SparseDGCNN: Recognizing Emotion From Multichannel EEG Signals

Guanhua Zhang, Minjing Yu, Yong-Jin Liu, *Senior Member, IEEE*, Guozhen Zhao, Dan Zhang, *Member, IEEE*, and Wenming Zheng, *Senior Member, IEEE*

Abstract—Emotion recognition from EEG signals has attracted much attention in affective computing. Recently, a novel dynamic graph convolutional neural network (DGCNN) model was proposed, which simultaneously optimized the network parameters and a weighted graph G characterizing the strength of functional relation between each pair of two electrodes in the EEG recording equipment. In this article, we propose a sparse DGCNN model which modifies DGCNN by imposing a sparseness constraint on G and improves the emotion recognition performance. Our work is based on an important observation: the tomography study reveals that different brain regions sampled by EEG electrodes may be related to different functions of the brain and then the functional relations among electrodes are possibly highly localized and sparse. However, introducing sparseness constraint into the graph G makes the loss function of sparse DGCNN non-differentiable at some singular points. To ensure that the training process of sparse DGCNN converges, we apply the forward-backward splitting method. To evaluate the performance of sparse DGCNN, we compare it with four representative recognition methods (SVM, DBN, GELM and DGCNN). In addition to comparing different recognition methods, our experiments also compare different features and spectral bands, including EEG features in time-frequency domain (DE, PSD, DASM, RASM, ASM and DCAU on different bands) extracted from four representative EEG datasets (SEED, DEAP, DREAMER, and CMEED). The results show that (1) sparse DGCNN has consistently better accuracy than representative methods and has a good scalability, and (2) DE, PSD, and ASM features on γ band convey most discriminative emotional information, and fusion of separate features and frequency bands can improve recognition performance.

Index Terms—Emotion recognition, multichannel EEG signals, graph convolutional neural network, sparse constraints

1 Introduction

Emotion is a human experience related to a particular pattern of physiological activity, which can be characterized by a flexible adaptation mechanism [1]. Recognizing human emotional states from their behavioral or physiological signals plays an important role in affective computing and human-machine interaction. Compared with behavioral signals (such as facial expression, vocal intonation, gesture and body posture), physiological signals are spontaneous and very difficult to conceal; therefore, they provide a direct and comprehensive means for emotion recognition [2], [3].

The human body produces various physiological signals, including brain electrical activity (electroencephalogram or EEG), heart rate change (electrocardiogram or ECG), muscle current (electromyogram or EMG), respiratory rate (capnogram), skin conductance and galvanic skin response, etc [4]. Among various types of physiological signals, EEG signals have received much attention recently [5], [6], [7], due to the popularization of new wireless headsets (e.g., Emotiv) which are portable, cost-effective, easy to use with increased practicability and less physical restriction.

To train and evaluate an emotion recognition system based on EEG signals, datasets with ground truth labels are required. In these datasets, to correctly label the EEG signals, standardised emotional stimuli (e.g., visual or auditory stimuli) are used to elicit target emotions. Some early datasets include the affective norms for English words and text [8], the international affective digitised sound system [9], the international affective picture system [10] and the Geneva affective picture database [11]. Recently proposed datasets extend traditional materials (such as word, text, picture and sound) to combine visual and auditory stimuli, which usually take the forms of film clip or music video [3], [12], [13]. Compared with static photos and slides, the presence of these combined stimuli can better capture real-life emotional experiences [14]. In this paper, we select four representative and publicly available datasets, i.e., DEAP [4], SEED [15], DREAMER [16] and CMEED [3], [17], [18], to evaluate our proposed emotion recognition method: in all of these datasets, EEG signals were collected and labeled by eliciting target emotions from watching standardized music videos or film clips.

Two different models can be used to characterize emotions in EEG data: the discrete model and dimensional model. The

- *Guanhua Zhang and Yong-Jin Liu are with the BNRist, MOE Key Laboratory of Pervasive Computing, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China. E-mail: zgh17@mails.tsinghua.edu.cn, liuyongjin@tsinghua.edu.cn.*
- *Minjing Yu is with the College of Intelligence and Computing, Tianjin University, Tianjin 300072, China. E-mail: minjingyu@tju.edu.cn.*
- *Guozhen Zhao is with the CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing 100101, P. R. China. E-mail: zhaogz@psych.ac.cn.*
- *Dan Zhang is with the Department of Psychology, Tsinghua University, Beijing 100084, P.R. China. E-mail: dzhang@tsinghua.edu.cn.*
- *Wenming Zheng is with the MOE Key Laboratory of Child Development and Learning Science, Southeast University, Nanjing 211100, China. E-mail: wenming_zheng@seu.edu.cn.*

Manuscript received 12 Apr. 2020; revised 28 Dec. 2020; accepted 9 Jan. 2021. Date of publication 13 Jan. 2021; date of current version 28 Feb. 2023. (Corresponding authors: Yong-Jin Liu and Wenming Zheng.) Recommended for acceptance by M. Soleymani. Digital Object Identifier no. 10.1109/TAFFC.2021.3051332

discrete model represents the emotional space as a limited number of basic emotions. For example, Ekman [19] proposed six universal emotions (joy, sadness, surprise, fear, anger, and disgust) and Plutchik [20] proposed eight discrete emotions by adding two more (curiosity and acceptance). The dimensional model represents the emotion as a two- or three-dimensional space, e.g., the three dimensions – valence, arousal and dominance – are widely used. Valence means the intrinsic attractiveness/goodness (positive valence) or averseness/badness (negative valence). Arousal reflects the mental vigilance level of emotion and the intensity of physiological activation that an individual feels. Dominance refers to an individual's status, i.e., in control or being controlled. Nowadays, the most commonly used model is the Circumplex Model of Affect, which only uses valence and arousal [6], [21].

Emotion recognition from EEG signals relies on discriminative EEG features. EEG signals are discrete time series and spatial, spectral and temporal EEG features that are consistently linked to cognitive processes [22] exist. In the time domain, some widely used statistical information such as entropy, the fractal dimension and higher order crossings can be used as EEG features [23], [24]. In the frequency domain, EEG signals are decomposed into several frequency ranges, each of which is prominent in certain brain activity; e.g., δ band (1-3 Hz), θ band (4-7 Hz), α band (8-13 Hz), β band (14-30 Hz) and γ band (>30 Hz) [25]. From each frequency band, some widely used features can be extracted, including the power spectral density (PSD) feature, the differential entropy (DE) feature, and the features of differential caudality (DCAU), asymmetry (ASM), differential asymmetry (DASM) and rational asymmmetry (RASM), etc [2], [4], [15], [26]. Compared with time domain features, frequency-based features are more recognized for emotion recognition. Some neuroscience studies revealed that emotion-related neural information mainly lies in higher frequency bands [27], [28], [29], but time domain features use information from all frequency bands.

The EEG is a measurement of brain electrical fields via electrodes placed on the scalp and it usually has sufficient density (>30 electrodes) to build topographical maps. So far the functional features that study the relationship between different EEG channels/electrodes are rarely considered in literature. Recently, Li *et al.* [30] proposed an adversarial neural network model to learn the discriminative emotional features for each of the left and right hemispheres. But their work did not investigate relationships between channels. Song *et al.* [31] built a weighted graph $\mathcal{G}$ which represents the connections between multiple EEG channels and proposed a *dynamic graph convolutional neural network* (DGCNN) model to automatically learn an optimal set of weights in $\mathcal{G}$. Each node in the graph $\mathcal{G}$ corresponds to an EEG channel and is represented by a scalar (for individual features) or a vector (for fusion features), and then the EEG data can be regarded as graph signals, i.e., signals defined on an irregular graph $\mathcal{G}$. The connectivity of nodes in $\mathcal{G}$ is predefined by the spatial relations of EEG channels. By learning an optimal set of weights defined on each edge in $\mathcal{G}$, the connection strength between nodes can be determined as a functional feature. To analyze graph signals, the techniques of signal processing on graphs (in particular, spectral graph filtering or graph convolution techniques) are needed [32], which we briefly summarized in Section 2.1.

To handle the signals on irregular graph structures such as those in social networks and brain connectomes, graph CNN (GCNN) was first proposed in [33] based on either a hierarchical clustering of the domain or the spectrum of the graph Laplacian. The GCNN was further improved [34] by introducing a fast localized spectral filter for graph convolution [35]. Some other variants of GCNN include [36], [37]. Song *et al.* [31] introduced the GCNN into emotion recognition using EEGs from multiple channels and proposed a DGCNN model. DGCNN considers the edge weights on the irregular graph and automatically learns an optimal set of such weights. These weights provide an effective way to reveal the intrinsic relations between EEG channels: the larger weight between the ith and jth electrodes, the more correlation at these two nodes in the graph signals of EEG. We briefly summarize DGCNN in Section 2.2.

Our work presented in this paper is based on an important observation: the DGCNN optimizes the weights of the graph $\mathcal{G}$ in an unconstrained way, while it is well known from tomography that different brain regions sampled by EEG electrodes may be related to different brain functions and then the weights (representing functional relations among electrodes) are possibly highly localized and sparse. In this paper, we improve the DGCNN by introducing a new sparseness constraint into the graph representation $\mathcal{G}$ and propose a solution to a sparseness-constrained minimization problem to ensure the convergence of the network model. We call our method *sparse DGCNN*. Experimental results demonstrate that compared with existing emotion recognition methods, the sparse DGCNN consistently achieves better performance based on different EEG features in time-frequency domain extracted from four datasets, DEAP [4], SEED [15], DREAMER [16] and CMEED [3]), averagely improving the accuracy by 8.88 percent on four datasets. Our results also show that DE, PSD and ASM features on γ band convey most important discriminative emotional information, which are consistent with previous findings.

2 Preliminaries

Our method is based on the spectral graph theory [32], [35], [38], [39], [40] and DGCNN [31]. We briefly introduce them before presenting our method.

2.1 Spectral Graph Theory

We are interested in analyzing multichannel EEG signals defined on an undirected and weighted graph $G = \{V, W\}$. $V = \{v_1, v_2, \ldots, v_n\}$ is the vertex set, where each vertex v_i corresponds to an electrode and n is the number of electrodes in the EEG recording equipment. The $n \times n$ matrix W is the adjacency matrix of G, whose entries $w_{ij} \geq 0$ measure the strength of functional relation between v_i and v_j. DGCNN [31] automatically learns an optimal adjacency matrix W from a training set.

The spectral graph theory generalizes the classical signal processing techniques to the graph spectral domain, which incorporates the irregular graph structure when processing signals on graphs. The Laplacian matrix L of the graph G, which plays a central role in the special graph theory, is defined as $L = D - W$, where D is an $n \times n$ diagonal matrix with entries $D_{ii} = \sum_{j=1}^{n} w_{ij}$. The Fourier basis U of the graph

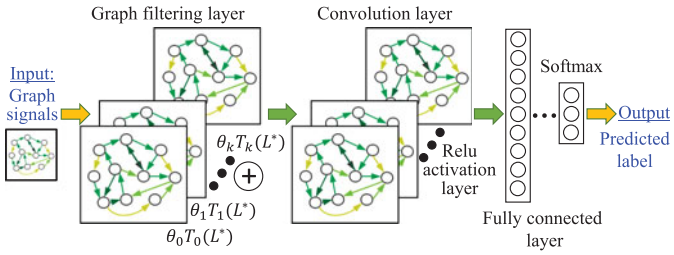


Fig. 1. The DGCNN model [31].

