Computers in Biology and Medicine 123 (2020) 103927

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network

Yu Liu [a], Yufeng Ding [a], Chang Li [a,*], Juan Cheng [a], Rencheng Song [a], Feng Wan [b], Xun Chen [c]

[a] *Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China*
[b] *Department of Electrical and Computer Engineering, University of Macau, Macau, China*
[c] *Department of Electronic Science and Technology, University of Science and Technology of China, Hefei 230027, China*

ARTICLE INFO

Keywords:
Deep learning
Electroencephalogram (EEG)
Emotion recognition
Capsule network

ABSTRACT

In recent years, deep learning (DL) techniques, and in particular convolutional neural networks (CNNs), have shown great potential in electroencephalograph (EEG)-based emotion recognition. However, existing CNN-based EEG emotion recognition methods usually require a relatively complex stage of feature pre-extraction. More importantly, the CNNs cannot well characterize the intrinsic relationship among the different channels of EEG signals, which is essentially a crucial clue for the recognition of emotion. In this paper, we propose an effective multi-level features guided capsule network (MLF-CapsNet) for multi-channel EEG-based emotion recognition to overcome these issues. The MLF-CapsNet is an end-to-end framework, which can simultaneously extract features from the raw EEG signals and determine the emotional states. Compared with original CapsNet, it incorporates multi-level feature maps learned by different layers in forming the primary capsules so that the capability of feature representation can be enhanced. In addition, it uses a bottleneck layer to reduce the amount of parameters and accelerate the speed of calculation. Our method achieves the average accuracy of 97.97%, 98.31% and 98.32% on valence, arousal and dominance of DEAP dataset, respectively, and 94.59%, 95.26% and 95.13% on valence, arousal and dominance of DREAMER dataset, respectively. These results show that our method exhibits higher accuracy than the state-of-the-art methods.

1. Introduction

Emotion is a psychological state that affects people's cognition, decision-making and behavior [1]. At present, affective computing plays a crucial role in artificial intelligence, particularly in the field of human–computer interaction. An artificial machine with the ability to analyze human emotion can better understand human beings so as to better meet human needs.

In general, human emotion identification are based on either non-physiological signals, such as facial expressions [2], body gestures [3], speech [4], or physiological signals, such as electrocardiogram (ECG) [5], electroencephalograph (EEG) [6], and electromyogram (EMG) [7]. Compared with non-physiological signals, physiological signals spontaneously produced by human body are not susceptible to the impact of subjective will, providing a reliable way for emotion recognition. From the view of neuroscience [8], there are some major brain cortex regions closely related with emotion, such as the orbital frontal cortex, ventral medial prefrontal cortex and amygdala [9]. Thus, among the various types of physiological signals, EEG signal has the advantage of reflecting the emotional states of human beings. Furthermore, with the rapid development of the techniques for collecting EEG signal, it is becoming more and more convenient to collect EEG signal. Therefore, EEG signal has been widely used for emotion recognition and achieved satisfactory results.

The process of EEG-based emotion recognition can be divided into two main stages. The first stage is extracting feature from EEG signals to effectively represent emotional state. EEG features can be extracted either from time domain or frequency domain. The time domain features mainly capture the temporal information of EEG signals, such as Hjorth feature [11], fractal dimension feature [12] and higher order crossing feature [13]. The frequency domain features aim to capture the EEG emotion information from the frequency domain, such as the power spectral density (PSD) feature, differential entropy (DE) feature [14], the rational asymmetry (RASM) feature [15] and the differential caudality (DCAU) feature [16], *etc.* The second stage is designing classifiers to predict the emotional labels according to those extracted features. Various machine learning algorithms have been used as classifiers for EEG-based emotion recognition with satisfactory accuracy, such as support vector machine (SVM) [16], k-nearest neighbors (k-NN) [17],

[*] Corresponding author.
E-mail addresses: yuliu@hfut.edu.cn (Y. Liu), changli@hfut.edu.cn (C. Li).

https://doi.org/10.1016/j.compbiomed.2020.103927
Received 28 March 2020; Received in revised form 19 July 2020; Accepted 19 July 2020
Available online 22 July 2020

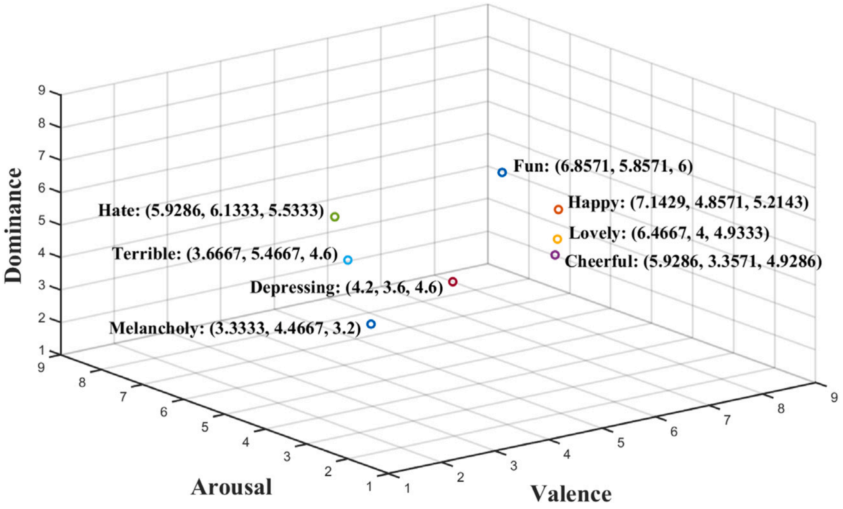

Fig. 1. Emotion distribution in Valence-Arousal-Dominance space [10].

linear discriminant analysis (LDA) [18], random forest, Naive Bayes (NB), *etc.*

In recent years, owing to the powerful ability of feature extraction, deep learning (DL) algorithms have achieved state-of-the-art performance in the fields of computer vision [19,20], natural language processing [21,22], speech recognition [23], as well as EEG-based emotion recognition [24–27]. Models that have been applied to EEG-based emotion recognition include deep brief networks (DBNs) [24], convolutional neural networks (CNNs) [25,26], graph convolutional neural networks (GCNN) [27], capsule network (CapsNet) [28], *etc.* Alhagry et al. proposed a 4-layer deep learning framework based on long-short term memory (LSTM) to learn features from raw 32-channel EEG signals, and then adopted the dense layer to classify these features into low/high arousal or low/high valence. This method achieved average accuracy of 85.65% and 85.45% for arousal and valence on DEAP dataset, respectively [29]. Tripathi et al. incorporated contemporary techniques (*e.g.*, dropout and rectilinear units) into a 2D CNN to effectively classify preprocessed 2D EEG data. They got classification accuracy of 81.41% and 73.35% for valence and arousal on DEAP dataset, respectively [26]. Chao et al. combined PSD features with spatial characteristics of original EEG signals to construct multiband feature matrix (MFM), and then utilized CapsNet as a classifier to determine the EEG emotional states from the MFM. They achieved average accuracy of 68.28% and 66.73% for arousal and valence on DEAP dataset, respectively [28].

Despite the great progress that has been achieved, there still exist some challenges in the study of DL-based EEG emotion recognition. First, many existing DL-based EEG emotion recognition methods have a feature pre-extraction stage for the raw EEG signals before network computing, such as the works mentioned above [24,27,28]. However, some feature extraction approaches require manual operation, and this stage is actually not in good accord with the data-driven principle of deep learning. Second, CNN has currently been one of the most popular DL models for EEG-based emotion recognition, while it typically needs a large-scale annotated dataset to achieve high performance. Unfortunately, unlike the problems in the field of computer vision (*e.g.*, natural image classification), it is very difficult to collect and annotate “sufficient” EEG signals for emotion recognition, especially for the subject-dependent case where the training and test data are both from the same subject. Besides, it is demonstrated that CNN cannot characterize the spatial relationship among different features well [30]. In multi-channel EEG-based emotion recognition, the intrinsic relationship among different channels is a crucial clue for identifying the emotional states [27,31]. There is a strong correlation between emotional states and brain functional connectivity patterns. Specifically, different emotional states cause different activities in various functional areas of the brain, and specific connections among different functional regions [32,33]. However, the intrinsic relationship among different channels may be neglected by the CNN models.

