GTSeption: A Deep Learning EEG Emotion Recognition Model Based on Fusion of Global, Time Domain and Frequency Domain Feature Extraction

Jian Zhao  
Changchun University of Science and Technology

ZhiWei Zhang (✉ zzw19971008@163.com)  
Changchun University: Chang Chun University

Jinping Qiu  
Changchun University: Chang Chun University

Lijuan Shi  
Changchun University: Chang Chun University

Zhejun KUANG  
Changchun University of Science and Technology  https://orcid.org/0000-0001-5632-7596

Weipeng Jing  
Northeast Forestry University

Zhan-peng Zhu  
Jilin University

Research Article

Keywords: Electroencephalogram, Convolution neural network, Emotion recognition, Deep learning

Posted Date: December 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1085276/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
GTSception: A deep learning EEG emotion recognition model based on fusion of global, time domain and frequency domain feature extraction

Jian ZHAO\textsuperscript{1,3}, Zhiwei ZHANG\textsuperscript{1,2}, Jinping QIU\textsuperscript{2,3}, Lijuan SHI\textsuperscript{2,3}, Zhejun KUANG\textsuperscript{1,3*}, Weipeng JING\textsuperscript{4} and Zhan peng ZHU\textsuperscript{5}

\textsuperscript{1*}College of Computer Science and Technology, Changchun University, Changchun, 130022, China.
\textsuperscript{2}College of Electronic Information Engineering, Changchun University, Changchun, 130022, China.
\textsuperscript{3}Jilin Provincial Key Laboratory of Human Health Status Identification and Function Enhancement, Changchun, 130022, China.
\textsuperscript{4}College of Information and Computer Engineering, Northeast Forestry University, Harbin, China.
\textsuperscript{5}Department of Neurosurgery, The First Hospital of Jilin University, Jilin, China.

*Corresponding author(s). E-mail(s): kuangzhejun@ccu.edu.cn;

Abstract

With the rapid development of deep learning in recent years, automatic electroencephalography (EEG) emotion recognition has been widely concerned. At present, most deep learning methods do not normalize EEG data properly and do not fully extract the features of time and frequency domain, which will affect the accuracy of EEG emotion recognition. To solve these problems, we propose GTSception, a deep learning EEG emotion recognition model. In pre-processing, the EEG time slicing data including channels were pre-processed. In our model, global convolution kernels are used to extract overall semantic features, followed by three kinds of temporal convolution kernels representing different emotional
2 GTSception

periods, followed by two kinds of spatial convolution kernels highlighting brain hemispheric differences to extract spatial features, and finally emotions are dichotomy classified by the full connected layer. The experiments is based on the DEAP dataset, and our model can effectively normalize the data and fully extract features. For Arousal, ours is 8.76% higher than the current optimal emotion recognition model based on Inception. For Valence, the best accuracy of our model reaches 91.51%.

Keywords: Electroencephalogram, Convolution neural network, Emotion recognition, Deep learning

1 Introduction

Being an essential way for people to communicate and express, emotion is a kind of people’s attitudinal responses and corresponding behavioral responses to objective things. With the wide application of human-computer interaction [1], emotion recognition has been widely studied by researchers. In order to analyze people’s emotions well, researchers need to collect and recognize the emotions expressed in speech [2, 3] facial expressions [4] peripheral physiological signals [5] and behavioral actions [6]. Because the emotional expression of facial expressions has strong subjectivity and camouflage, the emotion recognition through them cannot continuously feed back a reliable emotional state to ensure its authenticity. Physiological signals [7] are controlled by the autonomic nervous system of human beings, which can show related emotions for a long time more accurately and truly, compared with other emotional modes. In particular, EEG signals of physiological signals are generated by human cerebral cortical neurons and can truly reflect people’s emotional state stimulated by external stimuli. Therefore, it is relatively reliable to identify emotions through EEG [8].

