Learning Regional Attention Convolutional Neural Network for Motion Intention Recognition Based on EEG Data

Zhijie Fang[1,2], **Weiqun Wang**[1,2*], **Shixin Ren**[1,2], **Jiaxing Wang**[1,2], **Weiguo Shi**[1,2], **Xu Liang**[1,2], **Chen-Chen Fan**[1,2] and **Zengguang Hou**[1,2,3]

[1]The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China

[2]University of Chinese Academy of Sciences, Beijing, China

[3]Center for Excellence in Brain Science and Intelligence Technology, Beijing, China

{fangzhijie2018, weiqun.wang, renshixin2015, wangjiaxing2016, shiweiguo2017, liangxu2013, fanchenchen2018, zengguang.hou}@ia.ac.cn

Abstract

Recent deep learning-based Brain-Computer Interface (BCI) decoding algorithms mainly focus on spatial-temporal features, while failing to explicitly explore spectral information which is one of the most important cues for BCI. In this paper, we propose a novel regional attention convolutional neural network (RACNN) to take full advantage of spectral-spatial-temporal features for EEG motion intention recognition. Time-frequency based analysis is adopted to reveal spectral-temporal features in terms of neural oscillations of primary sensorimotor. The basic idea of RACNN is to identify the activated area of the primary sensorimotor adaptively. The RACNN aggregates a varied number of spectral-temporal features produced by a backbone convolutional neural network into a compact fixed-length representation. Inspired by the neuroscience findings that functional asymmetry of the cerebral hemisphere, we propose a region biased loss to encourage high attention weights for the most critical regions. Extensive evaluations on two benchmark datasets and real-world BCI dataset show that our approach significantly outperforms previous methods.

1 Introduction

The Brain-Computer Interface (BCI) can be defined as a technology that translates brain signals into commands for interactive applications [Cecotti and Graser, 2010]. Noninvasive Electroencephalography (EEG) is regarded as one of the most convenient ways to record brain activity. EEG-based motion intention recognition algorithms promise to revolutionize many application areas [Shan *et al.*, 2018], notably to enable severely motor-impaired users to control assistive technologies [Schiatti *et al.*, 2018] (e.g., mind-controlled exoskeletons, rehabilitation robotics, and entertainments). Due to low signal-to-noise ratio, limited public EEG datasets, it's very challenging to tackle this task.

*corresponding author.

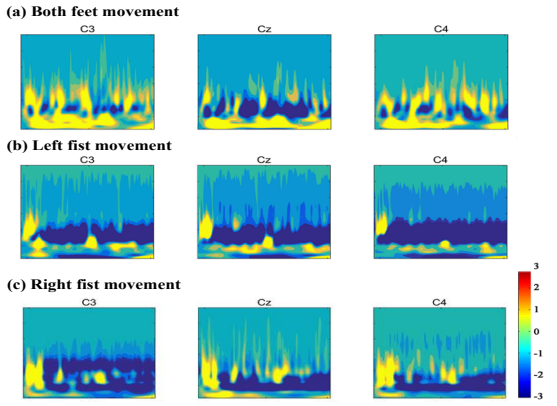

Figure 1: Time-frequency representations generated by our complex Morlet wavelet convolution and baseline subtraction. It can stably reveal many task-relevant dynamics in EEG data of different brain regions.

In the past few years, many EEG-based motion intention recognition algorithms have been proposed, which mainly focus on two parts: handcrafted feature-based models and deep learning-based models. Features extracted by common spatial pattern (CSP) algorithm [Wu *et al.*, 2014] or discrete wavelet transform (DWT) [Chen *et al.*, 2016] are fed to classical machine learning models such as support vector machine (SVM) [Lal *et al.*, 2005], linear discriminant analysis (LDA) [Kaneshiro *et al.*, 2015]. Deep learning-based models use deep neural networks to learn discriminative and robust features in an end-to-end manner [Sakhavi *et al.*, 2018]. For example, [Zhang *et al.*, 2018] try to learn spatiotemporal features by using cascade and parallel convolutional recurrent neural networks. [Chen *et al.*, 2018] extract EEG signals with different frequencies and introduce a novel multi-task deep learning model to learn human intentions. However, most of them fail to take full advantage of spectral-spatial-temporal features, which are all essential cues for EEG-based motion intention recognition [Bashivan *et al.*, 2016].

