

A Brain Network Analysis-Based Double Way Deep Neural Network for Emotion Recognition

Weixin Niu¹, Chao Ma¹, Xinlin Sun¹, Mengyu Li¹, and Zhongke Gao¹, *Senior Member, IEEE*

Abstract—Constructing reliable and effective models to recognize human emotional states has become an important issue in recent years. In this article, we propose a double way deep residual neural network combined with brain network analysis, which enables the classification of multiple emotional states. To begin with, we transform the emotional EEG signals into five frequency bands by wavelet transform and construct brain networks by inter-channel correlation coefficients. These brain networks are then fed into a subsequent deep neural network block which contains several modules with residual connection and enhanced by channel attention mechanism and spatial attention mechanism. In the second way of the model, we feed the emotional EEG signals directly into another deep neural network block to extract temporal features. At the end of the two ways, the features are concatenated for classification. To verify the effectiveness of our proposed model, we carried out a series of experiments to collect emotional EEG from eight subjects. The average accuracy of the proposed model on our emotional dataset is 94.57%. In addition, the evaluation results on public databases SEED and SEED-IV are 94.55% and 78.91%, respectively, demonstrating the superiority of our model in emotion recognition tasks.

Index Terms—Brain network, deep residual neural network, electroencephalogram (EEG), emotion recognition, spearman correlation coefficient.

I. INTRODUCTION

PHYSIOLOGICAL states are significant for health management and life quality improvement. Among these, emotions have a great impact on physical health and communication in our daily lives [1]. By properly recognizing emotion states, brain-computer interaction systems can be developed

Manuscript received 31 August 2022; revised 25 December 2022; accepted 9 January 2023. Date of publication 12 January 2023; date of current version 3 February 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61922062, Grant 61903270, and Grant 61873181; in part by the Innovation of Science and Technology Forward 2030 program “Brain Science and Brain-Inspired Intelligence Technology”; and in part by the National Natural Science Foundation of Tianjin, China, under Grant 21JCJQJC00130. (Corresponding author: Zhongke Gao.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Tianjin University, Tianjin, China, under Grant No. TJUE-2021-065.

The authors are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: zhongkegao@tju.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3236434

to a higher level, which can be used in many fields such as health care, nursing, entertainment, etc.

Judging emotions from external modalities has a high degree of complexity and uncertainty [2]. Recently, methods based on EEG signals have drawn a lot of attention due to the high temporal resolution and objectivity of EEG [3], [4]. Liang et al. [5] proposed a hybrid unsupervised deep convolutional recursive adversarial network-based EEG feature representation and fusion model for training and identifying emotional EEG features in an unsupervised manner. Liu et al. [6] proposed a subject-independent emotion recognition algorithm based on dynamic empirical convolutional neural networks. Jiang et al. [7] proposed a depression classification detection method using differential entropy and genetic algorithm for feature extraction and selection, and support vector machine for classification. These methods obtain temporal and spatial information from the EEG signals for emotion classification. However, due to the non-stationary feature and the high coupling between different channels of EEG signals, it is still a challenging task to detect the emotion-related information in EEG signals while maintaining the robustness of recognition under noise interference.

Considering the nonlinearity and complexity of EEG signals, the application of dynamic systems analysis methods to EEG-based emotion recognition will be potentially adaptive [8], [9]. Recently, to analysis EEG rhythms, Fourier-Bessel series expansion (FBSE) based methods [10], [11], [12], [13] have been widely applied to decompose EEG signals. Sharma et al. [14] proposed the discrete wavelet transform as decomposition method to explore the nonlinear dynamics of each sub-band signal. Nalwaya et al. [15] proposed tunable Q-factor wavelet transform to separate EEG rhythms. Das and Pachori [16] proposed a EEG rhythms segmentation method using multivariate iterative filtering. Alternatively, brain activity can also be described by constructing brain networks using EEG signals, which are built by taking electrodes (channels) as nodes and using correlation measures between channels as edges [17]. By using EEG, MEG, and fMRI to measure brain activity, Straaten and Stam [18] created functional networks for analyzing brain activity. Dang et al. [19] proposed a multilayer brain network to map the time, frequency and channel-related information from EEG signals into the multilayer network topology. The brain network method is an effective way to represent the inter-channel relationship

of EEG signals, which is important in emotion recognition task.

The significance of the attention mechanism in deep neural networks has been studied extensively in the previous literature [20], [21]. Attention not only tells where to focus, but it also improves the representation of interesting areas. Woo et al. [22] proved the efficiency of the attention mechanism used in computer vision tasks. Similarly, brain activities can be divided into many channels or regions. Each channel or region has a different intensity of response when the brain is engaged in different activities. To explore the effect of targeting specific regions and channels when analyzing brain activity, we introduce the attention mechanism into the emotion analysis model for further discussion.

Although existing works have been developed to some extent, there still remains a necessity to excavate new method for studying brain activities and to improve classification accuracy. The main contributions of this paper are as follow:

1) It proposes a novel double way feature extracting strategy to improve the utilization of temporal-frequency domain of EEG signals for emotion recognition task. The proposed method takes advantage of dynamic systems analysis by transforming raw EEG signals into brain networks. The deep learning techniques are adopted to attain feature embeddings. The outputs are combined to mitigate the information loss and to improve the classification performance.

2) In the process of dealing with the brain networks, the attention mechanism with residual connection is adopted to perform adaptive feature optimization. By applying it to the extraction of the brain networks, channels and areas that are useful for the emotion recognition are promoted, while the gradient dissipation and gradient explosion are prevented.

3) A series of experiments are conducted to collect emotional EEG data from multiple subjects. The proposed method is validated on our emotional dataset and achieves an average accuracy of 94.57% which is 1.4% higher than the highest baseline. Also, we evaluate our method on SEED and SEED-IV and it achieves 94.55% and 78.91% respectively.

4) We investigate the effect of different EEG rhythms on classification results. Moreover, a series of ablation experiments are carried out to demonstrate the necessity of the double way structure and the residual connection.

