## Title
Channel-adaptive Graph Convolution based Temporal Encoder Network for EEG Emotion Recognition



**Journal**
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

**Authors**
Guo, Renxi
Rao, Hong
An, Panfeng
et al.

**Publication Date**
2024

Peer reviewed

# Channel-adaptive Graph Convolution based Temporal Encoder Network for EEG Emotion Recognition

**Renxi Guo [1], Hong Rao [1], Shengbo Chen[2], Gang Luo[3], Panfeng An[4], Wenying Duan[3,5]**
[1]School of Software, Nanchang University, China
[2]School of Computer and Infomation Engineering, Henan University, China
[3]School of Mathematics and Computer Sciences, Nanchang University China
[4]Ningbo Artificial Intelligence Insititute, Shanghai Jiaotong University, China
[5]Jiangxi Provincial Key Laboratory of Intelligent Systems and Human-Machine Interaction, Nanchang University, China

## Abstract

Brain-computer interface technology has made significant progress in the field of intelligent human-computer interaction. Among them, electroencephalography based emotion recognition, as one of the important research directions in emotional brain-computer interaction, has received widespread attention. However, most previous studies were limited to feature extraction of global brain networks and local brain areas in the EEG spatial domain but ignored the channel-level dynamic features of EEG. To address this limitation, we proposed a Channel-Adaptive Graph Convolutional Network with Temporal Encoder (CAG-TEN). In CAG-TEN, the channel-adaptive graph convolutional module assigns a unique parameter space to each channel, focusing on channel-level dynamic features. Additionally, the temporal encoder module, inspired by the Encoders concept, is used to explore long-term temporal dependencies in EEG sequences. We conduct rigorous comparative experiments of CAG-TEN against several representative baseline models on the SEED dataset and achieve optimal performance.

**Keywords:** Emotional brain-computer, Emotion Recognition, EEG, Channel Adaptative, Temporal Encoder

## Introduction

Emotions, as psychological states closely related to daily life, are the inner expressions of an individual's perception and feelings (Panksepp, 2005; Vanderlind, Millgram, Baskin-Sommers, Clark, & Joormann, 2020; Wang et al., 2022). In recent years, the rapid development of brain-computer interface (BCI) technology has propelled human-computer interaction into a new phase. This revolutionary technology no longer requires additional physical movements like traditional human-computer interactions (Fallman, 2003; Bos et al., 2010), successfully establishing a direct communication bridge between the brain and computer devices. However, accurately and efficiently recognizing human emotions remains a key challenge researchers face. Typically, the subjects of emotion recognition studies can be divided into physiological signals (such as electroencephalography (EEG) and electrocardiogram) and non-physiological signals (such as facial expressions and body language) (Li et al., 2022; Canal et al., 2022). Compared to non-physiological signals, physiological signals offer significant advantages such as spontaneity and authenticity. Therefore, emotion recognition based on physiological signals has become mainstream over the past decade. In this paper, we have chosen EEG as our research subject and conducted the following studies.

From machine learning algorithms to deep learning networks, researchers have come up with a myriad of modeling solutions for EEG-based emotion recognition(Jenke, Peer, & Buss, 2014; Tzirakis, Trigeorgis, Nicolaou, Schuller, & Zafeiriou, 2017). Bazgir et al.(Bazgir, Mohammadi, & Habibi, 2018) utilized Discrete Wavelet Transform (DWT) techniques to manually extract a limited amount of channel and frequency band information from raw EEG data, followed by the application of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) classifiers for emotion classification. Tao et al. (Tao et al., 2020) targeted the spatio-temporal domain of EEG and proposed a cascaded Channel Attention and Self-Attention Network (ACRNN) based on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). This work eliminated the process of manual feature extraction, achieving end-to-end emotion classification. With the deepening of EEG data research and advancements in artificial intelligence technology, earlier simple algorithms and networks are increasingly reaching performance bottlenecks (Yang, Han, & Min, 2019). This indicates the need for designing more complex and focused models to adapt to various EEG-based emotion recognition scenarios.