To address the above challenges, we propose an end-to-end deep learning framework for multi-channel EEG-based emotion recognition based on the capsule network incorporating multi-level features, namely, MLF-CapsNet, which is fully data-driven without any feature pre-extraction stage. The MLF-CapsNet can be well trained with a much smaller scale of training data in comparison to CNNs, so it is very suitable for the EEG-based emotion recognition problem, especially for the subject-dependent task. Unlike convolutional neural networks, the MLF-CapsNet has strong ability to identify the positional relationship among local features in the spatial domain, which is beneficial to improving the classification accuracy of emotion. Moreover, the original CapsNet has limitation on some classification tasks, where the target objects have complex internal representations [30]. Thus, we incorporate multi-level feature maps learned by different convolution layers in forming the primary capsules, which is more efficient for feature representation in multi-channel EEG-based emotion recognition. In addition, in order to reduce the amount of parameters and accelerate the speed of calculation, we use a bottleneck layer to reduce the number of channel of concatenated feature maps. Experimental results on the popular DEAP [34] and DREAMER [35] datasets show that the proposed method can significantly outperform some state-of-the-art DL-based methods on subject-dependent EEG-based emotion recognition task. The main contributions of this paper can be summarized as follows:

1. We propose a DL framework *i.e.*, MLF-CapsNet for multi-channel EEG emotion recognition. The proposed MLF-CapsNet is an end-to-end framework, which can extract features from the raw EEG signals and determine the emotional states simultaneously. More importantly, it can effectively characterize the intrinsic relationship among various EEG channels due to its innovative structure.
2. In comparison to the original CapsNet, the MLF-CapsNet can combine multi-level features extracted from different convolution layers to form primary capsules, which can enhance the capacity of representation of capsule network. In addition, we add a bottleneck layer to reduce the amount of parameters and accelerate the speed of calculation.
3. We conduct experiments on two public datasets, *i.e.*, DEAP and DREAMER, for subject-dependent EEG-based emotion recognition. The proposed method achieves state-of-the-art performance

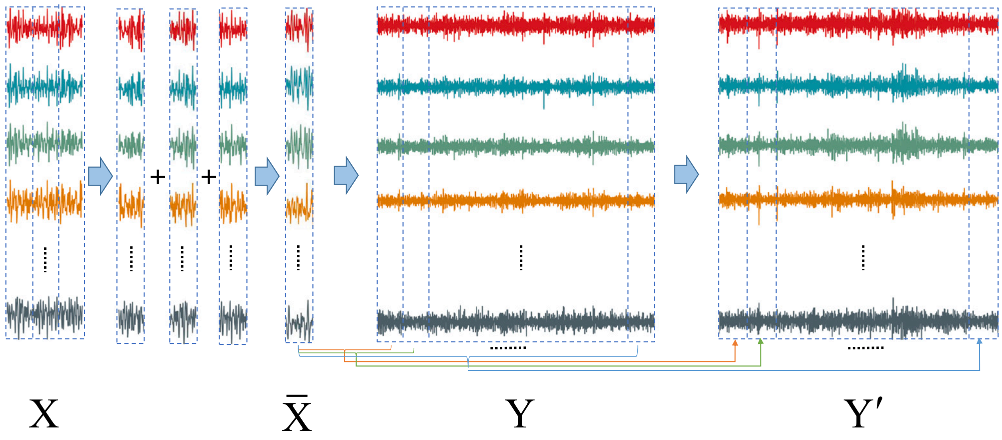

Fig. 2. Schematic diagram of preprocessing.

on both datasets. For DEAP, it achieves the mean accuracy of 97.97%, 98.31% and 98.32% for valence, arousal and dominance respectively. For DREAMER, it achieves the mean accuracy of 94.59%, 95.26% and 95.13% for valence, arousal and dominance, respectively. The performance on both two datasets shows significant improvement when compared with the performance of some recently proposed DL-based methods.

This paper is organized as follows. In Section 2, related works are described. Section 3 presents the proposed MLF-CapsNet-based emotion recognition method. Section 4 illustrates experiments conducted on DEAP and DREAMER datasets. Section 5 gives the discussion of the experimental results. Finally, we conclude this paper in Section 6.

2. Related works

In this section, we first introduce dimensional model to describe emotion, namely, valence-arousal-dominance model. Then, we introduce the theory and several applications of the CapsNet.

2.1. Emotion model

Emotion is traditionally classified by two basic models, *i.e.*, the discrete model and dimensional model [36]. Discrete theories of emotion propose that the existence of separate emotions are characterized by coordinated response patterns in physiology, neural anatomy, and morphological expressions. Happiness, sadness, anger, disgust, fear, and surprise are six basic emotions that are specified by discrete model to describe emotion [37,38]. However, dimensional theories believe that emotion is not discrete but continuous, and can be described by multiple consecutive values. The most famous dimensional theory is the three-dimensional space, namely valence, arousal and dominance [10]. The valence refers to the degree of human pleasure from positive to negative, the arousal characterizes the level of excitement from passive to active, and the dominance ranges from a helpless and weak feeling (without control) to an empowered feeling (in control of everything). Valence-arousal-dominance model has been widely used in EEG-based emotion recognition due to its simplicity and ability to describe emotion well [10]. Fig. 1 shows emotion distribution in valence-arousal-dominance space. This model adopts a three-dimensional coordinate system to express emotion, in which the emotional states are determined by the continuous values of arousal, valence and dominance. For example, the values of “Happy” in three dimensions are 7.1429, 4.8571 and 5.2143, respectively.

2.2. Capsule network

Neural networks based on CNN have achieved superior performance in computer vision tasks such as classification, object detection, and semantic segmentation. However, CNN still has a shortcoming caused by pooling operation. Pooling is a downsampling of the feature maps learned by convolution kernels, which can reduce the computational complexity, and deal with the changes in images caused by changes in viewpoint. Unfortunately, the benefits of pooling are at the expense of discarding precise spatial relationships between high-level parts, which is crucial for some recognition tasks. For example, face recognition requires precise spatial relationships between the five facial organs to identify human face correctly [39]. To overcome the shortcomings of CNN, a network called the capsule network (CapsNet) has been proposed [30]. The CapsNet can represent the relative spatial relationship between local parts and the whole object. The core units that make up the CapsNet are called capsules. Capsules are locally invariant groups of neurons, which learn to recognize the presence of visual entities and encode their properties into vector. The length of vector (between zero and one) represents the presence of the entity, and the orientation represents the instantiation parameters. Innovatively, the capsules of different layers are connected through an iterative routing-by-agreement mechanism: a lower-level capsule prefers to send its output to higher-level capsules whose activity vectors have a big scalar product with the prediction coming from the lower-level capsules. Furthermore, in the CapsNet, transformation matrices are used to encode the intrinsic spatial relationship between a part and a whole of an object so that the CapsNet can overcome the shortcomings of CNN caused by pooling. Due to these innovative ideas, the CapsNet is more robust to translation and rotation, and is considerably better at recognizing highly overlapping digits.

Considering the above mentioned advantages, the CapsNet has been applied to many fields in the past two years, such as natural language processing [40], medical image classification [41], hyperspectral image classification [42] and speech recognition [43], and has achieved superior performance. Wang et al. proposed the aspect-level sentiment model based on the CapsNet, which is capable of performing aspect detection and sentiment classification simultaneously [40]. Afshar et al. exploited the CapsNet to determine the correct type of brain tumor captured by Magnetic Resonance Imaging (MRI) [41]. Yin et al. proposed a new CapsNet-based architecture with three convolutional layers to adjust the CapsNet to hyperspectral image classification, achieving significantly superior performance in hyperspectral image classification [42]. Turan et al. extracted spectrogram representations from the short segments of an audio signals, and then used the CapsNet to recognize an infant’s cry [43].