G can be represented as an orthonormal matrix, obtained from the singular value decomposition of Laplacian matrix $L = U\Lambda U^T$, where $\Lambda = diag([\lambda_1, \lambda_2, \ldots, \lambda_n])$ is a diagonal matrix. Then the graph Fourier transform and its inverse can be expressed as $\hat{x} = U^T x$ and $x = U\hat{x}$. The graph convolution operator is defined in the graph spectral domain as

$$x * y = U\big[(U^T x) \odot (U^T y)\big], \tag{1}$$

where $\odot$ is the element-wise Hadamard product. It follows that a filtering function g can be designed as a diagonal matrix

$$g(\Lambda) = diag([\theta_1, \theta_2, \ldots, \theta_n]), \tag{2}$$

where $\{\theta_i\}_{i=1}^n$ is a vector of Fourier coefficients, such that a signal x filtered by $g(L)$ can be expressed as

$$y = g(L)x = g(U\Lambda U^T)x = Ug(\Lambda)U^T x. \tag{3}$$

The above filter design can be explained by that the filtered signal y equals to the graph convolution of the signals x and $Ug(\Lambda)$

$$\begin{aligned} y = g(L)x = Ug(\Lambda)U^T x = [Ug(\Lambda)] \odot (U^T x) \\ = U\{U^T[Ug(\Lambda)]\} \odot (U^T x) = x * [Ug(\Lambda)]. \end{aligned} \tag{4}$$

2.2 DGCNN

The input of DGCNN is a graph signal $G = \{V, W\}$, in which W is the adjacent matrix whose entries $w_{ij} \geq 0$ measure the strength of functional relation between every pair (v_i, v_j), $i, j \in \{1, 2, \ldots, n\}$, and the signal at each vertex $v_i \in V$ is an EEG feature f extracted from the corresponding electrode. In [31], each of the features PSD, DE, DASM, RASM and DCAU in five frequency bands (δ, θ, α, β and γ bands) is tested.

As illustrated in Fig. 1, the DGCNN model consists of a graph filtering layer, a convolution layer, a ReLu activation layer and a fully connected layer.

The graph filtering layer transfers the irregular graph signal into the frequency domain. However, the diagonal form of the filtering function g in Eq. (2) is not localized and then its computation is extremely time-consuming. DGCNN uses a polynomial filter proposed by Defferrard *et al.* [34], which is approximated by the Chebyshev expansion

$$g(\Lambda) = \sum_{k=0}^{K-1} \theta'_k \Lambda^k \simeq \sum_{k=0}^{K-1} \theta_k T_k(\widetilde{\Lambda}), \tag{5}$$

where θ'_k are polynomial coefficients, θ_k are Chebyshev coefficients and $T_k(\widetilde{\Lambda})$ is the Chebyshev polynomial of order k evaluated at $\widetilde{\Lambda} = 2\Lambda/\lambda_{\max} - I_n$, $\lambda_{\max}$ is the largest element among the diagonal entries of Λ and I_n is the $n \times n$ identity matrix.

Akin to the traditional CNN, the convolution layer detects possibly specific patterns in the frequency domain. Then the ReLu activation functions [41] are applied to realize the nonlinear mapping capability. The output of the ReLu activation layer is non-negative. The final fully connected layer uses a softmax function to predict desired class label information.

In addition to optimizing the network parameters like GCNN did [34], DGCNN simultaneously learns an optimal adjacent matrix W in the training process. To achieve this goal, DGCNN uses the following loss function and computes its partial derivatives with respect to both network parameters Θ and the matrix W

$$Loss = \psi(l, l^p) + \alpha\|\Theta\|_2, \tag{6}$$

where $\psi(l, l^p)$ is the average cross entropy of the ground truth labels l with respect to the predicted labels l^p, α is a regularization weight and $\|\cdot\|_2$ is the L_2 norm. DGCNN used the standard back propagation method to iteratively update the Θ and W.

3 The Sparse DGCNN Model

To date, the majority of the EEG-based emotion recognition studies have utilized the EEG features at individual electrodes (e.g., [3], [6]). In these methods, the underlying assumption is that each individual emotion experience is associated with a circumscribed set of cortical and subcortical brain regions and distinct brain regions are involved for different emotions, constituting the basis for recognition.

Recent advances in network neuroscience are bringing changes in our understanding of human emotion. Briefly speaking, network neuroscientists emphasize the importance of the neural connectivity rather than the neural response as a better description of neural mechanism underlying human cognitive functions [42]. As emotion is interlocked with perception, cognition, motivation and action, taking a network-based perspective is neurophysiological plausible and has gained increasing attention recently [43].

In line with the neuroscience advances, researchers in the field of affective computing are starting to explore the network-based features for emotion recognition, e.g., the DGCNN model proposed in [31]. However, the DGCNN model optimizes the adjacent matrix W in an unconstrained way and has not fully utilized the neurophysiological properties of the human neural network. Our key observation is that W has a high capacity involving $n \times n$ variables and we can reduce the variance by incorporating some specific prior knowledge. In this paper, the prior knowledge we considered is the sparsity of the matrix W: since W encodes the strength of functional relation between every pair (v_i, v_j), $v_i, v_j \in V$, the non-zero entries w_{ij} in W should be as sparse as possible.

The above sparse assumption of W is supported by some neuroscience research. It is well known from tomography that different brain regions may be related to different functions of the brain. As illustrated in Fig. 2, the scalp electrode placements are labelled according to different brain areas: the frontal lobe (F), the central lobe (C), the temporal lobe (T), the posterior lobe (P) and the occipital lobe (O). Some brain functions may

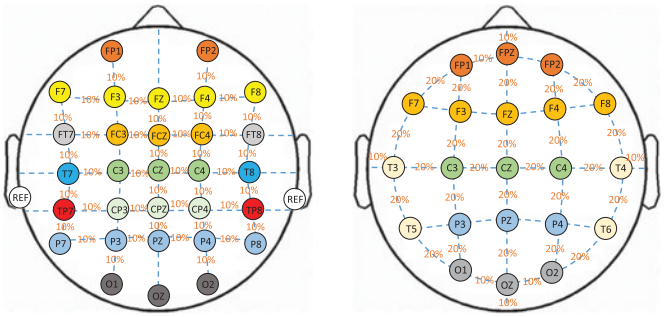


Fig. 2. The 10-20 electrode placement system of a 32-electrode Neuroscan quik-cap (left) and a 21-electrode EEG cap (right). The electrode placements are labelled with the frontal lobe (F), the central lobe (C), the temporal lobe (T), the posterior lobe (P) and the occipital lobe (O). Odd numbers/even numbers means the left/right side of the head. in the International 10-20 system, the 10 or 20 refers to the actual distances between adjacent electrodes being either 10 or 20 percent of the total front–back or right–left distance of the skull.

only activate a restricted brain area; e.g., emotional processors possibly locate at T3 and T4 [44]. Therefore, the functional relations among entries w_{ij} are possibly highly localized and accordingly the matrix W should be very sparse. Furthermore, brain networks usually exhibit a range of complex properties for efficient information processing, including the presence of hubs (highly interconnected nodes), small-world topology (dense local clustering with sparse long-range connections), etc [45]. These neuroscience findings thus call for the implementation of sparsity in the graph-based machine learning algorithms. Indeed, sparsity has been considered and proved to be an important factor for implementing reliable and high-performance EEG processing algorithms for general purpose brain-computer interfaces (e.g., [46]).

To impose sparsity on the matrix W, we expect that the non-zero entries in W should be as few as possible. To achieve this goal, it is well known in sparse coding research that we can use the L_1 or L_0 matrix norm. In our study, we choose the L_1 matrix norm due to its convexity property and formulate the following regularization term

$$\|W\|_1 = \sum_{i=1}^{n}\sum_{j=1}^{n}|w_{ij}|. \tag{7}$$

Then by introducing the sparseness constraint in Eq. (7) into the DGCNN loss function in Eq. (6), the loss function $\mathcal{L}$ in sparse DGCNN becomes

$$\mathcal{L}(\Theta, W) = \psi(l, l^p) + \alpha\|\Theta\|_2 + \lambda\|W\|_1, \tag{8}$$

where λ is the weight of the sparse constraint. The predicted labels l^p depends on the input graph signals, the dynamically adjusted matrix W and the network parameters Θ; therefore, it is a function $l^p(\Theta, W)$.

If we apply the same back propagation method in DGCNN to minimize the loss function $\mathcal{L}$, we need to compute both $\frac{\partial\mathcal{L}}{\partial\Theta}$ and $\frac{\partial\mathcal{L}}{\partial W}$, where $\frac{\partial\mathcal{L}}{\partial\Theta} = \frac{\partial Loss}{\partial\Theta}$, which is the same in DGCNN. However, different from DGCNN,

$$\frac{\partial\mathcal{L}}{\partial W} = \frac{\partial\psi(l, l^p)}{\partial W} + \lambda\frac{\partial\|W\|_1}{\partial W}, \tag{9}$$

which may become non-differentiable due to the new term $\frac{\partial\|W\|_1}{\partial W}$, e.g., at the singular points $w_{ij} = 0, \forall i, j$. In the case that $\|W\|_1$ is non-differentiable at some points, the subgradient method can be applied [47]. However, the iteration of the subgradient method become rare at the points of non-differentiability, where may often be the true minimum. Furthermore, the subgradient method sometimes cannot guarantee an accurate sparse solution [48]. In sparse DGCNN, we use the forward-backward splitting method [49], which can efficiently solve the non-differentiable and constrained optimization problem such as minimizing the objective function in Eq. (8). The detailed steps are as follows.

Sparse DGCNN has two steps in each iteration. The first step is the same as that in DGCNN by doing the back propagation using the $Loss$ function in Eq. (6)

$$W^{i+\frac{1}{2}} = W^i - \tau_i\frac{\partial Loss}{\partial W^i}, \tag{10}$$

where τ_i is another learning rate. In the second step, we compute a new matrix W^{i+1} that finds a good tradeoff between two goals: (1) stay closely to $W^{i+\frac{1}{2}}$ and (2) achieve a sparse representation with small $\|W\|_1$

$$W^{i+1} = \arg\min_W\left\{\|W - W^{i+\frac{1}{2}}\|_F^2 + \tau_{i+\frac{1}{2}}\lambda\|W\|_1\right\}, \tag{11}$$

where $\tau_{i+\frac{1}{2}}$ is a learning rate. In the above objective function (Eq. (11)), sparse coding is formulated as a minimization problem composed of a matrix data proximity term, characterized by the Frobenius norm,

$$\|W'\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}w'^2_{ij}} = \sqrt{trace(W'^TW')}, \tag{12}$$

and a l_1 regularizer that enforces sparsity on W.

Algorithm 1. Sparse DGCNN

Input: An graph signal $G = \{V, W\}$ representing multichannel EEG feature, the learning rate τ and the sparse constraint weight λ.

Output: The optimal adjacent matrix W^* and the optimal set of network parameters Θ^*.

