Residual GCB-Net: Residual Graph Convolutional Broad Network on Emotion Recognition

Qilin Li, *Student Member, IEEE*, Tong Zhang, *Member, IEEE*, C. L. Philip Chen, *Fellow, IEEE*, Ke Yi, and Long Chen, *Member, IEEE*

***Abstract*—Electroencephalogram (EEG) data are commonly applied in the emotion recognition research area. It can accurately reflect the emotional changes of the human body by applying graphical-based algorithms or models. EEG signals are nonlinear signals. Biological tissues' adjustment and adaptive ability will inevitably affect electrophysiological signals, making EEG have the typical nonlinear characteristics. The graph convolutional broad network (GCB-net) extracted features from nonlinear signals and abstract features via a stacked convolutional neural network. It adopted the broad concept and enhanced the feature by the broad learning system (BLS), obtaining sound results. However, it performed poorly with the increasing network depth, and the accuracy of some features decreased with BLS. This article proposed a residual graph convolutional broad network (Residual GCB-net), which promotes the performance on a deeper layer network and extracts higher level information. It substitutes the original convolutional layer with residual learning blocks, which solves the deep learning network degradation and extracts more features in deeper networks. In the SJTU emotion EEG data set (SEED), GCB-Res net could obtain the best accuracy (94.56%) on the all-frequency band of differential entropy (DE) and promote much on another feature. In Dreamer, it obtained the best accuracy (91.55%) on the dimension of Arousal. The result demonstrated the excellent classification performance of Residual GCB-net in EEG emotion recognition.**

***Index Terms*—Broad learning system (BLS), emotion recognition, graph convolutional broad network (GCB-net), graph convolutional neural network (CNN), residual graph convolutional broad network (Residual GCB-net).**

Manuscript received 2 September 2021; revised 11 December 2021; accepted 21 January 2022. Date of publication 31 January 2022; date of current version 11 December 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFA0706200; in part by the National Natural Science Foundation of China under Grant 62076102, Grant U1813203, and Grant U1801262; in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2020B1515020041; in part by the Science and Technology Major Project of Guangzhou under Grant 202007030006; in part by the Science and Technology Program of Guangzhou under Grant 202002030250; in part by the Program for Guangdong Introducing Innovative and Entrepreneurial Teams under Grant 2019ZT08X214; and in part by the Guangdong–Hong Kong–Macao Greater Bay Area Center for Brain Science and Brain-Inspired Intelligence Fund under Grant 2019016. *(Corresponding author: Tong Zhang.)*

Qilin Li, Tong Zhang, C. L. Philip Chen, and Ke Yi are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China, and also with Pazhou Lab, Guangzhou 510335, China (e-mail: tony@scut.edu.cn).

Long Chen is with the Department of Computer and Information Science, University of Macau, Macau, China.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCDS.2022.3147839.

Digital Object Identifier 10.1109/TCDS.2022.3147839

I. Introduction

EMOTION recognition plays an essential part in human–computer interaction [1], [2], making it possible for the computer to understand human feelings via the links of the brain responses to interior or exterior actions. It is necessary to obtain nonphysiological or physiological signals to accomplish the emotion recognition objective. The former mainly contains facial expression images [3], body gestures [4], and voice signals [5] and semantic information in natural language [6]. The latter generally includes electrocardiogram (ECG) [7], electromyogram (EMG) [8], and electroencephalogram (EEG) [9], [10]. Among physiological signals, due to high precision, low cost, and easy operation, the EEG signal is one of the most generally employed in automatic emotion recognition [11], [12].

EEG signal is a noninvasive, convenient and low-cost approach in the emotion recognition field [13], [14]. It is directly captured from the brain cortex, reflecting the mental relation between emotional states and brain activity [15]. EEG could capture subtle mood changes [16] through multielectrodes positioned on the scalp to record electrical activities of the brain. Besides, it could provide emotional analysis details via its high temporal resolution (can be accurate to within 100 ms). Nevertheless, it is limited by its insufficient spatial resolution, and it may also be affected by some external factors in the experiment. For instance, it can be impacted by the emotion complexity and individual differences, and susceptible to electro-oculogram (EOG) signals, ECG signals, muscular activities and power line interference [17]. Although predecessors have discerned those problems [18], [19] and can be eliminated during analysis, EEG emotion recognition is still a defiant work.

EEG emotion recognition could be generally parted into three phases: 1) preprocess; 2) feature extraction; and 3) classification. In a preprocess step, noise and artifacts (the recorded movement that is not of the cerebral origin) should be determined and removed from raw EEG signals. Due to some unexpected factors that may exist in the acquisition process, the collected EEG signal may appear incomplete. Methods [20] should be utilized to address incomplete data effectively. Wu *et al.* proposed data-characteristic-aware latent factor (DCALF) [21] addressing the issue of Web service selection and recommendation. To promote the convergence speed of nonnegative latent factor (NLF) [22], Luo *et al.* [23] proposed FNLF. Smooth L_1-norm-oriented

latent factor (SL-LF) model [24] has both strong robustness and high accuracy in predicting the missing data of a high-dimensional and sparse matrix. In the emotion recognition field, the EEG features are mainly separated into time-domain, frequency-domain features, and time–frequency-domain features. Most EEG devices collect the time domain features, so they are the most intuitive and easy to obtain. However, the time-domain feature cannot display the signal's frequency information. Consequently, the researchers [25] convert the original time-domain signal to the frequency domain to obtain the frequency spectrum via the Fourier transform (FT) and then decompose the frequency band into five sub-bands closely related to human psychological activities [26]: δ, θ, α, β, γ. The time-frequency domain can obtain the time and frequency information of signals simultaneously, improve the processing ability of unstable signals, and roughly calculate the time of the beginning and duration of emotions.

In the past few decades, many traditional statistical approaches, signal processing methods [27] and machine learning methods [28] lend themselves to the EEG emotion classification problem [29]. In the literature, the general approaches of EEG feature extraction consist of the Gaussian mixture model (GMM) [30], support vector machine (SVM) [31], deep learning (DL), visibility graph (VG) [32], linear discriminant analysis (LDA) [33], decision tree (DT) [34], ensemble [35], *K*-nearest neighbors (KNN) [36], and logistic regression [37]. Among them, DL has obtained the best capability. As one of the most famous DL methods, convolutional neural networks (CNNs) is commonly used in various classification problems [38], [39]. However, it fails to deal with the data with an irregular or non-Euclidean structure, for example, EEG [40]. To solve this problem, Song *et al.* proposed dynamical graph CNNs (DGCNNs) [41] based on GCN [42], which use a graph to model the multichannel EEG features and dynamically learn the intrinsic relationship between different EEG channels. However, it could not solve high computational burdens, large memory requirements, and intractable batch-processing problems [43]. Hong *et al.* proposed a VG-GCN [44] with hierarchically coarsened random walk (hcr-walk) method by taking advantage of the classic random walk and node/edge encapsulation, which combines advantages of GCN and GraphSAGE [45]. By integrating the advantages of DGCNN and broad learning system (BLS) [46]–[48], Wang *et al.* proposed the broad dynamical graph learning system (BDGLS) [49] to extract features on non-Euclidean domain and randomly generating nodes to find the desired connection weights simultaneously. Zhang *et al.* designed the graph convolutional broad network (GCB-net) [50] to explore the deeper level information of graph-structured data. With the enhancement of BLS, it could attain the highest accuracy of 94.24% on differential entropy (DE) feature. However, their work [41], [49], [50] does perform not well with the increasing number of CNN layers. In order to further improve the accuracy, residual learning [51] is adopted in this article to promote the performance of the deeper layer network.

Residual learning solves two deeper neural network training problems: 1) gradient vanishing/exploding and 2) network degradation. It utilizes the form of a jump layer connection. That is, it directly adds the unit's input to the unit's output and then activates it. Therefore, the residual network can be easily implemented with the mainstream automatic differential DL framework and directly use the BP algorithm to update the parameters. Experiments [51] show that the residual network solves the degradation problem of deep neural networks very well and has achieved excellent results on image tasks such as ImageNet and CIFAR-10. The residual network also converges faster under the premise of the same number of layers.

BLS has shown its promising performance [46], [47] on a small data set, and it promotes the performance of GCB-net on some features [50]. It mainly contains feature and enhancement nodes. The processing steps are in the following: inputs are stochastically mapped to feature nodes, and then features are created into the enhancement parts. Results of both nodes are saved and finally connected and sent into output. This flat network structure with broad random possibility makes it possible to find suitable features to predict. As it performs well on classification and cost less time, this article adopts it to enhance the residual graph convolutional broad network (Residual GCB-net) features.

These work contributions can be displayed in the following aspects.

1) Residual learning block is utilized to promote the performance of CNN feature extraction. The Residual GCB-net uses graph convolution layer to extract features on irregular data and feed them into residual convolution blocks to obtain deeper layer features than those in plain networks.
2) Both deep and broad approaches are adopted in the model. The Residual GCB-net uses residual blocks to assure the accuracy of the deeper network and gulp-concat the outputs of graph convolution and residual blocks. It not only extracts the high-level features but preserves the multilevel features of all hierarchical layers.
3) BLS is utilized to enhance the features from the Residual GCB-net. It generates learned space to a possible low-dimension vector space and search for suitable features with enhancement nodes in the broad vector space. It further improves the performance of the Residual GCB-net.

II. Preliminary

This section introduces some preliminary knowledge about the graph convolution neural networks, the residual learning structure, and the BLS, contributing to the construction of the Residual GCB-net model. Mathematical notations in the formula are listed in Table I.

A. Graph Convolutional Neural Networks

Graph FT (GFT) [52] is adopted to process the graph-structure data in the following.

