Convolutional LSTM: A Deep Learning Method for Motion Intention Recognition based on Spatiotemporal EEG Data

Zhijie Fang[1,2], Weiqun Wang[2(✉)], and Zeng-Guang Hou[1,2,3]

[1]University of Chinese Academy of Sciences, Beijing 100049, China
{fangzhijie2018,zengguang.hou}@ia.ac.cn
[2]The State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
{weiqun.wang}@ia.ac.cn
[3]CAS Center for Excellence in Brain Science and Intelligence Technology,
Beijing 100190, China

Abstract. Brain-Computer Interface (BCI) is a powerful technology that allows human beings to communicate with computers or to control devices. Owing to their convenient collection, non-invasive Electroencephalography (EEG) signals play an important role in BCI systems. Design of high-performance motion intention recognition algorithm based on EEG data under cross-subject and multi-category circumstances is a crucial challenge. Towards this purpose, a convolutional recurrent neural network is proposed. The raw EEG streaming is transformed into image sequence according to its location of the primary sensorimotor area to preserve its spatiotemporal features. A Convolutional Long Short-Term Memory (ConvLSTM) network is used to encode spatiotemporal information and generate a better representation from the obtained image sequence. The spatial features are then extracted from the output of ConvLSTM network by convolutional layer. The convolutional layer along with ConvLSTM network is capable of capturing the spatiotemporal features which enables the recognition of motion intention from the raw EEG signals. Experiments are carried out on the PhysioNet EEG motor imagery dataset to test the performance of the proposed method. It is shown that the proposed method can achieve high accuracy of 95.15%, which outperforms previous methods. Meanwhile, the proposed method can be used to design high-performance BCI systems, such as mind-controlled exoskeletons, prosthetic hands and rehabilitation robotics.

Keywords: Brain-Computer Interface · Convolutional LSTM
EEG · Motion intention recognition

This work is supported in part by the National Key R&D Program of China (Grant 2018YFC2001700), the Strategic Priority Research Program of Chinese Academy of Science (Grant No.XDB32000000), and Beijing Natural Science Foundation (Grant L172050 and 3171001).

1 Introduction

Brain science is one of the most challenging frontier research fields in the twenty-first century. The Brain-Computer Interface (BCI) is a kind of technology that helps human beings to communicate with computers or to control devices. Non-invasive [1] Electroencephalography (EEG) is regarded as one of the most convenient signal sources for BCI systems in practice. When a person is doing mental preparations of motor activity without any muscular motion, appropriate motor related EEG rhythms fluctuate from their scalp [2]. Many promising EEG-based BCI systems have been developed in the literature, such as mind-controlled exoskeletons [3], prosthetic hands [4], and rehabilitation robotics [5]. Therefore, EEG-based intention recognition has become a significant topic because of its industrial and medical applications.

Although a large number of scientists are trying to recognize motion intentions by analyzing EEG signals, this technology is facing several challenges. The first challenge in EEG-based intention recognition is the collected EEG signals themselves because of the low signal-to-noise ratio, coupled with a large quantity of noise, including external noise and physiological noise. The noise definitely presents a severe difficulty for interpretation and analysis of the EEG signal. Also, a typical EEG-based BCI system suffers from the high price, tolerability of the end user, so there are limited public EEG datasets compared with audio, image and video data. More over, most EEG-based intention recognition mainly focuses on manual feature selection, which is time-consuming and highly relys on human experience. For examples, some methods use multiscale principal component analysis [6] to eliminate noise or discrete wavelet transform [7] to extract features followed by a classification model. Finally, many research projects have a terrible classification accuracy, though they just classify EEG signals under the intra-subject or binary circumstances. Few research projects involve cross-subject and multi-category classifications, which is more consistent with real-world applications.

