Emotion Recognition from Few-Channel EEG Signals by Integrating Deep Feature Aggregation and Transfer Learning

Fang Liu, Pei Yang, Yezhi Shu, Niqi Liu, Jenny Sheng, Junwen Luo, Xiaoan Wang, and Yong-Jin Liu, *Senior Member, IEEE*

Abstract—Electroencephalogram (EEG) signals have been widely studied in human emotion recognition. The majority of existing EEG emotion recognition algorithms utilize dozens or hundreds of electrodes covering the whole scalp region (denoted as full-channel EEG devices in this paper). Nowadays, more and more portable and miniature EEG devices with only a few electrodes (denoted as few-channel EEG devices in this paper) are emerging. However, emotion recognition from few-channel EEG data is challenging because the device can only capture EEG signals from a portion of the brain area. Moreover, existing full-channel algorithms cannot be directly adapted to few-channel EEG signals due to the significant inter-variation between full-channel and few-channel EEG devices. To address these challenges, we propose a novel few-channel EEG emotion recognition framework from the perspective of knowledge transfer. We leverage full-channel EEG signals to provide supplementary information for few-channel signals via a transfer learning-based model *CD-EmotionNet*, which consists of a base emotion model for efficient emotional feature extraction and a cross-device transfer learning strategy. This strategy helps to enhance emotion recognition performance on few-channel EEG data by utilizing knowledge learned from full-channel EEG data. To evaluate our cross-device EEG emotion transfer learning framework, we construct an emotion dataset containing paired 18-channel and 5-channel EEG signals from 25 subjects, as well as 5-channel EEG signals from 13 other subjects. Extensive experiments show that our framework outperforms state-of-the-art EEG emotion recognition methods by a large margin.

Index Terms—Emotion recognition, few-channel EEG, transfer learning, cross-device.

1 Introduction

EEG-based emotion recognition has attracted considerable attention in the past decades. This technique is widely used in human-computer interaction (HCI) [1] scenarios. It can also help in the treatment of mental health conditions like depression, anxiety, etc, where the EEG signals are usually used to monitor people's emotional changes [2].

In this paper, we refer to the EEG devices with dozens or even hundreds of electrodes and the corresponding signals as *full-channel* EEG devices and signals, e.g., ESI NeuroScan System[1] used in [3] has 62 electrodes, Biosemi ActiveTwo System[2] used to construct DEAP [4] has 32 electrodes, and Wearable Sensing's wireless DSI-24 [3] has 18 electrodes. However, these devices are usually non-portable, expensive, and are mainly applicable to lab experiment scenarios, which limits the expansion of existing works to more general daily applications. Nowadays, more and more portable and miniature EEG devices with only a few electrodes are emerging, with some interesting research reported on emotion recognition [5], [6] and depression detection [7], [8], [9] using EEG signals from a few electrodes. We denote EEG devices with only a few electrodes as *few-channel* EEG devices in this paper, e.g., EMOTIV EPOC^{+}[4] has 14 channels [5], OpenBCI[5] has 8 channels [6], InteraXon MUSE[6] has 4 channels [10] and SleepUp2.0 Naolubrain has 5 channels [11]. While these few-channel EEG devices bring convenience to a variety of applications, they also make EEG-based emotion recognition more challenging, as fewer electrodes capture a lower density of EEG signals. More specifically, almost all few-channel EEG devices provide electrodes that cover only a portion of the brain area. Previous studies have shown that emotions are not a function of a single brain region, but rather exist in the activation or connectivity patterns of a whole-brain network [12], [13]. This means that emotion recognition using only part of the brain's EEG signals is more difficult than using all of them.

Most existing emotion recognition algorithms, including those based on handcrafted features (e.g., power spectral density (PSD), differential entropy (DE) [14], etc) and deep learning techniques (e.g., convolutional neural networks (CNN) [15], recurrent neural networks (RNN) [16], [17], graph convolutional neural networks (GCN) [18], [19], etc.), are designed to extract emotional patterns from the whole

- *F. Liu, P. Yang, Y. Shu, N. Liu, J. Sheng and Y.-J. Liu are with BNRist, the Department of Computer Science and Technology, Tsinghua University, and MOE-Key Laboratory of Pervasive Computing, Beijing 100084, China. E-mail: {lfang@, yangpei20@mails., shuyz19@mails., lnq22@mails., cqq22@mails., liuyongjin@}tsinghua.edu.cn.*
- *J. Luo and X. Wang are with BrainUp research lab. E-mail: {luojunwen, wangxiaoan}@naolubrain.com.*
- *Y.-J. Liu is the corresponding author.*
- *F. Liu and P. Yang contribute equally to this work.*

1. http://compumedicsneuroscan.com/
2. http://www.biosemi.com
3. https://wearablesensing.com/research/#publications
4. https://www.emotiv.com/epoc/
5. http://openbci.com/
6. https://choosemuse.com/

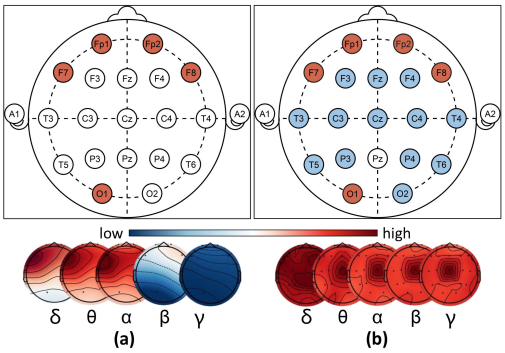


Fig. 1. (a) Few-channel (5-channel) EEG device with corresponding electrodes shown in red. (b) Full-channel (18-channel) EEG device with corresponding electrodes shown in red and light blue. Top: EEG electrode placements according to the International 10-20 system. Bottom: the energy distributions of the same subject in the "disgust" emotional state.

cerebral cortex. We cannot directly apply these full-channel methods on few-channel signals as, on the one hand, it is more challenging to extract efficient emotional patterns from EEG data with very limited sparse channels [20]. On the other hand, the inter-variation between these two types of EEG devices is large. For example, both the full-channel device (Wearable Sensing's wireless DSI-24) and the few-channel device (SleepUp2.0 Naolubrain) used in this study have frontal lobe electrodes F7, Fp1, Fp2, F8, and occipital lobe electrode O1 as shown in Figure 1; however, the distributions of the EEG signals in these channels of the two devices differ significantly. Specifically, the 5 electrodes in the SleepUp2.0 Naolubrain EEG device are Fp1/2, F7/8, and O1, and the electrodes in the 18-channel Wearable Sensing's wireless DSI-24 device are Fp1/2, Fz, F3/4, F7/8, Cz, C3/4, T3/4, A1/2, Pz, P3/4, T5/6, and O1/2, according to the international 10-20 system.

To address the above challenges, we aim to perform the emotion recognition task using few-channel EEG signals from the perspective of knowledge transfer. We propose to enhance the emotion recognition performance from few-channel EEG data with the help of knowledge learned from full-channel EEG data. Specifically, we formulate the full-channel-assisted few-channel EEG emotion recognition task as a transfer learning problem and propose a novel feature-aggregation-based cross-device EEG transfer network called *CD-EmotionNet* for emotion recognition. Our CD-EmotionNet consists of two parts: i) a base emotion model that extracts spatial-spectral-temporal EEG features and performs emotion classification, and ii) a cross-device transfer learning strategy. The base emotion model is first trained using the full-channel EEG dataset. Then, the learned emotional features and model parameters are transferred to the few-channel scenario to help enhance emotion recognition performance from few-channel EEG data.

Our base emotion model is inspired by the Vector of Locally Aggregated Descriptors (VLAD) [21], a successful feature descriptor for image classification(VLAD) that aims to convert local feature descriptors into a compact global representation based on a feature clustering process. Considering that the emotional EEG information has a whole-brain activation pattern, we build our base emotion model by integrating graph operations and a deep aggregation-and-fusion network to extract efficient spatial-temporal EEG features for emotion classification, because the graph model is effective at capturing spatial relationships. Through clustering and aggregating spatial-spectral EEG features in a learned way, our method can capture effective emotion-related temporal patterns in the EEG data. To further address the information sparsity issue caused by the lack of full-channel information, we integrate our base emotion model with a transfer learning strategy to utilize the knowledge from full-channel EEG signals. Specifically, our few-channel signals are collected within the subjects' pre-frontal region, while the full-channel signals are collected with devices that cover the whole brain. The knowledge learned from full-channel EEG signals can offer necessary supplementary information for few-channel signals, as previous studies indicate that all the pre-frontal, parietal, and occipital regions may provide emotion-related information [22].

Our cross-device EEG emotion transfer learning method is built upon the Model-Agnostic Meta-Learning (MAML) [23] framework, which is a type of meta-learning method designed for learning a good parameter initialization and for easy finetuning. By proposing a novel MAML-based network updating mechanism, we aim to learn a shared feature aggregation network between different devices. The core idea of MAML is to train a model on a variety of learning tasks (meta-training stage) so that it can swiftly adapt to new tasks using just a few examples (meta-testing stage). In this paper, both full-channel and few-channel paired EEG signals are leveraged in the meta-training stage to train the network, while only few-channel signals are used in model fine-tuning and testing stages. The most significant difference between the original MAML framework and our method is that MAML updates the same network parameters in both the inner loop (meta-training) and outer loop (meta-testing) stages, while our method sets different learning tasks for the inner and outer loops. In the inner loop, the model focuses on extracting spatial-temporal features from the full-channel EEG data. In the outer loop, the model adapts these features to the target domain with few-channel data. By sharing the feature aggregation and classification layers between loops, our model also incorporates temporal information, which proves to be effective in experiments. In summary, our proposed transfer learning approach integrates the base emotion model with our transfer learning strategy in a collaborative and progressive fashion. By designing separate learning objectives and network parameters for the meta-training and meta-testing stages, our method can better adapt to emotion recognition with few-channel EEG signals and achieve superior performance.

To evaluate our method, we construct a cross-device EEG dataset *CDEED*, which consists of (1) paired data from 25 subjects, where each paired data consists of 5-channel and 18-channel EEG signals collected with the same subjects using the same set of video stimuli and experimental protocol, and (2) 5-channel EEG signals from 13 other subjects (different from the 25 subjects mentioned above). Experimental results show that our CD-EmotionNet model can efficiently enhance the performance of EEG recognition with few-channel signals and achieves state-of-the-art performance on the CDEED dataset and three more publicly available datasets: SEED-IV [24], SEED-V [25] and DEAP [4].

The contributions of this paper are two-fold:

1) We address the challenging few-channel EEG emotion recognition task from the perspective of knowledge transfer. Specifically, we leverage full-channel EEG signals to provide supplementary information for few-channel signals by designing a transfer learning-based model *CD-EmotionNet*, which is comprised of a base emotion model for efficient emotional feature extraction and a cross-device transfer learning strategy that enhances the emotion recognition performance with few-channel EEG data using the knowledge acquired from full-channel EEG data.
2) We constructed a paired full-channel and few-channel EEG signal dataset containing EEG of the same subjects using the same set of stimuli and experiment protocol to validate our cross-device EEG emotion transfer network. The proposed dataset can also be set as a benchmark for the few-channel EEG emotion recognition task. Experiments on three other public datasets (SEED-IV, SEED-V, and DEAP) also demonstrate that our proposed CD-EmotionNet can achieve state-of-the-art performances.

2 Related Work

2.1 EEG Emotion Recognition with Deep Learning

PSD and DE of EEG signals are frequently used to distinguish emotions by capturing different characteristics of the brain's electrical activity associated with emotional states, and are usually incorporated with deep features together to recognize emotions. PSD effectively captures how signal power is distributed across different frequency components, and the core concept is to extract EEG features by depicting the energy variation within the signal's frequency domains [26], [27]. DE features are also employed to construct features in multiple frequency bands to measure the dynamics of uncertainty for continuous EEG signals [28].