Define a undirected or connected graph as $\mathcal{G} = (\mathcal{K}, \mathcal{Q}), \mathbf{W_A})$. $\mathcal{K}$ denotes a vertex set with $|\mathcal{K}| = n$ nodes and $\mathcal{Q}$ represents the set of edges connecting these nodes. $(k_i \in \mathcal{K}, k_j \in \mathcal{K})$. The adjacency matrix $\mathbf{W_A}$ describes the connections between any two nodes in $\mathcal{K}$, and the degree matrix with diagonal entries is defined as $\mathcal{D}_{ii} = \sum_j \mathbf{W_{A_{ij}}}$. According

TABLE I
TABLE OF MATHEMATICAL NOTATIONS IN PRELIMINARY

Mathematical Notations	Description
$\mathcal{G}$	a undirected or connected graph
$\mathcal{K}, \mathcal{Q}$	$\mathcal{K}$ defines nodes in $\mathcal{G}$, and $\mathcal{Q}$ defines edges
$\mathbf{W_A}$	the adjacency matrix of $\mathcal{G}$
$\mathcal{L}$	the graph Laplacian matrix
$\mathbf{U}$	the eigenvectors matrix
$\mathbf{\Lambda}$	the diagonal matrix with eigenvalues
$\odot$	element-wise Hadamard product
$T_k(\cdot)$	the Chebyshev polynomials
$\mathbf{\Theta}$	all learned parameters in Residual GCB-net

to the graph Laplacian matrix $\mathcal{L} = \mathcal{D} - \mathbf{W_A}$, the normalized version can be denoted as: $\mathcal{L} = \mathbf{I}_n - \mathcal{D}^{-1/2}\mathbf{W_A}\mathcal{D}^{-1/2}$, where $\mathbf{I}_n$ is the identify matrix. With the eigendecomposition (SVD) of the graph Laplacian matrix $\mathcal{L} = \mathbf{U}\Lambda\mathbf{U}^\top$, the eigenvectors matrix $\mathbf{U} = [\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}] \in \mathbb{R}^{n\times n}$ is obtained, where $\Lambda = \text{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{n-1}) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with corresponding eigenvalues.

Taking $\mathbf{U}$ as the Fourier basis, the GFT expression of a given spatial signal $\mathbf{x}$ is

$$\hat{\mathbf{x}} = \mathbf{U}^\top\mathbf{x}. \tag{1}$$

In (1), $\hat{\mathbf{x}}$ denotes the transformed frequency-domain signal, and the inverse of its GFT in (2) is

$$\mathbf{x} = \mathbf{U}\mathbf{U}^\top\mathbf{x} = \hat{\mathbf{U}}\mathbf{x}. \tag{2}$$

Then, define the convolution of signals $\mathbf{x}_1$ and $\mathbf{x}_2$ on the graph $*_\mathcal{G}$ in the following:

$$\mathbf{x}_1 *_\mathcal{G} \mathbf{x}_2 = \mathbf{U}\Big(\Big(\mathbf{U}^\top\mathbf{x}_1\Big) \odot \Big(\mathbf{U}^\top\mathbf{x}_2\Big)\Big) \tag{3}$$

where $\odot$ denotes elementwise Hadamard product. Define $d_\theta(\cdot)$ as a filtering function, and $\mathbf{x}$ filtered by $g_\theta(\cdot)$ can be represented as

$$d_\theta(\mathcal{L})\mathbf{x} = d_\theta\Big(\mathbf{U}\Lambda\mathbf{U}^\top\Big)\mathbf{x} = \mathbf{U}d_\theta(\Lambda)\mathbf{U}^\top\mathbf{x}. \tag{4}$$

To reduce the computational complexity ($\mathcal{O}(n^2)$ in (4)), the Chebyshev polynomials [53] $T_k(\cdot)$ are applied to compute $d_\theta(\mathcal{L})$ recursively from $\mathcal{L}$. The parameterized z−1 order Chebyshev filter can be expressed as

$$d'_\theta(\Lambda) = \sum_{z=0}^{Z-1}\theta_z T_z\Big(\tilde{\Lambda}\Big) \tag{5}$$

where θ_k denotes the coefficient of z-order Chebyshev polynomial and $\tilde{\Lambda} = 2\Lambda/\lambda_{\max} - \mathbf{I}_n$ is a diagonal matrix of rescaled eigenvalues. The Chebyshev polynomials can be described as $T_z(x) = 2xT_{z-1}(x) - T_{z-2}(x)$, where $T_0(x) = 1$ and $T_1(x) = x$. Then, (4) can be approximately expressed as

$$\begin{aligned} d'_\theta *_\mathcal{G} \mathbf{x} &= d'_\theta(\mathcal{L})\mathbf{x} \\ &= \sum_{k=0}^{Z-1}\theta_z T_z\Big(\tilde{\mathcal{L}}\Big)\mathbf{x}. \end{aligned} \tag{6}$$

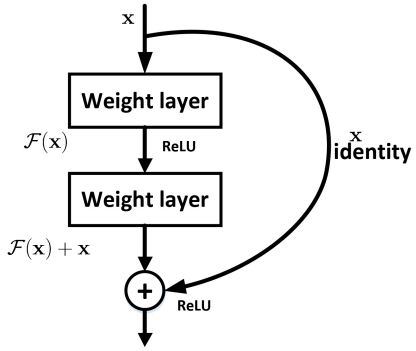

Fig. 1. Framework of shortcut connection. Denote $\mathbf{x}$ as input and $\mathcal{F}(\mathbf{x})$ is the residual function. By identity connection, the output of the previous layer (or several layers) can be added to the current layer.

Defining $\bar{\mathbf{x}}_z$ is equal to $T_k(\tilde{\mathcal{L}})\mathbf{x} \in \mathbb{R}^n$, recursively compute it as $\bar{\mathbf{x}}_z = 2\tilde{\mathcal{L}}\bar{\mathbf{x}}_{z-1} - \bar{\mathbf{x}}_{z-2}$ with $\bar{\mathbf{x}}_0 = \mathbf{x}$ and $\bar{\mathbf{x}}_1 = \tilde{\mathcal{L}}\mathbf{x}$. Rescale the Laplacian matrix via $\tilde{\mathcal{L}} = 2\mathcal{L}/\lambda_{\max} - \mathbf{I}_n$, (6) can be written as

$$\begin{aligned} d'_\theta(\mathcal{L})\mathbf{x} &= \sum_{z=0}^{Z-1}\theta_z\bar{\mathbf{x}}_z \\ &= \big[\bar{\mathbf{x}}_0, \ldots, \bar{\mathbf{x}}_{Z-1}\big]\boldsymbol{\theta}. \end{aligned} \tag{7}$$

Then, the time complexity of (7) is minimized to $\mathcal{O}(K|\mathcal{E}|)$.

As proposed by Song *et al.* [41] and Zhang *et al.* [50], DGCNN and GCB-net can capture the adjacency matrix $\mathcal{M}$ of graph dynamically to extract more discriminative features. Similarly, the loss $\mathcal{L}oss_{\text{RGCB}}$ with cross entropy $\mathcal{C}ross\_\mathcal{E}$ of Residual GCB-net is defined as

$$\mathcal{L}oss_{\text{RGCB}} = \mathcal{C}ross\_\mathcal{E}\big(\mathbf{r}, \mathbf{r}'\big) + \rho\|\Theta\| \tag{8}$$

where $\mathbf{r}$ and $\mathbf{r}'$ define true and predicted labels, respectively. Θ is the matrix of all parameters learned in the Residual GCB-net model and ρ is a regularization coefficient. The partial derivative of the loss function w.r.t. adjacency matrix $\mathcal{M}$ can be deduced in the following:

$$\frac{\partial\mathcal{L}oss_{\text{RGCB}}}{\partial\mathcal{M}} = \frac{\partial\mathcal{C}ross\_\mathcal{E}\big(\mathbf{r}, \mathbf{r}'\big)}{\partial\tilde{\mathcal{L}}} \cdot \frac{\partial\tilde{\mathcal{L}}}{\partial\mathcal{M}} + \rho\frac{\partial\|\Theta\|}{\partial\mathcal{M}} \tag{9}$$

and $\mathcal{M}$ can be updated by the following rule:

$$\mathcal{M} = (1-\delta)\mathcal{M} + \delta\frac{\partial\mathcal{L}oss_{\text{RGCB}}}{\partial\mathcal{M}} \tag{10}$$

and δ denotes the learning rate.

B. Residual Learning

Residual learning uses a shortcut-connection method, as shown in Fig. 1, which provides two mapping methods: 1) residual mapping and 2) identity mapping. It becomes a solution to the degradation problem. If the network has reached the optimum, continue to deepen the network, the residual mapping will be pushed to 0, only identity mapping is left, so theoretically, the network has always been in the optimal state, and the performance of the network will not decrease as the depth increases.

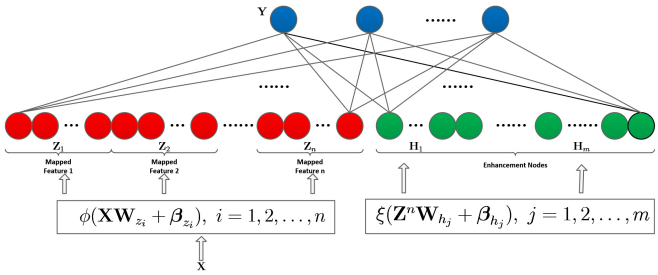

Fig. 2. Configuration of BLS. Denote $\mathbf{X}$ as input, $\mathbf{W}_{z_i}$ and $\mathbf{W}_{h_j}$ as weighting matrices, and $\mathbf{Y}$ in blue as a label matrix. The red nodes and green nodes denote the mapped features and enhancement nodes, respectively.

Denote $\mathbf{x}$ as input and $\mathbf{y}$ as output, and $\mathcal{H}(\cdot)$ is the function to be fitted in the neural network unit. Residual learning could be defined in the following:

$$\mathbf{y}^l = \mathcal{H}\left(\mathbf{x}^{l-1}\right) = \mathcal{F}\left(\mathbf{x}^{l-1}\right) + \mathbf{x}^{l-1} \tag{11}$$

where l denotes the number of network layers. Residual learning block transform the problem $\mathcal{H}(\mathbf{x}^{l-1}) - \mathbf{x}^{l-1} \rightarrow 0$ to $\mathcal{F}(\mathbf{x}^{l-1}) \rightarrow 0$, which makes it easier for the model to converge.