In this paper, we effectively and explicitly use the spectral-spatial-temporal features and propose a novel regional at-

tention convolutional neural network. Firstly, we use time-frequency-based approaches to analyze EEG data as a multidimensional signal that contains frequency as a prominent dimension, which provides many opportunities to link EEG data to experimental manipulations and ongoing subject behaviour. By doing so, we can explore spectral features of different brain regions while decreasing temporal precision to a certain extent. According to neuroscience, when a person performs motor execution (ME) or motor imagery (MI), the amplitude or energy of mu and beta rhythms in specific area of the primary sensorimotor will decrease, resulting in event-related desynchronization (ERD) [Dornhege *et al.*, 2003]. For example, when a person executes both feet, right-hand or left-hand motor movement, the energy of mu and beta rhythms in Cz, C3 or C4 decrease, separately. As shown in Figure 1, many task-relevant dynamics in EEG data are revealed by time-frequency-based approaches. Secondly, we propose a novel regional attention convolutional neural network (RACNN) to capture the importance of brain regions for EEG-based motion intention recognition. The RACNN consists of feature extraction module, self-attention module, and regional attention module. Given EEG data of different brain region, RACNN learns attention weights for each area in an end-to-end manner, and aggregates their CNN-based features. Finally, since functional asymmetry [Grosse-wentrup, 2009] of the cerebral hemisphere, we propose an attention loss to encourage a high attention weight for the most critical brain region. The attention loss constraint on the RACNN that the attention of the activated brain region should be larger than others.

Our main contributions can be summarized as below:

1. We propose a novel regional attention convolutional neural network (RACNN) to explore spectral-spatial-temporal features for EEG motion intention recognition. RACNN aggregates a varied number of spatial features of different brain regions produced by a backbone convolutional neural network into a compact fixed-length representation. The learned features are sensitive to task-relevant dynamics by distinguishing spectral-temporal features of different brain regions.
2. We adopt time-frequency based analysis to reveal spectral-temporal features in terms of neural oscillations of primary sensorimotor. It can stably reveal many task-relevant dynamics in EEG data of different brain regions.
3. Experiments on three datasets: PhysioNet EEG dataset, Upper Limb Movement EEG dataset and real-world BCI dataset demonstrate the effectiveness of our proposed RACNN and time-frequency based analysis. Based on them, our approach outperforms previous methods.

2 Related Works

2.1 Handcrafted Feature-Based Models

Previous EEG analysis systems learn hand-crafted features. [Chen *et al.*, 2016] use discrete wavelet transform (DWT) to extract features for epileptic focus localization. An optimal time-frequency resolution can be achieved in all frequency ranges by adopting DWT because of its varying window size. [Park and Chung, 2019] propose a novel motion intention recognition algorithm using filter-bank common spatial pattern (FBCSP) features. Those methods prove the effectiveness of spectral-spatial-temporal features for analyzing EEG signals. However, those features are manually designed, which is time-consuming and highly relies on the human experience. Different from them, our model exploits spectral-spatial-temporal information with deep learning model.

2.2 Deep Learning-Based Models

Deep Learning methods have been rising in popularity in the past few years, and are used as the most effective machine learning technology in dealing with EEG signals. Some deep learning-based models learn spatial-temporal representations of raw EEG data. For example, cascade and parallel convolutional recurrent network models [Zhang *et al.*, 2018] are used to extract spatial-temporal information and generate robust representations. [Chen *et al.*, 2018] extract EEG signals of different frequency band and introduce a novel Multi-task deep learning model to learn human intentions. Apart from those spatial-temporal features decoding method, the proposed approach [Sakhavi *et al.*, 2018] is used to preserve part spectral feature and spatial-temporal features of EEG data which leads to finding features that are less sensitive to variations within each dimension. Different from those deep learning-based models, which mainly focus on spatial-temporal features and fail to take advantage of spectral information, our regional attention convolutional neural network explicitly explores the spectral-spatial-temporal features.

2.3 Time-Frequency Based Analysis

Time-frequency-based approaches are used to conceptualize and analyze EEG data as a multidimensional signal that contains frequency as a prominent dimension, which provides many opportunities to link EEG data to experimental manipulations and ongoing subject behaviour. Recently, by using time-frequency-based approaches to reveal neural oscillations of primary sensorimotor, some methods [Tang *et al.*, 2016] have shown remarkable results in BCI. Different from them, our time-frequency-based approach is specifically designed for motion intention recognition, which adopts complex Morlet wavelets and baseline normalization, thus generating stable time-frequency representations.