Recently, various deep neural networks that have been designed for EEG emotion recognition tasks either focus on the design of feature extraction networks or fuse features from different domains. Zhang et al. [29] employs a heuristic Variational Pathway Reasoning (VPR) method based on pathway sampling and salient pathway reasoning. Suh et al. [30] leverages the concept of "separate to learn" to EEG signal analysis by dividing the embedding space into K-subproblems and building a model for each one of them. Thammasan et al. [31] studies the fusion of EEG and music features to identify the arousal and valence of emotion. Lu et al. [32] combines eye movements and EEG to classify emotions and reveal that they are complementary to emotion recognition. Similarly, Li et al. [33] also leverages multi-domain (i.e. temporal, frequency, and topology of channels) EEG information to classify emotions.

Multimodal information has been integrated to enhance emotion recognition. TSIN [34] incorporates temporal and semantic consistency into multimodal emotion recognition tasks. SAMS [35] is a multimodal approach to emotion recognition, achieving local and global alignment between modalities through a multi-spatial learning framework for each modality and a self-modal interaction module for cross-modal semantic learning.

Besides, EEG has also been incorporated with other modalities to conduct emotion recognition. For instance, DREAMER [5] is a multimodal database consisting of EEG and ECG signals from 23 participants, and the authors propose a participant-wise affect recognition method using EEG and ECG features. While leveraging multiple neural or physiological signals as complementary cues to improve emotion recognition performance, none of these approaches consider the variation across devices, which is the focus of our work. Our proposed CD-EmotionNet aims to address the device variation between full-channel and few-channel EEG by transferring knowledge learned from full-channel data to few-channel data.

In addition, another mainstream idea is to exploit the correlations between different channels or emotions by leveraging attention mechanisms. ACRNN [36] introduces an attention mechanism to recurrent neural networks for EEG feature extraction. SST-EmotionNet [37] integrates spatial-spectral-temporal EEG emotion features into an attention 3D dense network. More recently, TransEEG [38] addresses long-distance temporal dependencies, as well as inter-emotion correlations, to refine the emotional EEG feature extraction with a transformer-based architecture. Previous studies demonstrated the effectiveness of applying attention operations on EEG features to learn better emotion-related embeddings. The feature aggregation procedure that learns dynamic feature clusters in our base emotion model can also be seen as a kind of attention mechanism operating on the spatial-spectral EEG features.

GCN is suitable for modeling the spatial and temporal correlations across multiple EEG channels. In the field of applying GCN on EEG-based emotion recognition, existing studies usually model EEG signals collected with each electrode as a node in the graph, and the edge connecting two EEG-electrode nodes represents their correlation. DGCNN [39] explores the relationship between different EEG signal channels and uses graph operations to learn a dynamic adjacency matrix for emotion recognition. Zhang et al. [19] further improves DGCNN by employing a sparse constraint and proposes the SparseDGCNN framework for multichannel EEG analysis. Song et al. [18] presents a multi-level and multi-graph convolutional operation for EEG feature extraction. Other GCN-based EEG studies have also been proposed in [22], [40]. To enhance the robustness of emotion recognition, Li et al. [41] proposes to leverage multiple emotion-related spatial network topology patterns to capture discriminative graph topologies within EEG brain networks. In this paper, the proposed base emotion model is also constructed using a GCN to capture the whole-brain activation patterns of EEG signals and refine the extracted emotional features.

All the above-mentioned methods use full-channel EEG signals. Moreover, the EEG signals in the widely used EEG emotion benchmarks such as SEED [3] and DEAP [4] are also full-channel data, i.e., they are recorded using a 62-channel wet-electrode ESI NeuroScan System and 32 active AgCl electrodes respectively. In contrast, we focus on emotion classification with weaker supervision information

provided by few-channel EEG signals.

2.2 Transfer Learning in EEG Processing

Transfer learning, which aims at applying the knowledge learned from one domain to another different but related domain, has recently drawn increasing attention in the field of EEG signal analysis [42]. According to the amount of target data and whether the target data is labeled, transfer learning methods can be roughly classified into fine-tuning, multitask learning, domain adaptation, zero-shot learning, self-taught learning, self-taught clustering, etc. In EEG-based emotion recognition tasks, transfer learning techniques have been adopted to solve the performance limitations of algorithms caused by inherent EEG signal differences between different subjects [43], EEG signals collection scenarios [44], experiment sessions [45], datasets [46], etc. Among them, most attention has been paid to the *cross-subject* situation, where the main problem is the EEG data distribution domain gap between different individuals. Zheng et al. [47] explores two kinds of cross-subject transfer methods, i.e., network structure sharing and parameters transferring, to improve emotion recognition performance. Zhao et al. [48] divides EEG data into private and shared components and builds a plug-and-play emotion classifier via domain adaptation techniques. Personal-Zscore is proposed in [49] to enhance the robustness of cross-subject EEG emotion recognition models by improving the emotional representation ability for individuals. Li et al. [50] proposes a multisource transfer learning framework to eliminate the impact of personal differences in EEG emotion recognition, where existing people are seen as sources and the new person is the target; the algorithm contains a calibration session and a test session for each new target. Their method is similar to our transfer learning strategy in this paper to some extent; in our case, the full-channel EEG signals of existing people can be seen as the source, and the few-channel EEG signals of new people are set as the target. Moreover, there is also a base emotion model refining stage and a testing stage for new targets.

In the last few years, an increasing number of portable and miniature EEG devices equipped with only a limited number of electrodes are becoming available. Lakhan et al. [6] assesses OpenBCI's potential by comparing its performance to research-grade EEG systems and presenting evidence in support of the applicability of consumer-grade EEG device for emotion recognition research. Furthermore, there has been a growing body of research on depression detection using few-channel EEG signals, typically employing three electrodes [7], [8], [9]. However, very few studies have investigated the *cross-device* task based on transfer learning technology. Liu et al. [51] examines the brain-computer interface (BCI) between wet and dry EEG headsets and utilizes a transfer learning framework to exploit auxiliary wet-electrode EEG to improve the BCI implemented in dry electrodes. Different from our work, the task in [51] is steady-state visual-evoked potential-based BCI (SSVEP-BCI), while our goal is emotion recognition. Similarly, Nakanishi et al. [52] also solves the cross-device scalp-channel EEG calibration problem in SSVEP-BCI using transfer learning algorithms that applied spatial filtering to extract shared features across different devices. In addition, Feng et al. [53] tackles cross-subject and cross-device EEG signal transfer learning simultaneously. They construct two VR-induced EEG emotion datasets and propose a domain adaptation network (MSDAN) for cross-subject and cross-device EEG emotion recognition. However, there are only two emotion categories (positive and negative) in their datasets, and the authors build their transfer network by directly comparing the similarity of data from different domains. Closely related to ours, Xu et al. [54] proposes Armaiti, an EEG analysis model based on 5-channel EEG signals, but they aim to combine multi-modal signals (EEG and EOG) to build a lightweight EEG classification network to reduce the network depth and the training time, while we focus on boosting the performance of few-channel EEG emotion recognition. Few-channel EEG analysis for other mental state recognition also raised increasing attention recently, e.g. depression detection [55]. The MODMA dataset consists of both 128 and 3-channel EEG data for mental-disorder analysis [56]; a multi-modal fusion technique was also designed for depression detection with the MODMA dataset [57], [58]. The EEG devices used to collect few-channel signals in Armaiti is a 5-channel portable device, and the EEG data in MODMA is collected with a 3-channel device. These works solve the mental disorder detection task with few-channel EEG, while we concentrate on emotion recognition in this research.

3 METHOD

3.1 Overview

In order to perform the emotion recognition task using few-channel EEG signals, we need to learn discriminative emotion representation from EEG data with a very limited number of channels. These few channels convey neural activation information from only a portion of the cerebral cortex. Our proposed CD-EmotionNet mainly consists of two parts: (i) a base emotion model, including a graph convolutional module, a deep feature aggregation and fusion module, and an emotion classifier (see Figure 2), and (ii) a transfer learning strategy, which unitizes the full-channel EEG signals in the training data to facilitate the representation learning of the target few-channel EEG signals (see Figure 3).

Our base emotion model is trained with the proposed transfer learning strategy in a synergistic and progressive manner. Given that EEG signals have an inherent graph structure (electrode connectivity) that GCNs can leverage, we deploy separate GCN modules for the inner and outer loops of transfer learning. In the inner loop, the model concentrates on extracting spatial-temporal features from the full EEG data; while in the outer loop, GCN learns to adapt these features to the target domain. Our model also incorporates temporal information by sharing the feature aggregation and classification layers between loops, which has been demonstrated to be an effective approach in the experiments.

The proposed base emotion model and feature-aggregation-based cross-device transfer learning strategy are described in detail in Sections 3.2 and 3.3, respectively.

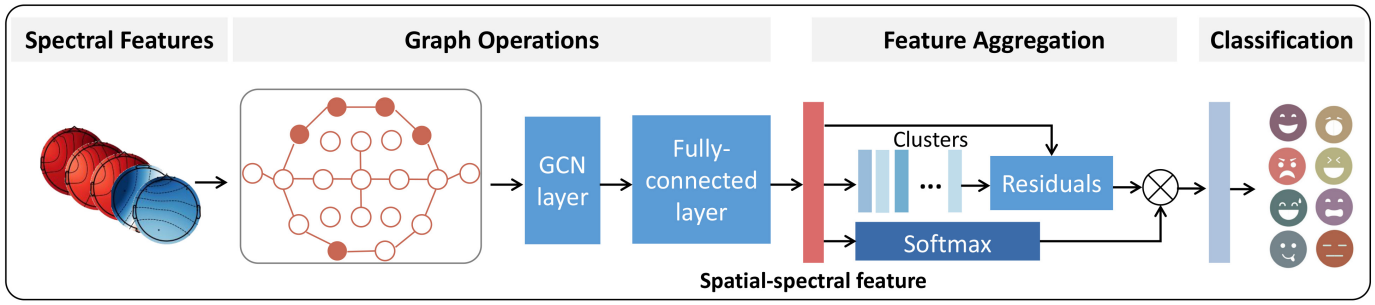


Fig. 2. The pipeline of our base emotion model. Our base emotion model mainly consists of a graph convolutional module, a deep feature aggregation and fusion module, and a classifier.

3.2 Base Emotion Model

Existing studies demonstrated that graph convolution-based methods can extract effective EEG features in the spatial domain by modeling the relationships between different EEG signal channels. This study combines the graph convolutional layer with feature aggregation networks in our base emotion model to learn efficient representations. Graphs are suitable for modeling the spatial dependencies in EEG data. By adopting GCN layers, the base emotion model can extract high-level emotional features that contain information about the dynamics and connectivity across different brain regions. It is worth noting that the feature aggregation network in the base emotion model is designed to capture the difference of EEG signal distributions in the time domain, which provides supplementary information to the spatial features extracted by GCNs.