The remainder of this paper is organized as follows. Section II explains the detailed experiment protocol and the processing of EEG signals. Section III explains the brain network construction block and the double way structure. Section IV shows the the classification results on the three datasets and compares the results with other baselines. Section V makes a discussion on ablation experiments. Finally, section VI summarizes this study.

II. EXPERIMENTS

The experiments are conducted in the Institute of Artificial Intelligence and Network Science of Tianjin University. The experiment process has been approved by the ethics committee of Tianjin University, Tianjin, China. A multi-channel electrode cap is used for data acquisition. The electrode distribution is in accordance with the 10-20 international standard.

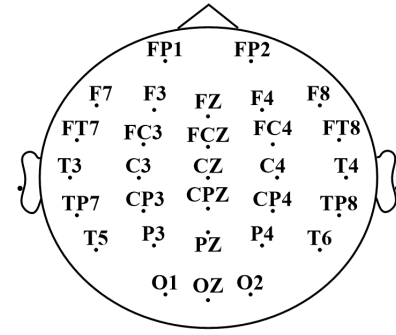


Fig. 1. Distribution of electrodes in international standard 10-20.

TABLE I
SOURCES OF THE MOVIE CLIPS AND THE
CORRESPONDING EMOTION LABELS

No.	Sources	Labels	Duration
1	Miracle in Cell No.7	sad	9 min 55 s
2	Dearest	sad	4 min 30 s
3	A Hero or Not	happy	5 min 30 s
4	Mr. Bean	happy	7 min 20 s
5	Mr. Bean	happy	5 min 52 s
6	Dead Silence	fear	8 min 24 s
7	The Conjuring	fear	4 min
8	Lights Out	fear	3 min 25 s

All the 30 electrodes used for EEG acquisition are shown in Fig. 1. With two posterior ear mastoid electrodes as reference, the number of channels is equal to 30 which is corresponding to the electrodes. In addition to the 30 electrodes, there are also 4 Electrooculogram (EOG) electrodes; NuAmps amplifier is used to amplify the collected data; A computer equipped with Scan4.5 is used to store and visualize enlarged EEG data, which can help to determine whether the electrodes are working properly.

We recruited eight right-handed healthy subjects (6 males and 2 females, aged 22-24 years). Before the experiment, participants were asked to remain calm and to avoid fatigue to ensure that their emotions were properly aroused. In addition, each subject was provided with detailed instructions during the experiment process to reduce deviation in the EEG signals.

The specific stimuli used to arouse the happy and sad emotions of the subjects were eight carefully selected movie clips, as shown in Table I. The participants were asked to watch these movie clips in an independent and undistracting environment to evoke corresponding emotional responses. According to the requirements of the experimental design, the subjects were asked to fill in the questionnaire during the interval between the movie clips to cooperate with the investigation of truly evoked emotional types and emotional intensity. At the end of each video clip, the computer went black for a period to calm the subjects down from the emotionally aroused state so that the awakening of the next emotional state was not interfered with.

The subjects were guided by the instructions on the screen throughout the experiment to ensure that the subjects were not disturbed by other factors. The experimental procedure were divided into four steps. In the preparation stage of emotional



Fig. 2. Realistic scene of experiment.

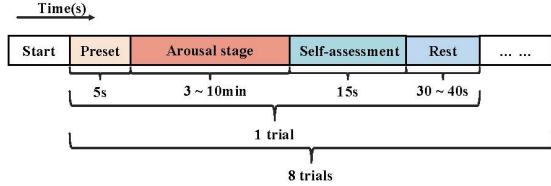


Fig. 3. Timing diagram of experiment.

arousal, the preset preparation time was 5s, and instruction on the screen was used to prompt the subjects to focus on the screen before the movie clips began. In the emotional arousal stage, the program randomly selected a movie clip to play to arouse the corresponding emotional state. In the questionnaire stage, the screen provided a simple survey to obtain the subjects' assessment of emotional arousal. In the emotional recovery stage, the subjects chose a recovery time by themselves, and then entered the next emotional arousal stage after complete recovery. The experiment scenario and the timing diagram are shown in Fig. 2 and Fig. 3 respectively.

After the EEG signals were obtained from the experiment, we down-sampled the data from 1000Hz to 250Hz to reduce the input dimension so that improve the efficiency of the analysis process. Then the processed data were filtered to 1-50Hz using a bandpass filter to reduce noise, artifact interference and some irrelevant components. The independent component analysis (ICA) [23] was used to eliminate blink artifacts. For the upper limit of the time window, directly using an entire trial as an input will cause a huge memory burden. For the lower limit of the time window, Ouyang et al. [24] demonstrated that a time window of 2-s can improve the classification accuracy using power spectral density (PSD) features and differential entropy (DE) features. Here we divide data into 5-s non-overlapping to seek a balance between accuracy and computation burden after preliminary experiments. Finally, our emotional dataset is obtained with 510 samples in total and 170 samples for each category.

III. METHODOLOGY

A. Overall Architecture

In this section, we will introduce the implementation of the whole method. The schematic of our proposed method is shown in Fig. 4. The whole model is divided into two ways, and both are fed with raw EEG signals. The first way begins

with a brain network construction block which establishes brain network matrices from the raw EEG signals. Then the matrices are fed into a deep convolutional neural network with several successively connected residual modules. In the second way the raw EEG signals are directly fed into a deep neural network block. Finally, the outputs of the two ways are concatenate and merged to obtain emotional categories.

B. Brain Network Construction Block

The brain network is employed to show correlations of EEG signals gathered in different areas of human brains [18]. To reveal the correlations on different frequency bands, we use the continuous wavelet transform (CWT) [25] to decompose the EEG signals into a superposition of a series of wavelets. Compared to the short-time fourier transform (STFT), CWT is able to attain higher frequency resolution in the low-frequency region and a higher time resolution in the high-frequency region. For any energy-limited signal $x(t)$, the CWT of $x(t)$ can be represented as:

$$\begin{aligned} W_f x(a, b) &= \{x(t), \Psi_{a,b}(t)\} \\ &= \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \Psi^* \left(\frac{t-b}{a} \right) dt, \quad (a > 0) \end{aligned} \quad (1)$$

where a and b denote the scale factor and the offset respectively. $\Psi^*(t)$ denotes the complex conjugate of mother wavelet $\Psi_{a,b}(t)$. $W_f x(a, b)$ is the wavelet coefficients of EEG signals. In this work, we choose Morlet wavelet as the mother wavelet because it can preserve the time and frequency resolution [26]. The Morlet wavelet can be represented as [27]:

$$\Psi(x) = e^{-\frac{x^2}{2}} \cos(5t) \quad (2)$$

To further analyse the time-frequency representation obtained by CWT, frequency should be determined by the scale factor a . The relationship between the scale factor a and the frequency can be expressed as follows [28]:

$$f = \frac{f_c}{a \times \Delta} \quad (3)$$

where f is the frequency corresponding to scale factor a , f_c is the center frequency of frequency window of $\Psi(x)$, and Δ is the sampling period of the EEG signals.

