Improving EEG-Based Motor Imagery Classification via Spatial and Temporal Recurrent Neural Networks

Xuelin Ma[1,2] , Shuang Qiu[1] , Changde Du[1,2], Jiezhen Xing[1,2] and Huiguang He[1,2,3,*]

***Abstract*— Motor imagery (MI) based Brain-Computer Interface (BCI) is an important active BCI paradigm for recognizing movement intention of severely disabled persons. There are extensive studies about MI-based intention recognition, most of which heavily rely on staged handcrafted EEG feature extraction and classifier design. For end-to-end deep learning methods, researchers encode spatial information with convolution neural networks (CNNs) from raw EEG data. Compared with CNNs, recurrent neural networks (RNNs) allow for long-range lateral interactions between features. In this paper, we proposed a pure RNNs-based parallel method for encoding spatial and temporal sequential raw data with bidirectional Long Short-Term Memory (bi-LSTM) and standard LSTM, respectively. Firstly, we rearranged the index of EEG electrodes considering their spatial location relationship. Secondly, we applied sliding window method over raw EEG data to obtain more samples and split them into training and testing sets according to their original trial index. Thirdly, we utilized the samples and their transposed matrix as input to the proposed pure RNNs-based parallel method, which encodes spatial and temporal information simultaneously. Finally, the proposed method was evaluated in the public MI-based eegmmidb dataset and compared with the other three methods (CSP+LDA, FBCSP+LDA, and CNN-RNN method). The experiment results demonstrated the superior performance of our proposed pure RNNs-based parallel method. In the multi-class trial-wise movement intention classification scenario, our approach obtained an average accuracy of 68.20% and significantly outperformed other three methods with an 8.25% improvement of relative accuracy on average, which proves the feasibility of our approach for the real-world BCI system.**

I. Introduction

Brain-Computer Interface (BCI) can provide an alternative communication channel and environmental control capability for severely disabled persons [1]. The Motor imagery (MI) is the only BCI paradigm that does not require external stimuli, which is actively controlled by the user and reflects the user's voluntary movement consciousness. EEG-based MI BCI system is popular because it is non-invasive, inexpensive and easily applied to human beings [2].

Recognizing the movement intention of a human from scalp EEG data is an open challenge, which has garnered significant attention since the early days of BCI. Researchers have generally approached intention recognition with performing handcrafted feature extraction and classifier learning separately [3]. However, the feature extraction leads to information loss and the two-stage approaches are difficult to optimize simultaneously.

With the development of deep learning, the intention recognition accuracy has been enhanced by replacing traditional classifier with deep neural network [4][5][6]. Furthermore, end-to-end deep learning approaches are used to learn feature representation and perform classification directly from raw EEG data. For example, Supratak et al. used CNNs to extract time-invariant features and used LSTMs to encode temporal information for automatic sleep stage scoring [7]. Schirrmeister et al. utilized pure CNNs for learning temporal and spatial representation step-by-step [8][9]. Zhang et al. used convolutional-recurrent neural networks (CNN-RNN) to learn the spatio-temporal representation of raw MI-based EEG data [10]. The existing end-to-end methods encode spatial information with CNNs. Compared with CNNs, RNNs are typically more expensive but allow for long-range lateral interactions between features in the same feature map [11]. In addition, the training and evaluation samples of machine learning methods should be independent identically distributed (i.i.d.) [12][13]. To obtain adequate samples, deep learning based methods generally segment the raw EEG data into clips by sliding window method [8][10][14]. However, some MI-based intention recognition researches shuffled all the clips and split them into training set and testing set [10][14]. By this way, the samples of training and testing sets are identically distributed but not independent.

To satisfy the i.i.d. character of samples and fully exploit the potential of RNNs in the context of EEG-based intention recognition, we propose the following items, and evaluate them empirically in the MI-based eegmmidb dataset. First, we rearrange the index of recorded electrodes according to their spatial positions so that the data can be viewed as spatial sequential streams, as shown in Fig. 1. Second, instead of shuffling all the samples, we split the samples according to the trial index to obtain i.i.d. training set and testing set. Third, we propose a parallel RNNs model, containing bidirectional-LSTM for modeling spatial sequential streams and standard LSTMs for modeling temporal sequential streams, illustrated schematically in Fig. 2.