Graph Convolutional Networks (GCNs) are often used to capture global dependency correlations in node spaces due to their inherent advantage of a global perception field (Kipf & Welling, 2016; S. Zhang, Tong, Xu, & Maciejewski, 2019; Manessi, Rozza, & Manzo, 2020). Additionally, GCNs have been widely applied in data domains of non-Euclidean spaces such as knowledge graphs and social networks. As EEG data is typically non-Euclidean, GCNs are commonly employed to extract key features from the spatial domain of EEG data. Song et al. (T. Song, Zheng, Song, & Cui, 2018) introduced a Dynamic Graph Convolutional Neural Network (DGCNN) that dynamically learns global dependencies in the EEG spatial domain by randomly initializing a custom adjacency matrix. Ultimately, DGCNN achieved good performance in emotion classification. Gu et al.(Gu, Zhong, Qu, Liu, & Chen, 2023) designed a Domain-level Graph Generative Adversarial Network (DGGN), which uses a Graph Convolutional Network (GCN) to extract features in the spatial domain and synchronously input sample data into a Long

Corresponding author: Panfeng An, Wenying Duan, Shengbo Chen and Gang Luo.



In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* 

Short-Term Memory network (LSTM) to learn temporal domain features. The outputs are then adversarially trained with a discriminator. DGGN also demonstrated excellent performance in classification. These approaches can be summarized as using GCN in the EEG spatial domain to extract dynamic features, followed by the targeted design of additional modules to extract deeper features.

Our main contributions in this paper are summarised as follows:

- We propose a spatio-temporal model for EEG emotion recognition named CAG-TEN. CAGCN is designed to extract global dynamic features in the spatial domain of EEG data while ensuring attention to channel-level features, addressing the insufficiencies in spatial domain exploration that exist in previous work. Additionally, we concurrently design a TE network tailored for temporal sequence EEG data to extract temporal domain information.
- We introduce a data augmentation method called Random Label Recombination. This method not only expands the quantity of the training set but also enhances the model's robustness and generalization.
- We conduct comprehensive and rigorous comparative experiments between CAG-TEN and several representative baseline models, effectively demonstrating the advancement of CAG-TEN.

# Related work

## GCNs work in EEG emotion recognition

In the early applications of GCNs in EEG emotion recognition(T. Zhang, Wang, Xu, & Chen, 2019; Zhong, Wang, & Miao, 2020), a pre-defined adjacency matrix was commonly used to represent the spatial relationships between EEG channels. However, these pre-defined matrices are generally created through manual calculations, such as measuring the 3D spatial distances between electrode points or computing the similarity between channel features (Nie, Ren, Nie, & Zhao, 2020). This method often adheres too closely to human subjective judgment and fails to comprehensively cover the spatial relationships between channels. Additionally, the relationships among EEG channels are not static, yet pre-defined adjacency matrices lack this dynamic adaptability. Subsequent research (T. Song et al., 2018) introduced a dynamically learning GCN, which replaced the pre-defined adjacency matrix with a custom adjacency matrix that could be continuously updated during the training iterations. This improvement eliminated the subjective biases introduced by manual pre-definition and also enabled the model to adapt to variations in channel relationships across different sample data. However, the custom approach often relies heavily on the initialization of the adjacency matrix, which can lead to issues such as exploding or vanishing gradients during the training process. Moreover, previous works utilizing GCNs in the spatial domain of EEG data have primarily focused on exploring either global brain networks or specific local brain regions (clusters of channels located in particular scalp areas). However, these approaches often overlook the channel-level features inherent in the EEG spatial domain, which implies that some spatial information may be lost during the model learning process.