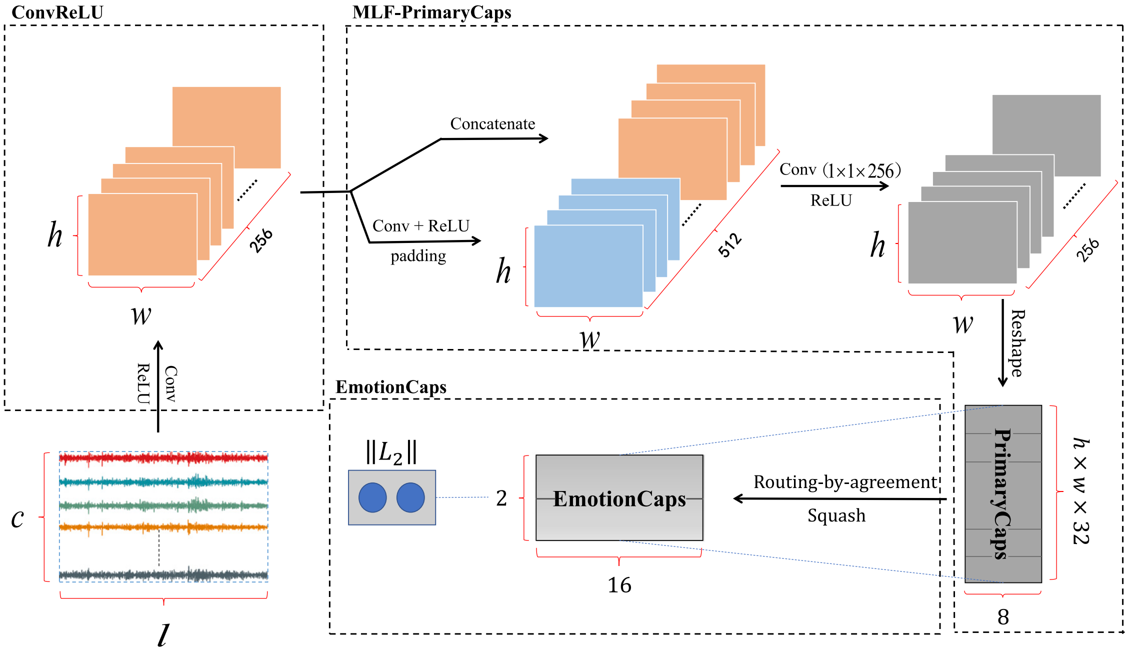

Fig. 3. The proposed MLF-CapsNet-based EEG emotion recognition framework.

3. The proposed method

In this section, we firstly introduce the preprocessing of raw EEG data, and then present the details of the proposed MLF-CapsNet-based emotion recognition method.

3.1. Preprocessing

Fig. 2 shows the process of preprocessing. For the sake of fair comparison with some recently proposed DL-based methods which provide source code for reproducible research, we conduct exactly the same data preprocessing approach as these previous works [25,44]. Each recording EEG signal contains a baseline signal recorded in relax state and an experimental signal recorded under stimulation. Let $\mathbf{X} \in \mathbb{R}^{C \times N_1}$ denote a baseline signal and $\mathbf{Y} \in \mathbb{R}^{C \times N_2}$ denote corresponding experimental signal. C is the number of electrode nodes. N_1 and N_2 denote the number of sampling points of $\mathbf{X}$ and $\mathbf{Y}$, respectively. First, we cut the baseline signal into M_1 segments with the same length L, each of which is recorded as X_i $(i = 1, 2, \dots, M_1) \in \mathbb{R}^{C \times L}$. Second, we do the element-wise addition for all of these segments, and calculate the mean value to get $\overline{\mathbf{X}}$, which is used to represent subjects' basic emotional state without any stimulation. This step can be formulated as:

$$\overline{\mathbf{X}} = \frac{\sum_{i=1}^{M_1} X_i}{M_1}. \tag{1}$$

Besides, the same segmentation is also used to cut experimental signal $\mathbf{Y}$. Therefore, we can get M_2 matrices, each of which is denoted as Y_j $(j = 1, 2, \dots, M_2) \in \mathbb{R}^{C \times L}$. Then, we subtract $\overline{\mathbf{X}}$ from Y_j to obtain Y_j'. This step can be formulated as:

$$Y_j' = Y_j - \overline{\mathbf{X}}. \tag{2}$$

After the above steps, we concatenate all of these Y_j' into a matrix recorded as $\mathbf{Y}'$, of which the size is the same as that of raw experimental EEG signals $\mathbf{Y}$.

Usually, DL-based emotion recognition methods require a large number of labeled EEG data to train the model. Therefore, most DL-based works divide the experimental signals into segments to increase the number of EEG samples. Specifically, according to the analysis results reported in [45], one second is the most suitable window length for emotion recognition. As a result, in this paper, we adopt one second slide window to cut the baseline signals $\mathbf{X}$ and the experimental signals $\mathbf{Y}$. Then, we also use one second slide window to segment the preprocessed signals $\mathbf{Y}'$. Each segment derived from $\mathbf{Y}'$ is regarded as a sample, which inherits the labels of the original experimental signals.

3.2. MLf-CapsNet-based emotion recognition

The proposed EEG emotion recognition framework based on MLF-CapsNet has three modules, namely, ConvReLU, multi-level features guided PrimaryCaps (MLF-PrimaryCaps) and EmotionCaps. The details are illustrated in Fig. 3.

ConvReLU is a convolutional layer, which has 256 convolutional kernels with a stride of 2 and ReLU activation. The size of these convolutional kernels is determined by the shape of input. The difference in the number of channels between the two datasets results in different size of convolutional kernels, and we adopt 9×9 for DEAP and 6×6 for DREAMER. This layer converts the value of sample points to the activities of local feature detectors, which are then used as inputs to the MLF-PrimaryCaps.

The MLF-PrimaryCaps is a convolutional capsule layer with 32 channels of convolutional 8D capsules (in other words each primary capsule contains 8 convolutional units with a 9×9 or 6×6 filter and a stride of 1). In total, the MLF-PrimaryCaps has k_1 $(k_1 = 32 \times h \times w)$ capsule outputs (each output is an 8D vector). The length and direction of each primary capsule represent the presence and property of the low-level features related to emotional states, respectively. In this module, we incorporate multi-level feature maps learned by different layers in forming the primary capsules so that primary capsules can contain more information. In addition, we add a bottleneck layer to reduce the amount of parameters and accelerate the speed of calculation. In the following, we provide the details of this module. First, convolution is employed to extract deeper features from the output of upper layer. The size of convolutional kernel is the same as that of ConvReLU layer. However, different from ConvReLU layer, we use stride of 1 and padding to ensure that the produced feature maps have the same size as that of the output of upper layer. Second, we concatenate the two-level features so that we can get 512 feature maps. Third, there is the bottleneck layer. We use 256 convolutional kernels of size 1×1 to

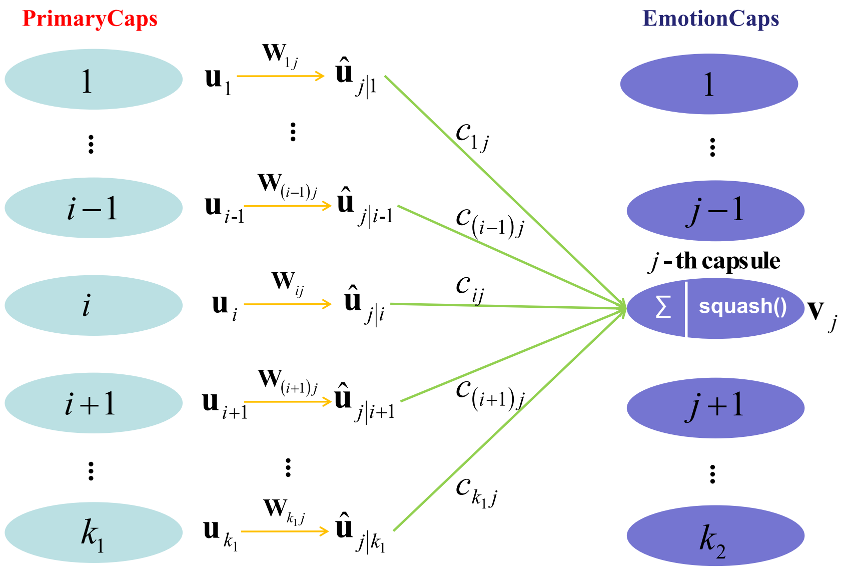

Fig. 4. Routing-by-agreement mechanism.

Table 1
The number of trainable parameters of two methods.

Methods		MLF-CapsNet (w/o)*	MLF-CapsNet
Params (M)	DEAP	200.35	110.85
	DREAMER	78.24	44.15

MLF-CapsNet (w/o)* means the MLF-CapsNet without bottleneck layer.

Table 2
Structural parameters of proposed model.