In summary, we make the following three contributions:

- We transform the EEG data into spatial sequence to learn more valuable information, which alleviates the problem of insufficient data.
- We are the first to introduce RNNs for modeling

*This work was supported by National Natural Science Foundation of China (91520202, 81701785), Youth Innovation Promotion Association CAS, the CAS Scientific Research Equipment Development Project (YJKYYQ20170050) and the Beijing AI Project.

[1]Research Center for Brain-inspired Intelligence and National Laboratory of Pattern RecognitionInstitute of AutomationChinese Academy of Sciences, Beijing, China

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

[3]Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing, China

*corresponding author huiguang.he@ia.ac.cn

EEG-based spatial sequential data. Experimental results demonstrate the practicability of the method.

- Our proposed approach significantly outperforms many methods by over 8.25% in intention recognition accuracy.

II. The Proposed Method

In this section, we introduce the data organization, training and testing strategies and the proposed parallel RNNs architecture.

A. Data Organization

We preprocess the raw data to obtain more samples for training deep networks. The overall preprocessing flowchart of our proposed method is shown in Fig. 1. As shown in the Fig. 1 EEG electrode map example, the recorded electrodes start from electrode FC5 and end to electrode Iz without an obvious orientation. The order of the recorded electrodes can be rearranged from frontal lobe (e.g. FP1) to occipital lobe (e.g. Iz), which considers the spatial location-relation between before and after electrodes.

Typically, after rearranging the channels, the given datasets can be denoted as $D^i = \{(X^1, y^1), ..., (X^{N_i}, y^{N_i})\}$, where N_i denotes the total number of recorded trials for subject i. The input matrix $X^j \in \mathbb{R}^{C\times T}$ of trial j, where $1 \leq j \leq N_i$, contains the signals of C recorded electrodes and T discretized time steps recorded per trial. The corresponding class label of trial j is denoted by y^j.

The sliding window approach is applied to divide the trial data into individual samples, which are used for later processing. Each sample has a fixed length S, with 50% overlapping between continuous neighbors.

A sample can be denoted as follow:

$$x_k^j = r_{mn} \in \mathbb{R}^{C\times S}, y_k^j = y^j \tag{1}$$

where $1 \leq j \leq N_i$, $1 \leq m \leq C$, $1 \leq n \leq S$ and $1 \leq k \leq [\frac{T}{(S/2)}] - 1$.

There are two separate viewpoints of a sample: A single sample can be viewed as S temporal sequential one-dimensional (1D) data vectors, denoted as $[r_{1n}, r_{2n}, ..., r_{Cn}]' \in \mathbb{R}^{C\times 1}$, each element of which contains C elements corresponding to C electrodes. These S vectors can be fed into temporal sequential model. From another point of view, there are C spatial sequential 1D data vectors, denoted as $[r_{m1}, r_{m2}, ..., r_{mS}] \in \mathbb{R}^{1\times S}$, each element of which contains S elements corresponding to S time stamps. These C vectors can be fed into spatial sequential model to improve the whole network performance.

B. Training and Testing strategy

In order to clarify the form of data input, the training and testing strategies are listed as follows:

1) Trial-Wise Training Strategy: In training, using trials $D^i = \{(X^1, y^1), ..., (X^{N_i}, y^{N_i})\}$ as input.

2) Sample-Wise Training Strategy: In training, using samples $\{(x_1^1, y_1^1), ..., (x_k^1, y_k^1), ..., (x_1^{N_i}, y_1^{N_i}), ..., (x_k^{N_i}, y_k^{N_i})\}$ as input.

