

Received 21 December 2023, accepted 2 January 2024, date of publication 4 January 2024,  
date of current version 19 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3349552

## RESEARCH ARTICLE

# MASTF-net: An EEG Emotion Recognition Network Based on Multi-Source Domain Adaptive Method Based on Spatio-Temporal Image and Frequency Domain Information

HONGXIANG XU<sup>1</sup>, ZIYI PEI<sup>2</sup>, QI HAN<sup>1</sup>, MINGYANG HOU<sup>1</sup>, XIN QIAN<sup>1</sup>,  
TENGFEI WENG<sup>3</sup>, YUAN TIAN<sup>1</sup>, ZICHENG QIU<sup>1</sup>, AND BAOBING ZHOU<sup>4</sup>

<sup>1</sup>School of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing 400030, China

<sup>2</sup>School of Materials Science and Engineering, Chongqing University of Arts and Sciences, Chongqing 402160, China

<sup>3</sup>School of Electrical Engineering, Chongqing University of Science and Technology, Chongqing 400030, China

<sup>4</sup>College of Information Engineering, Tarim University, Alar 843300, China

Corresponding author: Ziyi Pei (ziyi\_pei@163.com)

This work was supported in part by the West Light Foundation of the Chinese Academy of Sciences; in part by the Research Foundation of the Natural Foundation of Chongqing City under Grant cstc2021jcyj-msxmX0146 and Grant cstc2021jcyj-msxmX1212; in part by the Scientific and Technological Research Program of the Chongqing Municipal Education Commission under Grant KJQN202301517, Grant HZ2021015, Grant KJZD-K202100104, and Grant KJQN202301543; in part by the Chongqing Science and Technology Military-Civilian Integration Innovation Project, in 2022; in part by the Bingtuan Science and Technology Program in China under Grant 2021AB026; in part by the Shanxi Province Applied Basic Research Program, China, under Grant 202203021211116; and in part by the Oil and Gas Production Safety and Risk Control Key Laboratory of Chongqing Open Fund under Grant cqsrc202110.

**ABSTRACT** In the field of neuroscience, the electroencephalogram (EEG) is a crucial indicator of emotion. The EEG emotion recognition method based on domain adaptation (DA) has good objectivity and high time resolution and is the preferred method to study the brain's response to emotional stimuli. However, due to the obvious instability of EEG emotion characteristics, it is difficult to predict the emotion corresponding to EEG signals of cross-subjects by a model that combines all source domains into a single source. In order to solve the problem of cross-subject emotion analysis, we propose an EEG emotion recognition net with a cross-subject multi-source adaptive method (MASTF-net), where EEG features of different subjects are regarded as different domains. Through analyzing the invariance of the target domain and the uniqueness of the source domain, this method realizes the emotional analysis of different objects according to the spatio-temporal images and frequency domain information. First, features of EEG image are extracted from frequency and time dimensions. Secondly, combined with the serialized EEG frequency characteristics of local brain regions, independent classification module are established for different domains to recognize the emotion feature distribution of different subjects. In addition, a feature extraction method of differential entropy(DE) data of EEG is proposed based on frequency band division, which can provide stable feature input for our network structure. Finally, experiments are conducted on the SEED dataset. The experimental results show that our method has better classification accuracy in the experiment on the problem of cross multiple subjects. MASTF-net is superior to other relevant methods and models in multi-source domain. On the issue of cross subject emotion analysis, the highest accuracy of our method can reach 88.19%.

**INDEX TERMS** Cross-subject, domain adaptation, electroencephalogram, emotional stimuli, spatio-temporal image.

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

## I. INTRODUCTION

Emotion, as a form of physiological information, serves as a crucial indicator of individuals' current mental state during everyday communication [1], [2]. Its significance is evident in various contexts, such as mental state assessment [3], emotion classification, and the development of user-friendly interfaces.

From a health perspective, the frequent fluctuations in emotions often contribute to the onset of mental illness. In an exploration of this phenomenon, Jutta and Lan [4] discovered a strong correlation between depression and the implementation of emotion regulation strategies. Similarly, Radford [5] investigated the recognition of nonverbal cues and emotional signals in patients diagnosed with Alzheimer's disease. Furthermore, Weilong and Baoliang [6] directed their research towards understanding psychological issues through traditional means, such as the analysis of facial expressions and language, which enables the refinement and classification of emotional problems. However, these methods often fail to yield satisfactory testing results, as they can be influenced by the subjects' personal circumstances. Therefore, there is a need for a measurement method that can effectively capture the genuine sentiments of the subjects.

In recent years, non-invasive brain-computer interfaces, such as electroencephalography (EEG), have gained widespread usage in acquiring brain signals and analyzing psychological disorders. These techniques are appealing due to their reliability, accessibility, and high precision [7], [8]. Researchers have focused on developing a method for detecting mental health conditions based on EEG signals, with the goal of identifying the onset of mental illness and enabling early and effective treatment initiation. For instance, Yildirim et al. [9] extracted features based on spectral power and phase-locked values, and conducted emotion recognition research using swarm intelligence (SI) algorithm. Reference [10] utilized the spatial information extracted from EEG signals to classify individuals with depression. Similarly, Bingtao et al. [11] proposed a network framework for diagnosing severe depression based on EEG data. Additionally, Behshad and Mohammad [12] investigated the nonlinear characteristics of EEG signals to differentiate between patients with depression and healthy individuals. These studies highlight the potential of EEG-based approaches in diagnosing mental health disorders and facilitating timely intervention.

The emotion recognition method based on EEG signals can more accurately reflect the changes in the subject's emotions compared to methods based on facial and language information. However, considering the individual differences in EEG signals, it remains challenging to share a network model trained on specific subjects with other subjects for testing. Although this detection method utilizing EEG signals from a single subject improves the efficiency of model training, its application in cross-subject and cross-session scenarios is difficult.

To address the issue of emotion analysis among multiple subjects, domain adaptation (DA) has been widely employed in this research work. DA improves the learning in the unlabeled target domain through the transfer of knowledge from the source domains, which can significantly reduce the number of labeled samples [13].

In practical applications, we often encounter scenarios with multiple source domains, such as in cross-subject emotion analysis research, where EEG data from different subjects are considered as multi-source domain data. In recent years, many researchers have tended to merge multiple source domain data into one domain. For example, Yiming et al. [14] and He and Yiming [15] employed deep adaptive networks (DANs) [16] for emotion recognition based on EEG signals, using maximum mean difference (MMD) [17] as a distance metric between source and target domains.

The approach of combining multiple source domain data into one domain disrupts the unique characteristics of EEG signals from different subjects, neglecting the differences in EEG among different subjects, resulting in significant biases in cross-subject applications. To address this issue, we introduced a multi-branch structure approach [13], establishing one-to-one DA branches for all target domains and source domains to preserve the specific features of each source domain data. Many similar methods appeared in [13], [18], [19]. Pan and Zheng [18] proposed a MSFR-GCN model for EEG-based emotion and cognition recognition using graph convolutional networks (GCN) to obtain graph representations of EEG signals. Wenpeng and Bing [19] developed an end-to-end depth convolutional neural network (CNN) for cross-subject transfer using attention-based CNNs. However, these methods only considered the spatial features of different electrode locations in EEG signals, neglecting the temporal sequence characteristics of EEG signals, leading to suboptimal performance of the model.

To address the issues of multi-source domain adaptation and effectively utilize both temporal and spatial features in EEG-based emotion recognition, we propose an EEG emotion recognition network by use of multi-source domain adaptive method based on spatio-temporal image and frequency domain information (MASTF-net). In MASTF-net, a dual-branch shared feature extraction network is used to extract spatial and temporal features from EEG signals respectively, and a one-to-one DA branch feature extractor is established to capture domain-specific features. In this way, we consider the differences in EEG signals among different subjects, as well as the temporal and spatial features.

To sum up, the main contributions of our work are as follows:

- A method of extracting EEG DE feature based on frequency band division is proposed. The DE features are analyzed and extracted by time segment extraction, and these features are used as the input of MASTF-net to retain the time characteristics of DE features. The input of network with time information retained has better

performance than single DE input with single DE feature in the experiment. The network input that preserves time information has better performance than a single DE feature input.

- We propose a module(L-RNN) for extracting serial EEG temporal features from local brain regions, which is used to extract features of emotional state differences between left and right brain regions and learn the differences between brain regions under different emotions.
- A common feature extraction network with a dual branch structure is proposed for extracting low-level features from EEG data. In the network, spatial branches extract EEG image features from frequency and time dimensions, while temporal branches extract serialized EEG temporal features from local brain regions. By combining temporal, spatial, and frequency information, multi-source EEG features are enhanced to improve network performance.

## II. RELATED WORK

### A. EMOTIONAL ANALYSIS VIA DEEP LEARNING BASED ON BRAIN-COMPUTER INTERFACE (BCI)

Most of the research on emotion recognition has focused on analyzing the subject's voice, image, and text data. For instance, Rahman et al. developed a multimodal network that integrates image and text data for emotion analysis, which has demonstrated favorable outcomes in assessing emotions [20]. However, the majority of existing emotion assessment methods heavily rely on the completeness of subjects. Consequently, when conducting psychological evaluations, relying solely on image and text information for assessment can have a significant adverse impact on the evaluation results. EEG is a physiological signal that provides excellent insights into the psychological activity of subjects. Unlike text and image information, EEG signals are generally less prone to deception. As a result, there has been a growing focus on biomedical signal prediction research, particularly in the context of utilizing EEG data. For instance, the Multimodal Physiological Emotion Database (MPED) corpus [21] presents a dataset consisting of physiological records and establishes an initial benchmark for estimating emotional states. Zhao et al. introduced a linear graph convolution network (LGCN) based EEG signal detection model, aiming to address the deficiency in capturing the spatial relationships among different EEG channels [22]. Furthermore, researchers have discovered that tracking Electrocardiogram (EOG) signals can enhance detection accuracy in emotional analysis. By employing feature-level fusion and decision-level fusion of EOG and EEG signals, the capability of emotion recognition can be significantly enhanced [23].

In recent years, there has been a growing interest among researchers in the field of artificial intelligence (AI) and, more specifically, in algorithms based on machine learning (ML). This increased interest can be attributed to the higher robustness, improved accuracy, and technological diversification

offered by AI. However, most research was based on expensive large-scale equipment to collect experimental data, which has a huge obstacle to the research of biological signals. Recently, BCI plays an important role in emotion recognition, because BCI can use simple equipment to solve the research difficulties caused by large equipment. BCI is used to establish the correlation between the computer and the subject's brain physiological signals, this connection is non-invasive, so it will not cause harm to the subject when collecting data. In addition, researchers began to pay attention to the application of deep learning in emotion analysis based on BCI. Alexander and Yongtian [24] used deep learning (DL) algorithm for EEG, especially emotional assessment tasks. In addition, some models retrieved emotional states from EEG signals based on DL algorithm [6], [25]. These models adopted DL to analyze the conversion of EEG signals to different forms. For example, EEG signals were transformed into graphical representations in [26], and vectors were separated between each hemisphere [6].