By applying CWT to the raw EEG signals, we get a wavelet time-frequency diagram of each EEG channel. Based on different EEG rhythms [29] (δ (1-3 Hz), θ (4-7 Hz), α (8-13 Hz), β (14-30 Hz), and γ (31-50 Hz)), the wavelet coefficients can be divided into five frequency bands. For each EEG rhythm, the wavelet coefficients can be aggregated according to the boundaries of each frequency band by the following equation [30]:

$$W_{fave}(b) = \sum_{a=r_A}^{r_B} |W_f x(a, b)| \quad (4)$$

where $[r_A, r_B]$ denotes the boundary of different EEG rhythms and $W_{fave}(b)$ represents a aggregated coefficients sequence at the corresponding EEG rhythm.

To construct the brain networks in each frequency band, the channels of EEG signals are deemed as the nodes and

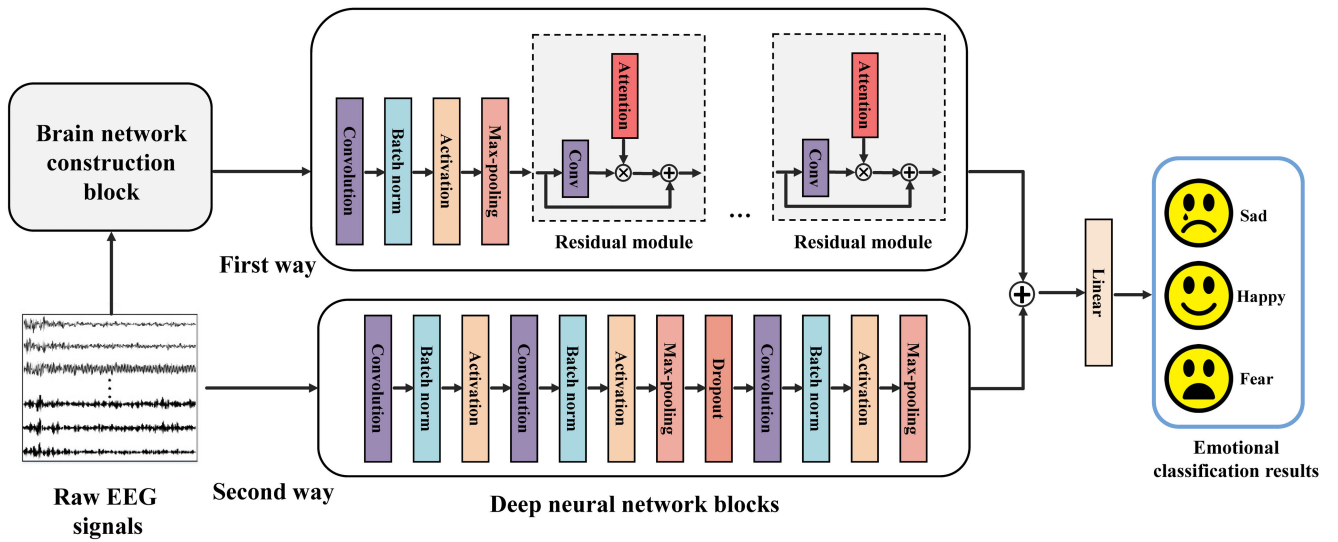


Fig. 4. Overall of the proposed double way model. The upper part is the first way of the model which transforms the EEG signals by the brain network construction block. The brain networks are then fed into the subsequent deep neural network block. The lower part is the second way of the model which comprises the other deep neural network block. The output features of the two ways are fused and fed into a linear layer to attain the classification results in the rightmost part.

the Spearman correlation coefficients between channels are applied to decide connecting edges of the brain networks. Spearman correlation coefficient [31] is a kind of non-parametric index that can measure the monotonic relationship between two variables. It is calculated according to the sorting position of the original data and is not sensitive to data errors. Therefore, the calculation of the Spearman correlation coefficient between the channels of EEG signals will be robust against the noise in EEG signals.

For two aggregated coefficients sequences of two EEG channels within same frequency band, they are first sorted according to the same sorting rules to obtain two ranking sets X and Y . A difference set D of ranked elements can be obtained by subtracting X and Y in element-wise order, in which the k^{th} element in D is calculated as [30]:

$$d_k = x_k - y_k \quad (5)$$

Then the Spearman correlation coefficient ρ can be obtained by the following equation [30]:

$$\rho = 1 - \frac{6 \sum d_k^2}{N(N^2 - 1)} \quad (6)$$

where N denotes the length of the input sequence. The obtained Spearman correlation coefficients range from $[-1, 1]$. The coefficient increases in absolute value as the correlation between variables increases and the symbol of the coefficient indicates the direction of correlation.

The initially constructed brain network is fully connected, and a threshold value needs to be set to convert it into a sparsely connected brain network. For each brain network, 50% is set as sparsity threshold, that is, half of the edges with lower absolute values are removed according to [32]. Finally, the brain network matrix in a shape of 30×30 is obtained, and the element in the i^{th} row and the j^{th} column denotes the

Spearman correlation coefficient between the i^{th} channel and the j^{th} channel.