## Shortcomings in GCNs work

From the perspective of a single channel, the operation of a GCN can be seen as transforming the features of a single channel through a graph convolution computation that applies weights and then produces an output. However, this also implies that all channels share the same set of parameters. In this shared-parameter pattern, GCN is more focused on extracting a prominent global feature pattern from all channel features, which overlooks the unique feature pattern that may exist within each channel individually. For instance, in two different channels, one might reflect features related to emotions, while the other could indicate features related to brain disorders. In such cases, the custom adjacency matrix approach of GCNs may fail to fully capture discriminative features that closely represent real emotional states. Therefore, in this paper, we maintain a unique parameter space for each channel to conduct channel-specific feature extraction, allowing for more tailored and accurate modeling of each channel's characteristics.

# Datasets

To validate the effectiveness of our model, we used the SJTU Emotion EEG Dataset (SEED) (Zheng & Lu, 2015; R.-N. Duan, Zhu, & Lu, 2013) provided by the Brain-like Computing and Machine Intelligence Laboratory team as our research subject. The SEED dataset was collected from 15 subjects (7 males and 8 females), and in this paper, we only used the EEG data of each subject.

# Method

## Overview

In this section, we first outline the EEG emotion recognition framework of the CAG-TEN model, as illustrated in Figure 1. After preprocessing the EEG data, we elaborate in detail on the construction of the CAGCN and TE modules within the CAG-TEN model. Finally, we list several details of the CAG-TEN model during the training process.

## Channel adaptive graph convolution

Generally, nearly all graph convolutional methods can be summarized by the following nonlinear function (Wu et al., 2019):

$$\hat{X}^{i+1} = f\left(X^{i}, A\right). \tag{1}$$

In this case, a batch of data samples $X^0 \in \mathbf{R}^{B\times W\times C\times F}$ and an adjacency matrix $A \in \mathbf{R}^{C\times C}$ serve as the initial inputs for the convolution function $f$. The only difference between different graph convolutional methods lies in the construction of the function $f$ and the adjacency matrix $A$. Therefore, we will

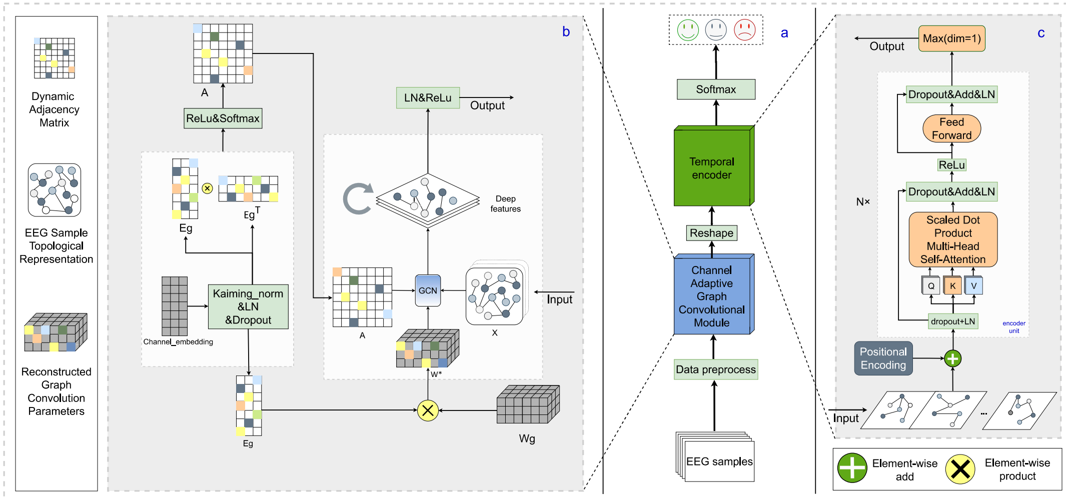


Figure 1: (a) The pipeline of the proposed CAG-TEN. (b) The CAGCN branch, which learns dynamic feature information in the spatial domain of DE features from both channel-level and global perspectives, then outputs to the next branch for further learning. (c) The TE branch, where the output from the CAGCN is fed into Encoders in a sequential format to learn long-term dependencies within the sequence samples. The subfigures on the left and bottom right provide explanations of some units within the CAG-TEN model.

introduce the Channel Adaptive Graph Convolution based on Equation 1.

