IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 16, 2018, accepted December 6, 2018, date of publication December 21, 2018,
date of current version January 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2889093

Wavelet Transform Time-Frequency Image and Convolutional Network-Based Motor Imagery EEG Classification

BAOGUO XU[1], LINLIN ZHANG[1], AIGUO SONG[1], (Senior Member, IEEE), CHANGCHENG WU[2], WENLONG LI[1], DALIN ZHANG[1], GUOZHENG XU[3], HUIJUN LI[1], AND HONG ZENG[1]

[1]State Key Laboratory of Bioelectronics, Jiangsu Key Lab of Remote Measurement and Control, School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China
[2]College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
[3]College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Corresponding author: Baoguo Xu (xubaoguo@seu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB1001303, in part by the National Natural Science Foundation of China under Grant 61673114, Grant 91648206, Grant 61803201, Grant 61673105, Grant 61502096, and Grant 61501431, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20170803 and Grant BK20141426, and in part by the Fundamental Research Funds for the Central Universities under Grant 2242016R30025.

ABSTRACT Feature extraction and classification play an important role in brain–computer interface (BCI) systems. In traditional approaches, methods in pattern recognition field are adopted to solve these problems. Nowadays, the deep learning theory has developed so fast that researchers have employed it in many areas like computer vision and speech recognition, which have achieved remarkable results. However, few people introduce the deep learning method into the study of biomedical signals, especially EEG signals. In this paper, a wavelet transform-based input, which combines the time-frequency images of C3, Cz, and C4 channels, is proposed to extract the feature of motor imagery EEG signal. Then, a 2-Layer convolutional neural network is built as the classifier and convolutional kernels of different sizes are validated. The performance obtained by the proposed approach is evaluated by accuracy and Kappa value. The accuracy on dataset III from BCI competition II reaches 90%, and the best Kappa value on dataset 2a from competition IV is greater than many of other methods. In addition, the proposed method utilizes a resized small input, which reduces calculation complexity, so the training period is relatively faster. The results show that the method using convolutional neural network can be comparable or better than the other state-of-the-art approaches, and the performance will be improved when there is sufficient data.

INDEX TERMS Brain computer interface (BCI), motor imagery (MI), wavelet transform time-frequency image, convolutional neural network (CNN).

I. INTRODUCTION

The brain is the central nervous system and activity control center of the human body. People rely on brain activity to control the body to respond differently to external stimulus. For people with movement disorder, their brain consciousness is normal but the motion intention cannot be implemented. The brain-computer interface (BCI) system has created a pathway which is independent of the surrounding nerves and muscles for them [1]–[3]. The BCI system converts the electroencephalogram (EEG) signal into the peripheral control signal, which functions like 'decoding'. It has been widely used in the field of rehabilitation training.

When people actually perform or only imagine unilateral limb movements (such as left or right hands), the activity state of different parts of the brain will change. The ipsilateral (relative to subjects' unilateral limbs) and the contralateral brain sensorimotor cortex will respectively show phenomena called event-related synchronization (ERS) and event-related desynchronization (ERD). It is manifested by the increase of mu and beta rhythm energy of ipsilateral sensory motor cortex and the decline of contralateral mu and beta rhythm energy [4]. This is the physiological basis for the classification of motor imagery (MI) EEG signals.

According to the traditional classification method of motor imagery EEG, feature extraction and recognition are two relatively independent parts. Most researchers focus on feature extraction algorithms. At present, the feature extraction algorithms mainly include adaptive auto regressive (AAR), short-time Fourier transform (STFT), wavelet transform (WT) and common spatial pattern (CSP) and other methods [5]–[9]. Feature recognition methods are mainly linear discriminant analysis (LDA), support vector machine (SVM) and Bayesian classifiers [10]–[12].

A lot of research shows that the traditional feature extraction algorithm is relatively mature, but there are still some problems. In the above methods, the AAR model is more suitable for analyzing stationary signals. For the STFT, the width of the window determines the time resolution and frequency resolution, the wider the window, the higher the frequency resolution and the worse the time resolution. On the contrary, the narrower the window, the worse the frequency resolution and the better the time resolution, that is to say, in the STFT, the time resolution or the frequency resolution cannot be obtained simultaneously, and the limited window width will cause a certain spectrum leakage. WT is a kind of time-frequency analysis method that can perform multi-scale analysis. However, most researchers use statistical features such as maximum, minimum, mean and standard deviation of wavelet coefficients as EEG features, which need manual screening and are based on a certain prior knowledge, while the pattern of motor imagery signals is not fixed and can cause unpredictable changes due to individuals and trails.

In recent years, deep learning theory, as an important branch of machine learning, has been effectively applied and enriched in computer vision, speech recognition and natural language processing [13]–[15]. The significant characteristic of deep learning is that the model can automatically extract effective features, which has certain advantages in large scale data. Furthermore, some researchers have applied deep learning theory to the study of EEG signals. In [16], the CSP algorithm is used to extract the characteristics of the original EEG signal, and a four-layer neural network is trained to classify the two types of movement imagination signals. The first and the second layer are stacked by automatic encoders, through which they learn to input new features of data that are more abstract. Through the comparison experiments of several models, it is found that the calculation amount will be greatly reduced when feature extracted by CSP is used as the input of the network. Reference [17] uses the deep belief network (DBN) and the implicit Markov model (HMM) to detect emotions (positive/negative), adopting the differential entropy of signals in five frequency bands and their sum in the EEG signal as the feature. In comparison with the traditional SVM and KNN models, the higher classification accuracy is obtained. This result shows that DBN assigns different weights to different feature components based on artificially provided features, and filters out invalid feature components. Reference [18] defines ERD/ERS as the percentage of signal power rise or fall in the reference time, using this percentage as the input as the 5-layer convolution network to complete the classification of left and right hand motion imaging EEG, and the accuracy improved by 5%-10% compared with traditional method [19].