Recently, deep learning [8] has shown strong capability when dealing with text, image, audio and video signals. Some researchers are trying to solve EEG-based intention recognition problem by using deep convolutional network or recurrent neural network. However, these methods only focus on spatial information [9] or temporal information [10]. Thus, current approaches can't deal well with EEG signals. To achieve high-performance BCI systems, an end-to-end convolutional recurrent neural network is proposed to recognize human intentions. We formulate EEG-based intention recognition as a spatiotemporal sequence classification problem. In particular, we transform the spatially distributed EEG signals into 2-D images by projecting the corresponding location of electrodes from a 3-dimensional space onto a 2-D surface [11]. The ConvLSTM network is used to encode EEG signals from spatiotemporal EEG movie. Several convolutional layers are applied to extract spatial features from the output of the ConvLSTM network. The major contributions of this paper can be outlined as follows:

• Firstly, we propose an end-to-end deep neural network model to recognize motion intentions based on raw spatiotemporal EEG data.

• It is shown that the proposed convolutional recurrent neural network is capable of encoding the spatiotemporal features from the raw EEG streaming and recognizing motion intentions under cross-subject and multi-category classification circumstances.

• The experimental results demonstrate that the proposed method outperforms previous methods and achieves high accuracy of 95.15% for EEG-based intention recognition.

The remainder of this paper is organized as follows: The detail of the proposed framework is demonstrated in Sect. 2. The data processing, model training, and the result analysis are discussed in Sect. 3. Lastly, we conclude this paper in Sect. 4.

2 Methods

The goal of the proposed convolutional recurrent neural network is to recognize motion intentions based on spatiotemporal EEG data. Fig. 1 shows an overview of the proposed method. The network is composed of a ConvLSTM layer for encoding spatiotemporal information and generating a better representation from raw EEG data and several convolutional layers for extracting spatial information.

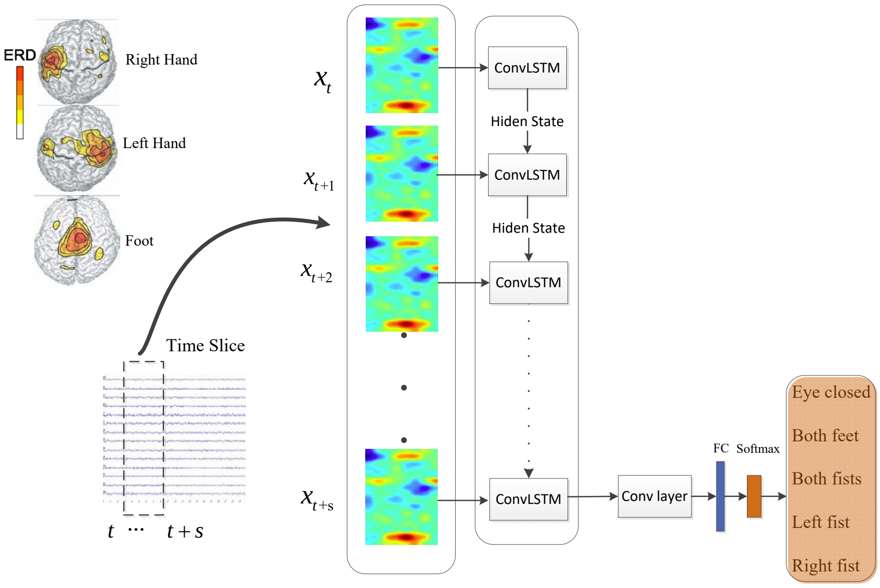

Fig. 1. The proposed convolutional recurrent neural network architecture.

2.1 Design of the Input Images from EEG Streaming

Neuroscience research found that the event-related desynchronization (ERD) starts before the motor imagery over the contralateral hemisphere then becomes bilaterally symmetrical with movement execution [12]. Specifically, when a person executes motor imagery, the specific area of the primary sensorimotor area is activated, in which the Rolandic mu and beta rhythms amplitude will decrease, resulting in event-related desynchronization [13]. The electrodes measure the EEG rhythms fluctuated from different areas of the brain. Hence, we transform the spatially distributed EEG signals into 2-D images by projecting the corresponding location of electrodes from a 3-dimensional space onto a 2-D surface [11]. Taking time into account, we can obtain a sequence of spatial information-preserving images. The detail will be discussed in Sect. 3.

