Journal of Neural Engineering

PAPER

DAGAM: a domain adversarial graph attention model for subject-independent EEG-based emotion recognition

To cite this article: Tao Xu *et al* 2023 *J. Neural Eng.* **20** 016022

Journal of Neural Engineering

CrossMark

RECEIVED
22 August 2022

REVISED
27 November 2022

ACCEPTED FOR PUBLICATION
22 December 2022

PUBLISHED
20 January 2023

PAPER

DAGAM: a domain adversarial graph attention model for subject-independent EEG-based emotion recognition

Tao Xu[1], **Wang Dang**[1], **Jiabao Wang**[1] and **Yun Zhou**[2,*]

[1] Northwestern Polytechnical University, School of Software, Xi'an, People's Republic of China
[2] Shaanxi Normal University, Faculty of Education, Xi'an, People's Republic of China
[*] Author to whom any correspondence should be addressed.

E-mail: zhouyun@snnu.edu.cn

Keywords: EEG, emotion recognition, subject independent

Abstract

Objective. Due to individual differences in electroencephalogram (EEG) signals, the learning model built by the subject-dependent technique from one person's data would be inaccurate when applied to another person for emotion recognition. Thus, the subject-dependent approach for emotion recognition may result in poor generalization performance when compared to the subject-independent approach. However, existing studies have attempted but have not fully utilized EEG's topology, nor have they solved the problem caused by the difference in data distribution between the source and target domains. *Approach.* To eliminate individual differences in EEG signals, this paper proposes the domain adversarial graph attention model, a novel EEG-based emotion recognition model. The basic idea is to generate a graph using biological topology to model multichannel EEG signals. Graph theory can topologically describe and analyze EEG channel relationships and mutual dependencies. Then, unlike other graph convolutional networks, self-attention pooling is used to benefit from the extraction of salient EEG features from the graph, effectively improving performance. Finally, following graph pooling, the domain adversarial model based on the graph is used to identify and handle EEG variation across subjects, achieving good generalizability efficiently. *Main Results.* We conduct extensive evaluations on two benchmark data sets (SEED and SEED IV) and obtain cutting-edge results in subject-independent emotion recognition. Our model boosts the SEED accuracy to 92.59% (4.06% improvement) with the lowest standard deviation (STD) of 3.21% (2.46% decrements) and SEED IV accuracy to 80.74% (6.90% improvement) with the lowest STD of 4.14% (3.88% decrements), respectively. The computational complexity is drastically reduced in comparison to similar efforts (33 times lower). *Significance.* We have developed a model that significantly reduces the computation time while maintaining accuracy, making EEG-based emotion decoding more practical and generalizable.

1. Introduction

The study of emotion recognition within computer science is a continuing endeavor. The findings and products of this emerging focus are increasingly being applied to education, digital games, e-commerce, advertising, e-health and many other areas. Electroencephalogram (EEG) has been suggested as a promising tool to investigate human emotions since it can directly and precisely reflect cognitive and emotional states at relatively low costs. As a result, EEG-based emotion recognition has attracted considerable research attention and interest.

However, studies in which deep learning algorithms are applied to EEG-based subject-independent emotion recognition are unsatisfactory. First, multichannel EEG signals have a structure based on biological topography belonging to a non-Euclidean domain. Directly applying deep learning methods to EEG-based recognition does not work well since these methods are designed for computer vision and natural language processing tasks. Second, EEG signals vary significantly between individuals, leading to different distributions of the source and target domains. This makes it challenging to achieve good performance across subjects.

The human brain's structural and functional systems have features of biological topography. Graph theory can topologically describe and analyze relationships and mutual dependencies between channels of EEG. The graph neural networks (GNNs) [1] make it promising to solve the classification problems on EEG data. Based on the graph, many researchers have made a great effort to solve these problems [2, 3]. These graph-based methods try to learn and extract the most salient features from the whole high-dimensional graph feature space generated by EEG data. Although some existing research has started to recognize the critical role of EEG channel topology, this structure is not fully utilized to effectively learn salient EEG features.

Most current recognition methods do not perform well when EEG training and testing data are from different individuals. For EEG-based emotion recognition, the data distributions of the source and target domains are different. This issue can be considered domain adaptation. Ganin *et al* [4] proposed domain-adversarial neural network (DANN) training to solve the cross-subject classification problem. Inspired by the idea of DANN, many studies have attempted to solve the problem and have achieved sigificant success [3, 5, 6]. However, there is still considerable room to improve the performance.

To address the two issues mentioned above in subject-independent emotion recognition, we propose the domain adversarial graph attention model (DAGAM), a novel EEG-based emotion recognition model. First, we model EEG signals using a graph based on biological topology. Graph convolutional networks (GCNs) with self-attention pooling are then used to extract EEG features strongly related to emotions. Finally, the graph-based domain adversarial model is used to identify emotions across subjects after graph pooling. The following are the significant contributions:

- The basic idea of DAGAM is to generate a graph to model multichannel EEG signals using biological topology. The use of graph attention neural networks (GANNs) effectively explores the relationships among multiple EEG channels for emotion recognition. Unlike other GCNs, self-attention pooling is applied to benefit salient EEG feature extraction from the graph, which effectively improves the performance.
- The domain adversarial (DA) model based on the graph is employed to identify and handle EEG variations across subjects. Combining DA and GANN, the source domain and the target domain can adapt to each other.

After evaluating DAGAM on two public emotion EEG data sets, SEED [7] and SEED IV [8], we found that our model achieves the state-of-the-art (SOTA) results in subject-independent emotion recognition, reaching a superior accuracy of performance with the lowest standard deviation (STD) (SEED: 92.59%/3.21%, SEED IV: 80.74%/4.14%) compared to other methods. Compared to other studies [3] in the same field, the computational complexity is drastically decreased (33 times lower). This model adds to the subject-independent method of emotion recognition using EEG, which is a significant improvement over the current approaches.

