Adaptive Hierarchical Graph Convolutional Network for EEG Emotion Recognition

Yunlong Xue
School of Biological Science and Medical Engineering
Southeast University
Nanjing, China
xueyunlong@seu.edu.cn

Wenming Zheng*
School of Biological Science and Medical Engineering
Southeast University
Nanjing, China
wenming_zheng@seu.edu.cn

Yuan Zong
School of Biological Science and Medical Engineering
Southeast University
Nanjing, China
xhzongyuan@seu.edu.cn

Hongli Chang
School of Biological Science and Medical Engineering
Southeast University
Nanjing, China
hongli_chang@seu.edu.cn

Xingxun Jiang
School of Biological Science and Medical Engineering
Southeast University
Nanjing, China
jiangxingxun@seu.edu.cn

2022 International Joint Conference on Neural Networks (IJCNN) | 978-1-7281-8671-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/IJCNN55064.2022.9892411

***Abstract*—Human emotion is closely related to multiple distributed brain regions, and functional connections exist between the regions. However, how to abstract the region-level information to improve electroencephalograph (EEG) emotion recognition performance has not been well considered. To address this problem, we proposed a novel Adaptive Hierarchical Graph Convolutional Network (AHGCN), which includes the basic channel-level graph of EEG channels and the region-level graph of brain regions. Different from previous methods, we propose an adaptive pooling operation to automatically partition brain regions rather than manually define them. To capture the intrinsic functional connections between the brain regions or EEG channels, we design a gated adaptive graph convolution operation. Besides, we develop a graph unpooling operation to integrate the region-level graph and channel-level graph to extract more discrimination features for classification. Experiments on two widely-used datasets show that our proposed method is superior to many state-of-the-art methods on EEG emotion recognition and could find some interesting combinations of EEG channels.**

***Index Terms*—EEG emotion recognition, graph convolutional neural network (GCNN), graph pooling**

I. INTRODUCTION

Emotion recognition is an important task in the emotional brain-machine interfaces, which plays a crucial role in human-machine interaction and health care [1]. Among various signals, electroencephalograph (EEG) signals can reflect the intrinsic emotion states with high temporal resolution. Therefore, EEG signals have drawn increasing attention in analyzing human emotion.

EEG channels are distributed on irregular grids. Traditional deep learning frameworks like convolutional neural networks (CNN) cannot well represent such non-Euclidean structured data. The graph-based methods provide an effective way to model EEG signals. Several recent works [2]–[4] employed graph convolutional neural networks (GCNN) to model the relations of different EEG channels, further extracting more discriminative features for EEG emotion recognition, in which the nodes represent EEG channels and the links mean relationships between EEG channels. However, these methods ignore the natural hierarchical structure of the brain, which not only includes the basic micro channel-level graph but also has the macro region-level graph with nodes representing the brain regions. Psychological study [5], [6] have revealed that human emotion perception draws on a distributed set of structures that include multiple brain regions and functional connections exist between the regions. These discoveries mean that human emotion can be represented by the information of the emotion-relevant brain regions and the connections between them to a certain extent. It is hard to automatically learn such a high-level structure from data by GCNN itself, so in order to capture region-level information for EEG emotion recognition, there are two critical problems to consider: i) how to detect the brain regions that may contain different numbers of EEG channels; ii) how to learn the intrinsic relations between the regions to capture the region-level information for classification.

Recently, some attempts have been made on this issue. Song et al. [4] manually aggregated the nodes into some macro nodes by a graph coarsening operation, which is designed according to spatial locations of EEG electrodes. Li et al. [7] also manually divided the electrodes into some brain regions. They first extract the features of each region and then integrate the features of the regions to obtain global features. However, imperceptible neuralmechanism in the brain makes it inaccurate to define the meaningful combination of EEG channels manually.

To address the aforementioned issues, we propose a novel Adaptive Hierarchical Graph Convolutional Network (AHGCN). The basic idea is to automatically aggregate EEG

978-1-7281-8671-9/22/$31.00 ©2022 IEEE

channels into brain regions for capturing region-level information. Based on this idea, an adaptive graph pooling operation is proposed to partition brain regions and construct the region-level graph flexibly. Specifically, this operation is based on a cluster assignment matrix that is adaptively generated by fusing spatial and frequency information of each EEG sample. To capture the intrinsic relations between the regions or the EEG channels, a gated graph convolution operation is defined, in which an adaptive graph connections operation and a gated structure are designed to acquire the most emotion-relevant functional connections and graph features, respectively. In addition, we develop a graph unpooling operation based on the cluster assignment matrix, which bridges the two graphs to get more rich information. In the experiment, our AHGCN achieves the state-of-the-art performance on two widely used public EEG emotion datasets, i.e., SEED [8] and SEED-IV [9].

II. RELATED WORK

This section reviews the related work from three aspects: EEG emotion recognition, graph convolutional neural network and hierarchical graph representation.

