IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 April 2023, accepted 18 April 2023, date of publication 25 April 2023, date of current version 3 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3270317

RESEARCH ARTICLE

EEG-Based Emotion Recognition Using Spatial-Temporal-Connective Features via Multi-Scale CNN

TIANYI LI, BAOLE FU, ZIXUAN WU, AND YINHUA LIU
Institute of Future, Qingdao University, Qingdao 266000, China

Corresponding author: Yinhua Liu (liuyinhua@qdu.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2020YFB1313600.

ABSTRACT Electroencephalography (EEG) signals from each channel mainly reflect activities of the brain region close to the channel position, and the activities cooperated by various brain regions are response to the emotion-induced stimuli. In this paper, temporal, spatial and connective features are extracted from EEG signals gotten around the head, and used for emotion recognition via a proposed model, spatial-temporal-connective muti-scale convolutional neural network (STC-CNN). The channel-to-channel connectivity is gotten to describe brain region-to-region cooperation under emotion stimuli. The model obtained an average accuracy of 96.79% and 96.89% in classifying the two emotional dimensions of valence and arousal.

INDEX TERMS Emotion recognition, EEG, connective features, STC-CNN.

I. INTRODUCTION

Emotion is an important component in daily life [1]. Negative emotions make people more prone to mental illnesses such as depression, schizophrenia. Therefore, emotion recognition plays an important role in the regulation of emotions. Emotional information is expressed mainly through facial expressions, voice tones, and physiological signals (EEG,EOG) [2]. Physiological signals are not affected by subjective factors and can accurately respond to emotional information. EEG signals have the advantage of being noninvasive and easy to use [3]. Therefore, EEG emotion recognition has been the focus of researchers with good results. In emotion recognition, the boundaries that distinguish emotional states are fuzzy but the changes in states are continuous. To better describe emotional states, a dimensional model is used. Emotional states are described as coordinate points in space using several basic dimensions with continuous values (e.g. arousal, valence), each of which is a measure of some aspect of emotion. Zheng used a bipolar dimensional model that includes arousal and valence dimensions [4].

The discrete emotion model describes emotions such as happiness, anger, sadness, disgust, and then all other emotions can be fused by these basic emotions [5]. The value of the valence axis from positive to negative refers to a measure of whether an individual's emotion is positive or negative. Similarly, positive values in arousal states indicate activated states (arousal), while negative values indicate inactive states (apathy). Dimensional emotion models are more intuitive and easier to accurately define the state of emotions by coordinates, and therefore are widely used in the task of emotion recognition. P.Bashivan proposed networks with feature reuse mechanism and used the dimensional emotion model to achieve high accuracy emotion recognition [6].

The advantages of deep learning are gradually manifested in the application of EEG. Deep learning, with its ability to automatically extract features and achieve end-to-end classification, is increasingly being used in emotion recognition. Kim proposed a model which based on a convolutional long short-term memory network [7]. Zhang proposed a spatial-temporal recurrent neural network (STRNN) for emotion recognition, and the results showed that STRNN significantly outperformed SVM [8]. In Zhang [9], [10], a convolutional neural network (CNN) is used to capture the relationship between channels by aggregating features

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

of adjacent channels using convolutional layers. Nakisa use a convolutional neural network (ConvNet) long short-term memory (LSTM) model to fuse the EEG and BVP signals [11]. Alhagry proposed a network to identify emotions from raw EEG signals, which uses LSTM-RNN to learn features from EEG signals and perform classification [12]. Tao proposed an attention-based convolutional recurrent neural network (ACRNN) to extract more informative discriminative features [13].

EEG signals are non-smooth and using a single size of convolution kernel may not sufficiently extract rich features for EEG classification tasks. Previous studies have shown that using convolutional kernels of different sizes can learn multi-scale EEG features that are beneficial for different EEG classification tasks [14]. Liu used ResNet as the backbone of EEG-based emotion recognition, where deep spatial features were extracted using pre-trained ResNet in the same way as the original EEG signal [15]. Although ResNet has a strong feature extraction capability, it cannot capture the contextual information of time-series signals and limits the correlation between channels due to the fixed size of its convolutional kernel. Zhang proposed a ResNet-based dynamic multiscale network for EEG signal classification, which can learn multi-scale features from different receptive domains at a finer level [9]. Li proposed a multi-scale fusion CNN model based on the attention mechanism. Multi-scale CNNs have also been introduced to emotion recognition [16]. Phan proposed a two-dimensional CNN model with convolutional kernels of different sizes for arousal and potentiated binary classification. They used kernel sizes of 5×5 and 7×7 to extract spatial features to describe the short-range and long-term relationships between EEG channels [14].

Emotion recognition based on EEG is not only improved on the model, but also studied on the extraction of EEG features in the temporal domain, frequency domain and spatial domain. Various handcrafted features have been used to extract the differences between different emotional states. Zheng manually extracted five features, power spectral density (PSD), differential entropy (DE), difference asymmetry, rational asymmetry, asymmetry, and differential causal features, to identify emotions using SVM and graph regularized extreme learning machine (GELM) classifiers, respectively [17]. Liu manually extracted the spatial-temporal feature of EEG signals and proposed the 3DCNN model to better extract the dynamic relation well between the multi-channel EEG signals and the internal spatial relation of the multi-channel EEG signals in view of the differences in EEG signals under different emotions [18].