3.2.1 Graph Convolutional Module

First, we employed a GCN layer proposed in [39] to extract the spatial-spectral channel features from input EEG signals. Each EEG sample is denoted as $x \in R^{t\times n\times d}$, where t is the length of the time sliding window, n is the number of signal channels, and d is the extracted feature dimension. Specifically, we extract power spectral density (PSD) [4] and differential entropy (DE) [59] features from multiple frequency bands (i.e., δ band: 1.5-4 Hz, θ band: 4-8 Hz, α band: 8-14 Hz, β band: 14-31 Hz, and γ band: 31-49 Hz) as representative spectral features. Then, we construct a graph $G = (V, E)$ for each EEG sample x to feed into the convolutional layer. More specifically, $V = \{v_i\}$ denotes the node-set representing the features of each channel, and $E = \{e_{i,j}\}$ is the edge set, in which $e_{i,j}$ represents a learnable weight built on nodes v_i and v_j. The original graph spectral filter [60] learns a function $g(\cdot)$ to extract EEG features from $G = (V, E)$ as:

$$y = g(L)x = Ug(\Lambda)U^T x, \tag{1}$$

where L is the Laplacian matrix of the graph, and U is the orthonormal matrix obtained by the singular value decomposition (SVD) of L. DGCNN [39] applies a Chebyshev polynomials framework to the regular graph operation by approximating $g(\Lambda)$ with $\sum_{k=0}^{K-1} \theta_k T_k(\tilde{\Lambda})$, in which θ_k and $T_k(\tilde{\Lambda})$ are the k-order Chebyshev coefficient and Chebyshev polynomial, respectively. The graph convolution of x can be further expressed as:

$$y = g(L)x \approx \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})x, \tag{2}$$

where $\tilde{L}$ can be obtained by normalizing L as $\tilde{L} = 2L/\lambda_{max} - I_N$, λ_{max} is the largest element of the diagonal entries of λ, and I_N is a $N \times N$ identity matrix.

Finally, we convert the output of the graph convolutional layer $y \in R^{t\times n\times d}$ into a spatial-spectral feature $\dot{x} \in R^{t\times D}$:

$$\dot{x} = FC_{gcn}(\sigma(y)), \tag{3}$$

where σ is a rectified linear unit (ReLU) activation function, and $FC_{gcn}(\cdot)$ represents a fully-connected layer. We set $D = 1024$ in our implementation. In summary, the GCN layers learn the spatial relationships between EEG channels, while the PSD and DE features characterize the spectral attributes of the signals.

3.2.2 Feature Aggregation and Fusion Module

As typical time series signals, temporal clues are also important in EEG emotion prediction. Aiming to extract more powerful EEG features with stronger representation ability in the time domain, we adopt a VLAD descriptor-based network, NeXtVlad [61], to conduct EEG feature aggregation and fusion. On the one hand, different emotions may have different EEG features as pointed out by previous work [37], and there may exist multiple discriminative local patterns for the emotional state in an EEG data segment. On the other hand, there are significant inter-variations in EEG signals, which means each subject's local pattern is different even under the same emotional stimuli. Our goal of introducing VLAD to emotion feature extraction is to construct feature descriptors that are universal for all subjects' EEG signals and invariant to the number of signal channels.

The VLAD-based feature aggregation module effectively models the temporal emotional clues in EEG signals and provides complementary information to the spatial features extracted by GCNs. VLAD compresses several local features into a global feature of a certain size by first conducting feature clustering and then accumulating all feature residuals for each cluster. We use the output of the graph convolutional module $\dot{x} \in R^{t\times D}$ as the input to the VLAD

module. This VLAD module then outputs a compact global feature $\mathbf{V} \in R^{K \times D}$:

$$\mathbf{V}(j,k) = \sum_{i=1}^{t} a_k(\dot{x}_i)(\dot{x}_i(j) - c_k(j)), k \in K, j \in D, \quad (4)$$

where $\dot{x}_i$ is the i_{th} local feature in $\dot{x}$, c_k is the k_{th} cluster center, and $a_k(\dot{x}_i)$ is a symbolic function, which is 1 if $\dot{x}_i$ belongs to cluster c_k and 0 otherwise. $\mathbf{V}(j,k)$ is the residual sum of all local features and the corresponding cluster centers in a class.

Original VLAD cannot be used in backpropagation networks since the symbolic function $a_k(\cdot)$ is non-derivable. To overcome this problem, Arandjelovic et al. [62] proposed NetVLAD, which replaces $a_k(\cdot)$ with a softmax function as in Equation 5. Then, the compact global feature $\mathbf{V}$ can be updated according to Equation 6.

$$\text{Softmax}_k(\dot{x}_i) = a_k(\dot{x}_i) = \frac{e^{w_k^T \dot{x}_i + b_k}}{\sum_{k'} e^{w_{k'}^T \dot{x}_{k'} + b'_k}}, \quad (5)$$

$$\mathbf{V}(j,k) = \sum_{i=1}^{t} \text{Softmax}_k(\dot{x})(\dot{x}_i(j) - c_k(j)), k \in K, j \in D, \quad (6)$$

where w_k, b_k, and c_k are the parameters to be learned. Different from VLAD, which needs to manually compute the clusters (e.g. using the K-means algorithm), NetVLAD calculates the c_k in a learned way and enables an end-to-end updating process. In our work, we use the improved version of NetVLAD, i.e., NeXtVLAD proposed in [61], to reduce the network parameters and improve the training speed. Specifically, the NeXtVLAD decomposes the input feature $\dot{x} \in R^{t \times D}$ into a group of lower-dimensional vectors $\tilde{x} \in R^{t \times C \times \frac{D}{C}}$ before aggregation, where C is the number of groups:

$$\mathbf{V}^g(j,k) = \sum_{i=1}^{t} \sigma_g(\dot{x}_i)\text{Softmax}_{gk}(\dot{x}_i)(\tilde{x}_{ig}(j) - c_k(j)),$$
$$g \in \{1,...,C\}, i \in \{1,...,t\}, k \in \{1,...,K\}, j \in \{1,...,\frac{D}{C}\}. \quad (7)$$

Then, we obtain the global spatial-spectral-temporal EEG emotion feature X by adding $V^g(j,k)$ over groups followed by a fully-connected fusion layer $FC_{fusion}(\cdot)$ as:

$$X = FC_{fusion}(\sum_{g=1}^{C} \mathbf{V}^g(j,k)). \quad (8)$$

The global spatial-spectral-temporal EEG emotion feature X can also be represented as:

$$X = (o_1, o_2, ..., o_c), \quad (9)$$

where o_i is the i_{th} neuron of the fully-connected layer $FC_{fusion}(\cdot)$ of the base emotion model, and c is the number of emotion classes.

Finally, we apply a softmax function to map the global spatial-spectral-temporal feature X to probabilities $\hat{Y} = \{\hat{Y}_1, \hat{Y}_2, ..., \hat{Y}_c\}$ corresponding to each emotion as:

$$\hat{Y}_i = \text{Softmax}(o_i). \quad (10)$$

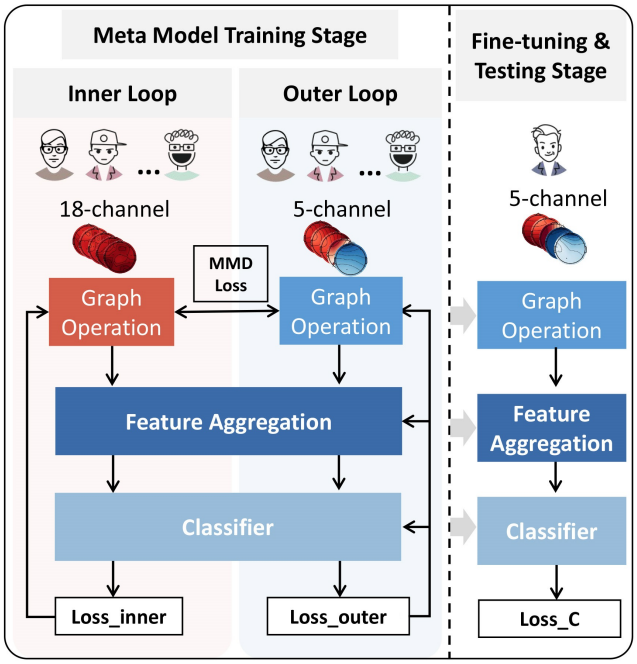


Fig. 3. The framework of our EEG emotion transfer learning includes two stages: (1) in the meta-model training stage, we aim to learn a good initialization for feature-aggregation-based cross-device EEG transfer network with 18-channel and 5-channel paired EEG signals of existing subjects. (2) In the model fine-tuning and testing stage, the feature aggregation network is adapted to new subjects with only few-channel data. Specifically, only a very small amount of 5-channel EEG signals are needed in the fine-tuning process, and the model can fit well on new subjects.

We use the cross-entropy loss for our base emotion model:

$$l_{base} = -\sum_{j=1}^{c} Y_j \log \hat{Y}_j. \quad (11)$$

where $Y = (Y_1, \cdots, Y_c)$ represents the ground truth emotion label.

3.3 Feature-Aggregation-Based Cross-Device Transfer Learning

We denote our EEG emotion recognition base model described in Section 3.2 as $\mathcal{M}$, which consists of a GCN module and a VLAD module, and maps an EEG signal sample x to an emotion state $\hat{Y}$:

$$\hat{Y} = \mathcal{M}(x). \quad (12)$$

In this section, we propose an EEG emotion transfer learning strategy by combining full-channel and few-channel EEG signals to jointly train a model $\mathcal{M}_{meta}$, which is then adapted to a new model $\mathcal{M}_{fine\_tune}$ to recognize a new subject's emotion states via fine-tuning with only few-channel EEG signals. Since GCN is highly related to the input sizes of EEG signal channels, we use separate graphs for full-channel and few-channel EEG signals and aim to learn a shared NeXtVLAD to fuse graph features of different subjects across different devices. The shared NeXtVLAD between different subjects and different EEG devices (1) has more data for stable training, and (2) can learn universal feature aggregations and fusion representations. As shown in

Algorithm 1 EEG Feature Aggregation Transfer Learning

Input: $D_{train} = \{C_i | i = 1, ..., M\}$: full-channel and few-channel paired dataset (C_{i_full} and C_{i_few})
Parameter: $\mathcal{M}_{meta}$: ϕ, graph module in the inner loop: $\theta_{g_{in}}$, graph module in the outer loop: $\theta_{g_{out}}$, all parameters of the inner loop: θ^i
Output: ϕ: Parameters initialization of the base model, including the graph module, the feature aggregation networks, and the classifier.

```
randomly initialize φ
while the iterations do not satisfy the predefined algorithm convergence condition do
    Sample batch of subjects from D_train
    for all subjects do
        Evaluate ▽_{θ^i_{g_in}} l(θ^i_{g_in}) with C_{i_full}
        Update the parameters θ of the graph operation in the inner loop with gradient descent
        Compute l(θ̂^i) and MMD(θ^i_{g_in}, θ^i_{g_out}) with C_{i_few}
    end for
    Compute L_meta(φ) and MMD(θ_{g_in}, θ_{g_out})
    Update the parameters φ of the graph operation in the outer loop with gradient descent
end while
```

Figure 3 and Algorithm 1, our cross-device transfer learning includes two stages:

1) Meta-model training stage: Using separated GCN modules, shared feature aggregation layers, and classifiers to learn common representation between full-channel and few-channel EEG signals to provide a good initialization for the next stage.
2) Fine-tuning and testing stage: Fine-tuning the pre-trained model with EEG signals of new subjects (unseen in the training data in the meta-model training stage), quickly adapting the meta-model into a user-specific model.

