Computers in Biology and Medicine 143 (2022) 105303

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Emotion recognition from EEG based on multi-task learning with capsule network and attention mechanism

Chang Li [a,b], Bin Wang [a], Silin Zhang [c,*], Yu Liu [a], Rencheng Song [a], Juan Cheng [a], Xun Chen [d,e]

[a] *Department of Biomedical Engineering, Hefei University of Technology, Hefei, 230009, China*
[b] *Anhui Province Key Laboratory of Measuring Theory and Precision Instrument, Hefei University of Technology, Hefei, 230009, China*
[c] *Reproductive Medical Center, Renmin Hospital of Wuhan University, Wuhan, 430060, China*
[d] *Department of Neurosurgery, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, 230001, Anhui, China*
[e] *Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, 230026, China*

ARTICLE INFO

Keywords:
Deep learning
Emotion recognition
Electroencephalogram
Multi-task learning
Capsule network

ABSTRACT

Deep learning (DL) technologies have recently shown great potential in emotion recognition based on electroencephalography (EEG). However, existing DL-based EEG emotion recognition methods are built on single-task learning, *i.e.*, learning arousal, valence, and dominance individually, which may ignore the complementary information of different tasks. In addition, single-task learning involves a new round of training every time a new task appears, which is time consuming. To this end, we propose a novel method for EEG-based emotion recognition based on multi-task learning with capsule network (CapsNet) and attention mechanism. First, multi-task learning can learn multiple tasks simultaneously while exploiting commonalities and differences across tasks, it can also obtain more data from different tasks, which can improve generalization and robustness. Second, the innovative structure of the CapsNet enables it to effectively characterize the intrinsic relationship among various EEG channels. Finally, the attention mechanism can change the weight of different channels to extract important information. In the DEAP dataset, the average accuracy reached 97.25%, 97.41%, and 98.35% on arousal, valence, and dominance, respectively. In the DREAMER dataset, average accuracy reached 94.96%, 95.54%, and 95.52% on arousal, valence, and dominance, respectively. Experimental results demonstrate the efficiency of the proposed method for EEG emotion recognition.

1. Introduction

Emotion is a complex behavioral phenomenon involving many levels of neural and chemical integration [1], which affects people's cognition, behavior, and interpersonal communication [2]. Emotional recognition is a person's psychological response to external or self-stimuli, including the physiological responses that accompany this psychological response [3,4]. This aspect is an important interdisciplinary research topic in the fields of neuroscience, psychology, computer science, and artificial intelligence [5,6].

Emotion recognition is divided into two types of signals: physiological and non-physiological. Non-physiological signals include facial expressions, vocal intonations and body postures. By contrast, physiological signals can more flexibly reflect the dynamic changes of the human nervous system, making it difficult to hide real emotions [7]. Physiological signals, such as electrooculogram, electrocardiogram, and electromyogram, are caused by an indirect response to emotional changes; accordingly, the recognition accuracy is low. Meanwhile, electroencephalography (EEG) signals record the electrical wave changes during brain activity with good temporal resolution [8], which can offer a direct, more convenient, and comprehensive means to collect EEG signals for emotion recognition with higher classification accuracy [9]. Therefore, EEG signals have been widely used in emotion recognition and achieved high accuracy [10].

Currently, emotion models for EEG-based emotion recognition are mainly divided into two: discrete and dimensional models. Six basic types of emotions are involved in most discrete models including happiness, sadness, fear, surprise, anger, and disgust, and other

[*] Corresponding author. Reproductive Medical Center, Renmin Hospital of Wuhan University, Wuhan, 430060, China.
E-mail addresses: changli@hfut.edu.cn (C. Li), binwang@mail.hfut.edu.cn (B. Wang), rm003305@whu.edu.cn (S. Zhang), yuliu@hfut.edu.cn (Y. Liu), rcsong@hfut.edu.cn (R. Song), chengjuan@hfut.edu.cn (J. Cheng), xunchen@ustc.edu.cn (X. Chen).

https://doi.org/10.1016/j.compbiomed.2022.105303
Received 11 September 2021; Received in revised form 20 January 2022; Accepted 20 January 2022
Available online 19 February 2022

emotions are combined from these basic emotions. The dimensional model defines emotions as points in a dimensional space [11]. It involves valence (whether the emotion is positive or negative), arousal (the intensity of emotion), and dominance (represents the degree of subjective control on an individual's emotional state) [12]. Researchers have proposed a 2D or 3D model: valence-arousal and valence-arousal-dominance. The 2D model is widely used, and the details are shown in Fig. 1. The dimensional model can show more feelings than the discrete model and is closer to human's true perception of external things [13]. Therefore, we adopt the dimensional model in this study.

EEG-based emotion recognition tasks are mainly divided into two stages. The first stage is to extract effective features from the EEG signal to accurately represent the emotional state [14]. The two types of feature extraction are time domain and frequency domain [15]. Time domain feature extraction is a method of directly analyzing a system in the time domain that has intuitive and clear physical meaning, such as fractal dimension feature [16], statistical features (power, mean, standard deviation, and the first difference) [17–19], and higher-order crossings [20], *etc*. The frequency domain information of the signal, such as power spectral density [21], differential entropy [22], rational asymmetry and differential causality features, are examined in the widely utilized frequency domain features [23,24], *etc*. The second stage is to design a classifier to predict emotion labels based on the emotion features extracted from the signal. Traditional algorithms, such as naive Bayes, support vector machine (SVM), *k*-nearest neighbor and multilayer perceptron (MLP), are widely used and have achieved high accuracy rates. However, traditional methods feature extraction largely depends on human manual extraction, resulting in low generalization ability and accuracy of classifiers.

In recent years, deep learning (DL) has been applied in many fields, such as computer vision, natural language processing [25], speech recognition [26], and emotion recognition, and achieved state-of-the-art performance. Convolutional neural networks (CNNs) [27], dynamic graph convolutional neural networks (DGCNN) [28], recurrent neural network (RNN) [29] deep forest [30], and capsule network (CapsNet) are among the models that have been applied to EEG-based emotion recognition tasks [31]. Song et al. proposed a novel DGCNN that could dynamically learn the internal relationship between different EEG channels represented by an adjacency matrix to classify EEG emotions. The accuracy rates for valence, arousal, and dominance on the DREAMER database were 86.23%, 84.54%, and 85.02%, respectively [32]. Alhagry et al. proposed method based on an end-to-end recurrent neural network (RNN) to identify emotions from raw EEG signals [29]. Cheng et al. proposed a deep forest-based gcForest model for EEG-based emotion recognition to extract spatial and temporal features from the constructed 2D EEG frames [30]. Chao et al. proposed a DL framework based on a multiband feature matrix (MFM) and a CapsNet [33]. In the framework, the frequency domain, spatial characteristics, and frequency band characteristics of the multi-channel EEG signals were combined to construct the MFM. The CapsNet model was then introduced to recognize emotion states according to the input MFM.

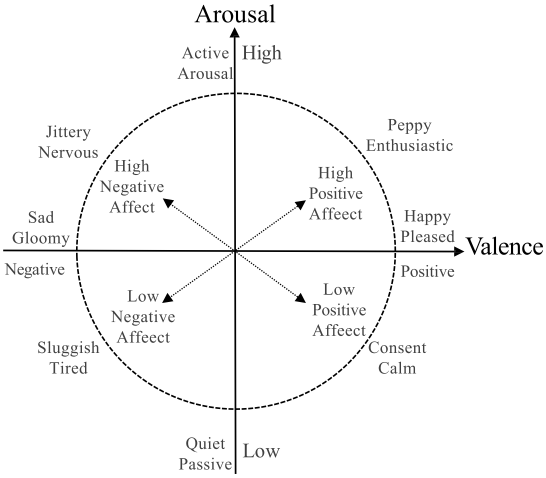

Fig. 1. The 2*D* valence-arousal model of emotions.

DL has shown desirable performance in EEG-based emotion recognition tasks. However, existing DL-based methods are based on single-task learning, *i.e.*, learning arousal, valence, and dominance individually, single-task learning involves a new round of training every time a new task appears, which not only ignores the complementary information of different tasks but also takes too much time. Humans can simultaneously learn multiple tasks in real life, and they can use the knowledge learned in a task to help learn another task, and vice versa, when these tasks are related. To solve the above mentioned shortcomings, we propose multi-task learning with CapsNet and attention mechanism. Multi-task learning varies from single-task learning, as shown in Fig. 2, since it aims at learning shared representations from multi-task supervisory signals. Multi-task DL has advantages over the single-task one as each task is solved by its own deep network [34]. First, multi-task learning can improve performance since related tasks share complementary information [35]. Second, the method can obtain more data from other related tasks to help alleviate the data scarcity problem [2]. Third, we introduce CapsNet, which can effectively characterize the intrinsic relationships between various EEG channels owing to their innovative structure. The capsule layer sharing representation can reduce the memory footprint and avoid the repeated calculation for features in the shared layers. Accordingly, CapsNet can yield fast learning speed and increase data efficiency. Finally, the attention mechanism can explore the information of the feature map by changing the weights of different channels. Thus, this mechanism can extract more important information about the EEG signal channels.