A. EEG Emotion Recognition

EEG emotion recognition algorithms have been evolved from traditional machine learning algorithms to deep learning. Traditional machine learning algorithms have been studied for many years. Liu et al. [10] designed an adaptive support vector machine to recognize different emotions and achieved promising results. Li et al. [11] extended the conventional linear regression and proposed a graph regularized sparse linear regression (GRSLR) model to improve the performance of EEG emotion recognition. As deep learning has achieved excellent performance in many fields, many deep learning methods have been applied to identify EEG emotions. Tripathi et al. [12] employed deep neural networks (DNN) and CNN to extract high-level information from EEG signals and achieved higher accuracy than traditional machine learning algorithms. Song et al. [2] used GCNN to model the intrinsic relations among different EEG channels and then perform EEG emotion classification. In this paper, we also model the EEG signals based on GCNN, but we consider both EEG channels features and brain regions features.

B. Graph Convolutional Neural Network

Graph neural network (GNN) aims to build neural networks to deal with data in graph domains, e.g., molecular structures, social networks, and knowledge graphs. Graph convolutional neural network is an extension of the conventional CNN, which is more advantageous in coping with feature extraction of data in discrete spatial domain [13]. GCNN has two different research directions: spatial domain methods and spectral domain methods [14]. The spatial domain methods operate on graph node features and neighbor aggregation to model spatial relationships. The spectral domain methods transform the signals into spectral domain through spectral graph theory [15]. Due to the excellent performance of GCNNs in modeling irregular data, such as skeleton-based action recognition [16], [17] and recommendation system [18], many works introduced GCNN into EEG emotion recognition. Song et al. [2] first introduced GCNN to EEG emotion recognition and employed a learnable adjacency matrix to characterize the relationship between EEG channels. Zhong et al. [3] further improved the adjacency matrix to capture both local and global inter-channel relations. However, there are functional regions in the brain, and the modeling of EEG channels by graph convolution cannot capture the high-level information well.

C. Hierarchical Graph Representation

There are several recent works focusing on constructing hierarchical graph structures to capture more information. In skeleton-based action recognition, Thakkar et al. [19] proposed a part-based graph convolutional network that manually divides the joint graph into several body parts subgraphs. They first extract the features of subgraphs and then propagate information between subgraphs. In traffic forecasting, Guo et al. [20] constructed hot traffic regions graph based on road network graph by spectral clustering, and then extracted features of the two graphs by graph convolution and fuse the features for classification. In EEG emotion recognition, Song et al. [21] fused local features extracted by CNN and global features extracted by GCNN to obtain more discriminative features for classification.

Our proposed model is also hierarchical structure. Specifically, considering that the generation of emotion is associated with a distributed network consisting of multiple brain regions [22], we design a graph pooling operation to aggregate EEG channels into meaningful regions in the scalp and then fuse the region-level information and the low-level relations between EEG channels.

III. METHOD

A. Overview

The framework of the proposed AHGCN is shown in Fig. 1. The input to our model is hand-crafted features extracted from EEG signals. The features are first represented in the form of graphs, i.e., the channel-level graph. To capture region-level information, we design an adaptive graph pooling to generate the region-level graph. After the two levels of graphs are constructed, they are respectively fed into a gated adaptive graph convolution layer to fuse the information among different graph nodes in order to obtain a better graph representation. To further combine the features of the two graphs, a graph unpooling operation is developed to align the features of the two graphs and get a more discriminative graph representation for classification. Then all node features are concatenated and fed to a full connection layer. Finally, a softmax layer is used to output the predicted labels. We will describe the key components of AHGCN in detail in the sequel.