C. Deep Neural Network Blocks

For the first way of the model, raw EEG signals are mapped to the brain network matrices $X \in R^{N \times C \times C}$ by the brain network construction block, where N represents the frequency band and C represents EEG channels. The matrices are then input into the subsequent deep neural network block. The deep neural network block starts with a convolutional layer. The calculation of the convolutional layer can be expressed by the following equation [33]:

$$X_j = f \left(\sum_{i \in M_j} X_i * k_{ij} + b_i \right) \quad (7)$$

where X_i is the i_{th} feature map, k is the weight matrix, and b is the deviation. The rectified linear units (ReLU) [34] is adopted as the activation function f of the convolutional layer. The convolutional layer is followed by a batch normalization layer [35]. Then an activation layer is introduced which also uses the ReLU function. Then we adopt a max-pooling layer speed up training and prevent over-fitting [36].

After completing the former operations, the processed features are fed into multiple successively connected residual modules. Each residual module has a residual connection that can effectively alleviate the problems of gradient dissipation and gradient explosion [37]. The schematic diagram of a residual module is shown in Fig. 5. We introduce both channel attention mechanism and spatial attention mechanism into the modules.

The main idea of the channel attention mechanism [22] is to use another neural network to obtain the importance of each channel in the feature maps and use the importance to assign weights to make the neural network focus

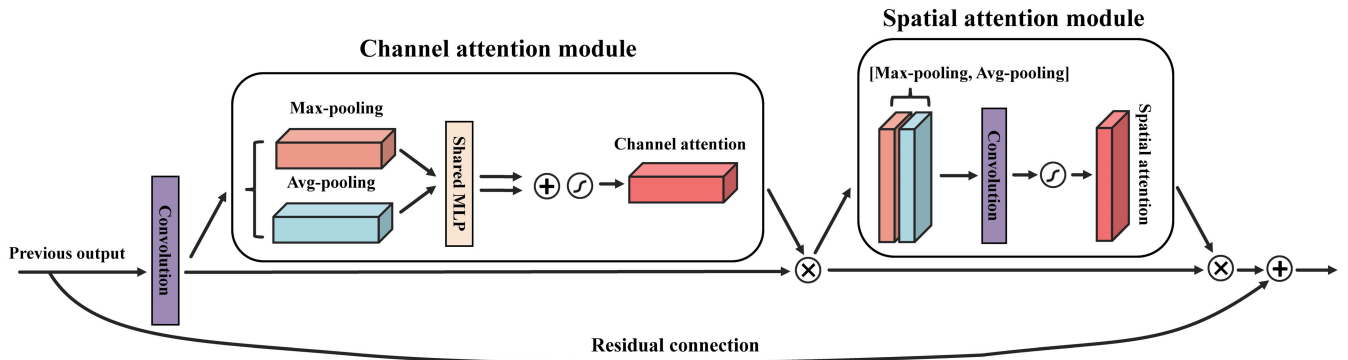


Fig. 5. Structure of a residual module in the first way of the model. The output of the previous is fed into several convolutional layers for preprocessing. The channel attention module extracts the channel attention which is then applied by element-wise product with the input features. Subsequently, the spatial attention module extracts the spatial attention and applies it to the features. A residual connection is set on the outermost to retain the information of the previous output.

on some feature channels. Specifically, the channel attention mechanism first needs to use both average-pooling and max-pooling operations to aggregate the spatial information of the feature maps. Here, the pooling operations are adaptive to input feature map and the size of the output is set to 1×1 . Then we calculate the correlation between channels through a shared multi-layer perceptron in which the dimension of output weights is the same as the number of channels in the input feature maps. Finally, the two output weights are summed to obtain a channel attention map which is normalized to 0-1. In short, channel attention can be expressed as the following formula [38]:

$$M_c = \sigma \left(W_1 \left(W_0 \left(F_{\text{avg}}^c(X) \right) \right) \right) + \sigma \left(W_1 \left(W_0 \left(F_{\text{max}}^c(X) \right) \right) \right) \quad (8)$$

where σ is the sigmoid activation function, W_1 and W_0 is the MLP weights, F_{avg}^c and F_{max}^c denotes the average-pooling operation and max-pooling operation for channel attention respectively, and X is the input feature map.

Compared to the channel attention mechanism, the spatial attention mechanism [22] assigns weights to spatial information, which is complementary to channel attention. Specifically, the spatial attention mechanism first aggregates channel information by applying average-pooling and max-pooling operations along the channel axis to get two weight maps. The pooling operations here are also adaptive which automatically fit and squeeze the input channel dimension into one. Then we use the convolution operation on the two stacked maps. The output map is normalized to 0-1. The calculation of spatial attention is shown as following equation [22]:

$$M_s = \sigma \left(F_{\text{conv}} \left(\left[F_{\text{avg}}^s(X); F_{\text{max}}^s(X) \right] \right) \right) \quad (9)$$

where σ is the sigmoid activation function, F_{conv} denotes the convolutional operation, F_{avg}^s and F_{max}^s denotes the average-pooling operation and max-pooling operation for spatial attention respectively, and X is the input feature map.

The second deep neural network block is adopted to gather the temporal features from raw EEG signals at a more comprehensive and adequate level. Inspired by the EEGNet [39], we adopt a deep neural network to enhance the performance.

TABLE II
PARAMETERS OF THE DEEP NEURAL NETWORK BLOCKS

Blocks	Layers	Kernel
	Convolution	7×7, 64
	Max-pooling	3×3, stride=2
	<i>For one residual module(total=3)</i>	
Block 1	Convolution	3×3, 64
	Convolution	3×3, 64
	<i>For channel attention</i>	
	Shared MLP	64
	<i>For spatial attention</i>	
	Convolution	7×7, 1, padding=3
	Convolution	1×100, 8
Block 2	Depthwise Convolution	30×1, 16, groups=2
	Max-pooling	1×8, stride=1
	Separable Convolution	1×16, 16, padding=(0, 8)
	Max-pooling	1×8, stride=1

Specifically, three convolutional layers are applied in the second deep neural network block. The first convolutional layer is calculated along channels to aggregate channel-wise information. The second convolutional layer is a depthwise convolution which doesn't connect to all the previous feature maps thus can learn a unique spatial filter for each temporal filter. And the separable convolution is chosen for the third convolutional layer which can synthesize the kernels of each feature map in the time dimension. After every convolutional layer there are batch normalization layer and activation layer. The max-pooling layer and the dropout layer adopted here can reduce the trainable parameters and alleviate overfitting.