2. Related work

EEG-based emotion recognition has received increased attention in recent years. The methods can be categorized into two groups. One group focuses on finding crucial features. Shi *et al* [9] proposed a novel feature called differential entropy (DE) for EEG-based vigilance estimation. Jenke *et al* [10] reviewed a wide range of features to attempt to find suitable features for relevant emotions. Wang *et al* [11] compared three existing EEG features: power spectrum, wavelet and nonlinear dynamical analysis for improving emotion recognition. The other group is committed to proposing better classification algorithms. Petrantonakis *et al* proposed a robust emotion recognition method based on higher-order crossing analysis. Zhang *et al* [12] proposed a heuristic variational pathway reasoning method to deal with EEG-based emotion recognition. Xu *et al* [13] proposed a dynamic adaptive convolutional quorum voting approach for variable-length EEG data.

EEG-based emotion recognition typically consists of two approaches: subject-dependent and subject-independent. When people are exposed to the same emotional stimulus, individual differences may cause different physiological responses and patterns. Thus, the classifier derived from a single subject was unable to perform well in the multiple-subject context, especially when physiological responses to the same emotion varied significantly between subjects. Subject-dependent refers to the way that each classification model is built by only a single person. This approach has never been widely accepted due to its specificity. Recently, the subject-independent approach has drawn more attention as a means of achieving universality. This cross-subject approach refers to the way that the classification model is derived from mixed subjects [14]. For example, Li *et al* [15] proposed a multisource transfer learning method for cross-subject EEG emotion recognition. Li *et al* [16] proposed a bi-hemisphere domain adversarial neural network (BiDANN) model, which achieves good performance on cross-subject recognition. However, compared to the performance of subject-dependent emotion recognition, there is considerable room for improvement in the subject-independent approaches.

After the first GNN was proposed in 2009 [1], different GNNs were applied to different fields. EEG data are considered to belong to non-Euclidean domains, which can be represented as a graph. Graph models contain rich relational information [17] and can reflect the connections between different regions of the brain. At present, many researchers attempt to apply it to the domain of EEG-based emotion recognition. Song *et al* [2] proposed a novel dynamical graph convolutional neural network. Zhong *et al* [3] proposed a regularized graph neural network (RGNN) for EEG-based emotion recognition, which extracts both local and global features among different EEG channels based on the biological topology among different brain regions. These methods only benefit some known biological features but do not try to use unknown crucial internal connections and features via learning. A direction for future research is how to use the graph models and attention mechanisms to find the salient features of EEG signals related to emotions.

DANN [4] was the first work and was demonstrated to be successful in solving two specific classification problems in 2016. After that, DANN gained increased attention and was used in emotion recognition. Li *et al* [18] applied deep adaptation network (DAN) to eliminate the individual differences in EEG signals. Luo *et al* [6] proposed a novel Wasserstein generative adversarial network domain adaptation framework for building cross-subject EEG-based emotion recognition models. Bao *et al* [5] proposed a two-level domain adaptation neural network (TDANN) to construct a transfer model for EEG-based emotion recognition. Zhao *et al* [19] proposed a plug-and-play domain adaptation method for dealing with individual differences. These studies point out the research direction for subject-independent emotion recognition.

3. DAGAM

The structure of the DAGAM is shown in figure 1. It contains three main parts: EEG data modeling based on the graph, GANNs, and domain adversarial based on the graph. First, the EEG data are modeled based on EEG channel dependencies. Next, a graph attention neural network is proposed to extract the core features and discard the unimportant channels in the graph structure. Finally, domain adversarial based on the graph helps us to handle cross-subject EEG variations, enabling good performance in subject-independent emotion recognition. The details of each part are provided as follows.

3.1. EEG modeling based on the graph

EEG data are collected using an EEG cap based on the 10–20 system. Each electrode position represents the surface division of the brain, which interconnects and influences each other. Traditional convolution neural networks cannot directly benefit from this biological topology. GNNs provide an opportunity to tackle this issue.

The first step is to model EEG signals using the graph. A basic graph can be expressed as a set of vertices and edges, denoted as $G = (V, E)$, where V is the set of vertices and E is the set of edges. In our study, the vertices set can be expressed as the matrix $X \in \mathbb{R}^{N\times D}$, and the edge set can be expressed as the adjacency matrix $A \in \mathbb{R}^{N\times D}$, where N represents the number of EEG channels and D represents EEG data over time.

A is used to reflect the biological topography of EEG and indicates the relationship between EEG channels. The 3D coordinates of each electrode are obtained based on a 3D spatial modeling of the 10–20 system based on calculating the Euclidean distance. To correctly reflect this kind of relationship, we attempt to define the elements of the adjacency matrix based on the method proposed by [3], as follows:

$$A_{ij} = \min\left(1, \frac{\sigma}{d_{ij}^2}\right), \qquad (1)$$

where d_{ij} denotes the physical distance between channels i and j, and σ is a constant, used to calibrate the weight A_{ij}, which can fall within $(0, 1)$. A suitable value of σ is set according to the experiment in subsection 4.3.1, as shown in table 4.

Nine pairs of global connections (FP1–FP2, AF3–AF4, F5–F6, FC5–FC6, C5–C6, CP5–CP6, P5–P6, PO5–PO6, O1–O2) are added to the adjacency matrix to improve network efficiency. Previous studies have reported that global channels reflect the asymmetry of neuronal activity between the left and right hemispheres, which is essential for EEG-based emotion recognition. For instance, Schmidt and Trainor [20] found that the pattern of asymmetrical frontal EEG activity helped to distinguish the valence of emotion. Work [21] proposed an EEG feature named DE for emotion recognition, which was extracted from symmetrical electrodes. Furthermore, based on the asymmetric properties in emotion processing, Zheng and Lu [22] found that four selected channels (FT7, FT8, T7 and T8) can achieve comparably high and stable accuracy in the emotion recognition task. We initialize the global inter-channel relationship in A to $[-1, 0]$ in order to use this information.