3 Regional Attention Convolutional Neural Network

3.1 Generating Time-Frequency Representation

To generate time-resolved frequency representation of EEG data, we adopt band-pass filter (4-30 Hz), complex Morlet wavelet convolution, and baseline normalization. With complex Morlet wavelet convolution, the mapping between the two vectors is represented in a 2-D space that allows extracting not only the bandpass filtered signal but also time-frequency power and phase information [Cohen, 2014].

A complex Morlet wavelet ω can be defined as the product of a complex sine wave and a Gaussian window.

$$w = e^{2i\pi ft} e^{\frac{-t^2}{2\sigma^2}} \tag{1}$$

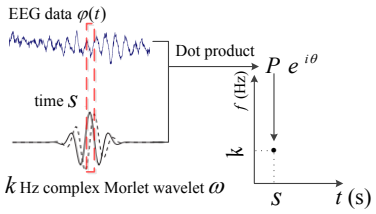

Figure 2: Graphical overview of how to extract power.

where i is the imaginary operator, f is frequency, and t is time. To avoid introducing a phase shift, t should be centred at $t = 0$, σ is the width of the Gaussian, which is defined as $\sigma = \frac{n}{2\pi f}$. When computing the dot product between a k Hz complex wavelet ω and a single-trial EEG data $\varphi(t)$ at time s, the result of the dot product contains a real part and an imaginary part.

$$P_{tf} = \omega \cdot \varphi(t) = p\cos\theta + ip\sin\theta = pe^{i\theta} \tag{2}$$

By extracting the magnitude of the result of convolution, we construct a time series of power one frequency band. Taking frequency into account, we can obtain time-resolved frequency representations of EEG data. It's still difficult to visualize activity from a large range of frequency bands simultaneously based on raw power values. Baseline normalization is used to convert time-resolved frequency representations to a scale that is suitable for quantitative statistical analysis. Decibel normalization and baseline subtraction are applied separately.

$$P_{tf} = 10\log_{10}\frac{a_{tf}}{b_{tf}} \tag{3}$$

$$P_{tf} = a_{tf} - b_{tf} \tag{4}$$

where a_{tf} is frequency-specific power, and b_{tf} is the baseline level of power in that same frequency. The generated time-resolved frequency representations are shown in Figure 1.

3.2 Regional Attention Convolutional Neural Network

After using complex Morlet wavelets and baseline normalization to extract the time-varying power from EEG data of each channel, the time-resolved frequency representations R are created as follows:

$$R = [r_1, r_2 \dots r_n] \tag{5}$$

where r_n is the time-frequency representation of channel n, and n is the number of channels. Regional attention convolutional neural network (RACNN) is used to recognize motion intentions from each time-resolved frequency representations R. RACNN consists of feature extraction module, self-attention module, and regional attention module as illustrated in Figure 3. The input to the model is the preprocessed representations of 3D data (e.g., R), containing both spectral-spatial-temporal information. We first extract the spectral-temporal features of each data representation and coarsely calculate the importance of each channel by a global average pooling layer conducted on its feature, which is called self-attention module. The second stage seeks to find more accurate attention weights by modeling the relationship between varied representations of different brain regions aggregated from the first stage, which is called regional attention module.

To extract the spectral-temporal features of each data, we apply a backbone CNN, as shown in Figure 3. The input time-resolved frequency representations are defined as $R = [r_1, r_2 \dots r_n]$. The feature representation r_n extracted by CNN is defined as below:

$$f_n = C(r_n) \tag{6}$$

Through the CNN spectral-temporal feature extraction step, the input time-resolved frequency representations are transformed into sequences of spectral-temporal feature representations.

$$F_c = [f_1, f_2 \dots f_n] \tag{7}$$

Self-attention module. To coarsely calculate the importance of each channel, we apply a global average pooling layer with a tanh activation. Mathematically, the attention weight of the n-th channel is defined as below:

$$\alpha_n = f(g(f_n)) \tag{8}$$

where g denotes the global average pooling function, and f denotes the tanh activation function. In this stage, we aggregate the spectral-temporal features of the same brain region with their attention weights into a global representation.

$$F_r^j = \frac{1}{\sum_{i=1}^{k}\alpha_i}\sum_{i=1}^{k}\alpha_i f_i, j = 1, 2 \dots m \tag{9}$$

where F_r^j is the spatial feature of the j-th brain region, which contains k channels.