Deep learning-based EEG features (DE and PSD) can represent the activity of each brain region. However, the brain is an organic whole. Brain is essentially a network, of which the function can be interpreted as the interaction between regions through the network. The expression of emotions is a result of cooperation between different areas of the brain [19]. Linhartov Phe have shown that people are able to regulate their brain activity in the presence of rt-fMRI-NF in areas of the brain associated with emotion regulation, including the amygdala, anterior insula and anterior cingulate cortex [20]. Therefore, emotion can be analyzed not only through the activities of each brain region, but also the connectivity between different brain regions, which motivates us to take advantage of the brain connectivity in EEG signals. Therefore, the connective feature of EEG signals, which are different from the conventional EEG features, provide the relevant activity information of emotion. Liu used Pearson correlation coefficient (PCC) to estimate the correlation between all channel pairs. They extracted PCC features and convolutional neural network (CNN) features in parallel to classify emotional states [21]. Lee and Hsieh used three types of connective feature to distinguish three different emotional states. Phase synchronization captures synchronous activity in the brain, allowing for comprehensive mining of effective structural and functional cognitive patterns [22]. Li used PLV for emotion recognition, which captures nonlinear phase synchronization between different brain regions [23].

This paper has two main contributions: (1) Features of channel-to-channel connectivity in each region of the brain to describe the mechanisms of cooperation between brain region-to-region, and combined with spatial-temporal features for modeling emotion recognition. (2) A multi-scale convolution kernel is proposed for CNN, which fully extracts the differential EEG signal features and capture the interactions between different brain regions in different emotional states. Experiments were conducted on the DEAP dataset, and we obtained higher accuracy than traditional methods on both valence and arousal dimensions. The results show that considering spatial-temporal feature and connective feature is more effective in EEG emotion recognition.

II. MODEL

As the embodiment of emotion changes, the spatial-temporal features of EEG signals significantly represent the activities of each brain region. The generation of emotions is the result of brain region-to-region cooperation. The correlation features between each brain area provide information about emotion related activities. Different from the spatial-temporal features, the connective features are used as both the supplement to the spatial-temporal features and the input of the (STC-CNN) model. As shown in Figure 1.

A. SPATIAL-TEMPORAL FEATURE EXTRACTOR

The expression of emotions is continuous.Therefore, we need to process EEG signals over a continuous period of time. The processed EEG signal will be divided into N segments by a length of the Ts window. EEG signals embody a lot of emotion-related features, however it also contains much noise. Putting the signals directly into the network will extract features unrelated and affect the accuracy of recognition.

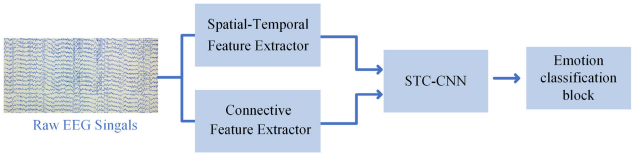


FIGURE 1. The structure of model.

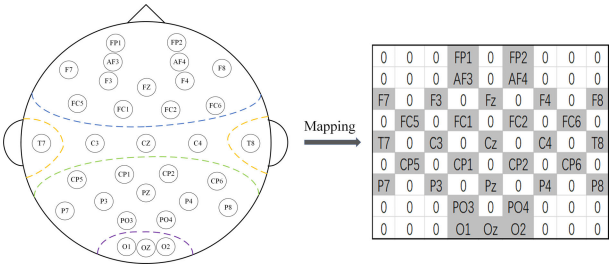


FIGURE 2. Mapping of electrodes distribution to matrix.

To extract more accurate features and reduce the influence of noise, it is necessary to further extract emotion-related features manually.

The model extracts the differential entropy (DE) features of θ (4-8Hz), α (8-12Hz),β (12-30Hz) and γ(30-50Hz) frequency bands from the EEG data of each length of Ts, which have been shown to better identify emotions [24]. The formula of DE is

$$\mathrm{D\,E}(x) = -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\delta^2}} \exp\frac{(x-\mu)}{2\delta^2} \log\frac{1}{\sqrt{2\pi\delta^2}} \cdot \exp\frac{(x-\mu)}{2\delta^2} dx = \frac{1}{2}\log 2\pi e\delta^2 \tag{1}$$

where the random variable x follows a Gaussian distribution $N\left(\mu, \delta^2\right)$.

EEG electrodes are equipped on the brain surface with spatial distribution according to the international 10-20 system standard. The brain is divided into frontal lobe, temporal lobe, parietal lobe, and occipital lobe. For frontal lobe, electrodes are FP1, FP2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, and FC6. For temporal lobe, electrodes are T7, T8. For parietal lobe, electrodes are CP5,CP1,CP2,CP6,P7,P3,Pz,P4,P8,PO3,PO4. For occipital lobe, electrodes are O1,Oz,O2. C3,Cz,C4, electrodes are located in the middle of the brain.To extract spatial information,the spatial distribution of electrodes is converted into matrix form, and the conversion process is shown in Figure 2. The format of two-dimensional transformation feature matrix, where FP1,FP2,AF3…,O1,Oz,O2 are the corresponding channels of EEG, and the remaining points are filled with zeros (2), as shown at the bottom of the next page.