C. Broad Learning System

BLS is a flat network, aiming to expand features to broad space through random methods, has been applied to numerous fields, including facial emotion recognition [3], natural language processing [54], [55], and time-series prediction [56]. The network mainly consists of feature nodes and enhancement nodes as is illustrated in Fig. 2. Denote graph-structure data and predicted label matrix as $\mathbf{X} \in \mathbb{R}^{N\times M}$ and $\mathbf{Y} \in \mathbb{R}^{N\times n_{\text{class}}}$, respectively. N denotes the sample number, M is described as the feature dimension, and n_{class} is the number of classes. Feature nodes can be formulated in the following:

$$\mathbf{Z}_i \triangleq \phi\left(\mathbf{X}\mathbf{W}_{z_i} + \boldsymbol{\beta}_{z_i}\right),\ i = 1, 2, \ldots, n \tag{12}$$

where the weight matrix $\mathbf{W}_{z_i} \in \mathbb{R}^{M\times N_f}$ and the bias matrix $\boldsymbol{\beta}_{z_i} \in \mathbb{R}^{1\times N_f}$ are randomly generated, and N_f denotes the dimension of generated features. Define the feature nodes as $\mathbf{Z}^n \triangleq [\mathbf{Z}_1, \ldots, \mathbf{Z}_n] \in \mathbb{R}^{N\times(n*N_f)}$, and the enhancement nodes can be computed in the following:

$$\mathbf{H}_j \triangleq \xi\left(\mathbf{Z}^n\mathbf{W}_{h_j} + \boldsymbol{\beta}_{h_j}\right),\ j = 1, 2, \ldots, m. \tag{13}$$

Similarly, $\mathbf{W}_{h_j}$ and $\boldsymbol{\beta}_{h_j}$ are also randomly initialized. Consequently, BLS can be deduced as

$$\begin{aligned} Y &= [\mathbf{Z}_1, \ldots, \mathbf{Z}_n|\mathbf{H}_1, \ldots, \mathbf{H}_m]\mathbf{W} \\ &= \left[\mathbf{Z}^n|\mathbf{H}^m\right]\mathbf{W}. \end{aligned} \tag{14}$$

III. Residual Graph Convolutional Broad Network

In this section, a Graph Convolutional Broad Residual model is proposed and utilized in the EEG emotion recognition problem. Besides, an improved algorithm BLS has been given in detail. Mathematical notations in the formula are listed in Table II.

TABLE II
Table of Mathematical Notations in Residual GCB-Net

Mathematical Notations	Description
$\omega \in \mathbb{R}^{K\times n_F}$	the matrix formed by coefficients of Chebyshev filters
n_F	the number of filters
$\mathcal{A}$	the activation function
$\mathcal{F}(x, \{W_i\})$	the residual mapping to be learned

A. Residual GCB-Net on EEG Emotion Recognition

The configuration of the Graph Convolutional Broad Residual model is illustrated in Fig. 3. Based on the GCB model, the Residual GCB-net model is designed further to discover the hidden information via a deeper layer network and solve the gradient diffusion or explosion problem. It first operates convolution via Chebyshev graph filters to obtain the results of the combination of the convolutional results with each of the Chebyshev polynomial components. Then, residual-structure convolutional blocks are applied to high-level feature extraction. Results of each block and the graph convolution filter are flattened and concatenated. In this model, the final result can be fed into the full-connection layer and predicted by the softmax function.

Define N and M as the number of samples and features, respectively. Input could be represented as $\mathbf{X} \in \mathbb{R}^{N\times M}$. Define C as the class of labels and Output could be represented as $\mathbf{Y} \in \mathbb{R}^{N\times C}$. For $Z-1$ order graph convolution layer, the output could be represented in the following:

$$\text{chebe}(\mathbf{x}) = \mathcal{A}\left(\left[T_0\left(\tilde{\mathcal{L}}\right)\mathbf{x}, \ldots, T_{Z-1}\left(\tilde{\mathcal{L}}\right)\mathbf{x}\right]\omega\right). \tag{15}$$

$\omega \in \mathbb{R}^{K\times n_F}$ represents the matrix formed by coefficients of Chebyshev filters, and n_F is the number of filters. $\mathcal{A}$ denotes the activation function, $\text{ReLU}(x) = \max(x, 0)$ in this article in order to add nonlinear factors. Then, the output $g(\mathbf{x}) \in \mathbb{R}^{N\times M\times n_F}$ is fed into the following residual learning blocks.

In this article, the residual learning block could be defined as

$$R(x) = \mathcal{F}(x, \{W_i\}) + W_s x. \tag{16}$$

In (16), x and $R(x)$ denote input and output, respectively. Function $\mathcal{F}(x, \{W_i\})$ is the residual mapping to be learned, and $W_s x$ is the identity mapping. The operation $\mathcal{F} + W_s x$ is realized by shortcut connection and elementwise addition. In this article, since each block has different dimensions of x and $\mathcal{F}$, a linear projection W_s is employed on the shortcut connections to make dimensions the same. It only changes the dimension of input but introduce neither extra parameter nor computation complexity.

For one thing, it assures the fair comparisons between plain and residual networks, as it has the same number of parameters, depth, width, and computational cost.

For another, it solves the vanishing gradient and degradation problem. Equation (16) could be defined as $x_{l+1} = \mathcal{F}(x_l, \{W_i\}) + W_s x_l$. Recursively, an L-layer unit could be represented as $x_L = \sum_{i=l}^{L-1} \mathcal{F}(x_i, \{W_i\}) + W_s x_l$. It shows that any units L and l have residual characteristics. To verify that the gradient can be passed without loss, according to the chain

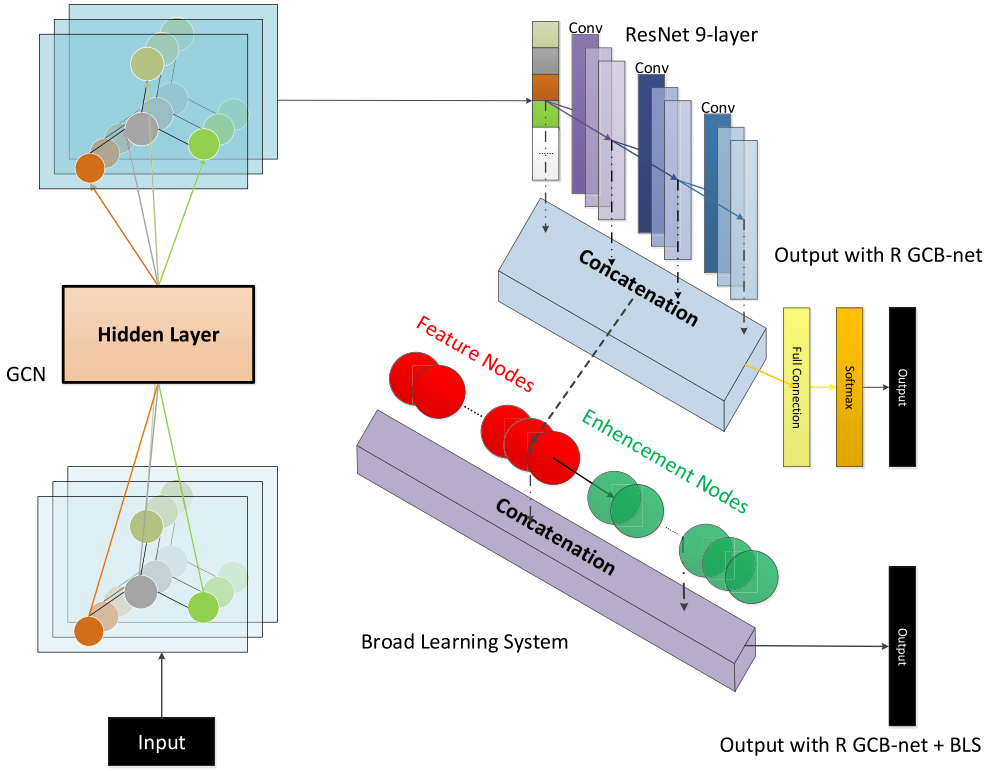

Fig. 3. Structure of Residual GCB-net + BLS. It could be separated in three parts: GCN, Residual block, and BLS. The nodes in GCN describe the graph-structure data. The GCN part shows that GCN will not change the node position and graph-structure. GCN extracts the graph features and transforms them into standardized structure. Different color layers indicate the corresponding residual blocks. In the BLS part, nodes in red are feature nodes and nodes in green are enhancement nodes.

rule, the gradient of the loss function with respect to x_l can be expressed as

$$\frac{\partial \text{Loss}}{\partial \mathbf{x_l}} = \frac{\partial \text{Loss}}{\partial \mathbf{x_L}} \cdot \frac{\partial \mathbf{x_L}}{\partial \mathbf{x_l}} = \frac{\partial \text{Loss}}{\partial \mathbf{x_L}} \cdot \left(1 + \frac{\partial}{\partial \mathbf{x_l}} \sum_{i=l}^{L-1} F(x_i, \{\omega_i\})\right). \tag{17}$$

It is clear that $(\partial \text{Loss}/\partial \mathbf{x_L})$ indicates that the gradient of the L layer can be directly passed to any layer l that is shallower than it. Due to $(\partial/\partial \mathbf{x_l}) \sum_{i=l}^{L-1} \mathcal{F}(x_i, \{\omega_i\}) > 0$, gradient will not vanish. W_s could be 1 if channels of both input and output are the same. The identity residual learning block could be defined as

$$R'(x) = \mathcal{F}(x, \{W_i\}) + x \tag{18}$$

With the process of a few residual learning blocks, outputs are flattened and linked together. Using $G(\mathbf{x}) = \text{flat}(\text{chebe}(\mathbf{x}))$ and $R_f(\mathbf{x})$ define the flatten function, respectively. For instance, $R_f 1 = \text{flat}(R_1(\mathbf{x}))$ and $R_1(\mathbf{x}) = R(\text{chebe}(\mathbf{x}), L_R, k_c)$, where $L_R = 1, 2, 3, \ldots, n$ indicates the length of residual mapping and k_c represents the dimension of convolutional kernel, and l is the number of residual blocks. The concatenation features of the Residual GCB block can be defined in the following:

$$\mathcal{R}GCB(\mathbf{x}) = \left[G, R_f 1, R_f 2, \ldots, R_f l\right] \tag{19}$$

$$\text{FC}(\mathbf{x}) = \mathcal{R}GCB(\mathbf{x})\mathbf{W}_f. \tag{20}$$

In (20), $\mathbf{W}_f \in \mathbb{R}^{N_{\text{RGCB}} \times C}$ is the affine matrix, where N_{RGCB} depends on the number of $\mathcal{R}GCB(\mathbf{x})$ features. Finally, the result of FC is fed to softmax function, processed to a probability distribution. The detailed algorithm and steps are illustrated in Algorithm 1.