### B. DOMAIN ADAPTATION VIA DEEP LEARNING

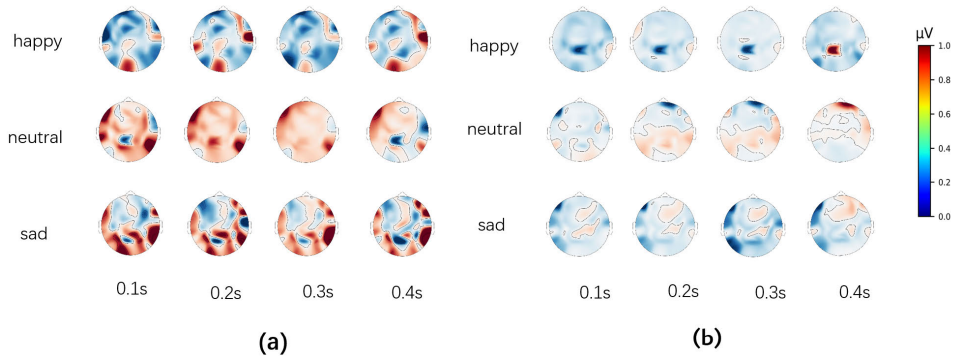
In EEG emotion analysis involving multiple subjects, many researchers use one-to-one domain adaptation (DA) to combine EEG signals from different subjects and a source domain, which ignores the difference of marginal distribution in different EEG domains. To solve this problem, Chen et al. proposed MS-MDA [13], which regards each data as a separate domain, thus avoiding the destruction of the edge distribution of multiple EEG source domains, and the domain invariant feature is also considered.

In order to solve the problem of EEG emotion classification of different subjects, MASTF-net is proposed for estimating emotions from EEG. The net is composed of a common feature extraction module with two branches and a domain feature extraction module with multilayer perceptron (MLP). In addition, EEG data comes from three domains of the frequency, time and space.

## III. MATERIALS

### A. DATASET

In this article, we use SEED and SEED-IV, In SEED, there are three emotion labels where 0 represents sadness, 1 represents neutral emotion, and 2 represents positive emotion. 15 people participated in the test, including 7 men and 8 women. EEG signals are recorded during watching 15 different movie clips by ESI NeuroScan system at a sampling rate of 1000Hz. In SEED-IV, there are four emotion labels where 0 represents happy, 1 represents sad, 2 represents neutral emotion and 3 represents fear. In these two datasets [19]. Each subject has gone through three session experiments, and the number of session is represented by  $n_s \in \{1, 2, 3\}$ . And  $n_p \in \{1, 2, 3, \dots, 15\}$  indicates a number of subject. In order to clearly express the meaning of symbols in the paper, symbols are listed in Table 1.



**FIGURE 1.** Select continuous EEG data for a specific time period from SEED, and obtain continuous EEG topographic map for three different tags (sad, neutral, happy). (a) The EEG topographic maps of the subject are shown when the emotional change is not active. (b) The EEG topographic maps of the subject are shown when the emotional change is active. A frame of EEG topographic map is obtained for the time interval is 0.1 seconds from SEED.

**TABLE 1.** Meaning of each symbol.

Symbol	Meaning
$n_s$	A number of session
$n_p$	A number of subject
$n_b$	Total number of frequency band
$n_c$	Total number of EEG electrodes
$n_t$	Total number of trials for each session of one person
$L_t$	Length of each sample
$r_{n_p}^{n_s}$	One data sample for each subject ( $r_{n_p}^{n_s} \in \mathbb{R}^{n_c \times n_b \times L_t}$ )
$S$	One sample for each of fifteen subjects ( $S \in \mathbb{R}^{15 \times n_c \times n_b \times L_t}$ )
$A_j$	The $j$ th region divided from the EEG topographic map
$n_{A_j}$	Feature number of $A_j$ region
$a_i^{A_j}$	The result of splitting $r_{n_p}^{n_s}$ by $L_t$ from EEG topographic map regions $A_j$ ( $a_i^{A_j} \in \mathbb{R}^{n_{A_j} \times n_b}$ )
$R^{A_j}$	The input of L-RNN, which consisted by $a_i^{A_j}$
$q_{n_p}^{n_s}$	EEG images obtained from FS ( $q_{n_p}^{n_s} \in \mathbb{R}^{n_{A_j} \times n_b}$ )
$Q$	Set composed of $q_{n_p}^{n_s}$
$p_t$	The result of splitting $q_{n_p}^{n_s}$ by $L_t$ ( $p_t \in \mathbb{R}^{n_c \times 32 \times 32}$ )
$x_{n_p}^{n_s}$	Common features
$Y_{n_p}^{n_s}$	Predicted result
$\hat{Y}_{n_p}^{n_s}$	Label
$\phi, \Phi$	$\phi$ represents DSFE $\Phi$ represents DSC

## B. SIGNAL PREPROCESSING

EEG is one of the physiological signals that can truly reflect people’s inner activities. However, in the process of EEG data collection, acquisition equipments and acquisition environment often lead to the interference of eye movement, electromyography (EMG) or other high-frequency noise. Therefore, the collected EEG data can only be used after being processed in a professional way. In the SEED dataset, the raw EEG signal is down-sampled at 200Hz, and then the noise interference is reduced by using 1Hz to 75Hz band-pass filtering. Next, the signal features are extracted by DE,

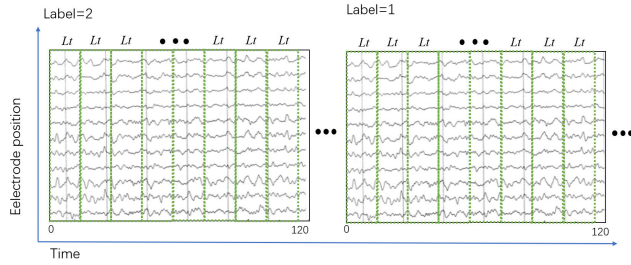
because DE features have the best performance in feature selection, so DE features are choose as the input of MASTF-net. SEED dataset provides DE features with five bands which are delta (1-4 Hz), theta (4-8Hz), alpha (8-14Hz), beta (14-31Hz) and gamma (31-50Hz). The extraction method of DE feature is as follows [26]

$$h(x) = - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx = \frac{1}{2} \log(2\pi e\sigma^2) \quad (1)$$

where the input  $x$  is subject to the Gaussian distribution  $N(\mu, \sigma^2)$ ,  $\mu$  and  $\sigma$  represent variance and standard deviation, respectively and  $e$  is the Euler constant. According to [26], raw EEG data are inputted in the form of vector and the DE feature  $h(x)$  can be obtained by transforming the input EEG raw data through equation 1. The dimension of a DE feature map is 310 (62 channels  $\times$  5 frequency bands) where 62 channels means that EEG signals are collected according to 62 channel ESI NeuroScan system, and the five frequency bands are represented by different letters  $\delta, \theta, \alpha, \beta$  and  $\gamma$  respectively.

## C. FEATURE SELECTION ANALYSIS

In order to obtain the regular of the strength of EEG signals with time, the data of three different emotions are analyzed, such as sad, neutral, and happy. Due to the different length of movie clips in SEED, we select the first 60 seconds and the last 60 seconds of each movie clip to judge which time period of emotional activity is the most obvious. Here, 100 EEG data are randomly selected, each of which consist of four frames and the interval of each frame is 0.1 second. Then, the EEG data are averaged and mapped into the EEG topographic map. The color in the EEG topographic map represents the intensity of the EEG signal, where the darker the color, the more obvious the emotional activity of the EEG as shown in Fig. 1. Fig. 1(a) shows the EEG topographic map that is averaged after 100 random samples in the first 60 seconds, and Fig. 1(b) shows the EEG topographic



**FIGURE 2.** Extract the first 120 values of each continuous DE feature in SEED. The 120 values are divided into samples of length  $L_t$ . Then, EEG data samples are drawn through the  $L_t$  second window (green dotted line window).

map that is averaged after 100 random samples in the last 60 seconds. From the comparison between Fig. 1(a) and (b), it can be seen that the period of emotional activity always appears in the early stage of the emotional test. As time goes on, the emotional activity becomes weaker and weaker.

Because different movie clips have different lengths of time, the dimensions of DE features are different. If the data sample is not filtered, the sample quantity of different labels will vary too much, which affects the final classification results. In order to maintain the balance of data samples, the number of samples for different labels should be consistent. The duration of each targeted experiment is not uniform, and from a psychological point of view, significant emotional changes often occur in the first part of a period of time. Therefore, in order to balance the number of samples and get better classification results, the first 120 DE features are obtained in each targeted experiment in the paper, which as shown in Fig. 2. The total number of training data for each session is  $1 \times 15 \times n_t \times \lfloor 120/L_t \rfloor$ , where 1 represents a session and 15 represents the total number of subjects,  $n_t$  represents total number of trials for each session of one person and  $L_t$  represents the length of each sample, which as shown in Fig. 2. Because there are three sessions in total, the total number of samples is  $3 \times 15 \times n_t \times \lfloor 120/L_t \rfloor$ , where 3 represents the total number of sessions. The range of  $L_t$  is from 4 to 8, which guarantees that the total number of samples at least 10000 for ensuring sufficient number of samples.

## IV. METHODS

A new architecture is proposed, which combines multi-source domain learning and spatio-temporal EEG image representation, as shown in Fig. 3. This model mainly includes three modules, which are two modules (TFEM and SFEM) with different common feature extraction and a module (DACM) with independent adaptive domain feature classification. TFEM evaluates the relationship of EEG data from the left and right hemisphere according to the brain electrode positions in different local areas and extracts the temporal features of EEG signals based on the contribution of each element at each spatial level. SFEM learns the spatial information contained in EEG data through a spatial extraction network, and extract the time frame features of EEG images by a gated recurrent neural network (GRU) [27].

In DACM, an independent discriminant branch is established for each different source domain data to learn the features of multiple specific domains. Next, we will describe the details of each module.

### A. INPUT DATA OF MASTF-NET

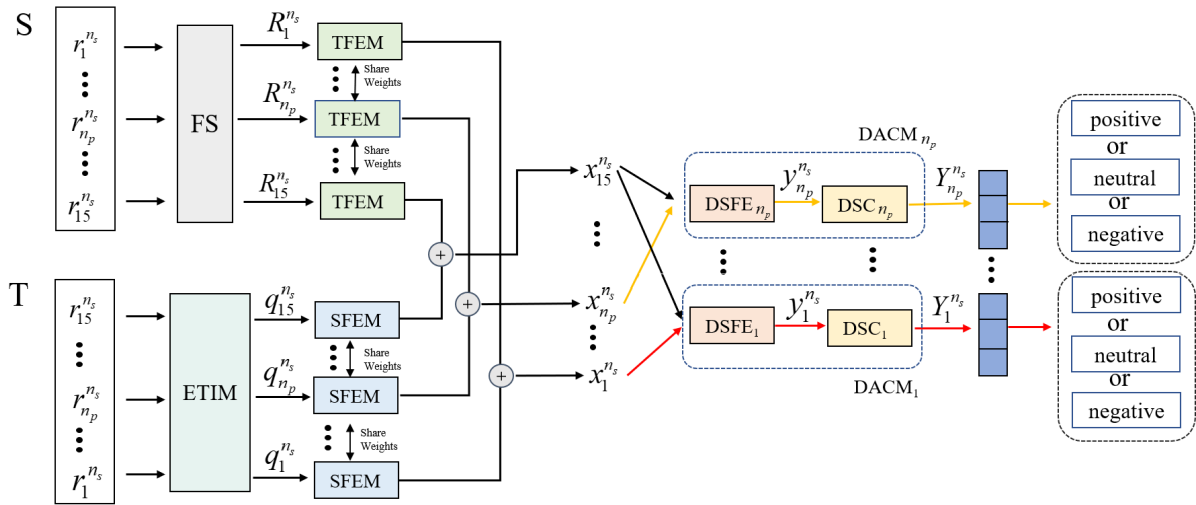
When training the network, we use labeled training data and unlabeled test data as the input of the network, and the input data is represented by a matrix  $S = \{r_1^{n_s}, r_2^{n_s}, \dots, r_{n_p}^{n_s}, \dots, r_{15}^{n_s}\}$ , as shown in Fig. 3.  $r_1^{n_s}, r_2^{n_s}, \dots, r_{14}^{n_s}$  are the labeled source domain data, which is used to train the model.  $r_{15}^{n_s}$  is the target domain data without label, which is used for model training and model accuracy test. The total number of samples is  $3 \times 15 \times n_t \times (120/L_t)$ , and each matrix  $r_{n_p}^{n_s}$  represents a training data containing  $n_c$  electrode positions in the  $n_b$  frequency band with length of  $L_t$ . The dimension of matrix  $r_{n_p}^{n_s}$  is represented as  $\mathbb{R}^{n_c \times n_b \times L_t}$ , where  $n_c$  is the total number of electrodes for collecting EEG, and  $n_b$  is total number of frequency bands.