Modules/Loss	Layers/Routing	Parameters	Shape/Value
Input	Input	–	DEAP: 32×128 DREAMER: 14×128
ConvReLU	Conv2D	Kernel	DEAP: $9 \times 9 \times 256$ DREAMER: $6 \times 6 \times 256$
MLF-PrimaryCaps	Conv2D	Kernel	DEAP: $9 \times 9 \times 256$ DREAMER: $6 \times 6 \times 256$
	Concatenate	–	–
	Bottleneck (Conv2D)	Kernel	$1 \times 1 \times 256$
	Reshape	–	–
EmotionCaps	Dynamic routing	$\mathbf{W}_{ij}$ c_{ij}	8×16 –
Loss	–	m^+ m^-	0.9 0.1

reduce the number of channel of concatenated feature maps from 512 to 256. Last, we group 256 feature maps into 8D capsules. After the above steps, we make the primary capsules contain more information, and enhance the representation capacity of capsules.

The final module is EmotionCaps. Because we use this framework to preform binary classification task, namely low/high valence, low/high arousal, low/high dominance, the EmotionCaps has k_2 ($k_2 = 2$) 16D emotional capsules that correspond to two types of emotional states. The length of the vector of each capsule in EmotionCaps layer indicates presence of an emotional class, and is used to calculate the classification loss. In this sense, a special mechanism has been implemented between MLF-PrimaryCaps and EmotionCaps, known as routing-by-agreement mechanism [30], which connects the current EmotionCaps layer with the previous MLF-PrimaryCaps layer. Its goal is to design better learning process in comparison with traditional pooling methods. This process not only captures part-whole spatial relationship by transformation matrices, but also routes the information between capsules by reinforcing connections of those capsules, which are allocated at different layers and obtain a high grade of agreement. In the following, we provide the details of this process and show it in Fig. 4.

First, we multiply the output $\mathbf{u}_i$ ($i = 1, 2, \ldots, k_1$) of the ith primary capsule by a weight matrix $\mathbf{W}_{ij}$ ($j = 1, 2, \ldots, k_2$) to get "prediction vector" (or high-level emotional feature) $\hat{\mathbf{u}}_{j|i}$. This step can be formulated as:

$$\hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij}\mathbf{u}_i, \tag{3}$$

where $\mathbf{W}_{ij}$ is a transformation matrix between $\mathbf{u}_i$ and $\hat{\mathbf{u}}_{j|i}$. It is used to describe the relative spatial relationship between the low-level emotional features and high-level emotional features.

Second, we sum all $\hat{\mathbf{u}}_{j|i}$ with different weights to obtain $\mathbf{s}_j$. This step can be formulated as:

$$\mathbf{s}_j = \sum_i c_{ij}\hat{\mathbf{u}}_{j|i}, \tag{4}$$

where c_{ij} is coupling coefficient between ith primary capsule and jth emotional capsule. The coupling coefficients between ith primary capsule and all emotional capsules sum to 1. c_{ij} is determined by a "routing softmax" as follows:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}, \tag{5}$$

where initial logit b_{ij} is the log prior probability that ith primary capsule should be coupled to jth emotional capsule.

Third, in order to ensure that the length of the output $\mathbf{v}_j$ of jth emotional capsule is between 0 and 1, a non-linear function called "squashing" is applied to squash $\mathbf{s}_j$. This step can be formulated as:

$$\mathbf{v}_j = \frac{\|\mathbf{s}_j\|^2}{1 + \|\mathbf{s}_j\|^2}\frac{\mathbf{s}_j}{\|\mathbf{s}_j\|}. \tag{6}$$

The initial coupling coefficients are then iteratively refined by measuring the agreement between the current output $\mathbf{v}_j$ and $\hat{\mathbf{u}}_{j|i}$ using the scalar product $\mathbf{v}_j \cdot \hat{\mathbf{u}}_{j|i}$, until reaching the preset maximum number of iterations,

$$b_{ij} \leftarrow b_{ij} + \mathbf{v}_j \cdot \hat{\mathbf{u}}_{j|i}. \tag{7}$$

The coupling coefficients decide how information flows between capsules in MLF-PrimaryCaps and EmotionCaps.

In addition, this model uses a separate margin loss for each emotion capsule. The margin loss L_k for a capsule representing class k is as

Table 3
The format of DEAP Dataset.

Array name	Array shape	Array contents
Data	40 × 40 × 8064	Video/trial × channel × data
Labels	40 × 4	Video/trial × label(valence, arousal, dominance, liking)

Table 4
The format of DREAMER Dataset.

Array name	Array shape	Array contents
ExperData	18 × 14 × 25 472*	Video/trial × channel × data
Baseline	18 × 14 × 7808	Video/trial × channel × data
Labels	18 × 3	Video/trial × label (valence, arousal, dominance)

25 472* means the average length of the trial.

follows:

$$L_k = T_k \max(0, m^+ - \|\mathbf{v}_k\|)^2 + \lambda(1 - T_k \max(0, \|\mathbf{v}_k\| - m^-)^2), \tag{8}$$

where T_k is an indication of the class. $T_k = 1$ if an emotion of class k is present, otherwise $T_k = 0$. m^+ and m^- are used to punish false positives and false negatives, respectively. We set $m^+ = 0.9$ and $m^- = 0.1$. In other words, if class k exists, $\|\mathbf{v}_k\|$ will not be less than 0.9, otherwise, $\|\mathbf{v}_k\|$ will not be greater than 0.1. The λ adjusts the proportion of the loss for absent emotional classes. We use $\lambda = 0.5$, which means punishment for false positives is roughly twice as important as punishment for false negatives. The total loss is simply the sum of the losses of all emotional capsules.

Table 1 shows the number of trainable parameters of the proposed model and the version without bottleneck. It can be seen that bottleneck layer makes our model have significantly less number of parameters. The structural parameters of the proposed model are shown in Table 2.

4. Experiments

In this section, we first introduce two popular public datasets for studying EEG-based emotion recognition and the results of data preprocessing. Then, the implementation details of our model are introduced. Next, we present the verification of preprocessing method. Finally, we compare our method with the state-of-the-art methods, and show the accuracy, training time, and testing time of different methods.

4.1. Datasets

The DEAP and DREAMER datasets have been widely used in the study of EEG-based emotion recognition. We also employ them in this paper to verify the effectiveness of the proposed MLF-CapsNet-based method. The DEAP dataset records 32-channel EEG signals and 8-channel peripheral physiological signals of 32 subjects when they are watching 40 one-minute long music videos. Just like most previous studies [25,26,29,44], only the EEG signals are used for emotion recognition. The DEAP dataset provides a pre-processed version, and the pre-processed version is used in the article. In the pre-processed version, EEG signals are down-sampled to 128 Hz. A bandpass frequency filter from 4.0–45.0 Hz is applied, and the eye artifacts are removed with a blind source separation technique such as independent component analysis (ICA). Each EEG signal contains a 3 s baseline signal recorded in relax state and a 60 s experimental signal recorded under stimulation. Participants rate their levels of arousal, valence, liking and dominance from 1 to 9 after watching each video. The format of DEAP dataset is shown in Table 3.

The DREAMER dataset contains EEG data of 23 subjects (14 males and 9 females), which are collected via 14 EEG electrodes from the subjects when they are watching 18 film clips. Each film clip lasts 65 s to 393 s, which is thought to be sufficient for eliciting single emotion. The average length of film clips is 199 s. The data collection begins with a neutral film clip watching to help the subjects return to the neutral emotion state in each new trial of data collection, and also to serve as the baseline signals. All the EEG signals are recorded at a sampling rate of 128 Hz, and have been filtered by bandpass Hamming sinc linear phase FIR filters. The artifact subspace reconstruction (ASR) method is used for artifacts removal. After watching a film clip, subjects rate their levels of arousal, valence and dominance from 1 to 5. Finally, there are experimental signals (ExperData), baseline signals and labels in DREAMER dataset. The format of DREAMER dataset is described in Table 4.

As most previous studies [25,26,29,44], we adopt the valence-arousal-dominance model for DEAP and DREAMER in our experiments. In these works, thresholds are set to divide each emotion dimension into two categories: low/high valence, low/high arousal and low/high dominance. The thresholds are 5 and 3 for DEAP and DREAMER respectively. For example, in DEAP, the label is low when the rating is less than 5, and the label is high when the rating is greater than or equal to 5. In DREAMER, the label is low when the rating is less than 3, and the label is high when the rating is greater than or equal to 3. In this way, the recognition task is actually a binary classification problem for each emotion dimension. By the way, the dominance scores of the 40 experimental signals of the 27th subject in DEAP dataset are all greater than 5, resulting in labels with only one category of high, and the model trained by such samples is invalid. Therefore, we do not use the samples of the 27th subject to do experiment on dominance of DEAP dataset.