These models using deep neural network as the classifier still extract the features manually as input. However, due to the characteristics of deep network nonlinearity, it can learn more abstract features, and use more information than traditional machine learning methods. Thus, there will be a raise in identify accuracy. In addition, some researchers no longer deal with feature extraction separately, instead, they use deep network to learn features automatically to complete classification. In [20], in order to improve the classification accuracy of P300 and non-P300 signals in the auxiliary spelling BCI system, a 5-layer convolutional neural network was established with 64 channels of time series as input, and CNN automatic feature learning was used to complete the classification. However, since each channel has rich frequency information, this model is less interpretable.

In order to provide a relatively general classification model of motor imagery EEG, we propose a new input form, which uses wavelet transform to convert multichannel EEG signals into two-dimensional time-frequency images, so as to obtain its comprehensive information, including both the time-frequency features and the relative position of the electrodes. Additionally, convolutional neural network is utilized as the classifier. Although reference [21] has proposed image as input, multi-scale analysis of wavelet transform avoids the problem of window size selection in STFT method. Moreover, convolutional network is more suitable for image classification than stacked automatic encoder. The presented method is tested and evaluated on Dataset A and Dataset B. The results are compared with those in the competition.

II. DATASETS

As shown in Table 1, our EEG data was from two sources:

(1) Dataset A was the dataset III from BCI competition II [22]. It was recorded from a normal subject (female, 25y). The experiment consists of 280 trials of 9s length in total, and the break between each trail is ranging from 0.5s to 2s. The experimental paradigm for each trial is the same, as is illustrated in Fig. 1. The first 2s was quite, at t=2s an acoustic stimulus indicates the beginning of the trial, and a cross "+" was displayed for 1s; then at t=3s, an arrow (left or right) was displayed as the cue. Meanwhile, the subject was asked to conduct the imagery task according to the arrow. The EEG was sampled with 128Hz, it was filtered between 0.5 and 30Hz.

(2) Dataset B was the dataset 2a from BCI competition ćô [23]. This dataset consists of EEG data from 9 subjects. The cue-based BCI paradigm consisted of four different motor imagery tasks, namely the imagination of the left hand, right hand, both feet, and tongue. Two sessions were recorded for each subject. Each session is comprised of 6 runs separated by short breaks. One run consists of 48 trials (12 for each of the four possible classes), yielding a total

TABLE 1. Datasets properties.

Name	Source	Subjects	Categories	Channels	Sampling frequency(Hz)
Dataset A	BCI competition II dataset III	1	2	3	128
Dataset B	BCI competition IV dataset 2a	9	4	22	250

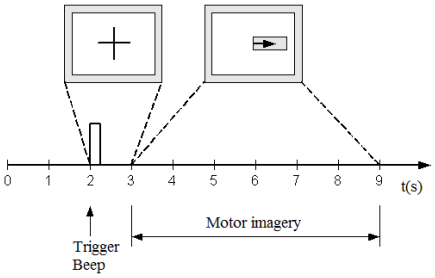


FIGURE 1. The experimental paradigm for each trial.

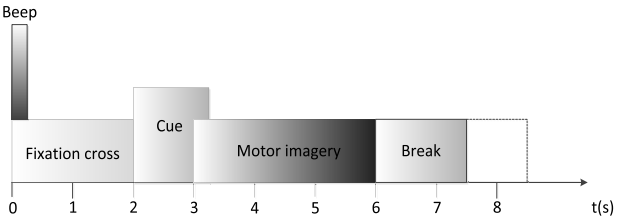


FIGURE 2. The experimental paradigm for each trial.

of 288 trials per session. The paradigm is illustrated in Fig. 2, at the beginning of a trial (t=0s), a fixation cross appeared on the black screen. In addition, a short acoustic warning tone was presented. After two seconds (t=2s), a cue in the form of an arrow pointing either to the left, right, down or up (corresponding to one of the four classes left hand, right hand, foot or tongue) appeared and stayed on the screen for 1.25s. This prompted the subjects to perform the desired motor imagery task. The signals were sampled with 250 Hz and band-pass filtered between 0.5 Hz and 100 Hz.

III. PREPROCESSING

The preprocessing of the EEG signal is mainly to remove the noise in the signal, to filter out the frequency components unrelated to the motor imagery task. Some digital signals contain low-frequency noise, which affects the analysis of the signal. It needs to be preprocessed to eliminate the signal baseline. By eliminating the trend from the data, the analysis can be concentrated on the fluctuation of data. A linear trend usually indicates a systematic increase or decrease. The non-linear trend can be measured by wavelet analysis, and signal is multi-scale analyzed by wavelet transform. The baseline trend of the signal can be seen from the low-frequency coefficient and it can be subtracted from the original signal.

The ERD/ERS pattern of EEG is mainly reflected in the frequency band of 8-30Hz. Different imagery tasks are different in the characteristic frequency band and brain area, and vary from person to person, even different trails of the same person. The ERD/ERS mode generated during the task is also different. Therefore, it is too complicated to extract the active segment of the relevant frequency band for each task of each person. We perform band-pass filtering on the original signal and select the frequency band of 8-30Hz.