Since GCN primarily focus on the interdependencies of channel features in the spatial domain, to provide a clearer explanation, we have selected representative GCN application models in the field of EEG emotion recognition, such as DGCNN, RGNN (T. Song et al., 2018; Zhong et al., 2020). In these models, a single layer of GCN applied to the spatial domain can be defined as:

$$\hat{X} = \sigma\left(\left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) X W + b\right). \tag{2}$$

In the equation, $I_N$ represents the identity matrix for the self-connections of all nodes. $D$ denotes the degree matrix. $W \in \mathbf{R}^{F \times O}$ is the trainable parameter matrix in graph convolution, and $b$ is the bias term. $\sigma(\cdot)$ is the activation function. In Equation 2, multiplying $A$ by $D^{-\frac{1}{2}}$ on both sides can prevent unfair weighting between nodes with high degrees and those with low degrees towards their neighboring nodes, thereby keeping the model's variance within a controllable range. However, this also introduces additional computational between matrices. In the following content, we will improve on this basis.

While such models may capture inter-channel correlations to some extent, their reliance on prior knowledge of graph structures could limit their capabilities, as many relationships in the brain are implicit and difficult to preset, and various types of spatial relationships coexist simultaneously (W. Duan, He, Zhou, Thiele, & Rao, 2023). Additionally, due to the complexity of information in the brain, significant differences may exist between spatial adjacent channels. Inspired by the concept of adaptive dependency matrices (Bai, Yao, Li, Wang, & Wang, 2020), We utilize channel-adaptive graph convolutions to learn spatial correlations as well as channel differences in the brain.

Therefore, in our approach, we prioritize assigning a unique parameter space to each channel (node). Initially, we transform the trainable parameter matrix W $\in \mathbf{R}^{F \times O}$ into $W^* \in \mathbf{R}^{C \times F \times O}$, where $C$ represents the number of channels. Thus, each channel is provided with its own dedicated parameter space. However, if $C$ is too large, the parameter $W^*$ also faces the issue of being too huge to optimize, which could lead to overfitting. Therefore, we apply matrix decomposition to $W^*$ to solve this issue (Bai et al., 2020).

$$W^* = E_g W_g \tag{3}$$

$E_g \in \mathbf{R}^{C \times d}$ is called the channel-embedding matrix, where $d$ is the embedding dimension, and $d \ll C$. $W_g \in \mathbf{R}^{d \times F \times O}$ is known as the shared parameter pool. This process can be interpreted as $E_g$ learning a specific set of feature patterns from a vast candidate set $W_g$, thereby gaining insights into the dynamic features of each channel. This is one of the core idea of our proposed channel-adaptive concept. The same operation can be applied to $b$ as well. Therefore, our design of assigning a parameter space to each channel, integrated with GCN, can be represented as follows:

$$\hat{X} = \left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) X E_g W_g + E_g b_g \tag{4}$$

Furthermore, the predefined adjacency matrix $A$ often leads to spatial graphs that do not comprehensively cover the interdependencies between nodes due to its overly intuitive nature, and without appropriate prior knowledge, the predefined $A$ generally lacks universality for different graph tasks. To solve this issue, we designed $E_A \in \mathbf{R}^{C\times d_a}$, an embedding matrix, where $d_a$ represents the embedding dimension. This is another core idea of our proposed channel-adaptive concept. We infer the implicit interdependencies between every pair of channels in the data by calculating the similarity between $E_A$ and $E_A^T$. Additionally, instead of directly generating $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, we have applied the $ReLU$ function and $softmax$ function sequentially to normalize $E_A * E_A^T$, directly producing $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. This design effectively saves the redundant and unnecessary computational overhead mentioned earlier during iterative training. Therefore, the symmetric normalization of the adjacency matrix $A$ can be represented as:

$$D^{-\frac{1}{2}}AD^{-\frac{1}{2}} = softmax\left(ReLU\left(E_A * E_A^T\right)\right) \tag{5}$$