### B. TEMPORAL FEATURE EXTRACTION MODULE (TFEM)

From Fig. 1, the level of brain activity is observed in different regions, where the action of EEG is high in the frontal lobe region, temporal lobe, and the occipital lobe, but the movement is relatively weak in the parietal lobe. It can be seen from the Fig. 1 that the most active regions of the brain are the left prefrontal lobe, right prefrontal lobe, left temporal lobe, right temporal lobe, left occipital lobe, and right occipital lobe, which show that EEG signals in these regions are more discriminative. In selecting EEG signal data, the relevant knowledge of EEG signal is considered. The frontal and temporal brain regions, especially the prefrontal regions, are mainly used to regulate emotions and memory functions. The occipital lobe region is mainly responsible for visual function, so the electrode signals in the occipital lobe brain region are also essential for visual emotion analysis. The parietal lobe brain region is mainly responsible for tactile and motor functions and plays a relatively weak role in emotional analysis. The results shown in Fig. 1 are consistent with physiological signals.

In order to explore the feature transformation of the frontal lobe, temporal lobe, and occipital lobe in the left and right hemispheres when emotions change, we propose TFEM based on LSTM [27] and RNN(L-RNN) to obtain the relationship between electrodes, physiological regions, and cerebral hemispheres. TFEM is an improvement on the hierarchical RNN(H-RNN) [28]

According to the different responses of EEG signals to emotional activities, we divide electrode positions of active region into eight regions. The electrode position can be divided according to the international standard 10-20 of EEG electrode position [29]. Therefore, eight-part DE features are selected in the paper, such as  $A_1$ (FP1,AF3,F1,F3,F5,F7),  $A_2$ (FP2,AF4,F2,F4,F6,F8),  $A_3$ (FT7,T7,TP7,P7,PO7),  $A_4$ (FT8,T8,TP8,P8,PO8),  $A_5$ (P5,



**FIGURE 3.** The structure of MASTF-net. This model mainly includes three modules, which are two modules(TFEM and SFEM) with different common feature extraction and multiple modules(DACM<sub>*n<sub>p</sub>*</sub>) with independent adaptive domain feature classification. *S* contains a sample  $r_{n_p}^{n_s}$  of each subject. TFEM evaluates the relationship of EEG data from the left and right hemispheres according to the brain electrode positions in different regions and extracts the temporal features of EEG signals based on the contribution of each region at each spatial level. SFEM learns the spatial information contained in EEG data through a spatial extraction network, and extracts the time frame features of EEG images by a gated recurrent neural network(GRU). DACM establishes separate domain feature extraction branches(DSFE<sub>*n<sub>p</sub>*</sub>) and classifiers(DSC<sub>*n<sub>p</sub>*</sub>) for different domains according to the common features extracted by TFEM and SFEM. FS represents feature selecting, which is detailed in Section IV-B1. ETIM is described in Section IV-C1.

P3, P1,PO3, PO5),  $A_6$ (P6, P4,P2,PO4, PO6),  $A_7$ (O1,CB1) and  $A_8$ (O2,CB2), where the eight-part DE features correspond to parts of a brain with obvious EEG activity. The eight regions are shown in Fig. 4.

### 1) FEATURE SELECTING (FS) OF EEG TOPOGRAPHIC MAP REGIONS

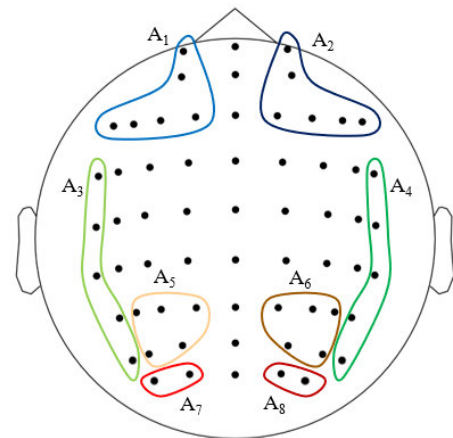
As shown in Fig. 3, the dimension of  $r_{n_p}^{n_s}$  is  $n_c \times n_b \times L_t$ . The features of  $r_{n_p}^{n_s}$  are extracted by FS in eight regions. The process of FS is as follows. We divide  $r_{n_p}^{n_s}$  according to  $L_t$ , where  $r_{n_p}^{n_s} = [a_1, a_2, \dots, a_i, \dots, a_{L_t}]$ ,  $i \in \{1, 2, \dots, L_t\}$  and  $a_i \in \mathbb{R}^{n_c \times n_b}$ .

As shown in Figure 4, 36 electrodes are selected from  $n_c$  electrodes, and the 36 electrodes are divided into eight regions, where the eight regions are  $A_1, A_2, \dots, A_j, \dots, A_8$ .  $a_i^{A_j}$  represents DE features of  $A_j$  region at  $i$ . The dimension of  $a_i^{A_j}$  is  $n_{A_j} \times n_b$ , where  $n_{A_j}$  represents the total number of electrodes in the  $A_j$  region.

In order to obtain the difference between the left and right brain EEG features and retain the time information of the EEG signal, we merge the EEG features of the same area at  $L_t$  into the same matrix  $R^{A_j} = [a_1^{A_j}, a_2^{A_j}, a_3^{A_j}, \dots, a_{L_t}^{A_j}]$ , which represents as  $R_{n_p}^{n_s} = \{(R^{A_j}, R^{A_{j+1}})\}$ ,  $j = \{1, 3, 5, 7\}$  and  $R_{n_p}^{n_s}$  is the result of FS.

### 2) L-RNN AND TFEM STRUCTURE CONFIGURATIONS

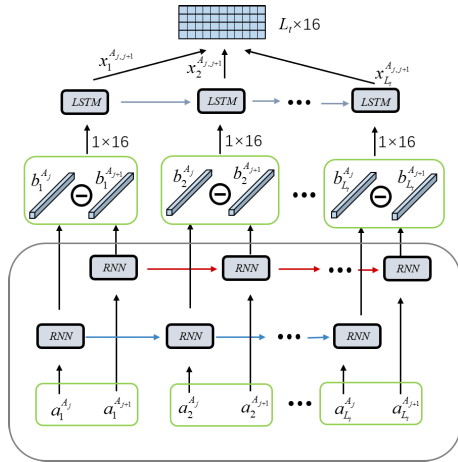
The features about time, space and frequency information can be extracted from EEG signals. At the same time, in order to consider the relationship between different cerebral hemispheres based on the same time period, an L-RNN module consisting of multiple parallel RNN and LSTM [30]



**FIGURE 4.** In order to explore the feature transformation of frontal lobe, temporal lobe and occipital lobe in left and right hemispheres when emotions change, eight regions with frequent EEG activity during emotional changes are divided.

was established, where each layer of L-RNN is used to extract information from different regions, including electrodes, physiological region information, and hemisphere-based relationships. The structure of L-RNN is shown in Fig. 5.

Two parallel RNNs with the same structure are used to initially extract the time information of DE features, and the information based on time is processed through the RNN hidden layer. The traditional RNN unit has the problem of gradient disappearance, which causes certain damage for processing of DE features of EEG. If the DE features are processed through simple RNN, the needs of EEG analysis may not be met. Therefore, after extracting feature information through RNN, RNN is no longer used to process



**FIGURE 5.** The structure of L-RNN. Each L-RNN input is a symmetrical EEG feature.  $t$  represents the input time sequence,  $A_j$  represents the location of the EEG region, and  $a_i^j$  represents the EEG feature in the  $A_j$  region of the corresponding time series  $t$ .  $\ominus$  represents the subtraction of the elements at the corresponding position of vectors  $b_i^{A_j}$  and  $b_i^{A_{j+1}}$ . The final output is the features matrix with dimension  $L_t \times 16$ .

EEG information of different regions but the proposed LSTM unit is used to further extract EEG feature of the local hemisphere. LSTM not only improves the memory ability of time information on the basis of RNN, but also improves the ability to solve the problem of gradient disappearance.

Each input of L-RNN module is a symmetrical EEG feature  $\{R^{A_j}, R^{A_{j+1}}\}$ , as shown in Fig. 5.  $L_t$  represents the input time sequence,  $A_j$  represents the location of the hemispheres region, and  $a_i^j$  represents the DE feature in the  $A_j$  region of the corresponding time series  $L_t$ . The final output result is a vector of length 16. We input  $a_i^j, a_i^{A_{j+1}}$  into the parallel RNN according to the time sequence  $L_t$  to get  $b_i^{A_j}$  and  $b_i^{A_{j+1}}$ , which is represented as

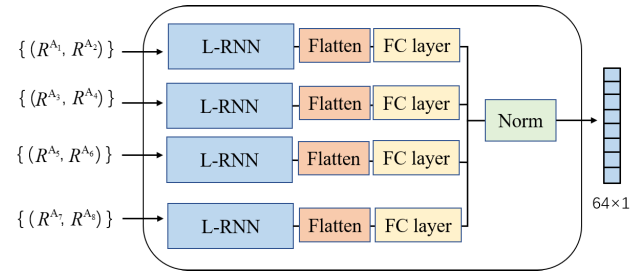
$$b_i^j = RNN(a_i^j), i \in [1, L_t], j \in \{1, 2, \dots, 8\} \quad (2)$$

After the features of each region are obtained, LSTM is used to obtain the feature information between the left and right brain in time series in order to obtain the activity difference between the left and right brain, which is represented as

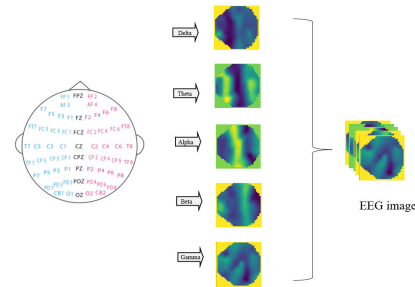
$$x_i^{A_j, A_{j+1}} = Relu(LSTM(b_i^{A_j} \ominus b_i^{A_{j+1}})), i \in [1, L_t], \quad (3)$$

where  $\ominus$  represents that the corresponding positions of matrices  $b_i^{A_j}$  and  $b_i^{A_{j+1}}$  are subtracted.  $x_i^{A_j, A_{j+1}}$  is the feature extracted from LSTM and Relu activation function is used after  $\ominus$ , and the final output result  $x_i^{A_j, A_{j+1}}$  of L-RNN is gotten. The final output is a matrix with dimension  $L_t \times 16$ .

