# Variational Instance-Adaptive Graph for EEG Emotion Recognition

Tengfei Song, Suyuan Liu, Wenming Zheng*, *Senior Member, IEEE*, Yuan Zong, Zhen Cui, *Member, IEEE*, Yang Li, Xiaoyan Zhou

***Abstract*—The individual differences and the dynamic uncertain relationships among different electroencephalogram (EEG) regions are essential factors that limit EEG emotion recognition. To address these issues, in this paper, we propose a variational instance-adaptive graph method (V-IAG) that simultaneously captures the individual dependencies among different EEG electrodes and estimates the underlying uncertain information. Specifically, we employ two branches, i.e., instance-adaptive branch and variational branch, to construct the graph. Inspired by the attention mechanism, the instance-adaptive branch generates the graph based on the input so as to characterize the individual dependencies among EEG channels. The variational branch generates the probabilistic graph, which quantifies the uncertainties. We combine these two types of graphs to extract more discriminative features. To present more precise graph representation, we propose a new operation named the multi-level and multi-graph convolution operation, which aggregates the features of EEG channels from different frequencies with different graphs. Furthermore, we design the graph coarsening and employ the sparse constraint to obtain more robust features. We conduct extensive experiments on three widely-used EEG emotion recognition databases, i.e., SJTU emotion EEG dataset (SEED), Multi-modal physiological emotion recognition dataset (MPED) and DREAMER. The results demonstrate that the proposed model achieves the-state-of-the-art performance.**

***Index Terms*—EEG emotion recognition, graph neural network, variational inference.**

## I. Introduction

EMOTION recognition has been an effective technology to capture human's emotional states, which is significant in human-machine interaction [1]. Many researches focused on emotion recognition from non-physiological signals, including facial images [2][3], human gestures [4], voices [5][6] and texts [7]. Recently, more physiological signals, such as electrocardiogram (ECG) [8], electromyogram (EMG) [9] and galvanic skin response [10], have been applied to predict human emotions. Generally, the responses of physiological signals, especially electroencephalogram (EEG) signals, are difficult to be disguised, which is helpful to obtain authentic emotions. EEG signals directly reflect the activities of human brain that is advantageous to measure the real mental states of humans [11][12].

Understanding the brain mechanism of emotion generation is significant to develop more effective algorithm to identify different emotions via EEG signals. EEG signals are recorded by multiple electrodes placed along the scalp, distributed in irregular grid. Therefore, EEG electrodes and their relationships can be represented as graph composed of vertexes and links. Neuroscience researches [13][14] have indicated that the functional connections of brain regions and their functional organizations are highly related to the generation of emotions. Studies on neuroimaging [15] have shown that emotion categories are uniquely related to the activity of distributed neural systems that span cortical and subcortical. Generally, graph-based methods model the relationships among nodes, which can be applied to characterize the functional dependencies among different EEG channels for EEG emotion recognition. However, due to the imperceptible neuromechanism and the limited understanding about brain mechanism, it is difficult to directly define the connections between different EEG electrodes from the prior knowledge. Recently, some studies [16][17] proposed to learn the graph in a data-driven manner, which provides an effective way to exploit the functional dependencies from data. They tend to learn an optimal graph to represent the dependencies among EEG channels since there are some common dependencies in terms of different people. But the constant graph ignores the individual differences of people, which are essential for EEG emotion recognition [18][19]. Inspired by the attention mechanism [20], Song et al. [21] employed additional branch to adaptively construct the individual graphs based on the input data for EEG emotion recognition so as to alleviate the influence of individual differences.

Although some studies proposed more flexible ways to represent the dependencies among EEG channels, they are easy to be affected by the noise. There are commonly two types of noise. One is the noise of the EEG signals and the other is the noisy labels. The recording process of EEG signals requires low noisy environment. The eye movement artifact, power-line interference and the EMG interference will affect the quality of EEG signals. Besides, many EEG emotion datasets employ videos as elicitation materials. Although these

Tengfei Song is with the Key Laboratory of Child Development and Learning Science (Southeast University), Ministry of Education, Southeast University, Nanjing 210096, China, and also with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China.

Suyuan Liu, Wenming Zheng and Yuan Zong are with the Key Laboratory of Child Development and Learning Science (Southeast University), Ministry of Education, Southeast University, Nanjing 210096, China, and also with the School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China. (E-mail: wenming_zheng@seu.edu.cn).

Zhen Cui is with the school of Computer Science, Nanjing University of Science and Technology, Nanjing, Jiangsu, P.R. China.

Yang Li is with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, the School of Artificial Intelligence, Xidian University, Xi'an, 710071, China.

Xiaoyan Zhou is with the school of electronic and information engineering, Nanjing University of Information Science and Engineering Technology, Nanjing, China.

*Corresponding author

elicitation materials are assessed by psychological methods, some emotions of subjects may not be elicited correctly. Generally, all the EEG signals during the subject watching a video will be annotated to the same type of emotional label, which means some labels may be not correct since the subject will not generate the target emotion at the beginning of the video. All these noises are not inevitable and will limit the performance on EEG emotion recognition. Besides, most of the existing methods for EEG emotion recognition are deterministic mode, which cannot effectively capture the data uncertainty. Compared with deterministic model, which gives point embedding without considering the data uncertainties, the probabilistic model has better capabilities to alleviate the influence of outliers and wrong labels. Probabilistic model provides the distributional estimation parameterized with estimated mean and variance, which can generate more robust result.

To alleviate the aforementioned issues, we propose a variational instance-adaptive graph neural network (V-IAG) to simultaneously capture the individual dependencies among EEG channels and the underlying uncertain information. Since neuroscience studies [22][14] have demonstrated that the emotion processing is the directed pattern, we employ two branches, i.e., instance-adaptive branch and the variational branch to construct the directed graph for graph convolution. The instance-adaptive branch as a deterministic model can generate the deterministic graph based on input data which adaptively captures the dependencies between EEG signal channels, while the variational branch as a probabilistic model assumes the graph as the random variable, which can capture and quantify uncertainties and mitigate noise. It is intractable to directly estimate the posterior distribution of the graph such that we use variational inference to approximate the posterior distribution of the graph. To provide more precise graphic representation, we propose the multi-level and multi-graph convolution operation, which employs different graphs to aggregate the features of different EEG channels in different frequency bands. Further, we define the spatial graph coarsening and employ the sparse constraint to extract discriminative features so as to depress the noise. Finally, we conduct extensive experiments on three widely used EEG emotion datasets, i.e., SEED [23], MPED [10] and DREAMER [24], and the results show that the proposed method achieves the state-of-the-art performance. Besides, we present insight analysis about learned graph on EEG emotion recognition.

The remainder of this paper is organized as follows: In section II, we briefly review the related works. In section III, we introduce the proposed variational instance-adaptive graph neural network for EEG emotion recognition. Extensive experiments and analysis are provided in section IV. Finally, we conclude the paper in section V.

## II. Related works

### A. EEG Emotion Recognition

Recently, capturing the real emotional states from EEG signals has achieved more attentions. Feature extraction from raw EEG signals and the classification are two primary steps to implement EEG emotion recognition.

Features extracted from EEG signals can be divided into time domain features [25][26], frequency domain features [27][28] and time-frequency domain features [29]. Frequency and time-frequency features are dominant feature extraction methods for EEG emotion recognition, e.g., the power spectral density (PSD) [30] and short-time Fourier transform (STFT) [31]. The raw EEG signals are commonly decomposed into five different frequency bands, i.e., $\delta$ (1-4Hz), $\theta$ (4-8Hz), $\alpha$ (8-14Hz), $\beta$ (14-30Hz) and $\gamma$ (30-50Hz) bands so as to extract features from different frequency bands separately. Hadjidimitriou et al. [29] analyzed the performance on EEG emotion recognition from different frequency bands and validated that the information related to human emotions are more likely existing in higher frequency bands, i.e., $\beta$ and $\gamma$ bands. For EEG emotion recognition, more researches employed the energy features from different frequency bands, which can reflect the electrical activity on the scalp. Zhong et al. [32] utilized the differential entropy energy features as the input of neural networks. The energy features from five frequency bands were provided for EEG emotion recognition. Li et al. [33] fed energy features extracted from five frequency bands into an adversarial neural network so as to capture the discrepancy information between two hemispheres. Zhang et al. [34] employed the EEG energy feature from five frequency bands as the input of the broad network. The results from each frequency band were also provided for analysis. Yang et al. [35] proposed a 3D representation of EEG segment to combine the energy features in different frequency bands. George et al. [36] also extracted time-frequency features in different frequency bands and achieved better performance, which validates the effectiveness of energy features.

Based on specific EEG features, classification algorithms are applied to distinguish various emotional states. Even though features from all EEG channels are obtained, not all channels are benefit to recognize emotions. To find out the EEG channels with major contributions for emotion recognition, Zheng et al. [37] developed a group sparse canonical correlation analysis method to use less channels and achieved better performance. Recently, more methods based on deep neural networks have been demonstrated to be powerful to deal with classification tasks. In [23], a deep belief network is applied to select useful channels to improve the classification accuracy. Many researches focused on LSTM-based methods for EEG emotion recognition. Alhagry et al. [38] employed long short-term memory network (LSTM) to learn features from EEG signals and a dense layer was used to classify these features. Song et al. [10] proposed an attention-LSTM, which combined the attention mechanism and LSTM so as to select the salient EEG channels and capture the dependencies among different channels for EEG emotion recognition. Xing et al. [39] combined the Stack AutoEncoder (SAE) with LSTM to capture both spatial and temporal information. Ma et al. [40] employed LSTM to fuse multi-model signals for emotion recognition. The complementary information can be helpful to achieve better performance. Further, some studies investigate the relationships between different EEG channels, which contain significant functional dependencies. Zhang et al. [41] proposed a spatial-temporal recurrent neural network

to build the dependencies among different EEG channels in spatial dimension and temporal dimension. Li et al. [42] proposed an algorithm to explore bi-hemispheric dependency for EEG emotion recognition. Developing an efficient algorithm characterizing the functional relationships is crucial to investigate human emotions. The individual difference is a key factor to limit the generalization of EEG emotion recognition algorithm. Hu et al. [43] provided a comprehensive investigation for different features in cross-subject EEG emotion recognition. Yin et al. [44] proposed a transfer learning method to overcome the individual differences. In this paper, we exploit the intrinsic dependencies among different EEG channels and capture some underlying uncertain information for EEG emotion recognition.

### B. Graph Neural Network

As an effective way to handle irregular data, graph can be used to represent various nodes and their relationships. Graph neural networks have been widely used in many applications, including image classification [45][46], object tracking [47], neural architecture search [48] and social science (social networks) [49][50].

Graph neural networks can be divided into three categories, i.e., token-based, graph-based and adjacency-matrix-based. Token-based methods [51][52] model graph as a sequence and build the relationships between different vertexes by RNNs. Graph-based method [53] represents a graph by adding vertexes and links according to their topological structure. Adjacency-matrix-based methods [49][45] employ the proxy matrix to characterize the dependencies among vertexes. With specific topological structures, each vertex updates their states by receiving information from neighbor nodes.