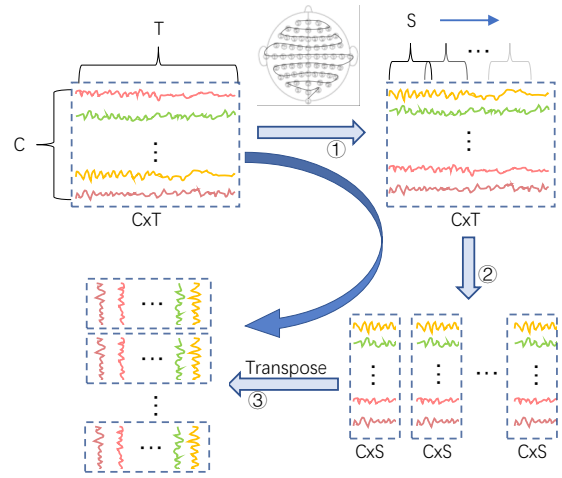

Fig. 1. The overall preprocessing flowchart. First, the electrode indexes are rearranged to make full use of the spatial location information. Second, the sliding window approach is applied to the trial data for obtaining more training samples. Finally, the temporal sequences are transposed to spatial sequences.

3) Trial-Wise Testing Strategy: While training with trial-wise strategy, we evaluate the models with trial data X^j and get $pred\_y^j$ for the trial-wise accuracy. While training with sample-wise strategy, we evaluate the models with sample data x_k^j , use the majority vote rule to fuse k predicted sample labels $pred\_y_k^j$ into predicted trial label $pred\_y^j$ and compute the trial-wise accuracy.

4) Sample-Wise Testing Strategy: While training with sample-wise strategy, we evaluate the models with sample data x_k^j and get $pred\_y_k^j$ for sample-wise accuracy.

C. Standard and Bidirectional Long Short-Term Memory

RNN is a class of neural network that maintains internal hidden states to model the dynamic temporal behaviour of sequences through directed cyclic connections between its units. LSTM extends RNN by adding three gates to an RNN neuron, which enable LSTM to learn long-term dependency in a sequence, and make it easier to optimize [15]. There are spatial and temporal sequential information containing in the EEG data. Therefore, LSTM is an excellent model for encoding sequential EEG data.

In a standard LSTM, information only flows in the forward time direction. The standard LSTM can be used for encoding temporal sequential EEG data. A bidirectional LSTM (bi-LSTM) [16] is a combination of two normal LSTMs, which allows dependencies in the reverse direction. So the temporal and spatial information of the segment data can be learned and combined through parallelling the standard LSTM and the bi-LSTM, as shown following.

D. Whole Parallel Architecture

The parallel architecture is illustrated in Fig. 2. It contains two parts, temporal LSTMs and spatial bi-LSTMs, for temporal and spatial feature extraction, respectively. We introduce a dense layer to denoise the temporal input vectors

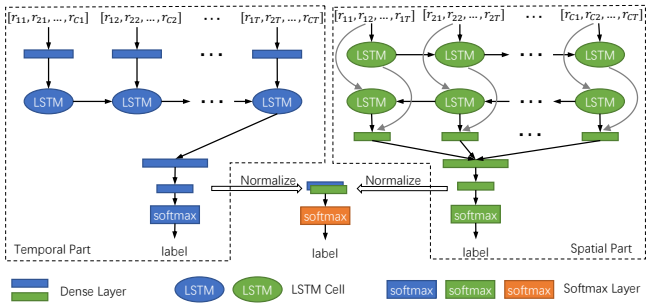

Fig. 2. The parallel architecture contains two parts, temporal LSTMs and spatial bi-LSTMs, for temporal and spatial feature extraction, respectively. The dense layers before softmax layers are normalized separately and added element-wise, along with another softmax layer to give the final predicted class label.

before feeding them to the LSTM cells. Then the temporal LSTMs model the denoised data sequentially and the last sequential output is connected to a dense layer. As for the spatial bi-LSTMs, it takes the raw spatial sequential data as input. All the sequential outputs of bi-LSTMs are concatenated and connected to a dense layer. Each dense layer after the output of LSTMs is followed by another size n (n is the number of target classes) dense layer, along with softmax output layers to generate the class labels.

The outputs of the two size n dense layers are normalized separately and added element-wise, along with another softmax layer to finally give the final predicted class labels, as shown in the middle of Fig. 2.