The traditional machine learning method of EEG emotion recognition requires manual feature extraction of time domain (Zheng [9]; Reuderink [10]), frequency domain (Gao [11]; Wang [12]) time-frequency domain (Dang [13]; Zhang [14]) and others, and then establish a machine learning model to identify emotions. On the other hand, deep learning algorithm can automatically extract high-dimensional features and classify them by using multi-layer neural network due to its good nonlinear learning ability [15]. Most researches only consider one of the features in the time-domain and frequency-domain, which will lead to insufficient feature extraction of EEG signal and the problem of low recognition accuracy. Only a few studies achieve high accuracy by fully combining time-domain and frequency-domain features [16]. For image and signal data, good data pre-processing can effectively reduce noise data, thus effectively improving accuracy [8]. The emergence of TScpction [17] solves the above problems, which effectively combines features in the time-domain and
frequency-domain. However, it is not perfect in global feature extraction and data pre-processing, whose recognition accuracy needs to be further improved.

In order to solve these problems, we propose a deep learning EEG emotion recognition model that fully extracts time-domain and frequency-domain features and has good data pre-processing effect. The advantages are as follows:

(1) Inspired by Inception, GTSception uses decomposed temporal and spatial convolutional kernels to fully extract time-domain and frequency-domain features. At the same time, GTSception additionally uses global convolution kernels to extract overall semantic features.

(2) The two normalization methods, Z-Score and Min-Max, are used for preprocessing comparison experiments on channels and time, because appropriate data normalization can effectively improve the accuracy of emotion recognition.

The rest of the article consists of the following parts: The second section describes the related work. The third section describes our methods in details. The fourth section describes the experiments and the results. Conclusions will be presented on the final section.

2 Relevant work

2.1 Inception structure

In ILSVRC2014, Google won with GoogleNet (Inception-V1)[18]. It uses multiple convolution kernels to extract multiscale features from images, and also used 1x1 convolution kernels to reduce the number of feature channels, which could ensure that the model could not overfit in the case of fully extracting features. In order to further reduce the number of parameters and the calculation cost, Inception v2[19] and Inception v3[20] have added operations such as decomposing convolution kernel (asymmetric convolution kernel, large scale convolution kernel to small scale), regularization (Dropout, Batch Normalization) and others to GoogleNet networks. Inception v4[21], proposed in 2016, ensures high accuracy and network depth by adding the ResNet module.

2.2 Application of EEG emotion recognition in artificial intelligence

In the research on EEG emotion recognition based on machine learning, Candra[22] showed that the recognition accuracy of Valence and Arousal in DEAP dataset was 65.13% and 65.33% by using time-frequency feature wavelet entropy; Ning[23] obtained that of Valence and Arousal in DEAP dataset was 69.1% and 71.9% by using frequency domain feature empirical mode decomposition (EMD). However, most of these methods extract only one of these features, which is a low-dimensional feature that contains insufficient electroencephalographic information to reliably identify emotions.

In the research of deep learning, there are mainly two kinds of methods: convolution neural network and recurrent neural network. Yanagimoto[24] could
4  

GTSception

achieve the accuracy of 81.85%, only adopting the relatively shallow CNN of Valence in DEAP. Rui[25] showed that the accuracy of Valence in the DEAP data could reach 87.27%, using a better input format, effective EEG channel filtering, and CNN. The researchers mentioned above mainly use the CNN, but it focuses more on the correlation between EEG channels. Alhagry[26] used the 2-layer LSTM to achieve the accuracy of 85.65% and 85.45% respectively in the Arousal and Valence for the DEAP dataset. LSTM is a classical network structure of recurrent neural network, which focuses more on the temporal timing of EEG signals.

In this paper, GTSception will adopt the global, temporal, spatial convolution kernels to effectively extract the relevant features, and our model achieves better performance in the DEAP dataset compared to TSception.

3 Method

3.1 GTSception EEG emotion recognition framework

The framework proposed in this paper is based on deep learning and applied to EEG emotion recognition. Each sub-module of GTSception is shown in Figure 1.

![Fig. 1 GTSception EEG Emotion Recognition Framework](image)

First, the model needs to collect or gather data related to EEG emotions. DEAP dataset from the public database is used[27] in our model. Secondly, in order to achieve faster convergence of our model, good data pre-processing methods are adopted. Z-Score and Min-Max, two excellent normalization methods, are selected. Then, the feature extraction was performed on the preprocessed dataset through the global convolution module, the temporal convolution module and the spatial domain convolution module. Finally, the classifiers analyzed the extracted features to classify emotions of Valence and Arousal.

