Deep Convolutional Neural Network for Decoding Motor Imagery based Brain Computer Interface

Jin Zhang, Chungang Yan, Xiaoliang Gong
College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China
Key Laboratory of Embedded System and Service Computing (Tongji University), Ministry of Education, Shanghai, China
Shanghai Electronic Transactions and Information Service Collaborative Innovation Center

Abstract—Deep convolutional neural networks (CNNs) have made tremendous development in the field of image recognition and natural language processing. However, there is still a lack of knowledge of using CNN models to decode motor imagery based Brain Computer Interface (BCI). This paper presents a method applying CNN to analyze the EEG signals, which are produced by left and right hand motor imagery tasks. EEG signals are transferred into time-frequency images using Short-Time Fourier Transform (STFT) and then those images are fed as input of the network for classification. A comparison is made to verify the performance of three different activation functions during the network`s learning procedure. They are rectified linear unit (ReLU), exponential linear unit (ELU) and the newly proposed scaled exponential linear unit (SELU). The results show that our method using CNN model can achieve better accuracy than the conventional method and SELU function shows superior ability for the network to convergence.

Keywords-deep convolutional neural network, motor imagery, brain computer interface, activation function

I. Introduction

Brain Computer Interface controls outer equipment, like cursors and wheelchairs, by recording and analyzing human`s physiological signals [1]. So it can help people communicate with the outside world without limbs and muscle. Collecting electroencephalography (EEG) during the motor imagery tasks is a very commonly used paradigm for BCI. Motor Imagery (MI), i.e. imaging movement of certain parts of the body, its neural foundation is the phenomenon of Event-Related Synchronization (ERS) and Event-Related De-synchronization (ERD) [2]. When people are doing certain MI tasks, energy increases in the contralateral area and simultaneously decreases around the ipsilateral area, so different kinds of signals will be obtained and recognized, and then transferred into commands accordingly.

Usually, there are three basic stages in the decoding work of EEG signals. First, pre-processing stage filters out some unrelated components, then feature extraction stage obtains discriminant features of different classes, at last, classification stage builds up a model to map between the features and their categories, and this model will be applied to make predictions for the new signals. Effective feature extraction and classification methods make great impact on the EEG decoding performance, and a lot of researches have illustrated that Common Spatial Pattern (CSP) and Support Vector Machine (SVM) handle the motor imagery based EEG classification problems very well [3,4].

Although a lot of studies have illustrated the ability of deep learning methods to optimize problems encountered in specific fields, there is still a lack of knowledge of using deep learning, like CNN to decode motor imagery based EEG signals. As the best of our knowledge, when applying CNN method to recognize EEGs, there are totally two kinds of input of the network. The first one is using raw EEG directly [5,6,7]. The CNN model enables the CNN model itself to learn features in the hidden layers. Another pattern regards CNN just as a classifier, it extracts features in advance using other methods like CSP [8,9] or Short-Time Fourier Transform(STFT) [10], then feeds those features to train the parameters of the network.

Recent years, significant advancement has been achieved to improve the capability of CNNs due to the boost in the filed of computer vision and natural language processing. For example, lots of activation functions has been proposed to address problems like vanishing gradient and bias shift in the learning procedure of the network [11,12,13]. In 2010, ReLU (Rectified Linear Unit) [14] is proposed to alleviate the gradient vanishing problem induced by the original sigmoid or tahn functions. However, there are no negative values in the units of the network after the ReLU transformation, which results a slower learning. In 2016, Djork-Arne introduced ELU (Exponential Linear Unit) [15] to make the network have negative values and the mean of the units closer to zero, and as a result, fasten the learning procedure. In 2017, Günter Klambauer [16] enables CNN to have self-

978-1-5386-3142-3/17/$31.00 ©2017 IEEE

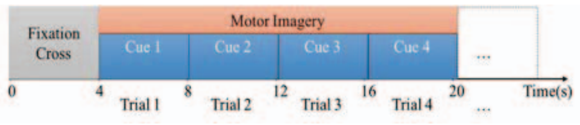

Fig.1 Experiment paradigm

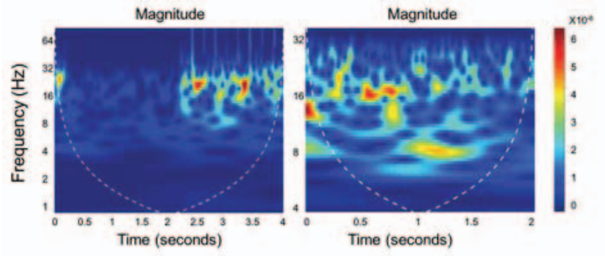

Fig.2 Energy distribution in the whole 4 seconds recording period from a single electrode

normalizing properties by proposing SELU (Scaled Exponential Linear Unit) function, and makes the network more robust to noise and permutations.