1: Initialize W and Θ.
2: **while** the termination conditions are not meet **do**
3: Compute the Laplacian matrix and the Chebyshev polynomials.
4: Compute the network results through the graph filtering layer, the graph convolution layer, the ReLu activation layer and the fully connected layer.
5: Compute the loss function in Eq. (8).
6: Update the network parameters by $\Theta \leftarrow \Theta - \tau\frac{\partial\mathcal{L}}{\partial\Theta}$.
7: Compute $W^{\frac{1}{2}} = W - \tau\frac{\partial Loss}{\partial W}$.
8: Compute $W = \arg\min_{W'}\{\|W' - W^{\frac{1}{2}}\|_F^2 + \tau\lambda\|W'\|_1\}$
9: Regularize each entry w_{ij} in W using the ReLu operation to ensure that all $w_{ij} \geq 0$.
10: **end while**
11: **return** W and Θ.

Theorem 1. *The solution to the problem in Eq. (11) is*

$$w_{j,k}^{i+1} = sign\left(w_{j,k}^{i+\frac{1}{2}}\right)\max\left\{0, |w_{j,k}^{i+\frac{1}{2}}| - \frac{1}{2}\tau_{i+\frac{1}{2}}\lambda\right\}, \tag{13}$$

where $w_{j,k}$ is the entry at the jth row and kth column of W and

$$sign(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases}.$$

Proof. See Appendix A1, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TAFFC.2021.3051332. □

Theorem 1 shows that if the absolute value of an entry in the matrix $W^{i+\frac{1}{2}}$ is smaller than $\frac{1}{2}\tau_{i+\frac{1}{2}}\lambda$, the corresponding entry in the optimal solution W^* will be set to zero and thus the sparsity is achieved. Note[1]. that in our second step for minimizing the objective in Eq. (11), an alternative method is to use the soft thresholding [50] that leads to the same solution as in Theorem 1.

Duchi and Singer [49] proves that if $\tau_{i+\frac{1}{2}} = \tau_{i+1}$ and under some mild assumptions (see Theorem 1 in [49]), the value of the loss function in each subsequent iteration will not increase and the forward-backward splitting method is converged. Accordingly, in Algorithm 1 which summarizes the pseudo-code of sparse DGCNN, we fix $\tau_{i+\frac{1}{2}} = \tau_{i+1} = \tau$, $\forall i$, and the value of τ is optimized in Section 4.

The iteration in Algorithm 1 terminates when the following condition is meet: the relative change $\frac{\mathcal{L}_k - \mathcal{L}_{k-1}}{\mathcal{L}_k}$ does not exceed a threshold ϕ in consecutive n_c iterations, where $\mathcal{L}_k$ is the loss function value at the kth iteration. In all our experiments, we used fixed parameters $\phi = 10^{-3}$ and $n_c = 5$.

4 Experiments

To evaluate the performance of sparse DGCNN, we compared it with four representative algorithms including two classic machine learning methods, i.e., Support Vector Machine (SVM) and Graph regularized Extreme Learning Machine (GELM), and two deep neural network models, i.e., Deep Belief Network (DBN) and the original DGCNN.

SVM uses some kernel functions to transform the input data into a feature space in which a hyperplane is found to best separate data into two classes with the largest distance. SVM is one of the most popular traditional machine learning methods and was used in [51] to investigate frequency bands in EEG signals. Peng *et al.* [52] proposed GELM that constructed a graph whose nodes were samples. GELM imposed graph regularization to basic ELM to enforce that for samples from the same class, their outputs are similar. Note that the graph in GELM does not depict multi-channel connections. GELM was used to recognize emotion and achieved better performance compared with SVM classifier [51]. DBN uses a stack of restricted Boltzmann machines and its last layer is a classifier. DGCNN is summarized in Section 2.2.

To sum up, for SVM, DBN and GELM, the data is embedded into sequences and matrices, i.e., EEG features from all channels in a time unit are concatenated to be a sequence and then all the time units composed a matrix. For DGCNN and sparse DGCNN, EEG data was embedded into an irregular graph structure, i.e., each node of the graph is the EEG data from one channel while the link between nodes represents the connection between channels. For fair comparison, we use the state-of-the-art implementations in [15] for SVM and DBN, [51] for GELM and [31] for DGCNN.

Evaluations were conducted on four representative EEG datasets which have different sizes and emotion classifications: SEED [15], DEAP [4], DREAMER [16] and CMEED [3]. To ensure the consistency of evaluation across different datasets, we applied the same protocols, performed the same emotion recognition tasks and calculated the same feature sets to the best extent. The details are as follows.

Two Protocols. One is leave-one-clip-out cross validation for subject dependent evaluation. That is, for each subject who watched m video clips, we trained a model for him/her using $m-1$ clips and tested on the left 1 clip. The final results were averaged over all the tests in which each clip was used for one test. The other protocol is leave-one-subject-out cross validation for subject independent evaluation. That is, data of one subject were used as the testing set while data from the other subjects were used as the training set. Likewise, the final results were averaged over all the tests in which data of each subject were used for one test.

Two Emotion Recognition Tasks. SEED has positive-negative-neutral valence emotions while DEAP, DREAMER and CMEED provide positive-negative valence and high-low arousal emotions. Thus, we identified positive versus negative valence and high versus low arousal emotions in this paper, i.e., the valence-arousal dimensional model was utilized.

Six Basic Features and Fusions. DE, PSD, DASM, RASM, ASM and DCAU from θ, α, β and γ frequency bands were extracted as data source. We further consider the feature fusion that may improve the classification accuracy. We first fused separate bands of a feature and used the feature from all bands as input data, e.g., DE from all bands

$$\text{DE\_all} = \{\text{DE\_}\theta, \text{DE\_}\alpha, \text{DE\_}\beta, \text{DE\_}\gamma\}. \tag{14}$$

Since (i) the features of DE and PSD are computed at each EEG electrode and (ii) the features of DASM, RASM and ASM are computed at each left-right pair of EEG electrodes, we then fuse different types of features and form two fusion features

$$\text{DE+PSD} = \{\text{DE\_all}, \text{PSD\_all}\} \tag{15}$$

$$\text{DASM+RASM+ASM} = \{\text{DASM\_all}, \text{RASM\_all}, \text{ASM\_all}\}. \tag{16}$$

Evaluation Metrics. We perform pair-sample t-tests (measured in p-value) to examine whether the differences in recognition accuracies between sparse DGCNN and the best method in SVM, DBN, GELM, or DGCNN are significant (***$p < .001$, **$p < .01$, *$p < .05$). We report the highest p-value for each experiment. Since all the test tasks in this paper are binary classification, in addition to classification accuracy, F1 score is another good measure that considers the balance between the precision and the recall. Moreover, the number of non-zero entries in the matrix is a widely used measure for the sparsity of that matrix. We also calculate the average percentage of non-zero entries in the learned adjacent matrices between DGCNN [31] and sparse DGCNN.

1. We thank one reviewer for pointing out this note to us.

TABLE 1
Comparison of Average (%) and Standard Deviation of Accuracies (mean/std) on SEED Using Five Classifiers, i.e., SVM, DBN, GELM, DGCNN, and Sparse DGCNN (SparseD)

Feature	Classifier	θ band	α band	β band	γ band	all bands	θ band	α band	β band	γ band	all bands
		Subject-dependent					Subject-independent				
DE	SVM	33.58/30.57	28.15/23.14	66.38/25.29	73.67/13.79	47.29/5.17	53.63/8.62	53.60/8.76	64.38/13.33	64.21/12.59	77.16/13.06
	DBN	47.54/5.67	45.47/7.41	50.95/5.68	48.81/8.18	43.16/10.36	65.94/14.41	72.40/16.46	67.22/11.71	64.28/10.96	58.50/10.94
	GELM	47.92/3.27	47.15/2.95	50.29/3.59	49.46/4.79	44.09/4.12	70.21/29.92	72.98/17.74	83.47/14.13	85.41/15.63	70.80/0.74
	DGCNN	71.73/11.42	67.93/16.29	88.47/12.40	92.25/6.65	92.27/5.32	65.62/13.06	65.62/13.32	82.20/13.52	86.85/13.77	80.11/16.27
	SparseD	**97.30*****/6.60	**97.22*****/5.36	**98.43****/2.74	**98.34*****/4.17	**98.53****/3.02	**77.30****/10.08	**79.29****/12.95	**90.62***/12.08	**90.65****/10.93	**89.71***/11.92
PSD	SVM	31.42/24.47	33.17/23.89	61.74/28.43	70.89/19.82	47.04/6.01	52.48/9.01	53.31/5.24	69.66/15.33	67.54/11.36	56.65/10.72
	DBN	48.95/6.05	46.93/7.38	48.71/6.41	47.64/9.46	42.28/11.03	55.46/10.99	72.99/15.13	75.14/14.83	74.62/15.47	59.30/14.23
	GELM	48.95/3.26	49.04/2.68	49.86/3.79	50.30/3.05	45.92/4.69	68.45/28.85	72.29/15.50	82.82/14.65	83.64/15.58	74.26/10.64
	DGCNN	69.32/11.46	70.83/16.17	87.60/13.62	91.20/6.79	93.15/7.27	69.44/11.76	65.10/12.91	85.12/13.04	80.42/16.31	70.28/20.77
	SparseD	**98.11*****/5.20	**98.28*****/4.06	**98.21****/4.98	**96.99*****/4.07	**98.26****/3.21	**77.62***/9.30	**76.41****/13.87	**86.73**/12.69	**85.77***/14.54	**88.06****/13.07
DASM	SVM	38.25/16.13	46.08/17.93	64.87/23.30	71.76/17.36	50.28/5.36	50.40/7.85	48.22/6.82	52.78/10.03	51.68/4.37	50.19/16.72
	DBN	43.44/5.83	46.78/8.98	46.94/6.74	49.15/3.98	53.68/9.57	59.22/10.21	49.14/12.50	54.15/26.31	55.55/25.88	46.52/13.71
	GELM	47.72/2.48	47.32/2.83	49.65/2.92	49.48/3.60	46.44/5.12	59.04/10.20	49.20/12.74	56.41/29.36	57.98/29.50	59.62/8.92
	DGCNN	69.17/12.12	67.06/15.84	83.94/15.59	87.69/8.90	86.44/11.97	55.71/6.57	53.94/6.00	59.25/12.62	62.72/14.32	66.01/14.63
	SparseD	**91.64*****/6.71	**95.54*****/8.24	**97.67****/4.24	**96.58*****/5.89	**98.27****/3.38	**65.67****/9.76	**66.96*****/11.23	**73.02***/17.06	**73.84*****/14.62	**75.07***/13.69
RASM	SVM	38.43/19.28	43.63/17.14	72.51/18.83	72.61/14.96	50.60/6.27	50.51/5.29	52.68/6.86	50.45/9.33	48.38/8.90	59.94/18.63
	DBN	44.63/6.48	44.55/8.79	47.47/4.94	48.94/8.26	55.35/8.65	59.17/11.67	50.44/12.60	53.56/26.04	57.70/21.68	51.14/4.55
	GELM	47.84/2.62	47.57/2.82	49.60/3.60	49.14/3.50	42.76/6.25	60.65/11.14	51.01/12.40	55.07/28.47	61.26/27.87	64.69/6.34
	DGCNN	70.59/18.70	65.64/11.49	83.71/11.49	87.51/10.93	87.79/12.17	55.14/5.28	57.83/7.40	58.43/12.28	65.40/17.34	69.30/15.70
	SparseD	**95.01*****/5.68	**97.05*****/4.86	**95.69*****/4.36	**96.56*****/5.28	**98.94****/2.71	**66.60*****/7.97	**66.34****/10.35	**71.69****/14.89	**72.48***/16.27	**76.51****/16.70
ASM	SVM	39.94/20.91	40.74/25.31	49.59/35.99	53.33/32.08	48.97/5.72	50.23/1.35	50.08/1.38	50.30/1.19	50.22/1.11	56.08/13.90
	DBN	46.72/8.31	44.14/6.60	48.39/8.25	51.11/8.29	49.35/12.71	52.54/13.43	43.18/15.51	54.23/22.86	54.44/20.50	50.29/1.17
	GELM	47.68/3.60	47.01/3.31	47.01/3.89	47.59/4.71	45.87/4.73	52.70/13.39	44.13/14.69	52.95/24.50	53.60/20.84	54.25/14.18
	DGCNN	67.77/17.22	71.19/14.51	71.40/16.12	74.17/22.86	74.15/16.73	52.04/2.56	51.63/2.02	51.52/1.61	52.24/4.28	59.73/10.08
	SparseD	**94.10*****/6.30	**93.73*****/5.62	**92.96*****/7.50	**94.79*****/8.32	**97.77*****/4.35	**61.74****/9.99	**58.48****/6.47	**63.10*****/10.31	**63.69*****/10.14	**65.78**/11.69
DCAU	SVM	38.62/20.86	37.47/19.91	68.00/20.64	68.66/21.11	48.83/6.11	49.50/4.05	51.24/5.37	51.94/6.66	53.77/7.37	58.51/11.13
	DBN	45.45/7.00	46.09/7.26	45.80/8.19	48.84/7.83	52.90/8.55	59.69/10.75	49.47/15.60	65.83/17.46	62.40/18.04	55.14/12.79
	GELM	47.49/3.39	47.06/2.60	48.07/2.77	48.47/3.31	46.77/4.72	62.31/12.20	49.88/17.24	71.13/20.79	63.17/21.77	63.12/8.03
	DGCNN	69.63/16.04	65.93/12.93	81.66/13.94	89.87/10.10	85.42/13.58	57.99/6.44	56.95/7.91	67.28/17.07	72.37/16.54	73.36/14.49
	SparseD	**94.30*****/10.62	**91.17*****/9.17	**96.29*****/4.78	**95.46****/4.91	**98.01*****/3.43	**63.84*****/7.25	**65.85****/7.83	**76.41***/14.52	**76.86**/14.92	**82.38***/11.03
Fusion features		Average (%) and standard deviation of accuracies (only available in all bands)									
		Subject-dependent					Subject-independent				
DE+PSD		SVM (47.88/2.82) DBN (33.45/1.30) GELM (46.05/3.77) DGCNN (95.16/4.73) SparseD (**99.01****/2.67)					SVM (74.15/7.18) DBN (50.80/0.74) GELM (84.90/10.92) DGCNN (84.45/13.59) SparseD (**88.67**/11.23)				
DASM+RASM+ASM		SVM (47.28/3.76) DBN (37.90/8.64) GELM (46.24/4.22) DGCNN (93.69/8.63) SparseD (**98.81*****/2.29)					SVM (52.43/18.24) DBN (52.31/7.51) GELM (63.37/7.48) DGCNN (60.65/9.85) SparseD (**74.89*****/11.31)				