III. Experiments and Results

Experiments are conducted on the publicly available EEG Movement/Imagery Database (eegmmidb) [17] of multi-class scenario for movement intention recognition. And the 7-fold cross validation experimental trial-wise results show that the proposed method yields relatively higher classification accuracies compared with a set of methods. Meanwhile, the component performance of the proposed method is shown to illustrate their necessities for making up the whole architecture.

A. Dataset

The eegmmidb dataset is collected using BCI2000 instrumentation [18]. The EEG signals were recorded using International System 10-20 with 64 electrodes placed on the scalp and 160Hz sampling rate. We select the first 12 subjects to carry out our study. Each subject contains 98 trials ($C = 64, T = 651$), including 14, 21, 21, 21 and 21 trials, belonging to five classes of eye closed (baseline), imagining moving both feet, both fists, left fist and right fist, respectively. The 98 trials are split into training set and testing set by 7-fold cross-validation so there are 84 training trials and 14 testing trials in each fold. Each trial is segmented into 129 samples by sliding window of length 10, with 50% overlap. The total 10,836 samples in the training set are shuffled to train the proposed method. The 1,806 samples for testing, on the contrary, keep their order to vote for the trial-wise predicted labels and gain accuracies.

B. Implementation details

The whole neural networks were implemented with the TensorFlow framework and trained on a Nvidia 1080Ti GPU from scratch in a fully-supervised manner. The Adam algorithm is used to optimize the cross-entropy loss function with a learning rate of $0.5 * 10^{-4}$. The dropout probability is 0.5. The hidden states number of the LSTM cell is 16. All fully connected layers have the same size of 512.

C. Model Evaluation

We compared our method with two baseline methods and a latest deep learning method for MI-based intention recognition. The overall trial-wise performance of all the methods is shown in table I. All the results were obtained by seven-fold cross-validation, which consists of dividing the 94 trials of one subject into seven different training and testing sets, so that each trial is included in one of the testing sets only once. Overall, our proposed parallel method outperforms these methods. In addition, the components' performance of our proposed method also indicates their necessities for making up the whole architecture.

1) Comparison with baseline methods: Both of Common Spatial Pattern (CSP) [19] and Filter Bank Common Spatial Pattern (FBCSP) [20] compute spatial filters that enhance class-discriminative band power features contained in the EEG.

The features are followed by a classifier such as Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) to finally perform the recognition.

FBCSP was the best-performing method for many EEG decoding competitions [8]. Therefore, we consider FBCSP with LDA classifier (FBCSP+LDA) as adequate baseline algorithm for the evaluation in the present study. In addition, CSP performs better than FBCSP for some subjects' data, so we consider CSP with LDA classifier (CSP+LDA) as another baseline method. Both of the baseline methods are trained and tested in trial-wise strategy.

Overall, our proposed method significantly outperforms the CSP+LDA and FBCSP+LDA methods with accuracy improvement of 9.19% and 8.25%, respectively (both $p < 0.05$). The results demonstrate that, given adequate input samples, the end-to-end deep learning method can reach or even exceed the performance of classical methods for intention recognition. Compared with classical methods, end-to-end deep learning methods can learn from raw MI-based EEG data adequately and optimize the feature extraction and classifier design steps simultaneously to obtain more stable and better performance.

2) Comparison with the newest method: We also compare the performance of our approach with the most recent published approach [10], proposed by Zhang et.al. Although they didn't evaluate their proposed CNN-LSTM architecture with i.i.d. samples, there are still some reasons for supporting their proposed models to work, such as converting one-dimension EEG sequences to two-dimension EEG meshes according to electrode distribution, learning spatial and temporal simultaneously. So we consider their parallel CNN-LSTM model

TABLE I

OVERALL PERFORMANCE. THE ACCURACIES ARE GIVEN IN PERCENT TRIAL-WISE.