The objective of this paper is applying CNN to analysis MI based EEG signals, and comparing the performance of the networks by employing three different activation functions. Specifically, raw EEGs will be transferred into time-frequency images using STFT, and these images will be the input of CNN. Three candidate activation functions are ReLU, ELU and SeLU, and the speed of convergence and accuracy of classification are used as the index to measure the performance for comparison. Results show that our proposing model achieves better accuracy than the conventional method using CSP and SVM, and SELU outperforms other activation functions.

The structure of this paper is as follows. Section II describes the procedures for data collection and the proposing EEG re-presentation pattern. Section III specifies the construction of our CNN model. Then section IV reviews the activation functions employed in this work. Experimental results are described in section V and finally Section VI concludes the paper.

II. Data Collection and Re-presentation

Our dataset was collected from fourteen different subjects, their ages varied from 24 to 28 and they were all right handed. The bio-signal amplifier used in this experiment was ActiCHamp with 64 channels and the recording electrode set consisted of C3, Cz and C4. The sample rate was 256Hz.

Each person should complete 60 trials, and there were 30 left and 30 right hand movement imagery tasks respectively. At the start of the experiment, there was fixation cross in the center of the screen in order to arouse the attention of the subject, then 60 cues, either left array or right array showed up successively, reminding the subject to do the corresponding MI tasks, and each task lasted for 4s. The paradigm of this experiment is shown in Fig.1. Fig.2 reveals the energy distribution in the whole 4 seconds recording period. It can be seen clearly that MI task related signals mainly show up in the last 2 seconds and within 8-30 Hz, so a band pass filter was applied and signals in the first 2 seconds were cut off.

After the pre-processing work, the EEGs from three electrodes were re-presented by Short Time Fourier Transform (STFT) into time-frequency images, as shown in Fig.3. The window size of STFT was 64 sample points and the overlap was 50% of the window. Each of the window obtained seven frequency components. Since there was 512 sample points in a single electrode, when the window slid forward over the temporal signal, 7x15 values would be obtained, and all the values from three electrodes (i.e. C3, Cz, C4) for each trial were concatenated together to construct the matrix with dimensions 21x15. The original image obtained from this matrix yielded large pixels, so it was resized into 48x36 pixels for the concern of computation complexity.

III. CNN Construction

After the temporal signals transformation work, the images were fed as input of the CNN model. There were totally seven layers in our network, and the construction was described in Fig.4. The first layer was the input layer; the second (C1) layer was a convolutional layer with kernel size 3x3; then followed a max pooling layer with kernel size 2x2; and then there was another convolutional layer (C3) with kernel size 3x3 and max pooling layer (S4) with kernel size 2x2 behind; the last two layers were fully connected layers, consisting of 100 and 2 neuros respectively. A softmax classifier was concatenated at the end of the layers to compute the predicted labels.

The objective function used in this model was cross entropy, and the learning procedure applied stochastic gradient descent with learning rate 1e-4. All the construction work was completed on the open source platform tensorflow.

IV. Review of the Activation Functions

There were totally three different activation functions employed in our work. The necessity of using these functions was to make a comparison in recognition accuracy and speed of convergence. Their expressions are shown as below:

ReLU (Rectified Linear Unit):

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{1}$$

ELU (Exponential Linear Unit):

TABLE I. Classification Accuracy of the Four Comparing Methods. Values in the Parentheses Are the Standard Derivatior

Subject	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	mear
CSP+SVM	89.17 (8.33)	68.06 (7.28)	100 (0)	90.28 (5.32)	100 (0)	77.32 (6.29)	70.00 (8.66)	82.37 (8.30)	82.17 (5.15)	100 (0)	74.44 (4.60)	92.78 (3.72)	75 (9.27)	94.79 (2.08)	83.68
CNN(ReLU)	88.88 (4.55)	84.58 (8.25)	100 (0)	66.67 (4.54)	100 (0)	77.78 (0)	84.43 (7.87)	71.94 (6.99)	81.95 (5.32)	100 (0)	77.78 (4.54)	93.12 (3.54)	88.61 (2.78)	98.61 (2.78)	86.74
CNN(ELU)	89.1 (3.83)	83.3 (7.12)	100 (0)	84.32 (5.28)	100 (0)	83.7 (6.45)	89.44 (2.28)	74.9 (7.51)	78 (2.31)	100 (0)	79.6 (4.3)	92.28 (4.12)	94.7 (3.21)	94.2 (2.14)	88.92
CNN(SELU)	95.78 (2.83)	88.61 (2.78)	100 (0)	87.49 (5.31)	100 (0)	84.44 (4.84)	93.08 (3.01)	80.83 (8.33)	88.89 (2.40)	100 (0)	87.5 (7.22)	95.83 (2.66)	95.71 (1.27)	100 (0)	92.73