For EEG signals, the response of different electrodes and their functional dependencies are particular significant in the cognitive generation of emotions [13] such that graph has been a promising way to deal with EEG emotion recognition. Many researches [45][49] proposed low-computation graph convolution operation with the predefined graph structures, which cannot be directly applied to EEG emotion recognition. Due to the imperceptible neuromechanism, it is extremely difficult to predefine the functional relationships among different EEG regions. In [17], a dynamical convolutional neural network was proposed to learn the connections among different EEG electrodes, which provides a promising way to exploit the relationships among different EEG channels. Zhang et al. [34] proposed to combine graph convolution network with board network, which is beneficial to learn the structure of graph. Liu et al. [54] developed an algorithm that combines attention mechanism and sparse constraint to learn a sparse graph. To alleviate the influence of individual differences, song et al. [21] proposed an instance-adaptive graph to adjust graph connections to represent different input instances for EEG emotion recognition. Besides employing the instance-adaptive module, in this paper, we further build a probabilistic graph that captures the individual differences and underlying uncertain information simultaneously.

### C. Variational Inference

Generally, there are two types of methods, i.e., Markov Chain Monte Carlo (MCMC) method [55] and variational inference [56], which can be used to estimate the posterior distribution of random variable.

Compared with MCMC method, variational inference requires small computation cost and can benefit from solidly theoretical property. Generally, to directly estimate the posterior distribution by Bayesian inference, i.e., $p(x|\mathcal{D}) = p(x)p(\mathcal{D}|x)/p(\mathcal{D})$, is intractable because the integral in $p(\mathcal{D}) = \int p(x, \mathcal{D})dx$ is computation-intractable. To this end, an auxiliary distribution function, denoted by $q(x)$, can be introduced to approximate the posterior distribution $p(x|\mathcal{D})$. $q(x)$ can be learned by optimization. The Kullback-Leibler (KL) divergence can be provided to reduce the gap between $q(x)$ and $p(x|\mathcal{D})$ such that the inference problem could be turned into optimization problem. $p(x|\mathcal{D})$ can be approximated more efficiently. Variational inference has been used to many applications, e. g., face alignment [57], neural network pruning [58] and neural network search [48]. Specifically, variational inference has been applied to EEG emotion recognition and achieves good property to mitigate the noise. Li et al. [59] proposed a variational autoencoder based latent factor decoding method for cross-subject EEG emotion recognition. Zhang et al. [60] employed the variational inference to obtain a distribution to select the right pathway for EEG emotion recognition. Variational inference is helpful to approximate a distribution to find a right way to propagate information among different EEG channels. In this paper, we employ the variational inference to approximate the posterior distribution of the graph, which is helpful to mitigate the noise and obtain more robust model.

## III. V-IAG for EEG Emotion Recognition

In this section, we first present a brief introduction to the attribute graph for EEG signals. Then we will introduce the details of the proposed V-IAG model.

### A. Attribute Graph for EEG Signals

Generally, we use the EEG energy features $\mathbf{X} \in \mathbb{R}^{n\times d}$ from five frequency bands, i.e., $\delta$ (1-4 Hz), $\theta$ (4-8 Hz), $\alpha$ (8-14 Hz), $\beta$ (14-30 Hz) and $\gamma$ (30-50 Hz), as the input, in which $n$ denotes the number of nodes (EEG channels) and $d$ denotes the number of frequency bands. A directed graph can be represented as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{A}\}$, where $\mathcal{V} = \{v_i\}_{i=1}^{n}$ is the set of nodes, $\mathcal{E}$ is the set of edges and $\mathbf{A} \in \mathbb{R}^{n\times n}$ is the adjacency matrix characterizing the connections in a graph. For graph neural network, the nodes in graph aggregate the information from neighbor nodes and obtain new states under the homogeneous topological structure. $\mathbf{A}_{i,j}$ denotes the connection from source node $v_i$ to the destination node $v_j$. If $\mathbf{A}_{i,j} = 0$, $v_i$ and $v_j$ are not connected, otherwise $\mathbf{A}_{i,j} \neq 0$.

Based on the graphic signal processing theory on directed graph [61], the Jordan decomposition of the adjacency matrix $\mathbf{A}$ can be defined as $\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}$, in which $\mathbf{F} = \mathbf{V}^{-1}$ is the graph Fourier Transform matrix and $\mathbf{J}$ is a block-diagonal matrix. The Fourier transform on $\mathbf{X}$ can be defined

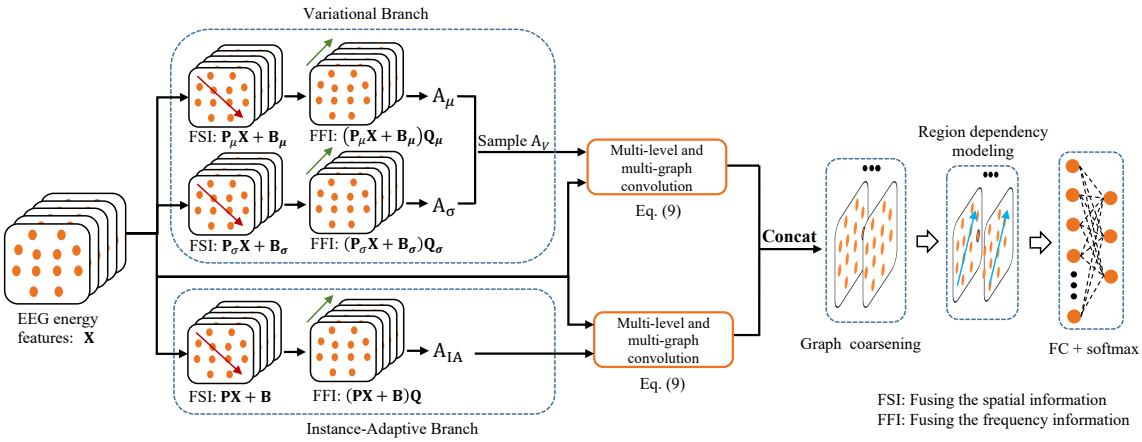



Fig. 1. The framework of the variational instance-adaptive graph (V-IAG). The instance-adaptive branch is employed to generate the deterministic graphs and the variational branch is employed to generate the sampled graphs from multiple probabilistic distributions. Then the features from two branches are concatenated and processed by graph coarsening, region dependency modeling, fully connected layer (FC) and softmax layer.

as $\hat{\mathbf{X}} = \mathbf{F}\mathbf{X}$, in which $\hat{\mathbf{X}}$ is the frequency domain signal of $\mathbf{X}$. And the inverse Fourier transform can be calculated by $\mathbf{X} = \mathbf{F}^{-1}\hat{\mathbf{X}}$. Let $h(.)$ be a filtering function and the filtering process can be written as:

$$\widetilde{\mathbf{X}} = h(\mathbf{A})\mathbf{X} = \mathbf{F}^{-1}h(\mathbf{J})\mathbf{F}\mathbf{X}, \tag{1}$$

in which $h(\mathbf{J})$ denotes the graph frequency response of the filter $h(\mathbf{A})$. The classical signal convolution theorem can be extended to graphic frequency representation.

### *B. The Variational Instance-Adaptive Graph Model*

The framework of the proposed variational instance-adaptive graph (V-IAG) model for EEG emotion recognition is shown in Fig. 1. In V-IAG model, the instance-adaptive branch is proposed to better characterize the dynamic graph connections between the EEG channels, where the graph connections can be adaptively changed along with input instance. The variational branch captures the underlying uncertain information. Then multi-level and multi-graph convolution is conducted to diffuse the information among different channels of EEG signals. Moreover, graph coarsening is adopted to achieve more robust features and reduce the computational complexity. After graph coarsening, long short-term memory networks (LSTM) is applied to build the dependencies of the features with sequential structure so as to obtain an emotion-discriminative feature vector, which can provide a good performance on EEG emotion recognition. Finally, all the hidden states of LSTM are connected to a fully connected layer and then a softmax layer is employed to output the predicted labels.

*1) Instance-Adaptive Branch:*

The dynamic functional relationships between EEG regions and the individual differences play an important role in EEG emotion recognition. Recently, attention mechanism has been extensively used for many different applications [62][63][64][65]. In this paper, attention mechanism refers to the technique of learning a mapping function to adaptively weight the EEG features according to the different inputs, such that most useful features associated with the important weights can be selected for emotion classification. Inspired by attention mechanism, we employed an additional branch (shown in Fig. 1) to adaptively generate different graphs for different instances such that the adaptive graph would be more powerful to represent the relationships among different EEG channels across different instances.

The graph generated from instance-adaptive branch is depended on input features such that we can exploit more useful information from input features. The spatial information and the frequency information are extremely important to exploit human emotions. In this paper, we propose the method of using both matrix left multiplication and right multiplication for the EEG feature matrix associated with each sample instance, resulting in the fusion of spatial and frequency information. Moreover, the operation of matrix right multiplication also provides us to dynamically generate the adjacency matrix of the graph associated with each instance. In this way, we are able to adaptively change the graph along with the input instance during training and classification processes.

The $t$th column of the $\mathbf{X}$, i.e., $\mathbf{X}_t$, corresponds to the EEG features across the various EEG electrodes in the $t$th frequency band such that we first conduct the left multiplication projection to fuse the spatial information, which can be written as

$$\mathbf{O}_t = \mathbf{P}\mathbf{X}_t + \mathbf{B}_t, (t = 1, 2, ..., d), \tag{2}$$

in which $\mathbf{P} \in \mathbb{R}^{n\times n}$ denotes the left multiplication matrix, $\mathbf{O} \in \mathbb{R}^{n\times d}$ denotes the output of the left multiplication projection and $\mathbf{B} \in \mathbb{R}^{n\times d}$ represents the bias matrix.

Since each row of $\mathbf{X}$ corresponds to the EEG features across the various frequency bands associated with one EEG electrode, we employ the right multiplication projection to fuse the emotional information among different frequency bands. The right multiplication projection can be estimated by

$$\mathbf{A}_{\text{IA}} = ReLu(\mathbf{O}\mathbf{Q}\mathbf{\Theta}), \tag{3}$$

in which $\mathbf{Q} \in \mathbb{R}^{d\times d}$ denotes the right multiplication matrix to aggregate frequency information, $\mathbf{\Theta} \in \mathbb{R}^{d\times nd}$ denotes the

projection matrix and $\mathbf{A}_{\text{IA}} \in \mathbb{R}^{n\times nd}$ represents the output. $ReLu$ function is applied to guarantee each element of $\mathbf{A}_{\text{IA}}$ non-negative and $\mathbf{A}_{\text{IA}}$ is reshaped into $d$ adjacency matrices, i.e., $[\mathbf{A}^*_{\text{IA}1}, ..., \mathbf{A}^*_{\text{IA}d}]$, to represent graphs in $d$ frequency bands.