The EEG signal itself has a low signal-to-noise ratio, and it greatly affects the accuracy of the final classification. There are two main types of noise sources for EEG signals, one is non-EEG artifacts such as myoelectric signals and eye movement signals, the other is the frequency component of the EEG signal that is unrelated to the motor imagery task. The two types of signals are different in frequency distribution from the useful components. The myoelectric signal has a relatively wide frequency range, and the larger amplitude portion is distributed in the frequency greater than 30 Hz; the eye movement signal is concentrated in the low frequency range of less than 5 Hz, especially in forehead. In this paper, we use a spatial filter to maximize the signal-to-noise ratio. Reference [24] pointed out that common average reference (CAR) filter emphasizes the common components in most channels, and it removes these components from the specified channels to increase the difference between different channels. Conversely, if some of the components are present in most channels but not in the interest channel, artifacts will be produced on interest channel. The Laplacian method has a similar effect to CAR, but uses only a small subset of neighborhood channels. Laplacian spatial filtering is calculated according to (1) and (2).

$$V_I^{LAP} = V_i^{ER} - \sum_{j \in Si} g_{ij} V_j^{ER} \tag{1}$$

$$g_{ij} = 1/d_{ij} / \sum_{j \in Si} d_{ij} \tag{2}$$

In equation (1) and (2), V_i^{LAP} is the value of the i-th channel filtered by the Laplacian method; V_i^{ER} is the potential difference between the i-th electrode and the reference electrode; S_i represents the set of neighboring electrodes of the i-th electrode, and d_{ij} indicates the distance between the i-th and j-th channels.

IV. INPUT DATA

The ERD/ERS phenomenon of left and right hand motor imagery occurs in the C3 and C4 regions of the cerebral

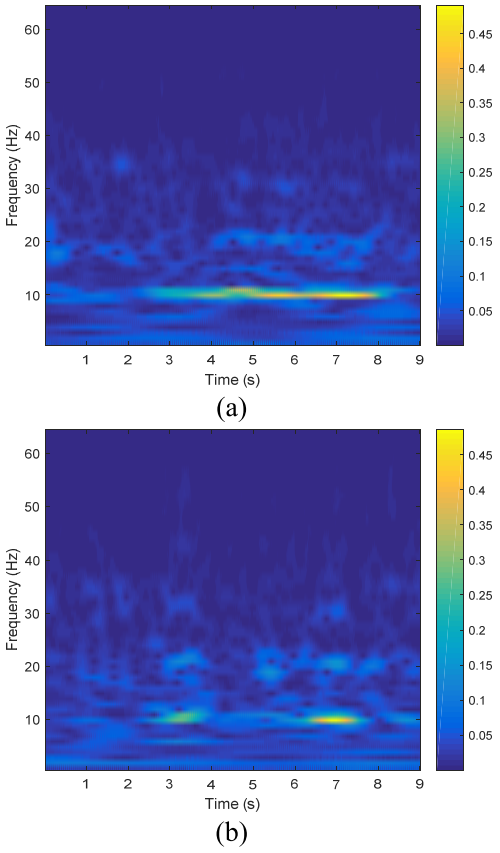


FIGURE 3. Wavelet time-frequency images of the C3 (a) and C4 (b) channels when the subject performing the left-hand movement imagery task. It can be seen that the energy of the C4 channel decreases significantly 3 seconds after the trail begins, then recovers after a period of time, which means the ERD phenomenon occurs. However, the energy of the C3 channel remains at a high level instead of decreasing, which is called the ERS phenomenon.

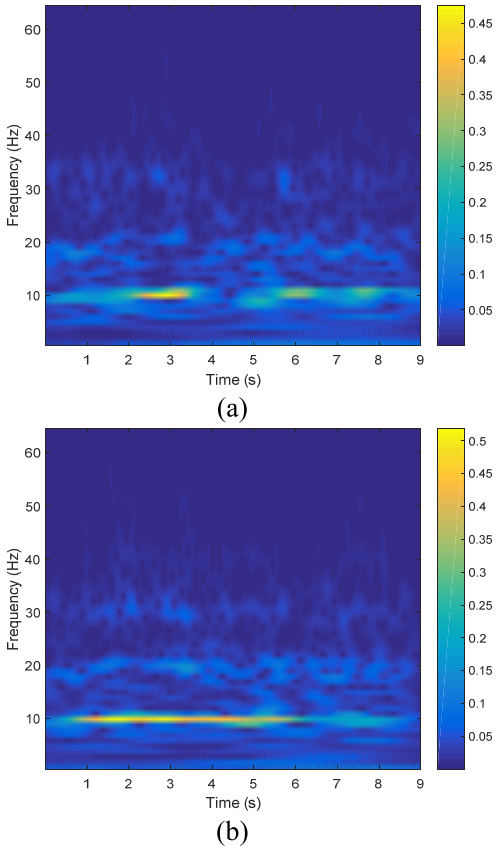


FIGURE 4. Wavelet time-frequency images of the C3 (a) and C4 (b) channels belonging to the trails of right hand imagery task. Obviously, this phenomenon is opposite to that of Fig.3 (a) and (b). The C4 channel displays ERS and the C3 channel presents ERD phenomenon.

cortex, which means the energy change of the mu rhythm and beta rhythm of EEG signals in these regions. Specifically, when the subject performs the left hand motor imagery, the power amplitude of the C3 channel increases in the 8-12 Hz band, that is, the ERS phenomenon occurs, and the power amplitude of the C4 channel in the 8-12 Hz band decreases, which means the ERD phenomenon occurs. Conversely, when the subject imagines the right hand movement, the ERD occurs in the C3 channel, and the ERS appears in the C4 channel. The corresponding wavelet time-frequency images are shown in Fig.3 and Fig.4. These images of C3, C4, and Cz were combined as the final input images, which are illustrated in Fig.5.

V. CONVOLUTIONAL NETWORK FRAMEWORK

A typical CNN network has the following levels: the input layer, the convolution layer, the pooling layer, the fully connected layer, and the output layer. The convolution kernel extracts the feature of the local receptive field of the input image, then the feature is used as the input of pooling layer. The abstraction ability of CNN is positively correlated to the number of layers and the parameters to be learned. The more layers included, the stronger the abstraction ability is. Meanwhile, the model is more likely to be over-fitting while the network is going deeper. The purpose of the pooling layer is to reduce the data dimension. Max pooling and average pooling are the common strategies employed in CNN, functioning as the sub-sampling layer. We calculate the convolution layer and pooling layer alternately, so that the size of feature map shrinks and the number of channels increases. When it comes to the fully connected layer, the previously obtained features are flattened, and the classification results are finally achieved through the feedforward network.