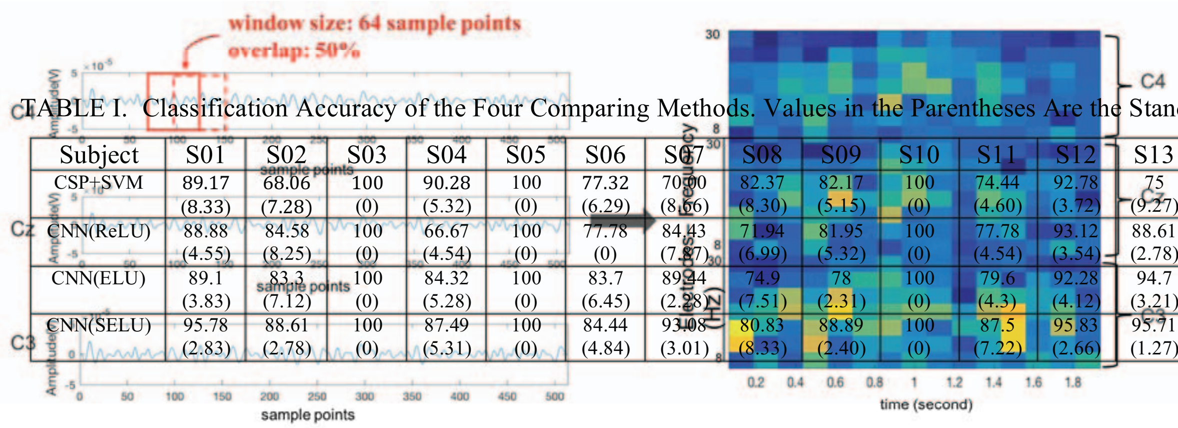

Fig.3 EEG data re-presentation by Short-Time Fourier Transform

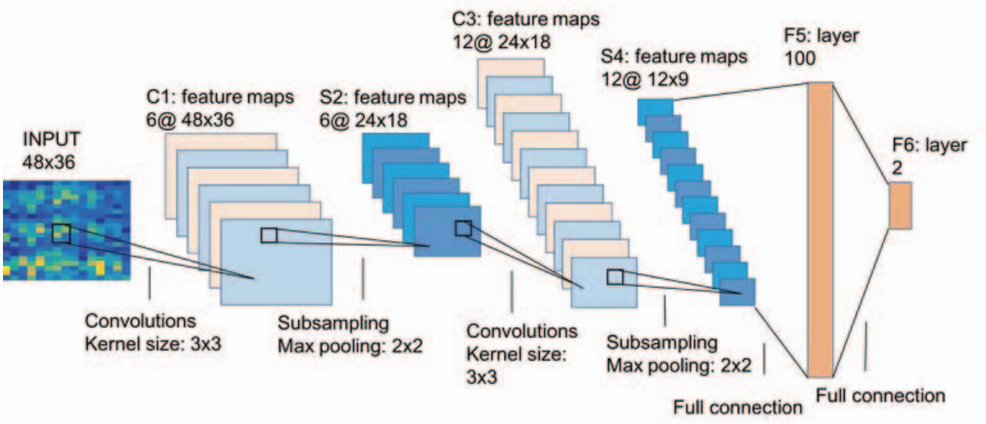

Fig.4 The construction of our CNN model

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha(\exp(x) - 1), & x \le 0 \end{cases} \quad (2)$$

since ELU saturates to a negative value when the input gets smaller, so the parameter α controls to which negative value ELU saturates.

SELU (Scaled Exponential Linear Unit):

$$f(x) = \lambda \begin{cases} x, x > 0 \\ \alpha(\exp(x) - 1), x \le 0 \end{cases} \quad (3)$$

so SELU multiplies the parameter $\lambda (\lambda > 1)$ to the ELU function, which ensures a slope larger than one for positive inputs.