In this work, we propose a new approach named multi-task channel attention CapsNet (MTCA-CapsNet), which integrates CapsNet and the attention mechanism into a multi-task learning framework. The contributions of this work can be summarized as follows:

1. We propose a framework for EEG-based emotion recognition based on multi-task learning with CapsNet and the attention mechanism. First, multi-task learning can exploit the complementary information and obtain more data from different tasks, and hence it can improve generalization and robustness. Second, the innovative structure of the CapsNet can effectively characterize the intrinsic relationship among various EEG channels. Third, the attention mechanism can change the weight of different channels to extract important information.
2. Our method is compared with several traditional methods on DEAP and DREAMER datasets. In the DEAP database, the average accuracy rates for valence, arousal, and dominance reach 97.25%, 97.41%, and 98.35%, respectively. In the DREAMER database, the average accuracy rates for valence, arousal, and dominance reach 94.96%, 95.54%, and 95.52%, respectively. These results achieve better performance than the other methods on both datasets and demonstrate the effectiveness of our method.

The rest of this paper is organized as follows. Section 2 presents the construction and implementation details of the proposed method. Section 3 introduces the DEAP and DREAMER datasets, experimental setup, and results. Section 4 gives a discussion of the experimental results. Section 5 concludes our study.

2. Methods

In this section, we propose a multi-task CapsNet (MTCA-CapsNet). The details are shown in Fig. 3. First, we introduce the MTCA

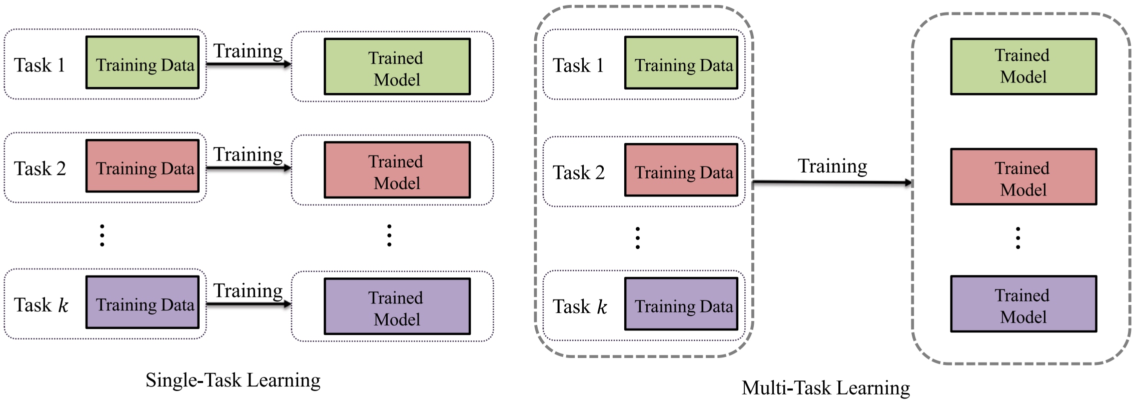

Fig. 2. This graph represents single-task learning compared to multi-task learning.

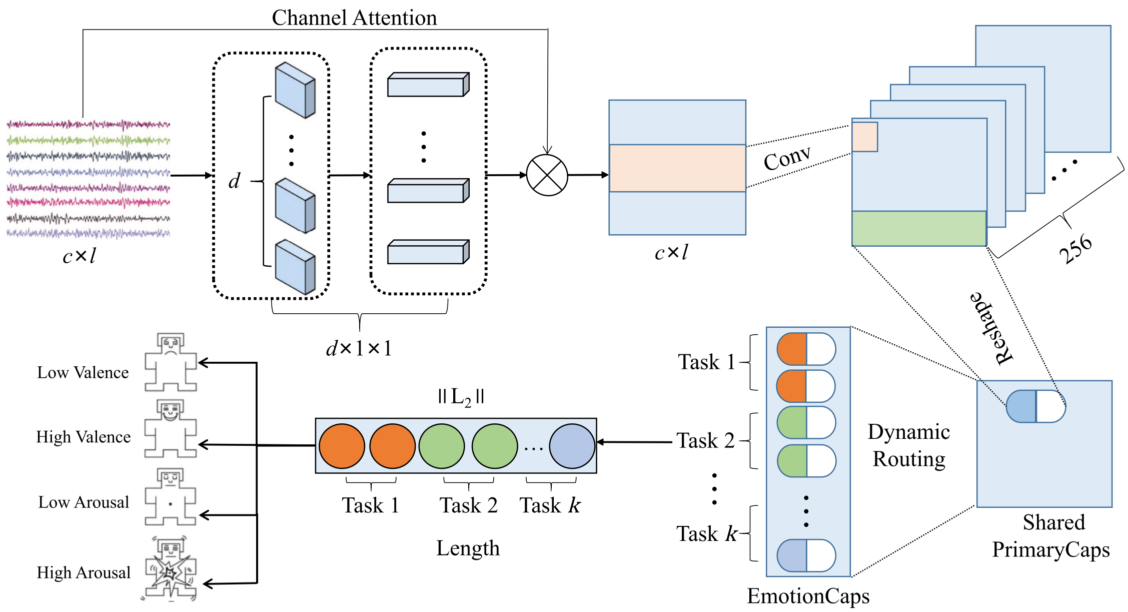

Fig. 3. MTCA-CapsNet for EEG signal classification, conv denotes the shared convolution.

mechanism of related work. We then describe the specific process of tailoring MTCA-CapsNet for multi-task EEG classification.

2.1. *Multi-task learning*

Single-task learning only focuses on a single task, where each individual task is solved separately by its own network. In a single classification task, the input EEG signal is represented as x, and the output is denoted as $\hat{y}$. Variable f denotes the deep neural network. The objective of single-task learning is to train f, which can transform each $\boldsymbol{x}$ into output $\hat{y}$. This step can be formulated as follows:

$$f(x) = \hat{y}, \tag{1}$$

The other data that may help optimize the metric are ignored during the learning process since single-task learning only considers a single task. Multi-task learning can obtain more data from different tasks, and it can learn more robust and universal representations for multiple tasks compared with single-task learning; hence, this method leads to better knowledge sharing among tasks. Collobert et al. [36] proposed the deep neural network for multi-task learning for the first time. They used CNNs to extract shared representations for different NLP tasks. The shared layer implicitly learns the relevant features of each word; thus, it is expected to improve the generalization performance when trained on the relevant task by enhancing these deeply generated features. In multi-task learning, this shared representation comes from related tasks, which can better generalize the original task. Given that k tasks $\boldsymbol{T} = [T_1, T_2, \ldots, T_k]$, the objective of multi-task learning is to train a jointly learning model F that will be able to make a prediction for samples $x^{(i)}(i = 1, 2, \ldots, k)$ from each task $T_i(i = 1, 2, \ldots, k)$, which can be denoted as follows:

$$F(x^{(1)}, x^{(2)} \cdots, x^{(k)}) = (\hat{y}^{(1)}, \hat{y}^{(2)}, \ldots, \hat{y}^{(k)}), \tag{2}$$

where $x^{(i)}$ denotes task instances from each task, and $\hat{y}^{(i)}(i = 1, 2, \ldots, k)$ represents the corresponding output [37].

2.2. *Attention mechanism*

Attention mechanism has made important breakthroughs in recent years in certain areas (e.g., image and natural language processing), and proven to be beneficial in improving model performance [38]. Raw EEG signal can contain spatial information through the intrinsic relationship between different channels [39]. The essence of attention mechanism is to locate the information of interest and suppress the useless information. The results are presented in the form of probability maps or feature vectors. In attention mechanism, the channel attention can consider the importance of different channels by paying attention to the channel and can squeeze the spatial information of multi-channel EEG signals to generate channel statistical information. Therefore, we use the channel attention mechanism in this work.

We adopt the channel attention mechanism for the raw EEG signal [40]. In this mechanism, we simultaneously use the average pool and max pool functions. The use of these two functions can greatly improve the network's presentation ability compared with utilizing them alone. First, the spatial information of the feature map is aggregated by using the average and max pooling operations. Next, this spatial information is fed to the shared network. Finally, we use element-wise summation and merge the output feature vectors to generate our channel attention map. The shared network consists of an MLP with a hidden layer, which uses the shared weights to reduce the training parameters of the network (Fig. 4). This step can be formulated as follows:

$$M_c(X) = \sigma(MLP(AP(X)) + MLP(MP(X))), \tag{3}$$

where $M_c \in \mathbb{R}^{d\times 1}$ denotes the channel attention map produced by the channel attention mechanism; d denotes the number of channels; AP denotes the average pooling; MP denotes the max pooling; and σ denotes the sigmoid function, which is used to output the probability of the event.