Regional attention module. Since the aggregated representation $F_r^j, \mathrm{j} = 1, 2 \dots m$ inherently represents the contents of all brain regions, the attention weights can be further refined by modeling the relationship between varied representations of different brain regions. With these region features, the region attention module applies an FC layer and a softmax function to estimate the attention weights β_j of j-th brain region.

$$\beta_j = f\left(\left[F_r^j\right]^{\mathrm{T}} q^1\right) \tag{10}$$

where f denotes the softmax function, and q^1 is the parameter of fully connected layer. The attention information along with the spectral-temporal features can be aggregated into a new compact feature.

$$F = \frac{1}{\sum_{i=1}^{m}\lambda_i\beta_i}\sum_{i=1}^{m}\lambda_i\beta_i F_r^i, i = 1, 2 \dots m \tag{11}$$

where λ_i is the minimum self-attention of i-th brain region, $\lambda_i = \min(\alpha_k \dots \alpha_d)$. The spectral-spatial-temporal representation F is fed into a standard softmax classifier. Then the classification loss can be formulated as:

$$\mathcal{L}_{cls} = -\sum_{c}\hat{Y}_c \log(P_c) \tag{12}$$

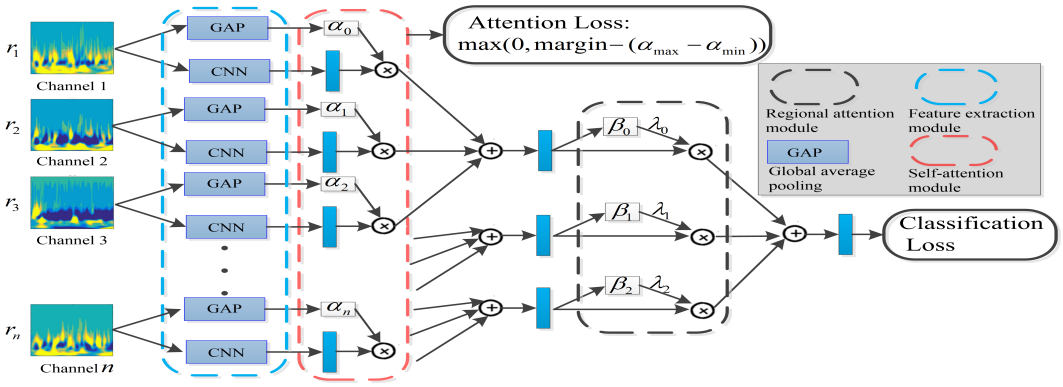

Figure 3: Architecture of our proposed regional attention convolutional neural network model. More details can be found in text.

Regional attention loss. Inspired by the neuroscience findings that functional asymmetry of the cerebral hemisphere, we propose a region biased loss to encourage high attention weights for the most important regions. The region attention loss can be formulated as:

$$\mathcal{L}_{ra} = \max\{0, \delta - (\alpha_{\max} - \alpha_{\min})\} \tag{13}$$

where δ is hyper-parameter served as margin, $\alpha_{\max}$ and $\alpha_{\min}$ is the maximum and minimum value of self-attention.

Finally, the overall loss function of our regional attention convolutional neural network can be formulated as:

$$\mathcal{L}_{all} = \mathcal{L}_{cls} + \mathcal{L}_{ra} \tag{14}$$

4 Experiment

4.1 Datasets

PhysioNet EEG dataset [Goldberger *et al.*, 2000] contains 109 subjects with 64 electrode channels and 160Hz [Schalk *et al.*, 2004] sampling rate. Each subject performed motor execution (ME) and motor imagery (MI) to finish four-class intentions, including eye closed, both feet, right fist and left fist open and close. Nevertheless, we found that the data of the #88, #89, #92, #100, and #104 subject were damaged in the data processing stage, so this participant's record was removed.

Upper Limb Movement EEG dataset [Ofner *et al.*, 2017] contains 15 subjects with 61 electrode channels and 512 Hz sampling rate. Each subject performed motor execution (ME) and motor imagery (MI) to finish seven-class intentions, including rest, elbow flexion, elbow extension, forearm supination, forearm pronation, hand close, and hand open.

4.2 Implementation Details

Nine electrode channels, which are shown in Figure 4, are chosen to generate time-resolved frequency representations. Through band-pass filter (4-30 Hz), complex Morlet wavelet convolution, and baseline normalization, the raw 4 trials of EEG data from the same session are used to generate spectral-spatial-temporal representation, as shown in Figure 1. Then three-quarters generated data chosen in random are used as the training set, and others are used as the validation set.