The structure of GTSception is shown in Figure 2, which is mainly divided into four parts: global convolution module, temporal convolution module, spatial convolution module and classifier module.
3.2 Data pre-processing

The original DEAP EEG data is multidimensional signal data including total number of experiments, EEG channels, time sequence signals, etc. In order to adapt to the training of GTSception, it needs to be transformed into two-dimensional data before it can be applied to the convolution module in our model. Algorithm 1 is the pre-processing algorithm for our model, which preforms well by using DEAP dataset.

**Algorithm 1 Data Preprocessing.**

**Input:** Original Data D\((\text{shape} = (32,40,40,8064))\); Sampel Rate \(\text{SR} = 128\); Subejects S\((\text{shape} = (32,))\); Channels C\((\text{shape} = (32,))\); Trials I\((\text{shape} = (40,))\); Times T\((\text{shape} = (60,))\); Baseline1 = \(D_{s,c,i,t}[0 : \text{SR} \ast 1]\), b2 = \(D_{s,c,i,t}[\text{SR} \ast 1 : \text{SR} \ast 2]\), b3 = \(D_{s,c,i,t}[\text{SR} \ast 2 : \text{SR} \ast 3]\).

**Output:** \(Z\) – Score: \(D' = Z(x) = \frac{x - x_{mean}}{x_{std}}\), Min – Max: \(D' = M(x) = \frac{x - x_{min}}{x_{max} - x_{min}}\), \(x = D_{s,c,i,t}\).

1: for \(s = 1,2,\ldots,S\) do
2:  for \(c = 1,2,\ldots,C\) do
3:    for \(i = 1,2,\ldots,I\) do
4:      for \(t = 1,2,\ldots,T\) do
5:        \(Y = x - \text{Ave}; (\text{Ave} = \frac{b1 + b2 + b3}{3})\)
6:        \(D' = Z(Y) OR M(Y)\)
7:      end for
8:    end for
9:  end for
10: end for

In algorithm 1, D represents the original DEAP dataset, SR represents the sampling rate, S represents the number of subjects, C represents the number of EEG electrodes, I represents the number of experiments, T represents the
6 **GTSception**

number of samples in each experiment, B1-B3 represents the baseline data, and Ave represents the baseline data after averaged in seconds.

According to algorithm 1, X is the two-dimensional EEG data obtained by the subject, the channel, the experiment and the time slice layer by layer, and then the average value of the first 3 seconds is subtracted to obtain the pure EEG data shown in Figure 3. In order to minimize the impact of noise data on the convergence of GTSception, we use two normalization methods, Z-Score and Min-Max, to conduct a comparative experiment and select the optimal normalization scheme.

Fig. 3 Pre-processed EEG data

Z-Score standardization uses the standard deviation and mean of the sequence data to standardize the data. To ensure the consistent overall distribution of the standardized data, the data of 0 in EEG data will be ignored. Z-Score standardized formula is shown in Formula 1:

$$
\hat{X}_i = \frac{X_i - Mean(X_i)}{Std(X_i)} \quad (1)
$$

Min-Max normalization uses the maximum and minimum values of sequence data to standardize data, whose formula is shown in Formula 2:

$$
\hat{X}_i = \frac{X_i - min(X_i)}{max(X_i) - min(X_i)} \quad (2)
$$

3.3 Global Convolution

Due to the improvement of hardware devices and the widespread application of image recognition, the convolutional neural networks are gradually developing and maturing. The Inception structure of GoogleNet is proposed for image recognition, and the model is also based on the implementation of CNN, so we also use a similar input format for images(length-width-channel).

The first part is the global convolution module, which adopts the global convolution kernel of 3X3 to extract the overall semantic features. Compared to TSception, the additional features extracted here provide richer details of the global EEG features in the latter part of the model.
3.4 Temporal convolution

The second part is the time-domain feature extraction module. The input data is a time sequence signal with a sampling rate of 128. In order to better integrate the EEG emotion features at different time periods, three temporal convolution kernels of 64x1, 32x1 and 16x1 were used to extract the feature layers containing different time-domain information, which respectively represent the relevant time-domain features of EEG data of 0.5s, 0.25s and 0.125s. Then, the features are average pooled using a 1x16 pooling kernel so that the features will be more compact and the computation will be smaller. To extract more complete and rich information later, the three feature layers will be merged into a feature layer with a size of 32x17x16. Finally, batch normalization is performed on the new feature layer, so that the data distribution is fixed and the model can converge better.