V. Experiment

In this section, we explain our experimental results. First we applied the proposed data re-presentation and CNN models to analyze our datasets, a comparison was made on the classification accuracy between CNNs and the conventional method. And then four confusion matrices were calculated to measure the performance of the classifiers. At last, the speed of convergence helped us

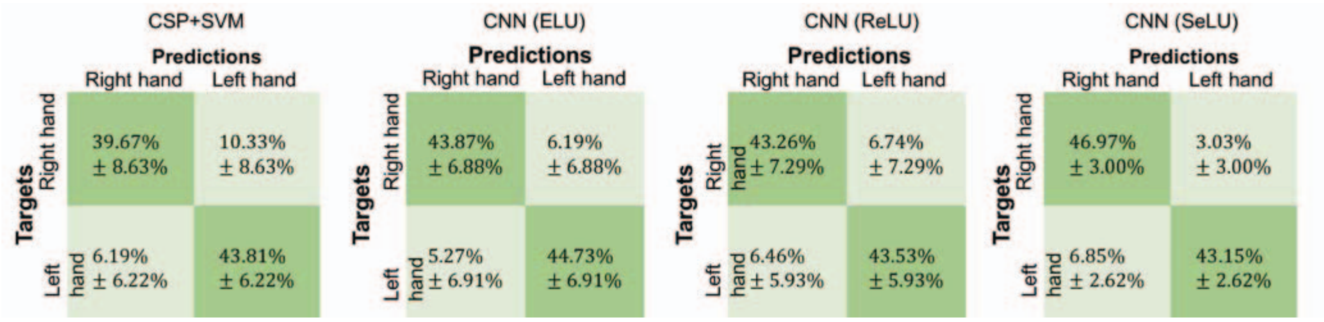

Fig.5 Confusion matrices of the four classifiers

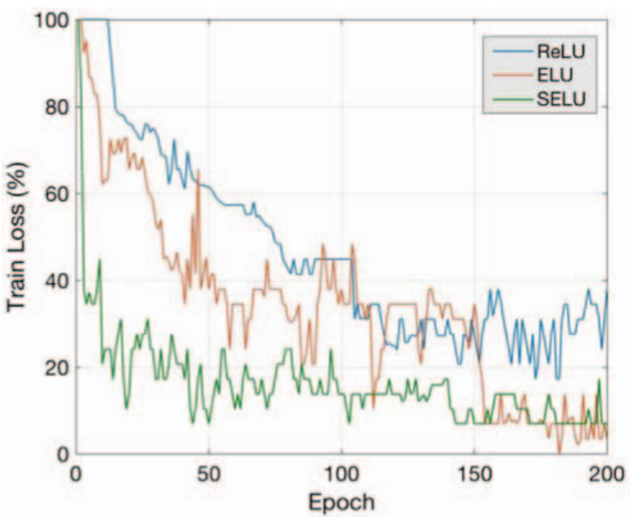

Fig.6 Training loss of the CNNs when they were employing ReLU, ELU and SELU as activation functions.

look insight into the behavior of the three activation functions.

For each subject, we separated the whole 60 trials into train set and test set, which included 42 trials and 18 trials respectively. Table 1 illustrated the classification accuracy among 14 subjects, the results were achieved under four-fold cross validation and each network was trained for 800 epochs.

Comparing to the first method (CSP for feature extraction and SVM for classification), the CNN models achieved higher mean accuracies by 3.06%, 5.14% and 9.05%. Specifically, for Sub.13, the accuracy for CNN with SELU was 30.71% higher than the first method. However, for Sub.4, the first method showed better performance.

Fig.5 explained the confusion matrices of four comparing methods. The right hand class was regarded as positive, and in contrast, the left hand class was regarded as negative. Each matrix included four values, and their expressions were shown as below. The upper left entry was the true positive rate (TPR), which meant the probability of detection; the upper right entry was the false negative rate (FNR), which meant the missing rate; the bottom left entry was the false positive rate, which measured the fall-out and the bottom right was the true negative rate (TPR), which measured the specificity. The four matrices could further prove that the CNN model with SELU was better for this two-class classification problem.

$$TPR = \frac{\sum TP}{\sum right\ hand\ samples} \tag{4}$$

$$FNR = \frac{\sum FN}{\sum right\ hand\ samples} \tag{5}$$

$$FPR = \frac{\sum FP}{\sum left\ hand\ samples} \tag{6}$$

$$TNR = \frac{\sum TN}{\sum left\ hand\ samples} \tag{7}$$

Additionally, we want to take the speed of convergence as the index to have an insight into the learning behavior of the activation functions. Three deep convolutional neural networks using ReLU, ELU($\alpha = 1.0$) and SELU ($\alpha = 1.0, \lambda = 2$) were trained on the dataset of a random subject (Sub.06) with learning rate 1e-4. The construction of the networks was consistent with the previous description. The training loss of each epoch was calculated to constitute the curve. As we can see in Fig.6, the training error of SELU decreased more rapidly than the other two methods.