For CNN part, we adopt three convolution layers with kernel sizes of 3×3 to extract spectral-temporal features. For the classification loss, a classifier includes a 1024-dim FC layer as the middle layer followed by dropout and an FC layer with identity number logits as output layer. The keep probability of the dropout operation is 0.6.

All experiments are established in TensorFlow framework with batch size 100. When jointly training with regional attention loss and classification loss, the default loss weight ratio is 2:1. The margin in regional attention loss is default as 0.15. The Adam method with 10^{-5} learning rate is used as the optimizer.

4.3 Ablation Study

To evaluate each component of our regional attention convolutional neural network, we conduct two variants trained under different settings. Firstly, we train the model under Eq.(12) as a baseline. Secondly, by comparing the whole model under Eq.(14) with the first variants, we can verify the effectiveness of attention loss. Both the first two variants are learning from time-frequency representations generated by complex Morlet wavelets and baseline subtraction under Eq.(4). Thirdly, by comparing model learning from time-frequency representations produced by complex Morlet wavelets and decibel normalization under Eq.(3) with the second variants, we can verify the effectiveness of complex Morlet wavelets and baseline subtraction. We mark them as $\mathcal{L}_{cls}(sub)$, $\mathcal{L}_{all}(sub)$, and $\mathcal{L}_{all}(db)$, respectively.

Methods	PhysioNet dataset		Upper Limb dataset	
	ME	MI	ME	MI
$\mathcal{L}_{cls}(sub)$	74.6	68.3	39.2	30.2
$\mathcal{L}_{all}(db)$	70.1	65.8	37.9	27.4
$\mathcal{L}_{all}(sub)$	76.9	70.2	42.6	33.1

Table 1: The classification accuracies of different model variants on PhysioNet EEG dataset and Upper Limb Movement EEG dataset.

Comparison with different variants of our regional attention convolutional neural network on PhysioNet EEG dataset and Upper Limb Movement EEG dataset as shown in Table 1, $\mathcal{L}_{cls}(sub)$ achieves 74.6% and 39.2% classification

accuracy of ME and 68.3% and 30.2% classification accuracy of MI on PhysioNet EEG dataset and Upper Limb Movement EEG dataset, respectively. This demonstrates that the regional attention convolutional neural network is capable of learning spectral-spatial-temporal features generated by time-frequency-based approaches. Secondly, $\mathcal{L}_{all}(sub)$ achieves 76.9% and 42.6% classification accuracy of ME and 70.2% and 33.1% classification accuracy of MI on PhysioNet EEG dataset and Upper Limb Movement EEG dataset, respectively. Compared with $\mathcal{L}_{cls}(sub)$, the attention loss improves the result by 2.3% and 3.4% classification accuracy of ME and 1.9% and 2.9% classification accuracy of MI on PhysioNet EEG dataset and Upper Limb Movement EEG dataset, respectively. This shows that attention loss contributes to learning spectral-spatial-temporal features. Finally, $\mathcal{L}_{all}(db)$ achieves 70.1% and 37.9% classification accuracy of ME and 65.8% and 27.4% classification accuracy of MI on PhysioNet EEG dataset and Upper Limb Movement EEG dataset, respectively. Compared with $\mathcal{L}_{all}(db)$, complex Morlet wavelets and baseline subtraction improves the result by 6.8% and 4.7% classification accuracy of ME and 4.4% and 5.7% classification accuracy of MI on PhysioNet EEG dataset and Upper Limb Movement EEG dataset, respectively. This shows that complex Morlet wavelets and baseline subtraction contribute to learning spectral-spatial-temporal features. In summary, our proposed approach by using spectral-spatial-temporal information is effective for motion intention recognition.

4.4 Evaluation of Time-Resolved Frequency Representation

Figure 1 displays the time-frequency representations generated by our complex Morlet wavelet convolution and baseline subtraction for ME of three movements from subject 3.

For each movement, time-frequency representations of channel C4 over the right sensorimotor cortex, C3 over the left sensorimotor cortex and Cz are demonstrated. In each time-frequency representation, the cue occurs at 0 s. The blue color means ERD (power decrease), and the yellow color means ERS (power increase). For both feet movement, the ERD in both theta (4-7 Hz) and alpha (8-13 Hz) bands is shown at Cz electrode and the ERS in both alpha and beta bands is shown at the hand area (C3 and C4). For the right fist and left fist movement, ERD is shown from around 0–4 s after the cue onset due to the response delay; ERD in both theta and alpha bands is shown over motor areas. What's more, ERD is most apparent over motor areas contralateral to the hand moves. In summary, our complex Morlet wavelet convolution can stably reveal many task-relevant dynamics in EEG data of different brain regions.