3.5 Spatial Convolution

The third part is the feature extraction module of frequency-domain. In the former module, the time-domain features have been fully extracted. Referring to the input of the model, 32 is the number of brain electrodes, i.e., the length of the input data. Due to the asymmetry of human brain, it will adopt 1x32 and 1x16 spatial domain convolution kernels to extract the spatial domain features of EEG signals. The features extracted by 1x32 convolution kernel takes into account the spatial domain features brought by all brain electrodes, while the frequency-domain features extracted by 1x16 convolution kernel represent those of two brain hemispheres. Subsequently, 1x8 pooling kernel is used to simplify the features of these two feature layers. Finally, the two feature layers are combined to form a new feature layer with a size of 3x2x64, where the batch normalization operation is also performed in consideration of model convergence.

3.6 Classifier

The last part is the classifier. EEG data has been processed into a feature layer with a size of 3x2x64 by former modules, which needs to be converted into a fully connected layer of size n. In order to reduce feature loss, we connect the fully connected layer consisting of 1024 neurons before classification, and Dropout is also used. Finally, the full connection layer of two neurons is used for classification.

The modules of our model are described above. Algorithm 2 contains the training Procedure of GTSception.
8  GTSception

Algorithm 2  Training Procedure for GTSception.

**Input:** Epoch E; Batch B; Classes C; Dropout D; Model GTSception.

**Output:** Initialization: ACC = 0, Loss = 0;

1: Train:
2: for e = 1, 2, ..., E do
3:     for i = 1, 2, ..., \( \frac{Y_{tr}}{B} \) do
4:         \( \text{Loss}_{train} = \text{GTSception}(x_{train}) \)
5:     end for
6: end for
7: Saved Model SAM;
8: Test:
9: for i = 1, 2, ..., \( \frac{Y_{te}}{B} \) do
10:    \( \text{ACC}_{test} = \text{SAM}(x_{test}) \)
11: end for

\( x_{train} \) represents training data, \( y_{train} \) represents training labels, \( x_{test} \) represents test data, \( y_{test} \) represents test labels, E represents the iteration cycle, B indicates batch size, C indicates emotion category, and Dropout represents the random discard rate of neurons. The model will be iterated for E cycles with \( y_{train}/B \) group samples, and the model will be saved after completion of the iterations. Finally, the saved model SAM will be used to test the model performance using \( x_{test} \).

4  Experiment and result analysis

4.1  Data

The experimental data used in this paper are from the multi-modal EEG emotion dataset of Queen Mary University of London, United Kingdom. It contains multi-modal emotional information such as video data of the subjects’ face, EEG, ECG, EMG, etc. Thirty-two participants were visually stimulated by 40 carefully selected one-minute music video clips to evoke relevant emotions. Each subject will then score the video in multiple emotional dimensions according to a score of 1-9. The length of each experiment in DEAP data was 63 seconds, with the first 3 seconds being the experiment wait time and the last 60 seconds being the experiment time while watching the video. The data has been sampled down to 128HZ, and the EOG signal artifact has been removed.

The detailed description of the DEAP dataset is shown in Table 1.

The DEAP dataset contains multiple dimensions of emotion such as pleasure, arousal, dominance, and liking, and they are not all widely used for emotion recognition. The two-dimensional Arousal-Valence emotion model proposed by emotionologist Russell[28] can combine multiple categories of emotions, and most researchers have studied both Arousal and Valence. Referring to Yang’s[29] method, this paper takes 5 as the threshold for the emotion
classification, which will be divided into HV/LV and HA/LA. The emotion classification is shown in Table 2.

| Labels  | High | Low  |
|---------|------|------|
| Valence | ≥ 5 HV | < 5 LV |
| Arousal | ≥ 5 HA | < 5 LA |

### 4.2 Experiment settings

Considering that the strong individual specificity of EEG emotion, the model will be trained and tested on a single subject. To optimize model performance, there are mainly the following aspects: in order to prevent the overfitting of the model, dropout randomly drops small neurons with a ratio of 0.5 in the full connection layer. Then, a learning rate of size 2e-4 as well as exponential decay of the learning rate decay method was used. The activation function is ELU, which is a good way to avoid the gradient disappearance problem that occurs when using ReLU. The Batch Normalization ensures that the feature layers before convolution have the same distribution and accelerates the convergence of models. Random truncation of standard normal distribution for weight parameter is adopted, which is similar to BN operation to accelerate model convergence and accelerates weight updating. In this paper, batch size is 80, the epoch is 200, and the model deployment is carried out on the GTX1070 8G Windows environment.