After segmenting the preprocessed experimental signals using 1s sliding window containing 128 sampling points, each signal of DEAP dataset can be divided into 60 segments. Because there are 40 experimental signals for each subject in DEAP, we obtain 2400 (40 trials × 60 segments) EEG samples for each subject. However, the length of each experimental signal in DREAMER datase is different because each film clip lasts from 65 s to 393 s. As a result, we get different number of EEG samples for each experimental signal of DREAMER dataset. But we finally get 3728 EEG samples for each subject in DREAMER, because each subject has the same total length of 18 experimental signals. As a result, every EEG sample in DEAP and DREAMER is a 32×128 matrix and a 14×128 matrix, respectively.

4.2. Implementation details

We adopt 10-fold cross validation [46] to evaluate the performance of our method and comparison methods. Specifically, the average accuracy of the 10-fold cross-validation is taken as the result of one subject, and then the average accuracy of all the subjects are reported as the final accuracy. For our method, we adopt Adam optimizer [47] to minimize the margin loss function, and set maximum number of iteration to 3. For DEAP, we set the learning rate, batch size and number of epochs to 10^{-5}, 100 and 40, respectively. For DREAMER, we set the learning rate, batch size and number of epochs to 10^{-4}, 100 and 30, respectively. We implement our method via the TensorFlow framework [48], and the code is available online.[1]

4.3. Verification of preprocessing method

Yang et al. proposed the preprocessing method of baseline removal and had validated the effectiveness for DT, MLP, CNN-RNN and Cont-CNN on DEAP dataset [25,44]. In order to validate that baseline removal is also suited with our model, we conduct two experiments. The first one is to perform the recognition task without baseline removal, and the other one is to perform the recognition task with baseline removal.

[1] https://github.com/2018110060ding/EmotionCaps.

Table 5
Average accuracy (%) of two experiments on DEAP and DREAMER (mean ± std. dev.).

Datasets	Experiments	Valence	Arousal	Dominance
DEAP	1*	62.57 ± 5.85	64.36 ± 7.96	64.54 ± 10.47
	2**	**97.97** ± **1.67**	**98.31** ± **1.24**	**98.32** ± **1.20**
DREAMER	1*	77.10 ± 6.33	78.39 ± 5.72	77.89 ± 5.64
	2**	**94.59** ± **3.77**	**95.26** ± **3.63**	**95.13** ± **3.81**

1* It means to perform the recognition task without baseline removal.
2** It means to perform the recognition task with baseline removal.

Table 6
Average accuracy (%) of different methods on valence, arousal and dominance classification tasks of DEAP dataset (mean ± std. dev.).

	Valence	Arousal	Dominance
DT	71.63 ± 4.64	73.70 ± 5.01	73.36 ± 7.70
MLP	87.82 ± 6.05	88.68 ± 4.96	88.59 ± 5.95
SVM	88.65 ± 6.18	89.07 ± 5.89	89.13 ± 6.59
Cont-CNN	89.45 ± 4.35	90.24 ± 4.02	90.25 ± 4.87
CNN-RNN	89.92 ± 2.96	90.81 ± 2.94	90.90 ± 3.01
DGCNN	92.55 ± 3.53	93.50 ± 3.93	93.50 ± 3.69
gcForest	97.69 ± 1.22	97.53 ± 1.52	97.62 ± 1.39
CapsNet	**98.22** ± **1.29**	98.05 ± 1.42	**98.44** ± **1.09**
Ours	97.97 ± 1.67	**98.31** ± **1.24**	98.32 ± 1.20

Table 7
Average accuracy (%) of different methods on valence, arousal and dominance classification tasks of DREAMER dataset (mean ± std. dev.).

	Valence	Arousal	Dominance
DT	75.53 ± 6.71	75.74 ± 6.44	76.40 ± 5.68
MLP	83.64 ± 5.97	83.71 ± 5.39	83.90 ± 5.32
SVM	87.14 ± 5.20	87.03 ± 4.88	87.18 ± 4.87
Cont-CNN	84.54 ± 5.00	84.84 ± 4.86	85.05 ± 4.96
CNN-RNN	79.93 ± 6.65	81.48 ± 6.33	80.94 ± 5.66
DGCNN	89.59 ± 5.13	88.93 ± 3.93	88.63 ± 5.13
gcForest	89.03 ± 5.56	90.41 ± 5.33	89.89 ± 6.19
CapsNet	93.94 ± 4.12	94.29 ± 4.39	94.45 ± 4.42
Ours	**94.59** ± **3.77**	**95.26** ± **3.63**	**95.13** ± **3.81**

As shown in Table 5, the baseline removal can significantly improve the recognition accuracy by nearly 34% and 17% on DEAP and DREAMER, respectively, which indicates that this approach is also suited with our model. By the way, baseline removal has a great influence on the recognition accuracy, especially the accuracy of the data-driven methods, which is consistent with the results of [25,44].

4.4. Comparison with the state-of-the-art methods

To further validate the proposed MLF-CapsNet-based method, we compared our method with the start-of-the-art methods on DEAP and DREAMER, respectively, including the decision tree (DT) [25], the multi-layer perceptron (MLP) [25], the support vector machine (SVM) [49], the continuous CNN (Cont-CNN) [25], the CNN-RNN [44], the dynamical graph convolutional neural network (DGCNN) [27], and the multi-grained cascade forest (gcForest) [50]. Yang et al. use DE features extracted from theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz) and gamma (31–50 Hz) frequency bands of preprocessed EEG signals as input to DT, MLP and SVM [25]. The Cont-CNN is a convolutional neural network without pooling operation. It takes the constructed 3D EEG cube as input, where the 3D EEG cube is a 3-dimensional representation that combines DE features with spatial information among electrodes [25]. The CNN-RNN is a hybrid neural network. It extracts spatial and temporal features from constructed 2D EEG frames and 1D EEG sequences, respectively. The DGCNN is proposed by Song et al. that can dynamically learn the internal relationship between different EEG channels represented by an adjacency matrix to classify EEG emotions [27]. The gcForest is a model based on deep forest, which is applied to EEG-based emotion recognition by Cheng et al. in 2020 [50]. The authors adopt it to extract spatial and temporal features from constructed 2D EEG frames like Yang et al. do [44]. Among these seven comparison methods, the CNN-RNN-based method and gcForest-based method both use 2D EEG frames as input, which are both end-to-end frameworks eliminating manual feature extraction. The remaining five methods have to manually extract DE features from four frequency bands. These seven methods use the same preprocessing

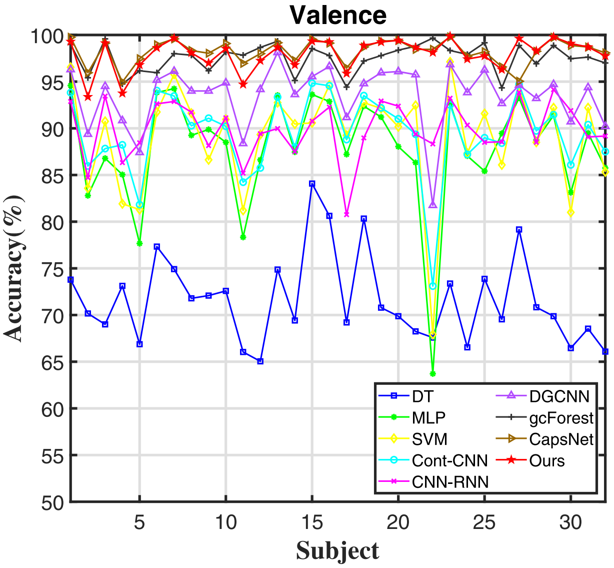

Fig. 5. Average accuracy (%) of each subject in DEAP dataset on valence classification task using different methods.