4.5 Visualizing Attention Weights

We present the visualized analysis of recognizing both feet, right-hand or left-hand motor execution on PhysioNet EEG dataset. Fig.4(a) shows the positions of different electrodes for PhysioNet EEG dataset; Fig.4(b) shows the attention weights of different electrodes for both feet, left-fist or right-fist motor execution (ME).

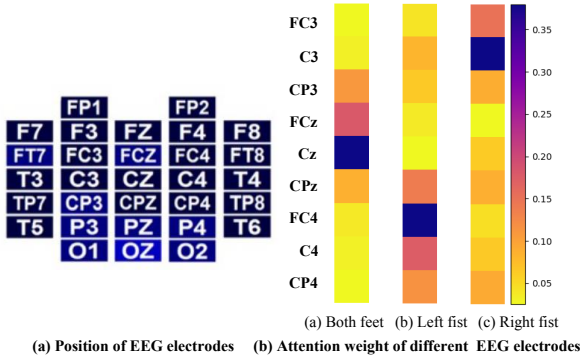

Figure 4: Visualization of attention weights of different EEG electrodes on PhysioNet EEG dataset.

Through self-attention module and regional attention module, our proposed regional attention convolutional neural network puts a high emphasis on Cz, FC4, and C3 when recognizing both feet, right-hand or left-hand motor execution, respectively. Compared with simple aggregation mechanism, which treats all time-frequency representations equally, the proposed attention mechanism automatically learns the priority of different brain region, which works better in feature fusion, and achieves better performance.

4.6 Comparison with Previous Methods

We compare our regional attention convolutional neural network with state-of-the-art, which can be grouped into two groups, $i.e.$ handcrafted feature-based models, and deep learning-based models. The compared methods are used to perform single-trial classification.

The experimental results on PhysioNet EEG dataset and Upper Limb Movement EEG dataset are shown in Table 2 and Table 3. Firstly, the handcrafted feature-based models contribute to motion intention recognition but achieve a poor classification performance. This is mainly because handcrafted feature-based models can't take full advantage of big data. Secondly, deep learning-based models significantly outperform the handcrafted feature-based models, which shows the effectiveness of deep learning. Finally, our regional attention convolutional neural network outperforms all those methods achieving new state-of-the-art. The experimental results verify the superiority of regional attention convolutional neural network over existing methods.

4.7 Case Study and Demonstration

We develop a BCI system for motion intention recognition and evaluate the proposed models on our real-world BCI dataset. Neuroscan's Grael EEG amplifier and 32-channel EEG cap are used with sampling rate of 256 Hz, which can be seen from Figure 5. Baseline correction, 50 Hz notch filter, and band-pass filter (4-30 Hz) are applied to eliminate baseline drift and noise. Each subject performed motor execution (ME) and motor imagery (MI) to finish four-class intentions, including eye closed, both feet, right fist and left fist open and close. In each recording session, the participants perform each task for 3 seconds, followed by rest for 3 seconds. Every

	Methods	Multi-class	ME			MI		
			Accuracy	Precision	Recall	Accuracy	Precision	Recall
1	SVM [Chen *et al.*, 2016]	Multi(4)	40.8	40.0	40.6	32.1	41.6	33.3
	LDA [Kaneshiro *et al.*, 2015]	Multi(4)	43.8	48.7	43.2	35.7	30.2	35.9
	FBCSP[Wu *et al.*, 2014]	Binary	-	-	-	56.3	55.9	56.1
2	Cascade [Zhang *et al.*, 2018]	Binary	63.4	63.6	63.7	57.3	57.4	57.9
	Parallel [Zhang *et al.*, 2018]	Binary	62.2	62.4	62.8	56.8	56.1	56.3
	Bashivan [Bashivan *et al.*, 2016]	Multi(4)	-	-	-	68.6	-	-
	EEGNet [Lawhern *et al.*, 2018]	Binary	73.8	73.2	73.5	-	-	-
	VG-HAM [Zhang *et al.*, 2019]	Binary	76.3	76.8	**76.5**	-	-	-
	$Ours(\mathcal{L}_{all}(sub))$	Multi(4)	**76.9**	**77.2**	75.9	**70.2**	**77.3**	**69.1**

Table 2: Comparison with previous models on PhysioNet EEG dataset. 1: handcrafted feature-based models. 2: deep learning-based models.