B. Enhanced by BLS

BLS is an enhancement method to promote the accuracy of the Residual GCB-net. Features from Residual GCB-net($\mathcal{R}(\mathbf{x})$) are first processed by feature mapping and enhancement mapping, reflecting the broad random space. The mapped feature can be represented in the following:

$$\mathbf{Z}_i \triangleq \mathcal{R}(\mathbf{x})\left(\mathbf{W}_{z_i} + \boldsymbol{\beta}_{z_i}\right), \ i = 1, 2, \ldots, n. \tag{21}$$

In this article, tansig function $\mathcal{T}(\cdot)$ is adopted as the nonlinear function of enhancement nodes. Then, the enhancement nodes can be represented in the following:

$$\mathbf{H}_j \triangleq \mathcal{T}(\mathbf{x})\left(\mathbf{Z}^n \mathbf{W}_{h_j} + \boldsymbol{\beta}_{h_j}\right) \ j = 1, 2, \ldots, m. \tag{22}$$

In order to prevent overfitting and make the optimization solution stable and fast, regular l_2-norm regularization is adopted to the loss function. The objective function can be represented as

$$\arg\min_{\mathbf{W}_b} \ : \left\|\left[\mathbf{Z}^n | \mathbf{H}^m\right]\mathbf{W}_b - \mathbf{Y}\right\|_2^2 + \lambda \|\mathbf{W}_b\|_2^2. \tag{23}$$

In (23), denote $\mathbf{W}_b$ as weight matrix and λ as the regularization parameter. Consider the derivative of (23) w.r.t $\mathbf{W}_b$ and arrange

Algorithm 1 Algorithm of Training Residual GCB-Net on EEG Emotion Recognition

Input: the graph-structured input data $\mathbf{x}$, the number of Chebyshev parameter Z, the layers of regular convolution n_l, the length of residual mapping L_R, the dimension of convolutional kernel k_c, the number of label class C;

Output: the predicted label matrix $\mathbf{Y}_p$;

Fed $\mathbf{x}$ into the graph convolutional layer and obtain $chebe(\mathbf{x})$ via (15);

Fed the result of $chebe(\mathbf{x})$ to the first residual block $R_1(\mathbf{x}) = R(chebe(\mathbf{x}), L_R, k_c) = \mathcal{F}(chebe(\mathbf{x}), \{W_i\})_{L_R} + W_s \cdot chebe(\mathbf{x})$;

The residual mapping operation times is depends on the L_R, and the output of R_1 is:

$R_1 = W_s \cdot chebe(\mathbf{x}) + \sum_{i=0}^{L_R-1} \mathcal{F}(x_i, \{W_i\})$

for $z = 2, \ldots, l$ **do**

    Compute the z-th residual convolutional block $R_z(\mathbf{x}) = R(R_{z-1}(\mathbf{x}), L_R, k_c)$;

Flatten the outputs of $chebe(\mathbf{x})$ to 1-d vectors G;

for $z = 1, \ldots, l$ **do**

    Flatten the outputs of $R_z(\mathbf{x})$ to 1d vectors $R_f z$;

Concatenate the flattened vectors of all layers to obtain $\mathcal{R}GCB(\mathbf{x}) = \left[G, R_f 1, R_f 2 \ldots, R_f l\right]$;

Then, compute the FC layer: $FC(\mathbf{X}) = \mathcal{R}GCB(\mathbf{x})\mathbf{W}_f$,

Softmax the result of FC layer, and the predicted label matrix $\mathbf{Y}_p$ can be computed as $\mathbf{Y}_s = \text{softmax}(FC(\mathbf{X}))$;

TABLE III
SUMMARY PARAMETER SETTING OF SEED AND DREAMER DATA SET

Dataset	Feature/ electrodes	Frequency band	Rating metric
SEED	PSD (62) DE (62) DCAU (23 pairs) RASM (27 pairs) DASM (27 pairs)	$\delta, \theta, \alpha, \beta, \gamma$ bands	positive neural negative
DREAMER	PSD (62)	θ, α, β bands	arousal valence dominance

it to 0, $\mathbf{W}_b$ can be approximately computed as

$$\mathbf{W}_b = \left(\left[\mathbf{Z}^n|\mathbf{H}^m\right]^\top \left[\mathbf{Z}^n|\mathbf{H}^m\right] + \lambda \mathbf{I}\right)^{-1} \left[\mathbf{Z}^n|\mathbf{H}^m\right]^\top \mathbf{Y}. \quad (24)$$

IV. Experiments and Analysis for EEG Emotion Recognition

This section describes the details of the experimental materials. It offers the results and analysis of experiments on the DREAMER [57] data set, and the SJTU emotion EEG data set (SEED) [58].

A. Data Set

SEED and DREAMER are applied to evaluate the model, the summaries are illustrated in Table III.

1) SEED Description: There are 15 Chinese subjects (eight females and seven males) who experimented by watching 15 movies of different emotion types (five positive, five neutral, and five negative [59]) three times at intervals of al least a week to arouse their corresponding emotions [60]. SEED provides files with the extracted features, involving power spectral density (PSD), DE, which is first proposed in [61], differential caudality (DCAU), rational asymmetry (RASM), and differential asymmetry (DASM), as the difference and ratio between the DE features of 27 pairs of hemispherical asymmetric electrodes. PSD feature and DE feature used 62 electrodes to extract information about frequency and energy spectrum. DCAU feature used 23 pairs and described the differences between channel pairs. RASM feature and DASM feature used 27 pairs electrodes and extracted asymmetrical information. Each feature are extracted in the frequency bands: δ band (1–3 Hz), θ band (4–7 Hz), α band (8–13 Hz), β band (14–30 Hz), and γ band (31–50 Hz). In order to extract the EEG feature, each trial of EEG signal flow is divisional into blocks, containing 1 s of EEG signals.

2) DREAMER: There are 23 participants (nine females and 14 males) who experimented by watching 18 movies of different emotion types, containing nine emotions: 1) happiness; 2) amusement; 3) excitement; 4) surprise; 5) calmness; 6) disgust; 7) fear; 8) sadness; and 9) anger. Participants can assess their emotion form arousal (level of emotional activation), valence (negative to positive), and dominance (strength of emotion) with five points. In this data set, the signal was recorded at a sampling 128-Hz rate with 16 contact-sensors and 14 electrodes for study. DREAMER provides files with the PSD features as θ band (4–7 Hz), α band (8–13 Hz), and β band (14–30 Hz).

TABLE IV
TABLE OF PARAMETERS SETTING

Parameter	Setting	Description
K	64	the number of Chebyshev filters
ρ	0.003	learning rate
m	100	the number of groups of enhancement node
n	10	the number of groups of feature node
N_f	10	the dimension of feature nodes
N_e	10	the dimension of feature enhancement nodes
λ	$10^{[-2:4]}$	the l_2-regularized coefficient

B. Experimental Setting

Parameters are mainly adopted the same experimental setting with GCB-net [50]. Detailed setting is illustrated in Table IV.

1) Experimental Setting on SEED: SEED contains 45 experiments as 15 subjects × 3 sessions, and each experiment consists of 15 trials. The first nine trails are chosen as the training, and the rest of the same one are considered testing set.

2) Experimental Setting on DREAMER: In this experiment, the subject-dependent leave-one-session-out (LOO) cross-validation approach was applied to appraise the property of Residual GCB-net. Each subject watched 18 film clips, with recording 18 sessions of their corresponding EEG data. Each experiment always retain one session of the same participant as the testing set and the remaining training sets. When repeating

TABLE V
COMPARISONS OF THE EEG EMOTION RECOGNITION ACCURACY/STAND DEVIATION (%) ON SEED AMONG SVM, DBN, GCNN, DGCNN, GCB-NET, GCB-NET+BLS, RESIDUAL GCB-NET, AND RESIDUAL GCB-NET + BLS