### 4.3 Experimental scheme

In this paper, the hyperparameters of the model such as batch size, epoch, learning rate, training set and test set size, and dropout rate are kept consistent so that we can use the control variable method to design accurate comparison experiments.
GTSception

For the comparison experiment of the model improvement, in order to maximize the effect of the model, it was found that the Valence emotion dimensions were more balanced than that of positive and negative samples of total Arousal data. Therefore, only the Valence category was used to carry out the comparison experiment. To that the experiment is more reliable, all subjects will be tested on the Valence for EEG recognition.

For the pre-processing part, because there was no relevant research on the Arousal in TSception, the comparison experiment was made specifically on the Arousal emotion dimension of dimension in the pre-processing. In the analysis of emotional categories of each subject, positive and negative samples of subjects S16 and S32 were the most balanced. For this reason, four different pre-processing schemes are designed in this paper.

Scheme 1, the Z-Score method was used to standardize the data containing a single sample point for the whole EEG channel; Scheme 2, the Z-Score method was used to standardize the data containing a single sample point for the whole EEG channel; Scheme 3, the Min-Max method was used to standardize the data containing a single sample point for the whole EEG channel; Scheme 4, the Min-Max method was used to standardize the data containing a single sample point for the whole EEG channel. By comparing four different pre-processing schemes, the optimal pre-processing scheme is selected to cooperate with the improved model in this paper to achieve the optimal recognition accuracy.

4.4 Analysis of experimental results

Table 3 shows the comparison results between the recognition accuracy of TSception and our model in Valence.

| Sub | Orig | Ours | Sub | Orig | Ours |
|-----|------|------|-----|------|------|
| 1   | 90.86% | 94.15% | 17 | 75.90% | 82.31% |
| 2   | 81.64% | 81.81% | 18 | 86.45% | 92.54% |
| 3   | 88.52% | 95.60% | 19 | 85.44% | 93.34% |
| 4   | 83.44% | 84.31% | 20 | 85.91% | 92.19% |
| 5   | 87.64% | 91.16% | 21 | 81.90% | 91.40% |
| 6   | 75.00% | 84.61% | 22 | 80.63% | 87.27% |
| 7   | 85.15% | 91.50% | 23 | 84.27% | 93.40% |
| 8   | 86.91% | 90.22% | 24 | 84.10% | 92.62% |
| 9   | 89.38% | 89.98% | 25 | 88.19% | 93.32% |
| 10  | 93.32% | 93.67% | 26 | 88.84% | 90.70% |
| 11  | 79.58% | 81.64% | 27 | 75.00% | 92.23% |
| 12  | 88.81% | 89.94% | 28 | 81.07% | 85.17% |
| 13  | 86.33% | 87.87% | 29 | 90.06% | 91.94% |
| 14  | 83.93% | 90.92% | 30 | 88.72% | 93.22% |
| 15  | 85.95% | 95.78% | 31 | 89.80% | 90.33% |
| 16  | 88.28% | 93.95% | 32 | 89.05% | 90.00% |
As shown in Table 3, compared with TSception, the full convolutional part of our model can bring more global feature details, and the results in the table also well verify the effect of our model. Ours indicates the accuracy of our model, Orig represents the accuracy of the TSception. Of the 32 subjects, most of the participants reported that the accuracy of our model was about 4% higher than TSception’s. However, the 27th subject showed the most significant difference, with a nearly 15% difference in accuracy, which could be attributed to the inconsistency of positive and negative samples from the 27th subject.

Figure 4 and Figure 5 show the performances of S16 and S32 in four data pre-processing schemes, respectively.