*The best performance is shown in bold. The differences in accuracies between SparseD and the best other method are highlighted with stars (***p $<$ .001, **p $<$ .01, *p $<$ .05). The max* p*-values are 0.057 and 0.199 for subject-dependent/independent evaluations, respectively. All results are for positive versus negative valence recognition.*

In all our experiments, we optimize the learning rate τ and the weight of the sparse constraint λ during the training process by searching in the range of [0.1,1].

4.1 Evaluation on SEED

The SEED dataset [15] collected EEG data from 15 subjects (7 males and 8 females, age mean=23.27, SD=2.37). These EEG data were elicited by audio and visual stimuli in the form of Chinese emotional film clips. During participants' watching 15 4-minute film clips with three emotion types (positive, negative and neutral), their EEG data were recorded by an ESI NeuroScan system from 62 channels at a sampling rate of 1,000 Hz. The EEG data were later down-sampled to 200 Hz and manually checked to remove EOG and EMG artifacts. For each subject, EEG recordings were captured across three chronologically disjointed sessions and each sessions repeated the same experiment. Because SEED is the only one that contains three repeated sessions in the four datasets, to ensure the consistency of evaluation, in our experiment we only use the first session for each subject since the first one reflects more reliable emotions than the later two sessions. In addition, SEED does not contain arousal information, so we only recognize positive and negative emotions on SEED. After preprocessing with a bandpass filter (between 0.3 to 50 Hz), five features (DE, PSD, DASM, RASM and DCAU) were extracted with a window of 1s on four bands, θ band (4-7 Hz), α band (8-13 Hz), β band (14-30 Hz) and γ band ($>$30 Hz).

The results of both subject-dependent and subject-independent evaluations were summarized in Table 1 (for the positive versus negative valence classification accuracy), Table A1 (for F1 scores) and Table A2 (for non-zero entries in the learned adjacent matrices) in Appendix, available in the online supplemental material, respectively, in which sparse DGCNN generally achieves the best performance. In these tables, "all bands" refers to fusion of all the four bands.

In subject-dependent evaluation, the accuracy of sparse DGCNN is averagely 16.73 percent higher than the accuracy of all other methods summarized on Table 1. The max p-value is 0.057 and all the other p-values are smaller than 0.05. The F1 score of sparse DGCNN is 15.76 percent higher than all the other methods. DGCNN produced 50.71 percent non-zero entries in adjacency matrix, while sparse DGCNN only produced 27.22 percent, demonstrating that sparse DGCNN can effectively sparsify the learned adjacent matrices by reducing the number of non-zero entries.

In subject-independent evaluation, the accuracy of sparse DGCNN is averagely 7.58 percent higher than the accuracy of all other methods. The max p-value is 0.199, and only 4 of 32 p-values are larger than 0.05 while all the others are smaller than 0.05. The F1 score of sparse DGCNN is averagely 6.32 percent higher than all the other methods. Meanwhile, DGCNN produced 51.99 percent non-zero entries in adjacency matrix, while sparse DGCNN only produced 23.84 percent non-zero entries.

4.2 Evaluation on DEAP

The DEAP dataset [4] collected EEG and peripheral signals from 32 subjects (16 female and 16 male, age mean=26.9, age ranged from 19 to 37). The EEG data were elicited during watching forty 1-min-long music videos and recorded at a sampling rate of 512 Hz using 32 electrodes in Biosemi ActiveTwo System. The EEG data were later down-sampled to 128 Hz and EOG artifacts were removed. The data was averaged to the common reference and segmented into 1-min-long clips. A 3-second-long pre-trial baseline was removed. According to self-assessment manikins (SAM), high/low binary labels in arousal, valence dimensions were assigned to each movie and related EEG signals. As aforementioned, we did binary classification on positive/negative valence and high/low arousal dimensions. Because the preprocessed data provided by DEAP was filtered in a 4.0 - 45.0

TABLE 2
Comparison of Average (%) and Standard Deviation of Accuracies (mean/std) on DEAP Using Five Classifiers, i.e., SVM, DBN, GELM, DGCNN, and Sparse DGCNN (SparseD)