2.2 Convolutional LSTM

By using the sliding window approach, the obtained image sequence can be divided into individual movie clips. The goal of the end-to-end deep neural network model is to classify motion intentions based on spatiotemporal features from EEG movie clips. For a model to recognize motion intentions based on EEG movie clips, it should be capable of identifying how the activated area of the primary sensorimotor is changing with time. Convolutional neural networks (CNN) is able to generate a spatial representation. Recurrent neural networks can encode temporal changes. Since the model should be able to deal with spatiotemporal information, ConvLSTM is a suitable option.

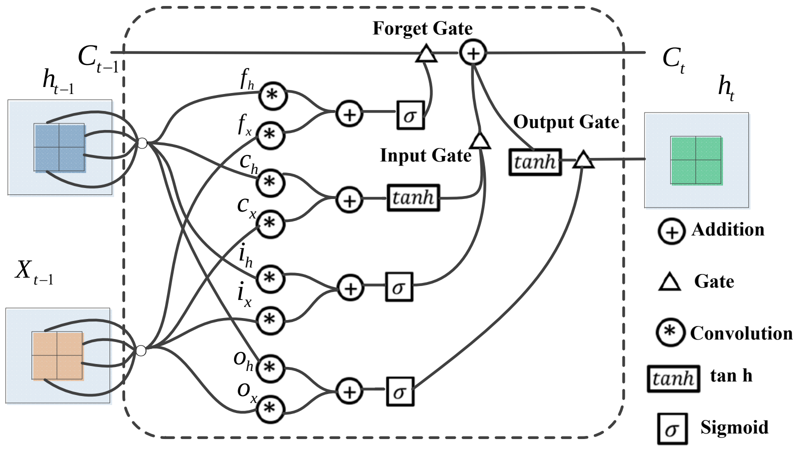

Fig. 2. The inner structure of a ConvLSTM cell [14].

ConvLSTM can encode spatiotemporal information and generate a better representation. The convLSTM model was first introduced to deal with precipitation nowcasting [15] due to its capacity of extracting spatiotemporal information. Fig. 2 shows the inner structure of a ConvLSTM cell. Different from Long Short-Term Memory (LSTM) network, the input feature of a ConvLSTM cell is a 3-D spatiotemporal tensor, and the state-to-state and input-to-state transitions are related to convolutional operations. The key equations of the ConvLSTM are shown as follows:

$$f_t = \sigma\left(U_f * X_t + W_f * h_{t-1} + b_f\right) \tag{1}$$

$$i_t = \sigma\left(U_i * X_t + W_i * h_{t-1} + b_i\right) \tag{2}$$

$$o_t = \sigma\left(U_o * X_t + W_o * h_{t-1} + b_o\right) \tag{3}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh\left(U_c * X_t + W_c * h_{t-1} + b_c\right) \tag{4}$$

$$h_t = o_t \circ \tanh\left(C_t\right) \tag{5}$$

In the equations, i_t, o_t, f_t are the outputs of input gate, output gate and forget gate at time step t. h_t stands for the hidden state of a cell at time step t. C_t stands for the cell output at time step t. The symbol "$*$" stands for the convolution operator, and "o" stands for the Hadamard product.

2.3 Network Architecture

After the spatial information-preserving image sequence is obtained, the end-to-end model is used to classify motion intentions based on the obtained image sequence. Fig. 1 shows an overview of the proposed method. By using the sliding window approach, which can preserve valuable spatiotemporal information, we divide the obtained image sequence into individual movie clips. The length of each clip is fixed, and there are overlapping between nearby neighbors, avoiding losing significant information. Then the proposed model is used to recognize the motion intentions form the EEG movie clips. ConvLSTM network has the capability to encode spatiotemporal information in its memory cell based on the obtained EEG movie clips. In the ConvLSTM, 256 filters are applied in all the gates, and the filter size are 3 × 3 with stride 1. Convolutional layers receive the output of the last time step of the ConvLSTM layer, and feeds to the fully connected layer, ending up with a softmax layer for motion intention prediction. ReLU is used as the non-linear activation function for the output of each convolutional layer.