Throughout the development of deep learning, the structure of the model is the priority in researches. With the beginning of LeNet-5, several classic CNN models emerged, such as AlexNet, VGG, GoogleNet and ResNet. Before the emergence of GoogleNet, the breakthrough of the mainstream network structure is to roughly increase the layers of network and the width of each layer. To handle the problem how to choose the kernel size and whether to use the pooling layer, the researchers has proposed the inception blocks, and the inception network works remarkably well. However, many of the mature networks are huge and not appropriate for our small size training set.

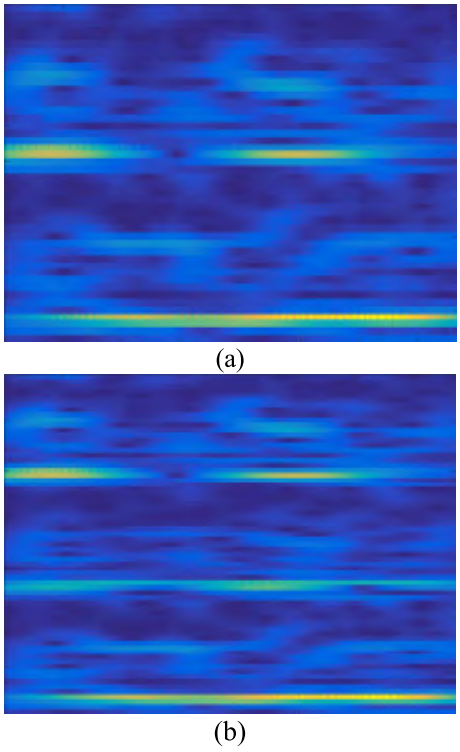


FIGURE 5. Final combined input image of 2 channels (C3 and C4) (a) and that of 3 channels (C3, Cz, and C4) (b).

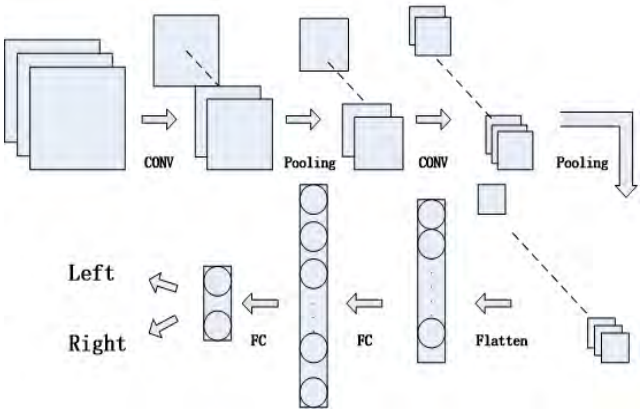


FIGURE 6. Structure of the 2-Layer convolutional network.

The amount of the data in BCI datasets is not sufficient for the classic models mentioned above, because they are difficult to train and more likely to be over fitting. A 2-Layer convolution network is proposed for the small dataset. Additionally, the input images are uniformly resized to 64*64, and the kernel size is adjusted accordingly. Two types of kernels is adopted in this paper, one is the 2D kernel, and the other is 1D kernel. The structure of the 2-Layer network is shown in Fig. 6.

The nonlinear characteristics of the neural network are largely derived from the nonlinear excitation function of the neuron. The sigmoid function used in the perceptron is also called the logistics function. The mathematical expression is shown in Eq. (3), which can map a real number in the interval

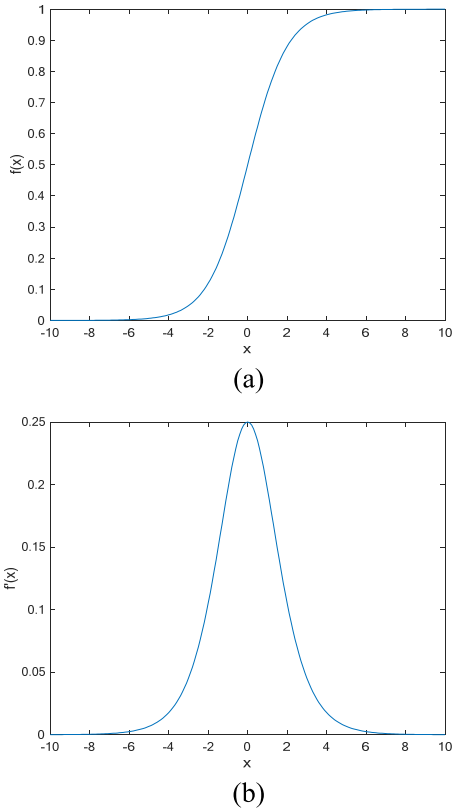


FIGURE 7. The curve of sigmoid function (a) and the derivative curve of sigmoid function (b).

of (0, 1). Nevertheless, when performing back propagation, derivative of the activation function needs to be calculated. The derivative of sigmoid is shown in Eq. (4). The curve of sigmoid function and its derivative are shown in Fig.7. It can be seen that the gradient starts from 0 and returns to 0, and its maximum is only 0.25, so the gradient disappears after chain-based derivation calculation, which is called gradient vanishing problem. As a result, relu function is chosen, its derivative is always equal to 1 if the input is greater than zero. Using relu in the deep network can avoid the above problem. The function representation and derivative of relu is shown in Eq. (5)-(6).

$$f(x) = \frac{1}{1+\exp(-x)} \tag{3}$$

$$f'(x) = f(x)(1-f(x)) \tag{4}$$

$$f(x) = \max(0, x) \tag{5}$$

$$f'(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{6}$$

In the training of neural network, the gradient descent algorithm is used iteratively. Gradient descent algorithm has 3 derivatives according to the size of samples included in the calculation of the loss function in each iteration: full-scale gradient descent (all samples are included), batch gradient descent (a batch of samples are included) and stochastic gradient descent (one sample is randomly selected to calculate

TABLE 2. Classification results using different channels and mother wavelets.