Each L-RNN module can extract the EEG feature of two symmetrical hemispheres regions  $A_j$  and  $A_{j+1}$ . In order to obtain the DE feature information of all important regions from  $A_1$  to  $A_8$ , multiple parallel L-RNNs are spliced together to form a TFEM module, which as shown in Fig. 6. The input of each TFEM module is the DE features of eight areas. The paired EEG features are input into four L-RNN modules



**FIGURE 6.** The structure of TFEM. The EEG features of eight areas are input into the L-RNN module in pairs. The output from the L-RNN module is flattened and spliced after FC layer, and finally, a vector with a length of 64 is obtained after Norm.



**FIGURE 7.** EEG data of five frequency bands are mapped to the corresponding positions of the EEG map to form a five-channel EEG map.

with the same structure, and the results obtained by L-RNN are flattened and spliced [31] after fully connected layer(FC layer) [32]. A vector with a length of 64 is obtained after Norm [32].

### C. SPATIAL FEATURE EXTRACTION MODULE (SFEM)

TFEM module plays an important role in extracting features of EEG signal, but TFEM only collects EEG features of local important brain regions and does not pay attention to the overall changes of global brain regions. Therefore, in order to supplement the TFEM module, we propose a feature extraction method for global EEG features(SFEM). The SFEM module contains two sub-modules, which are a module of converting EEG signals into images(ETIM) and a spatial convolution module(SCM), which are shown in Fig. 7 and Fig. 8. In ETIM, DE features are mapped into the EEG topographic map and SCM is used to obtain the global EEG spatial feature information. SFEM module aims to learn the feature representation of global EEG information.

#### 1) EEG TOPOGRAPHIC MAP FEATURE MAPPING

The DE features is transformed into an image representation of the EEG signal feature, and the spatial information is considered from the image-based EEG feature map [33]. The EEG mapping conversion method we use is as follows. Firstly, the specific position of the 3D brain electrode (the Cartesian coordinates of the scalp position ) is obtained [34], and the three-dimensional electrode coordinates are mapped to the two-dimensional space according to the isometric projection. Secondly, DE features is assigned to the corresponding electrode position, and then the EEG feature

image is constructed by cubic spline interpolation [35]. Finally, the dimension of EEG topographic map is  $n_b \times h \times w$ , where  $h$  is height,  $w$  is width,  $n_b$  is the channel of the EEG topographic map, which represents five different frequency bands. The height  $h$  and width  $w$  can be taken as required. In this paper, we set the size of EEG image to  $32 \times 32$ . Therefore, the size of each sample image is expressed as  $5 \times 32 \times 32$ . The specific conversion process is as follows [36]:

$$\text{elev} = \text{atan} 2 \left( z, \sqrt{x^2 + y^2} \right) = \begin{cases} \arctan \left( \frac{z}{\sqrt{x^2 + y^2}} \right) \\ \arctan \left( \frac{z}{\sqrt{x^2 + y^2}} \right) + \pi \\ \arctan \left( \frac{z}{\sqrt{x^2 + y^2}} \right) - \pi \\ \frac{\pi}{2} \\ -\frac{\pi}{2} \end{cases} \quad (4)$$

Equation (4) calculates the elevation angle  $\text{elev} \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ , formed by the  $z$  axis and  $xy$  plane from the three-dimensional space, where  $x$ ,  $y$  and  $z$  represent the three-dimensional coordinates of the input electrode.

$$\text{az} = \text{atan} 2 (x, y) = \begin{cases} \arctan \left( \frac{x}{y} \right) \\ \arctan \left( \frac{x}{y} \right) + \pi \\ \arctan \left( \frac{x}{y} \right) - \pi \\ \frac{\pi}{2} \\ -\frac{\pi}{2} \end{cases} \quad (5)$$

Equation (5) is the counterclockwise angle in the  $xy$  plane measured from the positive direction of the  $x$  axis, and the angle value is within the range of  $(-\pi, \pi)$ .

$$\hat{x} = \left(\frac{\pi}{2} - \text{elev}\right) \cos(\text{az}) \quad (6)$$

$$\hat{y} = \left(\frac{\pi}{2} - \text{elev}\right) \sin(\text{az}) \quad (7)$$

Equation (6) and Equation (7) respectively calculate electrode coordinates  $\hat{x}$  and  $\hat{y}$  mapped from 3D to 2D.

Through the above EEG to image method(ETIM), we generate five characteristic maps corresponding to five frequency subbands of 62 electrode positions, as shown in Fig. 7. Before retraining the model, the EEG features of 15 subjects are converted into images. We transform matrix  $S$  into matrix  $Q$ , where  $S = \{r_1^{n_s}, r_2^{n_s}, \dots, r_{15}^{n_s}\}$  and  $Q = \{q_1^{n_p}, q_2^{n_p}, \dots, q_{n_p}^{n_p}, \dots, q_{15}^{n_p}\}$  through the ETIM. For sample  $q_{n_p}^{n_p}$ , the dimension of  $q_{n_p}^{n_p}$  is  $n_c \times 32 \times 32 \times L_t$ .  $q_{n_p}^{n_p}$  is rewritten as  $[p_1, p_2, \dots, p_i \dots p_{L_t}]$ , where the dimension of  $p_i$  is  $n_c \times 32 \times 32$ ,  $n_c = 5$ .

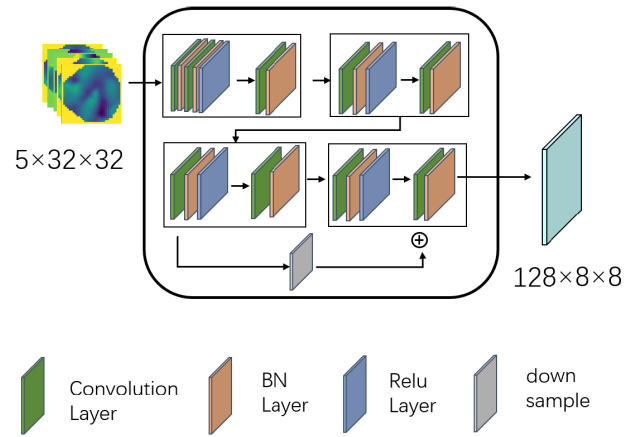


FIGURE 8. The structure of SCM. SCM is used to extract the global spatial feature of EEG, and the dimension of final output feature is  $128 \times 8 \times 8$ .

## 2) SPATIAL FEATURE EXTRACTION STRUCTURE BASED ON SCM AND GRU

In order to obtain the global EEG spatial feature, SCM and GRU are used to extract global EEG feature based on time frames. In SCM shown in Fig. 8, Resnet18 [37] are used as the basic network. The network first learns the spatial spectrum features in EEG images based on the 2D convolution layer and then uses GRU to learn the time series features. The reason for using GRU is that image data is difficult to process compared with signal data. In order to improve the operation speed, the faster GRU is selected to replace LSTM [38]. SFEM is used to make up for the incomplete global feature extraction of TFEM. If LSTM is used, the training efficiency will be reduced. Therefore, a more efficient GRU is used as a substitute for LSTM.

As shown in Fig. 9, this structure enables the GRU unit to discard useless outdated information and update its own state according to the newly obtained input. The input of each period of time  $t$  is expressed as  $x_t$ . The state of the reset gate and the update gate are controlled according to the input state at time  $t - 1$  and the input at time  $t$ , where the control door is represented as

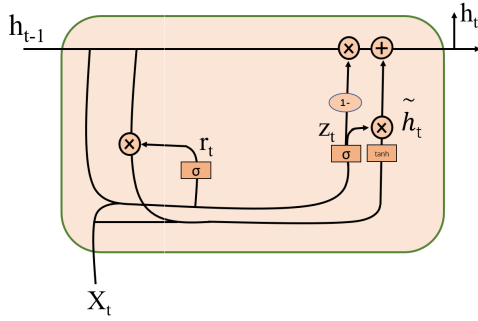
$$r_t = \sigma (W_r \cdot [h_{t-1}, x_t]) \quad (8)$$

where  $r_t$  indicates the door control of control reset,  $W_r$  is the weight matrix of reset gate,  $h_{t-1}$  represents the state information of previous GRU unit,  $\sigma$  represents the logistic sigmoid function. This function can be used to convert data to a value in the range of 0-1. Update door representation as

$$z_t = \sigma (W_z \cdot [h_{t-1}, x_t]) \quad (9)$$

where  $z_t$  represents the door control for updating,  $W_z$  represents the weight matrix of the update door. The closer the gate control signal is to 1, the more data will be memorized; The closer you get to 0, the more you forget. When the  $r_t$  is obtained, it can pass the current candidate state  $\tilde{h}_t$ ,  $\tilde{h}_t$  depends on the state information  $h_{t-1}$  at the previous time.  $\tilde{h}_t$





**FIGURE 9.** The traditional structural unit of GRU extracts the time information in EEG DE features.

**TABLE 2.** Parameter configuration used in SCM.

Kernel Size / Strides	Input/Output Channels	Output Size
Input(5×32×32)	–	–
3×3/1	5/64	32×32
3×3/1	64/64	32×32
3×3/1	64/64	32×32
3×3/1	64/64	32×32
3×3/1	64/64	32×32
3×3/2	64/128	16×16
3×3/1	128/128	16×16
1×1/2	64/128	8×8
3×3/1	128/128	8×8
3×3/1	128/128	8×8

is expressed as

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \quad (10)$$

where  $W_{\tilde{h}}$  represents the weight matrix of candidate states. Finally,  $z_t, h_{t-1}, \tilde{h}_t$  get the current state  $h_t$ .

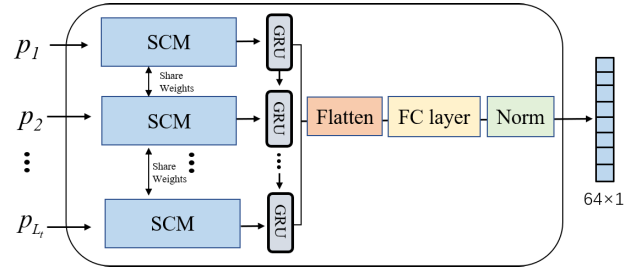
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (11)$$

In order to obtain the spatial information of the electrode position in EEG images, SCM is built to learn spatial features from multiple frequency dimensions as shown in Fig. 8. Table 2 shows the specific parameter configurations for the spatial convolution module.

In order to extract the global EEG features without ignoring the time features  $L_t$ , SCM modules are used to extract the spatial position information of EEG in different time sequences and GRU is used extract the time features. Each  $p_i$  is calculated by the SCM module with weight sharing, which is shown in Fig. 10. The input dimension of GRU is  $L_t \times 128 \times 8 \times 8$ , which consists of  $L_t$  parallel SCM outputs. Each GRU unit obtains a feature vector with a length of 256. The output of all GRU units is spliced and flattened to obtain a feature vector with a length of 64 through FC layer, and then the final result is obtained by Norm. The whole process is shown in Fig. 10. The configuration of GRU and FC layer is shown in Table 3.

