SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •
Special Focus on Brain Machine Interfaces and Applications

April 2022, Vol. 65 140405:1–140405:16
https://doi.org/10.1007/s11432-021-3439-2

Multi-channel EEG-based emotion recognition in the presence of noisy labels

Chang LI[1,2], Yimeng HOU[1], Rencheng SONG[1], Juan CHENG[1], Yu LIU[1] & Xun CHEN[3,4,5]*

[1]*Department of Biomedical Engineering, Hefei University of Technology, Hefei* 230009*, China;*
[2]*Anhui Province Key Laboratory of Measuring Theory and Precision Instrument, Hefei University of Technology, Hefei* 230009*, China;*
[3]*Epilepsy Center, Department of Neurosurgery, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei* 230001*, China;*
[4]*School of Information Science and Technology, University of Science and Technology of China (USTC), Hefei* 230027*, China;*
[5]*USTC IAT-Huami Joint Laboratory for Brain-Machine Intelligence, Institute of Advanced Technology, University of Science and Technology of China, Hefei* 230088*, China*

Received 12 August 2021/Revised 16 February 2022/Accepted 23 February 2022/Published online 11 March 2022

Abstract A large number of deep learning classification methods for emotion recognition tasks based on electroencephalogram (EEG) have achieved excellent performance, and it is implicitly assumed that all labels are correct. However, humans have natural bias, subjectiveness, and inconsistencies in their judgment, which would lead to noisy labels for the EEG emotion state. To this end, we propose a framework for multi-channel EEG-based emotion recognition in the presence of noisy labels. The proposed noisy labels classification method is based on the capsule network using a joint optimization strategy (JO-CapsNet) until convergence. Specifically, the network parameters are updated based on the loss function of the capsule network, and the pseudo label is updated by predicting the existence possibility of the class label based on the output of the capsule network. In this way, the alternate updating strategy can promote each other to correct the noisy labels. Experimental results demonstrate the advantage of our method.

Keywords electroencephalogram (EEG), emotion recognition, noisy labels, capsule network, joint optimization

Citation Li C, Hou Y M, Song R C, et al. Multi-channel EEG-based emotion recognition in the presence of noisy labels. Sci China Inf Sci, 2022, 65(4): 140405, https://doi.org/10.1007/s11432-021-3439-2

1 Introduction

Emotion is a general term for complex subjective cognitive experience, which includes human's psychological response and physiological response to external stimuli or self-stimulation. Emotion recognition occupies an important position in the field of artificial intelligence [1]. Correct emotion representation is a key step in emotion recognition research, which can be divided into non-physiological signals and physiological signals. Non-physiological signals include facial expressions [2], body movements [3], gestures [4], speech [5]. Physiological signals include heart rate (HR) [6], respiration rate (RR) [7], galvanic skin response (GSR) [8], electromyography (EMG) [9], and electroencephalography (EEG) [10–12]. EEG signals have the strong correlation with people's cognitive behavior and mental activities, and generally reflect emotions more directly compared with other non-physiological signals [13]. Besides, the multi-channel EEG signal has good time resolution [12], which can be used for emotion recognition more efficiently and accurately [14, 15].

Many researchers divide the emotion model into the discrete model and continuous model [16]. The discrete model points out that all emotions are composed of the eight basic emotions of anger, fear, expectation, sadness, disgust, surprise, acceptance, and joy. The continuous model is more extensible to

* Corresponding author (email: xunchen@ustc.edu.cn.)

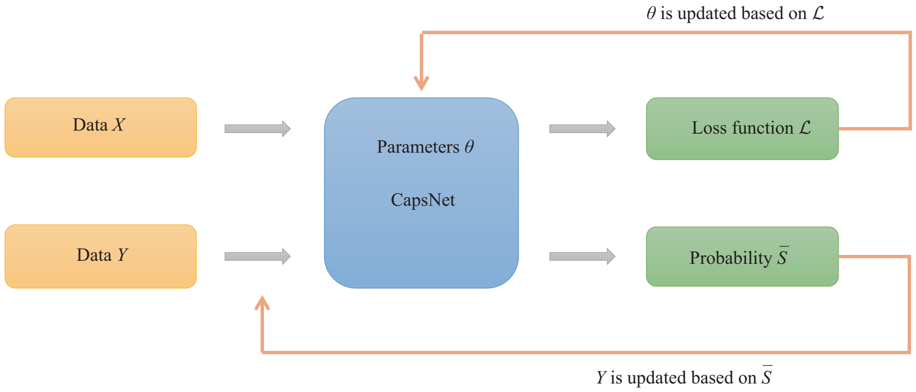
θ is updated based on $\mathcal{L}$
Data X
Data Y
Parameters θ
CapsNet
Loss function $\mathcal{L}$
Probability $\bar{S}$
Y is updated based on $\bar{S}$

Figure 1 (Color online) Proposed joint optimization framework in the presence of noisy labels.

vectorize emotion, and the research using this theory is more and more dominant in this field. Generally speaking, the steps of traditional EEG emotion recognition are to extract features from EEG signals and then use a classifier to classify the emotion features extracted from EEG signals [17]. For example, Li et al. [18] used the support vector machine (SVM) to classify features extracted from the gamma frequency band. Patil et al. [19] extracted higher-order crossing features from the EEG signal for classification, Shi et al. [20] used differential entropy (DE) features on different frequency bands, and combined K-nearest neighbor (KNN) and SVM for classification [21]. Recently, many researchers have confirmed that deep learning is superior to traditional methods in many fields, and many EEG emotion recognition methods combined with deep learning have demonstrated excellent results. For instance, Yang et al. [22] grouped DE features of multiple frequency bands, and used a continuous convolutional neural network (CNN) for classification. Song et al. [23] considered the electrode position relationship to design DE features, and used the graph convolutional neural network to recognize emotion. Besides, there are some data-driven and end-to-end EEG emotion recognition methods, which do not need to extract hand-crafted features. For instance, Alhagry et al. [24] presented LSTM-RNN to learn features from the original signal to identify emotions and then classify them in the dense layer. Dose et al. [25] put forward CNN to learn generalized features and adopted conventional fully connected layers to classify EEG signals. Fahimi et al. [26] developed an end-to-end deep CNN to decode attention information from EEG time series by feeding three different EEG representations into the network.

Owing to large-scale, well-labeled training data, the deep learning methods have achieved excellent performance in computer vision, pattern recognition, medical image analysis as well as EEG emotion recognition. However, humans have a natural bias, subjectiveness, and inconsistencies in their judgment, which would lead to noisy labels in the ratings of EEG emotion state [27, 28].

To solve the problem of noisy labels in EEG emotion recognition, we raise a method based on the capsule network using a joint optimization strategy (JO-CapsNet), which alternately updates network parameters and pseudo labels until convergence. Figure 1 shows the conceptual diagram of JO-CapsNet. Our goal is to automatically rectify noisy labels during iteration, and hence the performance of the classification can be increased in the presence of noisy labels.

The main contributions of this paper are as follows.

(1) In this paper, we come up with a JO-CapsNet framework for multi-channel EEG emotion recognition in the presence of noisy labels, which can avoid the accumulation of errors during the training process, so that the classification accuracy can be improved. The JO-CapsNet is an end-to-end framework, which can solve the task of extracting features from the original EEG signals and classifying the emotions simultaneously.

(2) We experiment on the DEAP dataset and DREAMER dataset, and the results demonstrate the effectiveness of the proposed method, which can improve the classification accuracy. For the arousal/valence dimension of DEAP, when the noise ratio is 10%, 20%, and 30%, classification accuracies of the JO-CapsNet method are increased by 0.4%/1.03%, 2.33%/3.29%, and 4.64%/4.56% compared with these of the original capsule network. For the arousal/valence/dominance dimension of DREAMER, when the noise ratio is 10%, 20%, and 30%, classification accuracies of the JO-CapsNet method increased by 0.88%/0.73%/0.6%, 1.72%/3.55%/3.34%, and 6.95%/7.72%/8.32% compared with these of the original capsule network.