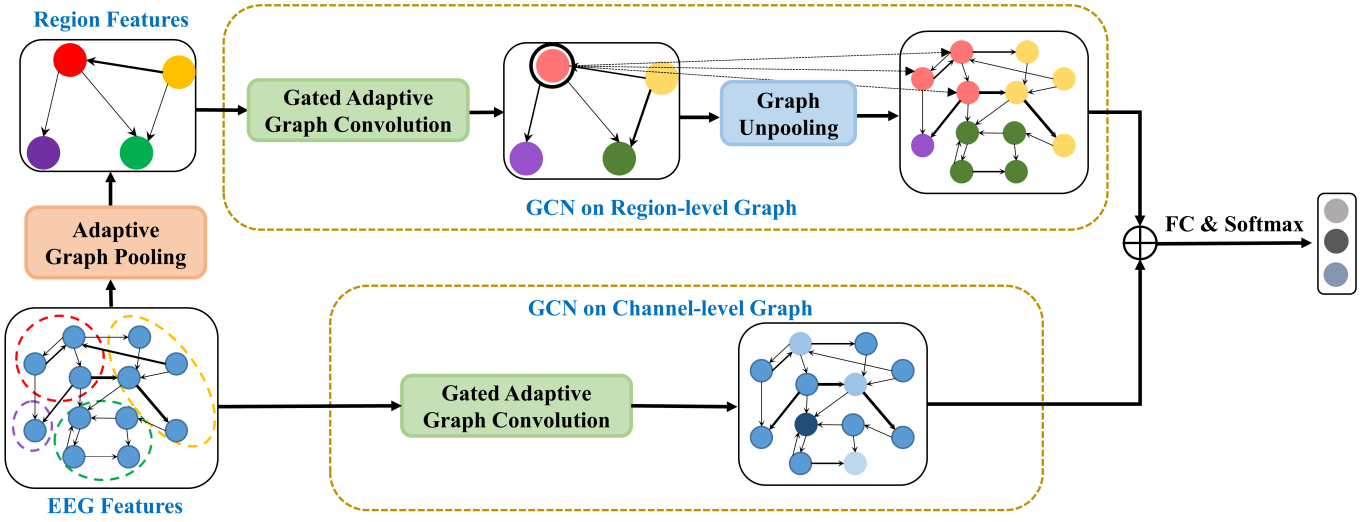

Fig. 1: The framework of the proposed AHGCN. The region-level graph is generated from the channel-level graph through adaptive graph pooling operation. The graph features of the channel-level graph and the region-level graph are fused to recognize the emotion classed.

B. Hierarchical Graphs Generated by Adaptive Graph Pooling

Our model contains two hierarchical graphs: channel-level graph and region-level graph, where the region-level graph is adaptively generated from the channel-level graph by adaptive graph pooling.

1) Channel-level graph: A weighted directed graph can be represented in this form: $\mathcal{G}=\{\mathcal{V},\mathcal{E},\boldsymbol{A}\}$, in which $\mathcal{V}$ represents the set of graph nodes, $\mathcal{E}$ represents the set of edges and $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ represents the adjacency matrix whose element $A_{ij}>0$ characterize the relation from node v_i to v_j. For EEG signals, let $\boldsymbol{X}\in\mathbb{R}^{n\times d}$ denote the EEG feature matrix, where n is the number of EEG channels and d is the number of frequency bands. For the construction of channel-level graph, we regard EEG channels as graph nodes and the relationships between EEG channels as adjacency matrix. However, Inadequate understanding of brain mechanisms makes it difficult to predefine these connections. Therefore, we set $\boldsymbol{A}$ to be the parameter that can learn the intrinsic connections between EEG channels dynamically. We use $\boldsymbol{A}=Relu\left(\boldsymbol{B}\right)$ to obtain a non-negative adjacency matrix, where $\boldsymbol{B}\in\mathbb{R}^{n\times n}$ is a parameter matrix.

2) Region-level graph: For constructing the macro graph of regions from EEG channels, the ideal way is to use functional brain regions as nodes. However, how to detect and recognize these regions in the complex brain network is another challenging issue. Complex neural mechanisms in the brain make it hard to predefine these regions.

Given the channel-level graph representation, we design an adaptive graph pooling method to aggregate the nodes to regions. Fig. 2 shows the general process of how to generate the region-level graph from the channel-level graph. We first propose to generate a weighted assignment matrix $\boldsymbol{R}\in\mathbb{R}^{n\times n_r}$, in which n_r is the number of regions. R_{ij} represents the probability that the i-th EEG channel belongs to the j-th brain region. $\boldsymbol{R}$ can be calculated as follows:

$$\boldsymbol{R}=softmax\left(\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}\boldsymbol{X}\boldsymbol{Q}\right), \tag{1}$$

where $\boldsymbol{D}\in\mathbb{R}^{n\times n}$ denotes the diagonal matrix with entries $D_{ii}=\sum_j A\left(i,j\right)$. $\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ is the normalized form of $\boldsymbol{A}$. $\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ is used to fuse spatial information and $\boldsymbol{Q}\in\mathbb{R}^{d\times n_r}$ is used to fuse the frequency information. The softmax operation is applied to the rows. In the case of comprehensive consideration of spatial and frequency domain information, the generation of the weight assignment matrix can be more reasonable. Moreover, the weight assignment matrix can be adaptively changed according to the EEG samples, which can characterize the differences between different emotions and different subjects to a certain extent. Besides, we impose an entropy minimization constraint on each row of $\boldsymbol{R}$, such that the rows of $\boldsymbol{R}$ are approximately one-hot vectors, i.e., the original graph is divided into non-overlapped subgraphs.

Given the weighted assignment matrix, the next step is to determine the node features and edge connections of the region-level graph. The new node features $\boldsymbol{X}_r\in\mathbb{R}^{n_r\times d}$ and the new coarsened adjacency matrix $\boldsymbol{A}_r\in\mathbb{R}^{n_r\times n_r}$ of the region-level graph can be formulated as:

$$\boldsymbol{X}_r=norm\left(\boldsymbol{R}^{\mathrm{T}},1\right)\boldsymbol{X}, \tag{2}$$

$$\boldsymbol{A}_r=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{R}. \tag{3}$$

In (2), L1-normalization is performed along columns of $\boldsymbol{R}$, and then the nodes are aggregated averagely and generate the new macro nodes. This process is similar to average pooling. In (3), when calculating the relations between two regions, we comprehensively consider the connection between two nodes in the regions and generate the region adjacency matrix denoting the connections between the regions.