Inspired by the MAML framework [23], we address the cross-device transfer learning problem by learning an optimal initialization of the EEG feature aggregation network, which could be quickly adapted to EEG signals of new subjects. The main differences between the original MAML framework and our cross-device transfer learning training strategy are twofold:

- The goal of MAML is to learn a good network initialization to achieve fast adaptation on new tasks with a small amount of training data and fast fine-tuning. MAML is essentially proposed to deal with the challenge of having unseen classes in the testing set, where the existing classes are the *meta-training dataset*, and the new classes compose the *meta-testing dataset*. However, in our case, we set the EEG signals of existing subjects as the source and the EEG signals of new subjects as the targets. The theoretical support behind this is that there exist common patterns of EEG emotional features among individuals [50].
- The original MAML framework keeps the entire network structure unchanged and uses two training loops to get good initialization parameters, while we aim to keep unique parameters of the graph operation module for each subject and learn a public feature-aggregation module common to all subjects and EEG devices in the outer loop.

Denote $D_{train} = \{C_i | i = 1, ..., M\}$ as the meta-training dataset to train the meta-model $\mathcal{M}_{meta}$, and $D_{test} = \{P_j | j = 1, ..., Q\}$ as the meta-testing dataset used to fine-tune and test the final emotion classification model $\mathcal{M}_{fine\_tune}$. Each C_i contains the paired full-channel (18-channel) and few-channel (5-channel) EEG signals of the i_{th} subject, and P_j only has the few-channel EEG signal of the j_{th} subject (different from those subjects who has paired data, i.e., the subjects of D_{train} and D_{test} are separate).

We define the emotion recognition of the i_{th} person as the task T_i of the meta-model. We consider the initialization parameters of our feature aggregation network $\mathcal{M}_{meta}$ as ϕ, and the model learned on the task T_i as θ^i. In our case, we split the model $\mathcal{M}_{meta}$ into three parts: the two graph operations (i.e. the graph modules in the inner loop and outer loop of the meta training process, $\theta^i_{g_{in}}$ and $\theta^i_{g_{out}}$), the feature aggregation networks, and the classifier. In the inner loop of our transfer learning process, we only update the parameters of the graph operations $\theta^i_{g_{in}}$ with gradient descent:

$$\hat{\theta}^i_{g_{in}} = \theta^i_{g_{in}} - \varepsilon \bigtriangledown_{\theta^i_{g_{in}}} l(\theta^i_{g_{in}}), \tag{13}$$

where $l(\theta^i_{g_{in}})$ is the cross-entropy loss (defined in Equation 11) of task T_i with model $\theta^i_{g_{in}}$ on the full-channel EEG signals from C_i, and ε is the learning rate.

$$l(\theta^i_{g_{in}}) = l_{base} | \theta^i_{g_{in}} = -\sum_{j=1}^{c} Y_j \log \hat{Y}_j | \theta^i_{g_{in}}, \tag{14}$$

where Y_j and $\hat{Y}_j$ are ground truth and predicted labels within task T_i, respectively.

In the outer loop, the loss function of $\mathcal{M}_{meta}$'s initialization parameters can be computed as the mean of the losses of all meta-training tasks on the few-channel EEG signals:

$$L_{meta}(\phi) = \frac{1}{M} \sum_{i=1}^{M} l(\hat{\theta}^i), \tag{15}$$

where $\hat{\theta}^i$ is the whole model (including the graph operation module, the feature aggregation module, and the classifier) of task T_i updated in the inner loop.

$$l(\hat{\theta}^i) = l_{base} | \hat{\theta}^i = -\sum_{j=1}^{c} Y_j \log \hat{Y}_j | \hat{\theta}^i. \tag{16}$$

We have also introduced a Maximum Mean Discrepancy (MMD) loss, which evaluates the distribution similarity between the features extracted by the graph operations of full-channel and few-channel EEG signals. Denote C_{i_full} and C_{i_few} as full-channel and few-channel EEG signals of the i_{th} subject respectively, the MMD loss can be computed as:

$$MMD(\theta_{g_{in}}, \theta_{g_{out}}) = || \frac{1}{M} \sum_{i=1}^{M} (g_{in}(C_{i\_full}) - g_{out}(C_{i\_few})) ||^2_H. \tag{17}$$

The final loss of the transfer learning training process is:

$$L(\phi) = L_{meta}(\phi) + MMD(\theta_{g_{in}}, \theta_{g_{out}}), \tag{18}$$

where $L_{meta}(\phi)$ is minimized by gradient descent:

$$\phi \leftarrow \phi - \eta \bigtriangledown_{\phi} L(\phi), \tag{19}$$

where η is the learning rate of the outer loop.

During the fine-tuning and testing stage, we first adopt the meta-model $\mathcal{M}_{meta}$ learned in the training stage (see the right part of Figure 3), then we use the few-channel signals of a new subject P_j for fine-tuning. The classification cross-entropy loss of $\mathcal{M}_{fine\_tune}$ on the support set of P_j is used to update the model. Finally, we obtain a specific emotion recognition model $\mathcal{M}^{j}_{fine\_tune}$ for subject j in D_{test}. In summary, the meta-model training stage aims to learn transferable feature aggregations by optimizing ϕ. The model fine-tuning stage adapts the meta-model to new subjects by optimizing the graph operation parameters $\theta^{n}_{g_{in}}$ and $\theta^{n}_{g_{out}}$ as well as the shared feature aggregation parameters ϕ. By leveraging transferable knowledge and conducting quick adaptation, the proposed framework can boost the performance of few-channel EEG emotion recognition with limited training data.

4 Experiment

4.1 Datasets

We build a new Cross-Device EEG Emotion Dataset (CDEED) with paired full-channel and few-channel EEG signals, which provides a benchmark for our CD-EmotionNet framework. Moreover, we also evaluate the performance of our model on three other publicly available EEG emotion datasets: SEED-IV [24], SEED-V [25] and DEAP [4]. These datasets are summarized in Table 1.

4.1.1 The CDEED dataset

38 subjects participated in data collection. Our CDEED dataset contains two sub-datasets: (1) the *BRK-KangII* sub-dataset contains paired data of 25 subjects. Each pair includes 18-channel (full-channel) and 5-channel (few-channel) EEG signals of the same subject under the same set of video stimuli. (2) The *KangII* sub-dataset contains 5-channel EEG data from the remaining 13 subjects. The BRK-KangII sub-dataset provides paired full-channel and few-channel EEG data, which enables cross-device transfer learning. The KangII sub-dataset acts as the testing set to evaluate the performance of the model adapted to the BRK-KangII sub-dataset.

All procedures of the study were carried out in accordance with the Declaration of Helsinki and were approved by the Ethics Committee of Tsinghua University. Before the experiment, all participants were thoroughly informed about the details of the experiment, including its goal, procedure, and the use of the collected data. Written informed consent was obtained from all participants.

Stimuli. In the experiments, emotional film clips were used as stimuli. The emotional film clips are from a standardized emotional film clips database [14], in eight types of emotions (disgust, fear, anger, sad, neutral, amusement, tenderness, and happiness). Each emotion was elicited by two movies, and each movie lasted about 90 seconds.

Experimental procedure of BRK-KangII sub-dataset. 25 subjects participated in the construction of the BRK-KangII sub-dataset, where two sessions were performed on different dates, and each session contained 16 trials. In each trial of the first session, the participants watched one of the film clips, while their EEG signals were collected with the 18-channel Wearable Sensing's wireless DSI-24 [7]. In each trial of the second session, the EEG signals were collected with the 5-channel SleepUp2.0 Naolubrain EEG device [11]. To reduce the confounding factors between the full-channel and few-channel EEG collection sessions, there was a one-week interval between the two sessions. Half of the subjects conducted the 18-channel EEG signal collection first and then the 5-channel EEG signal, and the other half of the subjects collected the data in the reverse order. The order of the eliciting materials was arranged in a Latin square design [63] among subjects during the emotional elicitation.

Experimental procedure of KangII sub-dataset. As for the construction of the other KangII sub-dataset, the rest of the 13 out of the 38 subjects participated in the experiment, where each of them only performed one session. They watched the 16 film clips in 16 trials, during which their EEG signals were collected with the 5-channel SleepUp2.0 Naolubrain EEG device.

Figure. 4 shows the time diagram of the experiment, which was used in the construction of both BRK-KangII and KangII sub-datasets. The experiment started with the participants filling in their personal information, followed by 16 trials. Participants were instructed to minimize head movements during the experiment. The steps in each trial were as follows: (1) The display of a film clip. (2) Self-assessment. (3) Addition and subtraction operations and a 15-second break before the next trial to eliminate the effect of the previous film on the current one.

The snapshots of the 5-channel SleepUp2.0 Naolubrain and the 18-channel Wearable Sensing's wireless DSI-24 EEG devices are shown in Figure 5. The electrode placements of the two devices are shown in 1. To point out, there are three electrodes (the reference electrode Fpz and two ear clip sensors A1/2) in the DSI-24 device that were not used in our experiments.

4.1.2 The SEED-IV dataset

The SEED-IV dataset [24] contains EEG data of 15 subjects. Each subject conducted 3 sessions, and there were 24 trials for each session. 72 films were used as the stimuli for eliciting four emotions (happiness, sadness, fear, and neutral). The EEG signals were collected with the 62-channel ESI NeuroScan System.

4.1.3 The SEED-V dataset

The SEED-V dataset [25] contains EEG data from 20 subjects, each of whom was asked to participate in three experiments. The subjects in each experiment needed to watch 15 films (3 pieces of each type of emotion). There are five emotional states in SEED-V (happy, sad, disgust, fear, and neutral). The EEG signals were recorded from 62 EEG electrodes according to the international 10-20 system. We use the EEG features of 16 subjects provided by the authors in our experiments.

7. https://wearablesensing.com/research/#publications

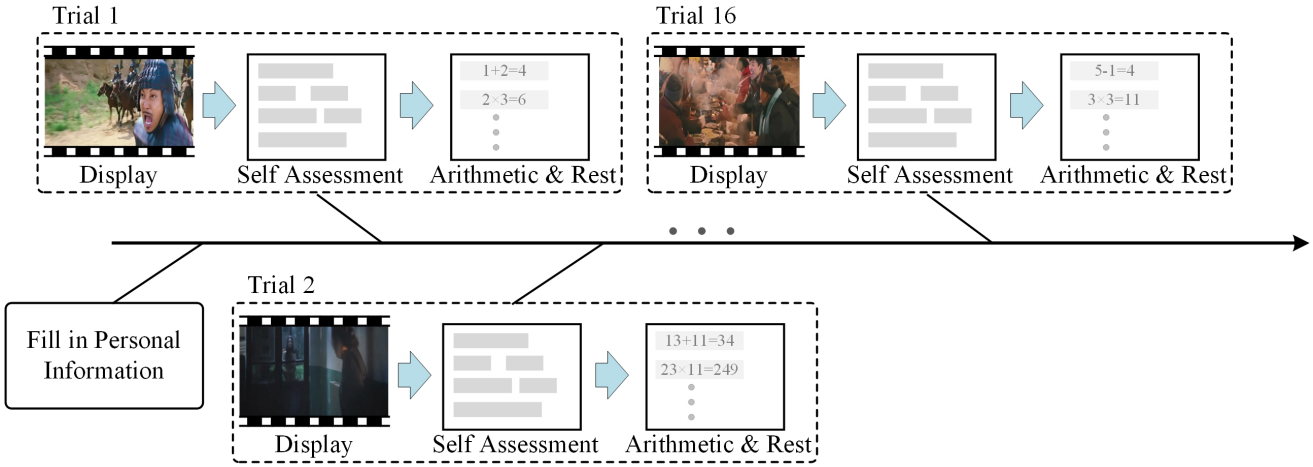


Fig. 4. The timing diagram of the experiment. The experiment procedure in each trial mainly consisted of: (1) the display of a film clip. (2) Self-assessment. (3) Addition and subtraction operations and a 15-second rest to eliminate the effect of the previous trial on the current one.