#### D. DOMAIN ADAPTIVE CLASSIFICATION MODULE (DACM)

Domain adaptation (DA) [39] is a special case of Transfer Learning. The idea is to map data features from different



**FIGURE 10.** The structure of SFEM. SFEM module contains  $L_t$  sub-modules, which are SCM. SCM is used to obtain the global EEG spatial feature, and weight sharing among  $L_t$  SCM modules. SFEM module is to learn the feature representation of global EEG information.

**TABLE 3.** Configuration of GRU and FC layer.

Type	Number of Cells	Output Size
Input	–	128×8×8
GRU	256	$L_t \times 256$
FC layer	$L_t \times 256$	64×1

domains (such as two different datasets) to the same feature space, so that target domain can be enhanced by the data from other domains. Due to the problem of multi-source domain adaptation in cross-topic EEG emotion analysis, we introduced domain specific feature extractor (DSFE) and domain specific classifier (DSC) from [13] to solve cross-subject emotion classification problem. Through TFEM and SFEM, two feature vectors with 64 dimensions can be obtained, and the two vectors are spliced which is the input of DACM. The structure of DACM is shown in Fig. 11. DACM contains two sub-modules, which are  $DSFE_{n_p}$  and  $DSC_{n_p}$  consisted of three FC layers.  $DSFE_{n_p}$  has three FC layers whose input dimension is 128 and output dimension is 32.  $DSC_{n_p}$  also has three FC layers whose input dimension is 32 and output dimension is 3.  $DSFE_{n_p}$  and  $DSC_{n_p}$  are represented as follows

$$y_{n_p}^{n_s} = \Phi_{n_p}(x_{n_p}^{n_s}) \quad (12)$$

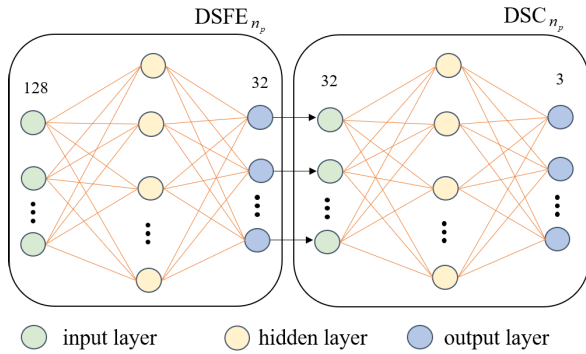
$$Y_{n_p}^{n_s} = \phi_{n_p}(y_{n_p}^{n_s}) \quad (13)$$

where  $\Phi_{n_p}$  represents the  $DSFE_{n_p}$ ,  $x_{n_p}^{n_s}$  is a vector with a length of 128,  $y_{n_p}^{n_s}$  is a vector with a length of 32,  $\phi_{n_p}$  represents the  $DSC_{n_p}$  and  $Y_{n_p}^{n_s}$  is the output of  $\phi_{n_p}$ , consisting of a vector of length 3.

In order to obtain the individual feature of each source domain, 14  $DACM_{n_p}$  ( $n_p=1,2,\dots,14$ ) are constructed, 14  $DSFE_{n_p}$  are to extract the corresponding 14 source domain features. At the same time, in order to map the source domain and corresponding target domain to the same space, the target domain  $x_{15}^{n_s}$  as input to train 14  $DSFE_{n_p}$ , as shown in Figure 3. 14  $DSC_{n_p}$  are used to provide classification results for the 14 source domain.

#### E. LOSS FUNCTION DESIGN AND OPTIMIZATION

In order to estimate the distance between the source domain and the target domain,  $L_{mmd}$  [40] is used to align the



**FIGURE 11.** The structure of DACM. DACM module contains two sub-modules, which are  $DSFE_{n_p}$  and  $DSC_{n_p}$  consisted of three FC layers. The input dimension of  $DSFE_{n_p}$  is 128, and the output dimension is 32. The input dimension of  $DSC_{n_p}$  is 32, and the output dimension is 3.

specific source domain to the target domain gradually without changing the spatial distribution of the target domain.  $L_{mmd}$  are represented as follows

$$L_{mmd}(x_{n_p}^{n_s}, x_{15}^{n_s}) = \frac{\sum_{n_p=1}^{14} \|\Phi_{n_p}(x_{n_p}^{n_s}) - \Phi_{n_p}(x_{15}^{n_s})\|}{14} \quad (14)$$

where the source domain feature  $x_{n_p}^{n_s}$  and target domain feature  $x_{15}^{n_s}$  are extracted by  $\Phi_{n_p}$ .

The classification results of the specified domain are obtained through the softmax classifier. The final result are composed of 14 outputs of  $DSC_{n_p}$ . In order to make the prediction convergence of each classifier result, cross-entropy loss  $L_{cls}$  is introduced to solve the prediction result loss, which is shown as follows

$$L_{cls} = - \sum_{c=1}^M \hat{Y}_{n_p}^{n_s} \log(Y_{n_p}^{n_s}) \quad (15)$$

where,  $M$  represents the total number of sample labels,  $\hat{Y}_{n_p}^{n_s}$  represents the sample label value.

Finally, if the target domain is not controlled and the prediction is simply averaged to the final result, the variance will be high. In order to make the gap between 14 target domains features in  $DSFE_{n_p}$  not too large, a method  $L_{disc}$  [13] is introduced to calculate the sum of feature differences of 14 independent target domains to control the feature distribution of the target domains. The equation of  $L_{disc}$  is as follows

$$L_{disc} = \frac{1}{14} \sum_{p=1, p' \neq n_p}^{14} |\Phi_p(x_{15}^{n_s}) - \Phi_{p'}(x_{15}^{n_s})| \quad (16)$$

The final loss is as follows

$$L = \epsilon L_{cls} + \mu L_{mmd} + \eta L_{disc} \quad (17)$$

where  $\epsilon$ ,  $\mu$  and  $\eta$  are three hyperparameters for the three losses, minimizing  $L_{mmd}$  can get domain-invariant features for each pair of the source and target domains, minimizing  $L_{cls}$  will bring more accurate classifiers for predicting the source domain data and minimizing  $L_{disc}$  will get more convergent multiple classifiers [13].

### Algorithm 1 Training Framework Based on the Proposed Method

**Input:** EEG features of 15 subjects  $S = \{r_1^{n_s}, r_2^{n_s}, \dots, r_{15}^{n_s}\}$  and labels  $\{\hat{Y}_1^{n_s}, \hat{Y}_2^{n_s}, \dots, \hat{Y}_{14}^{n_s}\}$

**Output:** forecast results  $\{Y_1^{n_s}, Y_2^{n_s}, \dots, Y_{14}^{n_s}\}$

```

1: Train:
2: for  $j$  in range(1,15) do
3:    $R_j^{n_s} = \text{FS}(r_j^{n_s}), q_j^{n_s} = \text{ETIM}(r_j^{n_s}),$ 
4:    $x_j^{n_s} = \text{Contact}(\text{TFEM}(R_j^{n_s}), \text{SFEM}(q_j^{n_s}))$ 
5:   if  $j \neq 15$  then
6:      $y_j^{n_s} = \text{DSFE}_j(x_j^{n_s}),$ 
7:      $y_{15}^{n_s} = \text{DSFE}_j(x_{15}^{n_s})$ 
8:      $Y_j^{n_s} = \text{DSC}_j(y_j^{n_s})$ 
9:      $L_{cls}(Y_j^{n_s}, \hat{Y}_j^{n_s})$ 
10:  end if
11:  if  $j = 15$  then
12:    for  $k$  in range(1,14) do
13:       $L_{mmd}(y_k^{n_s}, y_{15}^{n_s}), L_{disc}(x_{15}^{n_s})$ 
14:    end for
15:  end if
16: end for
17: Test:
18:  $R_{15}^{n_s} = \text{FS}(r_{15}^{n_s}), q_{15}^{n_s} = \text{ETIM}(r_{15}^{n_s}),$ 
19:  $x_{15}^{n_s} = \text{Contact}(\text{TFEM}(R_{15}^{n_s}), \text{SFEM}(q_{15}^{n_s}))$ 
20: for  $z$  in range(1,14) do
21:    $y_z^{n_s} = \text{DSFE}_z(x_{15}^{n_s}),$ 
22:    $Y_z^{n_s} = \text{DSC}_z(y_z^{n_s})$ 
23: return  $Y_z^{n_s}$ 
24: end for
25: save model

```

## V. EXPERIMENTS

In order to evaluate MASTF-net, different methods for validation are used on the SEED and SEED-IV. Model evaluation is mainly carried out through multi-source learning in the cross-subject situation. Using this evaluation method, one subject is used as the target domain data, and the other subjects are used as the source domain data. During training, the model is trained through all source domain data, and the target domain data is evaluated. In this way, all subjects are evaluated repeatedly and the corresponding mean value and standard deviation are calculated. At the same time, in the ablation experiment, the results are compared using DE features and PSD features as inputs, which is shown in Fig. 12.

### A. IMPLEMENTATION DETAILS

In the experiment process, because of the use of dual branch network,  $L_t$  must be consistent for SFEM and TFEM.

When converting EEG features into images through ETIM, the dimension of the image is set to  $32 \times 32$ . In the training phase, the cross entropy loss function is mainly used.  $\epsilon$  is set to 1,  $\mu$  is set to 0.3 and  $\eta$  is set to 0.03 during training. The Adam optimizer is used with a learning rate  $lr$  of 0.01.  $output_{SFEM}$  and  $output_{TFEM}$  represent the feature

**TABLE 4.** The hyperparameter values in the experiment in this paper.

parameters	Value/Initially Value
$\epsilon$	1
$\mu$	0.3
$\eta$	0.03
$L_t$	32
$lr$	0.01
$output_{TFEM}$	$1 \times 64$
$output_{SFEM}$	$1 \times 64$
batch size	32
epoch	3500

output dimensions of the SFEM module and TFEM module, respectively. During the experiment, we find that a significant amount of time and memory consumption is required for converting DE features into image data during data reading. Therefore, in order to improve experimental efficiency, the DE features of all subjects are preprocessed into images in advance and images are saved in *numpy* format. Then, the image file is directly read to reduce time. The specific parameters are shown in Table 4.

A training model is built based on the Python library, and EEG features is stored in the *numpy* format before training, which is executed on the 8 GB Nvidia GeForce RTX 3060 GPU. The main algorithm flow is shown in Algorithm 1.  $S$  represents the DE features of 15 subjects. The first fourteen subjects are labeled source domain data, and the fifteenth subject is unlabeled target domain data used for testing. Firstly, feature selection and image transformation are applied to  $r_{n_p}^{n_s}$  for obtaining  $R_{n_p}^{n_s}$  and  $q_{n_p}^{n_s}$ . Then,  $R_{n_p}^{n_s}$  and  $q_{n_p}^{n_s}$  are separately input into two branches (TFEM and SFEM) to extract common features. Contact indicates the concatenation of the common features extracted from the TFEM and SFEM branches along the second dimension, resulting in  $x_{n_p}^{n_s}$ . When the input is labeled test data, the model is trained using both the source domain data and the target domain data to calculate the classification loss ( $L_{cls}$ ) for each domain branch. However, when the input is unlabeled test data, the feature difference loss ( $L_{disc}$ ) between the target domain of all branches are used, as well as the domain loss ( $L_{mmd}$ ) between the source domain and the target domain are calculated. During testing, the common features of the test data are first obtained following the training process. Then, the results are sequentially predicted by passing through the domain adaptation branches, returning all the prediction results of the domain adaptation branches. Finally, the average of the final predictions is obtained.