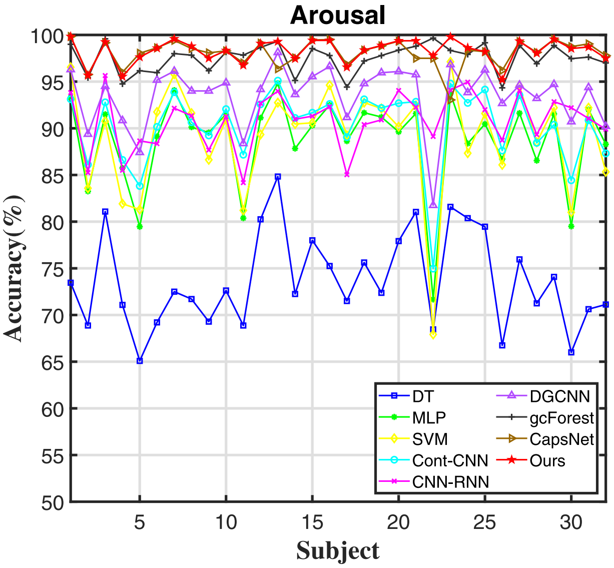

Fig. 6. Average accuracy (%) of each subject in DEAP dataset on arousal classification task using different methods.

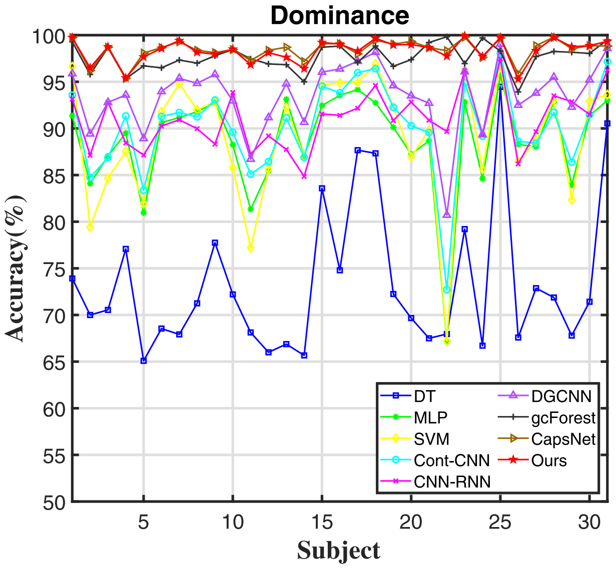

Fig. 7. Average accuracy (%) of each subject in DEAP dataset on dominance classification task using different methods.

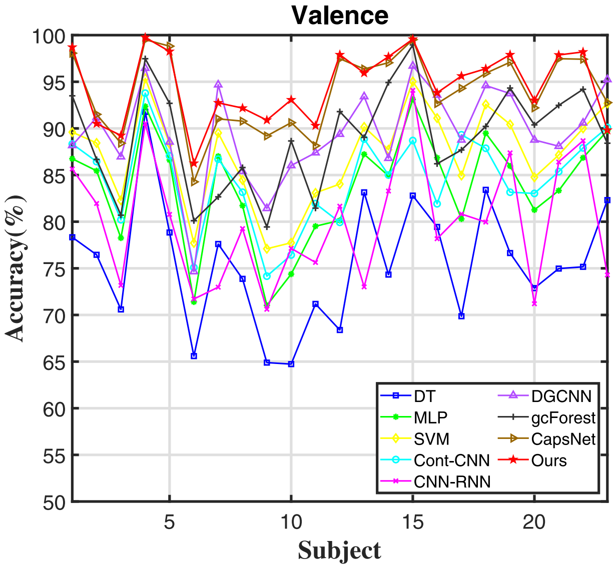

Fig. 8. Average accuracy (%) of each subject in DREAMER dataset on valence classification task using different methods.

method of baseline removal, the same slice length and the same 10-fold cross validation as our method, which ensures the fairness of comparison experiments.

Table 6 reports the average accuracy and standard deviation of 32 subjects on valence, arousal and dominance classification tasks on DEAP. It can be seen that our method significantly outperform the other seven methods except the original CapsNet. First, compared with the two latest CNN-based methods (Cont-CNN and CNN-RNN), our method improves the recognition accuracy by about 8% on three recognition tasks. Second, compared with the two data-driven methods (CNN-RNN and DGCNN), our method also shows advantages. Third, our method has the smallest standard deviation among all the methods, indicating its higher stability when applied to different subjects. Figs. 5–7 provide more detailed and distinct comparison among different methods. It can be seen that the accuracy of our method on valence, arousal and dominance classification tasks is higher than 94%, 95% and 95%, respectively, and our method has clear advantage over other methods with higher accuracy and stability. Moreover, compared to the original CapsNet-based method, the proposed MLF-CapsNet-based method improves accuracy and stability on arousal classification task.

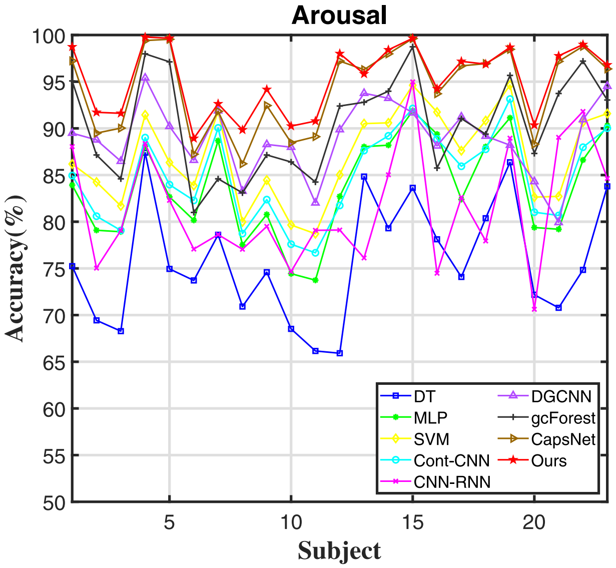

Fig. 9. Average accuracy (%) of each subject in DREAMER dataset on arousal classification task using different methods.

Table 8
Training time and testing time of different methods on DEAP and DREAMER datasets.

	DEAP		DREAMER	
	Training time (s)	Testing time (ms)	Training time (s)	Testing time (ms)
DT	0.2928	0.0015	0.4548	0.0008
MLP	33.7596	0.0063	52.3274	0.0025
SVM	0.7508	0.1846	1.4115	0.0988
Cont-CNN	12.6405	0.1141	20.6040	0.0767
CNN-RNN	656.3955	1.0554	602.0744	0.9210
DGCNN	7.0225	0.3208	10.2529	0.1820
gcForest	693.4861	10.5672	1307.406	7.4973
CapsNet	389.9597	7.6667	244.4818	1.7129
Ours	1338.3379	48.2138	635.9729	14.8009

Table 7 shows the average accuracy and standard deviation of 23 subjects on valence, arousal and dominance classification tasks on DREAMER. Similar to DEAP dataset, our method also significantly outperform other methods in both accuracy and standard deviation. Specifically, our method improves the recognition accuracy by about 10% on three recognition tasks compared with the two latest CNN-based methods (Cont-CNN and CNN-RNN). In particular, when compared with the two data-driven methods, our method has obvious advantages. Figs. 8–10 provide the recognition accuracy of 23 subjects on valence, arousal and dominance recognition tasks, respectively. It can be also seen that our method has clear advantage over other methods with higher accuracy and stability. Specifically, the accuracy of our method on valence and the other two classification tasks is greater than 85% and 86%, respectively. Furthermore, compared to the original CapsNet-based method, our method improves the recognition accuracy by about 1% on three recognition tasks.

4.5. Training and testing time

In our experiments, DT, SVM, and gcForest are trained on a INTEL i7-7800X CPU. MLP, Cont-CNN, CNN-RNN, and DGCNN are trained on a NVIDIA GPU with TensorFlow framework. The proposed method is trained on two NVIDIA GPU with TensorFlow framework. The training time and testing time of each method are shown in Table 8. It can be

Table 9
Details of several reported studies on DEAP dataset.

Studies	Inputs	Models	Evaluation Methods	Accuracy (%)		
				Valence	Arousal	Dominance
Salma et al. [29]	Raw EEG signals	LSTM	4-fold cross validation	85.45	85.65	–
Laura et al. [51]	Statistical features	KNN	Leave-one-out cross validation	89.61	89.84	–
Chao et al. [52]	Statistical, power, and HHS features	Ensemble DBN with Glia Chains	10-fold cross validation	76.83	75.92	–
Kwon et al. [53]	EEG Spectrograms and GSR features	Fusion CNN	Leave-one-out cross validation	80.46	76.56	–
Gao et al. [54]	DE features	Dense CNN	10-fold cross validation	92.24	92.92	–
Sharma et al. [55]	Higher order statistics	Bi-LSTM	10-fold cross validation	84.16	85.21	–
Chen et al. [56]	2D PSD mesh sequence	Cascaded CNN-RNN	10-fold cross validation	93.64	93.26	–
The proposed method	Raw EEG signals	MLF-CapsNet	10-fold cross validation	**97.97**	**98.31**	**98.32**

Table 10
Details of several reported studies on DREAMER dataset.