Feature	Model	δ band	θ band	α band	β band	γ band	all($\delta, \theta, \alpha, \beta, \gamma$)
DE	SVM [58]	60.50/14.14	60.95/10.20	66.64/14.41	80.76/11.56	79.56/11.38	83.99/9.72
	DBN [58]	64.32/12.45	60.77/10.42	64.01/15.97	78.92/12.48	79.19/14.58	86.08/8.34
	GCNN [41]	72.75/10.85	74.40/8.23	73.46/12.17	83.24/9.93	83.36/9.43	87.40/9.20
	DGCNN [41]	74.25/11.42	71.52/**5.99**	74.43/12.16	83.65/10.17	85.73/10:64	90.40/8.49
	GCB-net [50]	80.38/10.04	76.09/7.54	81.36/11.44	88.05/9.84	88.45/9.67	92.30/7.40
	GCB-net + BLS [50]	79.98/8.93	76.51/9.56	81.97/11.05	89.06/8.69	89.10/9.55	94.24/6.70
	Residual GCB-net	80.88/9.04	77.60/7.34	81.95/10.29	88.97/8.50	90.06/9.39	92.86/6.70
	Residual GCB-net + BLS	**81.20/7.97**	**79.85**/9.70	**84.03/9.93**	**90.70/7.28**	**91.70/7.41**	**94.56/6.61**
PSD	SVM [58]	58.03/15.39	57.26/15.09	59.04/15.75	73.34/15.20	71.24 /16.38	59.60/15.93
	DBN [58]	60.05/16.66	55.03/13.88	52.79/15.38	60.68 /21.31	63.42/19.66	61.90/16.65
	GCNN [41]	69.89/13.83	70.92/9.18	73.18/12.74	76.21/10.76	76.15/**10.09**	81.31/11.26
	DGCNN [41]	71.23/11.42	71.20/8.99	73.45/12.25	77.45/10.81	76.60/11.83	81.73/**9.94**
	GCB-net [50]	72.76/12.45	72.31/**8.52**	74.29/11.22	81.05/12.23	81.85/11.13	83.75/10.63
	GCB-net + BLS [50]	72.90/13.19	74.48/9.03	76.99/**10.36**	83.30/**10.73**	83.12/11.95	84.32/10.61
	Residual GCB-net	73.95/**11.93**	72.42/9.70	74.65/11.49	82.81/11.91	84.12/11.48	84.90/11.53
	Residual GCB-net + BLS	**74.38**/12.45	**75.75**/10.16	**77.69**/12.03	**83.55**/10.90	**83.87**/10.96	**84.94**/10.93
DASM	SVM [58]	48.87/10.49	53.02/12.76	59.81/14.67	75.03/15.72	73.59/16.57	72.81/16.57
	DBN [58]	48.79/9.62	51.59/13.98	54.03/17.05	69.51/ 15.22	70.06/18.14	72.73/15.93
	GCNN [41]	57.07/**6.75**	54.80/9.09	62.97/13.43	74.97/13.40	73.28/13.67	76.00/13.32
	DGCNN [41]	55.93/9.14	56.12/**7.86**	64.27/12.72	73.61/14.35	73.50/16.6	78.45/**11.84**
	GCB-net [50]	65.04/8.18	65.36/8.97	70.10/11.46	83.22/**10.36**	83.44/13.21	79.67/14.06
	GCB-net + BLS [50]	62.36/10.66	65.00/10.31	70.91/10.84	85.55/11.39	86.04/10.85	82.09/13.14
	Residual GCB-net	65.48/8.55	66.17/9.97	70.67/11.22	85.45/11.82	84.02/11.36	80.13/12.74
	Residual GCB-net + BLS	**66.51**/9.41	**66.40**/10.28	**72.12/10.44**	**86.46**/11.45	**88.06/10.21**	**82.25**/12.58
RASM	SVM [58]	47.75/10.59	51.40/12.53	60.71/14.57	74.59/16.18	74.61/15.57	74.74/14.79
	DBN [58]	48.05/10.37	50.62/14.02	56.15/15.28	70.31/15.62	68.22/18.09	71.30/16.16
	GCNN [41]	59.70/**5.65**	55.91/8.82	59.97/14.27	79.45/13.32	79.73/13.22	84.06/12.86
	DGCNN [41]	57.79/6.90	55.79/**8.10**	61.58/12.63	75.79/13.07	82.32/11.54	85.00/12.47
	GCB-net [50]	62.66/7.22	65.07/8.85	70.79/10.78	85.62/**10.11**	85.79/12.05	86.40/**9.96**
	GCB-net + BLS [50]	62.56/8.83	62.22/11.12	71.43/10.83	87.03/11.16	85.59/11.18	87.73/10.19
	Residual GCB-net	66.47/6.88	66.22/8.81	71.71/**9.47**	86.01/10.53	87.52/10.46	88.98/**9.96**
	Residual GCB-net + BLS	**67.59**/7.83	**66.42** /10.31	**74.79**/12.35	**89.14**/10.51	**89.24/10.26**	**90.45**/10.22
DCAU	SVM [58]	55.92/14.62	57.16/10.77	61.37/15.97	75.17/15.58	76.44/15.41	77.38/11.98
	DBN [58]	54.58/12.81	56.94/12.54	57.62/13.58	70.70/16.33	72.27/16.12	77.20/14.24
	GCNN [41]	62.60/12.88	65.05/8.35	66.41/11.06	77.28/11.55	78.68/13.00	79.02/11.27
	DGCNN [41]	63.18/13.48	62.55/**7.96**	67.71/10.74	78.68/10.81	80.05/13.03	81.91/10.06
	GCB-net [50]	70.86/11.27	69.63/9.42	70.66/10.45	82.85/11.65	82.53/11.77	83.32/9.40
	GCB-net + BLS [50]	70.63/10.75	69.82/9.46	73.43/11.38	87.08/9.52	85.56/10.64	86.50/8.63
	Residual GCB-net	72.65/**9.27**	71.98/10.43	73.53/**9.24**	85.16/11.08	86.24/12.08	85.00/9.30
	Residual GCB-net + BLS	**72.81**/11.30	**72.26**/9.74	**74.27** /9.91	**87.91/9.47**	**87.70/9.30**	**87.94/8.59**

18 times, it assures every session has been used for training and testing and the result of each participant is by calculating the mean recognition accuracies of the whole session times.

This article divides five points class into two classes to implement binary classifications: the first two rating ratios denote the low valence, arousal, or dominance, and the last three rating ratios denote the high valence, arousal, or dominance.

C. Results and Analysis

1) Experiments on SEED: In the SEED, Residual GCB-net is confronted with the currently advanced approaches, such as SVM, deep belief networks (DBNs), GCNN, GDCNN, and GCB. The performance of all methods are experimented on all features with five bands and the all band ($\delta, \theta, \alpha, \beta$, and γ bands and the all band). The same as the computation [50], the mean accuracy and standard deviation were calculated from the better two sessions, the result of all features is illustrated in Table V and the result of all band ($\delta, \theta, \alpha, \beta, \gamma$) is shown in Fig. 4. $F1$ score and AUC of Residual GCB-net are illustrated in Table VI.

Fig. 4 shows that Residual GCB + BLS achieves much higher classification accuracies than another convincing methods. From Table V, it is distinct that all models obtain

TABLE VI
F1 SCORE AND AUC OF THE EEG EMOTION RECOGNITION ON SEED OF RESIDUAL GCB-NET

Feature	Model	δ band	θ band	α band	β band	γ band	all($\delta, \theta, \alpha, \beta, \gamma$)
DE	**Residual GCB-net**	0.73/0.80	0.68/0.77	0.70/0.79	0.84/0.89	0.84/0.89	0.88/0.92
PSD	**Residual GCB-net**	0.64/0.75	0.67/0.77	0.63/0.75	0.73/0.82	0.75/0.83	0.76/0.83
DASM	**Residual GCB-net**	0.54/0.68	0.55/0.70	0.60/0.73	0.77/0.84	0.75/0.83	0.70/0.79
RASM	**Residual GCB-net**	0.59/0.70	0.61/0.73	0.57/0.70	0.61/0.73	0.81/0.87	0.81/0.86
DCAU	**Residual GCB-net**	0.62/0.74	0.62/0.74	0.65/0.76	0.76/0.84	0.76/0.84	0.79/0.85

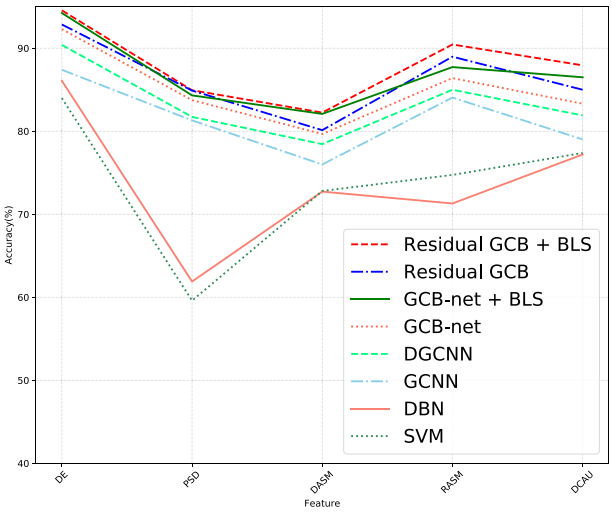

Fig. 4. Comparisons of the EEG Emotion Recognition Accuracy on SEED (all band) among SVM, DBN, GCNN, DGCNN, GCB-net, GCB-net + BLS, Residual GCB-net, and Residual GCB-net + BLS.

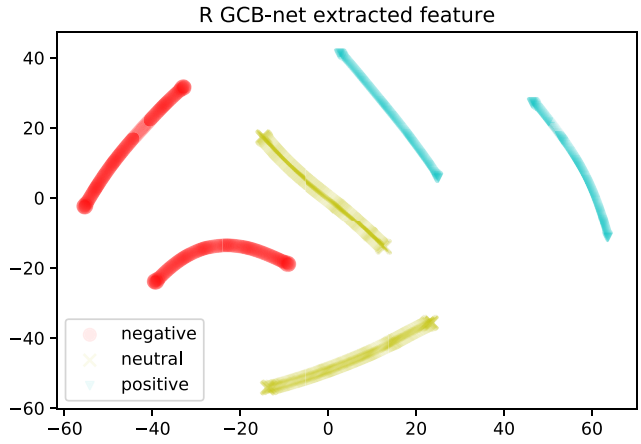

Fig. 5. DE features extracted by Residual GCB-net. This figure used the t-SNE [62] method to reduce the original features to two dimensions. In this figure, the red round dot defines the negative emotion, the yellow cross dot defines the neutral emotion, and the cyan triangle dot defines the positive emotion.

higher recognition accuracies on β band and γ band. There is a hypothesis that β band and γ band are more relevant to emotional responses. Compared to [41], [50], and [58], the Residual GCB-net model achieved the highest accuracy and promotion on all features with all bands. Different from the result of [50], Residual GCB-net + BLS had better performance on all features with all bands. It might suggest that Residual GCB-net could dig out more high-level features with its deeper layer network, allowing BLS to generate better nodes on the feature space. The best result of accuracy and standard deviation on DE with all band are updated to 94.56% and 6.61%, respectively.

Fig. 5 can visualize the results intuitively. It select the 12-DE features (the fourth person 3rd times) extracted by Residual GCB-net as an example. From this figure, it is clear that different emotion feature are distinctly separated in three different areas. It shows that Residual GCB-net has competitive performance on SEED emotion recognition.

2) Experiments on DREAMER: In experiments, Residual GCB-net is confronted with SVM, graph regularized sparse linear discriminant analysis (GraphSLDA) [63], group sparse canonical correlation analysis (GSCCA) [64], DGCNN, and GCB. Experiments are regulated PSD features. For conveniently comparing, experiments adopted the LOO method the same as [41] and [50], and BLS is not implemented. The result is illustrated in Table IX and Fig. 6.

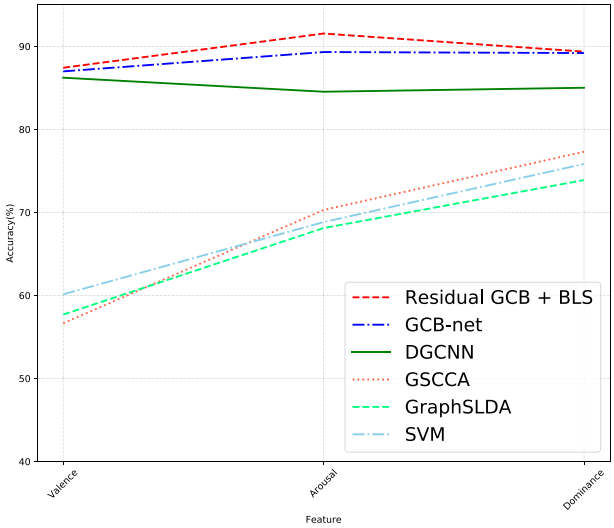

Fig. 6. Comparisons of the EEG Emotion Recognition Accuracy on on the DREAMER Data set among SVM, GraphSLDA, GSCCA, DGCNN, GCB-net, and Residual GCB-net.