2.3. Multi-task CapsNet

We first introduce the CapsNet, which we name as the single-task CapsNet (ST-Capsule), for better understanding [41]. The features of a capsule are shown in Fig. 5. The ST-Capsule consists of three layers: the convolutional layer, the primary capsule layer (PrimaryCaps) and the emotional capsule layer (EmotionCaps). First, an input enters the convolutional layer and extracts low-level features through the convolutional layer. Then, the PrimaryCaps layer encodes the instances that have passed through the convolution layer into a vector that contains the EEG signal attributes. Finally, we feed the representation generated from the PrimaryCaps layer into the EmotionCaps via dynamic routing. Each task is assigned a capsule in the EmotionCaps, which is a set of neurons in the CapsNet that identifies the task and encodes the attributes of the task as a vector. In comparison with MTCA-CapsNet, ST-Capsule can only be assigned one capsule for each emotional state, and the tasks are trained independently of each other, which does not allow for training multiple tasks at once and learning the correlation between related tasks. Specifically, our proposed MTCA-CapsNet creates a more general form of learning by using information from relevant training signals in other related tasks instead of training each task individually. The MTCA-CapsNet also contains three layers: a shared convolutional layer, a shared PrimaryCaps layer, and an EmotionCaps layer. The convolutional layer learns shared feature representation from multiple related tasks. Then, the shared feature representations are fed into the shared PrimaryCaps to generate a vector. The EmotionCaps layer (top-level layer) is comprised of C EmotionCaps, in which each one corresponds to a task. The length of each capsule represents the probability that the input sample belongs to this task. The direction of each set of parameters preserves the characteristics of the features. Finally, a dynamic routing mechanism is implemented between shared PrimaryCaps and EmotionCaps, which connects the current EmotionCaps layer to the previous PrimaryCaps layer. This process not only captures the part?whole spatial relationship through the transformation matrix but also transfers information between the capsules by strengthening the connection of these capsules, which are allocated at different layers and obtain a high level of agreement.

First, we assume that $x \in \mathbb{R}^{c\times l}$ is a representation of the input, where l is the length and c is the number of electrode nodes. We feed such data into a shared convolutional layer that has 256 convolutional kernels with a stride of two and ReLU activation. The size of the convolutional kernels is determined by the shape of the input dataset, which is set to 9 × 9 and 6 × 6 for the DEAP and DREAMER datasets, respectively. The sharing mechanism in multi-task learning can capture more information from the EEG signals of different tasks and share complementary information from the arousal, valence and dominance tasks that are beneficial to one another. Accordingly, more robust and generic representations are learned, leading to better knowledge sharing among tasks. We will simultaneously train each task, which embeds the data representation of multi-tasks into the same space. Consequently, the next layer can share the complementary information between multi-tasks. We name this sharing mechanism as hard sharing and this layer as shared convolutional layer. This layer transforms the values of the sample points into the activities of the local feature detectors, which are then used as inputs to the shared PrimaryCaps.

The shared PrimaryCaps is a convolutional capsule layer with 32 channels of convolutional 8D capsules, which means that each primary capsule contains eight convolutional units with 9 × 9 (DEAP dataset) or 6 × 6 (DREAMER dataset) filters and a stride of one. The shared PrimaryCaps has a total of $k_1(k_1 = 32 \times h \times w)$ capsule outputs, where every output is an 8D vector. The length and orientation of each primary capsule, respectively represent the presence and properties of low-level features associated with emotional states. In this module, we form our shared PrimaryCaps, which can be used to share complementary information while training multiple tasks by sharing different tasks; it enables primary capsules to contain more information. Hence, the representation of the capsule is enhanced.

The last module in the entire MTCA-CapsNet network is EmotionCaps. This framework can simultaneously classify low/high valence,

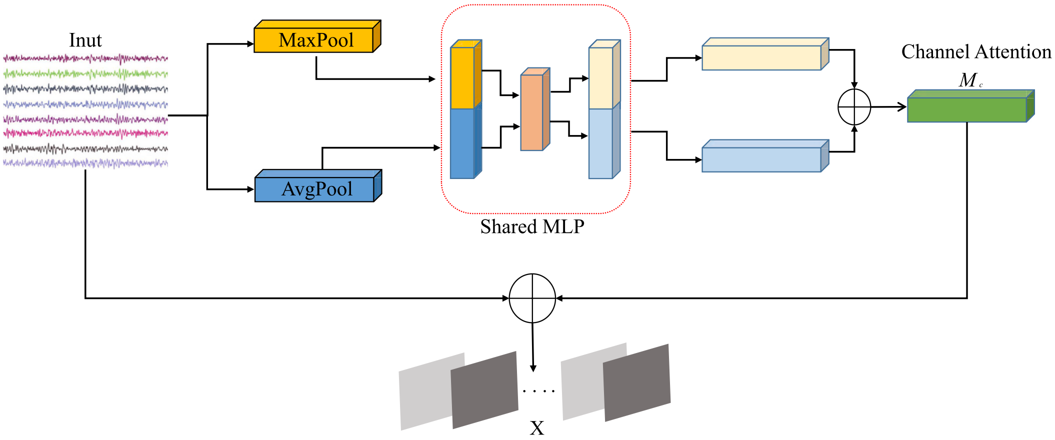

Fig. 4. The channel attention mechanism, X denotes EEG signals containing channel attention mechanisms.

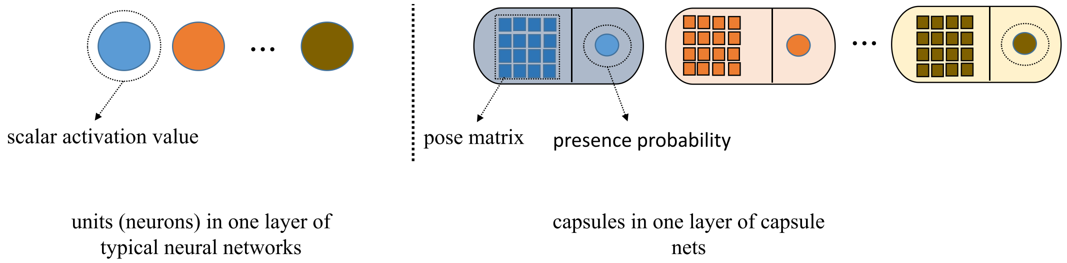
scalar activation value
pose matrix
presence probability
units (neurons) in one layer of typical neural networks
capsules in one layer of capsule nets

Fig. 5. The difference between capsules and conventional scalars.

arousal, and dominance since it is used to perform classification tasks. However, EmotionCaps in the DEAP and DREAMER datasets has k_2 16D emotion capsules, which means that k_2 emotion states are present. The shared PrimaryCaps and EmotionCaps are connected by a dynamic routing algorithm. Additional information about the dynamic routing protocols is provided in Fig. 6. The goal of dynamic routing is to design a better learning process than traditional pooling operations. This process preserves the positional relationships between the tasks through a transformation matrix that also captures the spatial relationships. The routing-by-agreement approach works well to cluster features into each class. We use this idea to cluster these features for different tasks and propose a task routing algorithm. First, we define a $\widehat{u}_{j|i}^{(k)}$, which represents a "prediction vector" (or higher-level emotional features) between two adjacent layers in the kth task. In multiple tasks, we multiply the output of the ith primary capsule $u_i(i = 1, 2, \ldots, k_1)$ by the weight matrix $W_{ij}(j = 1, 2, \ldots, k_2)$ to obtain the "prediction vector." This step can be formulated as follows:

$$\widehat{u}_{j|i} = W_{ij} \cdot u_i, \tag{4}$$

where W_{ij} is a transformation matrix between u_i and $\widehat{u}_{j|i}$. This transformation matrix is used to describe the spatial and positional relationships that exist between low-level and high-level emotional features.

We then sum up all the $\widehat{u}_{j|i}$ with different weights to obtain s_j as follows:

$$s_j = \sum_i c_{ij} \cdot \widehat{u}_{j|i}, \tag{5}$$

where c_{ij} is a coupling coefficient that represents the coefficient between the primary capsule i at the lth layer and the emotional capsule j at the lth layer. c_{ij} is obtained by computing the softmax function of b_{ij} as follows:

$$c_{ij} = \mathrm{softmax}(b_{ij}), \tag{6}$$

where c_{ij} is strictly limited to the range [0, 1], which denotes the probability that capsule i belongs to emotional capsule j. The initial logit b_{ij} is the log prior probability that the ith primary capsule should be coupled to the jth emotional capsule of multiple EEG-based emotion recognition tasks.

Finally, a nonlinear function called "squash" is applied to squash s_j between zero and one. This task is carried out to ensure that the output u_j has a length between zero and one. This step can be formulated as follows:

$$v_j = \frac{\|s_j\|_2^2}{1 + \|s_j\|_2^2} \frac{s_j}{\|s_j\|_2^2}. \tag{7}$$

We iteratively update the initial coupling coefficients by measuring the consistency between v_j and current output $\widehat{u}_{j|i}$, which uses the scalar product of $v_j \cdot \widehat{u}_{j|i}$, to refine the initially coupling coefficients. This step can be formulated as follows:

$$b_{ij} = b_{ij} + v_j \cdot \widehat{u}_{j|i}. \tag{8}$$

The above process determines how information flows between the capsules of the shared PrimaryCaps and the EmotionCaps.