3.2. GANNs

Graph structure helps us to model EEG data. However, long-period EEG data with a complicated graph structure bring a high computational cost for emotion recognition. The pooling method considers features both in the channels and the whole graph structure and removes the influence of unimportant nodes. It attempts to use a reasonable number of parameters to obtain better graph classification performance. Inspired by the work of [23], we adopted a graph

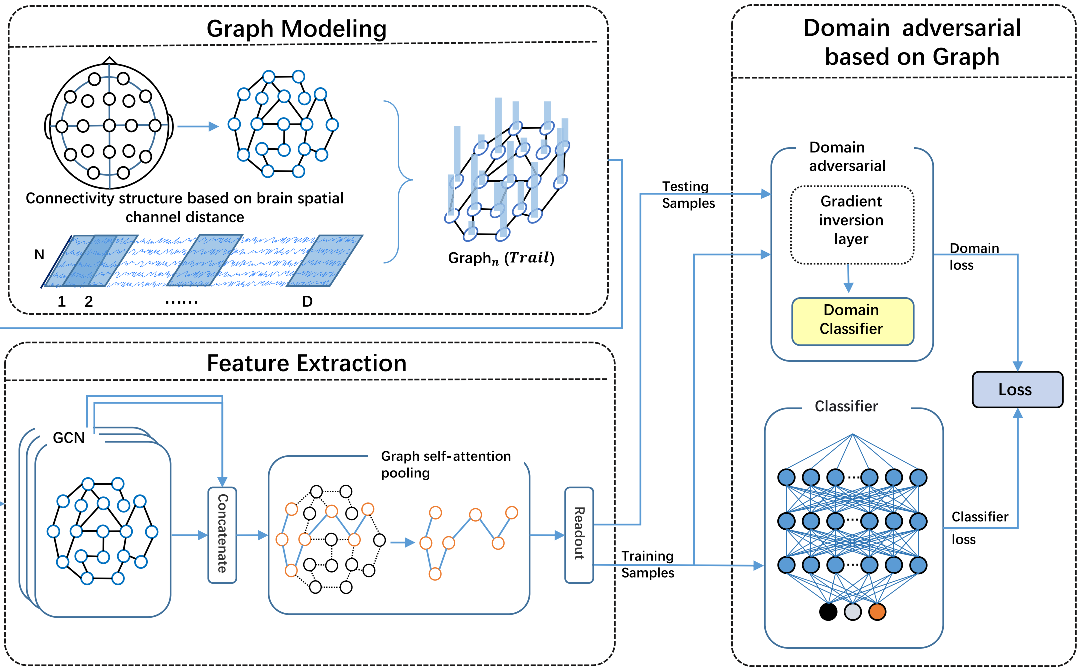


Figure 1. Structure of the DAGAM, consisting of three parts: graph modeling, feature extraction and domain adversarial based on the graph.

pooling method based on self-attention, called SAG-Pool, to extract crucial features from EEG data. The detailed steps are shown as follows:

First, the EEG data are processed by three layers of GNN to obtain the self-attention score. A widely used GNN model, GCN [24], which is implemented here, is formulated as follows:

$$\text{Score} = \sigma(\tilde{L}_{\text{sym}} X W_{\text{att}}), \tag{2}$$

where σ is the activation function of the layer network, and W_{att} is a weighted matrix used to perform the affine transformation on the input graph signal. $\tilde{L}_{\text{sym}}$ is the re-normalized Laplacian matrix following [24].

The index and $\text{Score}_{\text{mask}}$ of self-attention graph pooling can be obtained as follows:

$$\begin{aligned} \text{index} &= \text{top} - \text{rank}(\text{Score}, [kN]) \\ \text{Score}_{\text{mask}} &= \text{Score}_{\text{index}}, \end{aligned} \tag{3}$$

where $k \in (0, 1]$ is the proportion of nodes retained and top $-$ rank is the function that returns the indices of the top kN values based on the self-attention score. N is the total number of nodes. Score is used to select the nodes with the top proportion k and index is an indexing operation on Score to update the mask: $\text{Score}_{\text{mask}}$.

Next, we perform a pooling operation on feature data using GCN. The new feature matrix and the corresponding adjacency matrix are calculated as follows:

$$\begin{aligned} X' &= X_{\text{index}} \\ X_{\text{out}} &= X' \odot \text{Score}_{\text{mask}}, \\ A_{\text{out}} &= A_{\text{index},\text{index}} \end{aligned} \tag{4}$$

where $\odot$ represents the broadcasted element-wise product, X_{out} is the new feature matrix and A_{out} refers to the corresponding adjacency matrix.

Finally, a readout layer is provided to change features to a fixed size before graph classification. The representation results are concatenated by global average pooling and global max pooling, which is shown as follows:

$$s = \frac{1}{N}\sum_{i=1}^{N} x_i \| \max_{i=1}^{N} x_i, \tag{5}$$

where N is the number of nodes, X_i represents the feature value of the ith node and $\|$ is a concatenation operator.

3.3. Domain adversarial model based on the graph

Due to individual differences in emotion, the generalization of the emotion recognition model usually does not perform well. To improve the generalization performance of our model across subjects, we propose a method of domain adversarial based on the graph.

The main advantage of our method is to reduce the computational complexity significantly. Different to the work in [3], we apply the domain adversarial model to the graph after self-attention pooling instead of nodes since this graph after pooling contains the crucial features obtained by extraction from GCN. A detailed comparison is provided in section 4.6.

The core of domain adversarial based on the graph is the domain classifier. The input data of the domain adversarial based on the graph are from the readout layer, including the source and target domains. In the training process, the mixed samples from the source domain with labels and the target domain without labels are put into the domain adversarial model. To achieve effective domain migration, the domain classifier in this model is not expected to successfully distinguish between the source domain data and the target domain data. During training, the domain classifier discriminates between the source and the target domains, and non-crucial features of the source domain and target are removed.