From Table IX, it is distinct that all models acquire higher accuracies on Arousal and Dominance dimensions. It

TABLE VII
COMPARISON R GCB-NET EXPERIMENTS OF THE EEG EMOTION RECOGNITION ACCURACY/STAND DEVIATION (%) ON SEED AMONG 9-LAYER, 6-LAYER, 9-LAYER-ALL-CONCATENATE, 12-LAYER, AND 15-LAYER

Feature	Model	δ band	θ band	α band	β band	γ band	all($\delta, \theta, \alpha, \beta, \gamma$)
DE	**9-layer Residual GCB-net**	**80.88**/9.04	77.60/7.34	**81.95/10.29**	88.97/**8.50**	**90.06**/9.39	92.86/6.70
	6-layer Residual GCB-net	79.41/**8.90**	75.39/**7.28**	79.62/12.07	87.61/9.21	89.00/9.47	92.72/6.83
	9-layer-all-concatenate Residual GCB-net	79.59/9.06	**77.83**/8.65	80.26/11.25	87.58/8.88	88.56/8.93	91.75/8.13
	12-layer Residual GCB-net	78.50/10.08	75.45/6.45	78.86/12.49	**89.10**/8.70	88.57/**8.48**	**92.94**/6.58
	15-layer Residual GCB-net	80.39/9.39	76.20/7.91	80.78/10.76	88.65/10.31	89.48/10.07	92.78/**6.41**
PSD	**9-layer Residual GCB-net**	**73.95**/11.93	72.42/9.70	74.65/**11.49**	82.81/11.91	**84.12**/11.48	**84.90**/11.53
	6-layer Residual GCB-net	72.58/**11.42**	71.91/8.91	72.11/12.35	**82.91**/11.17	83.43/10.62	83.28/12.18
	9-layer-all-concatenate Residual GCB-net	72.45/12.19	71.72/9.19	**74.68**/14.08	82.01/**10.41**	82.39/**9.92**	83.25/13.06
	12-layer Residual GCB-net	73.32/12.55	71.50/**8.65**	72.24/13.23	81.84/10.98	83.43/10.36	83.08/10.82
	15-layer Residual GCB-net	73.35/12.30	**72.82**/9.47	67.50/13.12	77.09/12.63	77.57/12.56	77.37/**10.76**
DASM	**9-layer Residual GCB-net**	**65.48**/8.55	**66.17**/9.97	70.67/11.22	85.45/11.82	84.02/11.36	**80.13/12.74**
	6-layer Residual GCB-net	64.53/7.54	63.21/9.16	70.34/11.42	**86.18**/11.03	**86.18**/11.03	78.59/13.76
	9-layer-all-concatenate Residual GCB-net	64.09/**7.01**	64.74/10.14	68.96/11.75	84.53/11.57	83.27/11.05	78.11/14.04
	12-layer Residual GCB-net	65.20/7.38	64.00/**9.03**	69.50/12.75	84.05/**10.92**	85.16/12.19	78.31/13.62
	15-layer Residual GCB-net	64.61/8.09	65.71/10.02	**71.29/10.48**	85.91/11.11	85.59/**10.83**	78.32/12.78
RASM	**9-layer Residual GCB-net**	66.47/6.88	66.22/8.81	71.71/**9.47**	86.01/10.53	87.52/10.46	**88.98**/9.96
	6-layer Residual GCB-net	67.55/7.15	65.37/8.52	70.79/10.99	**87.68**/10.75	88.04/11.15	88.82/10.34
	9-layer-all-concatenate Residual GCB-net	67.62/7.03	68.24/8.67	72.72/10.45	72.31/10.36	**88.22/10.01**	88.57/10.27
	12-layer Residual GCB-net	**71.56**/9.61	**72.08/7.97**	**73.85**/9.76	83.59/11.21	85.44/11.27	85.05/**9.21**
	15-layer Residual GCB-net	56.83/**5.76**	62.83/15.04	59.93/9.99	61.88/10.10	61.88/**10.10**	81.41/13.06
DCAU	**9-layer Residual GCB-net**	72.65/**9.27**	71.98/10.43	**73.53**/9.24	**85.16**/11.08	**86.24**/12.08	**85.00/9.30**
	6-layer Residual GCB-net	71.47/9.97	71.56/**8.19**	73.52/9.69	84.58/9.46	85.43/11.27	84.85/10.20
	9-layer-all-concatenate Residual GCB-net	**72.69**/11.82	**72.38**/10.27	73.51/**8.82**	83.03/**11.06**	84.54/**11.16**	82.07/13.96
	12-layer Residual GCB-net	62.60/12.88	65.05/8.35	66.41/11.06	77.28/11.55	78.68/13.00	79.02/11.27
	15-layer Residual GCB-net	63.96/11.55	65.45/11.73	66.70/10.78	77.89/12.53	78.13/12.05	78.85/10.42

TABLE VIII
COMPARISON BLS EXPERIMENTS OF THE EEG EMOTION RECOGNITION ACCURACY/STAND DEVIATION (%) ON SEED AMONG BLS AND STACKED BLS

Feature	Model	δ band	θ band	α band	β band	γ band	all($\delta, \theta, \alpha, \beta, \gamma$)
DE	**R GCB-net + BLS**	81.20/7.97	79.85/9.70	84.03/9.93	90.70/7.28	91.70/7.41	94.56/6.61
	R GCB-net + stack BLS	79.71/8.54	78.89/9.87	82.59/9.94	90.22/7.35	90.72/9.89	93.59/7.47
PSD	**R GCB-net + BLS**	74.38/12.45	75.75/10.16	77.69/12.03	83.55/10.90	83.87/10.96	84.94/10.93
	R GCB-net + stack BLS	72.69/11.66	74.59/10.24	75.99/12.07	83.16/11.28	82.90/12.35	83.23/11.07
DASM	**R GCB-net + BLS**	66.51/9.41	66.40/10.28	72.12/10.44	86.46/11.45	88.06/10.21	82.25/12.58
	R GCB-net + stack BLS	65.54/9.00	65.42/10.76	72.36/10.77	86.30/11.17	86.90/10.51	80.49/12.93
RASM	**R GCB-net + BLS**	67.59/7.83	66.42/10.31	74.79/12.35	89.14/10.51	89.24/10.26	90.45/10.22
	R GCB-net + stack BLS	67.11/7.95	64.25/11.24	73.11/10.60	88.19/10.28	87.87/10.78	90.17/10.01
DCAU	**R GCB-net + BLS**	72.81/11.30	72.26/9.74	74.27/9.91	87.91/9.47	87.70/9.30	87.94/8.59
	R GCB-net + stack BLS	71.91/10.61	71.10/8.17	74.44/9.45	87.57/9.41	86.67/9.36	87.40/8.86

can suggest that those features have better performance on prediction. Residual GCB-net achieved the highest accuracy of 91.55%, 89.37%, and 87.43% on dimensions of arousal dominance and valence, respectively. However, it had poor performance on standard deviation (compared to [50]). It might indicate that the performance of Residual GCB-net is not steady on different films in DREAMER.

3) F1 Score and AUC Score Analysis: To further analyze the classification performance of Residual GCB-net, this section proposes the F1 score and AUC score as other evaluation methods. From Table VI, it demonstrates that the $F1$ score and AUC score result has a similar trend with that of accuracy. DE has a better $F1$ score than any other features and the AUC score of all features are close to 0.8. Since the DREAMER data set recorded only one emotion in each film cut, it is infeasible to calculate the AUC score on it. In order not to change the physical meaning of the corresponding features of each movie recorded in the data set, and for better

TABLE IX
COMPARISON OF THE EEG EMOTION RECOGNITION ACCURACY/STAND DEVIATION (%) ON THE DREAMER DATA SET AMONG SVM, GRAPHSLDA, GSCCA, DGCNN, GCB-NET, AND RESIDUAL GCB-NET

Model	Valence	Arousal	Dominance
SVM [41]	60.14/33.34	68.84/24.94	75.84/20.76
GraphSLDA [41]	57.70 / 13.89	68.12/17.53	73.90/15.85
GSCCA [41]	56.65/21.50	70.30/18.66	77.31/15.44
DGCNN [41]	86.23/12.29	84.54/10.18	85.02/10.25
GCB-net [50]	86.99/**6.21**	89.32/**5.01**	89.20/**4.33**
Residual GCB-net	**87.43**/14.89	**91.55**/14.78	**89.37**/16.78

comparison with previous work [41], [50], this article did not shuffle the subjects' features. The $F1$ score of valence, arousal, and dominance are 0.50, 0.43, and 0.52, respectively. This is consistent with the result of the larger standard deviation in the Table IX.

D. Comparison Experiments and Analysis

This section gives the comparison experiments of different layers of Residual GCB-net and types of BLS. Table VII describes the SEED experimental results of the Residual GCB-net with 9-layer, 6-layer, 9-layer-all concatenate, 12-layer, and 15-layer. Table VIII indicates the SEED experimental results of R GCB-net with BLS and stacked BLS [65]. The structure of stacked BLS is illustrated in Fig. 7.

Table VII demonstrates the trend of Residual GCB-net accuracy on SEED: it rises first then falls with the depth of residual block increasing. Due to the concatenation structure, the increase of residual blocks can result in the input of irrelevant noise. That could be the reason for the accuracy decrease. Considering the whole performance of the 9-layer, 6-layer, 12-layer, and 15-layer models, this article selects the 9-layer Residual GCB-net as the main model. Similar to the deeper models, the 9-layer-all concatenate model also has lower accuracy with more input. It might indicate that more output flattened and concatenated structure could not promote the performance of Residual GCB-net.

Table VIII indicates that the stacked BLS obtains a similar accuracy result based on the Residual GCB-net-extracted features. Stacked BLS utilizes n_{block} (the number of BLS blocks) BLS blocks and uses a residual short-cut connection, which could inherit the efficiency and effectiveness of BLS. From the result of Table VIII, the promotion of stacked BLS is limited but computation complexity is n_{block} times increased. In this article, it is proper to select the original BLS model. The results might indicate that the hyperparameter of stacked BLS should be adjusted to fit this data set. In addition, information might be limited due to the bottleneck of GCN, which prevents the advantages of stacked BLS from being effectively utilized.