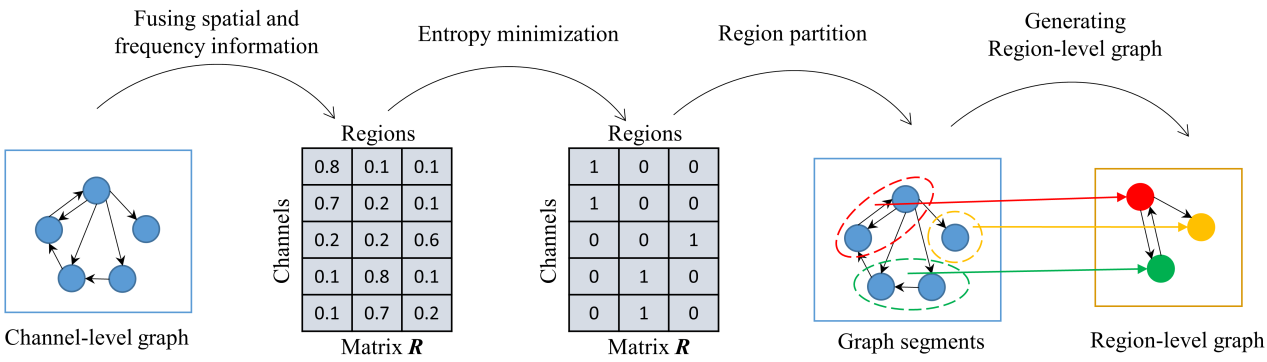

Fig. 2: The overall process of the region-level graph generation. By fusing the spatial and frequency information of EEG channels, the weight assignment matrix $\boldsymbol{R}$ is generated under the effect of entropy minimization. Based on $\boldsymbol{R}$, the region-level graph is generated from the channel-level graph.

C. Gated Adaptive Graph Convolution Layer

After the channel-level graph and region-level graph are constructed, they are respectively fed into a gated adaptive graph convolution layer to capture the intrinsic connections between EEG channels and regions. Here, we use the region-level graph as an example to illustrate this operation.

The fixed adjacency matrix $\boldsymbol{A}_r$ cannot represent the changes of different emotions and subjects and dynamics of EEG signals. To adaptively capture the important emotion-relevant edges, inspired by GAT [23] and self-attention mechanism [24], we use an adaptive mask to learn a data-dependent graph connections $\tilde{\boldsymbol{A}}_r \in \mathbb{R}^{n_r \times n_r}$ for each sample:

$$\tilde{\boldsymbol{A}}_r = \left({\boldsymbol{D}_r}^{-1/2} \boldsymbol{A}_r {\boldsymbol{D}_r}^{-1/2} \right) \odot \boldsymbol{M}, \tag{4}$$

$$\boldsymbol{M} = sigmoid\left(\boldsymbol{X}_r \boldsymbol{W}_1 (\boldsymbol{X}_r \boldsymbol{W_2})^{\mathrm{T}} \right), \tag{5}$$

where $\boldsymbol{D}_r \in \mathbb{R}^{n_r \times n_r}$ and ${\boldsymbol{D}_r}^{-1/2} \boldsymbol{A}_r {\boldsymbol{D}_r}^{-1/2}$ is the normalized form of $\boldsymbol{A}_r$. $\odot$ is the element-wise production. $\boldsymbol{M} \in \mathbb{R}^{n_r \times n_r}$ is a weighted mask adjusted according to the input features. $\boldsymbol{W}_1 \in \mathbb{R}^{d \times d_m}$ and $\boldsymbol{W}_2 \in \mathbb{R}^{d \times d_m}$ embed the node features into a higher-level directed graph embedding space and then calculate the dot product similarity between regions to measure the importance of the edges.

According to [4], graph convolutional operation can be repeated to model a high-level connections between nodes. ${\tilde{\boldsymbol{A}}_r}^k$, the k-order polynomial of $\tilde{\boldsymbol{A}}_r$, can express the relationship between nodes after k-step graph convolution operations. Thus, to capture different level information of the graph, the graph convolution operation f_{GCN} can be defined as

$$f_{GCN}\left(\boldsymbol{X}_r, \boldsymbol{A}_r\right) = \sum_{k=0}^{K} {\tilde{\boldsymbol{A}}_r}^k \boldsymbol{X}_r \boldsymbol{W}_k, \tag{6}$$

where $\boldsymbol{W}_k \in \mathbb{R}^{d \times d_1}$ is the weight matrix. This process provides a better way to integrate information from different levels of the graph.