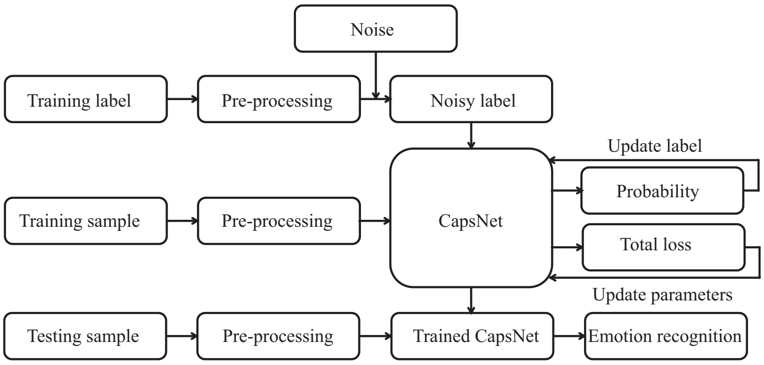

Figure 2 Flowchart of the JO-CapsNet.

The layout for the rest of this paper is as follows. Section 2 is the elaboration of the proposed method. Section 3 is the specific experimental process, results and analysis. Section 4 is the serious discussions, and Section 5 is the conclusion of our study.

2 Methods

In this section, we will recommend the steps of EEG emotion recognition based on JO-CapsNet, dataset and preprocessing, the existence of noisy labels, emotion recognition via capsule network, joint optimization strategy, respectively.

2.1 Steps of EEG emotion recognition based on JO-CapsNet

As shown in Figure 2, the steps of the proposed EEG emotion classification are mainly divided into three parts: the generation of training and test samples, the training process, and the testing part.

(i) An emotion is stimulated by music or movies [29], and electrodes are used to collect the brain electrical signals generated by the subject. EEG signals and other related information in the process (number of people, gender, number of electrodes) are recorded. Use artifact removal methods (such as independent component analysis and blind source separation) to preprocess the EEG signal [30]. Divide the obtained data into the training dataset and test dataset. In order to show the availability of our method, noise is introduced into the label of the training dataset to form a new training dataset with noisy labels, and the test dataset remains unchanged.

(ii) Input the obtained training data and noisy labels into the JO-CapsNet framework, and we can get the optimal solution via the joint optimization strategy. The network parameters are updated based on the loss function in the direction of gradient descent, and the pseudo label is updated based on the output of the capsule network.

(iii) Input the test data and true label into the trained model to obtain classification accuracy, and compare the JO-CapsNet with some traditional methods and deep learning methods to display the effectiveness.

2.2 Dataset and preprocessing

We show the advantage of the proposed method on the public datasets DEAP and DREAMER.

DEAP includes EEG and peripheral physiological signals recorded by 32 participants (16 males and 16 females) when watching 40 music videos [22, 24, 31]. The EEG signal containing 32 channels is used for emotion recognition, while the peripheral physiological signals are canceled. In addition, the DEAP dataset provides a preprocessed version that is used in this article. In preprocessing, the EEG signal is down-sampled to 128 Hz, and the band-pass frequency filter of 4.0–45.0 Hz is applied. In our experiments, the data of each subject contains 60 s test data and 1 s baseline data. EEG signals are easily affected by many factors, there is instability and sensitivity. In order to fairly compare with some other compared methods, we adopt the method to preprocess the dataset as same as [22, 32]. For the EEG emotion recognition task, removing the baseline signal in the emotionally relaxed state (no stimulus) is often used as a preprocessing step. The emotional music video consisted of 40 one-minute clips, and then the participants were asked to record their emotional levels. Score emotional level from 1 to 9 on the arousal and valence dimensions. In our experiment, we divided every dimension of data into two classes with a threshold of 5.

DREAMER includes EEG and peripheral physiological signals recorded by 23 participants (14 males and 9 females) when watching 18 film clips. The EEG signal containing 14 channels is used for emotion recognition and in preprocessing, the EEG signal is down-sampled to 128 Hz. The signal collected by each subject lasted from 65 to 393 s. The artifact subspace reconstruction (ASR) method is used to remove artifacts. The subjects rated their arousal, valence, and dominance levels from 1 to 5. Finally, there are experimental signals, baseline signals, and labels in the DREAMER dataset. In our experiment, we divided every dimension data into two classes with a threshold of 3.

We use the two datasets and 10-fold cross-validation to show the advantage of the proposed method as same as the comparison methods. The final experimental result is the average performance of the 10-fold cross-validation. That is, setting a threshold to divide each emotional dimension into two classes, such as low/high dominance, low/high valence, and low/high arousal. In DEAP, when the score is less than 5, the label is low, and when the score is greater than or equal to 5, the label is high. In DREAMER, when the score is less than 3, the label is low, and when the score is greater than or equal to 3, the label is high. In this way, the problem of EEG emotion recognition is actually a binary classification task. The experimental signal preprocessed by 1s segmentation contains a slicing window of 128 sampling points. For DEAP, one-minute clips can be divided into 60 1s segmented experimental data. Since each subject in DEAP has 40 experimental signals, we obtained 2400 $[32 \times 128]$ EEG experimental samples for each dimension about every subject. For DREAMER, one clip lasted from 65 to 393 s. We obtained 3728 $[14 \times 128]$ EEG experimental samples for each dimension about every subject. In order to simulate the situation of noisy labels, we inject symmetric noisy labels into the dataset, that is, the label is changed in a certain proportion in each dimension (arousal/valence), such as high → low, low → high. Ensure that the ratio of noisy labels is r, and the ratio of the true and pure label is $1 - r$.

2.3 The existence of noisy labels

Noisy labels widely exist in many different fields. For instance, in the field of computer vision, large-scale datasets used for face recognition usually contain noisy labels, especially when they are automatically collected through image search engines or movies [33]. In the field of hyperspectral image classification, noisy labels are caused by limited light in various frequency bands, atmospheric, and instrumental factors [34]. In the field of medical image analysis, datasets are usually small, labels require domain expertise, and the variability between and within observers is high, which leads to noisy labels [35]. In the field of medical applications, the description language may be too limited, which reduces the amount of available information. In some cases, the information is also poor or uneven. For example, the patient's recall may be incorrect [36]. Noisy labels may also exist in the field of data coding or communication. For example, in spam filtering, sources of noisy labels include misunderstandings of feedback mechanisms and accidental clicks [37].

In the field of EEG emotion recognition, humans have natural bias and inconsistencies in their judgment, which creates noise in the ratings [38, 39]. Besides, it is generally acknowledged that emotions are subjective, and studies have indicated that humans understand and perceive emotions varyingly [28]. Moreover, it would lead to a significant increase in mislabeled trials when participants become sleepy, bored, or distracted [27]. The goal of the traditional method is to fit the objective function without considering noisy labels. The parameters obtained after optimization deviate from the true optimal value, which leads to a decline in the classification result during testing. Therefore, it is necessary to take the noisy labels problem of EEG emotion recognition into consideration.

2.4 Emotion recognition via capsule network

CNN-based neural networks have achieved excellent performance in many fields, such as computer vision, pattern recognition, and medical image analysis. Nevertheless, in view of the pooling operation, CNN yet has shortcomings. Pooling improves the important feature information, compresses the feature, reduces the amount of calculation, and alleviates the overfitting. Unfortunately, it is likely to drop precise spatial relationships between some high-level features. For some unique classification problems, good results cannot be achieved. For example, in face recognition problems, the positional relationship of five facial organs, such an advanced spatial relationship, is very important [40]. In order to overcome the defect of the aforementioned CNN, a capsule network is proposed [41]. The capsule network uses encoding entity features and dynamic routing technology, and the core unit composed of it is called a capsule. A capsule is a group of neurons that learn to recognize visual entities and encode their attributes as vectors. The probability of entity existence is expressed by the length of the parameter vector of the instance, and the attribute of the entity is expressed by the direction of the vector. Innovatively, capsules of different layers are connected through the iterative dynamic routing-by-agreement mechanism, which is expressed as lower-level capsules tend to send their output to more correlated higher-level capsules, and the connection between the lower-level and higher-level is quantified by coupling parameters. In addition, the transformation matrix is used to express the local and global position relationship of the object in CapsNet. Based on these innovations, the shortcomings of CNN can be overcome [42]. Therefore, CapsNet has achieved good results in many fields: hyperspectral image classification [43], speech recognition [44], natural language processing [45], and medical image classification [46].