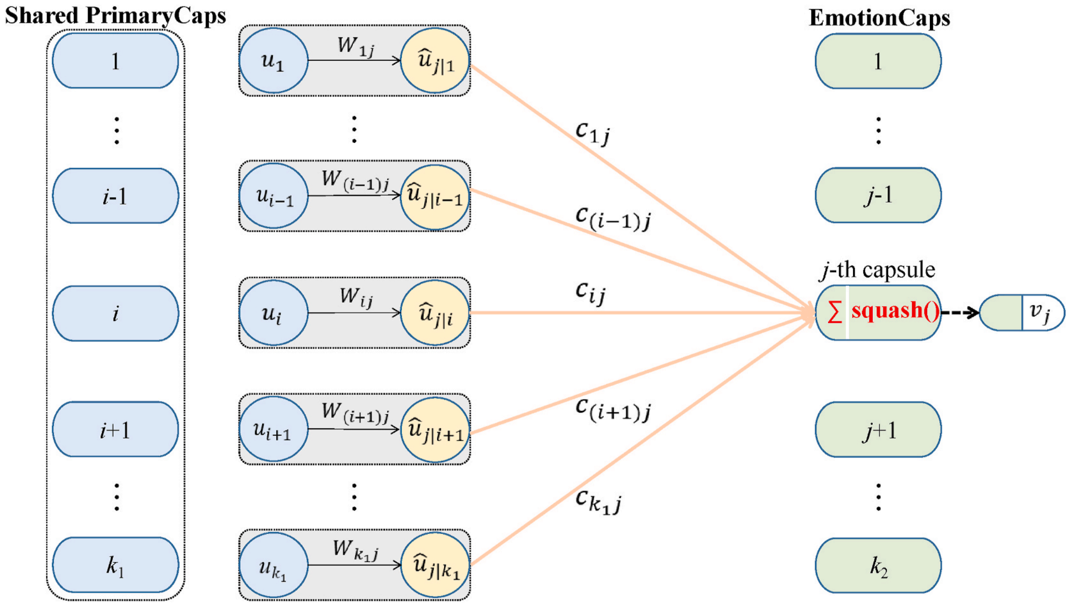

Fig. 6. The architecture of dynamic routing protocols.

2.4. Multi-task loss and training

In comparison with single-task learning, multi-task learning contains K tasks. We set $v_j = z_{kq}$, where v_j ($j = 1, 2, \ldots, k_2$), ($C = 2K$), and z_{kq} ($k = 1, 2, \ldots, K$), ($q = 1, 2$). We use margin loss for each EmotionCaps of the loss function of the MT-CapsNet. The goal is to increase the differences between classes. The detailed calculation is described as follows:

$$L_{kq} = \begin{aligned} & G_{kq}\max(0, m^{+} - \|z_{kq}\|_2)^2 + \\ & \lambda(1 - G_{kq})\max(0, \|z_{kq}\|_2 - m^{-})^2, \end{aligned} \tag{9}$$

where G_{kq} is the indicator function of the qth class for the kth task. If the qth class really exists in the kth task, then $G_{kq} = 1$; otherwise, $G_{kq} = 0$. m^{+} and m^{-} are the lower bound of the probability of existence and the upper bound of the probability of non-existence, which can be used to penalize positives and false negatives, respectively. We set $m^{+} = 0.9$ and $m^{-} = 0.1$ in this work, which means that $\|z_{kq}\|_2$ will not be greater than 0.1 if the qth class exists in the kth task. λ is the proportionality factor for adjusting the weight of losses caused by false positives and negatives. In this work, we set $\lambda = 0.5$, which means that the importance of penalizing false positives is roughly twice as important as penalizing false negatives. The total loss is the sum of the margin losses for each task in all classes as follows:

$$L = \sum_{1}^{K} \sum_{q=1}^{C} L_{kq}. \tag{10}$$

3. Experiments

In this section, we first introduce two public datasets for the proposed method. Then, we describe the relevant experimental setup in detail. Finally, we show the experimental results for the DEAP and DREAMER datasets and compare them with the state-of-the-art methods.

3.1. Databases

To better validate the performance of our proposed method, we use two public databases, namely, DEAP [42] and DREAMER [43], which are widely used in EEG-based emotion recognition. Table 1 demonstrates the form of the raw EEG data that we use in the two datasets.

The DEAP dataset recorded physiological signals from 32 healthy participants (16 males and 16 females). This dataset contained 32 channels of EEG signals and 8 channels of peripheral physiological signals. Each participant was asked to watch 40 (1-min) music videos. The EEG signals of the participants were recorded by 32 electrodes whose positions conformed to the international 10–20 system. The physiological signals were sampled at 512 Hz. The participants rated their valence, arousal, dominance, and liking on a continuous scale of one to nine after watching each video. A pre-processed version of the DEAP dataset was provided and employed in this work. In the pre-processed version, the EEG signals were down-sampled to 128 Hz. A bandpass frequency filter of 4.0–45.0 Hz was applied, and the eye artifacts were removed with a blind source separation technique, such as independent component analysis. The data recorded for each participant consisted of 40 segments of EEG data and the corresponding labels. Each EEG signal contained a 3s baseline signal recorded in the relaxed state and a 60s experimental signal recorded under stimulation.

Table 1
Database description.

Data format for each participant			
DEAP	Array	array shape	array content
	Data	40 × 32 × 8064	videos × channels × data
	Labels	40 × 2	videos × label (V, A)
DREAMER	ExperData	18 × 14 × 25 472 (M)	videos × channels × data
	BaseData	18 × 14 × 7808	videos × channels × data
	Labels	18 × 3	videos × label (V, A, D)

V means valence, A represents arousal, D refers dimonance, and M denotes mean.

The DREAMER dataset contained EEG data from 23 subjects (14 males and 9 females). These EEG signals were collected via 14 EEG electrodes while the subjects were watching 18 movie clips. Each participant was shown 18 movie clips. Every movie clip targeted one of the nine emotions: entertainment, excitement, happiness, calmness, anger, disgust, fear, sadness, and surprise. The movie clip played continuously from 65s to 393s, and this duration was considered sufficient to evoke a single emotion. The average time duration of the film clips was 199s. The data collection began with the subjects watching neutral movie clips to help them recover to a neutral emotional state in each new data collection experiment, which was used as a baseline signal. The EEG signals were recorded from 14 electrodes at a sampling rate of 128 Hz, and their positions conformed to the requirements of the international 10–20 system. The EEG signals were filtered with bandpass Hamming sinc linear phase FIR filters. The artifact subspace reconstruction method was used for artifact removal. After watching the movie clips, the participants rated their arousal, valence, and dominance levels on a scale from one to five. Finally, the data recorded for each participant consisted of three parts: 18 baseline signal segments corresponding to the relaxed state, 18 experimental signal segments, and 18 corresponding labels.

3.2. Experimental design

In this subsection, we present the specific details of the two datasets in the experiment. First, the size of the sliding window is set. Next, the preprocessing of data and labels is presented. Then, the detailed parameters in MTCA-CapsNet emotion recognition are described. Finally, the manner by which to divide the test and training sets is illustrated.

Window size setting: The duration of emotions is approximately 0.5–4s [44]. Numerous experiments have shown that dividing EEG data into segments of 1s yields the highest classification accuracy [45]. Therefore, we set the window size to 1s.

Data processing: In the DEAP dataset, each signal of the DEAP dataset can be divided into 60 segments after segmentation of the preprocessed experimental signals by using a 1s sliding window that contains 128 sampling points. Approximately 2400 (40 trials × 60 segments) samples can be acquired in each subject since every DEAP subject has 40 experimental signals. Accordingly, each EEG sample in the DEAP dataset is presented as a 32 × 128 matrix. The length of each experimental signal is different in the DREAMER database. Consequently, we obtained a different number of EEG samples for each experimental signal in the DREAMER dataset. We obtained 3728 EEG samples for each subject by using the same windowing technique. Hence, a 14 × 128 matrix is used for each EEG sample.

Label processing: Each segment of the signal divided by the sliding window is labeled with the corresponding segment of the signal. In the DEAP dataset, we divide the labels into scores from one to nine and set five as the threshold (high/low arousal and valence). Specifically, the label is high when the score is greater than five and low if it is not more than five (low: ≤ 5, high: > 5). In the DREAMER dataset, the labels are divided into one to five scores (high/low arousal, valence, and dominance), and the label is high when the rank is greater than three and low, otherwise (low: ≤ 3, high: > 3).