## B. EXPERIMENTAL RESULTS AND COMPARISON

A lot of emotion classification experiments are performed on the SEED and SEED-IV [6], and MASTF-net is compared with other models.

As shown in Table 5, we evaluate the performance of each participant based on SEED and SEED-IV. Each participant's

three sessions are divided into Session I, Session II and Session III.

The model in the paper is compared with other models which are DDC [41], DAN [13], DCORAL [42], MS-MDA [13], IAG [43], GMSS [44], MSFR-GCN [18]. By comparing MASTF-net with other models, it can be seen that MASTF-net is superior to other methods in both the experimental results for each session and average results, which is shown in Table 5.

In our study, experiments are conducted by using four different methods: DDC, DAN, DCORAL, and MS-MDA, for various conversations. The results for IAG, GMSS, and MSFR-GCN are directly cited from the literature. Notably, due to the use of a large sliding window for feature extraction in SEED-IV, there was a severe lack of training samples in the experiments. Therefore, we re-extract the DE features of five frequency bands from SEED-IV using a sliding window of length 1.

From the results in Table 5, it can be seen that MASTF-net achieved the best performance on both the SEED and SEED-IV datasets, surpassing MS-MDA and MSFR-GCN by 1.33% and 1.41% respectively on the SEED dataset. Simultaneously, on the SEED-IV dataset, our method outperformed MS-MDA, indicating that MASTF-net combining global spatial information and temporal feature information has a significant impact on emotion classification results. Compared to other models, MASTF-net's performance on SEED-IV is similar to GMSS, IG, and MSFR-GCN, but MASTF-net exhibits relatively smaller bias during training, which shows that, compared to others, MASTF-net demonstrates competitive performance and improved stability, with lower standard deviation and stronger generalization ability.

Different hyperparameters are tested to evaluate the MASTF-net and MASTF-net is compared with other models.

Multiple comparative experiments are conducted with different sample length  $L_t$ . The experimental results indicate that the larger the  $L_t$ , the better the TFEM module can learn time-based features from DE features, which is shown in Table 6. TFEM and SFEM cannot learn the key knowledge to be used from a short time series. In addition, it can be seen from Table 6 that as the batch size increases, the performance of MASTF-net decreases. When the  $L_t$  value is 4 and the batch size value is 128, model results have a significant decrease. When the  $L_t$  value is 8 and the batch size is adjusted to 32, the MASTF-net can achieve a highest average accuracy of 88.23%. The results also show that our method can still obtain better experimental results when the number of iterations is small, which indicates that our method can improve the classification accuracy of emotion recognition for multiple subjects in different scenarios.

As shown in Table 7, MASTF-net and other models are trained 1500, 2500, and 3500 times, respectively. It can be observed that with the increase of time, the accuracy of all models has also improved. From Table 7, it can be seen that MASTF-net can achieve the highest testing

**TABLE 5.** Comparison of different models on SEED and SEED-IV based on different sessions.

Datasets	Models	Session I	Session II	Session III	Session Average
SEED	DDC	65.43±8.15	67.99±7.63	66.97±7.93	66.80±7.98
	DAN	64.43±5.91	65.06± 8.14	66.51±8.14	65.34±8.05
	DCORAL	63.70±8.69	64.89±9.86	66.88±8.54	65.16±9.14
	MS-MDA	86.67±8.97	85.01±10.75	88.83±7.85	86.84±8.54
	IAG	—	—	—	86.30±6.91
	GMSS	—	—	—	86.52±6.22
	MSFR-GCN	—	—	—	86.78±5.40
MASTF-net	90.16±5.75	86.97±6.01	87.47±7.87	88.19±6.76	
SEED-IV	DDC	43.25±7.21	47.95±7.39	49.23±8.83	46.81±7.90
	DAN	35.73±6.35	38.28± 7.28	41.27±7.98	38.42±7.18
	DCORAL	42.31±7.90	39.85±10.61	41.35±9.71	41.17±9.40
	MS-MDA	67.25±5.24	62.73±9.37	63.71±8.39	64.56±7.67
	IAG	—	—	—	—
	GMSS	—	—	—	73.48±7.41
	MSFR-GCN	—	—	—	73.43±7.32
MASTF-net	72.57±6.33	73.88±7.51	74.08±6.85	73.51±6.89	

**TABLE 6.** Evaluation of MASTF-net with different batch size and Sample length  $L_t$  on SEED.

Times/ $L_t$	4	6	8
32	76.62±5.41	80.38±10.41	88.23±7.22
64	75.86±12.64	75.15±8.49	80.33±8.50
128	75.23±8.13	76.45±10.27	78.61±9.56

**TABLE 7.** Comparison of different models on SEED and SEED-IV based on different epochs.

Datasets	Models	1500	2500	3500
SEED	DDC	59.03±9.41	62.12±7.97	67.18±6.23
	DAN	62.39±8.19	66.33±8.05	66.81±8.56
	DCORAL	59.03±9.55	62.12±7.97	66.29±7.56
	MS-MDA	78.08±9.46	80.05±7.58	87.23±8.71
	MASTF-net	80.62±7.64	85.38±9.41	88.16±7.38
SEED-IV	DDC	40.68±10.41	42.29±7.75	46.26±8.83
	DAN	35.39±7.71	38.33±9.32	36.05±7.39
	DCORAL	38.05±7.05	38.73±6.68	40.25±9.87
	MS-MDA	60.24±9.35	60.25±10.35	64.71±7.39
	MASTF-net	68.25±7.25	72.38±6.51	73.35±7.27

accuracy. This also indicates that the MASTF-net has a more significant performance improvement in emotion recognition by use of EEG. During the experiment, we saved the results of each experiment. We used bar charts and confusion matrices to record the performance results of all participants in all sessions, as shown in Fig. 13. According to the experimental results, MASTF-net can effectively predict positive and negative emotions, but the predictive effect of neutral emotions not as good as that of positive and negative emotions, especially in the second session. It may be that the neutral mood change is relatively weak relative to the other two moods, thus reducing the average accuracy.

## VI. DISCUSSION

In this section, we investigate the proposed algorithm MASTF-net. We conduct ablation experiments on MASTF-net, including ablation experiments on TFEM and SFEM without using domain loss function, as well as ablation experiments on different frequency band data. Additionally, we analyze the performance of MASTF-net when using DE features and PSD features as inputs separately.

### A. ABLATION STUDY

In order to better understand whether each module can play a key role, the results of the ablation experiment are shown in Table 8.

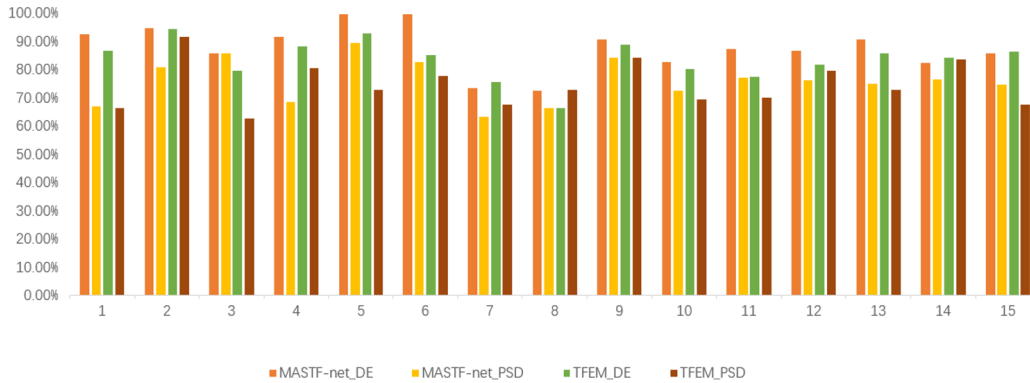
The TFEM module only learn the DE features of eight key regions, but some electrode positions are ignored by the TFEM module. In order to prove that the SFEM module compensates for the shortcomings of TFEM in the global space, in the ablation experiment, TFEM replaces the feature extraction module with a dual branch structure and is compared with the MASTF-net. The experimental results indicate that the SFEM module improves the overall experimental effect.

In order to prove that the domain loss function and difference loss are effective, in the ablation experiment, we removed these two loss functions and only used the cross-entropy loss to predict the results. The experimental results show that if only one prediction loss is used and the domain loss and difference loss are abandoned, the accuracy of the final prediction results will decline. The experimental results are shown in Table 8.

According to the results in Table 8, the testing accuracy decreases significantly without  $L_{mmd}$  and  $L_{disc}$ , indicating the importance of processing each source domain with DACM modules during the training process. In addition, if TFEM module is only used as a common feature extractor, classification accuracy will reduce, which proves that the

**TABLE 8.** Ablation experiment for Loss function and common feature extraction on SEED.

Models	Loss	batch size/32	batch size/64	batch size/128
TFEM+SFEM+DACM	$L_{cls} + L_{mmd} + L_{disc}$	88.23±7.22	80.33±8.50	78.61±9.56
TFEM+SFEM+DACM	$L_{cls}$	83.87±9.59	78.24±11.64	76.37±8.25
TFEM+DACM	$L_{cls} + L_{mmd} + L_{disc}$	84.65±9.52	81.59±9.31	77.24±7.85
TFEM+DACM	$L_{cls}$	80.25±10.37	74.64±10.79	73.58±10.53



**FIGURE 12.** In order to evaluate the impact of PSD features and DE features, repeated experiments are conducted on PSD features in the same experimental environment. In addition, if the TFEM module is only used as a common feature extractor, the classification accuracy will be reduced compared to the dual branch structure common feature extractor, which proves that the SFEM module compensates for the shortcomings of global spatial features in the TFEM module.

**TABLE 9.** The proportion of frequency bands for TFEM and SFEM on SEED and SEED-IV.

Frequency	SEED		SEED-IV	
	TFEM	SFEM	TFEM	SFEM
$\theta, \alpha, \beta, \gamma$	84.21±5.27	82.61±7.68	70.51±6.46	66.85±7.23
$\delta, \alpha, \beta, \gamma$	81.31±7.75	79.46±9.38	62.33±7.24	63.59±7.35
$\delta, \theta, \beta, \gamma$	80.21±7.36	80.25±9.76	60.75±5.71	62.28±9.31
$\delta, \theta, \alpha, \gamma$	77.39±9.59	76.61±7.83	63.25±5.54	61.59±7.38
$\delta, \theta, \alpha, \beta$	80.21±5.65	78.15±9.35	64.41±7.52	60.54±10.64

SFEM module compensates for the lack of global spatial features in the TFEM module.