Finally, a feature fusion strategy is implemented by concatenating the outputs of the two ways in row style. Then the fused features are fed into a linear layer which adopts softmax activation to turn the classification results between 0 and 1 and are summed to 1. Table II shows the specific parameter settings of the overall method. Floating-point operations per second (FLOPs) is a measure of compute performance used to

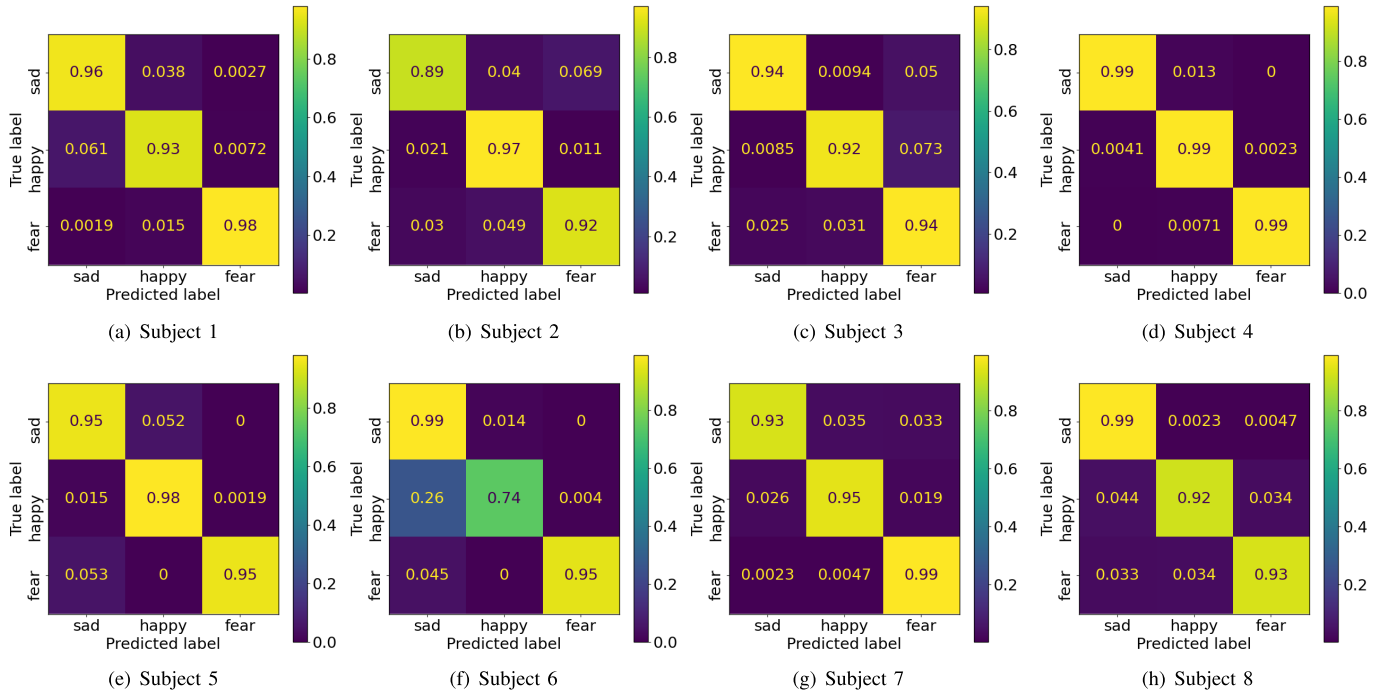


Fig. 6. Averaged confusion matrices of the eight subjects.

quantify the number of floating-point operations. We calculate the FLOPs of the whole method and result is 65.789M, which shows a competitive low cost of computational resource compared to other deep learning methods [37], [40].

IV. RESULTS

In this section, the experimental results are presented. Firstly we introduce the public dataset SEED and SEED-IV and the detailed experimental settings. Then the proposed method and other baseline models are scored on three datasets, that is, our emotional dataset collected in Section II and the two public datasets.

A. SEED and SEED-IV Datasets

The SEED [42], [44] dataset includes fifteen Chinese film clips (positive, neutral and negative emotions) and fifteen Chinese subjects (7 males and 8 females). The International 10-20 electrode positioning protocol was followed with 62 channels. The recorded EEG signals were down sampled at 200 Hz and further band-pass filtered between 0.5-70 Hz. We follow the same experimental settings with previous research [45], [46]. Specifically, the first 9 trials of EEG data serve as the training set and remaining 6 trials serve as the testing set. Then, the classification results corresponding to each session are obtained for each subject. Finally, the average classification accuracy and standard deviation over two sessions of all the 15 subjects are calculated.

The SEED-IV [47] dataset contains seventy-two film clips which have the tendency to induce happiness, sadness, fear or neutral emotions. A total of 15 subjects participated in the experiment. For each participant, 3 sessions were performed on different days, and each session contained 24 trials. The same

electrode positioning protocol and data preprocessing was followed as the SEED. We follow the same experimental settings with a previous research [48]. Specifically, the first 16 trials are used for training and the remaining 8 trials (containing all labels) for testing. All the three sessions are used for evaluation.

B. Evaluation Results

To evaluate the proposed method, our emotional dataset is first used for validation. Here we adopt the 10-fold cross-validation manner to evaluate the performance. Specifically, 510 samples are divided into ten sets for each subject, and every time we use nine sets to train the models while the remaining one set is used for validation. This process is done three times to average the results. The confusion matrices across ten folds for each subject is shown in Fig. 6.

To evaluate the effectiveness of the proposed model on our emotional dataset, we set up a group of comparative experiments on three baseline models. The implementation of these three models is as follows:

PSD + SVM: The energy spectral density PSD [41] features are extracted from the δ , θ , α , β , and γ bands of each EEG channel to form feature vectors. The classifier is set as a support vector machine.

DE + kNN: Feature vectors composed of differential entropy are classified by K-nearest neighbor method kNN, where K is equal to 5 [42].