The whole projection process in the instance-adaptive branch can be written as

$$\mathbf{A}_{\text{IA}} = f(\mathbf{X}, \mathbf{P}, \mathbf{B}, \mathbf{Q}, \mathbf{\Theta}), \tag{4}$$

in which $\mathbf{P}, \mathbf{B}, \mathbf{Q}$ and $\mathbf{\Theta}$ are the parameters in this model. The output $\mathbf{A}_{\text{IA}}$ is changed adaptively along with the input instance $\mathbf{X}$. This branch will adaptively adjust graph connections to fit every input sample, which provides a flexible way to capture the intrinsic functional relationships between EEG channels. To obtain the normalized graph, each element of adjacency matrix, i.e., $A_{ij}$ is multiplied by $\frac{1}{\sqrt{D_{ii}D_{jj}}}$, i.e., $\mathbf{A}^{norm} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$, in which $\mathbf{D}$ is a diagonal matrix estimated by $D_{ii} = \sum_j A_{ij}$. Unless otherwise specified, we employ the normalized adjacency matrix below.

The instance-adaptive branch provides a flexible way to adjust the graph connections across different instances or subjects. Specifically, instance-adaptive branch can generate some common graph connections and some individual graph connections during the training stage, where the common graph connections may have good generalization capabilities and hence would be helpful in non-user dependence experiment. The individual connections may provide more exact dynamic dependencies for each subject.

*2) Variational Branch:*

To capture the underlying uncertain information and mitigate the noise, we employ a probabilistic method to generate the matrix $\mathbf{A}_{\mathbf{V}} \in \mathbb{R}^{n\times nd}$ and assume each element of $\mathbf{A}_{\mathbf{V}}$ is random variable. And then, we can combine the result of graph convolution with deterministic graph and the probabilistic graph. Let $\mathcal{D}$ be the training data and we try to obtain the posterior distribution $p(\mathbf{A}_{\mathbf{V}}|\mathcal{D})$. Directly estimating the posterior distribution $p(\mathbf{A}_{\mathbf{V}}|\mathcal{D}) = p(\mathbf{A}_{\mathbf{V}})p(\mathcal{D}|\mathbf{A}_{\mathbf{V}})/p(\mathcal{D})$ is intractable since the integral in $p(\mathcal{D}) = \int p(\mathcal{D}, \mathbf{A}_{\mathbf{V}})d(\mathbf{A}_{\mathbf{V}})$ is computation-intractable. Thus, rather than using the MCMC method, we use variational inference [66] to approximate the posterior distribution $p(\mathbf{A}_{\mathbf{V}}|\mathcal{D})$.

In the variational branch, we use another distribution $q(\mathbf{A}_{\mathbf{V}})$ to approximate the real $p(\mathbf{A}_{\mathbf{V}}|\mathcal{D})$. By this way, we estimate the posterior distribution $p(\mathbf{A}_{\mathbf{V}}|\mathcal{D})$ by minimizing the Kullback-Leibler (KL) divergence, i.e., $KL[q(\mathbf{A}_{\mathbf{V}})||p(\mathbf{A}_{\mathbf{V}}|\mathcal{D})]$. We can minimize the KL divergence by maximizing the evidence lower bound (ELBO), which can be written as:

$$\begin{aligned}\mathcal{L}_{ELBO} = \sum_{(x,y)\in\mathcal{D}} & \mathbb{E}_{q(\mathbf{A}_{\mathbf{V}})}[log(p(y|x, \mathbf{A}_{\mathbf{V}})] \\ & - KL[q(\mathbf{A}_{\mathbf{V}})||p(\mathbf{A}_{\mathbf{V}})],\end{aligned} \tag{5}$$

in which the ELBO contains two terms, i.e., the expected log likelihood and KL divergence. The expected log likelihood aims to maximize the probability of the model prediction and the KL divergence is a regularization term. Because of the expectation in Equation (5), we cannot directly calculate the gradient to optimize the model such that the reparameterization trick [66] is employed to solve this problem.

We assume each element of $\mathbf{A}_{\mathbf{V}}$ follows Gaussian distribution, i.e., $\mathbf{A}_{\mathbf{V}i,j} \sim \mathcal{N}(\mathbf{A}_{\mu_{i,j}}, \mathbf{A}_{\sigma i,j})$. We assume $\epsilon \in \mathbb{R}^{n\times nd}$ is a random variable matrix and each element of $\epsilon$ follows standard normal distribution, i.e., $\epsilon_{i,j} \sim \mathcal{N}(0, 1)$. We generate the mean matrix $\mathbf{A}_{\mu} \in \mathbb{R}^{n\times nd}$ and the standard deviation (std) matrix $\mathbf{A}_{\sigma} \in \mathbb{R}^{n\times nd}$, which can be calculated by

$$\begin{aligned}\mathbf{A}_{\mu} &= (\mathbf{P}_{\mu}\mathbf{X} + \mathbf{B}_{\mu})\mathbf{Q}_{\mu}\mathbf{\Theta}_{\mu}, \\ \mathbf{A}_{\sigma} &= log(exp((\mathbf{P}_{\sigma}\mathbf{X} + \mathbf{B}_{\sigma})\mathbf{Q}_{\sigma}\mathbf{\Theta}_{\sigma}) + 1),\end{aligned} \tag{6}$$

in which $\mathbf{P}_{\mu}, \mathbf{P}_{\sigma} \in \mathbb{R}^{n\times n}$, $\mathbf{B}_{\mu}$, $\mathbf{B}_{\sigma} \in \mathbb{R}^{n\times d}$, $\mathbf{Q}_{\mu}, \mathbf{Q}_{\sigma} \in \mathbb{R}^{d\times d}$ and $\mathbf{\Theta}_{\mu}, \mathbf{\Theta}_{\sigma} \in \mathbb{R}^{d\times nd}$ are parameter matrices. Then we can obtain the sampled $\mathbf{A}_{\mathbf{V}}$ by reparameterization trick, which can be formulated as following:

$$\mathbf{A}_{\mathbf{V}} = \mathbf{A}_{\mu} + \mathbf{A}_{\sigma} \odot \epsilon, \tag{7}$$

in which $\odot$ is element-wise production. To guarantee each element of $\mathbf{A}_{\mathbf{V}}$ non-negative, we employ a $ReLu$ function $\mathbf{A}_V$=$ReLu(\mathbf{A}_{\mathbf{V}})$. Further, $\mathbf{A}_V$ can be reshaped into $d$ adjacency matrices, i.e., $[\mathbf{A}^*_{V1}, ..., \mathbf{A}^*_{Vd}]$, to characterize the graph connections in $d$ frequency bands.

The variational branch is a probabilistic model, which provides the distribution of the graph with estimated mean and variance. The estimated variance can capture the dependencies among EEG channels and the underlying uncertain information, which is helpful to mitigate the influence of outliers and noisy data.

*3) Multi-Level and Multi-Graph Convolutional Kernels:*

Generally, to extract high-level features in CNN, the convolutional operation with local square spatial kernels is conducted. While for graphs, we employ adjacency matrix $\mathbf{A}$ to model the connections between different nodes under the irregular structure, which is based on the graph filtering theories [61]. To extract more discriminative features based on the connections among EEG channels, we conduct repeated graph convolutional operations on the EEG signals by calculating $\mathbf{A}^k$. $\mathbf{A}^k$ represents the kernel of $k$-step graph convolution operations. Consequently, we can model these graphs by a polynomial of $\mathbf{A}$ to consider the graphs with different levels. Let $\varphi_k(\mathbf{A}) = \mathbf{A}^k$ denote the $k$-order polynomial and the multi-level graph convolution can be represented as following:

$$\mathbf{H} = \mathcal{G} * \mathcal{F} = \sum_{k=0}^{K-1} \varphi_k(\mathbf{A})\mathbf{X}, \tag{8}$$

in which $\varphi_k(\mathbf{A})$ is the $k$-th level graph and $\mathbf{H}$ is the output. This process provides a potential way to consider the information from different levels, i.e., different hops.

Specifically, the different relationships among EEG channels in terms of different frequency bands should also be taken into consideration. Therefore, we employ multiple graphs to model EEG features from different frequency bands respectively, via $d$ adjacency matrices $[\mathbf{A}^*_1, ..., \mathbf{A}^*_d]$, which provide more specific graph representations. Multi-Level and Multi-Graph convolution for EEG features in different frequency bands can be represented as:

$$\begin{aligned}\mathbf{H} &= g(\mathbf{A}, \mathbf{X}) \\ &= Cat\left[\sum_{k=0}^{K-1} \varphi_k(\mathbf{A}^*_1)\mathbf{X}_{\mathbf{1}}, ..., \sum_{k=0}^{K-1} \varphi_k(\mathbf{A}^*_d)\mathbf{X}_{\mathbf{d}}\right]\mathbf{U}\end{aligned} \tag{9}$$

in which $Cat[]$ denotes concatenation operation, $\mathbf{U} \in R^{d \times d'}$ represents the dimension transformation matrix and $\mathbf{H} \in R^{n \times d'}$ denotes the output. This graph convolution operation provides a multi-level and multi-graph representation, which aggregates the information between EEG channels in different frequency bands so as to extract more discriminative EEG features.

We use the adjacency matrices generated from instance-adaptive branch and variational branch to do multi-level and multi-graph convolution separately, which can be written as

$$\mathbf{H_{VIA}} = Cat[g(\mathbf{A_{IA}}, \mathbf{X}), g(\mathbf{A}_V, \mathbf{X})]. \tag{10}$$

$\mathbf{H_{VIA}} \in \mathbb{R}^{n \times 2d'}$ is the concatenated feature, which contains the result from deterministic graph and the result from probabilistic graph.

*4) Graph Coarsening:*

In CNNs, the pooling operation is used to reduce the redundant information and reduce computational complexity. Similarly, graph needs down-sampling to abstract the graph features and accelerate the computation. Pooling operation is usually conducted according to the spatial locations or the distance of features between different nodes. For EEG signals, clustering nodes with small feature distance fails to capture dynamic characteristics, which induces the bad performance on EEG emotion recognition. Thus, we implement the pooling operation based on the locations of EEG electrodes to avoid losing important dynamic information.

The EEG feature after graph convolution operation is divided into $n'$ groups and clustered into $n'$ nodes. The neighbor nodes are aggregated and the pooling operation on graphs, i.e., graph coarsening, can be written as:

$$\mathbf{Z}_p = \frac{1}{n_{te} - n_{ts}} \sum_{l=n_{ts}}^{n_{te}} \mathbf{H_{VIA}}l, (p = 1, 2, ..., n') \tag{11}$$

in which $\mathbf{Z}$ represents the nodes after graph coarsening and the $p$-th group includes nodes from number $n_{ts}$ to number $n_{te}$. This operation is to obtain the mean of neighbor nodes, which can be regarded as the specific region. The details about location division for graph coarsening are shown in Fig. 2. After graph coarsening, region dependency modeling (LSTM layer), fully connected layer (FC) and softmax layer are provided to predict the EEG emotion one-hot labels.