TABLE 1
Summary of the four datasets used in our experiments.

Dataset	Sub-Dataset	Emotion Categories	Number of Subjects	Is Paired?	Channels
CDEED	BRK-KangII	disgust, fear, anger, sad, neutral, amusement, tenderness, happiness	25 subjects	Yes	18 5
	KangII		13 subjects	No	5
SEED-IV [24]	-	happiness, sadness, fear, neutral	15 subjects	No	62
SEED-V [25]	-	happy, sad, disgust, fear, neutral	20 subjects	No	62
DEAP [4]	-	valence (positive/negative), arousal (high/low)	32 subjects	No	32

Fig. 5. The snapshots of the devices. (a) 5-channel SleepUp2.0 Naolubrain EEG device.(b) 18-channel Wearable Sensing's wireless DSI-24 device.

4.1.4 The DEAP dataset

The DEAP database contains EEG data of 32 healthy participants (16 males and 16 females) evoked by music video stimuli. Volunteers were asked to watch 40 one-minute videos, and the EEG signals of the subjects were collected at a 512 Hz sampling rate. The EEG signals were recorded from 32 electrodes according to the international 10-20 system [8]. All subjects were instructed to rate the valence, arousal, dominance, and liking of the viewed videos on a scale of 1 to 9 after watching the videos. We followed the partitioning strategy in [17], which transforms the dataset into binary emotion recognition tasks by segmenting the valence dimension to positive/negative and the arousal dimension to high/low arousal. The thresholds of both the two dimensions are 5. After preprocessing and segmenting, we obtained a balanced dataset to validate our proposed cross-device EEG emotion transfer model.

8. Besides the EEG data, the DEAP database also contains 8 peripheral channels, which were removed in our experiments in this paper.

4.2 Implementation Details

We trained our CD-EmotionNet on NVIDIA RTX 2080 GPU, and the batch size was set as 4. The learning rate was 0.01, and the Adam optimization was used to minimize the loss function.

For data preprocessing of the CDEED dataset, we employed the EEGLAB toolbox in MATLAB [64] to preprocess the collected raw EEG data. This preprocessing mainly contains data importation, electrode localization, electrode selection, re-referencing, filtering, baseline correction, manual identification and removal of bad segments and channels, independent component analysis, and manual exclusion of irrelevant components. Specifically, a bandpass filter (1-50 Hz) and a 50 Hz notch filter was first used to remove noise, then an independent principal component analysis was conducted to eliminate artifacts. For other datasets in the experiments, we utilized the preprocessed data provided by the authors directly.

We conducted experiments mainly with the subject-dependent scheme. The reason for adopting the subject-dependent setting is from the definition of our task. The application scenario of our task in the testing phase is to get a customized emotional model of a new user by fine-tuning the base emotion model with only a small amount of his/her few-channel EEG data. The **Dataset Splitting** in this section shows the implementation details for the subject-dependent setting, and the experimental results are shown in 4.3 and 4.4. Moreover, we also conducted additional experiments to

TABLE 2
Investigation of our CD-EmotionNet via ablation study on the CDEED, SEED-IV, SEED-V, and DEAP datasets. The mean and standard deviation (%) of accuracies are shown. Since there is no cross-device EEG data in the SEED-IV, SEED-V, and DEAP datasets, we only apply our full method containing the transfer learning training process on the CDEED dataset, where the 18-channel and 5-channel paired data of existing subjects in BRK-KangII are used as the training set, and the 5-channel data of new subjects are used to fine-tune and test.

Method	CDEED			SEED-IV	SEED-V	DEAP	
	BRK-KangII		KangII			Valence	Arousal
	18-channel ACC /STD (%)	5-channel ACC /STD (%)	5-channel ACC /STD (%)	ACC /STD (%)	ACC /STD (%)	ACC /STD (%)	ACC /STD (%)
Only Graph	63.14 / 14.17	41.85 / 6.46	40.93 / 12.75	65.97 / 15.03	58.54 / 20.73	75.42 / 10.85	73.97 / 12.61
Only Feature Aggregation	45.00 / 11.57	42.28 / 4.86	40.24 / 8.09	57.66 / 16.07	77.26 / 16.42	71.27 / 10.02	73.83 / 11.26
Graph+Feature Aggregation **(Base Emotion Model)**	**64.47** / 10.17	**45.67** / 5.12	41.15 / 8.64	**83.10** / 11.87	**82.67** / 16.51	**86.29** / 9.71	**84.16** / 10.86
Base Emotion Model+Transfer Learning **(Full Method)**	-	-	**47.50** / 8.92	-	-	-	-

validate the effectiveness of our proposed method in the subject-independent scenario. The experiment procedure and results are shown in Section 4.5.2.

For our main experiments with subject-dependent protocol on the CDEED dataset, we followed [37] to randomly shuffle all the data samples. For SEED-IV and SEED-V, we shuffled the samples within each session and within each trial, respectively. We ran each experiment 5 times and report the average value. Within each experiment, we computed the STDs of each subject and averaged them.

Dataset Splitting The data ratio for training and testing is set to 8:2. The EEG segment time length is 2s on the three datasets. (1) For the CDEED dataset, we only use the last 600 segments (480 for training, 120 for testing) of each subject watching each video. (2) For the Seed-IV, each subject conducted 3 sessions, and there were 24 films (16 for training and 8 for testing) in each session. The number of training and testing segments are 276/259/256 and 146/152/149 in the three sessions, respectively. (3) For the Seed-V, each subject watched 15 films (10 for training and 5 for testing) in each experiment. There are 247/176/215 training segments and 91/91/81 testing segments in the three experiments, respectively. (4) For the DEAP dataset, each participant watched a subset of 40 music videos. We used the preprocessed data provided by the authors, which was segmented into 60-second trials and downsampled to 128Hz. The data offered by the authors was also applied with a bandpass frequency filter from 4-45Hz and averaged to the common reference. PSD features were extracted before use in our experiments. We used each subject's data to evaluate different methods, and the final average recognition accuracy is reported. Within each subject, the data was randomly split as 80% for training and 20% for testing. There are 960 training segments and 240 testing segments for each subject.

4.3 Ablation study

In this section, we demonstrate the emotion recognition performance of our CD-EmotionNet on our proposed CDEED dataset by investigating the three main components of our CD-EmotionNet described in Section 3, i.e. the graph convolutional module, the feature aggregation and fusion module, and the cross-device transfer learning method, via an ablation study.

4.3.1 Experiment Results

We conduct an ablation study on the CDEED, SEED-IV, SEED-V, and DEAP datasets to explore the effectiveness of each component in our CD-EmotionNet. Specifically, we compare our full model with three variant models:

- Only Graph: only the Graph Convolutional Module in the base emotion model in Section 3.2, without the feature aggregation and fusion module and the transfer learning training procedure.
- Only Feature Aggregation: only the Feature Aggregation and Fusion Module in the base emotion model in Section 3.2, without the graph operations and the transfer learning training procedure.
- Graph+Feature Aggregation: the entire Base Emotion Model in Section 3.2, without the transfer learning training procedure.
- Full Method: applying the proposed transfer learning training procedure to the base emotion model.

The experiments are shown in Table 2. We observe that:

(1) *The graph convolutional module, the feature aggregation and fusion module, and the transfer learning schedule all contribute to the final performance of our CD-EmotionNet.* On the one hand, our full method consisting of the base emotion model and the transfer learning strategy performs best on the 5-channel EEG emotion recognition task on the KangII sub-dataset of the CDEED dataset, reaching a mean of emotion recognition accuracy of 47.5% with only few-channel EEG signals. On the other hand, our base emotion model outperforms both the *Only Graph* model and the *Only Feature Aggregation* model on all four datasets. Compared to the results of *Only Graph*, our base emotion model increases recognition accuracy by about 2%, 18%, 24%, 11%, and 10% on the BRK-KangII sub-dataset of CDEED, the SEED-IV, the SEED-V and the DEAP (Valence and Arousal) datasets, respectively (the 2nd and 4th rows in Table 2). The recognition accuracies of our base emotion model are about 12%, 25%, 15%, 15%, and 10% higher than those only using feature aggregation and fusion on the BRK-KangII sub-dataset of CDEED, the SEED-IV dataset, the SEED-V dataset, and the DEAP (Valence and Arousal) datasets (the 3rd and 4th rows in Table 2).

(2) *The Base Emotion Model can efficiently capture the emotional features in full channel EEG signals.* The fourth row in Table 3 shows the accuracy and standard deviation of our base emotion model built on GCNs and feature aggregation network (without transfer learning strategy) on the three

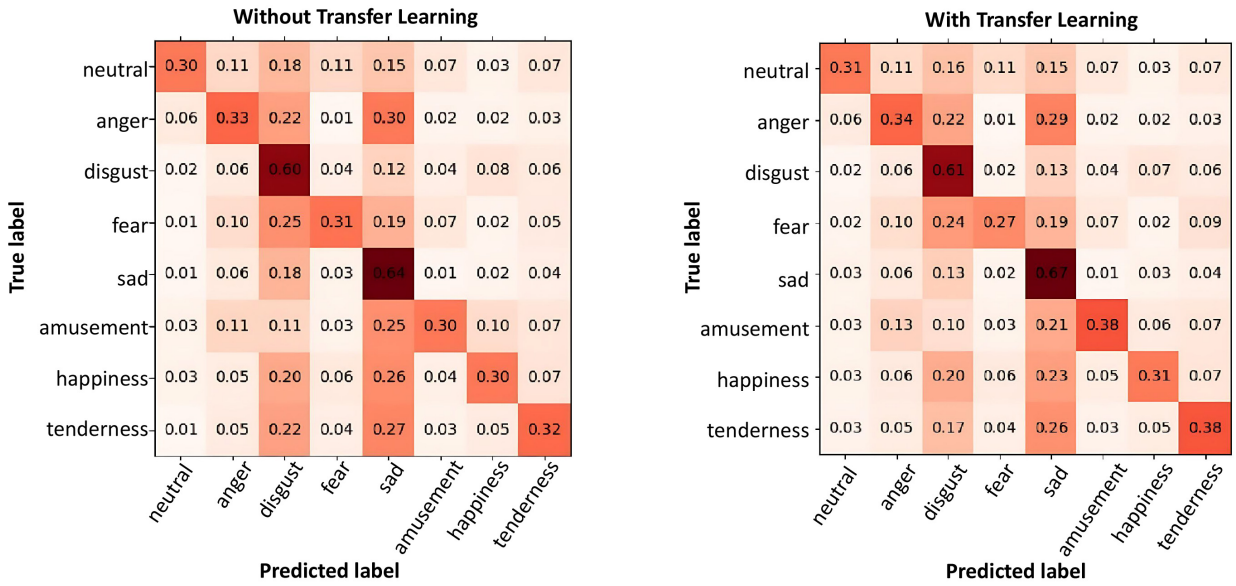


Fig. 6. The confusion matrix for the 8-class emotion classification of the KangII (13 subjects) dataset. We show the results without/with our transfer learning strategy. Except for the "fear" emotion, the recognition accuracy of all other emotions improved when applying the transfer learning training procedure, indicating the effectiveness of our transfer learning strategy.