For the loss function that needs to be optimized, we select X^{S}, $X^{\mathrm{T}} \subseteq \mathrm{GCN}_{\mathrm{feature}} \in \mathbb{R}^{G\times d}$ from $\mathrm{GCN}_{\mathrm{ferture}}$, where X^{S} is the source domain data, X^{T} is the target domain data, G is the number of graphs and d is the dimension of the graph data. The data label belonging to the source domain is set to 0 and the data label belonging to the target domain is set to 1. It is converted to a one-hot form as $Y_i^{\mathrm{S}} = [1,0]$, $Y_i^{\mathrm{T}} = [0,1]$ and then the cross entropy (CE) is employed to construct the following:

$$
\begin{aligned}
E_{\mathrm{D}} &= H\left(Y^{\mathrm{S}}, q^{\mathrm{S}}\right) + H\left(Y^{\mathrm{T}}, q^{\mathrm{T}}\right)\\
&= -\left(\sum_{i=1}^{G^{\mathrm{T}}}\sum_{j=1}^{C} Y_i^{\mathrm{S}}(x_j)\log(q_i(x_j))\right.\\
&\qquad \left. + \sum_{i=1}^{G^{\mathrm{S}}}\sum_{j=1}^{C} Y_i^{\mathrm{T}}(x_j)\log\left(q_i(x_j)\right)\right).
\end{aligned}
\tag{6}
$$

The classifier in the domain adversarial model based on the graph is a three-layer fully connected neural network for emotion recognition, which attempts to find the correct emotion based on features from graph self-attention pooling.

3.4. Defense against label noise

The label noise introduced by stimuli experiments is a common issue in emotion recognition. Emotion recognition labels are typically assigned based on the type of stimuli material. This assignment may be harmful if subjects exhibit unexpected physiological reactions. For example, if the video material is labeled as happy, so will the corresponding EEG data. However, the person watching this video may be feeling something other than happiness or a mixture of emotions. For example, the subject might not be completely happy, but he or she will be most content with a small number of neutral feelings. This noise has a negative impact on emotion recognition performance.

To address this issue, we start with the fundamental premise that an induced experiment will not result in opposite feelings, such as the belief that pleasant stimuli will not result in depressing ones. We alter each label's probability distribution according to [3]. Based on a prior probability distribution, the label Y is mapped into a new label $\hat{Y}$. For instance, SEED, an emotion EEG data set, has three categories: negative, neutral and positive. The following shows how label Y is translated into $\hat{Y}$:

$$
\hat{Y}_i = \begin{cases} (1-\epsilon, \epsilon, 0) & Y_i = 0\\ (\epsilon, 1-2\epsilon, \epsilon) & Y_i = 1\\ (0, \epsilon, 1-\epsilon) & Y_i = 2 \end{cases}, \tag{7}
$$

where i is the class indices, and 0, 1 and 2 represents negative, neutral and positive, respectively, and $\epsilon \in [0,1]$ represents a hyper-parameter controlling the noise level in the training labels.

Kullback–Leibler (KL) divergence is adopted as the loss function of the emotion recognition classifier since it can measure how one probability distribution is different from another.

$$
D_{\mathrm{KL}}(\hat{Y}\|q) = \sum_{i=1}^{N}[\hat{Y}(x_i)\log\hat{Y}(x_i) - \hat{Y}(x_i)\log q(x_i)], \tag{8}
$$

where $\hat{Y}$ represents the real data, a measured probability distribution. Distribution $q(x_i)$ represents instead a theoretical distribution of the data.

The training process is carried out to minimize E_{all}, which is the sum of the emotion recognition loss (L_y^i) and the domain classification loss (L_d^i).

$$
\begin{aligned}
L_y^i(\theta_f, \theta_y) &= D_{\mathrm{KL}}(\hat{Y}\|q)\\
L_d^i(\theta_f, \theta_d) &= E_{\mathrm{D}} = H(Y^{\mathrm{S}}, q^{\mathrm{S}}) + H(Y^{\mathrm{T}}, q^{\mathrm{T}}).\\
E_{\mathrm{all}} &= L_y^i(\theta_f, \theta_y) + L_d^i(\theta_f, \theta_d)
\end{aligned}
\tag{9}
$$

4. Experiments and evaluation

To evaluate our model, we apply DAGAM to two public emotion EEG-based data sets: the SJTU Emotion EEG Dataset (SEED) [7], and an evolution of the original SEED data set (SEED IV) [8].

4.1. Implementation details

In experiments in the two data sets, we set the hyperparameters of DAGAM as follows: the number of GCN layers L is 3; the pooling ratio k used in self-attention pooling is set to 0.5. The classifier for emotion recognition based on the graph is a three-layer fully connected neural network. Adam is used as the model's gradient descent optimizer with a value of 0.001. We implemented the whole model with PyTorch. The model runs on the server with Intel Core i9-9900K CPU @ 3.60 GHz, 32 GB memory, 512 GB SSD, and NVIDIA GeForce RTX 3090 running Linux Ubuntu 18.04.03 LTS.

4.2. Data set instruction

These data sets collect EEG signals from the same device: an ESI NeuroScan with 62 channel electrodes according to the international 10–20 system at a sampling rate of 1000 Hz. The raw EEG signals from these data sets are preprocessed, and different salient features are extracted based on previous studies [9]. The detailed information is provided as follows.