Besides, compared with CNN, the capsule network has a smaller training data scale and is equipped to recognize the spatial relationship between local and global features in the spatial domain [42], which is conductive to improve the performance of emotion recognition. The capsule network demonstrated excellent performance in EEG emotion recognition [42], so we choose the capsule network as a classifier. The capsule network in EEG emotion recognition is comprised of three modules, Conv1, PrimaryCapsules, and EmotionCapsules as shown in Figure 3.

The first layer Conv1 is a convolutional layer that has 256 kernels, ReLU activation, and the stride is 1. The size of these convolution kernels depends on the shape of the input. We use 9×9 for DEAP [29]. This layer converts the value of the sampling point into the activity of the local feature detector, which is then used as the input of the PrimaryCapsules layer.

The second layer (PrimaryCapsules) is a convolutional capsule layer. In detail, the task of PrimaryCapsules is to combine the basic features detected by Conv1. Finally, we group 256 feature maps into $32 \times 8D$ feature maps. In other words, we have 32 primary capsules, and 8 groups of 32 9×9 convolution kernels with the stride of 2 which are used to further extract features, and 8 groups of feature maps are obtained. PrimaryCapsules totally have $[56 \times 8 \times 32]$ capsule outputs, and all capsules share their weights.

The last module is EmotionCaps. We propose this framework for binary classification tasks, for example, low/high arousal and low/high valence, EmotionCaps have n $(n = 2)$ 16 D capsules representing two types of emotional states. The length of each capsule vector in the EmotionCaps layer expresses the existence of the emotional states. Thus the loss of emotion classification can be calculated. In order to ensure that the length of the EmotionCaps' output $\boldsymbol{v}_j$ is between 0 and 1, the non-linear function "squashing" is introduced. This step can be expressed as

$$\boldsymbol{v}_j = \frac{\|\boldsymbol{s}_j\|^2}{1 + \|\boldsymbol{s}_j\|^2} \frac{\boldsymbol{s}_j}{\|\boldsymbol{s}_j\|}, \tag{1}$$

where $\boldsymbol{v}_j$ $(j = 1, \ldots, n)$ is the vector output of emotion capsule j $(j = 1, \ldots, n)$, n is the number of emotion capsule as shown in Figure 4.

For all capsules except the first layer of capsules, the total input of capsule $\boldsymbol{s}_j$ is the weighted sum of all prediction vector $\hat{\boldsymbol{u}}_{j|i}$, which is obtained from the second layer of capsules. The weight matrix $\boldsymbol{W}_{ij}[8 \times 16]$ is multiplied by the output $\boldsymbol{u}_i$ $(i = 1, \ldots, k)$ of the lower layer capsule to get the prediction vector $\hat{\boldsymbol{u}}_{j|i}$, where k is the number of primary capsules. Mathematically, $\boldsymbol{s}_j$ and $\hat{\boldsymbol{u}}_{j|i}$ are expressed as

$$\boldsymbol{s}_j = \sum_i c_{ij} \hat{\boldsymbol{u}}_{j|i}, \tag{2}$$

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

Among them, c_{ij} is the coupling coefficient determined in the iterative dynamic routing process as follows:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_n \exp(b_{in})}, \tag{4}$$

where b_{ij} is the log prior probability of capsule i being coupled to emotion capsule j as shown in Figure 4, and the sum of the coupling coefficients between the i-th capsule of EmotionCaps layer and all capsules in the PrimaryCaps layer is 1.

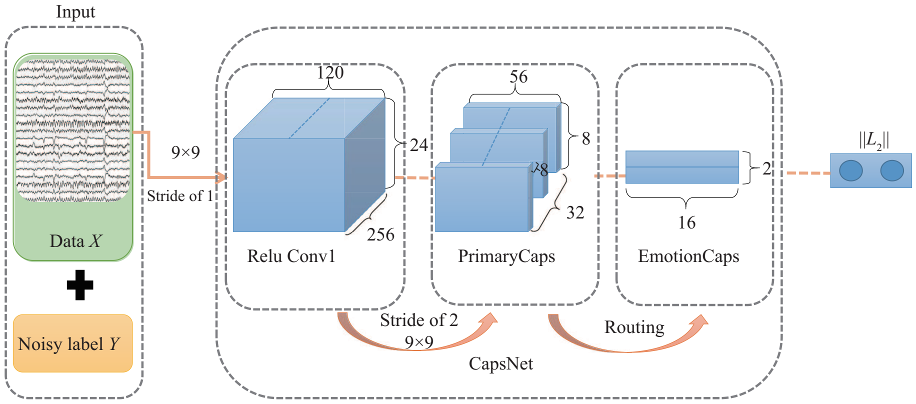

Figure 3 (Color online) The construction of capsule network.

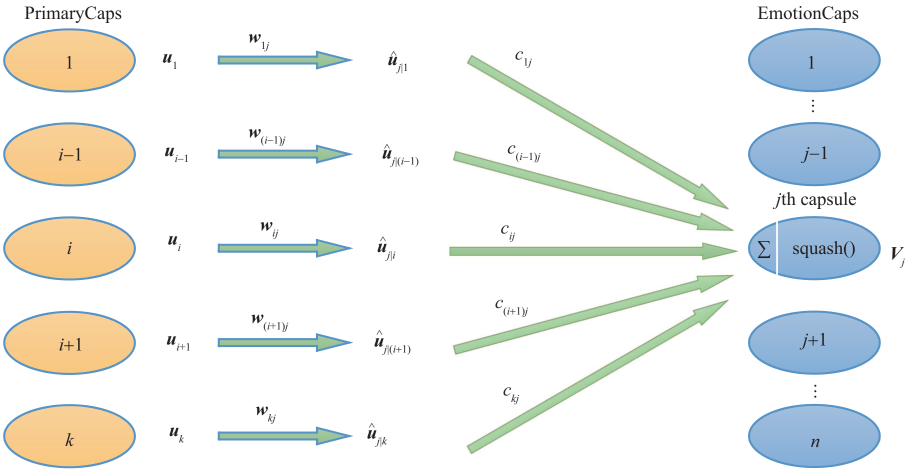

Figure 4 (Color online) Routing by agreement mechanism.

The capsules are connected by dynamic routing as shown in Figure 4. In the routing process, the lower capsule transmits the input vector to the upper capsule. The lower-level capsule calculates the prediction vector by multiplying its output with the weight matrix, which can be routed to the higher-level capsule. If the prediction vector and the output of the upper capsule have a large scalar product, there is top-down feedback, which has the effect of increasing the coupling coefficient of the upper capsule and reducing the coupling coefficient of other capsules. Conv1 output is one-dimensional and has no spatial direction, so no routing is used.

In the general supervised learning of clean data sets, the optimization problem is expressed as

$$\min_{\theta} \mathcal{L}(\theta|X,Y), \tag{5}$$

where $X=[X_1,\ldots,X_N]\in\mathbb{R}^{C\times HN}$ represents the training data matrix, N denotes the number of samples, C is the number of EEG electrode nodes and H is the sampling frequency. $Y=[Y_1,\ldots,Y_N]\in\mathbb{R}^{M\times N}$ represents the label matrix containing noisy labels, Y_i $(i=1,\ldots,N)$ is a one-hot vector representation of the class label and M is the number of classes.