EEG signals in period Ts are represented as a matrix $S(t)$ for feature extraction. In other words, $V_i(t)$ is the measured EEG data of the i-th channel, and the unit is μV. To combine the temporal features with the spatial features, the DE features is converted into a two-dimensional matrix as shown in Figure 3, where $C_j \in R^{X\times Y}(j \in \{1, 2, 3, 4, N\})$, X and Y are set to 9.

B. CONNECTIVE FEATURE EXTRACTOR

For the connective feature extractor, depending on the distribution of electrode channels in brain regions, the collaboration of brain regions is refected by the connectivity information we extracted from different electrode channels. Given the volume conduction effect, the brain signals obtained from adjacent brain regions tend to be similar, and the connectivity features can be represented as a smooth matrix. The sorting method for this is to start from the electrode in the left frontal area and select the electrode closest one to the current electrode as the next, which is shown as following: FP1, AF3, F3, FC5, T7, CP5, P7, P3, PO3, O1, Oz, O2, PO4, P4, P8, CP6, T8, FC6, F8, F4, AF4, FP2, Fz, FC1, C3, CP1, Pz, CP2, C4, FC2, Cz. The number (i, k) in the connective matrix represents the correlation between the EEG signals of the i-th channel and the k-th channel. As shown in Fig.4, EEG signals are first divided into four frequency bands (θ(4-8Hz), α(8-12Hz), β(12-30Hz) and γ(30-50Hz)). Then, the connective features will be extracted from the four frequency bands and connected together. The dimension of the connective feature matrix is $4 \times 32 \times 32$, where 4 denotes the four frequency bands extracted from the EEG signal, and 32×32 denotes the number of channels of the EEG signal. Connexity reflects the correlations between different brain regions, and each node represents the activity intensity between different EEG channels.

Connective features are divided into three categories 1)*PCC*

PCC is a linear correlation coefficient reflecting the linear correlation between EEG signals of different channels. The calculation formula is as follows

$$PCC = \frac{\mathrm{cov}(w, z)}{\sigma_w \sigma_z} \tag{3}$$

where w and z denote the two EEG signals from different channels, $\mathrm{cov}(w, z)$ is the covariance between w and z, σ_w and σ_z is the standard deviation of w and z. The value of PCC lies between -1 and 1, of which the larger the absolute value, the stronger the linear correlation.

2)*PLV*

PLV describes the phase synchronization between two EEG signals by calculating the average value of phase difference, and the calculation formula is

$$PLV = \frac{1}{M}\left|\sum_{n=1}^{M} e^{j\Delta\phi_n}\right| \tag{4}$$

where M is the sampling point of EEG signal, and $\Delta\phi_n$ is the phase difference of the NTH sampling point. It used the Hilbert-Huang transform to calculate the phase difference between two EEG signals. The value of PLV is between 0 and 1.

3)*PLI*

PLI is another measure of phase synchronization between two signals. By calculating the average value of phase

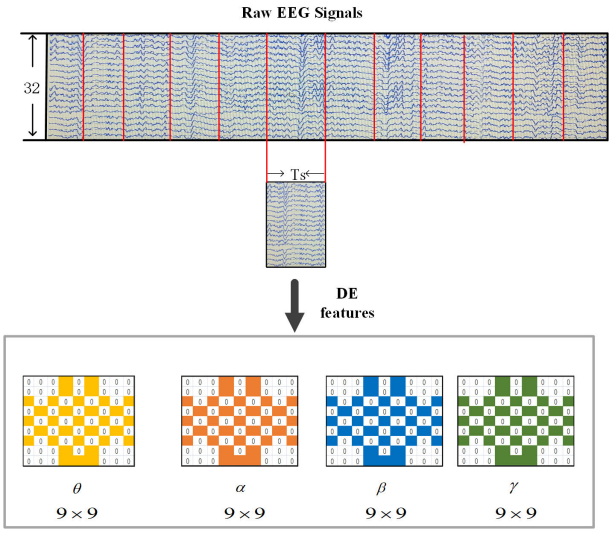


FIGURE 3. Spatial-temporal feature extraction.

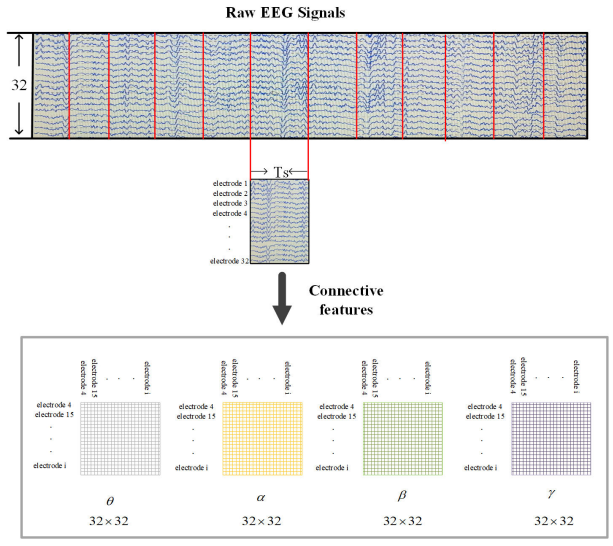


FIGURE 4. Connective feature extraction.

difference, the calculation formula is

$$PLI = \frac{1}{M}\left|\sum_{n=1}^{M} \text{sign}\left(\Delta\phi_n\right)\right| \tag{5}$$

where M is the number of sampling points of EEG signal, sign(.) denotes the sign function and $\Delta\phi_n$ is the phase difference of the nth sampling point.