Wavelet name	Accuracy (%)					
	2 channels			3 channels		
	Worst	Best	Mean	Worst	Best	Mean
db4	77.63	81.25	79.25	71.43	80.60	75.60
sym4	80.14	85.50	81.38	71.43	82.14	73.21
cmor3-3	87.50	92.75	89.56	78.25	83.50	82.37
haar	68.25	72.50	70.31	64.30	69.6	67.13

the loss). Usually, the mini-batch gradient descent is used to ensure convergence to a better global optimal solution when the loss function is a convex function, but the learning rate is a fixed value, and it directly affects the quality of the final solution. If learning rate is set too small, the convergence will be very slow, and setting it too large will cause the oscillation to converge. Moreover, the same learning rate is not applicable to the updating of all parameters. For non-convex problems, it can only converge to the local optimal solution using gradient descent, which cannot get rid of the local minimum and saddle points. Therefore, we utilize Adam algorithm to obtain the adaptive learning rate for each parameter [25].

In practical applications, compared with other adaptive optimization algorithms, Adam converges faster, and can avoid problems in other optimization algorithms, such as uniform learning rate, slow convergence and high variance in parameters updating, which cause obvious fluctuations in the loss function. Adam is essentially an RMSprop with a momentum term, which dynamically adjusts the learning rate of each parameter using the first-order moment estimate and the second-order moment estimate of the gradient. The main advantage of Adam is that after the offset correction, the learning rate in each iteration has a certain range, which makes the parameters relatively stable. The Adam adjusts its parameter according to Eq. (7)-(11).

$$m_t = \mu * m_{t-1} + (1-\mu) * g_t \tag{7}$$

$$n_t = v * n_{t-1} + (1-v) * g_t^2 \tag{8}$$

$$\hat{m} = \frac{m_t}{1-\mu^t} \tag{9}$$

$$\hat{n} = \frac{n_t}{1-v^t} \tag{10}$$

$$\Delta\theta_t = -\frac{\hat{m}_t}{\sqrt{\hat{n}_t}+\varepsilon} * \eta \tag{11}$$

Among the equations, m_t is the first moment estimation and n_t is the second moment estimation of the gradient. Respectively, $\hat{m}$ and $\hat{n}$ are the corrections to them.

VI. RESULTS AND DISCUSSION

A. PERFORMANCE VALIDATION ON DATASET A

In order to obtain a stable and reliable model, cross-validation is adopted to evaluate the performance of the classifier. Cross-validation, also known as loop estimation, captures as much information as possible when the training set is not large enough, and to some extent, can avoid the model from over-fitting. In this paper, the K-fold cross-validation is employed. K is set to be 5, and the average of these 5 models' results are calculated as the final result.

Reference [26] and other research papers show that the ERD/ERS mode during the right and left hand motor imagery occurs in the sensory motor cortex on the corresponding left and right sides, the regions beneath the electrodes C3 and C4. We compare the results of including Cz channel in the input image and those of not including it. The comparison is shown in Table 2. It can be seen that EEG signals of C3 and C4 channels are enough to classify different imagery tasks. Instead of improving the classification accuracy, it seems to introduce some noise while the Cz channel is included.

In addition, in the selection of the mother wavelet, it is generally desirable to obtain a smooth continuous wavelet amplitude when analyzing the time series, and thus the non-orthogonal wave function is suitable. The Morlet wavelet is not only non-orthogonal but also an exponential complex wavelet modified by Gaussian. In this paper, different mother wavelets are selected for comparison and the results are shown in Table 2. It can be seen that the classification results based on Morlet wavelet transform are relatively better. The time-frequency images of the signals after these wavelet transformation are demonstrated in Fig. 8. After the Mortlet wavelet transformation, the energy is more concentrated. The Daubechies wavelets, which usually called 'dbN' for short, have different vanishing moments. The letter N in 'dbN' represents the order of vanishing moment. The wavelet with higher-order vanishing moment has a better resolution in frequency domain, but the compact support in time domain is weakened, and there is an increase in calculation. The Symlets wavelet system was proposed by Daubechies, and its approximate symmetric wavelet function is an improvement of the Daubechies function. It is obvious to see from the Fig. 8 that the images after 'db4' and 'sym4' wavelet transformation are similar. In addition, Haar wavelet is a special case in Daubechies wavelets when its support length equals to 1.

As we can see from Table 2, wavelet transform based on 'cmor3-3' gets the highest accuracy on both types of input. Specifically, 92.75% for two channels input and 83.50% for