3 Experiments

3.1 Dataset

Experiments are carried out on the PhysioNet EEG motor imagery dataset [16], which contains 109 subjects. The dataset contains five motion intentions with eye closed, imagining moving both fists, both feet, right fist and left fist. And the dataset is collected by the BCI2000 instrumentation system, and this system has

64 channels and the sampling rate is 160 Hz. Each subject performed baseline runs and task runs.
Task 1: The subject will see an object either on the right or the left side of the monitor, and he should imagine closing and opening the corresponding fist until the object vanishes.
Task 2: The subject will see an object either on the bottom or the top of the monitor, and he should imagine closing and opening both feet if the object is on the bottom or imagine closing and opening both firsts if the object is on the top until the object vanishes.

3.2 Implementation Details

The collected EEG data has 64 channels, and we transform the EEG streaming into image sequence by projecting the corresponding location of electrodes from a 3-dimensional space onto a 2-D surface at each sampling moment. The obtained EEG image sequence is divided into clips with 10 sampling points and 5 sampling points overlap. Three-quarters data are chosen in random as the training set, and others are used as the validation set. The ConvLSTM layer is used to extract the spatiotemporal information, and several convolutional layers are used to extract spatial information. All experiments are established in Tensorflow framework with batch size 200. We adopt the Adam optimizer with 0.0005 learning rate.

3.3 Experiment Results

The performance of the proposed convolutional recurrent neural network is shown in this section. We compare the results with previous methods to evaluate the performance of the proposed model. Five convolutional recurrent neural network variants and the comparison models are shown in Table 1.

Table 1. Comparison between convolutional recurrent neural network and previous methods.

Method	Multi-class	Validation	Accuracy (%)
Wang [17]	Multi(3)	Intra-Sub	84.62
Sasweta Pattnaik [7]	Binary	Cross-Sub	80.71
Kbra Saka [18]	Binary	Cross-Sub	88.87
Pouya Bashivan [11]	Multi(4)	Cross-Sub	91.11
Jasmin [6]	Binary	Intra-Sub	**92.80**
ConvLSTM + 2 Conv layers	Multi(5)	Cross-Sub	89.39
ConvLSTM + 3 Conv layers	Multi(5)	Cross-Sub	94.05
ConvLSTM + 4 Conv layers	Multi(5)	Cross-Sub	**95.15**
ConvLSTM + 5 Conv layers	Multi(5)	Cross-Sub	95.10
ConvLSTM + 2 Conv + 2 pooling layers	Multi(5)	Cross-Sub	83.18

As is shown in Table 1, the proposed convolutional recurrent neural network achieves high accuracy of 95.15% and outperforms the previous methods.

ConvLSTM network along with four convolutional layers to extract spatiotemporal features can hit the best performance. Although Jasmin [6] centers on the intra-subject and binary circumstance, the proposed model still achieves higher accuracy than their method. Their model requires decomposing raw EEG signals, which may lose significant features while extracting the higher order statistic features. What's more, we add a max-pooling layer after the convolutional layer, but the validation accuracy decreases. Max-pooling layer may make convolutional recurrent neural network achieve translation invariance. Thus, the proposed model can not distinguish which area of the primary sensorimotor is activated.

The accuracy of the proposed method lies in the range between 89% and 95.15%. A ConvLSTM layer with four convolutional layers to extract spatial information can reach the best performance, with an improvement of 2.35% over the previous methods [6]. The validation accuracy of ConvLSTM layer with different convolutional layers to extract spatial features are shown in Fig. 3.