C. STC-CNN MODEL

To combine spatial-temporal features with connective features, this study proposes a convolutional neural network of STC-CNN. As shown in Fig.5, the designed model is composed of convolutional layer, fully connected layer and max-pooling layer, which adopts a two-branch model architecture. The inputs of the model are 3*D*-DE features and connective features. The CNN multi-scale method is used to extract high-level spatial-temporal and connective information from the features. Then, a fully connected layer is added after the fusion layer to extract the deep features. Finally, a softmax layer receives the output of the fusion layer for the final emotion prediction. For DE branches, the model inputs are 3*D*-DE features $D \in R^{9\times9\times4}$. The model consists of three convolutional layers, a scaled exponential linear unit (SELU) activation function, and a Flatten layer. The first layer is Convs1,there are 32 filters, the convolution size is 5×5 , stride is 1, the other two convolution operations are similar to Convs1, where Convc2 has 64 kernels of size 3×3 convolution and Convc3 has 128 kernels of size 3×3. Deep level features will be obtained through these operations and turned into one-dimensional features Fs through the Flatten layer.

The input of the model is connectivity feature $C \in R^{32\times32\times4}$, which reflects the correlation between different EEG channels. The multi-scale method is adopted, and the size of each convolution is different, indicating that different receptive domains can fully capture the connection feature of different channels and fully reflect the interaction of different brain regions. The model consists of three convolutional layers, two max-pooling layers and one Flatten layer. The first layer is Convc1, with 32 filters, the convolution size is 7×7 , the stride is 1, and the SELU activation function. The remaining two convolution operations are similar to Convc1, where Convc2 has 64 kernels of size 5×5 convolution and Convc3 has 128 kernels of size 3×3. The maximum pooling layer is respectively distributed behind the second convolutional layer and the third convolutional layer to reduce the feature dimension and the number of parameters. These Conv operation to obtain the feature of the deep, by will Flatten

$$S(t) = \begin{bmatrix} 0 & 0 & 0 & V_1(t) & 0 & V_{17}(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & V_2(t) & 0 & V_{18}(t) & 0 & 0 & 0 \\ V_4(t) & 0 & V_3(t) & 0 & V_{19}(t) & 0 & V_{20}(t) & 0 & V_{21}(t) \\ 0 & V_5(t) & 0 & V_6(t) & 0 & V_{22}(t) & 0 & V_{23}(t) & 0 \\ V_8(t) & 0 & V_7(t) & 0 & V_{24}(t) & 0 & V_{25}(t) & 0 & V_{26}(t) \\ 0 & V_9(t) & 0 & V_{10}(t) & 0 & V_{28}(t) & 0 & V_{27}(t) & 0 \\ V_{12}(t) & 0 & V_{11}(t) & 0 & V_{16}(t) & 0 & V_{29}(t) & 0 & V_{30}(t) \\ 0 & 0 & 0 & V_{13}(t) & 0 & V_{31}(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & V_{14}(t) & V_{15}(t) & V_{32}(t) & 0 & 0 & 0 \end{bmatrix} \tag{2}$$

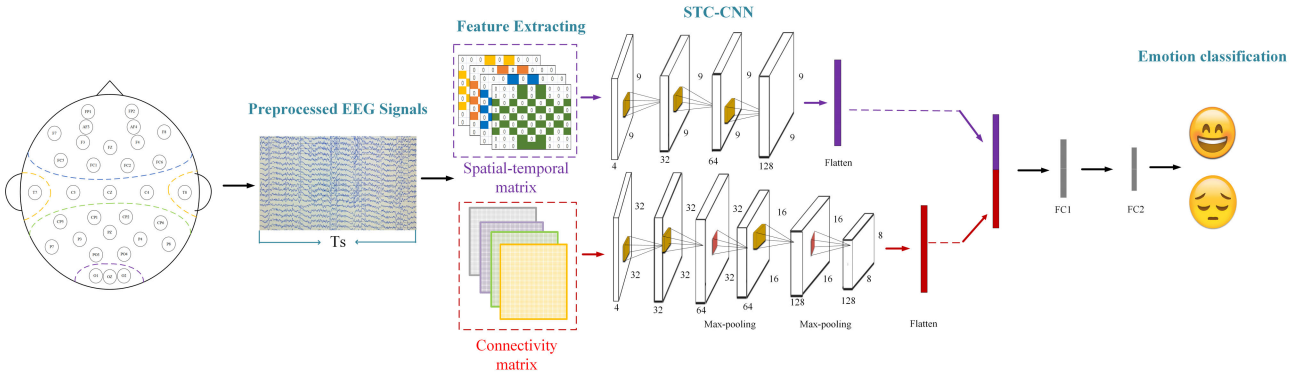


FIGURE 5. The structure of the proposed STC-CNN.