VI. Conclusion

This paper utilizes deep convolutional neural networks to decode motor imagery based EEGs. Signals are transferred into images using Short-Time Fourier Transform in advance, then those images are fed into a 7 layer CNN model for further classification.

In the meantime, our work has compared the performance of three different activation functions, namely ReLU, ELU, SELU. The results show that all convolutional neural networks achieve better accuracy than the conventional method, and SELU convergence faster than the other activation functions.

ACKNOWLEDGMENT

This work was supported by Shanghai Science and Technology Innovation Action Plan Project (Grant No. 16511100900), the State Key Program of National Natural Science Foundation of China (Grant No.61332008), and Shanghai NSF research project (Grant No. 16JC1401300).

REFERENCES

[1] Yi, Weibo, Qiu Shuang, Qi Hong zhi, Zhang Lixin, Wan Baikun, Ming Dong. "EEG feature comparison and classification of simple and compound limb motor imagery." Journal of neuroengineering and rehabilitation vol.10(1) 2013, pp. 106-118.

[2] Kevric J, Subasi A. "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system". Biomedical Signal Processing and Control, 2017, vol.31, pp. 398-406.

[3] Qiu, Zhaoyang, Brendan Z. Allison, Jing Jin, Yu Zhang, Xingyu Wang, Wei Li, and Andrzej Cichocki. "Optimized motor imagery paradigm based on imagining Chinese characters writing movement." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2017).

[4] Zhang, Yong, Bo Liu, Xiaomin Ji, and Dan Huang. "Classification of EEG signals based on autoregressive model and wavelet packet decomposition." *Neural Processing Letters* 45, no. 2 (2017): 365-378.

[5] Schirrmeister, Robin Tibor, et al. "Deep learning with convolutional neural networks for brain mapping and decoding of movement-related information from the human EEG." *arXiv preprint arXiv:1703.05051* (2017).

[6] Liu Jingwei, Cheng Yin, Zhang Weidong. "deep learning EEG response representation for brain computer interface." Proceedings of the 34th Chinese Control Conference, 2015.

[7] Tang, Xianlun, Na Zhang, Jialin Zhou, and Qing Liu. "Hidden-layer visible deep stacking network optimized by PSO for motor imagery EEG recognition." *Neurocomputing* 234 (2017): 1-10.

[8] Kumar, Shiu, Alok Sharma, Kabir Mamun, and Tatsuhiko Tsunoda. "A Deep Learning Approach for Motor Imagery EEG Signal Classification." In *Computer Science and Engineering (APWC on CSE), 2016 3rd Asia-Pacific World Congress on*, pp. 34-39. IEEE, 2016.

[9] Yang, Huijuan, Siavash Sakhavi, Kai Keng Ang, and Cuntai 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*, pp. 2620-2623. IEEE, 2015.

[10] Tabar, Yousef Rezaei, and Ugur Halici. "A novel deep learning approach for classification of EEG motor imagery signals." *Journal of neural engineering* 14, no. 1 (2016): 016003.

[11] Trottier, Ludovic, Philippe Giguère, and Brahim Chaib-draa. "Parametric exponential linear unit for deep convolutional neural networks." *arXiv preprint arXiv:1605.09332* (2016).

[12] Lin, M., Chen, Q., and Yan, S. (2013). Network in network. arXiv preprint arXiv:1312.4400. Maas, A. L., Hannun, A. Y., and Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. Citeseer.

[13] Maas, A. L., Hannun, A. Y., and Ng, A. Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the 30th International Conference on Machine Learning (ICML13), 2013.

[14] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In ICML, pages 807–814.

[15] Clevert, Djork-Arné, Thomas Unterthiner, and Sepp Hochreiter. "Fast and accurate deep network learning by exponential linear units (elus)." *arXiv preprint arXiv:1511.07289* (2015).

[16] Klambauer, Günter, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. "Self-Normalizing Neural Networks." *arXiv preprint arXiv:1706.02515* (2017).