	Methods	Multi-class	ME			MI		
			Accuracy	Precision	Recall	Accuracy	Precision	Recall
1	SVM [Chen *et al.*, 2016]	Multi(7)	37.5	36.1	37.8	23.2	24.8	22.7
	LDA [Kaneshiro *et al.*, 2015]	Multi(7)	38.1	37.1	39.8	27.2	26.5	27.7
2	Zhao [Zhao *et al.*, 2019]	Multi(7)	-	-	-	31.0	-	-
	Parallel [Zhang *et al.*, 2018]	Multi(4)	-	-	-	30.2	30.6	30.5
	$Ours(\mathcal{L}_{all}(sub))$	Multi(7)	**42.6**	**40.6**	**42.8**	**33.1**	**34.4**	**33.7**

Table 3: Comparison with previous models on Upper Limb Movement EEG dataset. 1: handcrafted feature-based models. 2: deep learning-based models.

volunteer performs 160 trials, and there are totally 4 male volunteers. Table 4 shows the classification results on our BCI dataset.

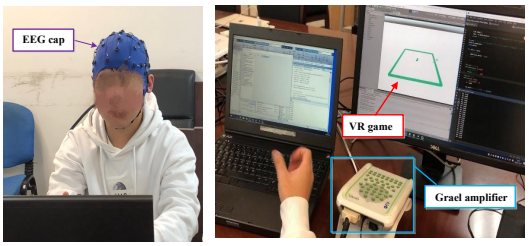

Figure 5: EEG signal recording process and EEG based BCI to control the movement of a virtual ball in 2D space.

Table 4 shows the evaluating results on our BCI dataset. Our method achieves 50.6% classification accuracy of ME and 35.1% classification accuracy of MI on our BCI dataset. Compared with the best baseline model, our methods improve the result by 8.5% classification accuracy of ME and 2.3% classification accuracy of MI, respectively. In summary, the proposed regional attention convolutional neural network is capable of learning spectral-spatial-temporal features and outperforms all those methods.

The proposed method was finally used to develop an EEG based BCI system, as shown in Figure 5. The movement of a virtual ball is controlled by the motion intention of the participant. For example, when the participant executes both feet, right fist and left fist movement, the virtual ball will go straight, turn right and turn left, separately.

Methods	Real-world BCI dataset	
	ME	MI
SVM [Chen *et al.*, 2016]	33.6	28.7
LDA [Kaneshiro *et al.*, 2015]	36.5	27.2
Cascade [Zhang *et al.*, 2018]	42.1	32.8
Parallel [Zhang *et al.*, 2018]	41.9	31.6
$Ours(\mathcal{L}_{all}(sub))$	**50.6**	**35.1**

Table 4: The classification accuracies of different models on our BCI dataset.

5 Conclusion

In this paper, we propose a novel approach to exploit spectral-spatial-temporal features for EEG motion intention recognition. Firstly, we propose a novel regional attention convolutional neural network, which learns spectral-spatial-temporal features by feature extraction module, self-attention module, and regional attention module. Secondly, we adopt complex Morlet wavelet convolution and baseline normalization to reveal many task-relevant dynamics in EEG data of different brain regions. Finally, experimental results on two benchmark datasets and our BCI dataset show the effectiveness of our method.

Acknowledgments

This work is supported in part by the National Key R&D Program of China (Grant 2018YFB1307804), the National Natural Science Foundation of China (Grants 91848110 and U1913601), the Strategic Priority Research Program of Chinese Academy of Science (Grant No.XDB32000000), and the Beijing Natural Science Foundation under Grant 4202074.

References

[Bashivan *et al.*, 2016] Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella. Learning representations from eeg with deep recurrent-convolutional neural networks. In *International Conference on Learning Representations, (ICLR)*, 2016.

[Cecotti and Graser, 2010] Hubert Cecotti and Axel Graser. Convolutional neural networks for p300 detection with application to brain-computer interfaces. *IEEE transactions on pattern analysis and machine intelligence*, 33(3):433–445, 2010.

[Chen *et al.*, 2016] Duo Chen, Suiren Wan, and Forrest Sheng Bao. Epileptic focus localization using discrete wavelet transform based on interictal intracranial eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(5):413–425, 2016.