4.2.1. SEED and SEED IV

In the SEED, 15 film clips were chosen to evoke three emotions: positive, neutral and negative. Fifteen subjects participated in the experiment. There were 15 trials for each subject in the experiment. In SEED IV, 72 film clips were chosen to evoke four kinds of emotions: happiness, sadness, fear, or neutrality. Fifteen subjects were also recruited to participate in this experiment. Three sessions, including 24 trials, were performed on different days for each subject. The raw EEG data were downsampled at 200 Hz to facilitate recognition. Then, a bandpass filter with 1–75 Hz was applied to remove the noise and artifacts. In our experiments, a time-frequency domain feature called DE [9] was extracted.

$$h(X) = -\int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \ell^{-\frac{(x-\mu)^2}{2\sigma^2}} \times \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \ell^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx = \frac{1}{2}\log\left(2\pi\ell\sigma^2\right), \tag{10}$$

where the time series X follows the Gauss distribution $N(\mu, \sigma^2)$.

4.3. Performance analysis

We compare DAGAM with other baseline methods to comprehensively evaluate our model, including the SOTA in mean accuracy and STD for SEED and SEED IV, respectively. The confusion matrix analysis is provided. Then, the results of the ablation study are presented to pinpoint the crucial elements. The section ends with a complex computation comparison.

4.3.1. Subject-independent emotion recognition

We conduct experiments on two data sets (SEED and SEED IV) using leave-one-out cross-validation (LOOCV) to evaluate the performance of DAGAM on subject-independent emotion recognition. The experimental settings are followed by [3, 25], which tests our DAGAM on one subject and trains on the remaining subjects for each fold. LOOCV evaluates each subject in the data sets. The mean accuracy (ACC) and STD are compared.

The performance of our DAGAM is shown in table 1, which lists the comparison between the DAGAM model and other methods in the subject-independent in SEED and SEED IV. The comparison includes 19 methods as follows: KLIEP [26], ULSIF [26], STM [27], SVM [28], TCA [29], SA [30], GFK [31], A-LSTM [32], T-SVM [33], EEG-GCN [34], DANN [4], DAN [18], BiDANN-S [16], BiHDM [35], RGNN [3], WGAN-DA [6], TDANN [5], DOGNN [36].

Table 1. Comparison of the performance accuracy on subject-independent emotion recognition in SEED and SEED IV. The top results are highlighted in bold. ACC/STD(%).

Method	SEED	SEED IV
KLIEP	45.17/17.76	31.46/9.20
ULSIF	51.18/13.57	32.99/11.05
STM	51.23/14.82	39.39/12.4
SVM	56.73/16.29	37.99/12.52
TCA	63.64/14.88	56.56/13.77
SA	69.00/10.89	64.44/9.46
GFK	71.31/14.09	64.38/11.41
A-LSTM	72.18/10/85	55.03/9.28
T-SVM	72.53/14	—
EEG-GCN	77.30/8.83	—
DANN	79.19/13.14	47.59//10.01
DGCNN	79.95/9.02	52.82/9.23
DAN	83.81/8.56	58.87/8.13
BiDANN-S	84.14/6.87	65.59/10.39
BiHDM	85.40/7.53	69.03/8.66
RGNN	85.3/6.72	73.84/8.02(SOTA)
WGAN-DA	87.07/7.14	—
TDANN	87.90/6.13	—
DOGNN	88.53/5.67(SOTA)	73.55/10.19
DAGAM	**92.59/3.21**	**80.74/4.14**

Obviously, our DAGAM performs better than the other 19 methods, including SOTA, on both SEED and SEED IV. The DAGAM achieves the highest accuracy with the lowest STD. This improves the accuracy of SOTA by 4.06% for SEED and 6.90% for SEED IV, respectively.

We directly quote emotion recognition results of other baselines from the work of [25]. Our model substantially improves the performance and is much better than others concerning accuracy, but with a relatively high STD. We have double-checked our results.

As shown in the experiments mentioned above, our DAGAM can further improve subject-independent emotion recognition compared with other methods. Among the methods compared with our model, there are two methods that use GNNs: DGCNN [2] and RGNN [3], and six methods that use domain adversarial training: TDANN [5], WGAN-DA [6], RGNN [3], DAN [18], BiDANN-S [16], DANN [4]. No one adopted the attention mechanism. Therefore, we assume that graph self-attention pooling effectively helps to extract crucial invariable features and remove irrelevant ones.

To further verify this assumption, we conducted further experiments. In each round of experiments on two data sets, we modified the core hyperparameters of GANNs: top proportion k. Table 2 shows the results. It can easily be found that the experimental results have undergone apparent changes, especially on

Table 2. Comparison of the performance accuracy at different top proportions k on SEED and SEED IV. The top results are highlighted in bold. ACC/STD(%).

k	SEED	SEED IV
0.9	87.70/5.25	80.64/3.63
0.8	88.59/4.35	80.18/4.68
0.7	87.56/3.22	80.09/4.55
0.6	88.15/3.60	80.18/4.87
0.5	**92.59/3.21**	**80.74/4.14**
0.4	89.57/4.62	79.53/4.71
0.3	88.57/4.02	80.62/6.22
0.2	87.70/3.88	79.72/4.41
0.1	88.15/4.19	80.55/4.13

Table 3. Comparison of the performance accuracy of the models with or without a global connection, Symbol '-' indicates the following component is removed, ACC/STD(%).

Method	SEED	SEED IV
DAGAM	92.59/3.21	80.74/4.14
-global connection	92.19/4.55	80.27/5.25

Table 4. Comparison of the performance accuracy in different σ on SEED and SEED IV. The top results are highlighted in bold. ACC/STD(%).

σ	SEED	SEED IV
1	84.00/3.16	73.05/5.26
2	85.77/3.95	77.31/6.26
3	89.48/4.17	78.51/5.45
4	89.77/3.87	78.98/4.75
5	**92.59/3.21**	**80.74/4.14**
6	87.25/3.93	78.51/4.69
7	88.59/4.33	79.72/5.48
8	89.63/3.86	78.61/5.24
9	88.89/3.71	79.72/4.85
10	86.96/5.18	79.25/5.83

SEED, by almost 5%. As a result, graph self-attention pooling does play a central role in our model.