However, when training the network on noisy labels, the result will deviate from the optimal solution after fitting the noisy labels. To this end, we propose JO-CapsNet to solve the problem of EEG emotion recognition using joint optimization strategy in the presence of noisy labels as follows:

$$\min_{\theta,Y} \mathcal{L}(\theta,Y|X). \tag{6}$$

Therefore, we assume that the network under the same learning rate will have difficulty in adapting to noisy labels compared with clean labels. In other words, the loss in (5) is higher for noisy labels and lower for clean labels. Under this assumption, we obtain the clean label by updating the label in the direction of decreasing the loss function expressed as (6), so that it is theoretically feasible to update the label to achieve the purpose of solving the problem of noisy labels. $\mathcal{L}$ is the total loss function as follows:

$$\mathcal{L}(\theta, Y|X) = \mathcal{L}_m(\theta, Y|X) + \alpha\mathcal{L}_p(\theta|X) + \beta\mathcal{L}_e(\theta|X), \tag{7}$$

where $\mathcal{L}_p$ and $\mathcal{L}_e$ represent two regularization losses, α and β denote their corresponding hyperparameters, which will be introduced later, and $\mathcal{L}_m$ is the classification loss using capsule network, which is the sum loss of all emotion capsules as follows:

$$\mathcal{L}_m = \sum_{j=1}^{n} \mathcal{L}_{m_j}, \tag{8}$$

$$\mathcal{L}_{m_j} = T_j \max(0, m^+ - \|\boldsymbol{v}_j\|)^2 + \lambda(1 - T_j)\max(0, \|\boldsymbol{v}_j\| - m^-)^2. \tag{9}$$

We use the length of the instantiation vector to indicate the probability that the capsule entity exists as shown in (9), T_j $(j = 1, \ldots, n)$ means emotion class. If the emotion of class j exists $T_j = 1$, otherwise $T_j = 0$, m^+, and m^- are used to reduce the loss when certain emotional states do not appear, and to prevent the initial learning of all emotion capsules to compress the vector length. In order to allow multiple emotions, we use a separate margin loss $\mathcal{L}_{m_j}$ $(j = 1, \ldots, n)$ for each emotion capsule. We set $m^+ = 0.9$ and $m^- = 0.1$. That is to say, if there is j class, T_j must be no less than 0.9, otherwise T_j will be no greater than 0.1. We use $\lambda = 0.5$ to adjust the loss percentage of the absent emotion class.

In the case of only minimizing (9), results fall into a trivial global optimal solution of the total network. For any training data X_i $(i = 1, \ldots, N)$, the one-hot label Y_i $(i = 1, \ldots, N)$ is always predicted as an instant one-hot label. In order to prevent all labels from being assigned to a class, we will introduce the regularization loss $\mathcal{L}_p$, which is the KL-divergence from $\bar{s}_j(\theta, X)$ $(j = 1, \ldots, M)$ to the prior probability distribution p_j $(j = 1, \ldots, M)$ as follows:

$$\mathcal{L}_p = \sum_{j=1}^{M} p_j \log \frac{p_j}{\bar{s}_j(\theta, X)}, \tag{10}$$

$$\bar{s}_j(\theta, X) = \frac{1}{N}\sum_{i=1}^{N} s_j(\theta, X_i) \approx \frac{1}{|B|}\sum_{X_i \in B} s_j(\theta, X_i), \tag{11}$$

where the p_j is a prior probability distribution, which is the distribution of emotion states among the whole dataset. If the prior distribution of the emotion states is known, the label updated in subsequent iterations should follow the same distribution. Since it is hard to obtain the prior distribution of the class in EEG emotion recognition, the average probability $\bar{s}_j(\theta, X)$ in the dataset is obtained by calculating each small batch B as shown in (11).

When $\alpha = \beta = 0$, the network parameters θ and label Y are both in the local optimum, we recommend an entropy term to group the probability distribution of all pseudo labels to overcome the problem as follows:

$$\mathcal{L}_e = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} s_j(\theta, X_i)\log s_j(\theta, X_i). \tag{12}$$

2.5 Joint optimization strategy

To solve the problem of noisy labels, we propose a framework using a joint optimization strategy, and the network parameters θ and pseudo label Y alternately update to optimize each other. Pseudo-labels are initially used to assign predictions from models trained on the clean dataset to unlabeled data. They are a form of self-training, often used in semi-supervised learning [47–50]. In semi-supervised learning, data is either labeled or unlabeled, then we complete the task of assigning pseudo-labels to unlabeled data, and update the pseudo labels iteratively. Considering that the specific location of the noise is not known in the task of EEG emotion recognition under noisy labels, we use the self-training method [51] to update the pseudo label in the iterative process to achieve the effect of correcting the noisy labels. Below we will elaborate on the update process and rules.

Update network parameters. In this step, we fix the pseudo label Y, optimize according to (7) and update θ through the stochastic gradient descent (SGD) of the loss function in (7).

Table 1 Detailed list of network paramrters

Modules/hyper-parameters	Layers	Parameters/routing	Shape/value
Input	Input	–	DEAP: 32 × 128
			DREAMER: 14 × 128
Conv1	Conv2D	Kernel	DEAP: 9 × 9 × 256
			DREAMER: 6 × 6 × 256
PrimaryCapsules	Conv2D	Kernel	DEAP: 9 × 9 × 256
			DREAMER: 6 × 6 × 256
EmotionCaps	Dynamic routing	$\boldsymbol{W}_{ij}$	8 × 16
		c_{ij}	–
Hyper-parameters	–	α	0.01
		β	0.02

Update label. In this step, we fix network parameter θ to update the pseudo label Y_i $(i = 1, \ldots, N)$. When the probability of the existence of the label is $s(\theta, X_i)$ $(i = 1, \ldots, N)$, the KL-divergence loss from $s(\theta, X_i)$ to Y_i is the smallest, so it is expressed as follows:

$$Y_i = s(\theta, X_i). \tag{13}$$

3 Experiments

In this section, we first introduce the related experimental process and parameter settings at length. Then, we introduce and analyze the experimental results on the DEAP and DREAMER datasets.

3.1 Experimental design

We use 10-fold cross-validation to evaluate our proposed method. The layer Conv1 parameters are set to 9 × 9 convolution kernels, the stride of 1. The layer PrimaryCapsule parameters are set to 9 × 9 convolution kernels, the stride of 2. For our method, we set the learning rate to 10^{-5}, and batch size to 100. The hyperparameters α and β of the two regularization losses in the loss function are set to 0.01 and 0.02, respectively. For DEAP and DREAMER, the total number of epochs to 70 and 36, respectively. In order to simulate different levels of noise, we set the noise ratio r to 0.1, 0.2, 0.3. The following will give a detailed list of network parameters, as shown in Table 1.

Partition of training/test sets. 10-fold cross-validation [52] is adopted in datasets.

3.2 Results of experiments

In order to demonstrate the effectiveness of JO-CapsNet in multi-channel EEG emotion recognition in the presence of noisy labels, we conduct experiments consisting of three parts: (i) illustrate the effect of different ratios of noisy labels on the original capsule network and other six compared methods; (ii) demonstrate the advantage of JO-CapsNet using a joint optimization strategy in comparison with the original capsule network and compared with other seven methods under different noise ratios; (iii) verify the actual correction effect of JO-CapsNet on noisy labels. We conducted extensive experiments on the arousal and valence dimensions of DEAP. In order to simulate the influence of different degrees of noise on the model, we injected different proportions of symmetrical noisy labels, such as high→ low, low→ high as follows:

$$Y = \begin{cases} \text{clean label with probability of } 1 - r; \\ \text{noisy labels with probability of } r. \end{cases} \tag{14}$$

Furthermore, we compared the proposed method with the latest deep learning methods: the CNN-RNN [32], and the multi-grained cascade forest (gcForest) [11]. CNN-RNN is a hybrid neural network, in which CNN extracts spatial features for space expansion and RNN extracts the temporal feature for time extension from EEG signal. Cheng et al. [11] put forward gcForest that the 2D frame sequences are constructed by obtaining the spatial position, and then the 2D frame sequences are input into the constructed classification model. In addition, we also used three traditional classifiers for comparison, including SVM [53], multi-layer perceptron (MLP) [22], and decision tree (DT) [22]. All methods have

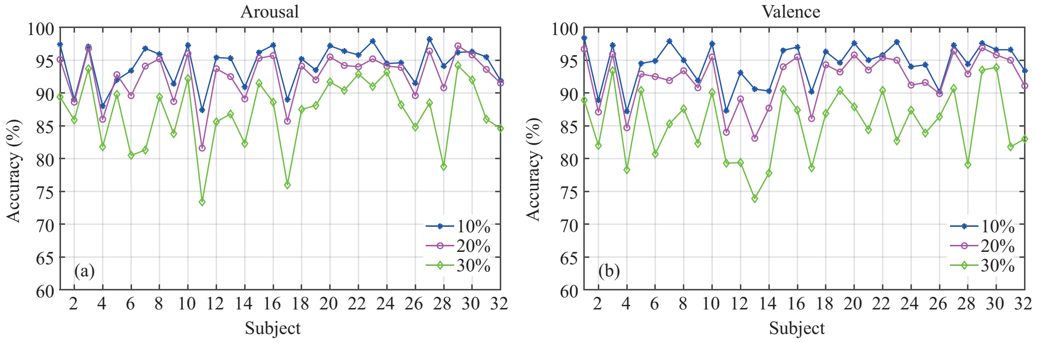

Figure 5 (Color online) The classification accuracies of the original capsule network under different ratios of noisy labels on (a) the arousal dimension and (b) the valence dimension.