Particularly, Different EEG channels and feature dimensions contribute differently to emotion recognition. In order to automatically acquire most remotion-relevant graph features, inspired by gated CNN [25], we design a gated graph convolution operation and the output of the region-level graph $\boldsymbol{Y}_r \in \mathbb{R}^{n_r \times d_1}$ is formulated as following:

$$\begin{aligned} \boldsymbol{Y}_r = Relu\left(\sum_{i=0}^{K} {\tilde{\boldsymbol{A}}_r}^i \boldsymbol{X}_r \boldsymbol{W}_i \right) \\ \odot sigmoid\left(\sum_{j=0}^{K} {\tilde{\boldsymbol{A}}_r}^j \boldsymbol{X}_r \boldsymbol{W}_j \right), \end{aligned} \tag{7}$$

where $\boldsymbol{W}_i \in \mathbb{R}^{d \times d_1}$ and $\boldsymbol{W}_j \in \mathbb{R}^{d \times d_1}$ is the learnable parameters. Here, The first part is the graph features after graph convolution. For the second part, we use an additional graph convolution operation to control which outputs should be concerned.

As for the channel-level graph, it shares the same structure to the region-level graph, except for different input data and learnable parameters. Here, we use $\boldsymbol{Y} \in \mathbb{R}^{n \times d_1}$ to represent the output of the channel-level graph after graph convolution.

D. Interaction between channel-level graph and region-level Graph by Graph Unpooling

After the gated adaptive graph convolution, we get better graph representations, and the next step is to combine the features of the two graphs. However, due to the adaptability of the region-level graph, i.e., different EEG samples may have different styles of brain regions partition, it is difficult to align features between the channel-level graph and the region-level graph. In graph pooling operation, we use $\boldsymbol{R}$ to generate the region-level graph, which contains the correspondence between EEG channels and brain regions. Here, we also can use it to develop a graph unpooling operation. This operation on the region-level graph can be formulated as:

$$\boldsymbol{Y}_r{}' = \boldsymbol{R}\boldsymbol{Y}_r, \tag{8}$$

where $\boldsymbol{Y}_r{}' \in \mathbb{R}^{n \times d_1}$. This process allows each node in a region to have the features of the region.

Now the channel-level graph and the region-level graph have the same dimension. Then we add the features of the two graphs. The final graph output features $\boldsymbol{Y}_{all} \in \mathbb{R}^{n \times d_1}$ can be formulated as:

$$\boldsymbol{Y}_{all} = \boldsymbol{Y} \oplus \boldsymbol{Y}_r{}', \tag{9}$$

where $\oplus$ is the element-wise addition. In this way, a node has both channel-level and region-level features. Then all node features are connected and fed to a full connection layer. Finally, a softmax layer is used to output the predicted labels.

E. Loss Function

The loss function used in our model includes two parts: cross entropy loss and entropy minimization loss of $\boldsymbol{R}$. The total loss function is formulated as follows:

$$\mathcal{L} = -\sum_{c=1}^{C} y_c log\hat{y}_c - \lambda \sum_{i=1}^{n} \sum_{j=1}^{n_r} R_{ij} log R_{ij}. \tag{10}$$

The first term is the cross entropy loss, where C denotes number of classes and $y_c \in \{0, 1\}$. If the sample belongs to the c-th class, then $y_c = 1$, otherwise $y_c = 0$. $\hat{y}_c \in [0, 1]$ represents the probability value predicted by our model that the sample belongs to c-th class. The second term is the entropy minimization loss, where λ is the coefficients for the entropy minimization. This loss function ensures that the rows of $\boldsymbol{R}$ are approximately one-hot vectors, i.e., each EEG channel belongs to only one region.

IV. EXPERIMENT

A. Datasets and Protocols

SEED [8] contains EEG data recorded from 15 healthy subjects (7 males) while they were watching 15 emotion-eliciting video clips. The EEG signals are collected at a sampling rate of 1000 Hz using a 62-channel electrode cap that conforms to the international 10-20 system. The video clips contain three types of emotions, namely positive, neutral and negative, with five video clips for each emotion. Each clip lasts about 4 minutes. Each subject participated in the experiment three times at different time periods, so there are three sessions of EEG data, and each session contains 15 trials for each subject. For comparison with previous literature, we follow the widely employed subject-dependent protocol in [8]. Specifically, the first nine trials of each session are used as training data and the remaining six trials are used as testing data. Then we calculate the average accuracy and standard deviation of 30 sessions (two sessions per subject) as evaluation metrics to measure our model.

SEED-IV [9] contains EEG data recorded from 15 healthy subjects (7 males) while they were watching 24 emotion-eliciting video clips. The EEG signals are collected at a sampling rate of 1000 Hz using a 62-channel electrode cap that conforms to the international 10-20 system. The video clips contain four types of emotions, including happy, sad, fear, and neutral, with six video clips for each emotion. Each clip lasts about 2 minutes. Each subject participated in the experiment three times at different time periods, so there are three sessions of EEG data and each session contains 24 trials for each subject. We follow the subject-dependent protocol used in [9]. Specifically, The last eight trials of each session containing all four emotions (two trials for each emotion) are used as test data, and the remaining 16 trials are used as training data. Then we calculate the average accuracy and standard deviation of all 45 sessions (three sessions per subject) as evaluation metrics to measure our model.

B. Preprocessing

For the SEED dataset, we follow the same preprocessing methods in [8]. The EEG signals are down-sampled with 200 Hz and the component in five frequency bands (δ: 1-3 Hz, θ: 4-7 Hz, α: 8-13 Hz, β: 14-30 Hz, γ: 31-50 Hz) is filtered. The Differential Entropy (DE) features, the input to our model, are pre-computed by using a 256-point short-time Fourier transform with a non-overlapped Hanning window of 1s.