[Chen *et al.*, 2018] Weitong Chen, Sen Wang, Xiang Zhang, Lina Yao, Lin Yue, Buyue Qian, and Xue Li. Eeg-based motion intention recognition via multi-task rnns. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 279–287. SIAM, 2018.

[Cohen, 2014] Mike X Cohen. *Analyzing neural time series data: theory and practice*. MIT press, 2014.

[Dornhege *et al.*, 2003] Guido Dornhege, Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Müller. Combining features for bci. In *Advances in Neural Information Processing Systems*, pages 1139–1146, 2003.

[Goldberger *et al.*, 2000] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[Grosse-wentrup, 2009] Moritz Grosse-wentrup. Understanding brain connectivity patterns during motor imagery for brain-computer interfacing. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 561–568. Curran Associates, Inc., 2009.

[Kaneshiro *et al.*, 2015] Blair Kaneshiro, Marcos Perreau Guimaraes, Hyung-Suk Kim, Anthony M Norcia, and Patrick Suppes. A representational similarity analysis of the dynamics of object processing using single-trial eeg classification. *Plos one*, 10(8):e0135697, 2015.

[Lal *et al.*, 2005] Thomas Navin Lal, Michael Schröder, N Jeremy Hill, Hubert Preissl, Thilo Hinterberger, Jürgen Mellinger, Martin Bogdan, Wolfgang Rosenstiel, Thomas Hofmann, Niels Birbaumer, et al. A brain computer interface with online feedback based on magnetoencephalography. In *Proceedings of the 22nd international conference on Machine learning*, pages 465–472. ACM, 2005.

[Lawhern *et al.*, 2018] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

[Ofner *et al.*, 2017] Patrick Ofner, Andreas Schwarz, Joana Pereira, and Gernot R Müller-Putz. Upper limb movements can be decoded from the time-domain of low-frequency eeg. *PloS one*, 12(8):e0182578, 2017.

[Park and Chung, 2019] Yongkoo Park and Wonzoo Chung. Selective feature generation method based on time domain parameters and correlation coefficients for filter-bank-csp bci systems. *Sensors*, 19(17):3769, 2019.

[Sakhavi *et al.*, 2018] Siavash Sakhavi, Cuntai Guan, and Shuicheng Yan. Learning temporal information for brain-computer interface using convolutional neural networks. *IEEE transactions on neural networks and learning systems*, 29(11):5619–5629, 2018.

[Schalk *et al.*, 2004] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043, 2004.

[Schiatti *et al.*, 2018] Lucia Schiatti, Jacopo Tessadori, Nikhil Deshpande, Giacinto Barresi, Louis C King, and Leonardo S Mattos. Human in the loop of robot learning: Eeg-based reward signal for target identification and reaching task. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4473–4480. IEEE, 2018.

[Shan *et al.*, 2018] Hongchang Shan, Yu Liu, and Todor P Stefanov. A simple convolutional neural network for accurate p300 detection and character spelling in brain computer interface. In *IJCAI*, pages 1604–1610, 2018.

[Tang *et al.*, 2016] Zhichuan Tang, Shouqian Sun, Sanyuan Zhang, Yumiao Chen, Chao Li, and Shi Chen. A brain-machine interface based on erd/ers for an upper-limb exoskeleton control. *Sensors*, 16(12):2050, 2016.

[Wu *et al.*, 2014] Wei Wu, Zhe Chen, Xiaorong Gao, Yuanqing Li, Emery N Brown, and Shangkai Gao. Probabilistic common spatial patterns for multichannel eeg analysis. *IEEE transactions on pattern analysis and machine intelligence*, 37(3):639–653, 2014.

[Zhang *et al.*, 2018] Dalin Zhang, Lina Yao, Xiang Zhang, Sen Wang, Weitong Chen, Robert Boots, and Boualem Benatallah. Cascade and parallel convolutional recurrent neural networks on eeg-based intention recognition for brain computer interface. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[Zhang *et al.*, 2019] Dalin Zhang, Lina Yao, Kaixuan Chen, Sen Wang, Pari Delir Haghighi, and Caley Sullivan. A graph-based hierarchical attention model for movement intention detection from eeg signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(11):2247–2253, 2019.

[Zhao *et al.*, 2019] Dongye Zhao, Fengzhen Tang, Bailu Si, and Xisheng Feng. Learning joint space–time–frequency features for eeg decoding on small labeled data. *Neural Networks*, 114:67–77, 2019.