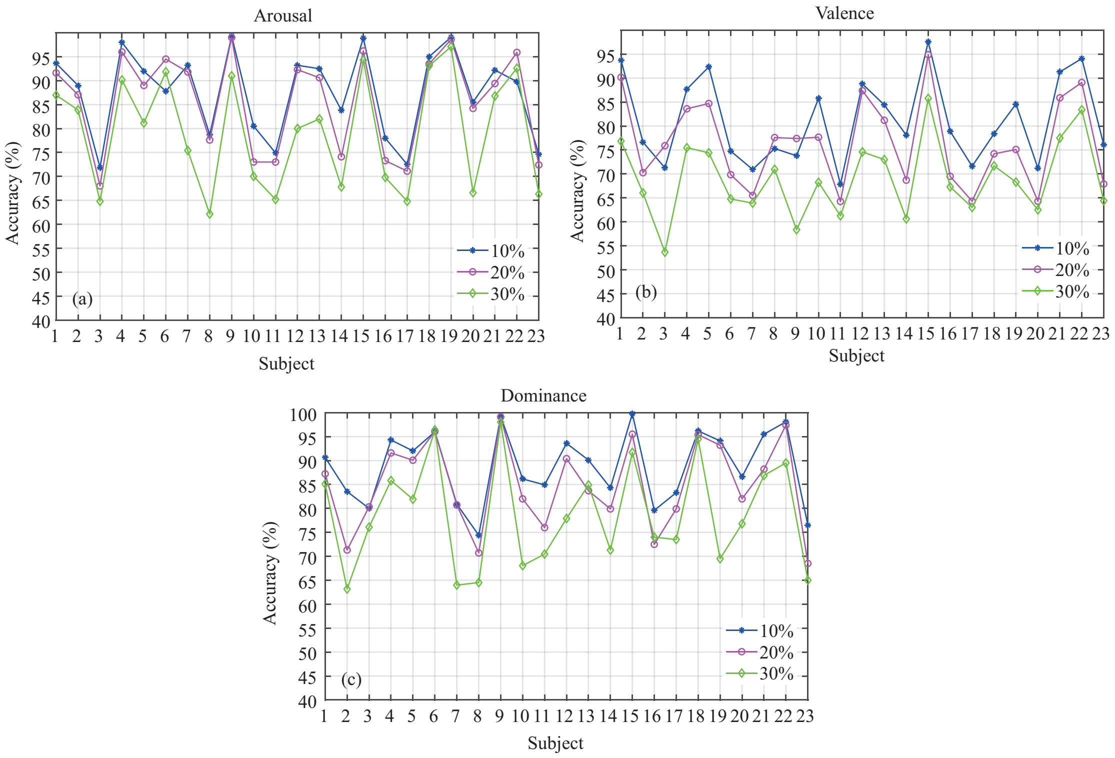

Figure 6 (Color online) The classification accuracies of the original capsule network under different ratios of noisy labels on (a) the arousal dimension, (b) the valence dimension, and (c) the dominance dimension.

been processed in the same preprocessing as our method, namely removing the baseline signal and sliding windows. For traditional classifiers, the input is DE feature, which is extracted from θ (4–7 Hz), α (8–13 Hz), β (14–30 Hz), and γ (31−50 Hz) frequency bands, and is often used in EEG emotion recognition with frequency domain features.

Figure 5 shows the classification accuracies of the original capsule network under different ratios of noisy labels on the arousal and valence dimensions of DEAP, respectively. Figure 6 shows the classification accuracies of the original capsule network under different ratios of noisy labels on the arousal, valence, and dominance dimensions of DREAMER, respectively. It can be seen that when we inject different ratios of symmetrical noisy labels into the training data of the dataset, it will lead the neural network to overfit noisy labels, thereby reducing the classification accuracy. For different subjects, the impact of classification accuracy under the same noise ratio is different, but the overall impact of different noise on different subjects is similar. For different noise ratios, the degree of classification accuracy decline is different, the higher the noise ratio, the lower the classification accuracy. Compared with the case of no noisy labels, the classification accuracy of the capsule network on the arousal dimension of DEAP is reduced by 1.21%, 4.83%, 12.78%, and the classification accuracy of the capsule network on the valence

Table 2 Average accuracies and standard deviations (%) of different methods under different ratios of noise on the arousal of DEAP

Method	10%	20%	30%
DT	68.45±5.15	65.26±3.50	60.94±2.81
SVM	86.60±8.45	85.90±6.21	82.52±6.69
MLP	83.35±7.53	78.84±7.30	73.76±2.60
CNN-RNN	87.27±6.79	84.23±3.04	77.38±2.93
gcForest	93.54±7.10	91.76±7.02	86.97±6.94
Original CapsNet	93.93±3.00	90.31±3.51	82.36±4.38
Ours (JO-CapsNet)	**94.33±3.04**	**92.64±3.65**	**87.00±5.21**

Table 3 Average accuracies and standard deviations (%) of different methods under different ratios of noise on the valence of DEAP

Method	10%	20%	30%
DT	67.66±3.97	61.02±3.34	60.11±2.51
SVM	87.41±6.52	85.24±6.83	81.16±7.50
MLP	83.33±6.36	78.45±7.20	73.34±2.21
CNN-RNN	86.99±2.70	82.40±2.61	76.14±2.55
gcForest	94.07±2.02	92.03±2.69	87.42±4.45
Original CapsNet	93.35±3.11	88.86±4.01	80.70±4.97
Ours (JO-CapsNet)	**94.38±3.22**	**92.15±3.89**	**85.26±5.26**

dimension of DEAP is reduced by 1.51%, 6.01%, 14.17% under the noise ratio of 0.1, 0.2, 0.3, respectively. Also compared with the case of no noisy labels, the classification accuracy of the capsule network on the arousal dimension of DREAMER is reduced by 6.73%, 8.55%, 15%, the classification accuracy of the capsule network on the valence dimension of DREAMER is reduced by 12.85%, 17.45%, 24.99% and the classification accuracy of the capsule network on the dominance dimension of DREAMER is reduced by 5.77%, 9.6%, 15.8% under the noise ratio of 0.1, 0.2, 0.3, respectively. Therefore, the noisy labels in the training data have a negative effect on the classification accuracy via the capsule network, which is due to that the original network easily overfits the noisy labels.

In addition, we demonstrate the performance of different ratios of noisy labels on classification accuracies using six compared methods, i.e., three traditional methods (SVM, MLP, and DT) and three latest deep learning methods (CNN-RNN, gcForest, and original CapsNet). All methods have been processed through the same preprocessing as our method. Tables 2 and 3 show the average accuracies and standard deviations of different methods under different ratios of noise on the arousal and valence dimensions of DEAP, respectively. Tables 4–6 show the average accuracies and standard deviations of different methods under different ratios of noise on the arousal, valence, and dominance dimensions of DREAMER, respectively. We can see that both traditional classification methods and deep learning classification methods are greatly affected by noisy labels. All supervised learning methods are driven by labeled data, and its classification accuracy is strongly dependent on the clean label to a large extent. As the proportion of noise increases, the classification accuracy gradually decreases that showing the huge influence of noisy labels on the different classification models.

We will demonstrate the effectiveness of JO-CapsNet to update pseudo labels using a joint optimization strategy. We input the preprocessed EEG data and noisy labels into the capsule network. After continuous iterative training, the network has learned the clean information in the dataset as well as the noisy information. When updating the noisy labels, it is expressed as a probability s, and we use the average output network probability $\bar{s}$ of the past 10 epochs. The averaging technique has a similar effect to ensemble learning, which can be used to prevent incorrectly updating pseudo labels to a large extent. First, we trained the network for dozens of epochs only to update parameters iteratively by fixing labels of training data. Then, we update the parameters and the pseudo-label alternately by setting the maximum epoch. In our method, we set the learning rate to 10^{-5}, $\alpha = 0.01$, and $\beta = 0.02$.