For the SEED-IV dataset, we follow the same preprocessing methods in [9]. The EEG signals are down-sampled with 128 Hz and sliced into 4-second non-overlapping segments. Then the component in five frequency bands (δ: 1-3 Hz, θ: 4-7 Hz, α: 8-13 Hz, β: 14-30 Hz, γ: 31-50 Hz) is filtered. Similar to SEED, the DE features are pre-computed over five frequency bands in each channel.

C. Implementation Details

In the experiment, the input to our model is the DE features extracted from five frequency bands, so the number of EEG channels (n) is 62, and the number of frequency bands (d) is 5. For graph pooling, the original 62 EEG channels are clustered into 10 regions for SEED and 14 regions for SEED-IV. For the graph convolution part, the order of graph convolution (K) is set to 2, and the transformed dimension (d_1) is set to 32. As for the coefficient λ, we fix it at 0.1 throughout the experiments. The proposed AHGCN is implemented by Pytorch. During the model training, we use the Adam to optimize the model with a learning rate of 0.001 and a batch size of 32.

V. RESULT

A. Result Analysis and Comparison

To verify the validity of our model, we compare our AHGCN with various existed algorithms on SEED and SEED-IV, respectively, including SVM [26], GSCCA [27], DBN [8], STRNN [28], DGCNN [2], BiDANN [29], BiHDM [30], R2G-STNN [7], RGNN [3], IAG [4] and V-IAG [31]. All these methods follow the same subject-dependent protocol with our AHGCN. Table I and Table II present the classification accuracy and standard deviation of our AHGCN and all baselines on SEED and SEED-IV using the pre-computed DE features.

Deep learning models outperform the traditional SVM. Our AHGCN improves the current state-of-the-art results on both SEED and SEED-IV. Our model achieves 96.72% on SEED and 82.78% on SEED-IV, which verifies the superiority of our AHGCN. Compared with other graph-based methods,

TABLE I: The results (mean accuracy / standard deviation) on SEED

Method	training /testing	ACC (%)	STD (%)
SVM [26]	9/6	83.99	9.72
GSCCA [27]	9/6	82.96	9.95
DBN [8]	9/6	86.08	8.34
STRNN [28]	9/6	89.50	7.63
DGCNN [2]	9/6	90.40	8.49
BiDANN [29]	9/6	92.38	7.04
BiHDM [30]	9/6	93.12	6.06
R2G-STNN [7]	9/6	93.38	5.96
RGNN [3]	9/6	94.24	5.95
IAG [4]	9/6	94.89	6.16
V-IAG [31]	9/6	95.64	5.08
AHGCN (our model)	9/6	**96.72**	**4.58**

TABLE II: The results (mean accuracy / standard deviation) on SEED-IV

Method	training /testing	ACC (%)	STD (%)
SVM [26]	16/8	56.61	20.05
GSCCA [27]	16/8	69.08	16.66
DBN [8]	16/8	66.77	7.38
DGCNN [2]	16/8	69.88	16.29
BiDANN [29]	16/8	70.29	12.63
BiHDM [30]	16/8	74.35	14.09
RGNN [3]	16/8	79.37	10.54
AHGCN (our model)	16/8	**82.78**	11.93

i.e., DGCNN, RGNN, IAG, and V-IAG, our model achieves better classification results. Although they are all graph-based methods, our hierarchical graph structure can capture both channel-level and region-level information and get better graph representation. Also, AHGCN has an improvement in contrast to other domain adaptive methods, including BiDANN, BiHDM, and R2G-STNN, which involve test data for network optimization. Our AHGCN does not use test data during training but still achieves better performance. This verifies the high generalization of our AHGCN.

Confusion matrices are presented in Fig. 3. For the SEED database, our model can recognize better for positive and neutral emotions than negative emotion. A similar phenomenon is observed in SEED-IV. For the SEED-IV database, our model performs better on happy and neutral emotion than sad and fear emotion. The findings indicate that participants watching positive or neutral movies may generate similar EEG patterns.

B. Ablation Studies

To further verify the validity of the vital component in our model, we conduct an ablation study. The results are shown in Table III.

Firstly, we estimate the impact of the region-level graph branch in our model. We present the result using our AHGCN

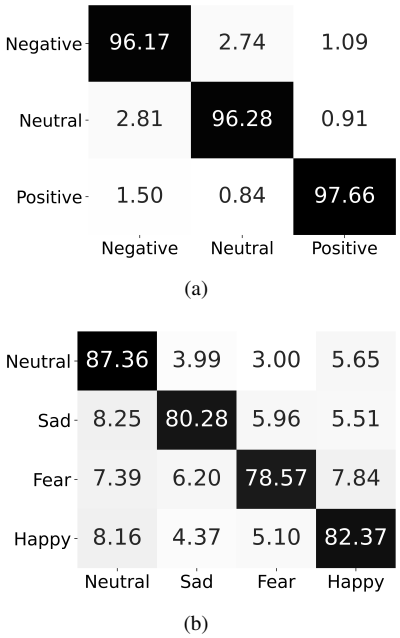

Fig. 3: The confusion matrices of EEG recognition results using the proposed AHGCN on the SEED and SEED-IV databases. (a) Confusion matrix on the SEED; (b) Confusion matrix on the SEED-IV.