Finally, we have refined the concept of channel adaptivity by setting $E_g = E_A$. By making improvements to the classic GCN. we have obtained the complete representation of the CAGCN module:

$$\hat{X} = \left(I_N + softmax\left(ReLU\left(E_g E_g^T\right)\right)\right) X E_g W_g + E_g b_g \tag{6}$$

## Temporal encoder

After obtaining the output $\hat{X} \in \mathbf{R}^{B\times W\times C\times O}$ from the CAGCN module, we synchronously designed the TE module to learn the temporal domain information of EEG data. Prior to this, we reshape $\hat{X}$ into $\bar{X} \in \mathbf{R}^{B\times W\times I}$, where $I = N \times O$, for the TE module.

Unlike the spatial domain of EEG data, the sequential nature and historical dependencies in the temporal domain are particularly prominent. During the collection of each subject's emotional EEG, the temporal information exhibits clear characteristics such as duration and intervals. Moreover, the current state's temporal information of a subject is often influenced by previous states, showing significant sequential dependencies. Therefore, learning from the temporal domain information can be simply viewed as a sequence modeling problem. Inspired by (Vaswani et al., 2017), based on Encoders, we designed a Temporal Encoder specifically tailored for EEG data to learn the temporal information of EEG.

First, the EEG data in the form of a sequence will be fused with position coding information in the sample by a position encoder in advance. This operation allows TE to learn the pre-arranged position information between sequences on a basis for subsequent training. This is also one of the reasons for the essential difference between TE and traditional sequential networks because TE can have a global perspective of all sequences and focus on and mine the truly important information in the sequence in parallel like the human attention mechanism. Subsequently, the EEG data will enter Encoders composed of multiple encoder units connected in series for attention learning. Each encoder is mainly composed of a multi-head self-attention (MHA) and feed-forward network. The process of MHA can be expressed as:

$$MHA\left(Q,K,V\right) = Concat\left(head_1,...,head_h\right)W^O \tag{7}$$

In the MHA process, $Q$, $K$, and $V$ serve as the inputs, representing the query, key, and value vectors of the EEG data, respectively. $W^O$ is the parameter matrix for the linear transformation in MHA. Concat$(\cdot,...,\cdot)$ is the concatenation function used for combining the outputs of multiple heads, and the $i$-th head, represented as $head_i$, is defined as:

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{8}$$

Please note that $W_i^Q$, $W_i^K$, and $W_i^V$ are the parameter matrices used for the linear transformations of the vectors $Q$, $K$, and $V$ respectively. Subsequently, the transformed $Q$, $K$, and $V$ serve as inputs for the scaled dot-product attention function, denoted as Attention(). The computation process for scaled dot-product attention can be represented as:

$$Attention\left(Q,K,V\right) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{9}$$

In this case, $d_k$ is a hyperparameter. $\sqrt{d_k}$ acts as a scaling factor, ensuring that the self-attention scores are maintained within a suitable range. This scaling helps improve the stability and learning effectiveness of the model.

Finally, to enhance the model's nonlinear fitting capability, we introduce more flexibility and complexity by adding a layer of feed-forward network after each MHA output sequence. After passing through $N$ encoder units, we obtain the final output from the Temporal Encoder (TE).

# Experiments

In this section, we first describe the experiment setup and then introduce a data augmentation method specifically designed for EEG data, named Random Label Recombination. To thoroughly evaluate our proposed model, we organized experiments from both within-domain and cross-domain perspectives, which include subject-dependent and subject-independent experiments. We then conducted a comprehensive analysis of the results from both types of experiments. In terms of comparative studies, we selected four representative models that have emerged in recent years in the field of EEG emotion recognition as baseline models. Additionally, we employed two widely used evaluation metrics for our analysis: Test Accuracy (ACC) and Standard Deviation (STD). Finally, we conducted an ablation study on the CAG-TEN and provided visualizations of CAG-TEN's training process as experimental support for this section.