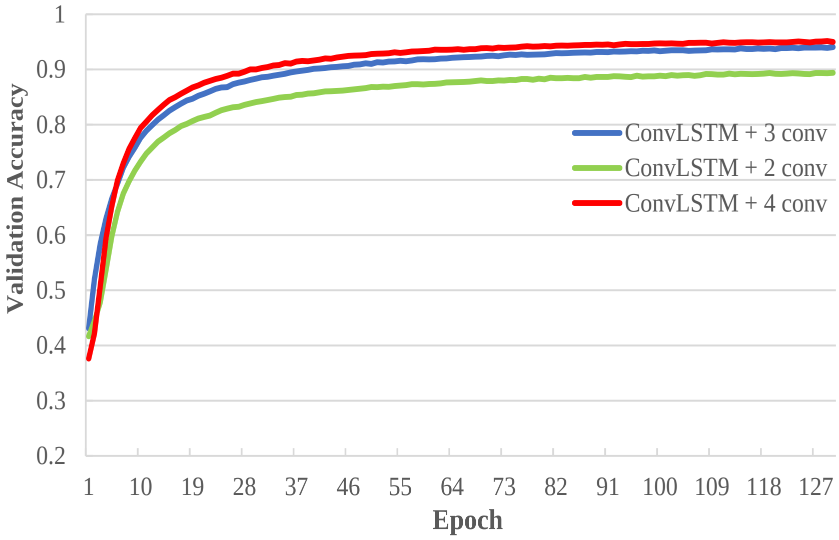

Fig. 3. The validation accuracy of three model variants based on the PhysioNet dataset. The horizontal axis stands for the number of epochs, and the left longitudinal axis stands for validation accuracy.

It can be seen from Fig. 3 that the validation accuracy of three convolutional recurrent neural network variants increases rapidly from the first epoch; the validation accuracy increases slowly when the epoch is from 15 to 70; all model variants converge after several fluctuations. Although the ConvLSTM network with four convolutional layers doesn't perform well after the first epoch, its convergence rate is faster than the other two model variants. With four convolutional layers to extract spatial features, the proposed model can achieve high accuracy of 95.15%.

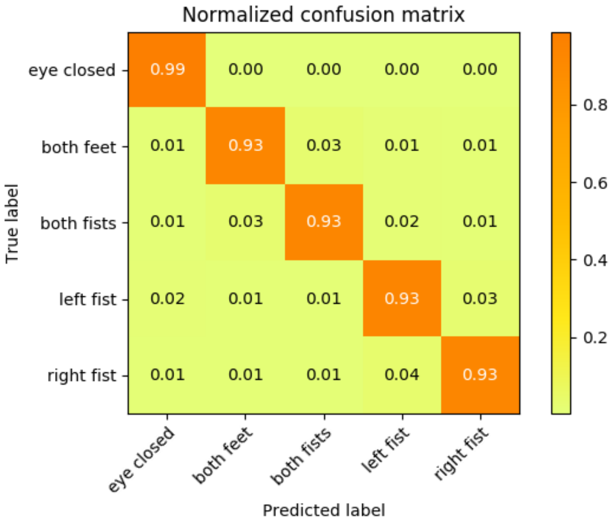

Fig. 4. Confusion matrix illustrating the per-class validation accuracy.

The result of best model variant is used to calculate the confusion matrix, which is shown in Fig. 4. When distinguishing both feet and both fists classes or left fist and right fist classes, the proposed model may make mistakes. However, the proposed model outperforms previous methods. The results show that the proposed model is capable of recognizing motion intentions under cross-subject and multi-category classification circumstances.

Table 2. The per-class performance of convolutional recurrent neural network.

Class	F1 (%)	Precision (%)	Recall (%)
eye closed	98.08	97.40	98.77
both feet	93.44	93.79	93.10
both fists	93.54	94.27	92.82
left fist	93.17	92.86	93.48
right fist	93.35	93.96	92.74
Mean±Std	94.32±3.0	94.46±2.0	94.18±4.0

What's more, F1 score, precision and recall are used to evaluate the per-class recognition performance. Precision can reflect the sensitivity of the proposed model. Recall is used to show the classifier's completeness. F1 score is the harmonic mean of precision and recall. It can be seen from Table 2 that for each class, F1 score, precision and recall are quite high with mean values of 94.32%, 94.46% and 94.18%, respectively.