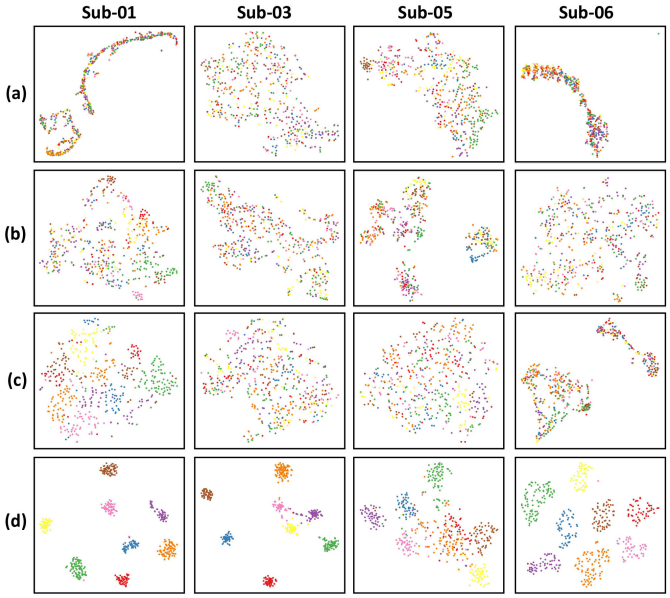


Fig. 7. Visualization of latent features using t-SNE on the KangII dataset. We presented the features extracted by four models: (a) Only Graph, (b) Only Feature Aggregation, (c) Base Emotion Model, and (d) Full Method. The different colors represent different emotions.

datasets. Our base emotion model achieves 64.47% and 45.67% in the eight-class emotional classification task on the full-channel (18-channel) and few-channel (5-channel) EEG data from the BRK-KangII sub-dataset of CDEED. Moreover, our base emotion model obtains an accuracy of 83.10% for four-class emotion discrimination on the SEED-IV dataset and 82.67% for five-class emotion recognition on the SEED-V dataset. These results indicate that our base emotion model is able to extract effective EEG features for emotion recognition.

(3) *The capability of using only few-channel EEG data to conduct emotion recognition is relatively weak for the models without the transfer learning framework.* Even the accuracy improvement of the base emotion model is limited on the KangII sub-dataset of CDEED, indicating the necessity of introducing extra information to help the feature learning of few-channel EEG signals (the 4th column in Table 2).

(4) *The transfer learning strategy improves the emotion recognition accuracy of most emotions.* We analyze the EEG emotion recognition results with and without transfer learning on the KangII sub-dataset of the CDEED dataset in Table 3. Recall that there is a meta-model training stage and a fine-tuning and testing stage in our transfer learning framework, as described in Section 3.3. In the experiments, the meta-model is trained with the 18-channel and 5-channel paired EEG signals in the BRK-KangII sub-dataset, and the model is fine-tuned and tested with the 5-channel EEG signals of new subjects in the KangII sub-dataset. We observe that the accuracy of emotion classification on the KangII sub-dataset increases from **41.15%** to **47.50%** when applying our transfer learning strategy to the base emotion model.

4.3.2 *Visualization Analysis*

Confusion Matrix Visualization. We analyze the confusion matrix of the KangII sub-dataset (13 subjects) in Figure 6. It is observed that regardless of transfer learning, the algorithms have high classification accuracies on the "sad" and "disgust" emotions. Similar to ours, the experiments in [65] also indicate that recognizing the "disgust" emotion has the best performance. Moreover, the classification accuracy of most emotions (except for the "fear" emotion) increases when applying our transfer learning strategy, among which recognizing the "amusement" emotion improved the most (with an accuracy increment of 8%), followed by the "tenderness" emotion (with an accuracy increment of 6%).

T-SNE Visualization. We further explore the effectiveness of our proposed CD-EmotionNet by visualizing the latent EEG patterns learned by different components of

our method (as shown in Section 4.3.1), i.e., Only Graph, Only Feature Aggregation, Base Emotion Model, and Full Method. Feature distributions of four subjects from the KangII dataset are shown in a 2-dimensional space with t-SNE. For each subject, we randomly visualize four latent features of length 1-s corresponding to the four methods from a trial in the testing set. As shown in Figure 7, we can observe that our full method that integrates the base emotion model and the transfer learning strategy is able to obtain more effective EEG features.

4.4 Comparison With State-Of-The-Art Methods

We compare our method with several mainstream traditional machine learning and state-of-the-art deep-learning-based methods. Then, we also conduct cross-device EEG emotion recognition tasks with our CD-EmotionNet and four representative transfer learning algorithms on the CDEED dataset.

4.4.1 Comparison with Representative EEG Emotion Recognition Methods

We compare our base emotion model with several representative EEG emotion recognition methods on all three datasets using traditional classifiers and deep learning-based networks. **(1) Traditional machine learning methods:** SVM [66] and Random Forest [67] are two classic machine learning classifiers that have been widely used to recognize emotion from EEG. SVM is a classic linear support vector machine classifier that has been used to recognize emotion from EEG data in various experiments. Random Forest is an ensemble classifier using multiple decision trees to train and predict samples. **(2) Popular convolutional neural network (CNN) based methods:** EEGNet [15] and DBN [3] are two representative CNN-based EEG signal analysis models. EEGNet is a flexible and easy-to-use CNN-based model designed for BCI. Deep Belief Networks (DBN) contain many layers of hidden variables and are able to fuse multi-channel and multi-frequency EEG features. SST-EmotionNet [37] is an attention 3D dense network based on the assumption that different emotions may have very different local spatial-spectral-temporal features. **(3) Graph convolutional network (GCN) based methods:** we compare with three state-of-the-art GCN-based methods: DGCNN [39], RGNN [22], and SparseDGCNN [19]. DGCNN explores the relationships between different channels via learning a dynamic adjacency matrix. RGNN considers the biological topology among different brain regions. SparseDGCNN is a sparse version of DGCNN. **(4) Recurrent neural network (RNN) based method:** We also compare with the recently proposed RNN-based method ACRNN [36], which adopts an attention-based RNN framework and combines features in the channel and time domains.

Table 3 presents the emotion recognition results of our CD-EmotionNet and all the baselines on the three datasets. We observed that: (1) our CD-EmotionNet outperforms all baseline methods with a large margin (the last row in Table. 3). (2) The graph-based methods (DGCNN, RGNN, and SpareDGCNN) outperform the classic methods (SVM and Random Forest) by a significant margin on all datasets. (3) EEGNet and DBN perform relatively worse on these datasets, possibly due to the fact that they were originally designed for fewer-category emotion recognition tasks, and their performance drops significantly when the number of emotion classes becomes larger. For instance, previous studies [17], [59] have shown that EEGNet performs well on the widely used EEG emotion dataset *SEED* [3], which contains three emotion classes (positive, neutral, and negative emotions). However, EEGNet's recognition accuracy significantly decreases on the CDEED, SEED-IV, and SEED-V datasets, which include a larger number of emotion categories, specifically 8, 4, and 5, respectively. Besides, we do not evaluate our method on the SEED dataset, but on the advanced dataset versions, i.e. SEED-IV and SEED-V. Similarly, our method performs much better than DGCNN on the SEED-V dataset. To a large extent, this is due to the fact that SEED-V contains more emotion categories than SEED-IV, and the emotion recognition task on SEED-V is more challenging. (4) Compared with using 5-channel EEG signals, most of the methods in Table 3 have obviously better results with 18-channel signals, indicating that the number of channels plays an important role in emotion recognition. (5) Apart from our approach, the DGCNN has gained a relatively high recognition precision. We have further applied the paired t-test to find whether there are significant differences between the results of the DGCNN and our method. The recognition results of the 13 subjects in the KangII dataset are used. We set the significant level as 0.05. The p-value is 0.000002, which indicates there is a significant difference between the two results.

4.4.2 Comparison with Transfer Learning Methods

We further compare our proposed EEG emotion recognition network with four representative transfer learning frameworks that are widely used in the field of EEG signal analysis. Since we focus on exploiting fewer-channel EEG emotion recognition with the help of full-channel EEG signals collected by different devices, we compare our CD-EmotionNet with these transfer learning baseline models on the CDEED dataset. **(1) Pretrain & Fine-tuning [68]:** the most general strategy of transfer learning that first trains a model with source domain data and then fine-tunes the model with the target data. In our case, we first use all the full-channel and few-channel paired EEG signals in BRK-KangII to pre-train the base emotion model, then fine-tune and test the model with the 5-channel data in KangII sub-dataset. **(2) MMD Loss [69]:** a domain adaptive network which introduces the MMD measure to reduce the distribution mismatch between source and target domains in the latent space. **(3) DANN [70]:** a domain-adversarial neural network that simultaneously learns the classifiers, feature extractors, and domain discriminators. **(4) SHOT [71]:** an unsupervised domain adaptation method that learns from labeled tasks and performs unlabeled but similar tasks.

As shown in Table 4, our CD-EmotionNet model achieves an eight-class emotion classification accuracy of 47.50%, which is 5% higher than the four other transfer learning baseline methods. The results indicate that our transfer learning strategy can capture invariant EEG emotion features shared between full-channel and few-channel EEG signals.

TABLE 3
Comparison of EEG emotion recognition performance with existing state-of-the-art methods on the CDEED, SEED-IV, and SEED-V datasets. The mean and standard deviation (%) of accuracies are shown.

Method	CDEED			SEED-IV	SEED-V
	BRK-KangII		KangII		
	18-channel	5-channel	5-channel	62-channel	62-channel
	ACC /STD (%)	ACC /STD (%)	ACC /STD (%)	ACC /STD (%)	ACC /STD (%)
SVM [66]	29.07 / 7.18	26.14 / 6.06	25.26 / 7.42	48.54 / 6.68	70.11 / 10.14
Random Forest [67]	46.43 / 13.34	31.53 / 6.41	35.73 / 9.23	57.58 / 13.19	57.85 / 13.87
EEGNet [15]	18.31 / 2.87	18.61 / 3.06	19.91 / 3.42	22.52 / 2.61	28.18 / 6.14
DBN [3]	14.99 / 3.60	14.98 / 4.31	13.81 / 4.26	25.83 / 11.64	22.65 / 2.49
DGCNN [39]	63.14 / 14.17	41.85 / 6.46	40.93 / 12.75	65.97 / 15.03	58.54 / 20.73
RGNN [22]	19.30 / 2.47	20.31 / 2.29	19.49 / 2.66	53.49 / 2.46	26.61 / 2.69
SpareDGCNN [19]	40.01 / 20.46	33.92 / 7.24	32.34 / 12.34	36.20 / 17.88	48.76 / 23.87
ACRNN [36]	23.10 / 3.14	24.71 / 4.21	24.87 / 4.22	35.09 / 10.51	39.05 / 8.15
SST-EmotionNet [37]	23.88 / 4.59	22.68 / 4.83	21.97 / 4.66	71.45 / 10.76	67.62 / 17.13
Our Method	**64.47** / 10.17	**45.67** / 5.12	**47.50** / 8.92	**83.10** / 11.87	**82.67** / 16.51

TABLE 4
Comparison with representative transfer learning methods that are widely used in the EEG analysis field. The mean and standard deviation (%) of accuracies are shown.

Method	KangII of CDEED (13 subjects) ACC / STD (%)
Pre-train & Fine-tune [68]	41.15 / 7.46
MMD Loss [69]	43.07 / 7.45
DANN [70]	42.17 / 10.79
SHOT [71]	42.82 / 8.71
Our Method	**47.50** / 8.92

TABLE 5
Subject-dependent evaluation on simulated datasets. The mean and standard deviation (%) of accuracies are shown.