Parameter settings in the MTCA-Capsule: In the whole MTCA-Capsule architecture, we use the margin loss as the loss function of the network. Adam optimizer is used to optimize the loss function. We utilize 256 convolutional filters in the shared convolutional layer. The filter size was set to 9 × 9 (DEAP dataset) and 6 × 6 (DREAMER dataset). Then, one capsule in the PrimaryCaps consisted of eight neurons while one capsule in the EmotionCaps layer had 16 neurons. We set the respective number of epochs, batch size, and learning rate to 30, 100, and 10^{-5} (DEAP dataset) and 30, 100, and 10^{-4} (DREAMER dataset). We also set the maximum number of iterations to 3. We implement our approach through the PyTorch framework. The coupling coefficient c_{ij} is updated by the dynamic routing algorithm, and the connectivity between the capsules is determined. The initial value of b_{ij} in dynamic routing is set to zero. The size of the transpose matrix W_{ij} is set to 8 × 16. We set the value of the routing iteration to three by experience to find the best update routing iteration for the coupling coefficient and obtain higher average classification accuracy.

Division of training/test sets: We use 10-fold cross-validation for the partitioning of the training and test sets of the two datasets [46].

3.3. Comparison of the DEAP datasets

We compare our method with three traditional methods and two DL methods on the DEAP dataset to further validate our proposed MTCA-CapsNet approach. The approach includes the decision tree (DT) [47], SVM [48], MLP [49], 3D-CNN [50], and DGCNN [28]. We also compare our approach with single-task learning CapsNet (ST-Capsule) and multi-task learning CapsNet (MT-Capsule) approaches to better demonstrate the effectiveness of MTCA-CapsNet. The inputs of the DT, SVM and MLP are DE features that were extracted from the θ (4–7 Hz), α (8–13 Hz), β (14–30 Hz), and γ (31–50 Hz) bands of the pre-processed EEG signal. DGCNN can dynamically learn the internal relationships between different EEG channels represented by the adjacency matrix. The internal relationship between different EEG channels is represented by the adjacency matrix, and it is used to classify EEG emotions. 3DCNN extracted 3D data representation from multi-channel EEG signals and sent it to the proposed 3DCNN model for spatiotemporal feature extraction. The ST-Capsule uses capsules to encode entities and a transformation matrix to encode the intrinsic spatial relationships between the parts and whole of EEG signals. Hence, the ST-Capsule can efficiently represent the relationship between the parts of EEG signals. The MT-Capsule only uses the CapsNet to simultaneously learn multiple tasks. Contrary to our approach, the MT-Capsule does not consider the importance of different channels of the raw EEG signals. A few of the abovementioned methods and our proposed method use the same data processing. These methods include removing the baseline and using the same slice length for training/test set division to ensure the fairness of the comparison experiments.

Table 2 shows the mean accuracy and standard deviation of the 32 subjects on the valence, arousal, and dominance classification tasks for the DEAP dataset. The results show that our method achieves the best performance in both dimensions compared with the other seven methods. First, our method improves the classification accuracy by approximately 7%, 4%, and 6% on three classification tasks compared with two state-of-the-art CNN-based methods (3DCNN and DGCNN). Second, our method improves the classification accuracy in three dimensions by approximately 17%, 15%, and 14% on average compared with the three traditional methods (DT, MLP, and SVM). Third, the MT-Capsule can achieve higher recognition accuracy than the ST-Capsule in terms of valence, arousal, and dominance dimensions. This result demonstrates the effectiveness of multi-task learning compared with single-task learning in the emotion recognition task. Table 2 illustrates that our method can achieve higher recognition accuracy than the MT-Capsule in terms of valence, arousal, and dominance dimensions. This result demonstrates that adding a channel attention mechanism for the MT-Capsule is effective. Besides, our approach has the smallest standard deviation among all approaches, indicating its high stability among different subjects. Figs. 7–9 show the recognition accuracies of the three dimensions for each participant by using the eight methods. The average accuracy of each participant is calculated via 10-fold cross-validation. By contrast, our method has a clear advantage over other methods in terms of higher classification accuracy and stability.

Table 2
Average accuracies and standard deviations (%) of different methods on deap database.

Method	Valence	Arousal	Dominance
SVM	86.60 ± 6.98	87.43 ± 6.62	89.14 ± 6.67
DGCNN	92.55 ± 3.53	93.50 ± 3.93	93.50 ± 3.75
3DCNN	89.45 ± 4.51	90.42 ± 3.72	90.25 ± 4.95
MLP	87.73 ± 6.30	88.88 ± 5.08	88.75 ± 5.53
DT	68.28 ± 4.12	71.16 ± 6.12	73.36 ± 7.82
ST-Capsule	96.36 ± 2.14	95.61 ± 3.06	97.65 ± 1.38
MT-Capsule	96.69 ± 1.73	96.84 ± 1.81	97.73 ± 1.15
Ours	**97.24 ± 1.58**	**97.41 ± 1.47**	**98.35 ± 1.28**

3.4. Comparison of the DREAMER datasets

In the DREAMER dataset, we validate the effectiveness of our proposed method by using the three dimensions of arousal, valence, and dominance.

Table 3 shows the average accuracy and standard deviation of the valence, arousal and dominance classification tasks on DREAMER for 23 participants. The recognition accuracy rates of our proposed method MTCA-CapsNet in the DREAMER database for the three dimensions of valence, arousal and dominance are 94.96%, 95.54%, and 95.52%, respectively. First, our method improves the classification accuracy in three dimensions by approximately 19%, 11% and 8% on average compared with the three traditional methods (DT, SVM, and MLP). Second, our method improves the classification accuracy in three dimensions by approximately 10% and 6% on average compared with the two state-of-the-art CNN-based methods (3DCNN and DGCNN). Finally, we compare the proposed approach with single-task and multi-task learning CapsNet. Table 3 illustrates that the accuracy rates of the MT-Capsule method for valence, arousal, and dominance were 94.54%, 95.01% and 95.12%, respectively. The MT-Capsule improved the recognition accuracy in all three dimensions compared with the ST-Capsule. This result demonstrates that multi-task learning is more effective than single-task learning in emotion recognition. MTCA-CapsNet improved the recognition accuracy in all three dimensions compared with the MT-Capsule. This result demonstrates that adding the channel attention mechanism for the MT-Capsule is effective.

Table 3 provides the mean accuracy and standard deviation of the 23 participants by using these methods. These results show that our method obtains a recognition accuracy that is higher than that of the ST-Capsule and much higher than those of other methods. Our method still has higher accuracy compared with the superior performance of the MT-

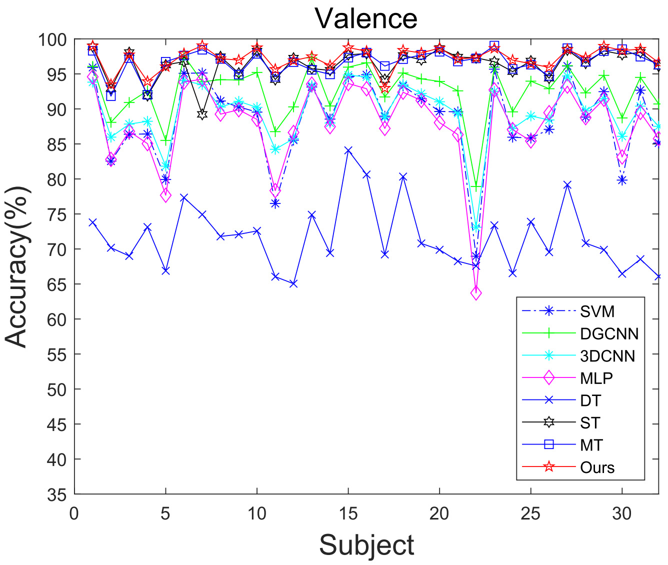

Fig. 7. Performance comparison of each subject using different methods for valence on DEAP database.

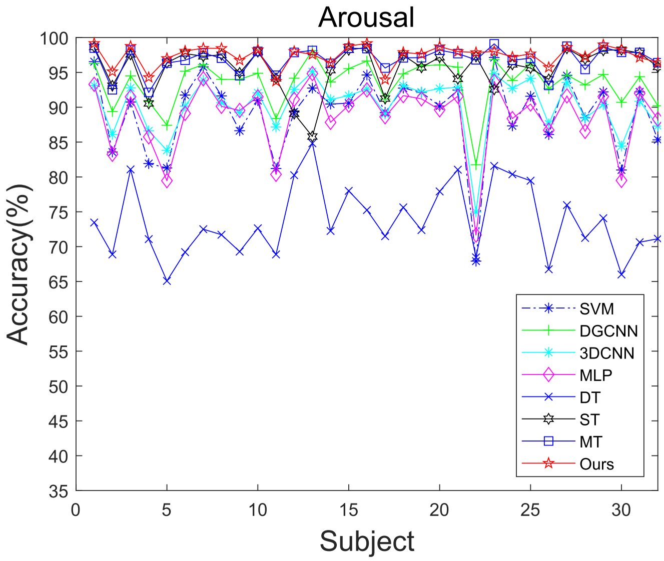

Fig. 8. Performance comparison of each subject using different methods for arousal on DEAP database.