TABLE III: The results (mean accuracy / standard deviation) of ablation studies

Method	SEED	SEED-IV
w/o region-level graph	94.66 / 5.11	80.79 / 12.76
w/o channel-level graph	93.04 / 6.27	79.61 / 14.17
Shared learnable matrix	96.30 / 5.34	81.63 / 13.21
Fixed 16 regions pooling	95.62 / 5.30	80.76 / 13.52
Fixed 17 regions pooling	96.20 / 4.91	81.89 / 13.46
w/o adaptive structure	95.57 / 5.21	81.44 / 13.53
w/o gated structure	95.98 / 5.37	82.26 / 13.51
AHGCN	**96.72 / 4.58**	**82.78 / 11.93**

'w/o' denotes 'without'

without the region-level graph. Without the region-level graph, we can see that the classification accuracy for EEG emotion recognition drops by about 2% on both databases, demonstrating that the region-level graph we designed can capture useful region-level information which cannot be easily captured by the regular graph convolution. Also, we remove the channel-level graph branch, then the performance also drops, indicating that channel-level information is also important for EEG emotion recognition.

The adaptive graph pooling operation is the key part of the region-level graph branch. To validate the efficiency of the adaptive graph pooling operation, we compare it with different variants. We first replace $\boldsymbol{R}$ with two different fixed regions partition methods in [4], [7], namely, fixed 16 regions pooling and fixed 17 regions in Table III. Specifically, they

manually defined regions based on spatial location or cortical regions. We also present the result of replacing $\boldsymbol{R}$ with a learnable matrix, i.e., all samples share the same learnable region partition. As shown in Table III, the two learnable region partition methods perform better than the fixed partition methods, proving that automatic partition is more reasonable, while manual partition may introduce some errors. Moreover, our adaptive pooling operation performs better than the learnable matrix, indicating that adaptive pooling can capture the differences between different EEG samples.

The two major designs in our gated adaptive graph convolution, i.e., adaptive graph connections and gated structures, are helpful in recognizing emotions. The adaptive connections can dynamically capture the important connections between EEG channels or brain regions. The gated structure, generated by a graph convolution operation, can better capture the most emotion-relevant graph features.

C. Visualization of Region Partition

The aforementioned experiments verify the competitive performance of our proposed AHGCN. To better illustrate the region partition our method learned, we here visualize some crucial regions. The degree centrality is a validated index measuring connectivity of a node with other nodes, which has been widely used to evaluate the importance of the nodes in the graph [32]. We use the degree centrality in the region-level graph to measure the importance of the regions. The degree centrality C_i of the i-th regions can be defined by

$$C_i = \sum_{n=1}^{n_r} A_r\left(i,n\right) + \sum_{m=1}^{n_r} A_r\left(m,i\right) - 2A_r\left(i,i\right). \qquad (11)$$

Fig. 4 shows the top 3 regions having the largest degree centrality. We have the following findings:

- There is a large variation of the region partition among subjects, where the regions have different locations and combinations of EEG channels.
- In statistics of positions, there are some commonalities between different subjects: EEG channels within the crucial regions frequently appear in the frontal lobe and temporal lobe, which means the channels in these regions contribute significantly to emotion recognition. These results coincide with the previous cognition observations of biological psychology [33].
- Most of the EEG channels within each region are concentrated in one area of the scalp, but a few are scattered, which indicates that the functional relationship between EEG channels is related to, but not entirely dependent on, the spatial distance.

VI. Conclusion

In this paper, we proposed a novel graph-based method for EEG emotion recognition called AHGCN to capture the information of both the channel-level graph and the region-level graph. The adaptive graph pooling operation provides a more flexible way to partition the brain regions dynamically.

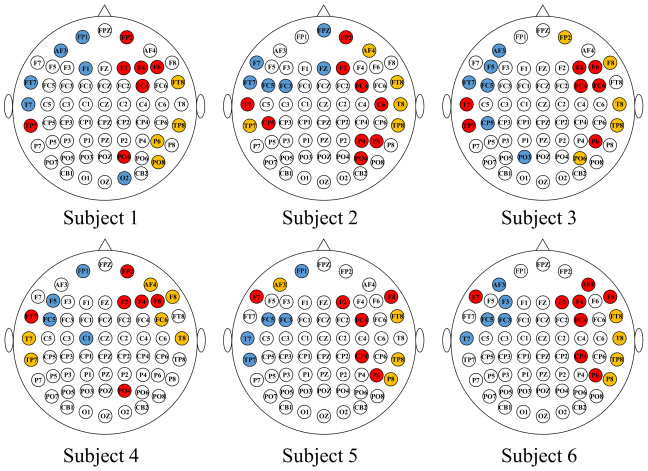

Fig. 4: The regions learned in the AHGCN are visualized by the top 3 regions having the largest degree centrality of different subjects on SEED. Each color represents a region, and the importance of these colors is in descending order: blue, red, yellow.