In order to explore the influence of different parameters on the model as much as possible, we conducted a comparative experiment on the global connection and σ. Table 3 shows the results of the comparison of the model with global connection and without connection. The results indicate that the global connection can slightly improve the performance of the model.

The effect of different σ on the graph modeling is shown in table 4. The value is set as five performing better than others on both SEED and SEED IV, respectively.

4.4. Confusion matrix analysis

To provide deep insight into our model for different emotions, we provide the confusion matrix for SEED and SEED IV. As shown in figure 2, these confusion matrices are represented in percentage with rows normalized.

For SEED, as shown in figure 2(a), our model performs with a high level of accuracy for all emotions. While it performs much better on neutral emotions than others, it is not very sensitive to negative emotions. Nearly 8% of negative emotions are misrecognized as neutral and positive emotions.

For SEED IV, as shown in figure 2(b), our model performs at around 80% for all four categorized emotions. It is good at distinguishing happiness but weak in recognizing neutral emotions. 8.15% of neutral emotions are categorized as sadness by mistake, and 7.04% and 6.30% of them are recognized as fear and happiness, respectively.

Overall, the model shows a fairly high level of emotion recognition.

4.5. Ablation study

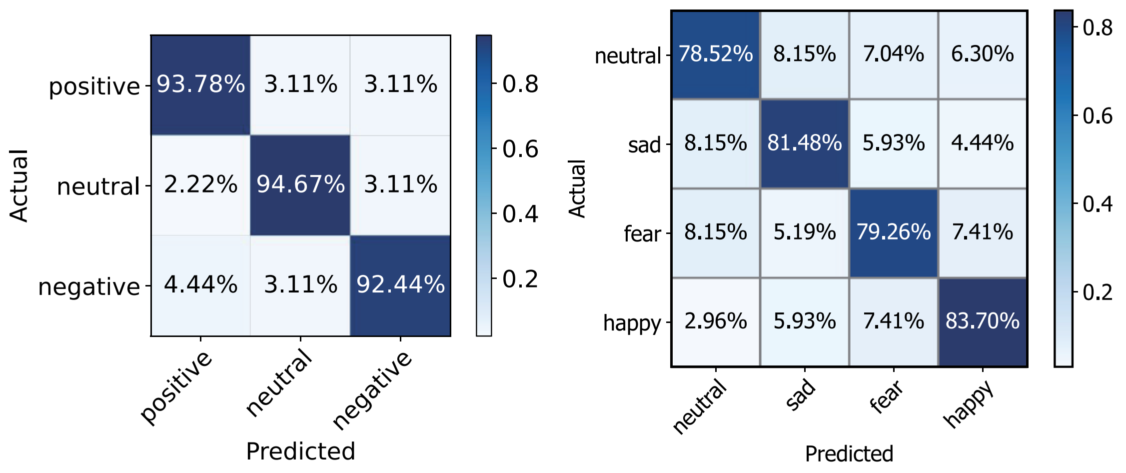
DAGAM mainly adopts three methods in different parts of the model to tackle individual differences. Graph self-attention pooling is adopted in the feature extraction phase to extract crucial features based on biological topology. In the phase of graph classification, KL divergence is adopted to handle inaccurate emotion labels, which can quantify differences between the probability distribution of the training set and the testing set. In the training phase, domain adversarial training based on the graph is an attempt to solve the problem of the same labels with different distributions, that is, domain adaptation.

To study the effects of the three core parts in the model, we conducted an ablation study (three further experiments) to verify them. In the first experiment, we disabled the domain adversarial part and only used other parts to recognize emotions to study the effects of domain adversarial training based on the graph. In the second experiment, we replaced the KL divergence with the another loss function: CE, to investigate the effect of KL divergence. In the third experiment, we attempted to discover the effects of graph self-attention pooling by disabling domain adversarial training and replacing KL divergence.

The results are shown in table 5. The KL divergence has a significant impact on the performance of the model, especially on SEED. Without the KL divergence, the accuracy of SEED drops by nearly 5.78%. The domain adversarial training based on the graph also affects the performance. Without domain adversarial training based on the graph, the accuracy decreases. We find that only domain adversarial training is applied, and its accuracy does not have a significant impact. If only graph self-attention pooling is used, it retains good accuracy on these two data sets. This result again verifies our previous assumption that graph self-attention pooling is a crucial part of our model.

4.6. Computational complexity comparison

As noted in section 3.3, our approach can significantly lower the computational complexity when compared to the work of [3]. It is dependent on two key factors. The first is to use self-attention pooling to pick out key features; it drastically reduces the parameters before



(a) (b)

Figure 2. Confusion matrices of the subject-independent EEG emotion recognition results using our DAGAM on SEED and SEED IV. (a) Confusion matrix of SEED; (b) confusion matrix of SEED IV.

Table 5. Ablation studies for subject-independent classification accuracy (mean/STD) on SEED, SEED IV, Symbol '-' indicates the following component is removed. ACC/STD(%).

Method	SEED	SEED IV
DAGAM	92.59/3.21	80.74/4.14
-Domain adversarial training	87.55/5.18	78.33/4.64
-KL divergence	86.81/3.48	77.50/5.17
-Domain adversarial training and KL divergence	88.00/3.70	79.53/4.71

Table 6. Comparison of the performance of training time(s).