Figure 7 shows the average accuracies of JO-CapsNet and the original capsule network on each subject on the arousal and valence dimensions of DEAP when the noise is injected with a percentage of 0.3, respectively. We can see that the overall classification accuracy of the original capsule network has decreased to about 82.36% and 80.70%. The accuracies are reduced by about 12.78% and 14.17% compared with a noise-free state. After using a joint optimization strategy, the classification accuracy of JO-CapsNet has

Table 4 Average accuracies and standard deviations (%) of different methods under different ratios of noise on the arousal of DREAMER

Method	10%	20%	30%
DT	79.13±8.73	72.62±7.77	65.73±5.83
SVM	87.29±7.20	86.60±7.26	76.02±7.52
MLP	87.01±8.15	82.57±8.72	75.61±8.71
CNN-RNN	76.81±13.57	75.88±12.30	66.03±11.18
gcForest	86.60±5.37	78.96±5.16	70.07±13.5
Original CapsNet	87.56±8.98	85.74±10.33	79.29±11.74
Ours (JO-CapsNet)	**88.44±9.30**	**87.46±9.64**	**86.24±10.14**

Table 5 Average accuracies and standard deviations (%) of different methods under different ratios of noise on the valence of DREAMER

Method	10%	20%	30%
DT	72.51±6.64	67.55±5.72	62.70±4.59
SVM	80.93±6.62	79.21±6.82	75.68±7.67
MLP	82.68±8.04	78.55±8.54	73.65±8.04
CNN-RNN	74.24±11.67	72.27±12.74	63.62±11.59
gcForest	85.52±5.57	79.06±5.61	68.15±8.95
Original CapsNet	81.09±8.88	76.49±9.24	68.95±7.94
Ours (JO-CapsNet)	**81.82±8.67**	**80.04±8.76**	**76.67±9.51**

Table 6 Average accuracies and standard deviations (%) of different methods under different ratios of noise on the dominance of DREAMER

Method	10%	20%	30%
DT	80.82±7.86	73.63±6.93	66.52±4.98
SVM	88.70±5.90	86.14±5.96	77.79±6.41
MLP	88.51±6.88	83.33±7.80	76.32±7.52
CNN-RNN	78.33±11.34	75.11±12.30	66.27±12.07
gcForest	85.22±5.50	79.26±4.94	70.93±6.19
Original CapsNet	88.68±7.54	84.85±9.26	78.65±10.99
Ours (JO-CapsNet)	**89.28±7.04**	**88.19±7.79**	**86.97±8.31**

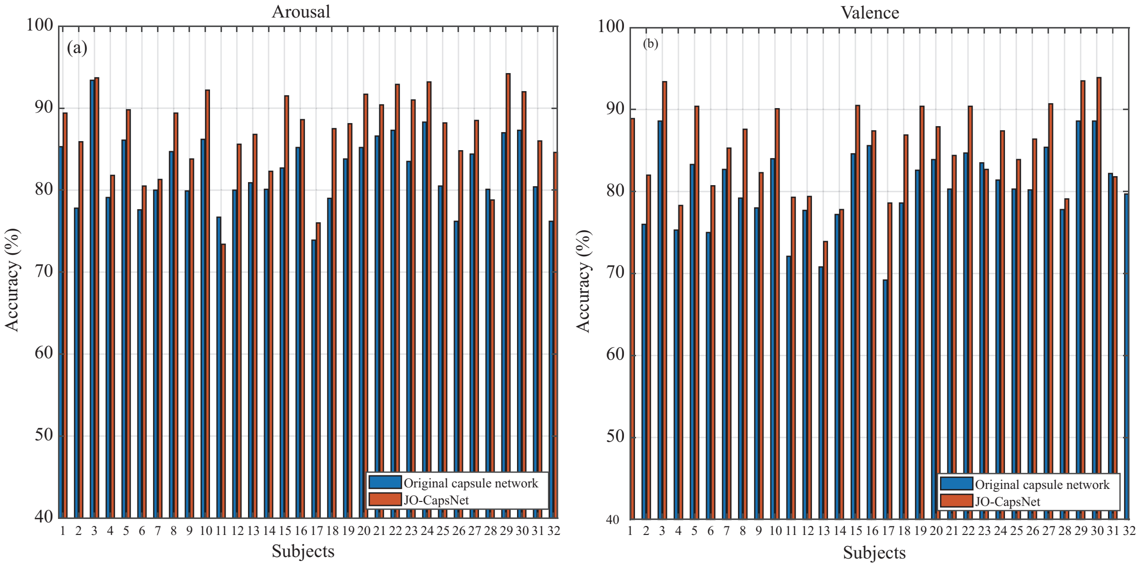

Figure 7 (Color online) Average accuracies (%) on each subject of JO-CapsNet and the original capsule network on (a) the arousal and (b) the valence of DEAP classification tasks when the noise ratio is 0.3.

increased by about 4.64% and 4.56% compared with the original capsule network as shown in Figure 7. Among them, individual subjects, such as s11 and s28, did not perform well in the two dimensions of

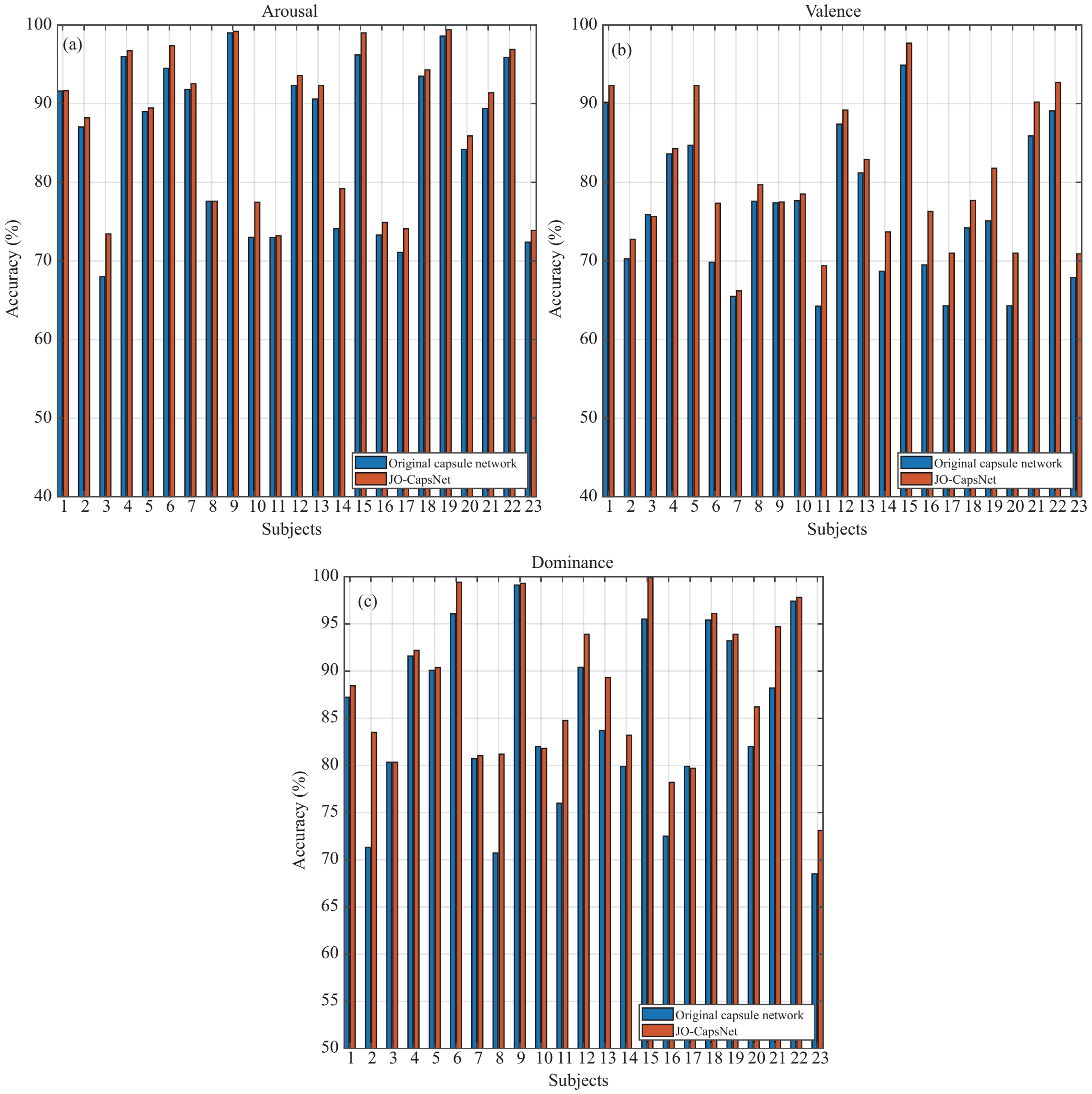

Figure 8 (Color online) Average accuracies (%) on each subject of JO-CapsNet and the original capsule network on (a) the arousal, (b) the valence, and (c) the dominance of DREAMER classification tasks when the noise ratio is 0.2.