DE + HCN: The differential entropy features of each frequency band are rearranged into a two-dimensional graph in a zero-fill manner to retain the information hidden in the electrode positions. Then, the feature maps are input into the six-layer CNN structure [43].

TABLE III
CLASSIFICATION RESULTS ON OUR EMOTIONAL DATASET

Subject	Accuracy (%) (mean \pm std.)			
	Proposed method	PSD + SVM [41]	DE + kNN [42]	DE + HCNN [43]
1	95.66 \pm 1.28	94.7 \pm 1.61	93.92 \pm 3.21	94.75 \pm 1.32
2	92.68 \pm 1.35	80.32 \pm 1.48	90.98 \pm 1.28	90.68 \pm 1.31
3	93.37 \pm 2.35	68.99 \pm 3.74	79.95 \pm 3.64	91.72 \pm 2.57
4	99.07 \pm 1.03	89.77 \pm 1.95	84.79 \pm 2.22	98.32 \pm 1.84
5	96.02 \pm 1.87	69.93 \pm 6.55	81.85 \pm 3.03	95.31 \pm 3.25
6	89.41 \pm 3.01	51.07 \pm 6.68	65.99 \pm 3.89	85.32 \pm 2.97
7	95.65 \pm 1.89	72.78 \pm 3.44	85.25 \pm 2.87	95.32 \pm 1.88
8	94.69 \pm 2.07	67.06 \pm 3.83	82.48 \pm 2.59	93.97 \pm 1.95
Average	94.57 \pm 1.86	74.34 \pm 3.661	83.15 \pm 2.84	93.17 \pm 2.14

TABLE IV
CLASSIFICATION RESULTS ON SEED AND SEED-IV DATASETS

Model	Accuracy (%) (mean \pm std.)	
	SEED	SEED-IV
SVM	83.99 \pm 9.92	56.61 \pm 20.05
GSCCA [51]	82.96 \pm 9.95	69.08 \pm 16.66
DBN [42]	86.08 \pm 8.34	66.77 \pm 7.38
BiDANN [46]	92.38 \pm 7.04	70.29 \pm 12.63
DGCNN [45]	90.40 \pm 8.49	69.88 \pm 16.29
BiHDM [48]	93.12 \pm 6.06	74.35 \pm 14.09
Our model	94.55 \pm 5.74	78.91 \pm 10.57

TABLE V
RESULT OF USING DECOMPOSED SIGNALS

Rhythm	Accuracy (%) (mean \pm std.)	
	Our Dataset	SEED
δ	66.16 \pm 5.34	75.23 \pm 7.54
θ	65.34 \pm 6.77	74.76 \pm 8.79
α	78.64 \pm 8.15	82.52 \pm 8.63
β	90.22 \pm 6.44	88.97 \pm 10.75
γ	91.97 \pm 3.68	89.06 \pm 8.31
All bands	94.57 \pm 1.86	94.55 \pm 5.74

These baseline methods are tested on our emotional data set in the same way of evaluation. The results are shown in Table III. Taken as a whole, the method we proposed achieves the highest accuracy among all the methods compared. The proposed method surpasses the accuracy of 90% for most subjects. In a comparative perspective it is about 20% higher than the PSD+SVM method and 11% higher than the DE+KNN method. Although the DE+HCNN method using a deep neural network structure achieve a higher result, there is still a 1.4% gap in the averaged accuracy compared with the proposed method. This indicates that our method can effectively perform emotion classification tasks.

Table IV gives the validation results on the SEED and SEED-IV dataset. For SEED dataset, Lu et al. [49] applied concatenation fusion, max fusion, and fuzzy integral to fuse multiple modalities and demonstrated that the fuzzy integral fusion method achieved accuracy of 87.6%. Song et al. [45] proposed DGCNN and obtained a classification accuracy of 90.4%. Yang et al. [50] built a single-layer feedforward network (SLFN) with subnetwork nodes and achieved an accuracy of 91.5%. From Table IV it can be seen that our method has achieved the best result among the five benchmarks. This exciting result illustrates the generality of our proposed double way structure. Compared to conventional model structures, the proposed structure can adaptively extract features in different frequency bands and capture fine-grained temporal features. Our model also obtains the best results on the four-class classification task SEED-IV dataset, which indicates that our model structure can be applied to multi-class emotion recognition scenarios.

We also investigate the effectiveness of each EEG rhythm by applying decomposed signals. We use FBSE to extract the five EEG rhythms from our emotional dataset. The proposed

method is used for testing with the first way removed because the attention mechanism is self-adaptive on different frequency bands, which is not suitable for comparing the effects of the rhythms. The 10-fold cross-validation is adopted on our dataset. Also, we process the SEED dataset with FBSE and test with the same model. The results are shown in Table V. It can be seen that among the five rhythms, γ rhythm attains the best results while δ and θ rhythms shows relatively poor performance. These results indicate that emotional information is more embedded in the higher frequency bands, especially in γ band, rather than the lower frequency bands of EEG signals. Similar results are also reported recently [45].

V. DISCUSSION

In this section, we make a comprehensive investigation of the proposed method. In Section V-A, we conduct a series of ablation experiments to demonstrate the necessity of the double way structure of our method. In Section V-B, we discuss the effect of residual modules on the classification performance.

A. Ablation Experiments

To demonstrate the necessity of the double way structure in our model, we conduct a series of ablation experiments. The specific settings of the ablation experiments are as follows:

Model 1: Use the double way structure as the proposed method.

Model 2: Only use the first ways, namely, the brain network construction block and the deep neural network block. At the end of the structure, the linear layer take the extracted features as input to attain classification result.

Model 3: Only use the second ways, that is, only use the neural network as the feature extractor with raw EEG input and linear is set as model 2.