Feature	Classifier	θ band	α band	β band	γ band	all bands	θ band	α band	β band	γ band	all bands
		Valence					Arousal				
		Subject-dependent									
DE	SVM	57.39/12.29	54.33/9.37	51.55/10.32	51.51/10.45	51.60/6.32	61.17/13.64	59.37/15.65	58.01/14.97	58.21/13.63	55.80/9.49
	DBN	54.41/10.77	53.73/10.80	54.13/10.36	55.78/8.90	55.04/8.82	61.57/13.41	61.37/13.54	61.31/12.88	62.30/11.86	62.43/11.87
	GELM	50.94/2.85	51.46/3.29	51.35/2.24	51.24/2.43	50.73/2.29	54.76/7.39	54.57/6.36	53.63/5.89	53.06/5.46	52.67/3.48
	DGCNN	82.53/8.37	83.41/8.90	83.98/9.29	84.25/10.45	86.32/6.04	79.77/7.49	80.13/8.76	81.44/9.20	80.42/8.49	83.68/5.68
	SparseD	**91.61*****/8.62	**92.20*****/9.03	**93.24*****/8.38	**92.09*****/9.52	**95.72*****/3.75	**88.30*****/8.32	**87.92*****/8.41	**88.07*****/8.16	**87.97*****/8.35	**91.75*****/5.23
PSD	SVM	57.38/12.29	54.33/9.37	51.55/10.33	51.51/10.47	51.58/6.35	61.17/13.65	59.37/15.64	58.01/14.97	58.21/13.63	55.80/9.51
	DBN	54.69/10.91	53.28/11.76	54.59/10.13	55.88/9.08	55.50/8.31	61.41/13.94	62.30/12.41	61.60/12.91	62.55/11.67	61.90/12.62
	GELM	51.27/3.11	51.32/3.11	50.92/2.61	51.26/2.59	51.14/2.37	54.91/7.22	54.40/6.33	53.33/5.63	53.11/5.43	52.90/3.63
	DGCNN	74.23/4.72	73.77/3.32	75.95/4.52	75.37/5.11	83.87/3.57	76.49/4.84	75.71/4.48	76.60/5.69	77.22/4.26	83.37/2.98
	SparseD	**97.69*****/3.63	**97.66*****/3.13	**98.42*****/2.10	**98.23*****/2.33	**98.74*****/1.61	**96.70*****/3.33	**96.12*****/4.23	**97.39*****/3.58	**97.80*****/2.14	**98.16*****/3.10
DASM	SVM	54.68/13.20	52.05/14.85	53.89/9.77	52.90/9.69	54.50/5.81	61.98/13.80	61.64/13.84	57.65/16.31	58.41/14.54	57.51/10.88
	DBN	54.61/10.90	55.41/9.46	53.77/9.10	56.31/9.27	56.70/9.28	60.81/13.30	60.53/13.21	60.06/11.95	59.54/10.15	60.88/11.42
	GELM	51.84/4.26	51.57/3.52	51.53/3.05	51.76/2.60	50.80/2.23	55.93/8.52	55.58/8.07	54.27/7.22	53.69/5.97	53.23/4.37
	DGCNN	74.57/8.73	74.76/8.11	77.59/9.31	79.14/9.95	79.98/7.65	73.99/7.56	73.87/7.74	74.40/9.11	75.70/8.68	76.87/6.45
	SparseD	**86.18*****/10.60	**86.32*****/10.11	**89.46*****/8.89	**89.71*****/9.01	**90.08*****/7.08	**83.35*****/9.29	**82.80*****/8.64	**84.16*****/8.67	**86.37*****/7.45	**85.90*****/6.22
RASM	SVM	54.52/12.69	53.97/12.37	52.75/11.06	51.94/11.32	52.15/6.63	61.73/13.91	61.12/14.49	57.60/16.68	59.54/13.85	58.51/11.40
	DBN	52.25/9.27	54.05/8.99	55.56/9.89	56.89/9.44	54.29/9.90	61.37/11.72	61.61/11.43	60.38/11.30	58.79/9.66	57.76/11.23
	GELM	52.18/4.13	51.30/3.10	51.43/2.99	51.62/2.39	51.20/2.27	55.74/8.58	55.56/7.98	54.03/6.80	53.91/5.44	52.95/4.17
	DGCNN	76.14/11.60	76.24/11.69	76.66/11.81	79.07/11.97	79.79/10.43	74.86/9.79	74.88/10.50	75.21/11.78	75.99/10.47	76.71/8.60
	SparseD	**87.25*****/12.44	**86.35*****/12.93	**86.71*****/13.25	**88.70*****/11.85	**90.08*****/9.77	**81.15*****/11.22	**80.86*****/11.49	**81.77*****/11.50	**82.86*****/10.78	**85.50*****/8.35
ASM	SVM	52.47/16.04	51.68/16.91	53.17/11.43	54.18/9.04	51.54/6.56	58.20/19.91	61.42/14.34	62.05/12.28	61.55/12.63	57.82/12.29
	DBN	52.44/11.02	53.01/9.25	50.81/8.05	52.04/8.36	50.50/8.69	59.20/13.43	58.54/13.82	58.44/12.01	57.52/10.02	55.73/11.70
	GELM	51.91/7.16	50.02/5.55	50.83/3.61	50.53/4.86	50.49/3.26	58.09/10.98	56.95/9.97	54.94/7.54	54.52/7.28	53.30/4.85
	DGCNN	73.88/8.85	73.74/8.56	75.37/8.78	78.27/9.79	79.81/7.50	79.74/8.20	74.08/7.98	75.49/8.62	73.27/8.72	80.21/8.65
	SparseD	**86.25*****/10.12	**85.46*****/10.48	**89.18*****/9.10	**89.26*****/7.82	**90.89*****/6.40	**98.24*****/2.90	**97.93*****/2.86	**98.70*****/1.82	**99.53*****/1.17	**95.17*****/5.80
DCAU	SVM	51.49/16.35	51.86/13.43	51.93/10.46	52.71/10.29	51.83/6.92	61.37/14.60	59.34/15.84	59.66/14.17	58.34/14.36	57.49/11.40
	DBN	54.29/10.84	53.77/10.65	55.65/8.19	56.96/9.07	57.07/8.26	61.16/13.47	62.22/11.26	61.56/11.42	59.52/10.67	62.38/11.16
	GELM	51.58/3.68	51.40/4.06	51.16/2.69	51.48/2.47	51.29/2.60	55.90/8.40	55.38/7.93	54.44/6.96	53.51/5.30	53.11/3.62
	DGCNN	73.88/8.85	73.74/8.56	75.37/8.78	78.27/9.79	79.81/7.50	73.51/8.58	72.74/7.62	74.38/8.22	75.81/8.80	77.68/5.55
	SparseD	**84.53*****/10.18	**83.82*****/10.34	**87.64*****/8.94	**88.05*****/7.31	**88.39*****/6.40	**82.05*****/8.99	**81.31*****/8.77	**83.34*****/8.75	**85.95*****/7.78	**86.94*****/6.54
Fusion features		Average (%) and standard deviation of accuracies (only available in all bands)									
		Valence					Arousal				
DE+PSD		SVM (51.82/5.65) DBN (55.16/8.43) GELM (51.39/2.29) DGCNN (86.93/4.04) SparseD (**96.18*****/2.54)					SVM (54.16/7.97) DBN (61.95/11.97) GELM (52.34/3.36) DGCNN (86.21/3.73) SparseD (**93.82*****/3.96)				
DASM+RASM+ASM		SVM (51.65/4.66) DBN (55.13/9.28) GELM (50.80/2.05) DGCNN (83.31/7.05) SparseD (**92.62*****/6.16)					SVM (55.08/8.97) DBN (59.26/13.93) GELM (52.96/3.85) DGCNN (80.91/6.15) SparseD (**90.35*****/6.41)				
		Subject-independent									
DE	SVM	56.56/9.22	56.56/9.22	56.56/9.22	56.57/9.22	48.58/4.24	58.91/15.45	58.87/15.39	58.91/15.45	58.91/15.45	50.75/4.87
	DBN	55.34/9.06	54.73/9.22	54.92/9.51	56.06/8.45	51.64/7.84	58.90/14.91	58.94/15.30	56.67/15.25	55.62/15.23	56.68/13.28
	GELM	55.71/8.55	55.30/7.67	55.63/8.47	54.17/9.63	54.12/9.12	58.99/15.08	58.45/14.86	58.87/15.19	54.88/14.77	58.64/14.65
	DGCNN	57.22/8.40	57.40/8.53	57.62/8.29	58.17/7.94	58.46/7.85	61.27/12.94	61.16/13.55	62.65/11.98	62.78/12.17	61.65/13.34
	SparseD	**59.44*****/6.67	**59.46*****/6.78	**60.25*****/6.75	**61.21*****/6.29	**60.65*****/6.24	**64.93**/9.68	**64.57***/10.04	**65.11***/9.74	**65.82****/9.01	**65.39**/9.41
PSD	SVM	56.56/9.22	56.56/9.22	56.56/9.22	56.57/9.22	48.57/4.24	58.91/15.45	58.87/15.39	58.91/15.45	58.91/15.45	50.75/4.86
	DBN	55.56/9.08	55.84/8.67	56.30/8.55	54.42/10.33	52.82/8.71	57.65/14.63	57.86/14.53	55.55/15.93	55.43/15.44	56.10/13.29
	GELM	56.51/9.00	52.01/5.55	55.84/9.53	55.64/9.57	56.33/8.73	56.82/14.81	57.49/15.21	58.27/15.22	58.02/15.56	58.82/15.19
	DGCNN	56.69/7.58	57.09/7.34	57.21/9.22	58.84/6.47	59.13/7.13	61.85/10.42	61.06/13.09	61.54/11.30	60.93/12.89	62.91/11.44
	SparseD	**59.98*****/5.67	**60.09*****/5.72	**61.97***/5.90	**63.51*****/5.93	**61.80*****/5.65	**65.10****/9.40	**64.94***/9.46	**65.81*****/8.95	**66.65****/8.16	**65.74****/9.05
DASM	SVM	56.56/9.22	56.56/9.22	56.56/9.22	56.56/9.22	56.56/9.22	58.91/15.45	58.91/15.45	58.90/15.16	58.90/15.44	49.93/8.72
	DBN	56.03/9.06	55.59/9.25	53.80/10.03	52.57/11.55	51.88/9.67	56.14/14.29	57.48/14.39	54.57/16.64	56.13/14.25	54.11/15.99
	GELM	56.55/9.07	56.05/9.04	56.56/9.07	56.56/9.07	56.56/9.07	56.98/15.24	58.82/15.12	56.65/14.73	58.90/15.20	56.51/15.98
	DGCNN	57.86/8.11	57.36/8.33	57.87/8.18	57.27/8.87	56.87/8.34	60.74/12.30	61.02/11.93	60.59/12.76	60.04/13.37	61.03/12.95
	SparseD	**59.72***/6.40	**58.54****/7.61	**60.55****/6.56	**61.17*****/6.92	**61.07*****/6.00	**64.80***/9.80	**64.15****/10.14	**65.06****/9.76	**65.75***/9.09	**65.12***/9.58
RASM	SVM	56.56/9.22	56.56/9.22	56.56/9.22	56.56/9.22	56.56/9.22	58.85/15.39	58.83/15.36	58.21/14.54	58.62/15.17	49.08/3.03
	DBN	48.68/10.51	51.41/9.68	53.73/10.41	53.64/9.74	52.56/9.31	43.18/15.60	55.13/16.06	56.13/14.15	49.43/16.54	52.47/14.76
	GELM	56.56/9.07	56.57/9.07	55.78/9.59	56.56/9.07	56.56/9.07	58.79/15.15	58.84/15.16	58.36/15.36	58.91/15.21	57.96/15.41
	DGCNN	58.32/7.55	57.72/8.18	57.07/8.86	58.56/7.50	57.55/8.26	62.65/12.16	60.61/13.02	62.55/12.19	61.69/13.24	61.53/13.02
	SparseD	**59.32***/6.68	**59.24***/6.74	**60.00***/6.81	**60.88***/6.78	**59.89****/6.68	**64.54**/10.04	**64.67***/9.91	**64.73**/9.92	**64.99**/9.65	**64.97***/9.67
ASM	SVM	56.56/9.22	56.56/9.22	56.43/9.34	55.77/8.77	53.60/7.98	58.80/15.09	58.66/14.99	58.21/14.87	58.27/14.90	58.17/14.27
	DBN	50.35/11.18	49.58/10.75	45.78/10.54	48.03/11.15	48.96/11.30	44.42/16.56	48.65/17.79	46.48/17.40	46.93/17.29	38.57/13.62
	GELM	56.56/9.07	55.74/9.50	56.54/9.04	56.53/9.09	56.64/8.92	58.22/14.55	58.14/15.44	59.97/14.41	58.83/15.19	58.16/14.89
	DGCNN	57.03/7.59	57.16/7.76	57.62/7.49	59.60/7.17	58.73/6.87	60.20/13.26	58.43/13.89	62.57/11.52	62.62/11.96	60.57/13.89
	SparseD	**59.85****/5.64	**59.77****/5.82	**62.52*****/6.00	**64.66*****/6.24	**60.40*****/5.98	**65.17***/9.36	**64.93****/9.57	**65.98***/8.92	**66.49****/8.51	**65.04**/9.61
DCAU	SVM	56.56/9.22	56.56/9.22	56.56/9.22	56.56/9.22	56.56/9.22	58.91/15.45	58.91/15.45	58.91/15.45	58.91/15.45	49.74/9.36
	DBN	53.10/9.44	52.27/9.03	56.43/8.52	54.17/10.67	48.60/10.71	51.71/13.96	54.49/14.87	54.07/14.81	53.10/16.41	50.66/15.10
	GELM	56.50/9.07	56.50/9.04	56.56/9.07	56.56/9.07	56.65/9.16	58.94/14.71	58.55/14.90	58.90/15.20	58.26/15.47	58.82/15.14
	DGCNN	56.93/8.83	56.32/9.20	57.05/8.83	58.40/7.41	56.11/9.12	60.09/12.21	60.14/13.19	61.52/12.97	60.83/13.61	60.66/13.70
	SparseD	**59.72****/6.47	**59.37****/6.80	**61.01*****/6.34	**61.01*****/6.34	**61.18****/5.59	**64.73****/9.60	**64.20***/10.38	**64.21**/10.48	**65.23***/9.08	**65.02***/9.67
Fusion features		Average (%) and standard deviation of accuracies (only available in all bands)									
		Valence					Arousal				
DE+PSD		SVM (47.85/3.37) DBN (51.66/8.92) GELM (56.53/8.85) DGCNN (59.64/7.43) SparseD (**61.93*****/6.03)					SVM (50.89/4.38) DBN (53.81/14.29) GELM (58.36/15.38) DGCNN (60.56/13.61) SparseD (**65.94****/8.82)				
DASM+RASM+ASM		SVM (48.08/4.87) DBN (50.32/10.47) GELM (56.14/8.96) DGCNN (58.69/7.45) SparseD (**60.85*****/5.90)					SVM (49.51/5.93) DBN (46.24/14.94) GELM (59.70/14.29) DGCNN (62.48/12.02) SparseD (**65.39*****/9.48)				

*The best performance is shown in bold. The differences in accuracies between SparseD and the best other method are highlighted with stars (***p < .001, **p < .01, *p < .05). All results are for positive versus negative valence recognition. The max* p*-values are 0.0001 and 0.193 for subject-dependent/independent evaluations, respectively.*

Hz band (i.e., δ band was removed), we extracted six features with windows of 2s on the other four bands, i.e., θ (4-8 Hz), α (8-12 Hz), β (12-30 Hz), and γ ($>$30 Hz).