*5) Loss Function:*

To optimize our model, we should calculate the KL divergence between $q(\mathbf{A_V})$ and $p(\mathbf{A_V})$ in Equation (5). We assume the distribution of each element of $\mathbf{A_V}$ is independent with each other such that the distribution of $q(\mathbf{A_V})$ and $p(\mathbf{A_V})$ can be written as

$$\begin{aligned} p(\mathbf{A_V}) &= \prod_{i,j} p(\mathbf{A_V}_{i,j}), \ \ \mathbf{A_V}_{i,j} \sim \mathcal{N}(\mathbf{A}^*_{\mu_{i,j}}, \mathbf{A}^*_{\sigma_{i,j}}), \\ q(\mathbf{A_V}) &= \prod_{i,j} q(\mathbf{A_V}_{i,j}), \ \ \mathbf{A_V}_{i,j} \sim \mathcal{N}(\mathbf{A}_{\mu_{i,j}}, \mathbf{A}_{\sigma_{i,j}}), \end{aligned} \tag{12}$$

in which $p(\mathbf{A_V})$ is the prior distribution and $\mathbf{A}^*_{\mu_{i,j}}$ and $\mathbf{A}^*_{\sigma_{i,j}}$ are the constant prior mean matrix and the prior standard deviation matrix. During the training process, the model tends to have small $\mathbf{A}_{\sigma_{i,j}}$ and the graph will be degraded to deterministic mode. Therefore, we assume the prior $p(\mathbf{A_V}_{i,j})$ follows the normal distribution with $\mathbf{A}^*_{\mu_{i,j}} = 0$ and $\mathbf{A}^*_{\sigma_{i,j}} = 1$. And then, the KL divergence term can be written as

$$\begin{aligned} &KL[q(\mathbf{A_V})||p(\mathbf{A_V})] \\ &\quad = \sum_{i,j} -\frac{1}{2}(log\mathbf{A}^2_{\sigma_{i,j}} - \mathbf{A}^2_{\sigma_{i,j}} - \mathbf{A}^2_{\mu_{i,j}} + 1). \end{aligned} \tag{13}$$

The first term in Equation (5) aims to maximize the probability of the model prediction and we can maximize this term by minimizing the cross entropy between the prediction and the ground truth. During the training process, we use the cross entropy loss function to estimate the dissimilarity between predicted labels and the ground truth. Besides, we employed sparse graphic connections between EEG channels to further improve the performance on EEG emotion recognition. Consequently, the loss function can be expressed as following:

$$\begin{aligned} \mathcal{L} =& cross\_entropy(\mathbf{Y}, \mathbf{Y}^p) \\ &+ KL[q(\mathbf{A_V})||p(\mathbf{A_V})] + \sum_{j=1}^{d} \alpha_j \left\| \mathbf{A}^*_{\mathbf{IA}j} \right\|_1 \end{aligned} \tag{14}$$

in which $\mathbf{Y}$ and $\mathbf{Y}^p$ denote the real label vector and the predicted one, respectively. $\alpha$ represents the trade-off weight and $\|\cdot\|_1$ is the $l_1$-$norm$. We employ $l_1$-$norm$ to constrain the deterministic adjacency matrices $[\mathbf{A}^*_{\mathbf{IA}1}, ..., \mathbf{A}^*_{\mathbf{IA}d}]$, which provides a type of sparse graphic representation. The sparsity of the adjacency matrices is controlled by $\alpha$.

Algorithm I summarizes the procedure of training the V-IAG model for EEG emotion recognition.

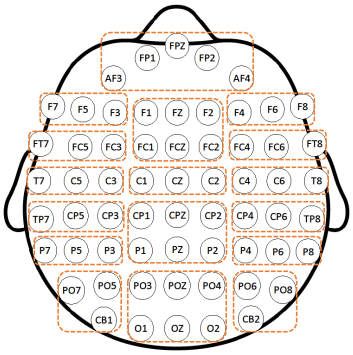



Fig. 2. The region partition for spatial graph coarsening. The 62 EEG electrodes are divided into 17 groups and the EEG electrodes of each group are clustered into one node.

## IV. Experiments

In this section, we perform our method on three widely used EEG emotion datasets, i.e., SJTU emotion EEG dataset (SEED) [23], Multi-modal physiological emotion recognition dataset (MPED)[1] [10] and DREAMER [24].

### A. EEG Emotion Dataset

SEED is a discrete EEG emotion dataset elicited by videos. The SEED dataset contains EEG signals of 15 subjects (8

[1]https://github.com/Tengfei000/MPED

**Algorithm 1** : Procedure of the V-IAG model.

Initialize model parameters
**for** each $i \in [0, epoch]$ **do**
  Calculate $\mathbf{A_{IA}}$ with $\mathbf{P}, \mathbf{B}, \mathbf{Q}, \mathbf{\Theta}$;
  Calculate $\mathbf{A}_\mu$ with $\mathbf{P}_\mu, \mathbf{B}_\mu, \mathbf{Q}_\mu, \mathbf{\Theta}_\mu$;
  Calculate $\mathbf{A}_\sigma$ with $\mathbf{P}_\sigma, \mathbf{B}_\sigma, \mathbf{Q}_\sigma, \mathbf{\Theta}_\sigma$;
  Sample $\mathbf{A_V}$ from the Gaussian distribution with $\mathbf{A}_\mu$ and $\mathbf{A}_\sigma$;
  Conduct multi-level and multi-graph convolution with $\mathbf{A_{IA}}$, $\mathbf{A_V}$ respectively, and concatenate them:

$$\mathbf{H_{VIA}} = Cat[g(\mathbf{A_{IA}}, \mathbf{X}), g(\mathbf{A_V}, \mathbf{X})].$$

  Do spatial pooling on the EEG features:

$$\mathbf{Z}_p = \frac{1}{n_{te} - n_{ts}} \sum_{l=n_{ts}}^{n_{te}} \mathbf{H_{VIA}}_l, (p = 1, 2, ..., n')$$

  Make prediction on the EEG data, calculate the loss function;
  Update parameters in the V-IAG model with gradient descent;
**end for**

females and 7 males) recorded by 62 EEG electrodes when they were watching fifteen Chinese film clips for three types of emotions, i.e., neutrality, positivity and negativity. Each film clip lasts about 4 minutes and there is enough resting time between two neighbor film clips so as to avoid long time fatigue. For each subject, the EEG signal recording process is repeated in 3 different periods corresponding to 3 sessions. Each session includes the EEG signals during watching the 15 film clips corresponding 15 trials such that there are total 45 trials of EEG signals. After watching a video, each subject conducts the self-assessment so as to guarantee that target emotion is elicited correctly. All the EEG signals at sampling rate of 1000 Hz are divided into 1-second samples and the energy features from five frequency bands are extracted.

MPED is a multi-model physiological emotion recognition dataset, which contains four types of physiological signals, i.e., electrocardiogram (EEG), electromyogram (ECG), galvanic skin response (GSR) and respiration (RSP) of 23 subjects. In this paper, we only use the EEG signals recorded by 62 EEG electrodes, which are elicited by 28 film clips with 7 types of emotions, i.e., joy, funny, angry, fear, sad, disgust and neutrality. All these elicitation materials are based on strict man-made labelling and the psychological assessment. The EEG signals of each subject contains 28 trials corresponding to 28 videos and there is enough resting time between two neighbor videos in order to avoid long time fatigue. The self-assessment is conducted after watching these elicited materials. The EEG signals at sampling rate of 1000 Hz are divided into many 1-second samples and the energy features from five frequency bands are extracted.

DREAMER is a multi-modal emotion dataset consisting of EEG and ECG signals. The EEG signals of 23 subjects were recorded using 14 electrodes at sampling rate of 256 Hz and each subject watched 18 film clips (18 trials). After watching each film clip, every subject assessed the valance and arousal by rating the self-assessment manikins (SAM). All the EEG signals are labelled with binary states (low/high valance, low/high arousal). For each trial, EEG signals of the last 60 seconds are provided for evaluation. Follow the same setting as previous works, 60 seconds signals are divided into 59 blocks by sliding a 2-second window, in which there is half overlap between two subsequent blocks. Under the experiment protocol of the previous works, we use the subject-dependent leave-one-trial-out cross-validation strategy to validate our method.

### *B. Implementation Details*

For our V-IAG model, the number of EEG channels $n$ is set to 62, the number of frequency bands $d$ is set to 5, the order of graph convolution $K$ is set to 8, and the hidden layers of LSTM is set to 64. The number of channel group $n^{'}$ after graph coarsening is 17. The trade-off parameters of the $\mathbf{A_{IA}}$ are set to $10^{-4}$, $10^{-5}$, $10^{-5}$, $10^{-5}$ and $10^{-5}$ respectively. During the training process, the learning rate is set to 0.001. The whole model is implemented on Tensorflow.

### *C. Experiment on SEED.*

To make a fair comparison on SEED dataset, we strictly follow the same experiment protocol with the previous works [23][17][67][68][21]. Subject-dependent and subject-independent classification protocols are conducted respectively. For subject-dependent experiment, we serve the first 9 trials of EEG data as training data and the remaining 6 ones as testing data. For each subject, two sessions of EEG signals are estimated, and then the mean accuracy of 15 subjects is calculated. For subject-independent protocol, we utilize the leave-one-subject-out cross validation strategy to evaluate our method. For each subject, one session of EEG data is involved in experiment. The energy feature that we use is differential entropy (DE) provided in [69].