Method	Modified SEED-IV 18-channel	Modified SEED-V 18-channel
	ACC / STD (%)	ACC / STD (%)
without transfer learning	82.04 / 13.57	73.87 / 16.12
with transfer learning	**89.41** / 6.64	**78.21**/ 9.74

4.5 More Evaluations

4.5.1 Evaluation on Simulated Datasets

In order to investigate the performance of our CD-EmotionNet with more few-channel EEG datasets, we modify the SEED-IV and SEED-V datasets by selecting partial electrodes of them to simulate the cross-device experiments, named **Modified SEED-IV** and **Modified SEED-V** respectively. Specifically, the SEED-IV and SEED-V datasets contain EEG signals of 62 channels [9], where we use the 62-channel of as the *full-channel* signals and making the *few-channel* signals by selecting 18 electrodes from 62 electrodes according to the distribution of electrodes in different brain regions. The 18 electrodes are Fp1/2, Fz, F3/4, F7/8, Cz, C3/4, T7/8, Pz, P3/4, P7/8.

Modified SEED-IV. As described in Section 4.1.2, SEED-IV contains EEG signals of 15 subjects. We simulate the cross-device EEG emotion recognition task by splitting the 15 subjects into two groups consisting of 12 and 3 subjects. Namely, the 12 subjects have paired 62-channel and 18-channel EEG data, while the other 3 subjects only have 18-channel EEG data.

Modified SEED-V. Similarly, we split the data of 16 subjects in SEED-V into two groups consisting of 12 and 4 subjects. Namely, 12 subjects in the Modified SEED-V dataset have paired 62-channel and 18-channel EEG data, while the other 4 subjects only have 18-channel EEG data.

We evaluate the performance of our method with/without transfer learning training on the Modified SEED-IV and Modified SEED-V datasets. The experiments are conducted in a subject-dependent setting, where experiments without/with transfer learning are conducted: (1) *without transfer learning:* there are 24 trials (16 for training and 8 for testing) in each session of Modified SEED-IV, and 15 trials (10 training and 5 for testing) in modified SEED-V. Accuracies are reported by averaging the recognition results of the corresponding subjects in each dataset. (2) *with transfer learning:* the full-channel (62-electrode) and few-channel (18-electrode) paired data of 12 subjects are used to train the meta-model in Figure 3, and the few-channel of the other 3 subjects are used in the fine-tuning and testing stages. The results are shown in Table 5. We observe that the recognition accuracies of our full method are about 7.4% and 4.4% higher than those not using our transfer learning strategy on the modified SEED-IV and modified SEED-V datasets, respectively. The transfer learning strategy is effective in enhancing the performance of few-channel EEG emotion recognition.

4.5.2 Subject-Independent Evaluation

We also evaluate our CD-EmotionNet with subject-independent evaluation on our proposed CDEED dataset and the two simulated datasets, i.e., modified SEED-IV and modified SEED-V. Similar to the subject-dependent setting, we conduct experiments without/with our transfer learning training: (1) *without transfer learning:* For our proposed CDEED/SEED-IV/SEED-V datasets, the EEG data of 25/12/12 subjects are used for training, and data of the other 13/3/4 subjects are used for testing. For the SEED-IV dataset, the experiments are carried out within each session, and the results are reported by averaging the recognition result of each session. (2) *with transfer learning:*

9. https://bcmi.sjtu.edu.cn/home/seed/seed-iv.html

TABLE 6
Subject-independent evaluation on KangII and simulated datasets. The mean and standard deviation (%) of accuracies are shown.

Method	KangII	Modified SEED-IV 18-channel	Modified SEED-V 18-channel
	ACC / STD (%)	ACC / STD (%)	ACC / STD (%)
without transfer learning	19.00 / 4.02	43.56 / 4.05	34.41 / 3.60
with transfer learning	**24.99 / 5.95**	**45.65 / 3.70**	**40.62 / 7.40**

TABLE 7
Cross-device cross-electrode evaluation on KangII and simulated datasets. The mean and standard deviation (%) of accuracies are shown.

Method	KangII	Modified SEED-IV 18-channel	Modified SEED-V 18-channel
	ACC / STD (%)	ACC / STD (%)	ACC / STD (%)
without transfer learning	41.15 / 8.64	82.04 / 13.57	73.87 / 16.12
with transfer learning	**45.89 / 7.21**	**84.29 / 11.39**	**76.38 / 12.78**

the full-channel and few-channel paired data (25, 12 and 12 subjects for CDEED, modified SEED-IV and modified SEED-V respectively) are used to train the meta-model in Figure 3, and the few-channel of the other subjects (13, 3 and 4 subjects for CDEED, modified SEED-IV and modified SEED-V respectively) are used for testing. There are no fine-tuning operations in this setting. The experiment with transfer learning under the subject-independent setting is a more challenging task, which involves *cross-device* and *cross-subject* at the same time. The results are presented in Table 6, showing that our proposed transfer learning strategy is able to optimize the network performance.

4.5.3 Cross-device Cross-electrode Evaluation

We also evaluate our CD-EmotionNet with cross-device cross-electrode evaluation on the three datasets, i.e., the CDEED dataset, the modified SEED-IV, and modified SEED-V datasets. In the cross-device cross-electrode setting, the meta-model training stage in Figure 3 was conducted using data from all electrodes in the few-channel EEG, as well as the remaining electrodes from the full-channel EEG that were not included in the few-channel EEG. On the other hand, the fine-tuning and testing stage was performed only with the few-channel EEG data. In other words, for the CDEED dataset, we utilized the 5-channel signals from the SleepUp2.0 Naolubrain EEG device (i.e., Fp1/2, F7/8, and O1), as well as the 13-channel signals from Wearable Sensing's wireless DSI-24 (i.e., Fz, F3/4, Cz, C3/4, T3/4, A1/2, Pz, P3/4, T5/6, and O2), to train the meta-model. Subsequently, testing was conducted using the 5-channel EEG data from the 5 electrodes in the testing dataset. The data setting of cross-device cross-electrode evaluation for the modified SEED-IV and modified SEED-V datasets follows a similar approach. We conduct the cross-device cross-electrode experiments under the subject-dependent setting. The experiment extends the application of our proposed transfer learning strategy to extract emotion-related EEG features that remain invariant not only across devices but also across electrodes. The results are shown in Table 7, demonstrating the capability of our proposed transfer learning strategy to acquire cross-device and cross-electrode EEG emotion features.

4.5.4 Application Scenarios

We further set our CD-EmotionNet method to solve the other cross-device EEG analysis task, i.e., depression detection with few-channel EEG. MODMA [56] is a mental-disorder dataset that contains 128-electrode of 53 subjects and 3-electrode EEG signals of 55 subjects. We use the 3-electrode EEG data offered by the authors [56], which has been filtered by a finite impulse response (FIR) filter of [1 Hz-45 Hz] and an adaptive noise canceller to remove eyeblink artifacts. We conduct subject-independent experiments using PSD features extracted from five frequency bands. Depression detection is performed as a binary classification task. Compared to the method in [72] with reported highest accuracy 72.25%, the classification accuracy is 75.96% and 80.08% using our model without and with transfer learning training, demonstrating that our proposed CD-EmotionNet method can be generalized to other applications. The full investigation of comprehensive comparisons in this application will be conducted in future work.

5 Conclusion

In this paper, we propose a transfer learning framework (CD-EmotionNet) to enhance the emotion recognition of few-channel EEG signals with the help of full-channel EEG: (1) We propose a base emotion model that integrates GCNs and the feature aggregation mechanism to extract efficient EEG emotion features. The model is first trained with full-channel and few-channel EEG data from existing subjects and then fine-tuned on few-channel EEG data from new subjects to capture individual characteristics. (2) We build a new CDEED dataset with 18-channel and 5-channel paired data to evaluate the proposed framework. The dataset contains EEG recordings from 38 subjects under emotional video stimuli. 25 subjects have paired 18-channel and 5-channel EEG, which are used for training the base model. 13 new subjects have only 5-channel EEG, which are used for model fine-tuning and testing. (3) Experiments on our proposed CDEED dataset and three existing datasets show that our method achieves state-of-the-art performance on the few-channel EEG emotion recognition task.

Our method is able to process EEG signals collected by portable and compact EEG devices, so as to promote the everyday application of EEG signals. We also show a specific application of our transfer learning method to tackle depression detection with 3-channel EEG signals, showing that our proposed method can be generalized to more mental disorder detection tasks. Future work includes exploring the relationship of the electrodes between full-channel and few-channel EEG devices and utilizing it to further enhance the emotion recognition performance across different EEG devices.

References

[1] H. Kim, S. Kim, H. Kim, Y. Ji, and C.-H. Im, "Modulation of driver's emotional states by manipulating in-vehicle environment: Validation with biosignals recorded in an actual car environment," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 1783–1792, 2022.

[2] H. Huang, Q. Xie, J. Pan, Y. He, Z. Wen, R. Yu, and Y. Li, "An eeg-based brain computer interface for emotion recognition and its application in patients with disorder of consciousness," *IEEE Transactions on Affective Computing*, vol. 12, no. 4, pp. 832–842, 2019.

[3] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[4] 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 Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[5] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 98–107, 2017.

[6] P. Lakhan, N. Banluesombatkul, V. Changniam, R. Dhithijaiyratn, P. Leelaarporn, E. Boonchieng, S. Hompoonsup, and T. Wilaiprasitporn, "Consumer grade brain sensing for emotion recognition," *IEEE Sensors Journal*, vol. 19, no. 21, pp. 9896–9907, 2019.

[7] J. Shen, S. Zhao, Y. Yao, Y. Wang, and L. Feng, "A novel depression detection method based on pervasive eeg and eeg splitting criterion," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2017, pp. 1879–1886.

[8] J. Shen, X. Zhang, G. Wang, Z. Ding, and B. Hu, "An improved empirical mode decomposition of electroencephalogram signals for depression detection," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 262–271, 2019.

[9] J. Shen, Y. Zhang, H. Liang, Z. Zhao, Q. Dong, K. Qian, X. Zhang, and B. Hu, "Exploring the intrinsic features of eeg signals via empirical mode decomposition for depression recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022.

[10] E. Ratti, S. Waninger, C. Berka, G. Ruffini, and A. Verma, "Comparison of medical and consumer wireless eeg systems for use in clinical trials," *Frontiers in human neuroscience*, vol. 11, p. 398, 2017.

[11] FCC, "Fcc id 2aynf-kang," https://fccid.io/2AYNF-KANG, 2021, accessed: 2022-08-16.

[12] R. M. Todd, V. Miskovic, J. Chikazoe, and A. K. Anderson, "Emotional objectivity: Neural representations of emotions and their interaction with cognition," *Annual review of psychology*, vol. 71, pp. 25–48, 2020.

[13] H. Saarimäki, E. Glerean, D. Smirnov, H. Mynttinen, I. P. Jääskeläinen, M. Sams, and L. Nummenmaa, "Classification of emotion categories based on functional connectivity patterns of the human brain," *NeuroImage*, vol. 247, p. 118800, 2022.

[14] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, "Real-time movie-induced discrete emotion recognition from eeg signals," *IEEE Transactions on Affective Computing*, vol. 9, no. 4, pp. 550–562, 2017.

[15] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[16] Y. Li, W. Zheng, Z. Cui, T. Zhang, and Y. Zong, "A novel neural network model based on cerebral hemispheric asymmetry for eeg emotion recognition." in *IJCAI*, 2018, pp. 1561–1567.

[17] X. Du, C. Ma, G. Zhang, J. Li, Y.-K. Lai, G. Zhao, X. Deng, Y.-J. Liu, and H. Wang, "An efficient lstm network for emotion recognition from multichannel eeg signals," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1528–1540, 2020.

[18] T. Song, S. Liu, W. Zheng, Y. Zong, and Z. Cui, "Instance-adaptive graph for eeg emotion recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 03, pp. 2701–2708, Apr. 2020. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/5656

[19] G. Zhang, M. Yu, Y.-J. Liu, G. Zhao, D. Zhang, and W. Zheng, "Sparsedgcnn: Recognizing emotion from multichannel eeg signals," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 537–548, 2023.