4 Conclusion

The work is motivated by the goal of achieving high-performance motion intention recognition algorithm under cross-subject and multi-category circumstances. The EEG streaming is transformed into image sequence according to its location of the primary sensorimotor area to preserve its spatiotemporal features. A convolutional recurrent neural network is proposed to learn features from raw EEG data. The proposed convolutional recurrent neural network is trained on PhysioNet EEG motor imagery dataset, and the results demonstrate that the proposed model outperforms the previous methods by achieving high accuracy of 95.15%. This results show that the proposed model can be used to design high-performance BCI systems, such as mind-controlled exoskeletons, prosthetic hands and rehabilitation robotics.

References

1. Shende, P.M., Jabade, V.S.: Literature review of brain computer interface (bci) using electroencephalogram signal. In: IEEE International Conference on Pervasive Computing (ICPC). pp. 1–5. IEEE (2015)
2. Kumar, S.U., Inbarani, H.H.: Pso-based feature selection and neighborhood rough set-based classification for bci multiclass motor imagery task. Neural Computing and Applications 28(11), 3239–3258 (2017)
3. Chowdhury, A., Raza, H., Meena, Y.K., Dutta, A., Prasad, G.: Online covariate shift detection-based adaptive braincomputer interface to trigger hand exoskeleton feedback for neuro-rehabilitation. IEEE Trans. Cogn. Dev. Syst. 10(4), 1070–1080 (2018)
4. Muller-Putz, G.R., Pfurtscheller, G.: Control of an electrical prosthesis with an ssvep-based bci. IEEE Trans. Biomed. Eng. 55(1), 361–364 (2008)
5. Ang, K.K., Chua, K.S.G., Phua, K.S., Wang, C., Chin, Z.Y., Kuah, C.W.K., Low, W., Guan, C.: A randomized controlled trial of eeg-based motor imagery brain-computer interface robotic rehabilitation for stroke. Clinical EEG and neuroscience 46(4), 310–320 (2015)
6. Kevric, J., Subasi, A.: Comparison of signal decomposition methods in classification of eeg signals for motor-imagery bci system. Biomedical Signal Processing and Control 31, 398–406 (2017)
7. Pattnaik, S., Dash, M., Sabut, S.: Dwt-based feature extraction and classification for motor imaginary eeg signals. In: IEEE International Conference on Systems in Medicine and Biology (ICSMB). pp. 186–201. IEEE (2016)
8. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436 (2015)
9. Dai, M., Zheng, D., Na, R., Wang, S., Zhang, S.: Eeg classification of motor imagery using a novel deep learning framework. Sensors 19(3), 551 (2019)
10. Michielli, N., Acharya, U.R., Molinari, F.: Cascaded lstm recurrent neural network for automated sleep stage classification using single-channel eeg signals. Computers in biology and medicine 106, 71–81 (2019)

11. Bashivan, P., Rish, I., Yeasin, M., Codella, N.: Learning representations from eeg with deep recurrent-convolutional neural networks. In: International Conference on Learning Representations, (ICLR) (2016)
12. Stancák Jr, A., Pfurtscheller, G.: The effects of handedness and type of movement on the contralateral preponderance of μ-rhythm desynchronisation. Electroencephalography and clinical Neurophysiology 99(2), 174–182 (1996)
13. Pfurtscheller, G., Neuper, C.: Motor imagery activates primary sensorimotor area in humans. Neuroscience letters 239(2-3), 65–68 (1997)
14. Yuan, Z., Zhou, X., Yang, T.: Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018. pp. 984–992 (2018)
15. Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: Advances in neural information processing systems. pp. 802–810 (2015)
16. Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation 101(23), e215–e220 (2000)
17. Wang, Z., Du, X., Wu, Q., Dong, Y.: Research on the multi-classifier features of the motor imagery EEG signals in the brain computer interface. In: Tenth International Conference on Digital Image Processing (ICDIP 2018). vol. 10806, p. 108066Z. International Society for Optics and Pho-tonics (2018)
18. Saka, K., Aydemir, Ö., Öztürk, M.: Classification of eeg signals recorded during right/left hand movement imagery using fast walsh hada-mard transform based features. In: IEEE International Conference on Tele-communications and Signal Processing (TSP). pp. 413–416. IEEE (2016)