The results of both subject-dependent and subject-independent evaluations were summarized in Table 2 (for positive versus negative valence and high versus low arousal classification accuracy), Table A3 (for F1 scores) and Table A4 (for non-zero entries in the learned adjacent matrices) in Appendix, available in the online supplemental material, respectively.

In subject-dependent evaluation, the accuracy of sparse DGCNN is averagely 12.12 percent higher than the accuracy of all other four methods, and the max p-value is 0.0001. The F1 score of sparse DGCNN is averagely 11.61 percent higher than all the other methods. DGCNN produced 47.75 percent non-zero entries in adjacency matrix, while sparse DGCNN only produced 37.77 percent, demonstrating that sparse DGCNN can effectively sparsify the learned adjacent matrices by reducing the number of non-zero entries.

In subject-independent evaluation, the accuracy of sparse DGCNN is averagely 3.39 percent higher than the accuracy of all other methods. The max p-value is 0.193, and only 8 of 64 p-values are larger than 0.05 while all the other 56 p-values are smaller than 0.05. The F1 score of sparse DGCNN is averagely 6.13 percent higher than all the other methods. DGCNN produced 51.75 percent non-zero entries in adjacency matrix, while sparse DGCNN only produced 25.37 percent non-zero entries.

4.3 Evaluation on DREAMER

The DREAMER dataset [16] collected EEG and ECG data from 23 subjects (14 males and 9 females, age mean=26.6, SD=2.7). The EEG data were elicited during watching 18 English emotional film clips (length between 65 to 393 seconds, mean=199 seconds), and recorded by an Emotiv EPOC system using 14 of 16 channels (the left 2 are references) at a sampling rate of 128 Hz. We followed the

preprocessing method in [16] to remove artifacts in EEG data, only kept the last 60s of each clip and further processed EEG data using a 2s window with an overlap of 1s in four bands as being consistent with other three datasets: the bands are θ band (4-8 Hz), α band (8-13 Hz), β band (13-30 Hz) and γ band (>30 Hz). For each band, the PSD feature was computed. Nine emotion types (amusement, excitement, happiness, calmness, anger, disgust, fear, sadness and surprise) are labeled with EEG data. These emotion types were further categorized into the valence/arousal rating scales and by following the same thresholding strategy as in [16], the problem was converted to a binary classification problem.

The results of both subject-dependent and subject-independent evaluations were summarized in Table A5 (for positive versus negative valence and high versus low arousal classification accuracy), Table A6 (for F1 scores) and Table A7 (for non-zero entries in the learned adjacent matrices) in Appendix, available in the online supplemental material, respectively.

In subject-dependent evaluation, the accuracy of sparse DGCNN is averagely 12.66 percent higher than the accuracy of all other four methods, and the max p-value is 0.006. The F1 score of sparse DGCNN is averagely 12.45 percent higher than all the other methods. DGCNN produced 44.47 percent non-zero entries in adjacency matrix, while sparse DGCNN only produced 39.77 percent non-zero entries.

In subject-independent evaluation, the accuracy of sparse DGCNN is averagely 4.69 percent higher than the accuracy of all other four methods. The max p-value is 0.876, and only 16 of 64 p-values are larger than 0.05 while all the others are smaller than 0.05. The F1 score of sparse DGCNN is averagely 8.51 percent higher than all the other methods. DGCNN produced 50.33 percent non-zero entries in adjacency matrix, while sparse DGCNN only produced 24.94 percent non-zero entries.

4.4 Evaluation on CMEED

The CMEED dataset [3] collected EEG data from 37 subjects (17 males and 20 females, age mean=23.95, SD=1.56). The EEG data were elicited during watching 16 film clips (length between 61 to 134 seconds) and recorded by a NeuroScan quik-cap[2] using 30 of 32 electrodes (left 2 are references) at a sampling rate of 1,024 Hz. Preprocess by filtering with a 1-45 Hz filter and independent component analysis was performed using EEGLAB toolbox in MATLAB. EOG artifacts were removed by experimenters. The PSD feature was extracted from four bands: θ band (4-7 Hz), α band (8-13 Hz), β band (14-30 Hz) and γ band (>30 Hz) using a 2s window with an overlap of 1s. Similar to DREAMER and DEAP, we also divide the emotions into positive/negative valence and high/low arousal according to ratings.

The results of both subject-dependent and subject-independent evaluations were summarized in Table A8 (for positive versus negative valence and high versus low arousal classification accuracy), Table A9 (for F1 scores) and Table A10 (for non-zero entries in the learned adjacent matrices) in Appendix, available in the online supplemental material, respectively.

In subject-dependent evaluation, the accuracy of sparse DGCNN is averagely 9.60 percent higher than the accuracy of all other four methods, and the max p-value is 0.006. The F1 score of sparse DGCNN is averagely 9.58 percent higher than all the other methods. DGCNN produced 46.23 percent non-zero entries in adjacency matrix, while sparse DGCNN only produced 39.25 percent non-zero entries.

In subject-independent evaluation, the accuracy of sparse DGCNN is averagely 4.28 percent higher than the accuracy of all other methods. The max p-value is 0.698, and only 16 of 64 p-values are larger than 0.05 while all the others are smaller than 0.05. The F1 score of sparse DGCNN is averagely 6.12 percent higher than all the other methods. DGCNN produced 51.76 percent non-zero entries in adjacency matrix, while sparse DGCNN only produced 28.90 percent non-zero entries.

4.5 Cross-Corpus Evaluation

To examine the generalization ability of sparse DGCNN, we conducted a cross-corpus evaluation using DEAP and MAHNOB-HCI datasets [53]. EEG data in these two datasets were collected by the same Biosemi Active II system, and then characterized by the same set of EEG channels. We preprocessed EEG data in MAHNOB-HCI in the same way as DEAP did, and applied the same threshold to categorize positive/negative valence and high/low arousal.

In the cross-corpus evaluation, we applied subject-independent protocol, i.e., data from all subjects in one dataset were used for training, and the data of each subject in the other dataset were used for testing. The accuracies were finally averaged on all subjects in the testing dataset. The results are summarized in Table 3, which contains two sets of cross-corpus evaluation results: (1) "D To M": training with DEAP and testing on MAHNOB-HCI and (2) "M To D": training with MAHNOB-HCI and testing on DEAP. Since comparing different bands of features is not the main goal of cross-corpus evaluation, we used all bands for each feature. The results on two traditional machine learning methods and three deep learning methods showed that mostly sparse DGCNN has the best performances.

5 Discussion

Table 4 summarized the maximum accuracies and F1 scores achieved by SVM, DBN, GELM, DGCNN and sparse DGCNN on four datasets in our experiments. The results showed that F1 score follows the same trend indicated by accuracy values. In addition to the maximal average accuracies among all individual and fused features in four datasets reported in Sections 4.1, 4.2, 4.3, and 4.4, we also summarized the average results of these average accuracies in four datasets, as shown in Figure A1 in the Appendix, available in the online supplemental material. All these average and maximal accuracy results showed that our sparse DGCNN model outperforms other methods. More importantly, sparse DGCNN showed better performances than its non-sparse counterpart in all the four datasets, demonstrating the effectiveness for the introduction of sparseness on the

2. The original CMEED dataset was recorded using 14-channel wireless Emotiv EPOC headset. The authors [3] further refined and re-collected the EEG data using a NeuroScan quik-cap [17], [18].

TABLE 3
Average (%) and Standard Deviation of Accuracies (mean/std) Using Five Classifiers, i.e., SVM, DBN, GELM, DGCNN, and Sparse DGCNN (SparseD), in Cross-Corpus Evaluation

Feature	Classifier	D To M		M To D	
		valence	arousal	valence	arousal
DE	SVM	50.26/19.32	55.06/9.61	51.90/9.26	58.91/15.45
	DBN	50.26/19.32	55.12/9.61	46.51/8.35	53.07/11.69
	GELM	50.37/19.34	55.06/9.61	56.56/9.22	58.91/15.45
	DGCNN	54.00/19.20	54.00/9.74	58.91/12.43	58.07/19.11
	SparseD	**54.97**/19.10	**55.35**/10.07	**58.92**/12.43	**58.94**/19.10
PSD	SVM	50.39/19.32	50.98/6.49	56.56/9.22	58.91/15.45
	DBN	49.26/17.87	53.66/10.57	43.44/9.22	41.09/15.45
	GELM	50.38/19.32	54.20/7.54	49.45/9.93	53.59/13.00
	DGCNN	**54.00**/19.20	**54.00**/9.74	58.91/12.43	58.94/19.09
	SparseD	**54.00**/19.20	**54.00**/9.74	**58.95**/12.43	**58.98**/19.31
DASM	SVM	50.39/19.32	55.06/9.61	49.94/9.85	58.91/15.45
	DBN	50.26/19.32	55.12/9.61	47.12/5.45	58.91/15.45
	GELM	50.39/19.32	54.27/9.66	48.18/8.86	58.91/15.45
	DGCNN	50.72/19.20	54.02/9.72	56.80/12.41	58.95/19.10
	SparseD	**52.75**/14.48	**55.64**/12.81	**66.91**/9.38	**59.01**/21.18
RASM	SVM	50.39/19.32	55.06/9.61	56.56/9.22	58.91/15.45
	DBN	50.26/19.32	55.12/9.61	49.31/3.62	49.96/4.69
	GELM	50.39/19.32	55.06/9.61	54.39/10.00	58.91/15.45
	DGCNN	50.72/19.20	54.12/9.65	56.56/12.43	58.97/19.06
	SparseD	**51.09**/18.72	**56.57**/11.43	**63.68**/12.50	**59.57**/20.10
ASM	SVM	50.39/19.32	52.08/10.55	56.56/9.22	58.91/15.45
	DBN	49.87/7.57	49.75/7.36	43.44/9.22	58.91/15.45
	GELM	50.44/19.20	49.71/8.14	44.44/9.97	58.91/15.45
	DGCNN	**50.72**/19.20	**54.00**/9.74	56.56/12.43	58.91/19.10
	SparseD	**50.72**/19.21	**54.00**/9.74	**56.57**/12.43	**58.99**/18.88
DCAU	SVM	50.39/19.32	55.06/6.61	49.80/9.20	58.91/15.45
	DBN	50.26/19.32	55.12/9.61	49.75/2.70	58.91/15.45
	GELM	50.39/19.32	55.06/9.61	47.49/9.30	58.91/15.45
	DGCNN	50.72/19.21	54.58/9.17	**56.97**/12.18	58.98/19.10
	SparseD	**53.00**/13.38	**55.20**/13.20	56.78/12.65	**59.45**/19.09

"D To M" means training with DEAP and testing on MAHNOB-HCI. The same interpretation holds for "M To D". The best performance is shown in bold.