arousal/valence compared with other subjects. Compared with Figure 5, it can be seen that the average classification accuracies of these subjects (s11, s28, etc.) are lower than other subjects in the noise-free state via the original capsule network. Since EEG signals vary from person to person, the personal factors of different subjects are quite different, and the EEG signals produced by them are also very different. Therefore, the accuracy reduction is different after the noise is introduced. However, on the whole, the classification accuracy of our method has greatly improved in the presence of noisy labels, with an average improvement of about 4.64% and 4.56% compared with the original capsule network, which verifies the effectiveness of JO-CapsNet using joint optimization strategy.

Figure 8 shows the average accuracies of JO-CapsNet and the original capsule network on each subject on the arousal, valence, and dominance dimensions of DREAMER when the noise is injected with a percentage of 0.2, respectively. The accuracies are reduced by about 8.55%, 17.45%, and 9.6% compared with a noise-free state. After using a joint optimization strategy, the classification accuracy of JO-CapsNet has increased by about 1.72%, 3.55%, and 3.34% compared with the original capsule network as shown in Figure 8. Similarly, the classification performance of different subjects is different, and the influence of noise on the valence dimension is greater than that on the arousal and dominance dimensions. However, the classification accuracy of each subject in the three dimensions is significantly improved after using JO-CapsNet, which proves the effectiveness of the proposed method.

In order to evaluate the performance of the JO-CapsNet under noisy labels with different degrees of noise, we trained the network using a joint optimization strategy on two datasets under different ratios of noisy labels (i.e., r = 10%, 20%, and 30%). Tables 2–6 present the average classification accuracies and standard deviations of JO-CapsNet using joint optimization strategy under different ratios of noisy labels compared with the original capsule network on the arousal/valence dimensions of DEAP and arousal/valence/dominance dimensions of DREAMER. When r is 30%, the increases of accuracies are about 4.64%/4.56% and 6.95%/7.72%/8.32%; when r is 20%, the increases of accuracies are about 2.33%/3.29% and 1.72%/3.55%/3.34%; when r is 10% and 5%, the classification accuracies of JO-CapsNet are close to that of the capsule network without noise, which means that the noisy labels can be roughly rectified. Besides, Tables 2–6 also illustrate the average accuracies and standard deviations of different ratios of noisy labels via JO-CapsNet and other six compared methods on the arousal/valence dimensions of DEAP and arousal/valence/dominance dimensions of DREAMER, which fully demonstrates the effectiveness of our method under various noise ratios, and the greater the noise in a certain range, the more obvious the improvement of classification accuracy.

Besides, the proposed joint optimization strategy can be used for any differentiable classification model, including SVM, MLP, CNN-RNN, and gcForest, but DT is not a differentiable classification model and can not be used.

4 Discussions

In the multi-channel EEG emotion recognition field, many researchers have raised valid classification models to achieve great results. There are still some shortcomings. For example, most of the existing methods belong to supervised learning, and the performance of supervised learning relies on clean labels to a large extent. However, labels tend to contain inaccurate labels that are termed as noisy labels under the realistic circumstance. Therefore, how to solve the classification of multi-channel EEG emotion recognition in the presence of noisy labels is particularly important. Our goal is to find a simple and effective method to realize emotion recognition of multi-channel EEG under noisy labels. In this article, we propose the JO-CapsNet method using a joint optimization strategy, which is customized for multi-channel EEG emotion recognition. We compared the proposed method with the other seven methods. Among them, DT, SVM, and MLP are traditional machine learning algorithms, and CNN-RNN, gcForest, and original capsule network are deep learning models to demonstrate the effectiveness of JO-CapsNet.

The production of emotion is related to the functions of the various parts of the brain. Different emotions lead to the different functional performance of various parts of the brain, and there are specific connections between different brain regions that are also related to emotions. Taking into account the ability of the capsule network to learn from special spatial information between the local part and the whole of the object, the capsule network is correspondingly applied to the multi-channel EEG emotion recognition. The capsule network can be used to characterize the connection among various channels of the EEG. Specifically, the primary capsule encodes the brain regions and the transformation matrix encodes the connections between various brain regions. These creative structures are conducive to extracting the intrinsic features of multi-channel EEG emotion recognition tasks. As shown in Tables 2–6, the original capsule network method can obtain significantly better performance than these of the three methods (CNN-RNN, gcForest, and DGCNN). The length of each capsule vector in the EmotionCaps layer expresses the existence of the emotional states and the classification loss can be calculated from it. We use the length of the capsule as the probability of predicting each class, the average of the probability of each class in the past 10 epochs is used as the pseudo label, and update the pseudo label and network parameters alternately. After the optimization of the network, the possibility of capsule prediction is more accurate, making the pseudo-label closer and closer to the clean label. Therefore, the alternate updating of the network parameters and pseudo labels can rectify noisy labels to improve the classification performance.

Among the seven methods of experimental comparison, the JO-CapsNet shows higher accuracy compared with other methods under different ratios of noisy labels (i.e., r = 0.1, 0.2, 0.3) in most circumstances, which can alternately update the pseudo label and parameters of capsule network until convergence. When the noise ratio is less than 0.3, they can correct the noisy labels to a certain extent, and finally achieve a classification accuracy of more than 90% for DEAP. For DREAMER, the classification accuracy is improved to the same extent, and the average classification accuracy in the dimensions of arousal and dominance can be improved to nearly 90%. In summary, the proposed JO-CapsNet is an excellent framework to solve the problem of multi-channel EEG emotion recognition in the presence of noisy labels.

5 Conclusion

We propose a framework using a joint optimization strategy based on a capsule network in the presence of noisy labels. Our proposed JO-CapsNet can well identify the internal connections between various EEG channels. With its good ability to "predict" class labels in the iterative process, the capsule network can promote updating pseudo labels correctly. The pseudo label and parameters of network update alternately until convergence during the training process, and hence improve the network's classification of multi-channel EEG emotion recognition. Besides, we conducted a lot of experiments on the DEAP and DREAMER datasets, which confirmed the effectiveness of our method. Within a certain range of noise ratios ($r = 0.1$, 0.2, 0.3), using our proposed JO-CapsNet can obtain better performance than these of the original capsule network. When the ratio of noise is 10%, the classification accuracies of capsule network using joint optimization strategy can increase to 94.33%/94.38% and 87.46%/80.04%/88.19%. When the noise ratio is 20% and 30%, the improvement of classification accuracies is more than 2% compared with the original capsule network under the same ratio of noisy labels, which also shows the excellent performance of our method.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61922075, 41901350, 32150017, 62176081, 62171176), National Defense Basic Scientific Research Program of China (Grant No. JCKY2019548B001), Fundamental Research Funds for the Central Universities (Grant Nos. JZ2021HGTB0078, JZ2021HGPA0061, PA2021KCPY0051), USTC Research Funds of the Double First-Class Initiative (Grant No. KY2100000123), Provincial Natural Science Foundation of Anhui (Grant No. 2008085QF285), and Anhui Key Project of Research and Development Plan (Grant No. 202104d07020005).