layer characteristic flat for a vector Fc. We fuse EEG features with different attributes and make use of the complementarity between features, and fuse the advantages between features, so as to improve the performance of the model. In the fusion stage, the spatial-temporal data and connectivity data are normalized to improve the data concentration and eliminate the adverse effects of sample outliers. The spatial-temporal DE features Fs and connectivity features Fc will be respectively connected to an integrated vector Fsc, which is the final comprehensive feature used for emotion classification. The final feature map will be used to visualize the model in the following analysis, and the classification part consists of two fully connected layers. The first layer FC1 contains 128 neurons and the activation mode is ReLU. There are two neurons in the output layer FC2. The softmax function maps the output y of the previous layer to the predicted probability p, and the output is as follows

$$P(l_i/y) = \frac{\exp(y)}{\sum_{i=1}^{N} \exp(y)} \tag{6}$$

At the same time, a dropout layer is added between the fully connected layers, so as to effectively suppress the overfitting problem.

III. DATA PROCESSING

A. DEAP DATASET

The DEAP dataset included EEG data from 32 participants of 16 males and 16 females. Participants were induced to feel similar emotions by watching 40 different one-minute music videos as emotional stimuli. During each subject's 40 trials, EEG signals were recorded using a 32-channel system. Music video playback, the experiment will last for 63 seconds, the first 3seconds is the time of each video conversion, the last 60 seconds is the actual music video playback time. During this process, participants should try to maintain their balance and reduce movements. Thus, the size of each participant's EEG was $60 \times 32 \times 40$. Each participant was asked to make 40 self-assessments on the SAM questionnaire and rate each video on a scale of 1 (low) to 9 (high) on arousal and pleasure levels. The details of the DEAP dataset are shown in Table 1.

TABLE 1. Description of the data set.

Type	Discription
Subject	32 (16 males and 16 females)
Channels	32
Sampling rate	128Hz
Movies	40 different movie clips
Data shape	(40,32,8064)

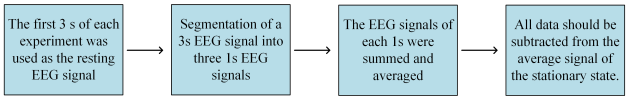


FIGURE 6. The diagram of baseline correction.

B. DATA PREPROCESSING

For data preprocessing, in order to reduce noise and improve the stability of EEG data, baseline correction and Z-score normalization are carried out on the dataset, which are common preprocessing methods of EEG. The baseline correction diagram is shown in Fig.6. First, data downsampling was used to reduce the frequency to 128Hz, and then blind source separation was used to remove electrical eye (EOG) artifacts and 3s baseline data. These operations can improve the accuracy of emotion recognition by reducing the interference to basic emotional states before the task cycle. The EEG signal is split into T=3s long segments with an overlap time of 2.5s to obtain a sufficient number of data samples to train the STC-CNN. Since the database has a trial time of 1 minute, we obtained 115 EEG signal segments per trial. In this experiment, valence and arousal were selected as emotion evaluation criteria, and the threshold of the two categories was set as 5 according to the level of arousal and valence (1-5 was negative emotion, 6-9 was positive emotion).

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A dimensional model of arousal, valence is used to better describe emotional states. This paper first describes the

TABLE 2. Relevant parameters in the network.

Parameter	Value
Epoch	100
Dropout	0.4
Batch size	128
Learning rate	0.0001
The number of classes	2

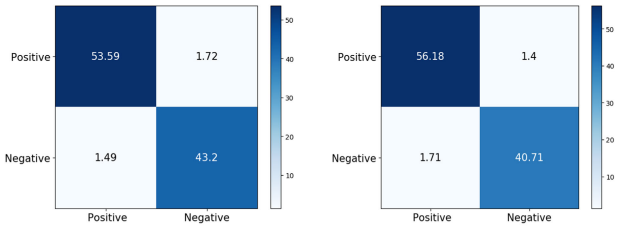


FIGURE 7. Accuracy of model STC-CNN (PLV) in valence(a) and arousal(b).

implementation of the model. Using the dimensional model, the accuracy of three different connectivity features for emotion recognition is compared. Finally, the results of the STC-CNN are compared with related research models, and the STC-CNN achieves high accuracy on the arousal and valence dimensions for the DEAP dataset.

A. IMPLEMENTATION OF THE MODEL

For the STC-CNN, cross entropy was used as the loss function. First, the Adam optimiser was used to minimise the loss function. For the first 20 learning cycles, the learning rate is 0.001, and the learning rate drops to 0.0001 for later learning cycles. Further,to solve the overfitting problem, we have set the dropout value to 0.4. For the dataset,10-fold cross-authentication was used and the performance of the model was assessed by average accuracy across all subjects.

B. EXPERIMENTAL RESULTS

The proposed model STC-CNN, combines the connectivity characteristics between different brain regions under different emotions and the spatial characteristics of EEG channels. Using multi-scale method, the spatial information of EEG signal can be fully extracted. Different sensory domains can capture the interaction of different brain regions in different emotional states. The model performed an experiment on valence and arousal dimensions of samples from the DEAP dataset, as shown in figure 7, and the average accuracy of 96.79% and 96.89% were obtained, respectively.