According to Tables V, VII, and VIII, it is notable that both residual block and BLS promote the accuracy of GCB-net. For those features with a higher accuracy baseline (such as the all-feature of DE), BLS performs better, and for those with a lower accuracy baseline, Residual block promotes more. This may be due to the different performance of different features on emotion recognition tasks. BLS has better performance on accuracy improvement on better features, and Residual block could further explore those features.

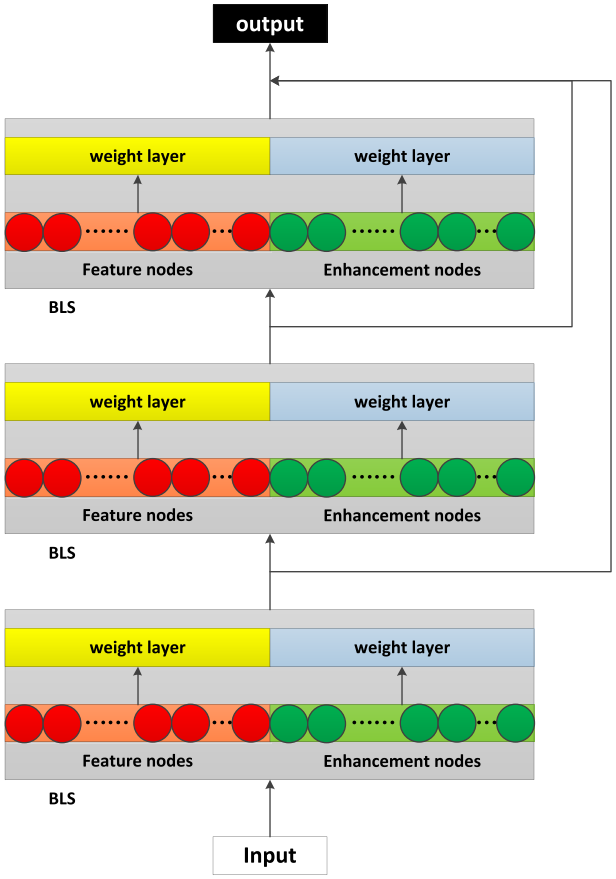

Fig. 7. Configuration of stacked BLS (stacked BLS). In this figure, the number of BLS blocks $n_{\text{block}} = 3$. The red nodes and green nodes denote the mapped features and enhancement nodes, respectively.

V. CONCLUSION AND DISCUSSION

In this article, a residual GCB-net (Residual GCB-net) was proposed to solve the EEG emotion recognition problem. It is designed to extract high-level features from graph-structure data. Residual GCB-net takes both deep and broad approaches, applying residual learning blocks to promote the performance on deeper networks and retaining the concatenation architecture to maintain performance on broad concepts. The residual learning block offers a shortcut connection to transform the problem "fitting an identity mapping" to "making the residual part close to 0," which solves the degradation problem. The residual learning block assures the accuracy will not decrease with the depth increasing, making it possible for the model to explore more high-level information. The Residual GCB-net model has experimented on SEED and DREAMER. Compared to SVM, DBN, GCNN, DGCNN, and GCB on SEED, Residual GCB-net + BLS obtained the best accuracy of 94.56% on the DE feature with the all-frequency band acquired the highest accuracy on all features with whole

bands. On DREAMER, it also obtained the highest accuracies as 91.55%, 89.37%, and 87.43% on dimensions of arousal dominance and Valence, respectively. The experimental results demonstrated the adoption of both broad and deep concept methods could obtain better results and promote the efficiency and classification capability. The accuracy promotion of Residual GCB-net might be due to the following major points.

1) Residual learning block solves the degradation problem of neural networks. It ensures that the network can be designed with sufficient layers to obtain deeper information, which remedies the defect that the GCN network cannot acquire enough global information.
2) This model uses a combination of DL and broad learning. It not only extracts the high-level features but preserves the multilevel features of all hierarchical layers by the output of GCN and each residual block.

It is worth noting that with the increase in depth, Residual GCB-net hardly improves the accuracy. It might indicate that GCN has the bottleneck that it could only aggregate information within two hops. A skip connection might be utilized to GCN to ensure it aggregates further hops information. It could attempt to balance the computation complexity and the max k-hops information in future work.

In addition, from Tables V and IX, it might indicate that the DE feature has a better performance on EEG emotion recognition. It is worth substituting the PSD feature with the DE feature on DREAMER in future work.

ACKNOWLEDGMENT

The authors would like to appreciate the support of the Data Science and Artificial Intelligence Laboratory (D-Sail), in South China University of Technology, Guangzhou, China. The group of Text and Emotion Recognition Analysis projects provides the necessary equipment and experimental data. They also thank the support of funding of the National Key Research and Development Program of China, the National Natural Science Foundation of China grant, the Guangdong Natural Science Funds for Distinguished Young Scholar, the Science and Technology Major Project of Guangzhou, the Program for Guangdong Introducing Innovative and Entrepreneurial Teams and Guangdong-Hong Kong-Macao Greater Bay Area Center for Brain Science and Brain-Inspired Intelligence Fund.

REFERENCES

[1] R. Cowie *et al.*, "Emotion recognition in human-computer interaction," *IEEE Signal Process. Mag.*, vol. 18, no. 1, pp. 32–80, Jan. 2001.
[2] Z. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, and G. Chen, "A channel-fused dense convolutional network for EEG-based emotion recognition," *IEEE Trans. Cogn. Develop. Syst.*, vol. 13, no. 4, pp. 945–954, Dec. 2021.
[3] T. Zhang, Z.-L. Liu, X.-H. Wang, X.-F. Xing, C. P. Chen, and E. Chen, "Facial expression recognition via broad learning system," in *Proc. IEEE Int. Conf. Syst. Man Cybern. (SMC)*, 2018, pp. 1898–1902.
[4] J. Van den Stock, R. Righart, and B. De Gelder, "Body expressions influence recognition of emotions in the face and voice," *Emotion*, vol. 7, no. 3, pp. 487–494, 2007.
[5] K. Wang, N. An, B. N. Li, Y. Zhang, and L. Li, "Speech emotion recognition using Fourier parameters," *IEEE Trans. Affect. Comput.*, vol. 6, no. 1, pp. 69–75, Jan.-Mar. 2015.
[6] X. Kang, F. Ren, and Y. Wu, "Exploring latent semantic information for textual emotion recognition in blog articles," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 1, pp. 204–216, Jan. 2018.
[7] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 102–115, Jan.-Mar. 2012.
[8] B. Cheng and G. Liu, "Emotion recognition from surface EMG signal using wavelet transform and neural network," in *Proc. 2nd Int. Conf. Bioinformat. Biomed. Eng. (ICBBE)*, 2008, pp. 1363–1366.
[9] S. Koelstra *et al.*, "Deap: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan.-Mar. 2012.
[10] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, Mar. 2010.
[11] 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 Trans. Biomed. Eng.*, vol. 62, no. 1, pp. 256–263, Jan. 2015.
[12] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016," *IEEE Trans. Cogn. Develop. Syst.*, early access, Jul. 7, 2020, doi: 10.1109/TCDS.2020.3007453.
[13] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Trans. Cogn. Develop. Syst.*, vol. 12, no. 2, pp. 344–353, Jun. 2020.
[14] Y. Liu, Z. Li, T. Zhang, and S. Zhao, "Brain–Robot interface-based navigation control of a mobile robot in corridor environments," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 8, pp. 3047–3058, Aug. 2020.
[15] V. Khurana *et al.*, "A survey on neuromarketing using EEG signals," *IEEE Trans. Cogn. Develop. Syst.*, vol. 13, no. 4, pp. 732–749, Dec. 2021.
[16] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography–Theory, instrumentation, and applications to noninvasive studies of the working human brain," *Rev. Mod. Phys.*, vol. 65, no. 2, p. 413, 1993.
[17] D. P. Subha, P. K. Joseph, R. Acharya, and C. M. Lim, "EEG signal analysis: A survey," *J. Med. Syst.*, vol. 34, no. 2, pp. 195–212, 2010.
[18] R. Srinivasan, "Methods to improve the spatial resolution of EEG," *Int. J. Bioelectromagn.*, vol. 1, no. 1, pp. 102–111, 1999.
[19] S. E. Kober, J. Kurzmann, and C. Neuper, "Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: An EEG study," *Int. J. Psychophysiol.*, vol. 83, no. 3, pp. 365–374, 2012.
[20] J. Pan, C. Li, Y. Tang, W. Li, and X. Li, "Energy consumption prediction of a CNC machining process with incomplete data," *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 5, pp. 987–1000, May 2021.
[21] D. Wu, X. Luo, M. Shang, Y. He, G. Wang, and X. Wu, "A data-characteristic-aware latent factor model for Web services QoS prediction," *IEEE Trans. Knowl. Data Eng.*, early access, Aug. 5, 2020, doi: 10.1109/TKDE.2020.3014302.
[22] X. Luo, M. Zhou, S. Li, D. Wu, Z. Liu, and M. Shang, "Algorithms of unconstrained non-negative latent factor analysis for recommender systems," *IEEE Trans. Big Data*, vol. 7, no. 1, pp. 227–240, Mar. 2021.
[23] X. Luo, Z. Liu, S. Li, M. Shang, and Z. Wang, "A fast non-negative latent factor model based on generalized momentum method," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 610–620, Jan. 2021.
[24] D. Wu and X. Luo, "Robust latent factor analysis for precise representation of high-dimensional and sparse data," *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 4, pp. 796–805, Apr. 2021.
[25] A.-C. Conneau and S. Essid, "Assessment of new spectral features for EEG-based emotion recognition," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2014, pp. 4698–4702.
[26] E. L. Van den Broek, "Ubiquitous emotion-aware computing," *Pers. Ubiquitous Comput.*, vol. 17, no. 1, pp. 53–67, 2013.
[27] L. I. Aftanas, A. A. Varlamov, S. V. Pavlov, V. P. Makhnev, and N. V. Reva, "Time-dependent cortical asymmetries induced by emotional arousal: EEG analysis of event-related synchronization and desynchronization in individually defined frequency bands," *Int. J. Psychophysiol.*, vol. 44, no. 1, pp. 67–82, 2002.
[28] E. Q. Wu, M. Zhou, P. Xiong, Z.-R. Tang, R. Hu, and Y.-W. Jie, "Inferring flight performance under different maneuvers with pilot's multi-physiological parameters," *IEEE Trans. Intell. Transp. Syst.*, early access, Sep. 3, 2021, doi: 10.1109/TITS.2021.3103068.
[29] M. M. Rahman *et al.*, "Recognition of human emotions using EEG signals: A review," *Comput. Biol. Med.*, vol. 136, Art. no. 104696, Sep. 2021.