TABLE 4
The Maximum Accuracies (and F1 Scores in Brackets) Achieved by SVM, DBN, GELM, DGCNN, and Sparse DGCNN on Four Datasets

Dataset	Classifier	Subject-dependent		Subject-independent	
		valence	arousal	valence	arousal
SEED	SVM	73.67(70.03)	N/A	77.16(76.14)	N/A
	DBN	55.35(50.05)	N/A	75.14(73.31)	N/A
	GELM	50.30(40.73)	N/A	85.41(85.46)	N/A
	DGCNN	95.16(95.82)	N/A	86.85(85.63)	N/A
	SparseD	**99.01**(**99.16**)	N/A	**90.65**(**90.30**)	N/A
DEAP	SVM	57.39(49.74)	62.05(58.87)	56.57(47.41)	58.91(48.26)
	DBN	57.07(50.27)	62.55(62.61)	56.43(41.42)	58.94(41.80)
	GELM	52.18(41.12)	58.09(48.01)	56.65(47.53)	59.97(38.22)
	DGCNN	86.93(90.13)	86.21(88.46)	59.64(49.85)	62.91(52.75)
	SparseD	**98.74**(**99.73**)	**99.53**(**99.69**)	**64.66**(**60.23**)	**66.65**(**59.57**)
DM	SVM	55.61(65.92)	56.53(67.45)	60.63(49.68)	56.29(50.04)
	DBN	60.87(62.60)	68.66(72.35)	62.29(43.38)	56.29(42.35)
	GELM	57.84(60.30)	57.41(60.60)	61.07(38.90)	56.43(37.26)
	DGCNN	81.01(83.41)	82.27(83.81)	64.43(48.80)	63.29(51.72)
	SparseD	**95.92**(**96.59**)	**94.73**(**95.48**)	**68.18**(**59.29**)	**67.12**(**60.31**)
CD	SVM	57.18(69.10)	68.60(73.76)	62.05(54.30)	62.21(50.56)
	DBN	58.39(65.93)	70.76(73.14)	62.48(45.21)	62.62(45.45)
	GELM	56.87(60.94)	70.33(71.99)	62.12(39.72)	62.23(39.96)
	DGCNN	79.70(81.70)	83.22(83.76)	63.93(53.22)	69.25(59.79)
	SparseD	**95.98**(96.09)	**94.59**(**94.59**)	**66.83**(**58.87**)	**75.47**(**68.04**)

DM and CD are for DREAMER and CMEED.

irregular-graph-based connectivity patterns of EEG features for emotion classification.

Compared with subject-dependent evaluation, the accuracies and F1 scores in subject-independent evaluation dropped significantly. This may possibly be attributed to the existence of individual differences. Individual differences have been widely concerned in behavioral science. In the current study, two sources of individual differences may contribute to the difference of recognition performance between subject-dependent and subject-independent evaluation. First, subjective emotional rating scores after watching emotional film clips could vary from individual to individual. That is, individuals generate different emotions in valence and arousal dimensions when presented with the same stimuli due to their inner psychological characteristics. Second, as a sensitive and real-time physiological model, EEG signals could vary from individual to individual due to their unique internal physiological characteristics. Thus, the emotion-related pattern learned from a participant may not well adapt to another participant. Compared with subject-dependent evaluation, subject-independent evaluation cannot deal with these two types of individual differences.

The cross-corpus accuracies shown in Table 3 are lower than accuracies of the evaluation in each single dataset, indicating that cross-corpus evaluation is a much more difficult task. The reasons may be two-fold. First, the two datasets used different types of audio-visual stimuli (i.e., DEAP used music videos while MAHNOB-HCI used movies) to elicit target emotions. These materials are different in multimedia contents (i.e., color, lightness, MFCC), which may elicit different emotional valence and intensity. Second, sophisticated experiment settings of data collection, such as instruction or distraction task between two successive emotional materials, are hard to be controlled at the same level. Hence, it is less possible that the EEG signals collected from these two datasets to share similar emotion-related patterns.

Table 5 summarized the features that achieve the highest accuracies among the five methods on four datasets, i.e., the features that obtain the highest accuracies shown as bold in Table 4. Among different frequency bands, γ band achieves the best recognition performance in most cases. This finding is consistent with previous studies, which showed that γ band is effective in EEG-based emotion classification [54], [55], [56]. It had also been proven that γ band is closely associated with emotional processing [27], [57]. For example, event-related γ band activities were associated with processing of emotional facial expressions [58] and stronger inter-hemispheric γ interval was found during the processing of unpleasant pictures [59].

Among all feature types, DE, PSD and ASM showed consistent and prominent contribution in emotion classification across various tasks. The neurophysiological basis of these features for emotion processing has already been well documented [6], [60], [61], [62]. DE is able to discriminate EEG

TABLE 5
The Features That Achieve the Highest Accuracies (Bold in Table 4) Among Five Methods on Four Datasets

Dataset	Subject-dependent		Subject-independent	
	valence	arousal	valence	arousal
SEED	DE+PSD	N/A	DE_γ	N/A
DEAP	PSD_all	ASM_γ	ASM_γ	PSD_γ
DREAMER	DE+PSD	DE+PSD	PSD_α	DE_β
CMEED	DE+PSD	DE_all	PSD_γ	PSD_γ

pattern between low and high frequency energies [15]. PSD has the merit of capturing a dynamic and multidimensional space of brain processing [63] and has been widely utilized in EEG-based emotion recognition as a sensitive and effective feature [64], [65]. As for ASM, a typical indicator of asymmetric brain activity in the frontal cortex has been reported to be sensitive to the valence as well as motivational direction of emotional states [17]. Specifically, it was effective in distinguishing emotional states (that vary from negative to positive [66]) and motivational direction (approach or withdrawal) [67].

Adjacency matrices learned in cross-corpus evaluation were further investigated. Weights of all edges for each node were summed up and sorted descendingly. Then the top 5 nodes with the highest sum values were recorded for every feature. We counted how many times a node was recorded for all features. The 5 electrodes with the largest recorded numbers, i.e., the nodes with the strongest associations with other nodes, were FC1, FC5, FC2, CP1 and CP5. These electrodes are located in the frontal-central and central-parietal lobes, which are known as related to emotion processing [18], [68]. Since sensory activities are mainly related to parietal lobe [69], we concluded that our model learned to characterize affective activities in brain rather than sensory activities caused by stimuli.

Several avenues exist for future work. First, although our work utilizes sparseness constraint that successfully reveals more discriminative emotional information and improves the recognition accuracy, Torfi *et al.* [70] found that applying excessive sparsity to neural networks may decrease the accuracy due to overmuch elimination of important elements. They proposed a method using attention mechanism to supervise the sparsity operation, in which a decoder determined to which the attention should be paid. In future work, we plan to investigate the role of attention mechanism and further guide our sparseness constraint in an appropriate level. Second, sparse DGCNN follows the original DGCNN [31] to model the adjacency matrix W using non-negative entries, since the output of the ReLu activation layer is non-negative. In future work, it is interesting to consider negative values due to the possible negative correlations found in emotional face processing (e.g., [71]). Third, the datasets of DEAP, DREAMER and CMEED used self-reported labels, while SEED used the categories of film clip as labels. These two types of labels are known as corresponding to felt and perceived emotions, respectively (e.g., [72]). Our sparse DGCNN can handle both types of labels. In future work, it is possible to improve the recognition accuracy by exploring the differences in these two label types.

6 Conclusion

Previous EEG studies on affective computing put emphasis on the analysis of individual EEG electrodes. Recently, researches such as the DGCNN method [31] paid attention to the strength of functional relations between each pair of EEG electrodes. In this paper, we introduce a sparseness constraint into the adjacent matrix W and propose a sparse DGCNN model. To efficiently train this model, we apply a novel solution to a constrained minimization problem. We show that compared with the existing recognition models including SVM, DBN, GELM and original DGCNN model, both the recognition accuracy and scalability can be significantly improved by the sparse DGCNN. Specifically, the recognition accuracy of sparse DGCNN is averagely 8.88 percent higher than all the other four methods consistently on SEED, DREAMER, DEAP and CMEED datasets under subject-dependent and subject-independent protocols. Our work also shows that DE, PSD and ASM features on γ band convey most important discriminative emotional information, and fusion of separate features and frequency bands can improve recognition performance, which are consistent with pervious findings.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China under Grants U1736220, 61725204, 62076064, 31771226. Guanhua Zhang and Minjing Yu are joint first authors.

References

[1] K. R. Scherer, "Toward a dynamic theory of emotion," *Geneva Stud. Emotion*, vol. 1, pp. 1–96, 1987.
[2] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis, "Toward emotion aware computing: An integrated approach using multichannel neurophysiological recordings and affective visual stimuli," *IEEE Trans. Inf. Technol. Biomedicine*, vol. 14, no. 3, pp. 589–597, May 2010.
[3] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, "Real-time movie-induced discrete emotion recognition from EEG signals," *IEEE Trans. Affective Comput.*, vol. 9, no. 4, pp. 550–562, Fourth Quarter 2018.
[4] S. Koelstra *et al.*, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, Firstquarter 2012.
[5] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," in *Proc. Int. Conf. Learn. Representations*, 2016, pp. 1–15.
[6] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affective Comput.*, vol. 10, no. 3, pp. 374–393, Third Quarter 2019.
[7] X. Hu, J. Chen, F. Wang, and D. Zhang, "Ten challenges for EEG-based affective computing," *Brain Sci. Advances*, vol. 5, no. 1, pp. 1–20, 2019.
[8] M. M. Bradley and P. J. Lang, "Affective norms for English text (ANET): Affective ratings of text and instruction manual," Univ. Florida, Gainesville, FL, Tech. Rep. D-1, 2007.
[9] M. M. Bradley and P. J. Lang, "The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual," Univ. Florida, Gainesville, FL, Tech. Report B-3, 2007.
[10] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," Univ. Florida, Gainesville, FL, Tech. Rep. A-8, 2008.
[11] E. S. Dan-Glauser and K. R. Scherer, "The geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance," *Behav. Res. Methods*, vol. 43, pp. 468–477, 2011.
[12] Y. Baveye, E. Dellandréa, C. Chamaret, and L. Chen, "LIRIS-ACCEDE: A video database for affective content analysis," *IEEE Trans. Affective Comput.*, vol. 6, no. 1, pp. 43–55, First Quarter 2015.
[13] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "MPED: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12 177–12 191, 2019.
[14] L. M. Jenkins and D. G. Andrewes, "A new set of standardised verbal and non-verbal contemporary film stimuli for the elicitation of emotions," *Brain Impairment*, vol. 13, no. 2, pp. 212–227, 2012.
[15] W. Zheng and B. 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.
[16] 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.