To better verify whether the model can improve the accuracy of emotion recognition under the three connectivity features. In this paper, three kinds of connectivity features are tested on the DEAP dataset and their accuracy rates are compared. In order to explore the relationship between

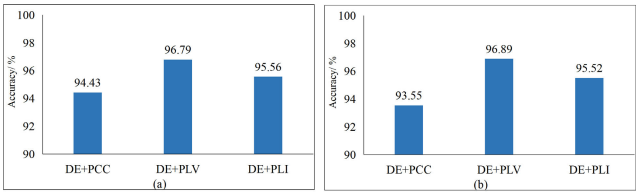


FIGURE 8. The performance of model STC-CNN.

TABLE 3. Accuracy of different feature.

Feature	Valence(%)	Arousal(%)
DE	86.35%	87.51%
PCC	80.28%	81.96%
PLV	87.36%	90.73%
PLI	86.74%	87.35%
DE+PLV	96.79%	96.89%

connectivity features and emotion recognition based on EEG, we conducted experiments on three connectivity features (PCC,PLV,PLI), and analyzed the results in two dimensions, namely valence and arousal. These three types of connection features are extracted from the EEG signals, and the details are as follows:

Figure 8 shows that in the dimension of valence, the final accuracy of PLV connection features is 2.36% higher than PCC and 1.23% higher than PLI. In the dimension analysis of arousal, PLV also had the highest connection characteristics, with an accuracy rate 3.54% higher than PCC and 1.37% higher than PLI.

It can be seen from the above analysis that different connectivity features have an impact on the final accuracy, which indirectly indicates that PLV connectivity characteristics considering phase synchronization are more closely related to emotional EEG. Although PLI also considers the phase synchronization between two signals, it only reflects the indication of the phase difference between the signals, not the size of the phase difference. In the connectivity features PLI, the loss of some EEG signal features leads to low accuracy and performance degradation.

With the purpose of explore the influence of spatial, frequency and connectivity information on EEG emotion recognition, we compared STC-CNN with two-dimensional CNN, and to verify whether the connectivity feature information and temporal and spatial DE feature information are complementary, so as to improve the performance of the model. Table3 shows the average accuracy of our model and the accuracy of two dimensional CNN. The STC-CNN model outperforms other models in both dimensions. STC-CNN was 10.44% higher in valence dimension and 9.38% higher in arousal dimension than the CNN+DE model, which indicated that the connectivity feature was complementary to the two-dimensional mesh DE feature. It could be considered that the connectivity feature could enhance the spatial information of EEG signals. Valence ratio of STC-CNN was

TABLE 4. Accuracy of STC-CNN and other methods.

Methods	Input data	Valence(%)	Arousal(%)
CNN+BiLSTM[25]	Temporal and frequency features	72.38%	72.38%
EEG-GCN[26]	Spatial-temporal features	81.77%	81.95%
GCNN[27]	DE features	90.45%	90.60%
RACNN[28]	Time–frequency and regional features	94.65%	95.55%
STC-CNN	Spatial-temporal and connective features	96.79%	96.89%

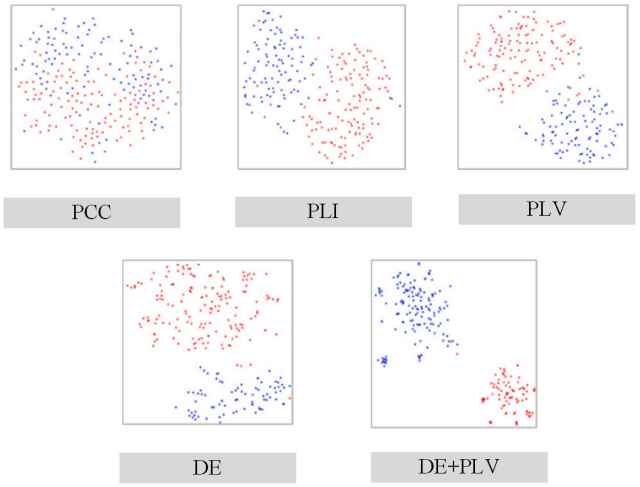


FIGURE 9. Diagram of five feature classification results.

16.51%, 9.43%, 10.05% higher than that of CNN+PCC model, CNN+PLV model and CNN+PLI model, respectively. In arousal dimension, STC-CNN evoked 14.93%, 6.16%, 9.54% higher. DE features are rich in spatio-temporal information, and the addition of emotions as a complement to the connectivity features helps to identify emotions.

To further investigate each model and STC-CNN results, we use t - SNE on final figure on the characteristics of visual features. The feature classification is shown in Figure 9. . It can be observed that the features extracted in the proposed STC-CNN have better separability than other methods.

In order to further verify the performance of the model STC-CNN, we compared the model with the other four models on the DEAP dataset, as shown in Table 3.

Table 4 shows that in the valence dimension and arousal dimension, the network model with the highest accuracy is STC-CNN, with an accuracy rate of 96.79% and 96.89%, respectively. Model STC-CNN was higher than model CNN+Bi-LSTM by 24.41%, model EEG-GCN by 15.02%, model GCNN by 6.34%, and model RACNN by 2.14%.