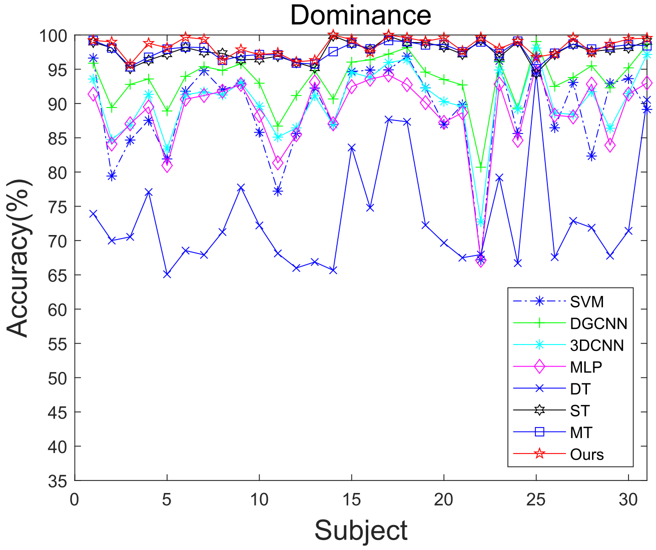

Fig. 9. Performance comparison of each subject using different methods for dominance on DEAP database.

Table 3
Average accuracies and standard deviations (%) of different methods on dreamer database.

Method	Valence	Arousal	Dominance
SVM	87.14 ± 5.20	87.03 ± 4.88	87.18 ± 4.87
DGCNN	89.59 ± 5.13	88.93 ± 3.93	88.63 ± 5.13
3DCNN	84.54 ± 5.00	84.84 ± 4.86	85.05 ± 4.96
MLP	83.64 ± 5.97	83.71 ± 5.39	83.90 ± 5.32
DT	75.53 ± 6.71	75.74 ± 6.44	76.40 ± 5.68
ST-Capsule	93.71 ± 4.19	94.03 ± 4.42	94.11 ± 4.63
MT-Capsule	94.54 ± 3.93	95.01 ± 3.88	95.12 ± 3.96
Ours	**94.96 ± 3.60**	**95.54 ± 3.63**	**95.52 ± 3.78**

Capsule. This result provides a stronger indication that our proposed MTCA-CapsNet approach is effective in the emotion recognition task. Our method is more stable than the other methods in terms of standard deviation. The recognition accuracy rates of each participant with these methods on the three dimensions of valence, arousal, and dominance are shown in Figs. 10–12. The proposed method achieves the highest recognition accuracy for most of the participants.

3.5. *Comparison with other studies*

Finally, we compared the performance of our proposed method with previous studies. To make the comparison experiments equitable, we selected studies related to emotion classification using the DEAP and DREAMER datasets. Tables 4 and 5 show the specific details in the DEAP and DREAMER datasets, respectively. From the Tables, we can see that our proposed method outperforms other studies on the DEAP and DREAMER datasets. On the DEAP dataset, our method achieves the highest accuracy rates of 97.24%, 97.41%, and 98.35% for valence, arousal, and dominance, respectively. Moreover, our method outperforms the second highest recognition study by approximately 5% [51]. On the DREAMER dataset, our method achieves the highest accuracy rates of 94.96%, 95.54%, and 95.52% for valence, arousal, and dominance in Table 4. The accuracy also improved by approximately 5% compared to the second highest accuracy [52] in Table 5. In addition, compared to other references [43,51–55,55–59], our method uses the raw EEG signal as input, which eliminates the complex process of manually extracting features.

3.6. *Experimental results using Kappa coefficients*

To further demonstrate the effectiveness of the proposed method, the standard Kappa coefficient was used in the experiment as a verification of the superiority of the method. Kappa coefficient is a metric used to test consistency and measure the effectiveness of classification. For the classification problems, the so-called consistency refers to whether the model prediction results and the actual classification results are consistent [60]. The closer the value is to 1, the better the classification performance is. As can be seen in Table 6, our Kappa coefficient is above 0.9, which can demonstrate the superiority of the proposed method and indicates that the predicted values are closer to the true values.

4. Discussion

EEG-based emotion recognition is a hot issue in the field of human-computer interaction in recent years. Many researchers have suggested a number of effective classification models that have produced good

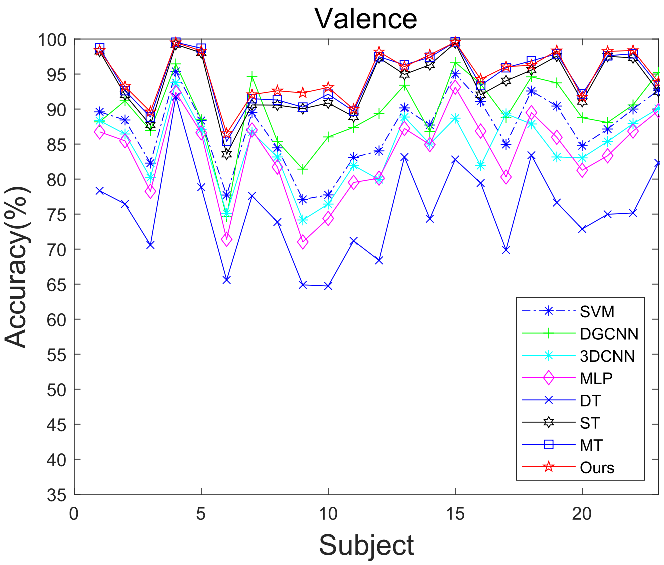

Fig. 10. Performance comparison of each subject using different methods for valence on DREAMER database.

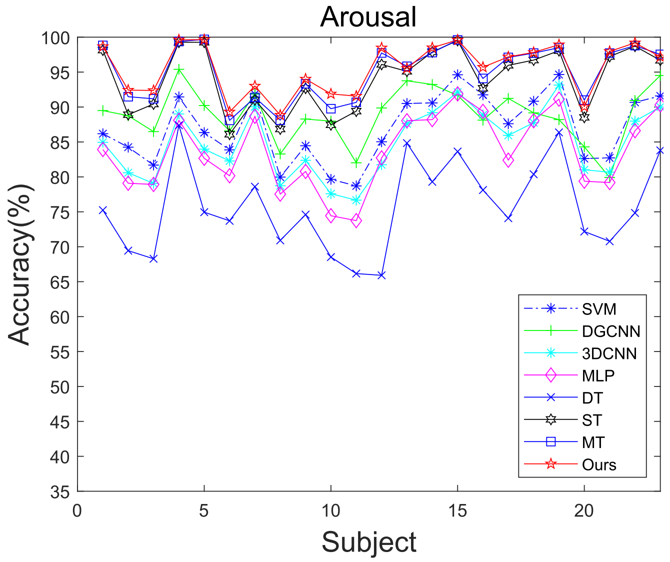

Fig. 11. Performance comparison of each subject using different methods for arousal on DREAMER database.

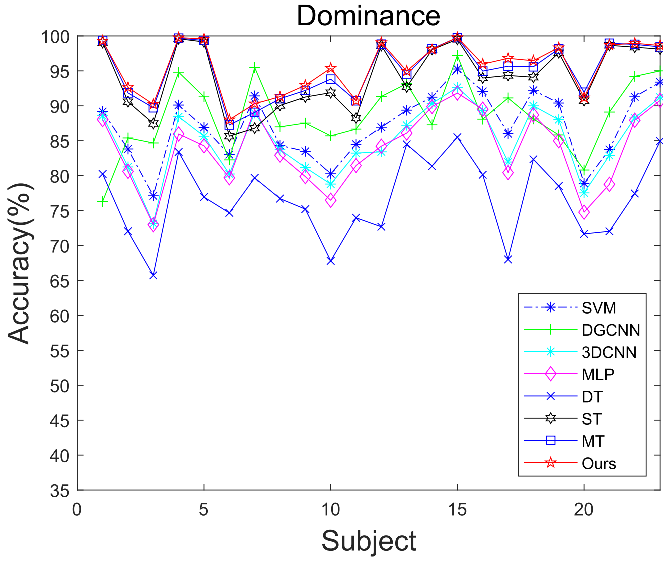

Fig. 12. Performance comparison of each subject using different methods for dominance on DREAMER database.

results. However, some limitations still persist. Humans can simultaneously learn multiple tasks, and they can use the knowledge learned in a task to help learn another task, and vice versa, when these tasks are related. However, existing DL methods were proposed on the basis of single-task learning. Single-task learning only focuses on a single task, where each individual task is separately solved by its own network for EEG-based emotion recognition. This study aims to find a system capable of recognizing multiple tasks and considering the correlation between tasks for the purpose of multi-task EEG-based emotion recognition. In this work, we provide an MTCA-Capsule model tailor for multi-task EEG-based emotion recognition considering the information between tasks. The superior recognition performance of our method during the experiment is most likely due to the following major points:

1. In contrast to traditional and DL methods, we adopt the idea of multi-task learning in the field of EEG-based emotion recognition. To achieve desirable performance, single-task learning often requires a large-scale annotated EEG dataset for supervised DL, which is almost impossible to acquire due to the high-cost of data acquisition and accurate annotation. By contrast, multi-task learning can improve the performance since related tasks share complementary information. This mechanism can also obtain more information from other related tasks to help alleviate the data scarcity problem.
2. Raw EEG signal can contain spatial information through the intrinsic relationship between different channels. In attention mechanism, the channel attention can consider the importance of different channels by paying attention to the channel and squeeze the spatial information of multi-channel EEG signals to generate channel statistical information. In the channel attention mechanism, we simultaneously use the average and max pooling functions, which can greatly improve the presentation ability of the network compared with utilizing them alone. These innovative structures are beneficial for extracting features in emotion recognition tasks.
3. Our framework MTCA-Capsule adopts capsules to encode entities. Capsules are local invariant groups of neurons that learn to recognize the presence of entities and encode their properties as vectors. The dynamic routing mechanism is implemented between shared PrimaryCaps and EmotionCaps, which connects the current EmotionCaps layer to the previous PrimaryCaps layer. This process not only captures the part-whole spatial relationship of the EEG signals through the transformation matrix but also transfers information between the capsules by strengthening the connection of these capsules. The dynamic routing mechanism approach effectively clusters features for different classes as well.

