optimized echo state network. In *25th International Conference on Neural Information Processing(ICONIP)*, eds. L. Cheng, A. C. S. Leung, and S. Ozawa (Springer, Cham), 138–149

Liu, Y.-J., Yu, M., Zhao, G., Song, J., Ge, Y., and Shi, Y. (2018c). Real - time movie - induced discrete emotion recognition from eeg signals. *IEEE Transactions on Affective Computing* 9, 550–562. doi:10.1109/TAFFC.2017.2660485

Luo, Y., Lu, Q., Liu, J., Fu, Q., Harkin, J., v, L. M., et al. (2018). Forest fire detection using spiking neural networks. In *Proceedings of the 15th ACM International Conference on Computing Frontiers* (ACM), 371–375

Nie, D., Wang, X.-W., Shi, L.-C., and Lu, B.-L. (2011). Eeg-based emotion recognition during watching movies. In *2011 5th International IEEE/EMBS Conference on Neural Engineering* (IEEE), 667–670

Orhan, U., Hekim, M., and Ozer, M. (2011). Eeg signals classification using the k-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications* 38, 13475–13481. doi:10.1016/j.eswa.2011.04.149

Petrantonakis, P. C. and Hadjileontiadis, L. J. (2010). Emotion recognition from eeg using higher order crossings. *IEEE Transactions on Information Technology in Biomedicine* 14, 186–197. doi:10.1109/TITB.2009.2034649

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of voice* 9, 235–248. doi:doi:10.1016/S0892-1997(05)80231-0

Song, T., Zheng, W., Song, P., and Cui, Z. (2018). Pnn for eeg-based emotion recognition. *IEEE Transactions on Affective Computing (Early Access)* in press. doi:10.1109/TAFFC.2018.2817622

Soroush, M. Z., Maghooli, K., Setarehdan, S. K., and Nasrabadi, A. M. (2019). A novel eeg-based approach to classify emotions through phase space dynamics. *Signal, Image and Video Processing* 13, 1149–1156. doi:10.1007/s11760-019-01455-y

Subasi, A. and Gursoy, M. I. (2010). Eeg signal classification using pca, ica, lda and support vector machines. *Expert Systems with Applications* 37, 8659–8666. doi:10.1016/j.eswa.2010.06.065

Tabar, Y. R. and Halici, U. (2016). A novel deep learning approach for classification of eeg motor imagery signals. *Journal of Neural Engineering* 14, 90–100. doi:10.1088/1741-2560/14/1/016003

Tripathi, S., Acharya, S., Sharma, R. D., Mittal, S., and Bhattacharya, S. (2017). Using deep and convolutional neural networks for accurate emotion classification on deap dataset. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. 4746–4752

Ullah, H., Uzair, M., Mahmood, A., Ullah, M., Khan, S. D., and Cheikh, F. A. (2019). Internal emotion classification using eeg signal with sparse discriminative ensemble. *IEEE Access* 7, 40144–40153. doi:10.1109/ACCESS.2019.2904400

van den Broek, E. L. (2013). Ubiquitous emotion-aware computing. *Personal and Ubiquitous Computing* 17, 53–67. doi:10.1007/s00779-011-0479-9

Wang, X., Zhang, T., Xu, X., Chen, L., Xing, X., and Chen, C. L. P. (2018). Eeg emotion recognition using dynamical graph convolutional neural networks and broad learning system. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (IEEE), 1240–1244

Wang, Y., Qiu, S., Li, J., Ma, X., Liang, Z., Li, H., et al. (2019). Eeg-based emotion recognition with similarity learning network. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE), 1209–1212

Zhang, D., Yao, L., Zhang, X., Wang, S., Chen, W., and Boots, R. (2018). Cascade and parallel convolutional recurrent neural networks on eeg-based intention recognition for brain computer interface. In *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*

Zhang, J., Chen, M., Hu, S., Cao, Y., and Kozma, R. (2016a). Pnn for eeg-based emotion recognition. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (IEEE), 002319–002323

Zhang, J., Chen, M., Zhao, S., Hu, S., Shi, Z., and Cao, Y. (2016b). Relieff-based eeg sensor selection methods for emotion recognition. *Sensors* 16, 40144–40153. doi:10.3390/s16101558

Zhang, Q., Chen, X., Zhan, Q., Yang, T., and Xia, S. (2017). Respiration-based emotion recognition with deep learning. *Computers in Industry* 92-93, 84–90. doi:10.1016/j.compind.2017.04.005

Zheng, W. (2017). Multichannel eeg-based emotion recognition via group sparse canonical correlation analysis. *IEEE Transactions on Cognitive and Developmental Systems* 9, 281–290. doi:10.1109/TCDS.2016.2587290

Zheng, W.-L. and Lu, B.-L. (2015). Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development* 7, 162–175. doi:10.1109/TAMD.2015.2431497

Zheng, W.-L., Zhu, J.-Y., Peng, Y., and Lu, B.-L. (2014). Eeg-based emotion classification using deep belief networks. In *2014 IEEE International Conference on Multimedia and Expo* (IEEE), 1–6

Zhou, Z. H. and Feng, J. (2017). Deep forest: Towards an alternative to deep neural networks. In *IJCAI International Joint Conference on Artificial Intelligence*. 3553–3559