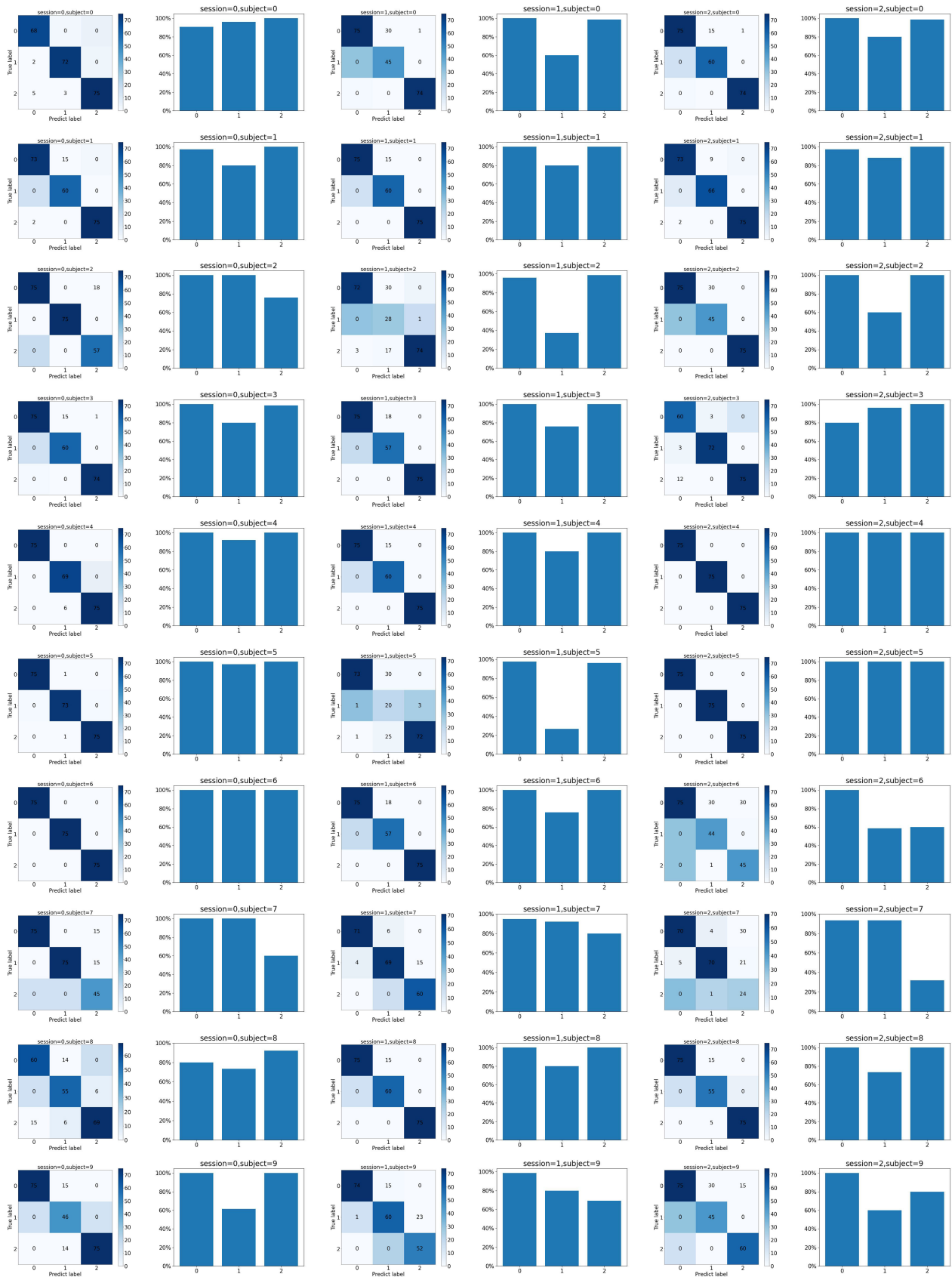
In order to investigate the extraction of low-level common feature information by MASTF-net, we conducted multiple experiments on the data of the five frequency bands, as shown in Table 9. According to Table 9, in seed and seed-IV,  $\delta$  has the least impact on the experiment, while  $\beta$  has a significant impact on the prediction results of both datasets. In SEED-IV, fear, negative and neutral emotions account for about one quarter of the data, and according to studies [18], and the  $\beta$  and  $\gamma$  bands are associated with positive emotions, while the  $\alpha$  band is associated with neutral emotions. Therefore, if high frequency bands are overly emphasized, the overall effect will not be good. Our findings are consistent with the physiological research results of MSFR-GCN [18].

**B. INPUT FEATURE COMPARISON**

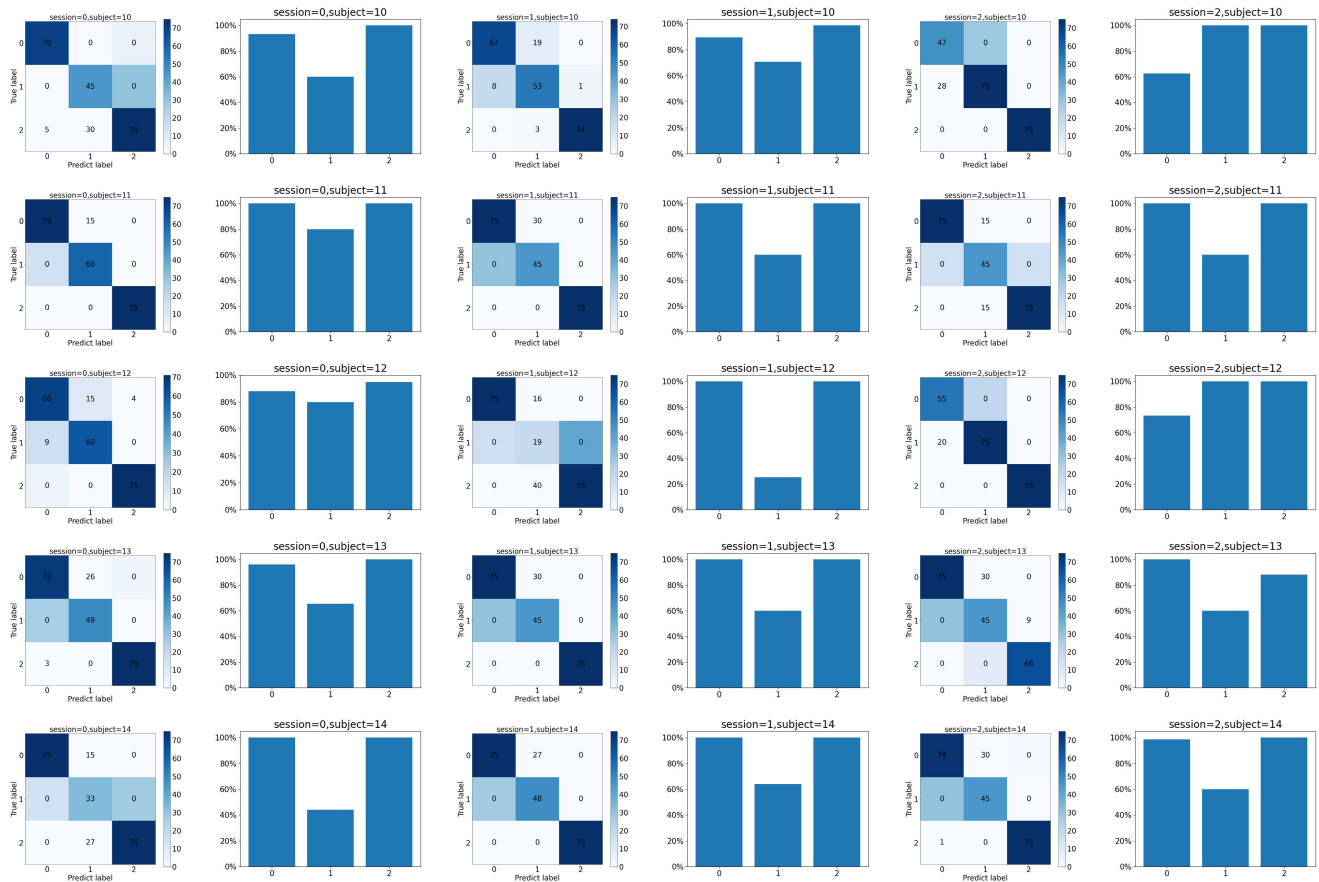
The SEED provides not only DE features but also PSD features. In order to improve the effectiveness of cross-subject emotion classification and evaluate the impact of PSD

features and DE features, repeated experiments are conducted on PSD features in the same experimental environment, and the results are shown in Fig. 12. We randomly select the DE and PSD feature data of all subjects in one session of three sessions. The abscissa in Fig. 12 represents the subject’s number  $n_p$ , and the ordinate represents accuracy. MASTF-net\_DE and MASTF-net\_PSD represent experimental results using DE and PSD features as MASTF-net inputs, respectively. TFEM\_DE and TFEM\_PSD represent using DE and PSD features as network inputs. However, unlike MASTF-net\_DE and MASTF-net\_PSD, in order to reflect ablation experiments, in the common feature extractor, we only use the TFEM module instead of the dual branch feature extraction module to demonstrate the high generalization of the TFEM module. The experimental results show that DE outperforms PSD in SEED under the same model and experimental environment. Therefore, we conclude that for SEED, DE is more conducive to the adaptation of EEG emotion recognition across subjects than PSD. From the experimental results can see that our method has shown good results in the field of adaptive EEG emotion analysis.

Furthermore, MASTF-net fails to achieve satisfactory results in cross-subject sentiment classification when using PSD features as input, as shown in Figure 12. Further optimization is needed to improve the generalization ability of MASTF-net, enabling it to adapt to DE features as network input and achieve satisfactory results when using PSD features as input. Finally, although MASTF-net demonstrates favorable analytical results compared to the comparative methods, its model complexity leads to longer training time. In future work, we will adjust the model structure of



**FIGURE 13.** For the 45 cross-subject emotion classification results of each subject in each conversation as the target domain, the confusion matrix represents the number of true and false predictions of each type, and the histogram represents the accuracy of each type of emotion prediction.



**FIGURE 13. (Continued.)** For the 45 cross-subject emotion classification results of each subject in each conversation as the target domain, the confusion matrix represents the number of true and false predictions of each type, and the histogram represents the accuracy of each type of emotion prediction.

MASTF-net based on the weighted proportion of different frequency bands in emotion prediction to enhance its efficiency.

## VII. CONCLUSION AND FUTURE WORK

In the paper, MASTF-net is compared with other models which are DDC [41], DAN [13], DCORAL [42], MS-MDA [13], and it is clear that MASTF-net has significant improvement in the emotional classification of cross-subjects. A dual branch network are used to build a shared network to extract the common time, space, and frequency information of EEG signals. On one hand, L-RNN module is used to obtain local area features of EEG in high-frequency active areas. On the other hand, 3D EEG data is mapped to a 2D EEG topographic map to extract EEG spatial global features.

In addition, in experiments, it can be seen that length of  $L_t$  have a significant impact on the cross-subject emotion classification results of the experiment. As can be seen from Table 6, when  $L_t$  is set to 8, the best experimental results are obtained, which shows that EEG signals have a relatively long time correlation. We also find that MASTF-net can achieve higher accuracy when choose a smaller batch number. In addition, when selecting the loss function, three aspects are considered. First, inter-domain loss is used to

calculate the difference between the source domain and the corresponding target domain. Second, domain invariant loss is used to improve the stability of the target domain after feature extraction. Third, prediction loss is used to predict the classification results.

Finally, the MASTF-net uses a relatively complex network structure. Compared with other simple architecture models, the MASTF-net has good performance in cross-subject emotion classification, but has a problem of longer training time. It can see that the training time increases linearly with the number of source domains, that is, the larger the number of source domains, the larger the model, and the longer the training time. In the next step, our goal is to improve the training efficiency without reducing the model accuracy. The main idea is to add a discriminator in the branches for domain features to discard unnecessary branches for reducing training time.

## REFERENCES

- [1] J. Dolan, "Emotion, cognition, and behavior," *Science*, vol. 298, no. 5596, pp. 1191–1194, Nov. 2002.
- [2] T. M. Chai, U. A. Hafeez, M. N. M. Saad, and A. S. Malik, "The influences of emotion on learning and memory," *Frontiers Psychol.*, vol. 8, p. 1454, Aug. 2017.