## Experiment setup

During the training process of the model, we adopt the cross-entropy loss function to measure the error between the true labels and the predicted results and use the Adam optimizer to update the model parameters during the training iterations to minimize the loss function. In addition, in a single experiment, we test the model after each epoch of training and record the test accuracy of the current epoch of training. In the list of test accuracy rates for a single experiment, we will use the item with the highest test accuracy rate as the final test result of the single experiment. Detailed statistical results can be found in our experimental protocol

## Data augmentation

In subject-dependent experiments, the limited EEG data collected from subjects often leads to overfitting. Moreover, considering the robustness and generalizability of the model, the sufficiency of the data volume is also a critical issue that can limit model performance. Inspired by the work in (Lotte, 2015), we designed the Random Label Recombination method to mitigate this problem. Through this method, we successfully doubled the original data volume. Before training the model, we extracted a batch of DE features from the training set, assuming a size of 72. Then, based on the label categories (three categories), we attempted to obtain an equal batch size of 72 augmented DE features from the training set (24 per category). For example, if the current label category is positive, we first locate all positive samples in the training set and divide the feature dimension (the last dimension) of these positive samples into $F_s$ equal parts (assuming $F_s = 5$). Keeping the original order of feature dimensions, we randomly select an equal amount of features in a sequential index order for recombination. Importantly, the augmented samples produced by this process do not repeat any samples already in the training set. Finally, we concatenated the initially extracted DE features with the enhanced DE features and used them as the training data input to the model.

## Subject-dependent experiment

Subject-dependent experiments focus solely on the EEG data of the current subject, meaning both the training and testing sets are derived from the same subject. This experimental approach aims to assess performance specific to the subject, providing a personalized evaluation that better adapts to the subject's physiological characteristics and emotional expression patterns. To objectively evaluate the CAG-TEN model, we employed a five-fold cross-validation method. Firstly, we divided the subject's 15 emotional labels equally into five parts, each containing all three emotional categories (positive, neutral, negative). We then rotated through these five parts, using four as the training set and the remaining one as the testing set, repeating this process five times. For a fair comparison, the same type of DE features were used as the original input across all models. Table 1 lists the performance of each model in the subject-dependent experiments.

From Table 1, it can be observed that all the baseline models achieved high test accuracies, particularly DGGN and Comformer, which are roughly on par 94.7%. Compared to other models, CAG-TEN exhibited a test accuracy lead of 6.87%, 5.13%, 0.61%, and 0.7% respectively, while also maintaining a lower standard deviation. These results highlight the significant advancements of CAG-TEN in emotion recognition tasks. Specifically, DGCNN relies solely on a custom adjacency matrix in the EEG spatial domain for dynamic learning, neglecting the dynamic features at the channel level within the spatial domain, and also lacks a process for learning temporal domain information. ACRNN employs a CNN network enhanced with channel attention scores within the EEG spatial domain. However, due to the limited receptive field of CNNs, they cannot focus on the global correlations of features and also fail to learn dynamic features at the channel level. Similarly, DGGN overlooks the dynamic features at the channel level from the EEG spatial domain, but it significantly enhances the spatio-temporal representation of subject samples through generative adversarial learning. Conformer uses the Encoders(Vaswani et al., 2017) network from the Transformer architecture to capture temporal information but still lacks consideration of dynamic features at the channel level in the EEG spatial domain. Overall, CAG-TEN demonstrates a deeper capability for exploring information in both the spatial and temporal domains, and its focus on dynamic features at the channel level within the EEG spatial domain is crucial and significant.