![Fig. 4 Accuracy variation of S16](image)

![Fig. 5 Accuracy variation of S32](image)

In Figure 5, it can be seen that for subject S32, the Z-Score and Min-Max methods performed almost identically in the model after preprocessing the EEG channel data due to the strong individual specificity of EEG.

In Figure 4 and Figure 5, the results of the Z-Score pre-processing method for channel and time dimensions are similar to those of the Min-Max method. To verify the robustness of our model, we observe the changes of the accuracy and the loss through a ten-fold cross-validation. The total data for individuals in DEAP would be divided into training set and test set in a ratio of 9:1.
All samples will be scrambled by non-return sampling. The results of cross-validation of subjects S16 and S32 using the Z-Score method with time and channel preprocessing and training of 200 epoch are shown in Table 4.

**Table 4** Comparison of cross-validation results for subjects S16 and S32.

| Person Type tr/ch | tr/ch/acc | te/ch/acc | tr/ch/loss | te/ch/loss |
|------------------|-----------|-----------|------------|------------|
| s16              | 98.22%    | 94.79%    | 0.3456     | 0.3762     |
| s32              | 96.29%    | 91.51%    | 0.3650     | 0.4025     |

| Person Type tr/T | tr/T/acc | te/T/acc | tr/T/loss | te/T/loss |
|-----------------|----------|----------|-----------|-----------|
| s16             | 86.65%   | 76.02%   | 0.4654    | 0.5384    |
| s32             | 88.96%   | 79.79%   | 0.4447    | 0.5114    |

From table 4, it is obvious that cross-validation is a good way to improve the robustness and accuracy of the model. According to the above chart, the Z-Score method of EEG channel data pre-processing ensures that the model improves the accuracy of emotion recognition steadily during the training process.

## 5 Conclusion

In this paper, we proposed GTSception, a deep learning model for EEG emotion recognition, which can fully extract time-domain and frequency-domain features. Our model performs well on Valence and Arousal in the DEAP dataset in two main ways: First, compared with TSception, the convolution kernel and full convolutional structure were improved, making the model adapt to DEAP dataset to achieve high accuracy. Secondly, an appropriate normalization method was proposed to make the distribution of the model input data more balanced, leading the model to converging and fitting faster. Compared with TSception, a good normalization method and full convolution module were introduced to effectively extract time-domain and frequency-domain EEG features, thus reducing the noise data of the signal, which has a greater accuracy improvement for the model.

## Statements and Declarations

This work was supported by National Natural Science Foundation of China[Grant No.61502052] and Jilin Provincial Science and Technology Department[Grant No.20200601004JC, 20200301054RQ and 20200404207YY]. Author zhao jian has received research support from National Natural Science Foundation of China and Jilin Provincial Science and Technology Department; This work was supported by Science and Technology Project of Jilin Provincial Department of Education[Grant No.20190201195JC]. Author Shi Lijuan
has received research support from Science and Technology Project of Jilin Provincial Department of Education; This work was supported by Natural Science Foundation of Jilin Province(Grant No.20210101477JC). Author Kuang Zhejun has received research support from Natural Science Foundation of Jilin Province; This work was supported by Science and Technology Development Plan of Jilin Province(Grant No.20200403120SF). Author Zhu Zhanpeng has received research support from Science and Technology Development Plan of Jilin Province. All authors read and approved the final manuscript. The authors have no relevant financial or non-financial interests to disclose.

Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in Soft Computing.

Authorship contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Zhang Zhiwei, Qiu Jinping, Shi Lijuan, Kuang Zhejun and Zhu Zhanpeng. The first draft of the manuscript was written by Zhao Jian and all authors commented on previous versions of the manuscript.

References

[1] Fiorini L, Mancioppi G, Semeraro F, et al. Unsupervised emotional state classification through physiological parameters for social robotics applications[J]. Knowledge-Based Systems, 2020, 190: 105217.

[2] Chen M, He X, Yang J, et al. 3-D convolutional recurrent neural networks with attention model for speech emotion recognition[J]. IEEE Signal Processing Letters, 2018, 25(10): 1440-1444.

[3] Liu Z T, Xie Q, Wu M, et al. Speech emotion recognition based on an improved brain emotion learning model[J]. Neurocomputing, 2018, 309: 145-156.