[30] S. K. Prabhakar and H. Rajaguru, "Conceptual analysis of epilepsy classification using probabilistic mixture models," in *Proc. 5th Int. Winter Conf. Brain-Comput. Interface (BCI)*, 2017, pp. 81–84.
[31] S. Sengupta, D. Chanda, A. Mitra, and S. Dutta, "Computer aided technique for epilepsy classification using cross wavelet transform and RBF-Kernel based support vector machine," in *Proc. 2nd Int. Conf. Next Generat. Comput. Technol. (NGCT)*, 2016, pp. 501–505.
[32] L. Wang, X. Long, J. B. Arends, and R. M. Aarts, "EEG analysis of seizure patterns using visibility graphs for detection of generalized seizures," *J. Neurosci. Methods*, vol. 290, pp. 85–94, Oct. 2017.
[33] M. Murugappan, R. Nagarajan, and S. Yaacob, "Comparison of different wavelet features from EEG signals for classifying human emotions," in *Proc. IEEE Symp. Ind. Electron. Appl.*, vol. 2, 2009, pp. 836–841.
[34] S. Wang, Y. Li, P. Wen, and G. Zhu, "Analyzing EEG signals using graph entropy based principle component analysis and J48 decision tree," in *Proc. 6th Int. Conf. Signal Process. Syst. (ICSPS)*, 2014, pp. 1–6.
[35] S. Bhattacharyya, A. Konar, D. Tibarewala, A. Khasnobish, and R. Janarthanan, "Performance analysis of ensemble methods for multiclass classification of motor imagery EEG signal," in *Proc. Int. Conf. Control Instrum. Energy Commun. (CIEC)*, 2014, pp. 712–716.
[36] F. Bahari and A. Janghorbani, "EEG-based emotion recognition using recurrence plot analysis and k nearest neighbor classifier," in *Proc. 20th Iran. Conf. Biomed. Eng. (ICBME)*, 2013, pp. 228–233.
[37] A. Subasi and E. Ercelebi, "Classification of EEG signals using neural network and logistic regression," *Comput. Methods Programs Biomed.*, vol. 78, no. 2, pp. 87–99, 2005.
[38] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
[39] J. Cao, J. Zhu, W. Hu, and A. Kummert, "Epileptic signal classification with deep EEG features by stacked CNNs," *IEEE Trans. Cogn. Develop. Syst.*, vol. 12, no. 4, pp. 709–722, Dec. 2020.
[40] D. Zeng, K. Huang, C. Xu, H. Shen, and Z. Chen, "Hierarchy graph convolution network and tree classification for epileptic detection on electroencephalography signals," *IEEE Trans. Cogn. Develop. Syst.*, vol. 13, no. 4, pp. 955–968, Dec. 2021.
[41] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul.-Sep. 2020.
[42] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
[43] X. Liu, M. Yan, L. Deng, G. Li, X. Ye, and D. Fan, "Sampling methods for efficient training of graph convolutional networks: A survey," 2021, *arXiv:2103.05872*.
[44] X. Hong, T. Zhang, Z. Cui, and J. Yang, "Variational gridded graph convolution network for node classification," *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 10, pp. 1697–1708, Oct. 2021.
[45] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1025–1035.
[46] C. L. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.
[47] C. P. Chen, Z. Liu, and S. Feng, "Universal approximation capability of broad learning system and its structural variations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1191–1204, Apr. 2019.
[48] X. Gong, T. Zhang, C. L. P. Chen, and Z. Liu, "Research review for broad learning system: Algorithms, theory, and applications," *IEEE Trans. Cybern.*, early access, Mar. 17, 2021, doi: 10.1109/TCYB.2021.3061094.
[49] X.-H. Wang, T. Zhang, X.-M. Xu, L. Chen, X.-F. Xing, and C. P. Chen, "EEG emotion recognition using dynamical graph convolutional neural networks and broad learning system," in *Proc. IEEE Int. Conf. Bioinformat. Biomed. (BIBM)*, 2018, pp. 1240–1244.
[50] T. Zhang, X. Wang, X. Xu, and C. P. Chen, "GCB-Net: Graph convolutional broad network and its application in emotion recognition," *IEEE Trans. Affect. Comput.*, early access, Aug. 27, 2019, doi: 10.1109/TAFFC.2019.2937768.
[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[52] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
[53] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 3837–3845.
[54] T. Zhang, X. Gong, and C. L. P. Chen, "BMT-NET: Broad multitask transformer network for sentiment analysis," *IEEE Trans. Cybern.*, early access, Mar. 4, 2021, doi: 10.1109/TCYB.2021.3050508.
[55] X.-R. Gong, J.-X. Jin, and T. Zhang, "Sentiment analysis using autoregressive language modeling and broad learning system," in *Proc. IEEE Int. Conf. Bioinformat. Biomed. (BIBM)*, 2019, pp. 1130–1134.
[56] M. Xu, M. Han, C. L. P. Chen, and T. Qiu, "Recurrent broad learning systems for time series prediction," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1405–1417, Apr. 2020.
[57] 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 J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 98–107, Jan. 2018.
[58] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.
[59] Y. Yang, Q. J. Wu, W.-L. Zheng, and B.-L. Lu, "EEG-based emotion recognition using hierarchical network with subnetwork nodes," *IEEE Trans. Cogn. Develop. Syst.*, vol. 10, no. 2, pp. 408–419, Jun. 2018.
[60] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain adaptation techniques for EEG-based emotion recognition: A comparative study on two public datasets," *IEEE Trans. Cogn. Develop. Syst.*, vol. 11, no. 1, pp. 85–94, Mar. 2019.
[61] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, 2013, pp. 81–84.
[62] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.
[63] Y. Li, W. Zheng, Z. Cui, and X. Zhou, "A novel graph regularized sparse linear discriminant analysis model for EEG emotion recognition," in *Proc. Int. Conf. Neural Inf. Process.*, 2016, pp. 175–182.
[64] W. Zheng, "Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis," *IEEE Trans. Cogn. Develop. Syst.*, vol. 9, no. 3, pp. 281–290, Sep. 2017.
[65] Z. Liu, C. L. P. Chen, S. Feng, Q. Feng, and T. Zhang, "Stacked broad learning system: From incremental flatted structure to deep model," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 209–222, Jan. 2021.

Qilin Li (Student Member, IEEE) received the B.Eng. degree in electrical engineering and automation from Wuhan University of Science and Technology, Wuhan, China, in 2018, and the M.S. degree in computer science from the University of Wollongong, Wollongong, NSW, Australia, in 2020.

He is currently a Research Assistant with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China, also with Pazhou Lab, Guangzhou. His current research interests include emotion recognition, broad learning system, and natural language processing.

Tong Zhang (Member, IEEE) received the B.S. degree in software engineering from Sun Yat-sen University, Guangzhou, China, in 2009, and the M.S. degree in applied mathematics and the Ph.D. degree in software engineering from the University of Macau, Macau, China, in 2011 and 2016, respectively.

He is currently a Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, and also with Pazhou Lab, Guangzhou. His research interests include affective computing, evolutionary computation, neural network, and other machine learning techniques and their applications.

Prof. Zhang has been working in publication matters for many IEEE conferences.

C. L. Philip Chen (Fellow, IEEE) received the M.S. degree in electrical and computer science from the University of Michigan at Ann Arbor, Ann Arbor, MI, USA, in 1985, and the Ph.D. degree in electrical and computer science from Purdue University, West Lafayette, IN, USA, in 1988.

He is the Chair Professor and the Dean of the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China, and also with Pazhou Lab, Guangzhou. Being a Program Evaluator of the Accreditation Board of Engineering and Technology Education in the U.S., for computer engineering, electrical engineering, and software engineering programs, he successfully architects the University of Macau's Engineering and Computer Science programs receiving accreditations from Washington/Seoul Accord through Hong Kong Institute of Engineers (HKIE), Hong Kong, of which is considered as his utmost contribution in engineering/computer science education for Macau as the former Dean of the Faculty of Science and Technology. His current research interests include cybernetics, systems, and computational intelligence.

Dr. Chen was a recipient of the 2016 Outstanding Electrical and Computer Engineers Award from his alma mater, Purdue University in 1988. He received the IEEE Norbert Wiener Award in 2018 for his contribution in systems and cybernetics, and machine learning. He is also a highly cited researcher by Clarivate Analytics in 2018, 2019, and 2020. He is currently the Editor-in-Chief of the IEEE TRANSACTIONS ON CYBERNETICS and an Associate Editor of the IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE and IEEE TRANSACTIONS ON FUZZY SYSTEMS. He was the IEEE Systems, Man, and Cybernetics Society President from 2012 to 2013, and the Editor-in-Chief of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS from 2014 to 2019. He was the Chair of TC 9.1 Economic and Business Systems of International Federation of Automatic Control from 2015 to 2017. He is a Fellow of AAAS, IAPR, CAA, and HKIE; and a member of Academia Europaea, European Academy of Sciences and Arts.

Ke Yi is currently pursuing the bachelor's degree with the South China University of Technology, Guangzhou, China, and also with Pazhou Lab, Guangzhou.

His research interests are deep learning, convolutional neural networks, and hardware development.

Long Chen (Member, IEEE) received the M.S. degree in computer engineering from the University of Alberta, Edmonton, AB, Canada, in 2005, and the Ph.D. degree in electrical engineering from the University of Texas at San Antonio, San Antonio, TX, USA, in 2010.

From 2010 to 2011, he was a Postdoctoral Fellow with the University of Texas at San Antonio. He is currently an Associate Professor with the Department of Computer and Information Science, University of Macau, Macau, China. His current research interests include computational intelligence techniques and their applications.

Dr. Chen has been working in publication matters for several IEEE conferences. He was the Publications Co-Chair of the IEEE International Conference on Systems, Man and Cybernetics in 2009, 2012, and 2014.