*Ablation Study on Subject-dependent Experiment:* In this part, we provide the ablation study on subject-dependent experiment to validate the effectiveness of each module in our method. As shown in Table I, IAG (w/o IAG, pooling) is the baseline model that removes the instance-adaptive graph and

TABLE I
THE MEAN ACCURACIES (ACC) AND STANDARD DEVIATIONS (STD) FOR ABLATION STUDY OF SUBJECT-DEPENDENT EXPERIMENT ON SEED.

| Method | Subject-dependent ACC / STD (%) | Subject-independent ACC / STD (%) |
|---|---|---|
| IAG (w/o IAG, pooling) | 90.25 / 08.46 | 81.61 / 08.57 |
| IAG (w/o pooling) | 93.87 / 05.95 | 85.01 / 08.24 |
| IAG (k-means pooling) | 86.39 / 09.45 | 74.36 / 11.11 |
| IAG (spatial pooling) | 94.89 / 06.16 | 85.24 / 06.86 |
| V-IAG (Only Instance) | 95.44 / 05.48 | 86.30 / 06.91 |
| V-IAG (Only Variational) | 94.78 / 05.38 | 86.81 / 05.72 |
| V-IAG | **95.64 / 05.08** | **88.38 / 04.80** |

'w/o' denotes 'without'.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3064940, IEEE Transactions on Affective Computing



TABLE II
COMPARISONS OF THE AVERAGE ACCURACIES AND STANDARD DEVIATIONS (%) OF SUBJECT-DEPENDENT EEG-BASED EMOTION RECOGNITION EXPERIMENTS ON SEED DATASET USING DIFFERENT METHODS.

| Method | $\delta$ band | $\theta$ band | $\alpha$ band | $\beta$ band | $\gamma$ band | all ($\delta, \theta, \alpha, \beta, \gamma$) |
|---|---|---|---|---|---|---|
| SVM [70] | 60.50 / 14.14 | 60.95 / 10.20 | 66.64 / 14.41 | 80.76 / 11.56 | 79.56 / 11.38 | 83.99 / 09.72 |
| DBN [23] | 64.32 / 12.45 | 60.77 / 10.42 | 64.01 / 15.97 | 78.92 / 12.48 | 79.19 / 14.58 | 86.08 / 08.34 |
| GCNN [45] | 72.75 / 10.85 | 74.40 / 08.23 | 73.46 / 12.17 | 83.24 / 09.93 | 83.36 / 09.43 | 87.40 / 09.20 |
| DGCNN [17] | 74.25 / 11.42 | 71.52 / 05.99 | 74.43 / 12.16 | 83.65 / 10.17 | 85.73 / 10.64 | 90.40 / 08.49 |
| BiDANN-S [68] | 76.97 / 10.95 | 75.56 / 07.88 | 81.03 / 09.59 | 89.65 / 09.59 | 88.64 / 09.46 | 92.38 / 07.04 |
| R2G-STNN [71] | 77.76 / 09.92 | 76.17 / 07.43 | 82.30 / 10.21 | 88.35 / 10.52 | 88.90 / 09.97 | 93.38 / 05.96 |
| GCB-net+BLS [34] | 79.98 / 08.93 | 76.51 / 09.56 | 81.97 / 11.05 | 89.06 / 08.69 | 89.10 / 09.55 | 94.24 / 06.70 |
| IAG [21] | 80.10 / 09.14 | 81.34 / 09.14 | 83.95 / 09.89 | 91.21 / 07.25 | 92.54 / 07.61 | 95.44 / 05.48 |
| V-IAG | **81.14 / 09.46** | **82.37 / 07.44** | **84.51 / 09.68** | **92.15 / 08.90** | **92.96 / 06.19** | **95.64 / 05.08** |

TABLE III
THE MEAN ACCURACIES (ACC) AND STANDARD DEVIATIONS (STD) OF SUBJECT-DEPENDENT EEG EMOTION RECOGNITION EXPERIMENT WITH EEG FEATURES FROM TWO FREQUENCY BANDS ON SEED DATASET

| Method | $\alpha$ and $\beta$ bands (ACC/STD) | $\alpha$ and $\gamma$ bands (ACC/STD) | $\beta$ and $\gamma$ bands (ACC/STD) |
|---|---|---|---|
| IAG [21] | 90.72 / 07.78 | 90.95 / 07.92 | 92.36 / 07.20 |
| V-IAG | **92.97 / 07.59** | **92.51 / 07.87** | **93.04 / 06.98** |

TABLE IV
THE MEAN ACCURACIES (ACC) AND STANDARD DEVIATIONS (STD) OF SUBJECT-DEPENDENT EEG EMOTION RECOGNITION EXPERIMENT WITH DIFFERENT TYPES OF EEG FEATURES ON SEED DATASET.

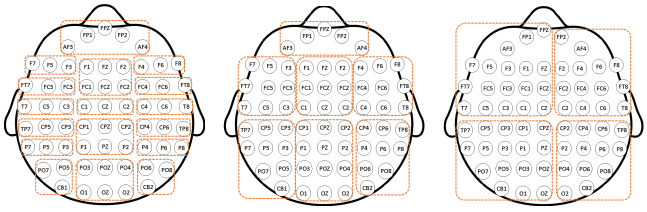
| Method | DE (ACC/STD) | PSD (ACC/STD) | DASM (ACC/STD) | RASM (ACC/STD) |
|---|---|---|---|---|
| SVM [70] | 83.99 / 09.72 | 59.60 / 15.93 | 72.81 / 16.57 | 74.74 / 14.79 |
| DBN [23] | 86.08 / 08.34 | 61.90 / 16.65 | 72.73 / 15.93 | 71.30 / 16.16 |
| DGCNN [17] | 90.40 / 08.49 | 81.73 / 09.94 | 78.45 / 11.84 | 85.00 / 12.47 |
| IAG [21] | 95.44 / 05.48 | 85.69 / 12.00 | 89.29 / 09.48 | 89.47 / 09.81 |
| V-IAG | **95.64 / 05.08** | **86.71 / 10.25** | **90.10 / 08.73** | **90.53 / 09.22** |

the spatial pooling module from IAG. IAG (w/o pooling) is the model that removes the spatial pooling module from IAG. Compared with IAG (w/o IAG, pooling), IAG (w/o pooling) achieves better performance, which validates the effectiveness of the instance-adaptive graph. For IAG (spatial pooling) and IAG (k-means pooling), spatial pooling strategy in Fig. 2 is more effective than k-means pooling strategy. The k-means graph pooling is based on the distance between different node features, which may destroy the dynamic information of EEG signals and induce the low classification results. Specifically, to validate the effectiveness of each branch in our model, we provide the results of V-IAG (Only Instance), i.e., IAG, V-IAG (Only Variational) and V-IAG. The combination of instance-adaptive branch and variational branch, i.e., V-IAG, achieves better result. Compared with instance-adaptive branch, variational branch performs better in subject-independent experiment. All these three models have the sparse constraint term.

*Comparisons with Different Methods on Subject-dependent Experiments*: As shown in TABLE II, for subject-dependent experiment on SEED, we provide the average accuracies and standard deviations in each of the five frequency bands and all frequency bands using state-of-the-art methods including support vector machine (SVM), deep belief network (DBN), graph convolutional neural network (GCNN), dynamical graph convolutional neural network (DGCNN), the improved bi-hemisphere domain adversarial neural network (BiDANN-S) and Graph convolutional broad network (GCB-net+BLS). Besides, we also provide the result of IAG and the proposed V-IAG for comparison. The proposed V-IAG and IAG achieve the best performance among these methods. Compared with traditional methods, like DBN and SVM, V-IAG performs much better. Compared with graph-based methods, like GCNN, DGCNN and GCB-net+BLS, the proposed V-IAG also achieves better result, which benefits from the instance-adaptive graph and the variational graph. The instance adaptive graph is helpful to exploit the individual dependencies among EEG channels and the variational graph can capture the underlying uncertain information. The experiment results using five frequency bands features are better than that using single-band features. We can see that the results in high frequency bands, i.e., $\beta$ and $\gamma$, are better than that in low frequency bands, i.e., $\delta, \theta$ and $\alpha$, which is consistent with previous works [17][68]. In particular, as shown in TABLE III, we provide the result of IAG and V-IAG using the energy features from two frequency bands in subject-dependent protocol on SEED. We provide three types of combinations among high frequency bands, i.e., $\alpha$ band and $\beta$ band, $\alpha$ band and $\gamma$ band, $\beta$ band and $\gamma$ band. Compared with IAG, V-IAG achieves better performance since V-IAG captures the uncertainties. Compared with the performance of V-IAG using energy features from two frequency bands, V-IAG using energy features from all five frequency bands achieves better classification accuracies. Although V-IAG with low frequency band energy features, i.e., $\delta$ and $\theta$, has lower classification accuracies, there are still some complementary information in $\delta$ and $\theta$, which is helpful for EEG emotion recognition.

*Comparisons with Different Features on Subject-dependent*

*Experiments*: To analyze the influence of different EEG features, we conduct the subject-dependent experiment with different EEG features on SEED in Table IV. We employ the different features, i.e., DE, PSD, DASM and RASM, provided by SEED. Different features will directly affect the performance on EEG emotion recognition. DE feature achieves better performance than other features. Our IAG and V-IAG perform well with higher classification accuracies. In particular, the results of V-IAG are better than IAG since V-IAG considers the uncertain information. All these experiments validate the effectiveness of the proposed IAG and V-IAG.



(a) (b) (c)

Fig. 3. Different region partitions for graph coarsening. (a) 17 regions; (b) 7 regions; (c) 4 regions.

TABLE V
THE MEAN ACCURACIES (ACC) AND STANDARD DEVIATIONS (STD) OF SUBJECT-DEPENDENT EEG EMOTION RECOGNITION EXPERIMENT WITH DIFFERENT GRAPH COARSENING METHODS.

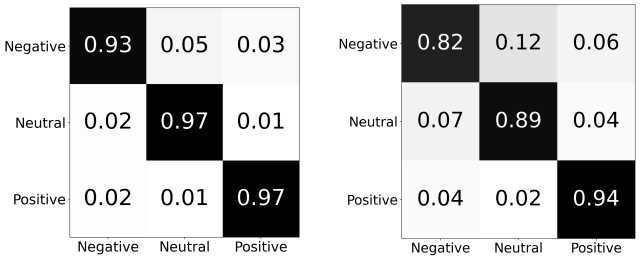
| Method | 17 regions (ACC/STD) | 7 regions (ACC/STD) | 4 regions (ACC/STD) |
|---|---|---|---|
| IAG [21] | 95.44 / 05.48 | 94.31 / 06.52 | 94.23 / 04.88 |
| V-IAG | **95.64 / 05.08** | **95.16 / 05.80** | **94.82 / 05.70** |

*Comparisons with Graph Coarsening Methods on Subject-dependent Experiments*: Some neuroscience researches [13][14] show that the functional connection between local regions is important to distinguish different human emotion patterns. We employ a graph coarsening method by aggregating the local EEG features so as to obtain more robust features. As shown in Fig. 3, we provide three types of divisions, i.e., 17 regions, 7 regions and 4 regions to estimate whether there are significant differences to the emotion recognition performance with respect to different scales of coarsening regions. In TABLE V, we present the subject-dependent result experiment with different types of graph coarsening methods. The division in Fig. 3 (a) achieves the best performance. Overall, the performance with small number of regions will have lower classification accuracy. Compared with IAG, V-IAG has better performance with any type of division, which also validates the advantages of V-IAG.

*Comparisons with Different Methods on Subject-independent Experiments*: In TABLE VI, we present the subject-independent results on SEED using different state-of-the-art methods, i.e., SVM, kernel principal component analysis (KPCA), transfer component analysis (TCA), transductive parameter transfer (TPT), DGCNN, domain-adversarial neural networks (DANN), BiDANN, BiDANN-S and R2G-STNN. For the subject-independent experiments, we emphasize the comparison with some transfer learning methods. Specifically, the proposed V-IAG achieves the remarkable result with the average accuracy $88.38\%$ and the standard deviation $4.8\%$, which is much better than other methods. Compared with subject-dependent experiment, the recognition result with our V-IAG model improves a lot in subject-independent experiment because the variational graph takes into consideration of the uncertain information and the probabilistic model is helpful to mitigate the noise.