Method	Time
DAGAM	0.0208
RGNN	0.0561

classifier input. The other is that the entire graph is applied to the domain adversarial training rather than just the node [3]. A computational complexity comparison is provided as follows. $O(G)$ represents the computational complexity of our work.

$$O(G) = O\Big(\text{GraphNum}\Big(\text{HiddenDim} \times 3 \times 2 \times \text{HiddenDim} + \text{HiddenDim} \times \frac{\text{HiddenDim}}{2} + \frac{\text{HiddenDim}}{2} \times 2\Big)\Big). \tag{11}$$

$O(N)$ of [3] is as follows:

$$O(N) = O\left(\text{NodeNum} \times \text{GraphNum}\left(\text{HiddenDim} \times 3 \times \text{HiddenDim} + \text{HiddenDim} \times \frac{\text{HiddenDim}}{2} + \frac{\text{HiddenDim}}{2} \times 2\right)\right). \tag{12}$$

We compared $O(G)$ with $O(N)$ as follows:

$$\frac{O(N)}{O(G)} = \frac{7 \times \text{NodeNum}}{13} + \frac{12}{217 \times \text{HiddenDim} + 14}. \tag{13}$$

From formula (13), it is evident that there is still a difference in the computational cost of $\frac{7\times \text{NodeNum}}{13}$ times even when the hidden layer dimension increases to infinity, causing the second portion to approach zero. When there are 62 nodes (due to the EEG cap with 62 electrodes used in SEED and SEED IV), the node-domain adversarial computational complexity is approximately 33 times greater than the graph-domain adversarial nodes (our method).

To evaluate the performance of DAGAM in practical applications, we record the average training time for each epoch of our algorithm and the comparison algorithm. Table 6 indicates that our model runs nearly three times faster than [3]. However, there is a difference in the theoretical value. According to the analysis, we found that the main reason is the difference in the implementation method. In work [3,] implementation is based on the framework of PYG, which provides better optimization of the implementation process, while we manually implement the core GCN part based on the Numpy. This indicates the direction for our future work.

4.7. Brain region analysis through visualization

We visualized connections in order to study the role of various brain regions in different emotions. Since our model adopts SAGpooling, it is difficult to directly visualize the model based on electrodes. We use LIME [37] to solve this problem. The results are shown in figure 3.

We find that the asymmetry in frontal and parietal EEG activity may reflect alterations in different

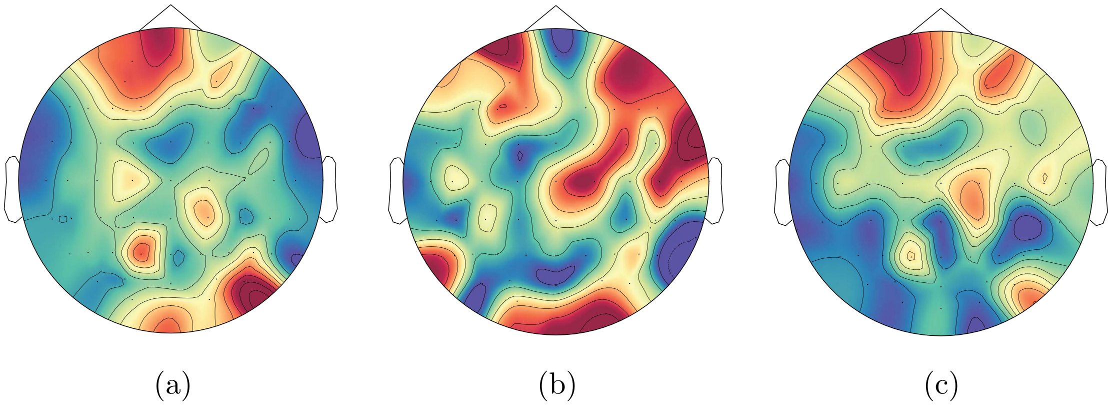


Figure 3. Heat map learned from DAGAM according to the LIME method: (a) positive; (b) neutral; (c) negative.

emotions. This supports the previous research [7, 20, 21] that the asymmetry in EEG activity between the left and right hemispheres is a crucial feature for emotion recognition.

5. Conclusion

This study contributes to the growing area of EEG-based subject-independent emotion recognition by proposing a DAGAM. DAGAM is powerful for learning the relationships between EEG channels based on graph pooling using self-attention pooling benefits to extract salient features for the emotion recognition task. Domain adversarial training based on the graph contributes significantly to tackling the cross-subject EEG variation issue. Extensive experiments on two public data sets (SEED and SEED IV) show that the performance of our model achieves SOTA, providing the highest accuracy and lowest STD than other competitive baselines with low computational complexity. In future work, we will continue to move along the line of graph models.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://bcmi.sjtu.edu.cn/home/seed/.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62077036) and the National Key Research and Development Program of China (Grant Nos. 2022YFC3303600 and 2018AAA0100501; We gratefully acknowledge the support of Ascend CANN (Compute Architecture for Neural Networks) and Ascend 910 AI Processor used for this research. Artificial intelligence computing power and technical support were provided by the Xi'an Future Artificial Intelligence Computing Center.

ORCID iDs

Tao Xu https://orcid.org/0000-0002-1721-561X
Wang Dang https://orcid.org/0000-0001-6795-9636
Jiabao Wang https://orcid.org/0000-0002-8889-3223
Yun Zhou https://orcid.org/0000-0002-2306-8986

References

[1] Scarselli Fand G M, Tsoi A, Hagenbuchner M and Monfardini G 2009 The graph neural network model *IEEE Trans. Neural Netw.* **20** 61–80
[2] Song T, Zheng W, Song P and Cui Z 2020 EEG emotion recognition using dynamical graph convolutional neural networks *IEEE Trans. Affective Comput.* **11** 532–41
[3] Zhong P, Wang D and Miao C 2022 EEG-based emotion recognition using regularized graph neural networks *IEEE Trans. Affective Comput.* **13** 1290–301
[4] Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M and Lempitsky V 2016 Domain-adversarial training of neural networks *J. Mach. Learn. Res.* **17** 2096–30
[5] Bao G, Zhuang N, Tong L, Yan B, Shu J, Wang L, Zeng Y and Shen Z 2021 Two-level domain adaptation neural network for EEG-based emotion recognition *Front. Hum. Neurosci.* **14** 620
[6] Luo Y, Zhang S, Zheng W and Lu B 2018 WGAN domain adaptation for EEG-based emotion recognition *Int. Conf. on Neural Information Proc.* (*Siem Reap, Cambodia, 13–16 December*) pp 275–86
[7] Zheng W and Lu B 2015 Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks *IEEE Trans. Auton. Ment. Dev.* **7** 162–75
[8] Zheng W, Liu W, Lu Y, Lu B and Cichocki A 2019 EmotionMeter: a multimodal framework for recognizing human emotions *IEEE Trans. Cybern.* **49** 1110–22
[9] Shi L, Jiao Y and Lu B 2013 Differential entropy feature for EEG-based vigilance estimation *2013 35th Annual Int. Conf. IEEE (EMBC)* pp 6627–30
[10] Jenke R, Peer A and Buss M 2014 Feature extraction and selection for emotion recognition from EEG *IEEE Trans. Affective Comput.* **5** 327–39