To further highlight the contribution of our proposed method. We have collected results from recently published literature, and they all use the DEAP and DREAMER datasets. Tables 7 and 8 show the detailed comparison results on the DEAP and DREAMER datasets, respectively. Compared with CNN-based methods, the proposed MTCA-CapsNet achieves the highest average recognition accuracy. It is verified that each capsule in the capsule network contains more useful information than neurons in CNN. In addition, our method can obtain better results than the ACRNN, which can demonstrate the efficiency of our attention

Table 4
Details of the DEAP dataset in previous studies.

Algorithms	Features	subject number	channel number	Cross Validation	Accuracy (%)		
					Valence	Arousal	Dominance
Bagging [53]	The Mean Energy	32	32	5-fold cross validation	63.97	63.97	63.97
AdaBoosting [53]	The Mean Energy	32	32	5-fold cross validation	66.95	66.95	66.95
Rotation Forest [54]	Tunable Q Wavelet Transform	32	32	Holdout validation	93.1	-	-
LSTM [29]	Raw EEG signals	32	32	4-fold cross validation	85.39	85.04	-
KNN [55]	Statistical features	32	16	Leave-one-out cross validation	89.91	89.74	-
Ensemble DBN [56]	Statistical, power, and HHS features	32	40	10-fold cross validation	75.97	76.07	-
Fusion CNN [57]	EEG Spectrograms and GSR features	32	42	Leave-one-out cross validation	80.32	75.97	-
Dense CNN [51]	DE features	32	36	10-fold cross validation	92.04	92.96	-
Bi-LSTM [58]	Higher order statistics	32	40	10-fold cross validation	83.66	84.00	-
MTCA-CapsNet	Raw EEG signals	32	32	10-fold cross validation	**97.24**	**97.41**	**98.35**

Table 5
Details of the DREAMER dataset in previous studies.

Algorithms	Features	subject number	channel number	Cross Validation	Accuracy (%)		
					Valence	Arousal	Dominance
SVM [43]	PSD	25	14	10-foldcross validation	62.49	62.17	61.84
VGG-16 network [59]	EEG-PSD images-based Deep-Learning features	23	14	Leave-one-out cross validation	78.99	79.23	-
Deep CCA [52]	DE and HHS features	32	40	Leave-one-out with Glia Chains	90.57 cross validation	88.99	90.67
MTCA-CapsNet	Raw EEG signals	23	14	10-fold cross validation	**94.96**	**95.54**	**95.52**

Table 6
Kappa coefficients and standard deviations of the proposed method for deap and dreamer database.

database	Valence	Arousal	Dominance
DEAP	0.947 ± 0.032	0.939 ± 0.037	0.953 ± 0.022
DREAMER	0.920 ± 0.056	0.934 ± 0.044	0.929 ± 0.047

Table 7
Details of the DEAP dataset in the state-of-the-art studies.

Algorithms	Features	Cross Validation	Accuracy (%)		
			Valence	Arousal	Dominance
Rotation Forest [54]	Tunable Q Wavelet Transform	Holdout validation	93.1	-	-
Dense CNN [51]	DE features	10-fold cross validation	92.04	92.96	-
CNN [62]	EMD/IMF CNN and VMD feature	8-fold cross validation	95.20	95.49	-
DNN [63]	CNN and SAE feature	80% for training and 20% for testing	89.49	92.86	-
ACRNN [39]	DE feature	10-fold cross validation	93.72	93.38	-
MTCA-CapsNet	Raw EEG signals	10-fold cross validation	**97.24**	**97.41**	**98.35**

Table 8
Details of the DREAMER dataset in the state-of-the-art studies.

Algorithms	Features	Cross Validation	Accuracy (%)		
			Valence	Arousal	Dominance
ARF [61]	multi-scale spectral and temporal entropies	10-fold validation	86.20	85.40	84.50
CNN and SVM [64]	HOLO-FM	10-fold cross validation	88.20	90.43	-
MTCA-CapsNet	Raw EEG signals	10-fold cross validation	**94.96**	**95.54**	**95.52**

mechanism. Moreover, these recently published methods use single task for learning. While our method can use multiple tasks for learning to improve the recognition accuracy. Bhattacharyya et al. proposed the sparse autoencoder based random forest (ARF), but it needs to extract the multi-scale spectral and temporal entropy feature manually [61]. It can be seen from Tables 7 and 8 that our proposed method achieves the highest accuracy on two public datasets, *i.e.,* DEAP and DREAMER datasets, verifying the effectiveness of the proposed method.

Owing to the advantages of multi-task learning, it allows multiple tasks to be learned together and in this way improves the accuracy rate. The proposed method only takes the multiple recognition tasks into consideration, and so it is necessary to add other auxiliary tasks to further improve the recognition accuracy. For example, it is desirable to consider both recognition and reconstruction tasks simultaneously. In addition, the proposed method adopts the hard sharing mechanism for multi-task learning. In the future, it is recommended to adopt a more suitable sharing mechanism to improve the emotion recognition performance, such as the soft sharing mechanism. We will also adapt our method to the EEG-based emotion recognition in the subject-independent scenario.

5. Conclusion

In this work, we propose emotion recognition from EEG based on multi-task learning with CapsNet and attention mechanism (MTCA-CapsNet). Our proposed method can efficiently recognize the intrinsic relationship between various EEG tasks. First, we extract the feature probability maps from different channels and combine them with the original input. The shared parameters are then generated by entering the shared convolutional layer and shared PrimaryCaps layer. Finally, the parameters are classified after the EmotionCaps layer. In this way, the MTCA-CapsNet is formed to improve the recognition accuracy between tasks. The proposed framework is used to conduct experiments on the DEAP and DREAMER datasets. The average accuracy rates of our framework on the DEAP dataset for valence, arousal, and dominance are 97.24%, 97.41%, and 98.35%, respectively. The average accuracy rates on the DREAMER dataset for valence, arousal, and dominance are 94.96%, 95.54%, and 95.52%, respectively. The experimental results show that MTCA-CapsNet can significantly outperform several other compared models, the ST-Capsule, and the MT-Capsule.

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, 32150017 and 62176081), the National Defense Basic Scientific Research Program of China (Grant JCKY2019548B001), the Fundamental Research Funds for the Central Universities (Grants JZ2021HGTB0078 and PA2021K-CPY0051), the Provincial Natural Science Foundation of Anhui (Grant 2008085QF285), the Anhui Key Project of Research and Development Plan (Grant 202104d07020005).

References

[1] L.F. Barrett, B. Mesquita, K.N. Ochsner, J.J. Gross, The experience of emotion, Annu. Rev. Psychol. 58 (2007) 373–403.