Studies	Inputs	Models	Evaluation methods	Accuracy (%)		
				Valence	Arousal	Dominance
Katsigiannis et al. [35]	PSD	SVM	10-fold cross validation	62.49	62.17	61.84
Siddharth et al. [57]	EEG-PSD images-based Deep-Learning features	VGG-16 network	Leave-one-out cross validation	78.99	79.23	–
Liu et al. [58]	DE	Deep CCA	Leave-one-out cross validation	90.57	88.99	90.67
The proposed method	Raw EEG signals	MLF-CapsNet	10-fold cross validation	**94.59**	**95.26**	**95.13**

seen that DT, MLP, SVM, Cont-CNN and DGCNN have lower computational cost of training and testing than CNN-RNN, CapsNet and our model. This is because the former are all feature-driven methods, and the latter are all data-driven methods. Compared with the data-driven methods, our model requires longer training time since our multi-level features increase the complexity of the model. This is the inadequacy of our model and needs further improvement in the future.

4.6. Comparison with several existing studies

Finally, we compare the proposed method with several existing studies using the same datasets, *i.e.*, DEAP dataset and DREAMER dataset. Tables 9 and 10 show the details of the existing studies on DEAP and DREAMER, respectively. Items crossed in tables are not indicated in the corresponding references. From the results of EEG emotion recognition summarized in Tables 9 and 10, we can see that our method improves the current state-of-the-art results on both DEAP and DREAMER. Specifically, on DEAP dataset, our method achieves the highest accuracy of 97.97%, 98.31% and 98.32% for valence, arousal and dominance, respectively. The accuracy achieved by our method is 5% higher than the second highest accuracy [56] listed in Table 9. On DREAMER dataset, our method achieves the best performance of 94.59%, 95.26% and 95.13% for valence, arousal and dominance, respectively, which also improves the accuracy by about 5% compared with the second highest accuracy [58] listed in Table 10. Moreover, compared with the methods in references [56] and [58], our method uses raw EEG signals as inputs, which eliminates the complicated process of manually extracting features.

5. Discussion

It can be seen from the experimental results that the proposed method significantly outperforms some state-of-the-art methods on the subject-dependent task, in particular, our method is data-driven, eliminating complex feature engineering. It is necessary to analyze why the proposed method can achieve such a superior performance on multi-channel EEG-based emotion recognition. The superior recognition performance of our method is most likely due to the following major points:

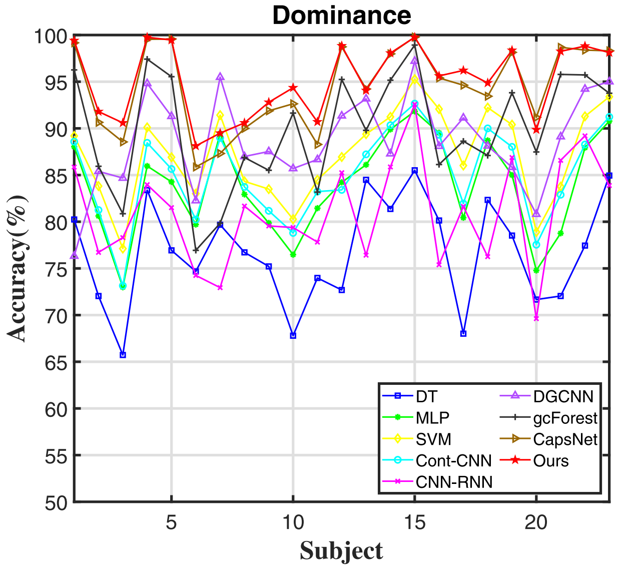

Fig. 10. Average accuracy (%) of each subject in DREAMER dataset on dominance classification task using different methods.

1. There is a strong correlation between emotional states and brain functional connectivity patterns. Specifically, different emotional states cause different activities in various functional areas of the brain, and specific connections between different functional regions [32,33]. Interestingly, the proposed framework adopts capsules to encode entities, and adopts transformation matrices to encode the intrinsic spatial relationship between

a part and a whole of an object, which makes itself can well represent the relationship between the parts of a object. Therefore, our method provides a useful way to characterize the intrinsic relationships among the various EEG channels. Specifically, the primary capsules encode the brain areas, and the transformation matrices encode the connection among different brain areas. These innovative structures are advantageous for extracting the most discriminative features for the emotion recognition task.

2. Compared to handwritten digital images in the MINIST dataset which is used to evaluate the performance of original CapsNet, EEG signals have more complex internal representation related to emotion. The proposed framework incorporates multi-level feature maps learned by different layers in forming the primary capsules so that the capability of feature representation can be enhanced, which makes our method achieve better performance than that of the original CapsNet on multi-channel EEG-based emotion recognition task.
3. The size of multi-channel EEG signal is usually large, which will increase the amount of parameters and computational complexity of network. The proposed framework adopts a bottleneck layer to reduce the number of channel of concatenated feature maps, which makes our method have significantly less number of parameters without sacrificing performance on multi-channel EEG-based emotion recognition task.

6. Conclusion

In this paper, we propose an end-to-end MLF-CapsNet framework for multi-channel EEG emotion recognition. Our proposed framework can identify the intrinsic relationship among various EEG channels well. We combine multi-level features extracted from different convolution layers to form primary capsules. Besides, we add bottleneck layer to reduce the amount of parameters and accelerate the speed of calculation. Finally, experiments on DEAP dataset and DREAMER dataset are conducted. Our method achieves average accuracy of 97.97%, 98.31% and 98.32% for valence, arousal and dominance on the DEAP dataset, and achieves average accuracy of 94.59%, 95.26% and 95.13% for valence, arousal and dominance on the DREAMER dataset, respectively. Experimental results demonstrate that the proposed MLF-CapsNet-based method achieves higher accuracy than some state-of-the-art DL-based methods on the subject-dependent task, such as the Cont-CNN, CNN-RNN, DGCNN and gcForest methods. Moreover, compared with original CapsNet, our method improves the accuracy by 0.3% on arousal classification task of DEAP dataset, and improves the accuracy by about 1% on three classification tasks of DREAMER dataset, which validates the efficiency of our method. In the future, we will study the effectiveness of the proposed MLF-CapsNet-based framework in subject-independent EEG-based emotion recognition through domain adaptation and domain generalization, and reduce the complexity of the network through sharing parameters between capsule layers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (Grant 2019YFA0706203), the National Natural Science Foundation of China (Grants 61922075, 41901350 and 61701160), the Provincial Natural Science Foundation of Anhui (Grants 2008085QF285 and 1808085QF186) and the Fundamental Research Funds for the Central Universities (Grants JZ2020HGPA0111 and JZ2019HGBZ0151).

References

[1] R.J. Dolan, Emotion, cognition, and behavior, Science 298 (5596) (2002) 1191–1194.

[2] G. Recio, A. Schacht, W. Sommer, Recognizing dynamic facial expressions of emotion: Specificity and intensity effects in event-related brain potentials, Biol. Psychol. 96 (2014) 111–125.

[3] H. Gunes, M. Piccardi, Bi-modal emotion recognition from expressive face and body gestures, J. Netw. Comput. Appl. 30 (4) (2007) 1334–1345.

[4] K. Han, D. Yu, I. Tashev, Speech emotion recognition using deep neural network and extreme learning machine, in: Fifteenth Annual Conference of the International Speech Communication Association, 2014.

[5] F. Agrafioti, D. Hatzinakos, A.K. Anderson, Ecg pattern analysis for emotion detection, IEEE Trans. Affect. Comput. 3 (1) (2011) 102–115.

[6] C. Li, W. Tao, J. Cheng, Y. Liu, X. Chen, Robust multichannel eeg compressed sensing in the presence of mixed noise, IEEE Sens. J. 19 (22) (2019) 10574–10583.

[7] B. Cheng, G. Liu, Emotion recognition from surface emg signal using wavelet transform and neural network, in: Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE), 2008, pp. 1363–1366.