TABLE VI
THE MEAN ACCURACIES (ACC) AND STANDARD DEVIATIONS (STD) OF SUBJECT-INDEPENDENT EEG EMOTION RECOGNITION EXPERIMENT ON SEED DATASET.

| Method | ACC / STD (%) |
|---|---|
| SVM [70] | 56.73 / 16.29 |
| KPCA [72] | 61.28 / 14.62 |
| TCA [73] | 63.64 / 14.88 |
| TPT [67] | 76.31 / 15.89 |
| DANN [74] | 75.08 / 11.18 |
| DGCNN [17] | 79.95 / 09.02 |
| BiDANN [75] | 83.28 / 09.60 |
| BiDANN-S [68] | 84.14 / 06.87 |
| R2G-STNN [71] | 84.16 / 07.63 |
| IAG [21] | 86.30 / 06.91 |
| V-IAG | **88.38 / 04.80** |




(a) (b)

Fig. 4. The confusion matrices of EEG emotion recognition results using the proposed V-IAG on the SEED dataset. (a) Confusion matrix in the subject-dependent protocol; (b) Confusion matrix in subject-independent protocol.

In Fig. 4., we present the confusion matrices of EEG emotion recognition result using the proposed V-IAG model on SEED dataset. For both subject-dependent experiment and subject-independent experiment, we can see that positive emotions are easier to be distinguished than negative emotions. This conclusion is consistent with previous works [68]. Due to the individual differences, subject-dependent experiment achieves higher average classification accuracies than subject-independent experiment.

### D. Experiment on MPED

On MPED dataset, we follow the same experiment protocols in [10]. There are three types of protocols. 21 trials of EEG data are employed as training data and the remaining ones

are employed as the testing data. For protocol one, eight types of combinations in form of positive-neutral-negative are presented and the average accuracy of these combinations are evaluated for comparison. The training data with three emotional categories in protocol one is balanced. To evaluate the EEG emotion recognition with imbalanced training data, protocol two is introduced. For protocol two, seven emotions are divided into three types of emotions, i.e., positive, neutral and negative emotions, which is imbalanced. To investigate the performance on EEG emotion recognition using multiple emotional categories, i.e., seven emotional categories, protocol three is provided. For protocol three, seven types of emotions are presented for comparison. STFT energy features are presented for EEG emotion recognition.

TABLE VII
EXPERIMENT RESULTS(%) IN PROTOCOL ONE AND PROTOCOL THREE (AVERAGE ACCURACIES AND STANDARD DEVIATIONS), AS WELL AS IN PROTOCOL TWO (AVERAGE ACCURACIES AND F1 SCORES) ON MPED DATABASE.

| Method | Protocol one (ACC/STD) | Protocol two (ACC/F1) | Protocol three (ACC/STD) |
|---|---|---|---|
| SVM [70] | 59.86 / 16.29 | 57.06 / 24.43 | 31.14 / 08.06 |
| DBN [23] | 65.83 / 13.20 | 65.98 / 59.19 | 29.26 / 09.19 |
| LSTM [76] | 72.09 / 14.94 | 71.92 / 65.12 | 38.55 / 08.43 |
| STRNN [41] | 65.38 / 13.20 | 66.84 / 60.57 | 35.64 / 09.57 |
| DGCNN [17] | 71.13 / 15.77 | 68.02 / 61.11 | 36.92 / 12.78 |
| A-LSTM [10] | 72.93 / 13.19 | 71.57 / 67.74 | 38.74 / 07.75 |
| IAG [21] | **74.77** / **10.75** | 73.58 / 68.41 | 40.38 / 08.75 |
| V-IAG | 74.72 / 10.91 | **74.13** / **67.83** | **40.40** / **09.35** |

As shown in TABLE VII, we provide the comparison with the state-of-the-art methods includes SVM, DBN, long short-term memory network (LSTM), spatial-temporal recurrent neural network (STRNN), DGCNN and attention LSTM (A-LSTM) on MPED dataset. With protocol two and protocol three, the proposed V-IAG achieves better performance than other methods. Since all these protocols are subject-dependent protocols, the improvement is marginal. With protocol one, IAG achieves the best result, which demonstrates the effectiveness of the instance-adaptive graph. The experiment in protocol one is based on balanced data and only three types of emotions and the contribution of variational graph is small. The experiment in protocol two is based on unbalanced data and the experiment in protocol three involves more emotion categories, which indicates that the variational graph is helpful to solve the unbalanced data and the multi-emotion recognition.

To further analyze the specific performance in terms of different emotion categories, we provide the confusion matrices on MPED using the proposed V-IAG model in Fig. 5. In Fig. 5 (a) under protocol one, we can see that the positive emotion has better performance than negative emotion, which is consistent with the conclusion in SEED. However, as shown in Fig. 5 (b) under protocol two, negative emotion achieves higher average classification accuracy than positive emotion and neutral emotion since the data in protocol two is unbalanced and the negative emotion contains more data. Neutral emotion has the smallest amount of data such that neutral emotion achieves the lowest average classification accuracy. In protocol three, we distinguish seven different emotions and the confusion matrix is shown in in Fig. 5 (c). With the balanced data, generally, positive emotions achieve better performance than negative emotions. Joy and funny are easier to be confused, which is consistent with [10]. Overall, the data distribution is a key factor to affect the performance of EEG emotion recognition.

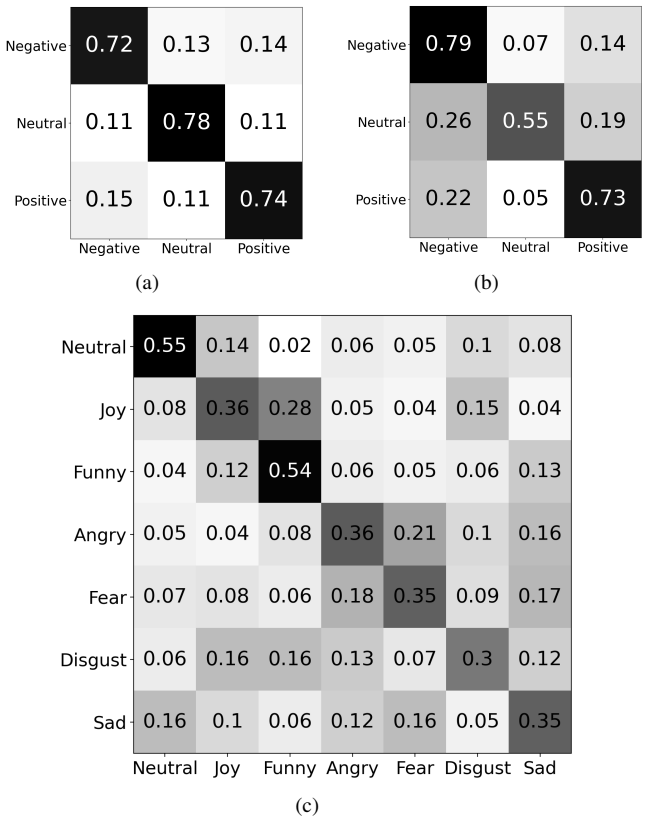



Fig. 5. The confusion matrices of EEG emotion recognition using the proposed V-IAG on MPED dataset. (a) Confusion matrix in protocol one; (b) Confusion matrix in protocol two; (c) Confusion matrix in protocol three.

## *E. Experiment on DREAMER*

TABLE VIII
THE MEAN ACCURACIES (ACC) OF SUBJECT-DEPENDENT EEG EMOTION RECOGNITION EXPERIMENT WITH PSD EEG FEATURES ON DREAMER.

| Method | Valence (ACC) | Arousal (ACC) |
|---|---|---|
| SVM [70] | 60.14 | 68.84 |
| GraphSLDA [77] | 57.70 | 68.12 |
| GSCCA [37] | 70.30 | 77.31 |
| DGCNN [17] | 86.23 | 84.54 |
| A-LSTM [10] | 77.00 | 83.99 |
| STRNN [41] | 68.89 | 79.65 |
| IAG [21] | 90.75 | 91.03 |
| V-IAG | **92.82** | **93.09** |

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3064940, IEEE Transactions on Affective Computing



To further estimate the effectiveness of the proposed method, we provide the experiment on DREAMER, which contains the annotation of valence and arousal. On DREAMER dataset, we follow the subject-dependent leave-one-trial-out cross-validation experiment protocol. We use SVM [70], GraphSLDA [77], GSCCA [37], DGCNN [17], A-LSTM [10] and STRNN [41] as the comparison methods. PSD features are provided for evaluation. As shown in TABLE VIII, IAG and V-IAG achieve better performance than other methods. In particular, V-IAG achieves higher classification accuracies in both valence and arousal. The results are shown in Table VI. Our method works well on valence and arousal annotations, which validates the effectiveness of our method.

TABLE IX
THE ACCURACIES (ACC) OF CROSS-DATASET EEG EMOTION RECOGNITION EXPERIMENT WITH SEED AND MPED.

| Method | SEED → MPED (ACC) | MPED → SEED (ACC) |
|---|---|---|
| SVM [70] | 33.42 | 32.53 |
| TKL [78] | 33.42 | 37.10 |
| TCA [73] | 40.18 | 49.38 |
| IAG [21] | 40.90 | 60.89 |
| V-IAG | 42.35 | 62.12 |

### F. Cross-datasets Experiments

To further evaluate our method, we conduct cross-dataset experiments, i.e., one dataset for training and another for testing. As shown in Table IX, the results of the cross-dataset experiments using SEED dataset and MPED dataset are provided. SEED → MPED denotes that SEED is used for training and MPED is used for testing. For SEED dataset, we employ one session of every subject for evaluation. For MPED, we use joy, neutrality and sadness to represent positive, neutral and negative emotions. We use SVM [70] and some transfer learning methods, i.e., transfer kernel learning (TKL) [78] and TCA [73], as comparison methods. Compared with the experiment within one dataset, the classification accuracies of cross-dataset experiments are lower. SVM obtains the random results. In particular when SEED is used for training, the best performance of V-IAG is 42.35%. For different datasets, the elicitation materials and experiment environments are different, which will make EEG emotion recognition be a more challenging task. However, our IAG and V-IAG still achieve better result than other methods.

### G. The Influence of Hyperparameters

As shown in Fig. 6, to analyze the influence of different hyperparameter values ($\alpha_5$), we change the value of $\alpha_5$ and estimate the accuracies of IAG under subject-dependent protocol on SEED. If we employ large values for $\alpha_5$, i.e., $0.1$ and $0.01$, the accuracy will decrease. $\alpha_5 = 1e^{-5}$ tend to provide a better result. Overall, $\alpha$ with small values tend to achieve better and stable performance. For IAG, we set $\{\alpha_i\}_{i=1}^{5}$ to be $\{1e^{-4}, 1e^{-5}, 1e^{-5}, 1e^{-5}, 1e^{-5}\}$. For V-IAG, we directly use the same hyperparameter values as IAG.