For the selection of features, the first three models all extract single-level emotional features, and feature extraction is not sufficient. The STC-CNN model extracts multi-level emotion features, and the accuracy of emotion recognition is greatly improved. But the accuracy of RACNN is relatively

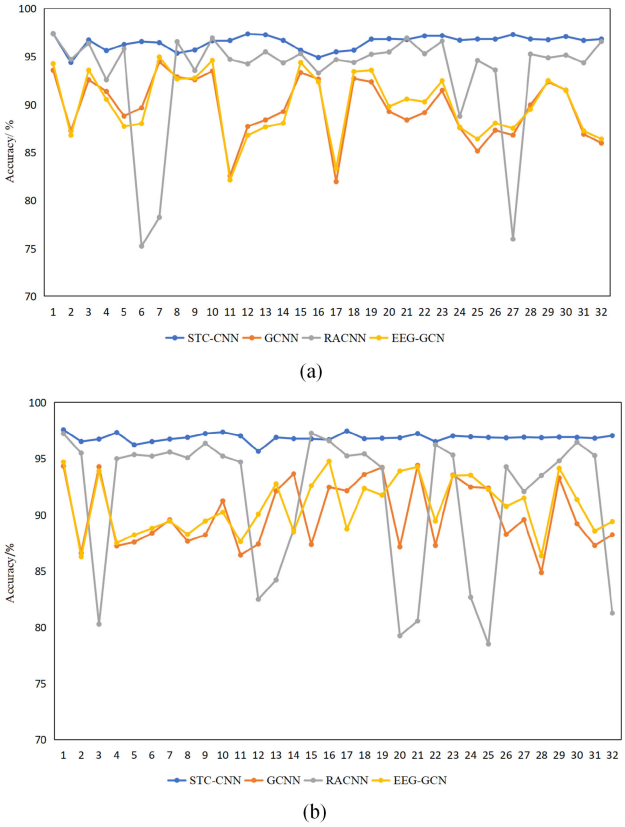


FIGURE 10. Average accuracy of valence (a) and arousal (b) dimensions for 32 subjects.

close, and both combine two different emotional feature. Compared with STC-CNN, the connectivity feature of the entire brain region is more suitable for the recognition of emotions than the distinguishing feature between the two hemispheres.

For the construction of the models, the STC-CNN model not only extracts multi-level features but also has a simple model structure with few parameters, and a multi-scale strategy is more conducive to mining deeper features.

To observe the accuracy performance of the model STC-CNN on each subject. Figure 10 shows that the figures are valence dimension and arousal dimension respectively, and the accuracy of each subject. It can be found that with the change of subjects, STC-CNN has better performance than

other models, which can produce higher accuracy across all subjects.

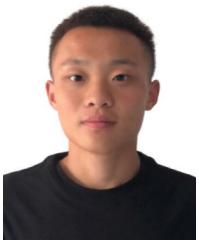
V. CONCLUSION

An emotion recognition method, STC-CNN, is proposed to extract the connective features and the spatial-temporal features by using multi-scale convolution kernel. It is proved that the method can achieve the average accuracy of 96.79% and 96.89% in valance and arousal of DEAP dataset,which is superior to the existing models.

This work is our active attempt at EEG-based emotion recognition. We will continue to explore the impact of the number of EEG signal segments on the accuracy of feature extraction, explore the characteristics of functional brain network connectivity, and intensity differences between the left and right hemispheres of the brain will be considered.