[20] N. Jatupaiboon, S. Pan-ngum, and P. Israsena, "Emotion classification using minimal eeg channels and frequency bands," in *The 2013 10th international joint conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2013, pp. 21–24.

[21] H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *IEEE computer society conference on computer vision and pattern recognition*. IEEE, 2010, pp. 3304–3311.

[22] P. Zhong, D. Wang, and C. Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1290–1301, 2020.

[23] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 1126–1135.

[24] W. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE Transactions on Cybernetics*, pp. 1–13, 2018.

[25] W. Liu, J.-L. Qiu, W.-L. Zheng, and B.-L. Lu, "Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 715–729, 2022.

[26] T. Åkerstedt and M. Gillberg, "Sleep duration and the power spectral density of the eeg," *Electroencephalography and clinical neurophysiology*, vol. 64, no. 2, pp. 119–122, 1986.

[27] R. R. Vázquez, H. Velez-Perez, R. Ranta, V. L. Dorr, D. Maquin, and L. Maillard, "Blind source separation, wavelet denoising and discriminant analysis for eeg artefacts and noise cancelling," *Biomedical signal processing and control*, vol. 7, no. 4, pp. 389–400, 2012.

[28] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for eeg-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2013, pp. 81–84.

[29] T. Zhang, Z. Cui, C. Xu, W. Zheng, and J. Yang, "Variational pathway reasoning for eeg emotion recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 03, pp. 2709–2716, Apr. 2020. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/5657

[30] Y.-J. Suh and B. H. Kim, "Riemannian embedding banks for common spatial patterns with eeg-based spd neural networks," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, pp. 854–862, May 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/16168

[31] N. Thammasan, K.-i. Fukui, and M. Numao, "Multimodal fusion of eeg and musical features in music-emotion recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/11112

[32] Y. Lu, W.-L. Zheng, B. Li, and B.-L. Lu, "Combining eye movements and eeg to enhance emotion recognition." in *IJCAI*, vol. 15. Buenos Aires, 2015, pp. 1170–1176.

[33] R. Li, Y. Wang, and B.-L. Lu, "A multi-domain adaptive graph convolutional network for eeg-based emotion recognition," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 5565–5573.

[34] B. Chen, Q. Cao, M. Hou, Z. Zhang, G. Lu, and D. Zhang, "Multimodal emotion recognition with temporal and semantic consistency," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3592–3603, 2021.

[35] M. Hou, Z. Zhang, C. Liu, and G. Lu, "Semantic alignment network for multi-modal emotion recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 9, pp. 5318–5329, 2023.

[36] W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, and X. Chen, "Eeg-based emotion recognition via channel-wise attention and self attention," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 382–393, 2023.

[37] Z. Jia, Y. Lin, X. Cai, H. Chen, H. Gou, and J. Wang, "Sst-emotionnet: Spatial-spectral-temporal based attention 3d dense network for eeg emotion recognition," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2909–2917.

[38] X. Du, X. Deng, H. Qin, Y. Shu, F. Liu, G. Zhao, Y.-K. Lai, C. Ma, Y.-J. Liu, and H. Wang, "Mmpose: Movie-induced multi-label positive emotion classification through eeg signals," *IEEE Transactions on Affective Computing*, pp. 1–14, 2022.

[39] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2018.

[40] 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 Transactions on Affective Computing*, 2022, doi=10.1109/TAFFC.2022.3170428.

[41] C. Li, P. Li, Y. Zhang, N. Li, Y. Si, F. Li, Z. Cao, H. Chen, B. Chen, D. Yao *et al.*, "Effective emotion recognition by learning discrimi-

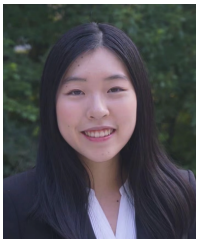
native graph topologies in eeg brain networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
[42] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in eeg signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, 2021.
[43] H. Tang, G. Jiang, and Q. Wang, "Deep neural network for emotion recognition based on meta-transfer learning," *IEEE Access*, vol. 10, pp. 78 114–78 122, 2022.
[44] Y.-P. Lin, "Constructing a personalized cross-day eeg-based emotion-classification model using transfer learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1255–1264, 2020.
[45] X. Chai, Q. Wang, Y. Zhao, Y. Li, D. Liu, X. Liu, and O. Bai, "A fast, efficient domain adaptation technique for cross-domain electroencephalography (eeg)-based emotion recognition," *Sensors*, vol. 17, no. 5, p. 1014, 2017.
[46] Y. Cimtay and E. Ekmekcioglu, "Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition," *Sensors*, vol. 20, no. 7, p. 2034, 2020.
[47] W.-L. Zheng and B.-L. Lu, "Personalizing eeg-based affective models with transfer learning," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, ser. IJCAI'16. AAAI Press, 2016, p. 2732–2738.
[48] L.-M. Zhao, X. Yan, and B.-L. Lu, "Plug-and-play domain adaptation for cross-subject eeg-based emotion recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, pp. 863–870, May 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/16169
[49] H. Chen, S. Sun, J. Li, R. Yu, N. Li, X. Li, and B. Hu, "Personalzscore: Eliminating individual difference for eeg-based cross-subject emotion recognition," *IEEE Transactions on Affective Computing*, 2021, doi=10.1109/TAFFC.2021.3137857.
[50] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, and H. He, "Multisource transfer learning for cross-subject eeg emotion recognition," *IEEE transactions on cybernetics*, vol. 50, no. 7, pp. 3281–3293, 2019.
[51] B. Liu, X. Chen, X. Li, Y. Wang, X. Gao, and S. Gao, "Align and pool for eeg headset domain adaptation (alpha) to facilitate dry electrode based ssvep-bci," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 2, pp. 795–806, 2022.
[52] M. Nakanishi, Y.-T. Wang, C.-S. Wei, K.-J. Chiang, and T.-P. Jung, "Facilitating calibration in high-speed bci spellers via leveraging cross-device shared latent responses," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 4, pp. 1105–1113, 2020.
[53] F. Kuang, L. Shu, H. Hua, S. Wu, L. Zhang, X. Xu, Y. Liu, and M. Jiang, "Cross-subject and cross-device wearable eeg emotion recognition using frontal eeg under virtual reality scenes," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2021, pp. 3630–3637.
[54] H. Xu and X. Xu, "Lightweight eeg classification model based on eeg-sensor with few channels," in *2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, 2019, pp. 464–473.
[55] J. Shen, X. Zhang, X. Huang, M. Wu, J. Gao, D. Lu, Z. Ding, and B. Hu, "An optimal channel selection for eeg-based depression detection via kernel-target alignment," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2545–2556, 2020.
[56] H. Cai, Y. Gao, S. Sun, N. Li, F. Tian, H. Xiao, J. Li, Z. Yang, X. Li, Q. Zhao *et al.*, "Modma dataset: a multi-modal open dataset for mental-disorder analysis," *arXiv preprint arXiv:2002.09283*, 2020.
[57] T. Chen, R. Hong, Y. Guo, S. Hao, and B. Hu, "Ms ˆ{2}-gnn: Exploring gnn-based multimodal fusion network for depression detection," *IEEE Transactions on Cybernetics*, 2022.
[58] J. Shen, Y. Zhang, H. Liang, Z. Zhao, K. Zhu, K. Qian, Q. Dong, X. Zhang, and B. Hu, "Depression recognition from eeg signals using an adaptive channel fusion method via improved focal loss," *IEEE Journal of Biomedical and Health Informatics*, 2023.
[59] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for eeg-based emotion classification," in *International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2013, pp. 81–84.
[60] F. R. Chung and F. C. Graham, *Spectral graph theory*. American Mathematical Soc., 1997, vol. 92.
[61] R. Lin, J. Xiao, and J. Fan, "Nextvlad: An efficient neural network to aggregate frame-level features for large-scale video classification," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, September 2018.
[62] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "Netvlad: Cnn architecture for weakly supervised place recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5297–5307.
[63] D. A. Grant, "The latin square principle in the design and analysis of psychological experiments." *Psychological bulletin*, vol. 45, no. 5, p. 427, 1948.
[64] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.
[65] X. Shen, X. Liu, X. Hu, D. Zhang, and S. Song, "Contrastive learning of subject-invariant eeg representations for cross-subject emotion recognition," *IEEE Transactions on Affective Computing*, 2022, doi=10.1109/TAFFC.2022.3164516.
[66] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.
[67] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[68] F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, and S. Coleman, "Emotion recognition with convolutional neural network and eeg-based efdms," *Neuropsychologia*, vol. 146, p. 107506, 2020.
[69] M. Ghifary, W. B. Kleijn, and M. Zhang, "Domain adaptive neural networks for object recognition," in *Pacific Rim international conference on artificial intelligence*. Springer, 2014, pp. 898–904.
[70] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The journal of machine learning research*, vol. 17, no. 1, pp. 2096–2030, 2016.
[71] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation," in *International Conference on Machine Learning*. PMLR, 2020, pp. 6028–6039.
[72] Q. Shi, A. Liu, R. Chen, J. Shen, Q. Zhao, and B. Hu, "Depression detection using resting state three-channel eeg signal," *arXiv preprint arXiv:2002.09175*, 2020.

Fang Liu is currently a Lecturer at Communication University of China. She received her Ph.D. degree from the University of the Chinese Academy of Sciences (UCAS), Beijing, China, in 2021. She worked as a postdoctoral fellow at Tsinghua University from 2021 to 2023. Her research interests include computer vision, sketch interaction, and affective computing.

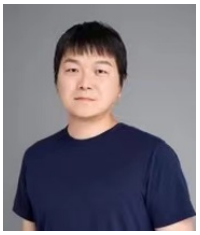
Pei Yang is is currently a Ph.D student in Department of Computer Science and Technology, Tsinghua University. He received his bachelor's degree from Inner Mongolia University, China, in 2009 and his master's degree from Minzu University of China in 2016. His research interests include emotion recognition and machine learning.

Yezhi Shu is a Ph.D student with Department of Computer Science and Technology, Tsinghua University. She received her B.Eng. degree from Shandong University, China, in 2019. Her research interests include affective computing, computer vision, deep learning algorithms and applications.

Niqi Liu is an M.S. student with Department of Computer Science and Technology, Tsinghua University. She received her B.Eng. degree from Tsinghua University, China, in 2022. Her research interests include cognitive computation and deep learning.



Jenny Sheng is currently a master student with the Department of Computer Science and Technology, Tsinghua University, China. She received her B.S.E. degree in Computer Science from Princeton University, USA, in 2022. Her research interests include intelligent media processing, human-computer interaction, and computer vision.



Junwen Luo obtained his PhD degree from The University of Newcastle Upon Tyne, UK. He was an academic visitor at MIT in 2011. He leads the BrainUp Research Lab now. His research interest is brain-machine Interface and neuromorphic computing.

Xiaoan Wang received her master degree from Harvard University in Cambridge, US in 2015. She is the founder of BrainUp Inc. Her business focus is on brain-computer interface and its applications.

Yong-Jin Liu is a Professor with Department of Computer Science and Technology, Tsinghua University, Beijing, China. He received the B.Eng. degree from Tianjin University, Tianjin, China, in 1998, and the M.Phil. and Ph.D. degrees from the Hong Kong University of Science and Technology, Hong Kong, China, in 2000 and 2004, respectively. His research interests include computational geometry, computer vision, cognitive computation and pattern analysis. For more information, visit http://cg.cs.tsinghua.edu.cn/people/~Yongjin/Yongjin.htm