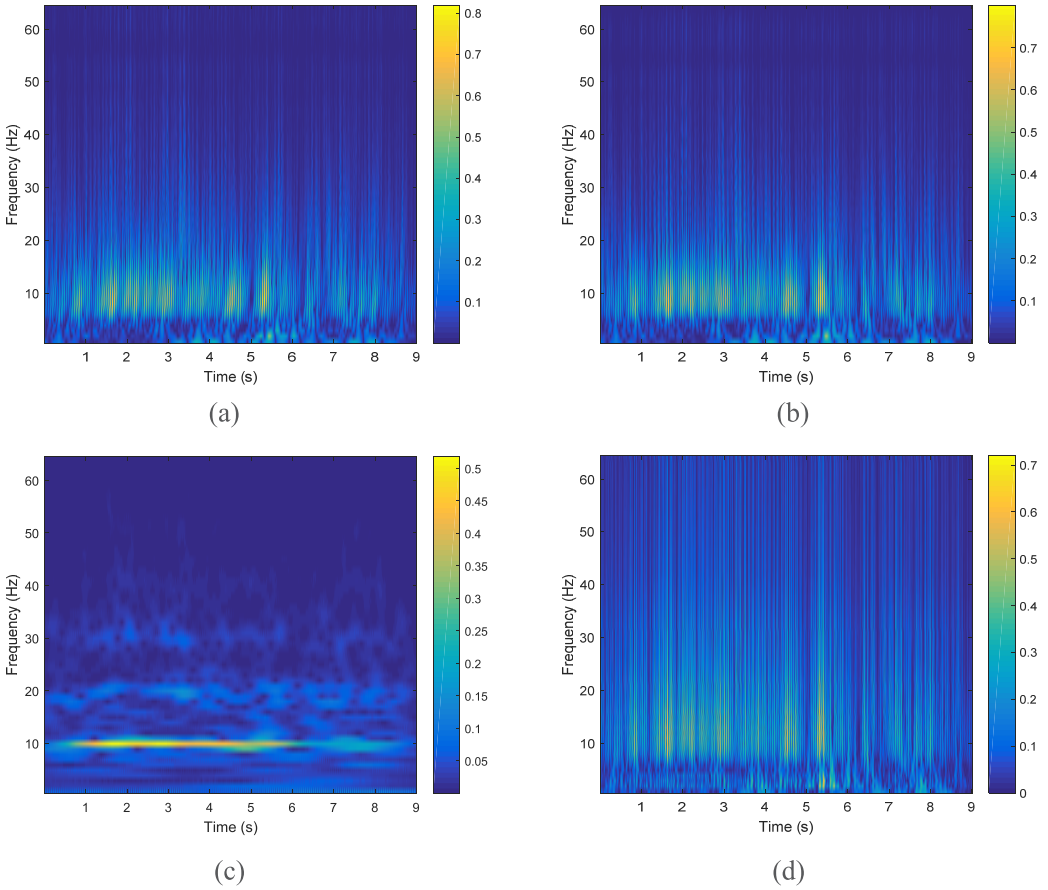


FIGURE 8. Transformed images using mother wavelet: db4 (a), sym4 (b), cmor3-3 (c) and haar (d).

TABLE 3. Results of different time duration.

Duration (s)	Accuracy (%)		
	Worst	Best	Mean
3.00-6.00	78.50	84.09	81.49
3.25-6.25	87.50	92.75	89.56
3.50-6.50	81.63	87.22	83.21
3.75-6.75	73.55	86.30	77.50
4.00-7.00	70.10	78.34	74.26

three channels input. From Fig. 8, it can also be observed that the transform based on 'cmor3-3' wavelet (Fig. 8c) is more concentrated in frequency, while the transform based on 'haar' wavelet is the least concentrated one in Fig. 8(d). According to the images Fig. 8(a) and Fig. 8(b), it can be assumed that the result based on 'db4' is similar to 'sym4', and the result in Table 2 verifies this assumption. What's more, the accuracy of 3 channels is mostly lower than that of 2 channels, we can come to the conclusion that the effective components are mostly included in C3 and C4.

The experimental paradigm, as shown in Fig. 1, shows that the cue appears at 3rd s, and then the corresponding imagery task starts, so the time period selected in this paper is the data after 3rd s. However, if the whole period from 3 to 9 seconds are adopted, it is bound to introduce noise unrelated to the task. At the same time, considering the subject has a certain delay after seeing the arrow at the beginning of the imagination task, the sliding window is proposed to select the most appropriate time period. Our method starts from the 3rd s, with 0.25s as the sliding step and 3s as the window width, sliding to the 4th s.

According to the Table 3, the results demonstrate that the classification performance based on the 3.25-6.25s period is the best. It means there is indeed a delay when people perform motor imagery according to the cue. A smaller sliding step and different window width can be employed to precisely find the active segment. However, the active segment for each subject is different, and the effective duration where ERD/ERS mode occurs is uncertain, so the selection of time period is quite a difficult problem need to be settled.

B. PERFORMANCE VALIDATION ON DATASET B

Since Dataset B contains data from 9 subjects, we train the model separately for each subject, and test the two

TABLE 4. Results of 2D and 1D kernel on Dataset B.

Subjects	2D kernel			1D kernel		
	Best Accuracy (%)	F1 score	Time(s)	Best Accuracy (%)	F1 score	Time(s)
S1	85.71	0.86	90.52	82.14	0.82	21.59
S2	78.57	0.78	80.31	75.05	0.75	17.52
S3	92.15	0.93	86.23	90.10	0.89	18.81
S4	95.67	0.95	85.37	94.61	0.95	15.33
S5	89.20	0.88	74.23	92.30	0.92	18.36
S6	85.12	0.85	80.46	86.75	0.85	14.77
S7	79.23	0.80	80.90	76.60	0.77	14.82
S8	81.28	0.82	77.83	80.50	0.80	18.30
S9	80.67	0.82	82.09	79.55	0.80	17.11
Average	85.59	0.85	81.99	84.18	0.84	17.40

TABLE 5. Comparison of best Kappa values.