Table 1: subject-dependent experiments in SEED

| Model | Acc | Std |
|---|---|---|
| DGCNN(T. Song et al., 2018) | 88.53% | 9.6 |
| ACRNN(Tao et al., 2020) | 90.27% | 10.3 |
| Conformer(Y. Song et al., 2022) | 94.7% | 9.7 |
| DGGN(Gu et al., 2023) | 94.79% | **8.3** |
| **Ours** | **95.4%** | 9.3 |

## Subject-independent experiment

In contrast, subject-independent experiments take into account the EEG data of all subjects in the dataset. This experimental setup aims to evaluate the model's generalizability and universality in recognizing emotional states across different subjects, and the model does not need to be trained specifically for unknown subjects, making it more suitable for practical applications. To fully assess the model's generalizability, we employed a leave-one-out cross-validation approach. Likewise, in the comparative experiments, the same input data were used for all models. Table 2 lists the performance of each model in the subject-independent experiments.

From Table 2 we can see that since the individual differences under cross-domain conditions cannot be ignored, the overall test accuracy of the subject-independent experiment is lower than that of the subject-dependent experiment. But compared to other baseline models, CAG-TEN still maintains its lead. And on the premise of having a large amount of subject data, the stability in CAG-TEN is also the best. This is because assigning a unique parameter space to each channel can better adapt to the differences between different subject

Table 2: subject-independent experiments in SEED

| Model | Acc | Std |
|---|---|---|
| DGCNN(T. Song et al., 2018) | 79.95% | 10.02 |
| ACRNN(Tao et al., 2020) | 80.27% | 13.32 |
| Conformer(Y. Song et al., 2022) | 86.39% | 6.32 |
| DGGN(Gu et al., 2023) | 83.84% | 6.92 |
| **Ours** | **87%** | **5.56** |

samples, and the complexity of the model can also be enhanced by adjusting the depth $d$ of the parameter space, and the model can also accommodate more information and details. Overall, CAG-TEN's generalization in cross-domain context is promising.

## Conclusion

In this paper, we introduce a spatio-temporal model for EEG-based emotion recognition, named CAG-TEN. The CAG-TEN consists of two core modules: the CAGCN (Channel Adaptive Graph Convolutional Network) and the TE (Temporal Encoder). In the spatial domain, we offer a novel graph convolutional perspective by designing the CAGCN to focus on exploring channel-level dynamic features of EEG data while also addressing the extraction of global dynamic features. This represents the first application of channel-adaptive concepts in the domain of emotion recognition that we are aware of. We restructured and decomposed the traditional GCN parameter matrix to obtain two key inputs: the channel embedding matrix $E_g$ and the parameter pool matrix $W_g$, which are used to explore the channel-level dynamic features of EEG data. Additionally, we designed the adjacency matrix $A$ as $E_g \times E_g^T$ to explore global dynamic features. In the temporal domain, inspired by the Encoders concept from the Transformer architecture, we designed the TE to perform temporal sequence modeling on the outputs from the CAGCN. TE possesses significant advantages in global parallel computation and the exploration of long-term dependencies, playing an indispensable role in the exploration of sequential features. Moreover, we also designed a data augmentation method called Random Label Recombination, which doubles the training samples. Finally, comparative experimental results show that CAG-TEN has higher test accuracy and stability compared to the other four baseline models. It is worth discussing whether CAG-TEN, as a synchronized spatio-temporal model, also possesses similar decoding performance in other EEG tasks, which will be one of the themes of our future work.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant no 62102133, the High-level and Urgently Needed Overseas Talent Programs of Jiangxi Province under Grant no: 20232BCJ25026, the Kaifeng Major Science and Technology Project under Grant no: 21ZD011, the Ji'An Finance and Science Foundation under Grants no: 20211–085454, 20222–151746, 20222–151704, Ji'an key core common technology "reveal the list" Project under Grant no: 2022-1. Wenying Duan's work was supported in part by Technical Leaders in Major Disciplines-Leading Talents Project (Grant No. 20225BCI22016), Jiangxi Double Thousand Plan (Grant No. JXSQ2023201022).

## References


Bai, L., Yao, L., Li, C., Wang, X., & Wang, C. (2020). Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, *33*, 17804–17815.

Bazgir, O., Mohammadi, Z., & Habibi, S. A. H. (2018). Emotion recognition with machine learning using eeg signals. In *2018 25th national and 3rd international iranian conference on biomedical engineering (icbme)* (pp. 1–5).