[8] E. Lotfi, M.-R. Akbarzadeh-T., Practical emotional neural networks, Neural Netw. 59 (2014) 61–72.

[9] K.A. Lindquist, L.F. Barrett, A functional architecture of the human brain: emerging insights from the science of emotion, Trends Cogn. Sci. 16 (11) (2012) 533–540.

[10] G.K. Verma, U.S. Tiwary, Affect representation and recognition in 3d continuous valence–arousal–dominance space, Multimedia Tools Appl. 76 (2) (2017) 2159–2183.

[11] B. Hjorth, Eeg analysis based on time domain properties, Electroencephalogr. Clin. Neurophysiol. 29 (3) (1970) 306–310.

[12] Y. Liu, O. Sourina, Real-time fractal-based valence level recognition from eeg, in: Transactions on Computational Science XVIII, Springer, 2013, pp. 101–120.

[13] P.C. Petrantonakis, L.J. Hadjileontiadis, Emotion recognition from eeg using higher order crossings, IEEE Trans. Inf. Technol. Biomed. 14 (2) (2009) 186–197.

[14] L.-C. Shi, Y.-Y. Jiao, B.-L. Lu, Differential entropy feature for eeg-based vigilance estimation, in: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2013, pp. 6627–6630.

[15] Y. Lin, C. Wang, T. Jung, T. Wu, S. Jeng, J. Duann, J. Chen, Eeg-based emotion recognition in music listening, IEEE Trans. Biomed. Eng. 57 (7) (2010) 1798–1806.

[16] W.-L. Zheng, B.-L. Lu, Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks, IEEE Trans. Auton. Ment. Dev. 7 (3) (2015) 162–175.

[17] M. Naji, M. Firoozabadi, P. Azadfallah, Emotion classification during music listening from forehead biosignals, Signal Image Video Process. 9 (6) (2015) 1365–1375.

[18] J.R. Estepp, J.C. Christensen, Electrode replacement does not affect classification accuracy in dual-session use of a passive brain-computer interface for assessing cognitive workload, Front. Neurosci. 9 (2015) 54.

[19] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[20] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[21] X. Liu, P. He, W. Chen, J. Gao, Multi-task deep neural networks for natural language understanding, 2019, arXiv:1901.11504.

[22] R. Collobert, J. Weston, A unified architecture for natural language processing: Deep neural networks with multitask learning, in: Proceedings of the 25th International Conference on Machine Learning, Association for Computing Machinery, 2008, pp. 160–167.

[23] Z. Yao, Z. Wang, W. Liu, Y. Liu, J. Pan, Speech emotion recognition using fusion of three multi-task learning-based classifiers: Hsf-dnn, ms-cnn and lld-rnn, Speech Commun. 120 (2020) 11–19.

[24] W.-L. Zheng, J.-Y. Zhu, Y. Peng, B.-L. Lu, Eeg-based emotion classification using deep belief networks, in: 2014 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2014, pp. 1–6.

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

[26] S. Tripathi, S. Acharya, R.D. Sharma, S. Mittal, S. Bhattacharya, Using deep and convolutional neural networks for accurate emotion classification on deap dataset, in: Twenty-Ninth AAAI Conference, 2017.

[27] T. Song, W. Zheng, S. Peng, C. Zhen, Eeg emotion recognition using dynamical graph convolutional neural networks, IEEE Trans. Affect. Comput. PP (99) (2018) 1.

[28] H. Chao, L. Dong, Y. Liu, B. Lu, Emotion recognition from multiband eeg signals using capsnet, Sensors 19 (9) (2019) 2212.

[29] S. Alhagry, A.A. Fahmy, R.A. El-Khoribi, Emotion recognition based on eeg using lstm recurrent neural network, Emotion 8 (10) (2017) 355–358.

[30] S. Sabour, N. Frosst, G.E. Hinton, Dynamic routing between capsules, in: Advances in Neural Information Processing Systems, 2017, pp. 3856–3866.

[31] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, R. Boots, Eeg-based intention recognition from spatio-temporal representations via cascade and parallel convolutional recurrent neural networks, 2017, arXiv preprint arXiv:1708.06578.

[32] S.-E. Moon, S. Jang, J.-S. Lee, Convolutional neural network approach for eeg-based emotion recognition using brain connectivity and its spatial information, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 2556–2560.

[33] Y.-Y. Lee, S. Hsieh, Classifying different emotional states by means of eeg-based functional connectivity patterns, PLoS One 9 (4) (2014) e95415.

[34] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, Deap: A database for emotion analysis using physiological signals, IEEE Trans. Affect. Comput. 3 (1) (2011) 18–31.

[35] S. Katsigiannis, 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 Inf. 22 (1) (2017) 98–107.

[36] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, A. Al-Wabil, Review and classification of emotion recognition based on eeg brain-computer interface system research: a systematic review, Appl. Sci. 7 (12) (2017) 1239.

[37] E.L. Broek, Ubiquitous emotion-aware computing, Pers. Ubiquitous Comput. 17 (1) (2013) 53–67.

[38] P. Ekman, An argument for basic emotions, Cogn. Emot. 6 (3–4) (1992) 169–200.

[39] G.E. Hinton, A. Krizhevsky, S.D. Wang, Transforming auto-encoders, in: International Conference on Artificial Neural Networks, Springer, 2011, pp. 44–51.

[40] Y. Wang, A. Sun, M. Huang, X. Zhu, Aspect-level sentiment analysis using as-capsules, in: The World Wide Web Conference, ACM, 2019, pp. 2033–2044.

[41] P. Afshar, A. Mohammadi, K.N. Plataniotis, Brain tumor type classification via capsule networks, in: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, 2018, pp. 3129–3133.

[42] J. Yin, S. Li, H. Zhu, X. Luo, Hyperspectral image classification using capsnet with well-initialized shallow layers, IEEE Geosci. Remote Sens. Lett. (2019).

[43] M.A.T. Turan, E. Erzin, Monitoring infant's emotional cry in domestic environments using the capsule network architecture, in: Interspeech, 2018, pp. 132–136.

[44] Y. Yang, Q. Wu, M. Qiu, Y. Wang, X. Chen, Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network, in: 2018 International Joint Conference on Neural Networks (IJCNN), IEEE, 2018, pp. 1–7.

[45] X.-W. Wang, D. Nie, B.-L. Lu, Emotional state classification from eeg data using machine learning approach, Neurocomputing 129 (2014) 94–106.

[46] G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (2) (1979) 215–223.

[47] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[48] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems, 2016, arXiv preprint arXiv:1603.04467.

[49] J.A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300.

[50] J. Cheng, M. Chen, C. Li, Y. Liu, R. Song, A. Liu, X. Chen, Emotion recognition from multi-channel eeg via deep forest, IEEE J. Biomed. Health Inf. (2020) 1.

[51] L. Piho, T. Tjahjadi, A mutual information based adaptive windowing of informative eeg for emotion recognition, IEEE Trans. Affect. Comput. (2018) 1.

[52] H. Chao, H. Zhi, L. Dong, Y. Liu, J. Dauwels, Recognition of emotions using multichannel eeg data and dbn-gc-based ensemble deep learning framework, Intell. Neurosci. 2018 (2018).

[53] Y.-H. Kwon, S.-B. Shin, S.-D. Kim, Electroencephalography based fusion two-dimensional (2d)-convolution neural networks (cnn) model for emotion recognition system, Sensors 18 (5) (2018).

[54] Z. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, G. Chen, A channel-fused dense convolutional network for eeg-based emotion recognition, IEEE Trans. Cogn. Dev. Syst. (2020) 1.

[55] R. Sharma, R.B. Pachori, P. Sircar, Automated emotion recognition based on higher order statistics and deep learning algorithm, Biomed. Signal Process. Control 58 (2020) 101867.

[56] J. Chen, D. Jiang, Y. Zhang, P. Zhang, Emotion recognition from spatiotemporal eeg representations with hybrid convolutional recurrent neural networks via wearable multi-channel headset, Comput. Commun. 154 (2020) 58–65.

[57] S. Siddharth, T. Jung, T.J. Sejnowski, Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing, IEEE Trans. Affect. Comput. (2019) 1.

[58] W. Liu, J.-L. Qiu, W.-L. Zheng, B.-L. Lu, Multimodal emotion recognition using deep canonical correlation analysis, 2019, arXiv:1908.05349.