[2] X. Chen, C. Li, A. Liu, M. J. McKeown, R. Qian, Z. J. Wang, Toward open-world electroencephalogram decoding via deep learning: a comprehensive survey, IEEE Signal Processing Magazine, doi: 10.1109/MSP.2021.3134629.
[3] C. Li, Z. Zhang, R. Song, J. Cheng, Y. Liu, X. Chen, Eeg-based emotion recognition via neural architecture search, IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2021.3130387.
[4] H. Cui, A. Liu, X. Zhang, X. Chen, K. Wang, X. Chen, Eeg-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network, Knowledge-Based Systems 205 (2020) 106243.
[5] Y. Liu, Y. Ding, C. Li, J. Cheng, R. Song, F. Wan, X. Chen, Multi-channel eeg-based emotion recognition via a multi-level features guided capsule network, Computers in Biology and Medicine 123 (2020) 103927.
[6] C. Li, W. Tao, J. Cheng, Y. Liu, X. Chen, Robust multichannel eeg compressed sensing in the presence of mixed noise, IEEE Sensors Journal 19 (22) (2019) 10574–10583.
[7] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, X. Yang, A review of emotion recognition using physiological signals, Sensors 18 (7) (2018) 2074.
[8] 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 transactions on affective computing 3 (1) (2011) 18–31.
[9] A. Subasi, E. Ercelebi, Classification of eeg signals using neural network and logistic regression, Computer methods and programs in biomedicine 78 (2) (2005) 87–99.
[10] K. S. Kamble, J. Sengupta, Ensemble machine learning-based affective computing for emotion recognition using dual-decomposed eeg signals, IEEE Sensors Journal, doi: 10.1109/JSEN.2021.3135953.
[11] T. Eerola, J.K. Vuoskoski, A comparison of the discrete and dimensional models of emotion in music, Psychology of Music 39 (1) (2011) 18–49.
[12] H. Rabitz, Ö.F. Aliş, General foundations of high-dimensional model representations, Journal of Mathematical Chemistry 25 (2) (1999) 197–233.
[13] T.J. Trull, C.A. Durrett, Categorical and dimensional models of personality disorder, Annu. Rev. Clin. Psychol. 1 (2005) 355–380.
[14] W. Ting, Y. Guo-Zheng, Y. Bang-Hua, S. Hong, Eeg feature extraction based on wavelet packet decomposition for brain computer interface, Measurement 41 (6) (2008) 618–625.
[15] V. Srinivasan, C. Eswaran, N. Sriraam, Artificial neural network based epileptic detection using time-domain and frequency-domain features, Journal of Medical Systems 29 (6) (2005) 647–660.
[16] C. Traina Jr., A. Traina, L. Wu, C. Faloutsos, Fast feature selection using fractal dimension, Journal of Information and data Management 1 (1) (2010), 3–3.
[17] K.G. Jöreskog, D. Sörbom, S. Du Toit, LISREL 8: New Statistical Features, Scientific Software International, 2001.
[18] M. Müller, F. Kurth, M. Clausen, Audio matching via chroma-based statistical features, ISMIR 2005 (2005) 6.
[19] M. Zhang, Statistical features of human exons and their flanking regions, Human molecular genetics 7 (5) (1998) 919–932.
[20] P.C. Petrantonakis, L.J. Hadjileontiadis, Emotion recognition from eeg using higher order crossings, IEEE Transactions on information Technology in Biomedicine 14 (2) (2009) 186–197.
[21] R. Martin, Noise power spectral density estimation based on optimal smoothing and minimum statistics, IEEE Transactions on speech and audio processing 9 (5) (2001) 504–512.
[22] R.-N. Duan, J.-Y. Zhu, B.-L. Lu, Differential entropy feature for eeg-based emotion classification, in: 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE, 2013, pp. 81–84.
[23] J.S. Hughes, J. Liu, J. Liu, Information asymmetry, diversification, and cost of capital, The accounting review 82 (3) (2007) 705–729.
[24] K. Grill-Spector, T. Kushnir, S. Edelman, G. Avidan, Y. Itzchak, R. Malach, Differential processing of objects under various viewing conditions in the human lateral occipital complex, Neuron 24 (1) (1999) 187–203.
[25] G.G. Chowdhury, Natural language processing, Annual review of information science and technology 37 (1) (2003) 51–89.
[26] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, et al., The kaldi speech recognition toolkit, in: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, No. CONF, IEEE Signal Processing Society, 2011.
[27] A. Vedaldi, K. Lenc, Matconvnet: convolutional neural networks for matlab, in: Proceedings of the 23rd ACM International Conference on Multimedia, 2015, pp. 689–692.
[28] T. N. Kipf, M. Welling, Semi-supervised Classification with Graph Convolutional Networks, arXiv preprint arXiv:1609.02907.
[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] J. Cheng, M. Chen, C. Li, Y. Liu, R. Song, A. Liu, X. Chen, Emotion recognition from multi-channel eeg via deep forest, IEEE Journal of Biomedical and Health Informatics 25 (2) (2020) 453–464.
[31] S. Sabour, N. Frosst, G. E. Hinton, Dynamic Routing between Capsules, arXiv preprint arXiv:1710.09829.
[32] T. Song, W. Zheng, P. Song, Z. Cui, Eeg emotion recognition using dynamical graph convolutional neural networks, IEEE Transactions on Affective Computing 11 (3) (2018) 532–541.
[33] H. Chao, L. Dong, Y. Liu, B. Lu, Emotion recognition from multiband eeg signals using capsnet, Sensors 19 (9) (2019) 2212.
[34] K. Lei, Q. Fu, Y. Liang, Multi-task learning with capsule networks, in: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–8.
[35] M. Crawshaw, Multi-task Learning with Deep Neural Networks: A Survey, arXiv preprint arXiv:2009.09796.
[36] 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, 2008, pp. 160–167.
[37] R. Caruana, Multitask learning, Machine learning 28 (1) (1997) 41–75.
[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.
[39] W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, X. Chen, Eeg-based emotion recognition via channel-wise attention and self attention, IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2020.3025777.
[40] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.
[41] Y. Zhang, Q. Yang, An overview of multi-task learning, National Science Review 5 (1) (2018) 30–43.
[42] 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 IAAI Conference, 2017.
[43] S. Katsigiannis, N. Ramzan, Dreamer: a database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices, IEEE journal of biomedical and health informatics 22 (1) (2017) 98–107.
[44] V. Gupta, M.D. Chopda, R.B. Pachori, Cross-subject emotion recognition using flexible analytic wavelet transform from eeg signals, IEEE Sensors Journal 19 (6) (2018) 2266–2274.
[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] R. Kohavi, et al., A study of cross-validation and bootstrap for accuracy estimation and model selection, Ijcai 14 (1995) 1137–1145. Montreal, Canada.
[47] J.R. Quinlan, Induction of decision trees, Machine learning 1 (1) (1986) 81–106.
[48] J. Suykens, L. Lukas, P. Van Dooren, B. De Moor, J. Vandewalle, et al., Least squares support vector machine classifiers: a large scale algorithm, in: European Conference on Circuit Theory and Design, ECCTD, vol. 99, Citeseer, 1999, pp. 839–842.
[49] 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.
[50] S. Ji, W. Xu, M. Yang, K. Yu, 3d convolutional neural networks for human action recognition, IEEE transactions on pattern analysis and machine intelligence 35 (1) (2012) 221–231.
[51] Z. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, G. Chen, A channel-fused dense convolutional network for eeg-based emotion recognition, IEEE Transactions on Cognitive and Developmental Systems, doi: 10.1109/TCDS.2020.2976112.
[52] W. Liu, J.-L. Qiu, W.-L. Zheng, B.-L. Lu, Multimodal Emotion Recognition Using Deep Canonical Correlation Analysis, arXiv preprint arXiv:1908.05349.
[53] V.Q. Huynh, T. Van Huynh, et al., An investigation of ensemble methods to classify electroencephalogram signaling modes, in: 2020 7th NAFOSTED Conference on Information and Computer Science (NICS), IEEE, 2020, pp. 203–208.
[54] A. Subasi, T. Tuncer, S. Dogan, D. Tanko, U. Sakoglu, Eeg-based emotion recognition using tunable q wavelet transform and rotation forest ensemble classifier, Biomedical Signal Processing and Control 68 (2021) 102648.
[55] L. Piho, T. Tjahjadi, A mutual information based adaptive windowing of informative eeg for emotion recognition, IEEE Transactions on Affective Computing 11 (4) (2018) 722–735.
[56] H. Chao, H. Zhi, L. Dong, Y. Liu, Recognition of emotions using multichannel eeg data and dbn-gc-based ensemble deep learning framework, Computational intelligence and neuroscience (2018).
[57] 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) 1383.
[58] R. Sharma, R.B. Pachori, P. Sircar, Automated emotion recognition based on higher order statistics and deep learning algorithm, Biomedical Signal Processing and Control 58 (2020) 101867.
[59] S. Siddharth, T.-P. Jung, T. J. Sejnowski, Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing, IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2019.2916015.
[60] M.L. McHugh, Interrater reliability: the kappa statistic, Biochemia medica 22 (3) (2012) 276–282.
[61] A. Bhattacharyya, R.K. Tripathy, L. Garg, R.B. Pachori, A novel multivariate-multiscale approach for computing eeg spectral and temporal complexity for human emotion recognition, IEEE Sensors Journal 21 (3) (2020) 3579–3591.
[62] R. Alhalaseh, S. Alasasfeh, Machine-learning-based emotion recognition system using eeg signals, Computers 9 (4) (2020) 95.
[63] J. Liu, G. Wu, Y. Luo, S. Qiu, S. Yang, W. Li, Y. Bi, Eeg-based emotion classification using a deep neural network and sparse autoencoder, Frontiers in Systems Neuroscience 14 (2020) 43.
[64] A. Topic, M. Russo, Emotion recognition based on eeg feature maps through deep learning network, Eng. Sci. Technol. An Int. J. 24 (6) (2021) 1442–1454.