TABLE VI
RESULTS OF THE ABLATION EXPERIMENT

Subject	Accuracy (%) (mean \pm std.)		
	Model 1	Model 2	Model 3
1	95.66 \pm 1.28	86.97 \pm 2.11	93.27 \pm 1.57
2	92.68 \pm 1.35	81.28 \pm 3.25	91.22 \pm 1.33
3	93.37 \pm 2.35	85.35 \pm 3.75	91.84 \pm 2.47
4	99.07 \pm 1.03	89.95 \pm 1.98	96.35 \pm 1.78
5	96.02 \pm 1.87	88.64 \pm 2.37	95.37 \pm 1.92
6	89.41 \pm 3.01	78.93 \pm 4.25	88.94 \pm 2.97
7	95.65 \pm 1.89	82.32 \pm 2.94	93.21 \pm 1.71
8	94.69 \pm 2.07	85.97 \pm 3.07	93.07 \pm 2.19
Average	94.57 \pm 1.86	84.93 \pm 2.97	92.90 \pm 1.99

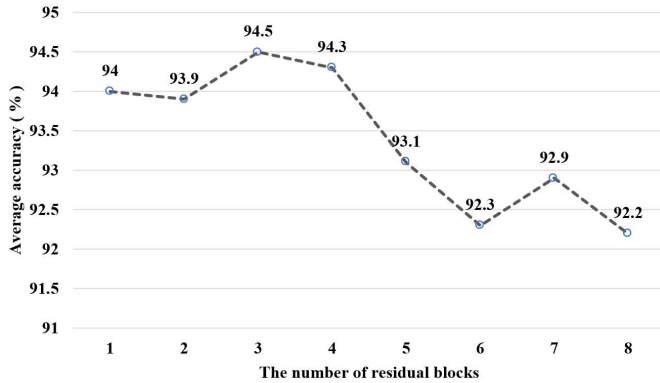


Fig. 7. Classification results of the model with different numbers of residual modules.

We perform 10-fold cross-validations on our emotional dataset. The results are shown in Table VI. Among them, the averaged accuracy of model 1 is 94.57% which is higher than that of model 2 and model 3. The standard deviation of model 1 is 1.86% which is also better than the other two models. The above results indicate that the proposed double way structure can extract information from multiple perspectives, and retain robustness among subjects, which demonstrated its superiority.

B. Residual Models

In this work, we employ residual modules in the deep neural network blocks to enhance the effectiveness of emotion classification. The residual modules are connected in sequence and the number of the modules should be discussed. Thus, a series of experiments are conducted to find out how the number influences performance. We increase the number of residual modules in the proposed method gradually and a 10-fold cross-validation is performed for each modified model on our emotional dataset. The results are shown in Fig. 7. It is shown that the accuracy gets higher as the number of the residual modules increases and hits the best result when the number is three. However, the accuracy does not improve significantly when the number is further increased and it decreases when the number is six. This indicates that the residual modules of the deep neural network with a certain depth can improve the total performance, but a too deep structure will result in degeneration of performance.

VI. CONCLUSION

In this work, we have proposed a double way deep neural network model based on brain networks. Because of the deep correlation between EEG signals gathered from different channels, it is important for a method to extract and merge effective information from different perspectives. Firstly, we transform EEG signals into five frequency bands by wavelet transform, and construct brain networks for each frequency band based on Spearman correlation coefficients between EEG channels. Brain networks are good at reflecting the correlation of brain activity regions. Then we use deep neural networks with residual connections to extract features from the obtained brain networks. Secondly, considering that the brain network may lose temporal information of the EEG signals when implementing the Spearman correlation coefficient, another deep neural network block is adopted. Finally, the features extracted by the two ways are concatenated and classified. The performance of the proposed model are verified by several comparative experiments on three datasets. A series of ablation experiments are conducted to prove the necessity of the particular structures. Moreover, the influence of residual modules on the performance is discussed and we find out the optimal value for the task of emotion recognition in the proposed method.

The research in this work can be applied to a variety of emotion recognition scenarios. To address the channel reduction and data quality issues associated with portable EEG acquisition devices in practical applications, further work will be conducted on small sample problems and subject-independent problems. With the help of progressing algorithms and hardware, we expect the BCI to be applied to real scenarios sooner and better.

REFERENCES

- [1] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 374–393, Jul./Sep. 2019.
- [2] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychol. Sci. Public Interest*, vol. 20, no. 1, pp. 1–68, Jul. 2019.
- [3] Z. Gao et al., "EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2755–2763, Sep. 2019.
- [4] P. Chen, H. Wang, X. Sun, H. Li, C. Grebogi, and Z. Gao, "Transfer learning with optimal transportation and frequency mixup for EEG-based motor imagery recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2866–2875, 2022.