- [3] H. Ke, D. Chen, B. Shi, J. Zhang, X. Liu, X. Zhang, and X. Li, "Improving brain e-health services via high-performance EEG classification with grouping Bayesian optimization," *IEEE Trans. Services Comput.*, vol. 13, no. 4, pp. 696–708, Jul. 2020.
- [4] J. Joormann and I. H. Gotlib, "Emotion regulation in depression: Relation to cognitive inhibition," *Cognition Emotion*, vol. 24, no. 2, pp. 281–298, Feb. 2010.
- [5] A. Radford, "Emotion processing in Alzheimer's disease," *Neurobiol. Aging*, vol. 21, no. 3, pp. 222–232, 2000.
- [6] 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.
- [7] S.-H. Lee, M. Lee, J.-H. Jeong, and S.-W. Lee, "Towards an EEG-based intuitive BCI communication system using imagined speech and visual imagery," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 4409–4414.
- [8] A. Frisoli, C. Loconsole, D. Leonardi, F. Banno, M. Barsotti, C. Chisari, and M. Bergamasco, "A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 42, no. 6, pp. 1169–1179, Nov. 2012.
- [9] E. Yildirim, Y. Kaya, and F. Kiliç, "A channel selection method for emotion recognition from EEG based on swarm-intelligence algorithms," *IEEE Access*, vol. 9, pp. 109889–109902, 2021.
- [10] 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.
- [11] B. Zhang, G. Yan, Z. Yang, Y. Su, J. Wang, and T. Lei, "Brain functional networks based on resting-state EEG data for major depressive disorder analysis and classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 215–229, 2021.
- [12] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal," *Comput. Methods Programs Biomed.*, vol. 109, no. 3, pp. 339–345, Mar. 2013.
- [13] C. Hao, J. Ming, and L. Zhunan, "MS-MDA: Multisource marginal distribution adaptation for cross-subject and cross-session EEG emotion recognition," *Frontiers Neurosci.*, vol. 15, p. 778488, 2021.
- [14] Y.-M. Jin, Y.-D. Luo, W.-L. Zheng, and B.-L. Lu, "EEG-based emotion recognition using domain adaptation network," in *Proc. Int. Conf. Orange Technol. (ICOT)*, Dec. 2017, pp. 222–225.
- [15] H. Li, Y.-M. Jin, W.-L. Zheng, and B.-L. Lu, "Cross-subject emotion recognition using deep adaptation networks," in *Proc. Int. Conf. Neural Inf. Process. Cham, Switzerland: Springer*, 2018, pp. 403–413.
- [16] L. Mingsheng, C. Yue, W. Jianmin, and J. Michael, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 97–105.
- [17] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, Jul. 2006.
- [18] D. Pan, H. Zheng, F. Xu, Y. Ouyang, Z. Jia, C. Wang, and H. Zeng, "MSFR-GCN: A multi-scale feature reconstruction graph convolutional network for EEG emotion and cognition recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3245–3254, 2023.
- [19] W. Yin, H. Schütze, B. Xiang, and B. Zhou, "ABCNN: Attention-based convolutional neural network for modeling sentence pairs," *Trans. Assoc. Comput. Linguistics*, vol. 4, pp. 259–272, Dec. 2016.
- [20] M. A. Rahman, A. Anjum, M. M. H. Milu, F. Khanam, M. S. Uddin, and M. N. Mollah, "Emotion recognition from EEG-based relative power spectral topography using convolutional neural network," *Array*, vol. 11, Sep. 2021, Art. no. 100072.
- [21] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "MPED: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12177–12191, 2019.
- [22] Y. Zhao, C. Dong, G. Zhang, Y. Wang, X. Chen, W. Jia, Q. Yuan, F. Xu, and Y. Zheng, "EEG-based seizure detection using linear graph convolution network with focal loss," *Comput. Methods Programs Biomed.*, vol. 208, Sep. 2021, Art. no. 106277.
- [23] H. Cai, X. Liu, A. Jiang, R. Ni, X. Zhou, and A. Cangelosi, "Combination of EOG and EEG for emotion recognition over different window sizes," in *Proc. IEEE 2nd Int. Conf. Human-Machine Syst. (ICHMS)*, Sep. 2021, pp. 1–6.
- [24] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.
- [25] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.
- [26] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 2, pp. 344–353, Jun. 2020.
- [27] A. Gasparin, S. Lukovic, and C. Alippi, "Deep learning for time series forecasting: The electric load case," *CAAI Trans. Intell. Technol.*, vol. 7, no. 1, pp. 1–25, Mar. 2022.
- [28] Y. Du, W. Wang, and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1110–1118.
- [29] P. Zhang, X. Wang, W. Zhang, and J. Chen, "Learning spatial-spectral-temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 1, pp. 31–42, Jan. 2019.
- [30] J. Donahue and L. A. Hendricks, "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2625–2634.
- [31] M. Suchetha, R. Madhumitha, M. M. Sorna, and R. Sruthi, "Sequential convolutional neural networks for classification of cognitive tasks from EEG signals," *Appl. Soft Comput.*, vol. 111, Nov. 2021, Art. no. 107664.
- [32] R. T. Schirrmeyer, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggenberger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.
- [33] E. Q. Wu, P. Xiong, Z.-R. Tang, G.-J. Li, A. Song, and L.-M. Zhu, "Detecting dynamic behavior of brain fatigue through 3-D-CNN-LSTM," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 1, pp. 90–100, Jan. 2022.
- [34] V. Delvigne, A. Facchini, H. Wannous, T. Dutoit, L. Ris, and J.-P. Vandeborre, "A saliency based feature fusion model for EEG emotion estimation," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 3170–3174.
- [35] S. McKinley and M. Levine, "Cubic spline interpolation," *College Redwoods*, vol. 45, no. 1, pp. 1049–1060, 1998.
- [36] S. Siddharth, T.-P. Jung, and T. J. Sejnowski, "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 96–107, Jan. 2022.
- [37] A. Ullah, H. Elahi, Z. Sun, A. Khatoun, and I. Ahmad, "Comparative analysis of AlexNet, ResNet18 and SqueezeNet with diverse modification and arduous implementation," *Arabian J. Sci. Eng.*, vol. 47, no. 2, pp. 2397–2417, Feb. 2022.
- [38] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2017, pp. 1597–1600.
- [39] K. You, M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Universal domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2715–2724.
- [40] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5018–5027.
- [41] Y. Liu, A. Liu, N. Zhang, X. Liu, M. Ma, and Y. Hu, "DDC: Dynamic duty cycle for improving delay and energy efficiency in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 131, pp. 16–27, Apr. 2019.
- [42] S. Baochen and S. Kate, "Deep CORAL: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer*, 2016, pp. 443–450.
- [43] T. Song, S. Liu, W. Zheng, Y. Zong, and Z. Cui, "Instance-adaptive graph for EEG emotion recognition," in *Proc. AAAI*, 2020, pp. 2701–2708.
- [44] Y. Li, J. Chen, F. Li, B. Fu, H. Wu, Y. Ji, Y. Zhou, Y. Niu, G. Shi, and W. Zheng, "GMSS: Graph-based multi-task self-supervised learning for EEG emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 2512–2525, Jul. 2023.

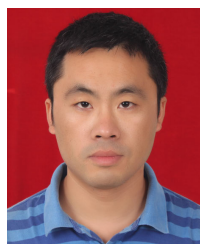




**HONGXIANG XU** received the B.S. degree from Sanjiang University, Jiangsu, China, in 2020. He is currently pursuing the master's degree with the Chongqing University of Science and Technology. His research interests include deep learning, brain-computer interface, and medical image analysis.



**ZIYI PEI** is currently pursuing the degree in material forming and control engineering with the Chongqing University of Arts and Sciences. His current research interests include artificial intelligence and deep learning.



**QI HAN** received the B.S. degree in computer science and technology from Shandong University, China, in 2005, and the M.S. and Ph.D. degrees from Chongqing University, China, in 2009 and 2012, respectively. He is currently an Associate Professor with the Chongqing University of Science and Technology. His current research interests include artificial intelligence, system optimization, neural networks, and chaos control.



super-resolution, and medical image processing.

**MINGYANG HOU** received the B.S. degree from the School of Information Engineering, Institute of Disaster Prevention, Hebei, China, in 2018, and the M.S. degree from the School of Intelligent Engineering, Chongqing University of Science and Technology, Chongqing, China, in 2023, where he is currently pursuing the Ph.D. degree. He has published several articles in reputable journals, including *Computers in Biology and Medicine*. His research interests include deep learning, image



**XIN QIAN** received the bachelor's degree in computer science from Jinggangshan University, China, in 2021. He is currently pursuing the master's degree with the Chongqing University of Science and Technology. His research interests include machine learning and medical image analysis.



**TENGFEI WENG** received the B.S. and M.S. degrees in chemistry and chemical engineering from Chongqing University, China, in 2007 and 2010, respectively. She is currently with the Chongqing University of Science and Technology. Her current research interests include artificial intelligence and neural networks.



impulsive systems, discontinuous dynamical systems, and multi-agent systems.

**YUAN TIAN** received the B.S. degree in computer science and technology from Chongqing Normal University, Chongqing, China, in 2009, the M.S. degree in computer application technology from Chongqing University, Chongqing, in 2012, and the Ph.D. degree in computational intelligence and information processing from Southwest University, Chongqing, in 2020. She is currently with the Chongqing University of Science and Technology, Chongqing. Her current research interests include



he was an Assistant Professor with the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China. From 2015 to 2020, he was an Associate Professor with the College of Information Engineering, Tarim University, Alar, Xinjiang, China. He is currently an Associate Professor with the School of Intelligent Technology and Engineering. He is the author of more than ten articles and three inventions. His research interests include AI, intelligent speech technologies, and the security of AI.

**ZICHENG QIU** was born in Wuhan, Hubei, China, in 1983. He received the B.S. degree in optoelectronic engineering from the Huazhong University of Science and Technology, Wuhan, in 2005, and the Ph.D. degree in optical engineering (advanced lithography technologies) from the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai, China, in 2010. From 2010 to 2011, he was a Software Engineer with Synopsys. From 2011 to 2015,



Research and Development Program, the National Natural Science Foundation of China, and the National Natural Science Foundation of Xinjiang Joint Key Project. He is the author of seven books and more than 50 articles. His research interests include applied mathematics, natural speech and language intelligent technologies, agricultural information processing and computer intelligent management, and agricultural information technologies. He is currently a member of the Xinjiang Organizing Committee of China Undergraduate Mathematical Contest in Modeling and a member of the "Smart Ecology" Professional Technical Group of the Internet of Things Committee of the Chinese Communications Society. One of his books was awarded the 12th Five-Year National Planning Textbook. He was a recipient of the Excellent Talent of XPCC, in 2013, and the Bronze XPCC Talent, in 2014, for his research work in Southern Xinjiang, China. He is also the Director of the Xinjiang Uyghur Autonomous Region Mathematical Society and the Physics Society.

**BAOBING ZHOU** received the B.S. degree in mathematics from Kashgar Teachers College, in 1995, and the M.S. degree in agricultural electrification and automation from Northeast Agricultural University, in 2004. Since 2010, he has been a Professor with the College of Information Engineering, Tarim University. He has been the person in charge of more than ten national and Xinjiang Production and Construction Corps (XPCC) projects, which include the National Key