Bos, D. P.-O., Reuderink, B., van de Laar, B., Gürkök, H., Mühl, C., Poel, M., … Nijholt, A. (2010). Human-computer interaction for bci games: Usability and user experience. In *2010 international conference on cyberworlds* (pp. 277–281).

Canal, F. Z., Müller, T. R., Matias, J. C., Scotton, G. G., de Sa Junior, A. R., Pozzebon, E., & Sobieranski, A. C. (2022). A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*, *582*, 593–617.

Duan, R.-N., Zhu, J.-Y., & Lu, B.-L. (2013). Differential entropy feature for EEG-based emotion classification. In *6th international ieee/embs conference on neural engineering (ner)* (pp. 81–84).

Duan, W., He, X., Zhou, Z., Thiele, L., & Rao, H. (2023). Localised adaptive spatial-temporal graph neural network. In *Proceedings of the 29th acm sigkdd conference on knowledge discovery and data mining* (pp. 448–458).

Fallman, D. (2003). Design-oriented human-computer interaction. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 225–232).

Gu, Y., Zhong, X., Qu, C., Liu, C., & Chen, B. (2023). A domain generative graph network for eeg-based emotion recognition. *IEEE Journal of Biomedical and Health Informatics*.

Jenke, R., Peer, A., & Buss, M. (2014). Feature extraction and selection for emotion recognition from eeg. *IEEE Transactions on Affective computing*, *5*(3), 327–339.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Li, X., Zhang, Y., Tiwari, P., Song, D., Hu, B., Yang, M., … Marttinen, P. (2022). Eeg based emotion recognition: A tutorial and review. *ACM Computing Surveys*, *55*(4), 1–57.

Lotte, F. (2015). Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces. *Proceedings of the IEEE*, *103*(6), 871–890.

Manessi, F., Rozza, A., & Manzo, M. (2020). Dynamic graph convolutional networks. *Pattern Recognition*, *97*, 107000.

Nie, W., Ren, M., Nie, J., & Zhao, S. (2020). C-gcn: Correlation based graph convolutional network for audio-video emotion recognition. *IEEE Transactions on Multimedia*, *23*, 3793–3804.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and cognition*, *14*(1), 30–80.

Song, T., Zheng, W., Song, P., & Cui, Z. (2018). Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, *11*(3), 532–541.

Song, Y., Zheng, Q., Liu, B., & Gao, X. (2022). Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 710–719.

Tao, W., Li, C., Song, R., Cheng, J., Liu, Y., Wan, F., & Chen, X. (2020). Eeg-based emotion recognition via channel-wise attention and self attention. *IEEE Transactions on Affective Computing*.

Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B. W., & Zafeiriou, S. (2017). End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of selected topics in signal processing*, *11*(8), 1301–1309.

Vanderlind, W. M., Millgram, Y., Baskin-Sommers, A. R., Clark, M. S., & Joormann, J. (2020). Understanding positive emotion deficits in depression: From emotion preferences to emotion regulation. *Clinical psychology review*, *76*, 101826.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., … others (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, *83*, 19–52.

Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., & Weinberger, K. (2019). Simplifying graph convolutional networks. In *International conference on machine learning* (pp. 6861–6871).

Yang, H., Han, J., & Min, K. (2019). A multi-column cnn model for emotion recognition from eeg signals. *Sensors*, *19*(21), 4736.

Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: a comprehensive review. *Computational Social Networks*, *6*(1), 1–23.

Zhang, T., Wang, X., Xu, X., & Chen, C. P. (2019). Gcb-net: Graph convolutional broad network and its application in emotion recognition. *IEEE Transactions on Affective Computing*, *13*(1), 379–388.

Zheng, W.-L., & Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162-175. doi: 10.1109/TAMD.2015.2431497

Zhong, P., Wang, D., & Miao, C. (2020). Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, *13*(3), 1290–1301.