Subject	Best Kappa value			
	2L-CNN	CNN-SAE	DDFSB	RQNN
S1	0.696	0.738	0.710	0.640
S2	0.587	0.458	0.310	0.590
S3	1.000	0.845	0.750	0.650
S4	1.000	1.000	0.470	0.990
S5	0.830	0.750	0.190	0.460
S6	0.800	0.796	0.200	0.510
S7	0.716	0.699	0.780	0.810
S8	0.626	0.751	0.770	0.800
S9	0.640	0.550	0.730	0.770
Average	0.766	0.732	0.546	0.691

convolution kernels mentioned above, then record the training time, as demonstrated in Table 4. It can be seen from the table that the results based on one-dimensional convolution kernel and the two-dimensional convolution kernel are not much different from each other, and could always achieve a similar performance by adjusting parameters. Nevertheless, the model with the two-dimensional convolution kernel has more parameters to be learned and requires more time to train. Furthermore, no matter which convolution kernel is used, the time is quite saved due to the small size of the network. The 2D kernel size in the proposed approach is 4*4 for the first convolution layer and 2*2 for the second convolution layer. The 1D kernel size is 1*5 and 1*3. The number of kernels in the first and second convolution layer is 8 and 16 separately.

The best Kappa values of the proposed model are compared with the methods in the other literature, as is shown in Table 5. From this table, we can see that 2-layer convolution network presented in this paper performs better on several subjects. Meanwhile, the quality of the motor imagery EEG signals from each subject are different, resulting in a big difference in the accuracy of prediction.

VII. CONCLUSION

In this study, a wavelet transform time-frequency image and convolutional network based approach is proposed to classify motor imagery (MI) EEG signals, here MI tasks are left and right hand movement imagery. Dataset III from BCI competition ćò and dataset 2a from BCI competition IV are used to train and test the proposed model.

In the proposed method, wavelet transform is introduced to generate the input images of model. Wavelet transform is the priority option, because of its multi-scale analysis ability, avoiding selecting the window size in STFT method. The input in the form of image also retains the relative position between the channels. Taking the form of image as input, the abstraction ability of convolution network is utilized, which provides a new means for EEG signal classification. Experiments show that the method using convolutional neural network can achieve better results than traditional methods. The choice of neural network is based on the scale of training data and a two-layer convolution network is established. Furthermore, the image is re-sized to reduce the amount of parameters, thereby shortening the training time and reducing the risk of over-fitting.

It can be seen from the comparison of the above results that the proposed method can achieve the best accuracy of 92.75% in Dataset A, which is higher than 90% in [20]. And its best Kappa values among 9 subjects on Dataset B are mostly greater than those of other methods, which can be interpreted as the stronger ability of feature extracting. As the results show, the proposed method is efficient, especially using the 1D kernel, with less parameters to learn.

Obviously, there are two main factors affecting the performance of classification of motor imagery signals, one is the quality of the signal itself, such as the effective duration

and intensity of the ERD/ERS mode, and the other is the amount of examples. The larger scale of training data means the smaller proportion of noise, so the model can learn more useful information. As the training data of EEG signals is augmented, the performance will be improved accordingly.

REFERENCES

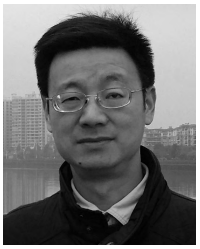
[1] I. Iturrate, J. M. Antelis, A. Kubler, and J. Minguez, "A noninvasive brain-actuated wheelchair based on a p300 neurophysiological protocol and automated navigation," *IEEE Trans. Robot.*, vol. 25, no. 3, pp. 614–627, Jun. 2009.
[2] J. R. Wolpaw, D. J. McFarland, T. M. Vaughan, and G. Schalk, "The wadsworth center brain-computer interface (BCI) research and development program," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 1–4, Jun. 2003.
[3] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
[4] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzerb, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 6, pp. 642–651, Dec. 1997.
[5] B.-H. Yang, G.-Z. Yan, R.-G. Yan, and W. Ting, "Adaptive subject-based feature extraction in brain–computer interfaces using wavelet packet best basis decomposition," *Med. Eng. Phys.*, vol. 29, no. 1, pp. 48–53, Jan. 2007.
[6] M. Shannon, H. Zen, and W. Byrne, "Autoregressive models for statistical parametric speech synthesis," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 3, pp. 587–597, Mar. 2013.
[7] H.-W. Sun, Y.-F. Fu, X. Xiong, J. Yang, C.-W. Liu, and Z.-T. Yu, "Identification of EEG induced by motor imagery based on Hilbert–Huang transform," *Acta Automatica Sinica*, vol. 41, no. 9, pp. 1686–1692, 2015.
[8] Y.-F. Fu, B.-L. Xu, Y.-C. Li, H.-Y. Li, C.-Y. Wang, and Z.-T. Yu, "Recognition of actual grip force movement modes based on movement-related cortical potentials," *Acta Automatica Sinica*, vol. 40, no. 6, pp. 1045–1057, 2014.
[9] E. A. Mousavi, J. J. Maller, P. B. Fitzgerald, and B. J. Lithgow, "Wavelet Common Spatial Pattern in asynchronous offline brain computer interfaces," *Biomed. Signal Process. Control*, vol. 6, no. 2, pp. 121–128, Apr. 2011.
[10] S. Bhattacharyya, A. Khasnobish, S. Chatterjee, A. Konar, and D. N. Tibarewala, "Performance analysis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data," in *Proc. Int. Conf. Syst. Med. Biol.*, Dec. 2010, pp. 126–131.
[11] S. Singla, S. N. Garsha, and S. Chatterjee, "Characterization of classifier performance on left and right limb motor imagery using support vector machine classification of EEG signal for left and right limb movement," in *Proc. 5th Int. Conf. Wireless Netw. Embedded Syst. (WECON)*, Oct. 2016, pp. 1–4.
[12] L. He, D. Hu, M. Wan, Y. Wen, K. M. von Deneen, and M. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.
[13] M. Ahmed, P. C. Shill, K. Islam, M. A. S. Mollah, and M. A. H. Akhand, "Acoustic modeling using deep belief network for Bangla speech recognition," in *Proc. 18th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2015, pp. 306–311.
[14] T. Dobhal, V. Shitole, G. Thomas, and G. Navada, "Human activity recognition using binary motion image and deep learning," *Procedia Comput. Sci.*, vol. 58, pp. 178–185, Dec. 2015.
[15] T. Mikolov, A. Deoras, S. Kombrink, L. Burget, and J. Černocký, "Empirical evaluation and combination of advanced language modeling techniques," in *Proc. 12th Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, vols. 1–5, 2011, pp. 605–608.
[16] S. Kumar, A. Sharma, K. Mamun, and T. Tsunoda, "A deep learning approach for motor imagery EEG signal classification," in *Proc. 3rd Asia–Pacific World Congr. Comput. Sci. Eng. (APWC-CSE)*, Dec. 2016, pp. 34–39.
[17] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "EEG-based emotion classification using deep belief networks," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.
[18] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11–18, Feb. 2017.
[19] Y. Zhang, B. Liu, X. Ji, and D. Huang, "Classification of EEG signals based on autoregressive model and wavelet packet decomposition," *Neural Process. Lett.*, vol. 45, no. 2, pp. 365–378, Apr. 2017.
[20] M. Liu, W. Wu, Z. Gu, Z. Yu, F. Qi, and Y. Li, "Deep learning based on batch normalization for P300 signal detection," *Neurocomputing*, vol. 275, pp. 288–297, Jan. 2018.
[21] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, p. 016003, Nov. 2017.
[22] B. Blankertz *et al.*, "The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1044–1051, Jun. 2004.
[23] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain–computer communication: Motivation, aim, and impact of exploring a virtual apartment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 473–482, Dec. 2007.
[24] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *Proc. 27th Annu. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 5, Jan. 2006, pp. 5392–5395.
[25] P. K. Diederik and B. Jimmy, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015.
[26] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. H. Lopes da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.