[11] Wang X, Nie D and Lu B 2014 Emotional state classification from EEG data using machine learning approach *Neurocomputing* **129** 94–106
[12] Zhang T, Cui Z, Xu C, Zheng W and Yang J 2020 Variational pathway reasoning for EEG emotion recognition *Proc. Conf. AAAI Artificial Intelligence* vol 34 pp 2709–16
[13] Xu T, Zhou Y, Hou Z, Zhang W and Yuan Y 2020 Decode brain system: a dynamic adaptive convolutional quorum voting approach for variable-length EEG data *Complexity* **2020** 1–9
[14] Chen J, Hu B, Wang Y, Moore P, Dai Y, Feng L and Ding Z 2017 Subject-independent emotion recognition based on physiological signals: a three-stage decision method *BMC Med. Inf. Decis. Making* **17** 167
[15] Li J, Qiu S, Shen Y, Liu C and He H 2020 Multisource transfer learning for cross-subject EEG emotion recognition *IEEE Trans. Cybern.* **50** 3281–93
[16] Li Y, Zheng W, Zong Y, Cui Z, Zhang T and Zhou X 2018 A Bi-hemisphere domain adversarial neural network model for EEG emotion recognition *IEEE Trans. Affective Comput.* **12** 494–504
[17] Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C and Sun M 2020 Graph neural networks: a review of methods and applications *AI Open* **1** 57–81
[18] Li H, Jin Y, Zheng W and Lu B 2018 Cross-subject emotion recognition using deep adaptation networks *ICONIP 18* (*Siem Reap, Cambodia, 13–16 December*) pp 403–13
[19] Zhao L, Yan X and Lu B 2021 Plug-and-play domain adaptation for cross-subject EEG-based emotion recognition *Proc. Conf. AAAI Artificial Intelligence* vol 35 pp 863–70
[20] Schmidt L A and Trainor L J 2001 Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions *Cogn. Emot.* **15** 487–500
[21] Duan R N, Zhu J Y and Lu B L 2013 Differential entropy feature for EEG-based emotion classification 2013 *6th Int. IEEE/EMBS Conf. on Neural Engineering (NER)* pp 81–84
[22] Zheng W-L and Lu B-L 2015 Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks *IEEE Trans. Auton. Ment. Dev.* **7** 162–75
[23] Lee J, Lee I and Kang J 2019 Self-attention graph pooling *Proc. of the 36th ICML* (*Proc. of Machine Learning Research*) vol 97 (PMLR) pp 3734–43
[24] Kipf N and Welling M 2017 Semi-supervised classification with graph convolutional networks *ICLR 2017 (Toulon, France, 2017)*
[25] Li Y, Zheng W, Zong Y, Cui Z, Zhang T and Zhou X 2021 A Bi-hemisphere domain adversarial neural network model for EEG emotion recognition *IEEE Trans. Affective Comput.* **12** 494–504
[26] Kanamori T, Hido S and Sugiyama M 2009 A least-squares approach to direct importance estimation *J. Mach. Learn. Res.* **10** 1391–445
[27] Chu W, De la Torre F and Cohn J F 2017 Selective transfer machine for personalized facial expression analysis *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 529–45
[28] Suykens J A K and Vandewalle J 1999 Least squares support vector machine classifiers *Neural Process. Lett.* **9** 293–300
[29] Pan J, Tsang W, Kwok T and Yang Q 2011 Domain adaptation via transfer component analysis *IEEE Trans. Neural Netw.* **22** 199–210
[30] Fernando B, Habrard A, Sebban M and Tuytelaars T 2013 Unsupervised visual domain adaptation using subspace alignment *2013 IEEE ICCV* pp 2960–7
[31] Gong B, Shi Y, Sha F and Grauman K 2012 Geodesic flow kernel for unsupervised domain adaptation *2012 IEEE CVPR* pp 2066–73
[32] Song T, Zheng W, Lu C, Zong Y, Zhang X and Cui Z 2019 MPED: a multi-modal physiological emotion database for discrete emotion recognition *IEEE Access* **7** 12177–91
[33] Collobert R, Sinz F, Weston J and Bottou L 2006 Large scale transductive SVMs *J. Mach. Learn. Res.* **7** 26
[34] Gao Y, Fu X, Ouyang T and Wang Y 2022 EEG-GCN: spatio-temporal and self-adaptive graph convolutional networks for single and multi-view EEG-based emotion recognition *IEEE Signal Process. Lett.* **29** 1574–8
[35] Li Y, Wang L, Zheng W, Zong Y, Qi L, Cui Z, Zhang T and Song T 2021 A novel Bi-hemispheric discrepancy model for EEG emotion recognition *IEEE Trans. Cogn. Dev. Syst.* **13** 354–67
[36] Li H, Zhang X and Xia Y 2022 EEG emotion recognition based on dynamically organized graph neural network Int. Conf. on Multimedia Modeling 2022 pp 344–55
[37] Ribeiro M T, Singh S and Guestrin C 2016 "Why Should I Trust You?": explaining the predictions of any classifier *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (KDD'16) (New York: Association for Computing Machinery) pp 1135–44