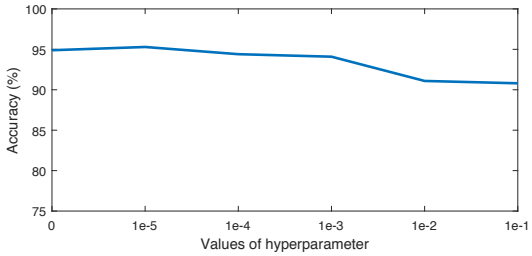



Fig. 6. The experiment results using IAG with different hyperparameters under subject-dependent protocol on SEED. We fix $\alpha_1$, $\alpha_2$, $\alpha_3$ and $\alpha_4$ and change the value of $\alpha_5$.

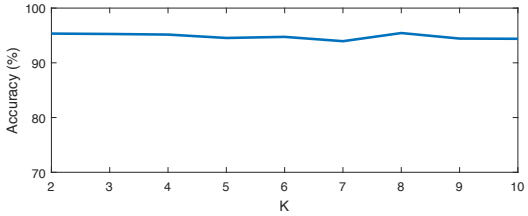



Fig. 7. The experiment results using IAG with different K values under subject-dependent protocol on SEED.

In Fig. 7, we show the results of IAG with different $K$ values for EEG emotion recognition under subject-dependent protocol on SEED. Compared with other parameters, the result using different $K$ values are more stable. We select the best one, i.e., $K = 8$. For V-IAG, we employed the same K value as IAG.

### H. T-SNE Visualization of V-IAG

To further exploit the training process of V-IAG, we employ the T-SNE technology to visualize the training process of

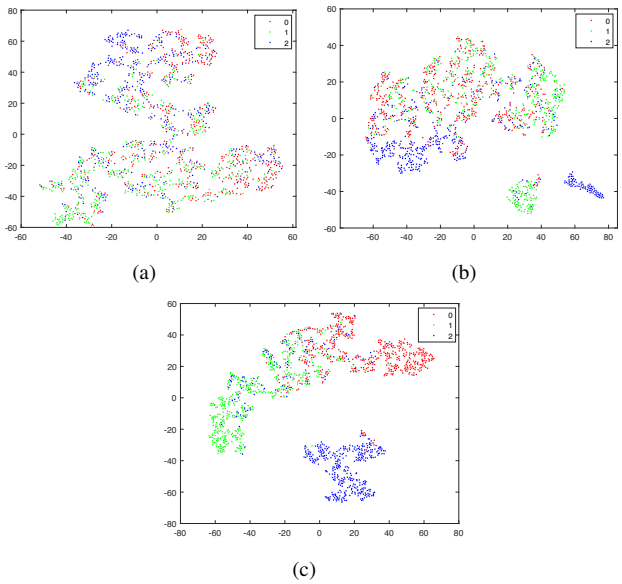
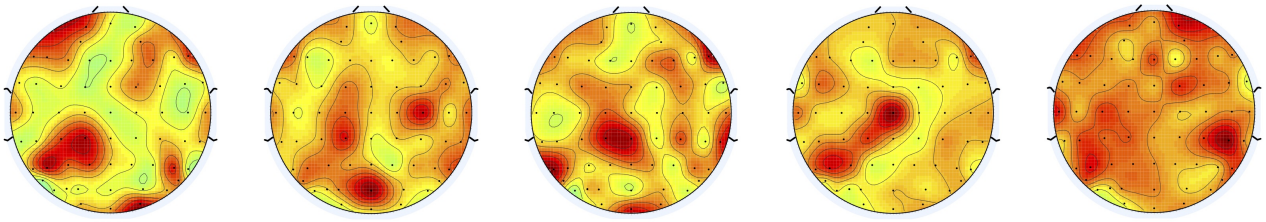



Fig. 8. The T-SNE visualization of features of V-IAG for subject-independent experiment. Red points (‘0’) denote the negative emotion, green points (‘1’) denote neutral emotion and blue points (‘2’) denote the positive emotion. (a) Initial state; (b) Middle state; (c) Trained state.



(a) (b) (c) (d) (e)

Fig. 9. The visualization of degree centrality distribution of $\mathbf{A}_V$ on scalp in our V-IAG model for neutral emotion on SEED dataset. (a) 1-4Hz; (b) 4-8Hz; (c) 8-14Hz; (d) 14-30Hz; (e) 30-50Hz.

the proposed model. As shown in Fig. 8, three types of emotions are mixed for initial state in Fig. 8(a). After training, three types of emotions can be distinguished well for the final state in Fig. 8(c). For subject-independent experiment, there are still some failed examples that cannot be recognized correctly. The individual difference will limit the performance of EEG emotion recognition. However, most of instances can be predicted correctly, which can validate the effectiveness of our model.

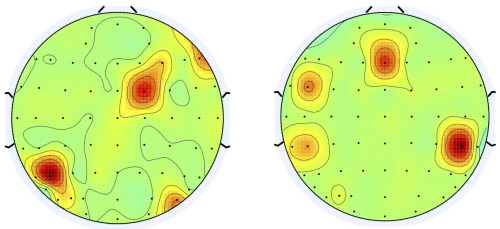
### *I. The Visualization of Degree Centrality in the Graph for EEG Emotion Recognition*

To further investigate the connections in variational graph, we try to visualize the graph connections. The connection in our variational graph is dense such that we visualize the degree centrality distribution on scalp of the variational graph. The degree centrality [79] evaluates the connectivity of a node with the other nodes, which has been widely used to measure the importance of the nodes in a graph. In our model, the adjacency matrix $\mathbf{A}$ represents the connections among different EEG channels and the values of the $i$-th row and the $j$-th column in $\mathbf{A}$ characterize the weights connected with the $i$-th node. Therefore, we can calculate the degree centrality of the $i$-th node, i.e., the $i$-th EEG channel, by

$$\mathcal{C}_i = \sum_{n=1}^{62} \mathbf{A}_{i,n} + \sum_{m=1}^{62} \mathbf{A}_{m,i} - 2\mathbf{A}_{i,i}, (i = 1, ..., 62). \quad (15)$$

As shown in Fig. 9, we visualize the degree centrality distribution of $\mathbf{A}_V$ based on V-IAG model for neutral emotion on SEED dataset. The degree centrality distribution on scalp in five frequency bands are visualized separately. We can see that the degree centrality distributions in different frequency bands are extremely different. Due to the multi-graph module, we can capture different dependencies among EEG channels in different frequency bands with different graphs, which can better capture more precise functional dependencies. In high frequency band, i.e., $\gamma$ band, we can see the connectivity is stronger than that in low frequency band. Especially in the prefrontal lobe, the degree centrality is high since the prefrontal lobe has high relations with emotion generation.

Further, in Fig. 10, we show the degree centrality distribution on scalp of two subjects in same frequency band. Although two subjects have the same emotional state, the learned graph connection patterns are extremely different. The instance-adaptive structure in our model captures the individual differences so as to achieve better performance on EEG emotion recognition.



(a) (b)

Fig. 10. The degree centrality distribution of $\mathbf{A}_V$ on scalp in our V-IAG model on SEED dataset for different subjects. (a) Subject1; (b) Subject2.

## V. Conclusion

In this paper, we proposed a variational instance-adaptive graph model that captures the individual differences and the uncertainties simultaneously. Further, the multi-level and multi-graph convolution provides a better graph representation and the graph coarsening is applied to mitigate the noise. Extensive experiments demonstrate the efficiency of the proposed method and the visualization of degree centrality has validated that our method can learn different graphs in different frequency bands for different subjects. Since the probabilistic model is helpful to quantify the uncertainties so as to mitigate the noise, it is a promising direction to quantify the uncertainties of the EEG signals and the uncertainties of model for EEG emotion recognition.

## 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] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie, "Facial expression recognition via learning deep sparse autoencoders," *Neurocomputing*, vol. 273, pp. 643–649, 2018.

[3] B. Ryu, A. R. Rivera, J. Kim, and O. Chae, "Local directional ternary pattern for facial expression recognition," *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 6006–6018, 2017.

[4] J. Yan, G. Lu, X. Bai, H. Li, N. Sun, and R. Liang, "A novel supervised bimodal emotion recognition approach based on facial expression and body gesture," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 101, no. 11, pp. 2003–2006, 2018.

[5] Z.-T. Liu, Q. Xie, M. Wu, W.-H. Cao, Y. Mei, and J.-W. Mao, "Speech emotion recognition based on an improved brain emotion learning model," *Neurocomputing*, vol. 309, pp. 145–156, 2018.

[6] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke, "Prosody-based automatic detection of annoyance and frustration in human-computer dialog." in *INTERSPEECH*, 2002.

[7] Z. Hu, Z. Yang, R. Salakhutdinov, and E. Xing, "Deep neural networks with massive learned knowledge," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 1670–1679.

[8] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "Ecg pattern analysis for emotion detection," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 102–115, 2012.

[9] B. Cheng and G. Liu, "Emotion recognition from surface emg signal using wavelet transform and neural network," in *Proceedings of the 2nd international conference on bioinformatics and biomedical engineering (ICBBE)*, 2008, pp. 1363–1366.

[10] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "Mped: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12 177–12 191, 2019.

[11] Y. M. Chi, Y.-T. Wang, Y. Wang, C. Maier, T.-P. Jung, and G. Cauwenberghs, "Dry and noncontact eeg sensors for mobile brain–computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 2, pp. 228–235, 2012.

[12] Y.-J. Huang, C.-Y. Wu, A. M.-K. Wong, and B.-S. Lin, "Novel active comb-shaped dry electrode for eeg measurement in hairy site," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 1, pp. 256–263, 2015.

[13] H. Kober, L. F. Barrett, J. Joseph, E. Bliss-Moreau, K. Lindquist, and T. D. Wager, "Functional grouping and cortical–subcortical interactions in emotion: a meta-analysis of neuroimaging studies," *Neuroimage*, vol. 42, no. 2, pp. 998–1031, 2008.

[14] M. J. Kim, R. A. Loucks, A. L. Palmer, A. C. Brown, K. M. Solomon, A. N. Marchante, and P. J. Whalen, "The structural and functional connectivity of the amygdala: from normal emotion to pathological anxiety," *Behavioural brain research*, vol. 223, no. 2, pp. 403–410, 2011.

[15] P. A. Kragel and K. S. LaBar, "Decoding the nature of emotion in the brain," *Trends in cognitive sciences*, vol. 20, no. 6, pp. 444–455, 2016.

[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 Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 026–12 035.

[17] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, 2018.

[18] Y.-Y. Lee and S. Hsieh, "Classifying different emotional states by means of eeg-based functional connectivity patterns," *PloS one*, vol. 9, no. 4, 2014.