BAOGUO XU was born in Nantong, China, in 1981. He received the B.S. degree in measurement and control from the China University of Mine and Technology, Xuzhou, China, in 2004, and the Ph.D. degree in measurement and control from Southeast University, Nanjing, China, in 2009.

He is currently an Associate Professor with the School of Instrument Science and Engineering, Southeast University. His current research interests include brain computer interface, natural human–robot interaction, and rehabilitation robot.

LINLIN ZHANG was born in Jiangsu, China, in 1993. She received the B.S. degree in sensor network technology from Dalian Maritime University, Dalian, China, in 2016. She is currently pursuing the M.S. degree in instrument science and technology with Southeast University, Nanjing, China.

Her current interests include motor imagery EEG analysis and machine learning.



AIGUO SONG was born in Huangshan, China, in 1968. He received the B.S. degree in automatic control and the M.S. degree in measurement and control from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1990 and 1993, respectively, and the Ph.D. degree in measurement and control from Southeast University, Nanjing, in 1996.

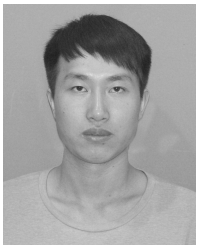
From 1996 to 1998, he was an Associate Researcher with the Intelligent Information Processing Laboratory, Southeast University. From 1998 to 2000, he was an Associate Professor with the Department of Instrument Science and Engineering, Southeast University. From 2000 to 2003, he was the Director of the Robot Sensor and Control Laboratory, Southeast University. From 2003 to 2004, he was a Visiting Scientist with the Laboratory for Intelligent Mechanical Systems, Northwestern University, Evanston, IL, USA. He is currently a Professor with the School of Instrument Science and Engineering, Southeast University. His current interests include teleoperation, haptic display, the Internet telerobotics, and distributed measurement systems.

CHANGCHENG WU was born in Liancheng, China, in 1987. He received the B.S. degree in measurement and control from Jilin University, Changchun, China, in 2010, and the Ph.D. degree in instrument science and technology from Southeast University, Nanjing, China, in 2016.

He is currently a Lecturer with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics. His current research interests include human–robot interaction and robotic control.

WENLONG LI was born in Henan, China, in 1995. He received the B.S. degree in measurement and control technology from Zhengzhou University, Zhengzhou, China, in 2014. He is currently pursuing the M.S. degree in instrument science and technology with Southeast University, Nanjing, China.

His current interests include brain computer interface and human–robot interaction.



DALIN ZHANG was born in Henan, China, in 1995. He received the B.S. degree in measurement and control technology from Henan University, Kaifeng, China, in 2018. He is currently pursuing the M.S. degree in instrument science and technology with Southeast University, Nanjing, China.

His current interests include brain computer interface and machine learning.

GUOZHENG XU received the B.S. degree in mechatronics from Anhui Polytechnic University, Wuhu, China, in 2002, the M.S. degree in mechatronics from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2005, and the Ph.D. degree in instrumental science and technology from Southeast University, Nanjing, in 2010.

He is currently an Associate Professor with the College of Automation and Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing. His current research interests include rehabilitation robot and human–robot interaction.

HUIJUN LI was born in Pingdingshan, China, in 1976. She received the B.S. degree in instrument science and the M.S. degree in condensed matter physics from Zhengzhou University, Zhengzhou, China, in 1999 and 2002, respectively, and the Ph.D. degree in measurement and control from Southeast University, Nanjing, China, in 2005.

She is currently an Associate Professor with the School of Instrument Science and Engineering, Southeast University. Her current research interests include teleoperation and rehabilitation robotics.

HONG ZENG was born in Langzhong, China, in 1981. He received the M.E. degree in signal and information processing from the Radio Department, Southeast University, in 2006, and the Ph.D. degree from the Department of Computer Science, Hong Kong Baptist University, in 2010.

He is currently an Associate Professor with the School of Instrument Science and Engineering, Southeast University. His current interests include machine learning, pattern recognition, and bio-signal processing.