References

1 Zhang T, Wang X H, Xu X M, et al. GCB-Net: graph convolutional broad network and its application in emotion recognition. IEEE Trans Affect Comput, 2019. doi: 10.1109/TAFFC.2019.2937768

2 Shojaeilangari S, Yau W Y, Nandakumar K, et al. Robust representation and recognition of facial emotions using extreme sparse learning. IEEE Trans Image Process, 2015, 24: 2140–2152

3 Castellano G, Villalba S D, Camurri A. Recognising human emotions from body movement and gesture dynamics. In: Proceedings of International Conference on Affective Computing and Intelligent Interaction, 2007. 71–82

4 Vu H A, Yamazaki Y, Dong F, et al. Emotion recognition based on human gesture and speech information using RT middleware. In: Proceedings of IEEE International Conference on Fuzzy Systems, 2011. 787–791

5 Razak A, Yusof M H M, Komiya R. Towards automatic recognition of emotion in speech. In: Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology, 2003. 548–551

6 Guo H W, Huang Y S, Lin C H, et al. Heart rate variability signal features for emotion recognition by using principal component analysis and support vectors machine. In: Proceedings of the 16th International Conference on Bioinformatics and Bioengineering, 2016. 274–277

7 Silva D C, Vinhas V, Reis L P, et al. Biometric emotion assessment and feedback in an immersive digital environment. Int J Soc Robot, 2009, 1: 307–317

8 Liu M Y, Fan D, Zhang X H, et al. Human emotion recognition based on galvanic skin response signal feature selection and SVM. In: Proceedings of International Conference on Smart City and Systems Engineering, 2016. 157–160

9 Yang G, Yang S. Emotion recognition of electromyography based on support vector machine. In: Proceedings of the 3rd International Symposium on Intelligent Information Technology and Security Informatics, 2010. 298–301

10 Chen X, Li C, Liu A P, et al. Toward open-world electroencephalogram decoding via deep learning: a comprehensive survey. IEEE Signal Process Mag, 2022, 39: 117–134

11 Cheng J, Chen M Y, Li C, et al. Emotion recognition from multi-channel EEG via deep forest. IEEE J Biomed Health Inform, 2021, 25: 453–464

12 Li C, Tao W, Cheng J, et al. Robust multichannel EEG compressed sensing in the presence of mixed noise. IEEE Sens J, 2019, 19: 10574–10583

13 Adolphs R, Tranel D, Damasio A R. Dissociable neural systems for recognizing emotions. Brain Cognition, 2003, 52: 61–69

14 Li C, Wang B, Zhang S L, et al. Emotion recognition from EEG based on multi-task learning with capsule network and attention mechanism. Comput Biol Med, 2022, 143: 105303

15 Nie D, Wang X W, Shi L C, et al. EEG-based emotion recognition during watching movies. In: Proceedings of the 5th International IEEE/EMBS Conference on Neural Engineering, 2011. 667–670

16 Li C, Zhang Z Z, Song R C, et al. EEG-based emotion recognition via neural architecture search. IEEE Trans Affect Comput, 2021. doi: 10.1109/TAFFC.2021.3130387

17 Tao W, Li C, Song R C, et al. EEG-based emotion recognition via channel-wise attention and self attention. IEEE Trans Affect Comput, 2020. doi: 10.1109/TAFFC.2020.3025777

18 Li M, Lu B L. Emotion classification based on gamma-band EEG. In: Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009. 1223–1226

19 Patil A, Deshmukh C, Panat A. Feature extraction of EEG for emotion recognition using Hjorth features and higher order crossings. In: Proceedings of Conference on Advances in Signal Processing, 2016. 429–434

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

21 Duan R N, Zhu J Y, Lu B L. Differential entropy feature for EEG-based emotion classification. In: Proceedings of the 6th International IEEE/EMBS Conference on Neural Engineering, 2013. 81–84

22 Yang Y L, Wu Q F, Fu Y Z, et al. Continuous convolutional neural network with 3D input for EEG-based emotion recognition. In: Proceedings of International Conference on Neural Information Processing, 2018. 433–443

23 Song T F, Zheng W M, Song P, et al. EEG emotion recognition using dynamical graph convolutional neural networks. IEEE Trans Affect Comput, 2020, 11: 532–541

24 Alhagry S, Fahmy A A, El-Khoribi R A. Emotion recognition based on EEG using LSTM recurrent neural network. Emotion, 2017, 8: 355–358

25 Dose H, Møller J S, Iversen H K, et al. An end-to-end deep learning approach to MI-EEG signal classification for BCIs. Expert Syst Appl, 2018, 114: 532–542

26 Fahimi F, Zhang Z, Goh W B, et al. Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. J Neural Eng, 2019, 16: 026007

27 Porbadnigk A K, Görnitz N, Sannelli C, et al. When brain and behavior disagree: tackling systematic label noise in EEG data with machine learning. In: Proceedings of International Winter Workshop on Brain-Computer Interface, 2014. 1–4

28 Fayek H M, Lech M, Cavedon L. Modeling subjectiveness in emotion recognition with deep neural networks: ensembles vs soft labels. In: Proceedings of International Joint Conference on Neural Networks, 2016. 566–570

29 Koelstra S, Muhl C, Soleymani M, et al. DEAP: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput, 2012, 3: 18–31

30 Chen X, Xu X Y, Liu A P, et al. Removal of muscle artifacts from the EEG: a review and recommendations. IEEE Sens J, 2019, 19: 5353–5368

31 Tripathi S, Acharya S, Sharma R D, et al. Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017. 4746–4752

32 Yang Y L, Wu Q F, Qiu M, et al. Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network. In: Proceedings of International Joint Conference on Neural Networks, 2018. 1–7

33 Wu X, He R, Sun Z N, et al. A light CNN for deep face representation with noisy labels. IEEE Trans Inform Forensic Secur, 2018, 13: 2884–2896

34 Jiang J J, Ma J Y, Wang Z, et al. Hyperspectral image classification in the presence of noisy labels. IEEE Trans Geosci Remote Sens, 2019, 57: 851–865

35 Karimi D, Dou H, Warfield S K, et al. Deep learning with noisy labels: exploring techniques and remedies in medical image analysis. Med Image Anal, 2020, 65: 101759

36 Frenay B, Verleysen M. Classification in the presence of label noise: a survey. IEEE Trans Neural Netw Learn Syst, 2014, 25: 845–869

37 Zhu X Q, Wu X D. Class noise vs. attribute noise: a quantitative study. Artif Intell Rev, 2004, 22: 177–210

38 Ringeval F, Eyben F, Kroupi E, et al. Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data. Pattern Recogn Lett, 2015, 66: 22–30

39 Zhong P X, Wang D, Miao C Y. EEG-based emotion recognition using regularized graph neural networks. IEEE Trans Affect Comput, 2020. doi: 10.1109/TAFFC.2020.2994159

40 Hinton G E, Krizhevsky A, Wang S D. Transforming auto-encoders. In: Proceedings of International conference on artificial neural networks, 2011. 44–51

41 Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules. In: Proceedings of Advances in Neural Information Processing Systems, 2017. 3856–3866

42 Liu Y, Ding Y F, Li C, et al. Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network. Comput Biol Med, 2020, 123: 103927

43 Yin J H, Li S, Zhu H M, et al. Hyperspectral image classification using CapsNet with well-initialized shallow layers. IEEE Geosci Remote Sens Lett, 2019, 16: 1095–1099

44 Turan M A T, Erzin E. Monitoring infant's emotional cry in domestic environments using the capsule network architecture.

In: Proceedings of Interspeech, 2018. 132–136

45 Wang Y Q, Sun A X, Huang M L, et al. Aspect-level sentiment analysis using as-capsules. In: Proceedings of the World Wide Web Conference, 2019. 2033–2044

46 Afshar P, Mohammadi A, Plataniotis K N. Brain tumor type classification via capsule networks. In: Proceedings of the 25th IEEE International Conference on Image Processing, 2018. 3129–3133

47 Haffari G R, Sarkar A. Analysis of semi-supervised learning with the Yarowsky algorithm. 2012. ArXiv:1206.5240

48 Lee D H. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks. In: Proceedings of Workshop on Challenges in Representation Learning, 2013

49 Zhu X J. Semi-supervised learning literature survey. Computer Sci TR 1530, 2008

50 Du C D, Du C Y, Wang H, et al. Semi-supervised deep generative modelling of incomplete multi-modality emotional data. In: Proceedings of the 26th ACM International Conference on Multimedia, 2018. 108–116

51 Reed S, Lee H, Anguelov D, et al. Training deep neural networks on noisy labels with bootstrapping. 2014. ArXiv:1412.6596

52 Moore A W. Cross-validation for Detecting and Preventing Overfitting. Pittsburgh: School of Computer Science Carneigie Mellon University, 2001

53 Suykens J A K, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett, 1999, 9: 293–300