REFERENCES

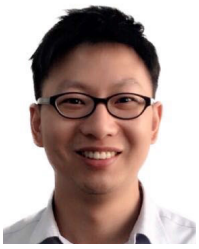
[1] U. Retkoceri, "Remembering emotions," *Biol. Philosophy*, vol. 37, no. 1, p. 5, Feb. 2022.
[2] M. S. Hossain and G. Muhammad, "Emotion recognition using deep learning approach from audio–visual emotional big data," *Inf. Fusion*, vol. 49, pp. 69–78, Sep. 2019.
[3] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Review and classification of emotion recognition based on EEG brain–computer interface system research: A systematic review," *Appl. Sci.*, vol. 7, no. 12, p. 1239, Dec. 2017.
[4] W. Zheng, "Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis," *IEEE Trans. Cogn. Devel. Syst.*, vol. 9, no. 3, pp. 281–290, Sep. 2017, doi: 10.1109/TCDS.2016.2587290.
[5] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 417–429, Jul. 2019.
[6] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," 2015, *arXiv:1511.06448*.
[7] B. H. Kim and S. Jo, "Deep physiological affect network for the recognition of human emotions," *IEEE Trans. Affect. Comput.*, vol. 11, no. 2, pp. 230–243, Apr./Jun. 2020, doi: 10.1109/TAFFC.2018.2790939.
[8] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2018.
[9] G. Zhang, J. Luo, L. Han, Z. Lu, R. Hua, J. Chen, and W. Che, "A dynamic multi-scale network for EEG signal classification," *Frontiers Neurosci.*, vol. 14, Jan. 2021, Art. no. 578255.
[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[11] B. Nakisa, M. N. Rastgoo, A. Rakotonirainy, F. Maire, and V. Chandran, "Automatic emotion recognition using temporal multimodal deep learning," *IEEE Access*, vol. 8, pp. 225463–225474, 2020, doi: 10.1109/ACCESS.2020.3027026.
[12] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 1–4, 2017.
[13] W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, and X. Chen, "EEG-based emotion recognition via channel-wise attention and self attention," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 382–393, Jan. 2023, doi: 10.1109/TAFFC.2020.3025777.
[14] T.-D.-T. Phan, S.-H. Kim, H.-J. Yang, and G.-S. Lee, "EEG-based emotion recognition by convolutional neural network with multi-scale Kernels," *Sensors*, vol. 21, no. 15, p. 5092, Jul. 2021.
[15] S. Liu, X. Wang, L. Zhao, B. Li, W. Hu, J. Yu, and Y.-D. Zhang, "3DCANN: A spatio-temporal convolution attention neural network for EEG emotion recognition," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 11, pp. 5321–5331, Nov. 2022, doi: 10.1109/JBHI.2021.3083525.
[16] D. Li, J. Xu, J. Wang, X. Fang, and Y. Ji, "A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of EEG signals decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2615–2626, Dec. 2020.
[17] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auto. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.
[18] T. Liu and D. Yang, "A three-branch 3D convolutional neural network for EEG-based different hand movement stages classification," *Sci. Rep.*, vol. 11, no. 1, p. 10758, May 2021.
[19] S.-E. Moon, C.-J. Chen, C.-J. Hsieh, J.-L. Wang, and J.-S. Lee, "Emotional EEG classification using connectivity features and convolutional neural networks," *Neural Netw.*, vol. 132, pp. 96–107, Dec. 2020.
[20] P. Linhartová, A. Látalová, B. Kóša, T. Kašpárek, C. Schmahl, and C. Paret, "FMRI neurofeedback in emotion regulation: A literature review," *NeuroImage*, vol. 193, pp. 75–92, Jun. 2019, doi: 10.1016/j.neuroimage.2019.03.011.
[21] N. Liu, Y. Fang, L. Li, L. Hou, F. Yang, and Y. Guo, "Multiple feature fusion for automatic emotion recognition using EEG signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 896–900.
[22] Y-Y. Lee and S. Hsieh, "Classifying different emotional states by means of EEG-based functional connectivity patterns," *PLOS ONE*, vol. 9, no. 4, 2014, Art. no. e95415.
[23] P. Li, H. Liu, Y. Si, C. Li, F. Li, X. Zhu, X. Huang, Y. Zeng, D. Yao, Y. Zhang, and P. Xu, "EEG based emotion recognition by combining functional connectivity network and local activations," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2869–2881, Oct. 2019.
[24] H. Chao and L. Dong, "Emotion recognition using three-dimensional feature and convolutional neural network from multichannel EEG signals," *IEEE Sensors J.*, vol. 21, no. 2, pp. 2024–2034, Jan. 2021.
[25] A. Samavat, E. Khalili, B. Ayati, and M. Ayati, "Deep learning model with adaptive regularization for EEG-based emotion recognition using temporal and frequency features," *IEEE Access*, vol. 10, pp. 24520–24527, 2022.
[26] Y. Gao, X. Fu, T. Ouyang, and Y. Wang, "EEG-GCN: Spatial–temporal and self-adaptive graph convolutional networks for single and multiview EEG-based emotion recognition," *IEEE Signal Process. Lett.*, vol. 29, pp. 1574–1578, 2022.
[27] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106954.
[28] H. Cui, A. Liu, X. Zhang, X. Chen, K. Wang, and X. Chen, "EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network," *Knowl.-Based Syst.*, vol. 205, Oct. 2020, Art. no. 106243.
[29] W. Ko, E. Jeon, S. Jeong, and H.-I. Suk, "Multi-scale neural network for EEG representation learning in BCI," *IEEE Comput. Intell. Mag.*, vol. 16, no. 2, pp. 31–45, May 2021.
[30] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 793–799.
[31] J. Zhang, N. Wang, H. Kuang, and R. Wang, "An improved method to calculate phase locking value based on Hilbert–Huang transform and its application," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 125–132, Jan. 2014.
[32] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Hum. Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, Nov. 2007.



TIANYI LI was born in 1999. He is currently pursuing the master's degree in automation with Qingdao University. His research interest includes affective computing.

BAOLE FU received the bachelor's degree in automation from the Qilu University of Technology, in 2022. He is currently pursuing the master's degree with Qingdao University.

ZIXUAN WU received the bachelor's degree in automation from Yanshan University, in 2019. He is currently pursuing the master's degree with Qingdao University. From 2021 to 2023, he was a Research Assistant with the Affective Computing Project. He has authored two patents.



YINHUA LIU received the Ph.D. degree in mechanical engineering from Sophia University, Japan, in March 2013. He joined Qingdao University, in August 2017. His research interests include intelligent wearables, offshore photovoltaic, deep learning, and artificial intelligence. He has participated in a number of key research and development projects of the Ministry of Science and Technology. He has published more than 20 papers in international and domestic journals and conferences. He has obtained or accepted more than 30 patents in the State Intellectual Property Office. In the field of offshore photovoltaic, intelligent parking, smart city, and other fields, they have developed a number of industry-university-research innovation projects. He won a number of innovation and entrepreneurship awards. They have accumulated rich project resources and technical teams, among which more than 30 team research and development personnel are specialized in software, machinery, electromechanical, embedded, design, and other fields, with interdisciplinary ability and experience. He has a number of product research and development experience and project management experience, won the outstanding mentor of RoboMaster University single competition, the first prize of Shandong New Generation Mobile Internet Innovation Application Skills Competition, and many other awards.