[17] G. Zhao, Y. Zhang, Y. Ge, Y. Zheng, X. Sun, and K. Zhang, "Asymmetric hemisphere activation in tenderness: Evidence from EEG signals," *Sci. Rep.*, vol. 8, no. 1, pp. 1–9, 2018.
[18] G. Zhao, Y. Zhang, and Y. Ge, "Frontal EEG asymmetry and middle line power difference in discrete emotions," *Front. Behav. Neurosci.*, vol. 12, 2018, Art. no. 225.
[19] P. Ekman, *Basic Emotions*. Hoboken, NJ, USA: Wiley, 1999.
[20] R. Plutchik, "The nature of emotions," *Amer. Scientist*, vol. 89, 2001, Art. no. 344.
[21] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Develop. Psychopatho.*, vol. 17, no. 3, pp. 715–734, 2005.
[22] M. Cohen, "Where does EEG come from and what does it mean?" *Trends Neurosci.*, vol. 40, no. 4, pp. 208–218, 2017.
[23] E. Bullmore *et al.*, "Fractal analysis of electroencephalographic signals intracerebrally recorded during 35 epileptic seizures: Evaluation of a new method for synoptic visualisation of ictal events," *Electroencephalogr. Clin. Neurophysiol.*, vol. 91, no. 5, pp. 337–345, 1994.
[24] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. Biomed. Eng.*, vol. 14, no. 2, pp. 186–197, Mar. 2010.
[25] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Hoboken, NJ, USA: Wiley, 2007.
[26] Y.-P. Lin *et al.*, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.
[27] M. M. Müller, A. Keil, T. Gruber, and T. Elbert, "Processing of affective pictures modulates right-hemispheric gamma band EEG activity," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1913–1920, 1999.
[28] J. Kortelainen, E. Väyrynen, and T. Seppänen, "High-frequency electroencephalographic activity in left temporal area is associated with pleasant emotion induced by video clips," *Comput. Intell. Neurosci.*, vol. 2015, 2015, Art. no. 31.
[29] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG features in cross-subject emotion recognition," *Front. Neurosci.*, vol. 12, 2018, Art. no. 162.
[30] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, and X. Zhou, "A bihemisphere domain adversarial neural network model for EEG emotion recognition," *IEEE Trans. Affective Comput.*, to be published, doi: 10.1109/TAFFC.2018.2885474.
[31] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affective Comput.*, vol. 11, no. 3, pp. 532–541, Third Quarter 2020.
[32] 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.
[33] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," 2014, *arXiv:1312.6203*.
[34] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3837–3845.
[35] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI, USA: Amer. Math. Soc., 1997.
[36] D. K. Duvenaud *et al.*, "Convolutional networks on graphs for learning molecular fingerprints," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 2224–2232.
[37] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Representations*, 2017, pp. 1–14.
[38] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 260–291, 2016.
[39] H. Behjat, U. Richter, D. V. D. Ville, and L. Sörnmo, "Signal-adapted tight frames on graphs," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 6017–6029, Nov. 2016.
[40] L. A. Stankovic, M. Dakovic, and E. Sejdic, "Vertex-frequency analysis: A way to localize graph spectral components," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 176–182, Jul. 2017.
[41] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
[42] M. Bola and V. Borchardt, "Cognitive processing involves dynamic reorganization of the whole-brain network's functional community structure," *J. Neurosci.*, vol. 36, no. 13, pp. 3633–3635, 2016.
[43] L. Pessoa, "A network model of the emotional brain," *Trends Cogn. Sci.*, vol. 21, no. 5, pp. 357–371, 2017.
[44] P. L. Nunez, *Neocortical Dynamics and Human EEG Rhythms*. New York, NY, USA: Oxford Univ. Press, 1995.
[45] A. Fornito, A. Zalesky, and M. Breakspear, "The connectomics of brain disorders," *Nat. Rev. Neurosci.*, vol. 16, no. 3, pp. 159–172, 2015.
[46] P. Krishnaswamy *et al.*, "Sparsity enables estimation of both subcortical and cortical activity from MEG and EEG," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 48, pp. E10 465–E10 474, 2017.
[47] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
[48] H. Zhou, J. M. Alvarez, and F. Porikli, "Less is more: Towards compact CNNs," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 662–677.
[49] J. C. Duchi and Y. Singer, "Efficient learning using forward-backward splitting," in *Proc. 22nd Int. Conf. Neural Inf. Process. Syst.*, 2009, pp. 495–503.
[50] N. Parikh *et al.*, "Proximal algorithms," *Found. Trends® Optim.*, vol. 1, no. 3, pp. 127–239, 2014.
[51] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affective Comput.*, vol. 10, no. 3, pp. 417–429, Third Quarter 2019.
[52] Y. Peng, S. Wang, X. Long, and B.-L. Lu, "Discriminative graph regularized extreme learning machine and its application to face recognition," *Neurocomputing*, vol. 149, pp. 340–353, 2015.
[53] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 42–55, First Quarter 2012.
[54] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 1223–1226.
[55] L. Aftanas, L. Savotina, V. Makhnev, and N. Reva, "Analysis of evoked EEG synchronization and desynchronization during perception of emotiogenic stimuli: Association with autonomic activation processes," *Neurosci. Behavioral Physiol.*, vol. 35, no. 9, pp. 951–957, 2005.
[56] J.-Y. Zhu, W.-L. Zheng, Y. Peng, R.-N. Duan, and B.-L. Lu, "EEG-based emotion recognition using discriminative graph regularized extreme learning machine," in *Proc. Int. Joint Conf. Neural Netw.*, 2014, pp. 525–532.
[57] A. Keil, M. M. Müller, T. Gruber, C. Wienbruch, M. Stolarova, and T. Elbert, "Effects of emotional arousal in the cerebral hemispheres: A study of oscillatory brain activity and event-related potentials," *Clin. Neurophysiol.*, vol. 112, no. 11, pp. 2057–2068, 2001.
[58] M. Balconi and C. Lucchiari, "Consciousness and arousal effects on emotional face processing as revealed by brain oscillations. A gamma band analysis," *Int. J. Psychophysiol.*, vol. 67, no. 1, pp. 41–46, 2008.
[59] N. Martini *et al.*, "The dynamics of EEG gamma responses to unpleasant visual stimuli: From local activity to functional connectivity," *NeuroImage*, vol. 60, no. 2, pp. 922–932, 2012.
[60] R. J. Davidson, "Affective neuroscience and psychophysiology: Toward a synthesis," *Psychophysiology*, vol. 40, no. 5, pp. 655–665, 2003.
[61] W. J. Ray and H. W. Cole, "EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes," *Science*, vol. 228, no. 4700, pp. 750–752, 1985.
[62] W. Klimesch, M. Doppelmayr, H. Russegger, T. Pachinger, and J. Schwaiger, "Induced alpha band power changes in the human EEG and attention," *Neurosci. Lett.*, vol. 244, no. 2, pp. 73–76, 1998.
[63] F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie, "The brainweb: Phase synchronization and large-scale integration," *Nat. Rev. Neurosci.*, vol. 2, no. 4, pp. 229–239, 2001.
[64] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. Müller-Putz, "Unsupervised feature learning for EEG-based emotion recognition," in *Proc. Int. Conf. Cyberworlds*, 2017, pp. 182–185.
[65] R. E. Yohanes, W. Ser, and G.-B. Huang, "Discrete wavelet transform coefficients for emotion recognition from EEG signals," in *Proc. Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2012, pp. 2251–2254.
[66] B. Schmidt and S. Hanslmayr, "Resting frontal EEG alpha-asymmetry predicts the evaluation of affective musical stimuli," *Neurosci. Lett.*, vol. 460, no. 3, pp. 237–240, 2009.
[67] E. Harmon-Jones and P. A. Gable, "On the role of asymmetric frontal cortical activity in approach and withdrawal motivation: An updated review of the evidence," *Psychophysiology*, vol. 55, no. 1, 2018, Art. no. e12879.

[68] D. M. Tucker and S. L. Dawson, "Asymmetric EEG changes as method actors generated emotions," *Biol. Psychol.*, vol. 19, no. 1, pp. 63–75, 1984.
[69] J. W. Bisley and M. E. Goldberg, "Attention, intention, and priority in the parietal lobe," *Annu. Rev. Neurosci.*, vol. 33, pp. 1–21, 2010.
[70] A. Torfi, R. A. Shirvani, S. Soleymani, and N. M. Nasrabadi, "GASL: Guided attention for sparsity learning in deep neural networks," 2019, *arXiv: 1901.01939*.
[71] V. Soto, J. Tyson-Carr, K. Kokmotou, H. Roberts, S. Cook, N. Fallon, T. Giesbrecht, and A. Stancak, "Brain responses to emotional faces in natural settings: A wireless mobile EEG recording study," *Front. Psychol.*, vol. 9, p. 2003, 2018.
[72] K. Kallinen and N. Ravaja, "Emotion perceived and emotion felt: Same and different," *Music. Sci.*, vol. 10, no. 2, pp. 191–213, 2006.

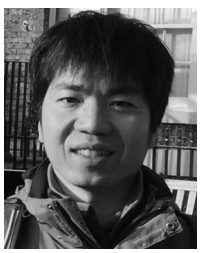
Guanhua Zhang received the BEng degree from the Beijing University of Posts and Telecommunications, China, China, in 2017, and the master's degree from the Department of Computer Science and Technology, Tsinghua University, China, in 2020. She is currently working toward the PhD degree with the Institute for Visualisation and Interactive Systems, University of Stuttgart, Germany. Her research interests include affective computing and machine learning.

Minjing Yu received the BE degree from Wuhan University, Wuhan, China, in 2014, and the PhD degree from Tsinghua University, Beijing, China, in 2019. She is currently an assistant professor with the College of Intelligence and Computing, Tianjin University, China. Her research interests include cognitive computation, artificial intelligence and computer graphics.

Yong-Jin Liu (Senior Member, IEEE) received the BEng degree in mechano-electronic engineering from Tianjin University, China, in 1998, and the PhD degree from the Hong Kong University of Science and Technology, Hong Kong, in 2003. He is currently a professor with the Department of Computer Science and Technology, Tsinghua University, China. His research interests include intelligent media processing, pattern analysis and human-computer interaction. For more information please visit: https://cg.cs.tsinghua.edu.cn/people/~Yongjin/Yongjin.htm.

Guozhen Zhao received the BS degree in industrial engineering, from Tianjin University, China, in 2007, and the MS and PhD degrees in industrial and systems engineering from the State University of New York, Buffalo, New York, in 2009 and 2011, respectively. He is currently an associate professor with the Institute of Psychology, Chinese Academy of Sciences, China. His current research interests include the mathematical modeling of human cognition and performance, emotion recognition, emotional interaction and augmentation.

Dan Zhang (Member, IEEE) received the BE degree in automation and the PhD degree in biomedical engineering both from Tsinghua University, China, in 2005 and 2011, respectively. He is currently an associate professor with the Department of Psychology, Tsinghua University, China. He was a postdoctoral fellow with the School of Medicine, Tsinghua University, China from 2011 to 2013. His research interests include social neuroscience, engineering psychology, and brain-computer interfaces. He is an associate editor of the *IEEE Transactions on Affective Computing*.



Wenming Zheng (Senior Member, IEEE) received the PhD degree in signal and information processing from Department of Radio, Southeast University, China, in 2004. He is currently a professor and the director of Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Science & Medical Engineering, Southeast University (SEU), China. His research interests include affective computing, neural computation, pattern recognition, machine learning, and computer vision. He is an associate editor of the *IEEE Transactions on Affective Computing*.