- [5] Z. Liang et al., "EEGFuseNet: Hybrid unsupervised deep feature characterization and fusion for high-dimensional EEG with an application to emotion recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1913–1925, 2021.
- [6] S. Liu, X. Wang, L. Zhao, J. Zhao, Q. Xin, and S.-H. Wang, "Subject-independent emotion recognition of EEG signals based on dynamic empirical convolutional neural network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 5, pp. 1710–1721, Sep. 2021.
- [7] C. Jiang, Y. Li, Y. Tang, and C. Guan, "Enhancing EEG-based classification of depression patients using spatial information," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 566–575, 2021.
- [8] O. De Wel et al., "Complexity analysis of neonatal EEG using multiscale entropy: Applications in brain maturation and sleep stage classification," *Entropy*, vol. 19, no. 10, p. 516, 2017.
- [9] Z. K. Gao, Q. Cai, Y. X. Yang, N. Dong, and S. S. Zhang, "Visibility graph from adaptive optimal kernel time-frequency representation for classification of epileptiform EEG," *Int. J. Neural Syst.*, vol. 27, no. 4, Jun. 2017, Art. no. 1750005.
- [10] A. Anuragi, D. S. Sisodia, and R. B. Pachori, "EEG-based cross-subject emotion recognition using Fourier–Bessel series expansion based empirical wavelet transform and NCA feature selection method," *Inf. Sci.*, vol. 610, pp. 508–524, Sep. 2022.
- [11] A. Bhattacharyya, R. K. Tripathy, L. Garg, and R. B. Pachori, "A novel multivariate-multiscale approach for computing EEG spectral and temporal complexity for human emotion recognition," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3579–3591, Feb. 2021.
- [12] A. Nalwaya, K. Das, and R. B. Pachori, "Automated emotion identification using Fourier–Bessel domain-based entropies," *Entropy*, vol. 24, no. 10, p. 1322, Sep. 2022.
- [13] K. Das, P. Verma, and R. B. Pachori, "Assessment of chanting effects using EEG signals," in *Proc. 24th Int. Conf. Digit. Signal Process. Appl. (DSPA)*, Mar. 2022, pp. 1–5.
- [14] R. Sharma, R. B. Pachori, and P. Sircar, "Automated emotion recognition based on higher order statistics and deep learning algorithm," *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, Art. no. 101867.
- [15] A. Nalwaya, K. Das, and R. B. Pachori, "Emotion identification from TQWT-based EEG rhythms," in *AI-Enabled Smart Healthcare Using Biomedical Signals*. Hershey, PA, USA: IGI Global, 2022, pp. 195–216.
- [16] K. Das and R. B. Pachori, "Schizophrenia detection technique using multivariate iterative filtering and multichannel EEG signals," *Biomed. Signal Process. Control*, vol. 67, May 2021, Art. no. 102525.
- [17] W.-M. Huang, L.-C. Cao, Q.-J. Chen, C.-W. Su, Y.-J. Li, and Z.-G. Huang, "Modelling and analysis of brain functional network," *Scientia Sinica Phys., Mechanica Astronomica*, vol. 50, no. 1, Jan. 2020, Art. no. 010506.
- [18] E. C. van Straaten and C. J. Stam, "Structure out of chaos: Functional brain network analysis with EEG, MEG, and functional MRI," *Eur. Neuropsychopharm.*, vol. 23, no. 1, pp. 7–18, Jan. 2013.
- [19] W. Dang, D. Lv, L. Rui, Z. Liu, G. Chen, and Z. Gao, "Studying multi-frequency multilayer brain network via deep learning for EEG-based epilepsy detection," *IEEE Sensors J.*, vol. 21, no. 24, pp. 27651–27658, Dec. 2021.
- [20] J. Yang, J. S. Franco, F. Hétyroy Wheeler, and S. Wuhler, "Estimation of human body shape in motion with wide clothing," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 439–454.
- [21] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black, "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 561–578.
- [22] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
- [23] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, Jun. 2000.
- [24] D. Ouyang, Y. Yuan, G. Li, and Z. Guo, "The effect of time window length on EEG-based emotion recognition," *Sensors*, vol. 22, no. 13, p. 4939, Jun. 2022.
- [25] P. Jadhav, G. Rajguru, D. Datta, and S. Mukhopadhyay, "Automatic sleep stage classification using time-frequency images of CWT and transfer learning using convolution neural network," *Biocybern. Biomed. Eng.*, vol. 40, no. 1, pp. 494–504, Jan. 2020.
- [26] M. X. Cohen, "A better way to define and describe Morlet wavelets for time-frequency analysis," *NeuroImage*, vol. 199, pp. 81–86, Oct. 2019.
- [27] A. Bernardino and J. Santos-Victor, "A real-time Gabor primal sketch for visual attention," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.* Cham, Switzerland: Springer, 2005, pp. 335–342.
- [28] J. Zhang, W. Liu, and Z. Zhang, "Study on characteristics location of pantograph–catenary contact force signal based on wavelet transform," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 2, pp. 402–411, Feb. 2019.
- [29] J. J. Newson and T. C. Thiagarajan, "EEG frequency bands in psychiatric disorders: A review of resting state studies," *Frontiers Hum. Neurosci.*, vol. 12, p. 521, Jan. 2019.
- [30] Z. Gao, R. Li, C. Ma, L. Rui, and X. Sun, "Core-brain-network-based multilayer convolutional neural network for emotion recognition," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
- [31] M. Stephanou and M. Varughese, "Sequential estimation of Spearman rank correlation using Hermite series estimators," *J. Multivariate Anal.*, vol. 186, Nov. 2021, Art. no. 104783.
- [32] C. R. Buchanan et al., "The effect of network thresholding and weighting on structural brain networks in the UK biobank," *NeuroImage*, vol. 211, May 2020, Art. no. 116443.
- [33] C. Mouton, J. C. Myburgh, and M. H. Davel, "Stride and translation invariance in CNNs," in *Proc. Southern African Conf. Artif. Intell. Res. Cham, Switzerland: Springer*, 2021, pp. 267–281.
- [34] L. Weng et al., "Towards fast computation of certified robustness for ReLU networks," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 5276–5285.
- [35] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [36] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2010, pp. 92–101.
- [37] 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.
- [38] G. Thimm and E. Fiesler, "High-order and multilayer perceptron initialization," *IEEE Trans. Neural Netw.*, vol. 8, no. 2, pp. 349–359, Mar. 1997.
- [39] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.
- [40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [41] S. Barua, M. U. Ahmed, C. Ahlström, and S. Begum, "Automatic driver sleepiness detection using EEG, EOG and contextual information," *Expert Syst. Appl.*, vol. 115, pp. 121–135, Jan. 2019.
- [42] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.
- [43] J. Li, Z. Zhang, and H. He, "Hierarchical convolutional neural networks for EEG-based emotion recognition," *Cogn. Comput.*, vol. 10, no. 2, pp. 368–380, 2018.
- [44] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Nov. 2013, pp. 81–84.
- [45] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul/Sep. 2020.
- [46] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, and X. Zhou, "A bi-hemisphere domain adversarial neural network model for EEG emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 494–504, Apr. 2021.
- [47] W. L. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki, "EmotionMeter: A multimodal framework for recognizing human emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, Mar. 2019.
- [48] Y. Li et al., "A novel bi-hemispheric discrepancy model for EEG emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, vol. 13, no. 2, pp. 354–367, Jun. 2021.
- [49] Y. Lu, W. L. Zheng, B. Li, and B. L. Lu, "Combining eye movements and EEG to enhance emotion recognition," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 1–7.
- [50] Y. Yang, Q. M. J. Wu, W.-L. Zheng, and B.-L. Lu, "EEG-based emotion recognition using hierarchical network with subnetwork nodes," *IEEE Trans. Cogn. Devel. Syst.*, vol. 10, no. 2, pp. 408–419, Jun. 2017.
- [51] 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.