[4] Huang Y, Yang J, Liao P, et al. Fusion of facial expressions and EEG for multimodal emotion recognition[J]. Computational intelligence and neuroscience, 2017, 2017.

[5] Wang X W, Nie D, Lu B L. Emotional state classification from EEG data using machine learning approach[J]. Neurocomputing, 2014, 129: 94-106.

[6] Ross P D, Polson L, Grosbras M H. Developmental changes in emotion recognition from full-light and point-light displays of body movement[J]. 2012.
[7] Al-Nafjan A, Hosny M, Al-Ohali Y, et al. Review and classification of emotion recognition based on EEG brain-computer interface system research: a systematic review[J]. Applied Sciences, 2017, 7(12): 1239.

[8] Cui H, Liu A, Zhang X, et al. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network[J]. Knowledge-Based Systems, 2020, 205: 106243.

[9] Zheng W L, Zhu J Y, Lu B L. Identifying stable patterns over time for emotion recognition from EEG[J]. IEEE Transactions on Affective Computing, 2017, 10(3): 417-429.

[10] Reuderink B, Mühl C, Poel M. Valence, arousal and dominance in the EEG during game play[J]. International journal of autonomous and adaptive communications systems, 2013, 6(1): 45-62.

[11] Gao Z, Cui X, Wan W, et al. Recognition of emotional states using multiscale information analysis of high frequency EEG oscillations[J]. Entropy, 2019, 21(6): 609.

[12] Wang X W, Nie D, Lu B L. EEG-based emotion recognition using frequency domain features and support vector machines[C].International conference on neural information processing. Springer, Berlin, Heidelberg, 2011: 734-743.

[13] Zhang J, Chen P, Nichele S, et al. Emotion Recognition Using Time-frequency Analysis of EEG Signals and Machine Learning[C].2019 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2019: 404-409.

[14] Gao Z, Dang W, Liu M, et al. Classification of EEG signals on VEP-based BCI systems with broad learning[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020.

[15] Sheykhiand S, Mousavi Z, Rezaei T Y, et al. Recognizing emotions evoked by music using CNN-LSTM networks on EEG signals[J]. IEEE Access, 2020, 8: 139332-139345.

[16] Nath D, Singh M, Sethia D, et al. A comparative study of subject-dependent and subject-independent strategies for EEG-based emotion recognition using LSTM network[C].Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis, 2020: 142-147.

[17] Ding Y, Robinson N, Zeng Q, et al. Tsception: a deep learning framework for emotion detection using EEG[C].2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-7.
[18] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[19] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. International conference on machine learning. PMLR, 2015: 448-456.

[20] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.

[21] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]. Thirty-first AAAI conference on artificial intelligence. 2017.

[22] Candra H, Yuwono M, Chai R, et al. Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine[C]. 2015 37th Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015: 7250-7253.

[23] Zhuang N, Zeng Y, Tong L, et al. Emotion recognition from EEG signals using multidimensional information in EMD domain[J]. BioMed research international, 2017, 2017.

[24] Yanagimoto M, Sugimoto C. Recognition of persisting emotional valence from EEG using convolutional neural networks[C]. 2016 IEEE 9th International Workshop on Computational Intelligence and Applications (IWClA). IEEE, 2016: 27-32.

[25] Qiao R, Qing C, Zhang T, et al. A novel deep-learning based framework for multi-subject emotion recognition[C]. 2017 4th International Conference on Information, Cybernetics and Computational Social Systems (ICCSS). IEEE, 2017: 181-185.

[26] Alhagry S, Fahmy A A, El-Khoribi R A. Emotion recognition based on EEG using LSTM recurrent neural network[J]. Emotion, 2017, 8(10): 355-358.

[27] Koelstra S, Muhl C, Soleymani M, et al. Deap: A database for emotion analysis; using physiological signals[J]. IEEE transactions on affective computing, 2011, 3(1): 18-31.

[28] Russell J A, Barrett L F. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant[J]. Journal of personality and social psychology, 1999, 76(5): 805.
[29] Yang Y, Wu Q, Qiu M, et al. Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network[C]. 2018 International Joint Conference on Neural Networks. IEEE, 2018: 1-7.