Given the two levels of graphs, the gated adaptive graph convolution is employed to diffuse information between graph nodes so as to obtain a better graph representation. Besides, integrating the features of the region-level and channel-level graph, achieved by the graph unpooling operation, can abstract richer information which is helpful for EEG emotion recognition. Extensive experiments on SEED and SEED-IV demonstrated the superiority of our model, and some regions are shown for intuitive understanding, which may also be rather meaningful from the view of human emotion cognition.

VII. Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) under the Grants 62076064 and 61921004, and in part by the Fundamental Research Funds for the Central Universities under Grant 2242022k30036.

References

[1] R. B. Knapp, J. Kim, and E. André, "Physiological signals and their use in augmenting emotion recognition for human–machine interaction," in *Emotion-oriented systems*. Springer, 2011, pp. 133–159.

[2] 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.

[3] P. Zhong, D. Wang, and C. Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, 2020.

[4] T. Song, S. Liu, W. Zheng, Y. Zong, and Z. Cui, "Instance-adaptive graph for EEG emotion recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 03, 2020, pp. 2701–2708.

[5] R. Adolphs, "Neural systems for recognizing emotion," *Current opinion in neurobiology*, vol. 12, no. 2, pp. 169–177, 2002.

[6] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature reviews neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.

[7] Y. Li, W. Zheng, L. Wang, Y. Zong, and Z. Cui, "From regional to global brain: A novel hierarchical spatial-temporal neural network model for eeg emotion recognition," *IEEE Transactions on Affective Computing*, 2019.

[8] 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.

[9] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 1110–1122, 2018.

[10] Y. H. Liu, C. T. Wu, Y. H. Kao, and Y. T. Chen, "Single-trial eeg-based emotion recognition using kernel eigen-emotion pattern and adaptive support vector machine," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2013, no. 2013, pp. 4306–4309, 2013.

[11] Y. Li, W. Zheng, Z. Cui, Y. Zong, and S. Ge, "Eeg emotion recognition based on graph regularized sparse linear regression," *Neural Processing Letters*, vol. 49, no. 2, pp. 555–571, 2019.

[12] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on deap dataset." in *Twenty-ninth IAAI conference*, 2017.

[13] F. Petroski Such, S. Sah, M. Dominguez, S. Pillai, C. Zhang, A. Michael, N. Cahill, and R. Ptucha, "Robust spatial filtering with graph convolutional neural networks," *arXiv e-prints*, pp. arXiv–1703, 2017.

[14] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.

[15] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, 2014.

[16] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 12 026–12 035.

[17] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," 2018.

[18] J. Zhao, Z. Zhou, Z. Guan, W. Zhao, W. Ning, G. Qiu, and X. He, "Intentgc: a scalable graph convolution framework fusing heterogeneous information for recommendation," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2347–2357.

[19] K. Thakkar and P. Narayanan, "Part-based graph convolutional network for action recognition," *arXiv preprint arXiv:1809.04983*, 2018.

[20] K. Guo, Y. Hu, Y. Sun, S. Qian, J. Gao, and B. Yin, "Hierarchical graph convolution networks for traffic forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, 2021, pp. 151–159.

[21] T. Song, W. Zheng, S. Liu, Y. Zong, Z. Cui, and Y. Li, "Graph-embedded convolutional neural network for image-based eeg emotion recognition," *IEEE Transactions on Emerging Topics in Computing*, 2021.

[22] J. E. LeDoux, "Emotion circuits in the brain," *Annual review of neuroscience*, vol. 23, no. 1, pp. 155–184, 2000.

[23] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[25] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *International conference on machine learning*. PMLR, 2017, pp. 933–941.

[26] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[27] W. Zheng, "Multichannel eeg-based emotion recognition via group sparse canonical correlation analysis," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 3, pp. 281–290, 2016.

[28] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 839–847, 2018.

[29] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, and X. Zhou, "A bi-hemisphere domain adversarial neural network model for eeg emotion recognition," *IEEE Transactions on Affective Computing*, 2018.

[30] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, and T. Song, "A novel bi-hemispheric discrepancy model for eeg emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 2, pp. 354–367, 2020.

[31] T. Song, S. Liu, W. Zheng, Y. Zong, Z. Cui, Y. Li, and X. Zhou, "Variational instance-adaptive graph for eeg emotion recognition," *IEEE Transactions on Affective Computing*, 2021.

[32] X. Zhang, G. Cheng, and Y. Qu, "Ontology summarization based on rdf sentence graph," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 707–716.

[33] J. A. Coan and J. J. B. Allen, "Frontal eeg asymmetry as a moderator and mediator of emotion," *Biological Psychology*, vol. 67, no. 1-2, pp. 7–49, 2004.