Subject	1	2	3	4	5	6	7	8	9	10	11	12	Mean
CSP+LDA	**70.41**	63.27	65.31	53.06	60.20	41.84	**94.90**	54.08	52.04	42.86	59.18	51.02	59.01
FBCSP+LDA	69.39	62.24	64.29	61.22	57.14	44.90	88.78	60.20	48.98	46.94	**63.27**	52.04	59.95
Zhang et al. 2017	53.06	61.22	66.33	44.90	59.18	63.27	79.59	46.94	53.06	57.14	46.94	58.16	57.48
Temporal-RNN	44.90	58.16	70.41	43.88	58.16	68.37	77.55	51.02	59.18	63.27	52.04	67.35	59.52
Spatial-RNN	53.06	65.31	71.43	52.04	47.96	66.33	60.21	50.00	62.25	53.06	54.08	65.31	58.42
Our Method-Parallel	65.31	**72.45**	**74.49**	**61.23**	**60.20**	**70.41**	82.65	**65.31**	**65.31**	**67.35**	61.22	**72.45**	**68.20**

as compared method. We utilize their open source code to reproduce the work and evaluate it with the same strategy as our proposed method (see section II).

The compared deep learning method acquires a average accuracy of 57.48%, and our proposed method outperforms it with a significant accuracy difference of 10.72% ($p < 0.01$). In the other way, the number of parameters in our model is much less than that of the compared method, so less computing resources are needed. Therefore, compared with the parallel CNN-RNN methos, our proposed method is better suited to further on-line BCI systems.

3) Component evaluation: The component performance of our parallel method is also shown in table I. Both spatial part and temporal part perform no significant differences with the three compared methods (all $p > 0.05$). Fusing the prediction of these two parts significantly improves the classification accuracy (both $p < 0.01$). These results indicate that the parallel model can learn useful and distinguishing feature representations through two pathways, by being fed with different organization data of same samples.

IV. CONCLUSION AND FUTURE WORK

In this paper, we address the EEG-based intention recognition problem. We introduce a novel viewpoint for EEG data to easily expand the data for further process. To make full use of the limited data, we propose a parallel LSTMs network for learning spatial and temporal information simultaneously. By 7-fold cross-validation evaluation, our proposed method outperformed a set of methods in trial-wise multi-class scenario. Since the training and testing sets are split trial-wise, our improved results demonstrate that, the proposed method gains better generalization between trials, which is critical to the feasibility of our method for real-world BCI system. In the future, we will further study the spatial and temporal information containing in the raw EEG data, improve our proposed method on the EEG data with a little channels, and apply it to online system.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[2] N. F. Ince, S. Arica, and A. Tewfik, "Classification of single trial motor imagery eeg recordings with subject adapted non-dyadic arbitrary time–frequency tilings," *Journal of neural engineering*, vol. 3, no. 3, p. 235, 2006.

[3] S. Sun and J. Zhou, "A review of adaptive feature extraction and classification methods for eeg-based brain-computer interfaces," in *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2014, pp. 1746–1753.

[4] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted boltzmann machines," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 25, no. 6, pp. 566–576, 2017.

[5] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, vol. 2014, 2014.

[6] H. Yang, S. Sakhavi, K. K. Ang, and C. Guan, "On the use of convolutional neural networks and augmented csp features for multi-class motor imagery of eeg signals classification," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, 2015, pp. 2620–2623.

[7] A. Supratak, H. Dong, C. Wu, and Y. Guo, "Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017.

[8] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[9] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: A compact convolutional network for eeg-based brain-computer interfaces," *arXiv preprint arXiv:1611.08024*, 2016.

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

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[12] V. N. Vapnik, "An overview of statistical learning theory," *IEEE transactions on neural networks*, vol. 10, no. 5, pp. 988–999, 1999.

[13] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.

[14] X. Zhang, L. Yao, C. Huang, Q. Z. Sheng, and X. Wang, "Intent recognition in smart living through deep recurrent neural networks," in *International Conference on Neural Information Processing*. Springer, 2017, pp. 748–758.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[17] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[18] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *IEEE Transactions on biomedical engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.

[19] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[20] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface," in *2008 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2008, pp. 2390–2397.