[19] R. J. Davidson, H. Abercrombie, J. B. Nitschke, and K. Putnam, "Regional brain function, emotion and disorders of emotion," *Current opinion in neurobiology*, vol. 9, no. 2, pp. 228–234, 1999.

[20] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," in *Advances in neural information processing systems*, 2015, pp. 577–585.

[21] T. Song, S. Liu, W. Zheng, Y. Zong, and Z. Cui, "Instance-adaptive graph for eeg emotion recognition." in *AAAI*, 2020, pp. 2701–2708.

[22] T. C. Ho, C. G. Connolly, E. H. Blom, K. Z. LeWinn, I. A. Strigo, M. P. Paulus, G. Frank, J. E. Max, J. Wu, M. Chan *et al.*, "Emotion-dependent functional connectivity of the default mode network in adolescent depression," *Biological psychiatry*, vol. 78, no. 9, pp. 635–646, 2015.

[23] 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.

[24] 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.

[25] B. Hjorth, "Eeg analysis based on time domain properties," *Electroencephalography and clinical neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.

[26] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from eeg using higher order crossings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186–197, 2010.

[27] M. Li and B.-L. Lu, "Emotion classification based on gamma-band eeg," in *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 1223–1226.

[28] R. J. Davidson, "What does the prefrontal cortex "do" in affect: perspectives on frontal eeg asymmetry research," *Biological psychology*, vol. 67, no. 1, pp. 219–234, 2004.

[29] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "Eeg-based classification of music appraisal responses using time-frequency analysis and familiarity ratings," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 161–172, 2013.

[30] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis, "Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 589–597, 2010.

[31] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "Eeg-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

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

[33] 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*, 2020.

[34] T. Zhang, X. Wang, X. Xu, and C. P. Chen, "Gcb-net: Graph convolutional broad network and its application in emotion recognition," *IEEE Transactions on Affective Computing*, 2019.

[35] Y. Yang, Q. Wu, Y. Fu, and X. Chen, "Continuous convolutional neural network with 3d input for eeg-based emotion recognition," in *International Conference on Neural Information Processing*. Springer, 2018, pp. 433–443.

[36] F. P. George, I. M. Shaikat, P. S. Ferdawoos, M. Z. Parvez, and J. Uddin, "Recognition of emotional states using eeg signals based on time-frequency analysis and svm classifier." *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 9, no. 2, 2019.

[37] 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, 2017.

[38] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on eeg using lstm recurrent neural network," *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.

[39] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "Sae+ lstm: A new framework for emotion recognition from multi-channel eeg," *Frontiers in Neurorobotics*, vol. 13, p. 37, 2019.

[40] J. Ma, H. Tang, W.-L. Zheng, and B.-L. Lu, "Emotion recognition using multimodal residual lstm network," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 176–183.

[41] 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.

[42] Y. Li, W. Zheng, L. Wang, Y. Zong, L. Qi, Z. Cui, T. Zhang, and T. Song, "A novel bi-hemispheric discrepancy model for eeg emotion recognition," *arXiv preprint arXiv:1906.01704*, 2019.

[43] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring eeg features in cross-subject emotion recognition," *Frontiers in neuroscience*, vol. 12, p. 162, 2018.

[44] Z. Yin, Y. Wang, L. Liu, W. Zhang, and J. Zhang, "Cross-subject eeg feature selection for emotion recognition using transfer recursive feature elimination," *Frontiers in neurorobotics*, vol. 11, p. 19, 2017.

[45] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in Neural Information Processing Systems*, 2016, pp. 3844–3852.

[46] A. Nicolicioiu, I. Duta, and M. Leordeanu, "Recurrent space-time graph neural networks," in *Advances in Neural Information Processing Systems*, 2019, pp. 12 818–12 830.

[47] Z. Cui, Y. Cai, W. Zheng, C. Xu, and J. Yang, "Spectral filter tracking," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2479–2489, 2018.

[48] M. Zhang, S. Jiang, Z. Cui, R. Garnett, and Y. Chen, "D-vae: A variational autoencoder for directed acyclic graphs," in *Advances in Neural Information Processing Systems*, 2019, pp. 1588–1600.

[49] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[50] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems*, 2017, pp. 1024–1034.

[51] M. J. Kusner, B. Paige, and J. M. Hernández-Lobato, "Grammar variational autoencoder," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 1945–1954.

[52] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik, "Automatic chemical design using a data-driven continuous representation of molecules," *ACS central science*, vol. 4, no. 2, pp. 268–276, 2018.

[53] W. Jin, R. Barzilay, and T. Jaakkola, "Junction tree variational autoencoder for molecular graph generation," in *International Conference on Machine Learning*, 2018, pp. 2328–2337.

[54] S. Liu, W. Zheng, T. Song, and Y. Zong, "Sparse graphic attention lstm for eeg emotion recognition," in *International Conference on Neural Information Processing*. Springer, 2019, pp. 690–697.

[55] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, "An introduction to mcmc for machine learning," *Machine learning*, vol. 50, no. 1-2, pp. 5–43, 2003.

[56] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[57] S. Liu, Y. Huang, J. Hu, and W. Deng, "Learning local responses of facial landmarks with conditional variational auto-encoder for face alignment," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. IEEE, 2017, pp. 947–952.

[58] C. Zhao, B. Ni, J. Zhang, Q. Zhao, W. Zhang, and Q. Tian, "Variational convolutional neural network pruning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2780–2789.

[59] X. Li, Z. Zhao, D. Song, Y. Zhang, C. Niu, J. Zhang, J. Huo, and J. Li, "Variational autoencoder based latent factor decoding of multichannel eeg for emotion recognition," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2019, pp. 684–687.

[60] T. Zhang, Z. Cui, C. Xu, W. Zheng, and J. Yang, "Variational pathway reasoning for eeg emotion recognition." in *AAAI*, 2020, pp. 2709–2716.

[61] 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.

[62] 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.

[63] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.

[64] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion aware facial expression recognition using cnn with attention mechanism," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439–2450, 2018.

[65] C. Yan, Y. Tu, X. Wang, Y. Zhang, X. Hao, Y. Zhang, and Q. Dai, "Stat: spatial-temporal attention mechanism for video captioning," *IEEE transactions on multimedia*, vol. 22, no. 1, pp. 229–241, 2019.

[66] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[67] E. Sangineto, G. Zen, E. Ricci, and N. Sebe, "We are not all equal: Personalizing models for facial expression analysis with transductive parameter transfer," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 357–366.

[68] 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.

[69] 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)*. IEEE, 2013, pp. 81–84.

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

[71] 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.

[72] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[73] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2010.

[74] 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.

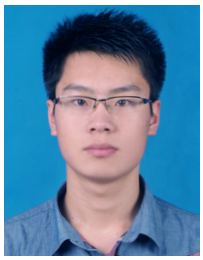
[75] Y. Li, W. Zheng, Z. Cui, T. Zhang, and Y. Zong, "A novel neural network model based on cerebral hemispheric asymmetry for eeg emotion recognition." in *IJCAI*, 2018, pp. 1561–1567.

[76] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," 2014.

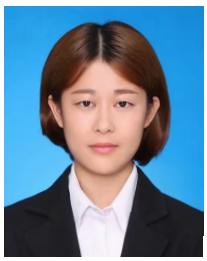
[77] Y. Li, W. Zheng, Z. Cui, and X. Zhou, "A novel graph regularized sparse linear discriminant analysis model for eeg emotion recognition," in *International Conference on Neural Information Processing*, 2016, pp. 175–182.

[78] M. Long, J. Wang, J. Sun, and S. Y. Philip, "Domain invariant transfer kernel learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 6, pp. 1519–1532, 2014.

[79] 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.

**Tengfei Song** received the B.S. degree in communication engineering from Hohai University, Jiangsu, China, in 2016. He is currently working towards his Ph.D. degree in the Department of information and communication engineering of Southeast University, China. His research interests include computer vision, machine learning and pattern recognition.



**Suyuan Liu** received the B.S. degree in automation engineering from Hohai University, Jiangsu, China, in 2017. She is currently working towards her M.S. degree in the Department of Biological Sciences and Medical Engineering of Southeast University, China. Her research interests include affective computing, machine learning and pattern recognition.

**Wenming Zheng** (SM'18) received the B.S. degree in computer science from Fuzhou University, Fuzhou, China, in 1997, the M.S. degree in computer science from Huaqiao University, Quanzhou, China, in 2001, and the Ph.D. degree in signal processing from Southeast University, Nanjing, China, in 2004. Since 2004, he has been with the Research Center for Learning Science, Southeast University, where he is currently a Professor with the School of Biological Science and Medical Engineering and the Key Laboratory of Child Development and Learning Science of the Ministry of Education. His current research interests include affective computing, pattern recognition, machine learning, and computer vision. Dr. Zheng served as an Associate Editor of several peer-reviewed journals, such as the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, Neurocomputing, and Visual Computer.

**Yuan Zong** received the BS and MS degrees in electronics engineering from Nanjing Normal University, Nanjing, China, in 2011 and 2014, respectively, and the PhD degree in Biomedical Engineering from Southeast University, Nanjing, China, in 2018. He is currently a Lecturer with the Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Sciences and Medical Engineering, Southeast University. From 2016 to 2017, he was a Visiting Student with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. His research interests include affective computing, pattern recognition, and computer vision.

**Zhen Cui** received the Ph.D. degree in computer science from Institute of Computing Technology (ICT), Chinese Academy of Science (CAS), Beijing, in Jun. 2014. He was a Research Fellow in the Department of Electrical and Computer Engineering at National University of Singapore (NUS) from Sep. 2014 to Nov. 2015. He also spent half a year as a Research Assistant on Nanyang Technological University (NTU) from Jun. 2012 to Dec. 2012. Now he is a professor of Nanjing University of Science and Technology, China. His research interests cover computer vision, pattern recognition and machine learning, especially focusing on deep learning, manifold learning, sparse coding, face detection/alignment/recognition, object tracking, image super resolution, emotion analysis, etc.

**Yang Li** received the B.S. degree in electronic information and science technology from the School of Physics and Electronics, Shandong Normal University, Jinan, China, in 2012, the M.S. degree in electronic and communication engineering from the School of Electronic Engineering, Xidian University, Xi'an, China, in 2015, and the Ph.D. degree from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2020. He was also a visiting student with the University of Wollongong, Wollongong, NSW, Australia, from August 2018 to August 2019. He is currently an Associate Professor with the School of Artificial Intelligence, Xidian University, Xi'an, China. His research interests focus on affective computing, pattern recognition, and computer vision.

**Xiaoyan Zhou** received her BS degree in communication engineering from Hohai University, Changzhou, Jiangsu, China, in 1997. She received the MS and PhD degrees in signal and information processing from Southeast University, Nanjing, Jiangsu, China, in 2005 and 2011, respectively. Currently, she is an associate professor of electrical and information engineering at Nanjing University of Information and Science and Engineering. Her research interests include pattern recognition and machine learning.