Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938
Online ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3128

A SPATIOTEMPORAL DUAL-BRANCH NETWORK BASED ON CHANNEL ATTENTION AND RESIDUAL CONNECTION IN MOTOR IMAGERY EEG CLASSIFICATION

HuiXi Mo
School of Computer Science and Artificial Intelligence, Beijing Technology and Business University, Beijing 102488, China.

Abstract: Accurate decoding of motor imagery (MI) electroencephalogram (EEG) signals is a core bottleneck for the practical application of brain-computer interface (BCI) technology. Due to the characteristics of low signal-to-noise ratio, non-stationarity, and multi-dimensional feature coupling of these signals, a single feature extraction method is difficult to fully mine discriminative information. To address this problem, this study proposes a spatiotemporal dual-branch MI-EEG decoding method based on hierarchical channel attention and learnable residual connections. Taking a 3D EEG tensor as input, the method captures scalp electrode topological features and temporal dynamic dependencies through parallel spatial and temporal branches, respectively. It also introduces a channel attention mechanism with learnable residual connections to adaptively enhance the representation ability of key feature channels. Experimental results on the public BCI Competition IV 2A dataset show that the proposed method achieves an average classification accuracy of 79.63% and a Kappa coefficient of 0.7284 in four types of motor imagery tasks, significantly outperforming mainstream models such as EEGNet and FBCNet. Research indicates that the organic combination of the spatiotemporal dual-branch structure and the hierarchical channel attention mechanism can effectively improve the ability to extract spatiotemporal features and channel dependency relationships of EEG signals. The method exhibits excellent classification performance and cross-subject generalization in motor imagery EEG decoding tasks, providing strong support for feature construction and model design in related fields.
Keywords: Motor imagery; Electroencephalogram; Channel attention; Residual connection

1 INTRODUCTION

As an interactive paradigm directly connecting brain neural activities with external devices, the brain-computer interface (BCI) breaks the dependence on peripheral neuromuscular pathways and demonstrates irreplaceable application value in fields such as neural function rehabilitation, auxiliary communication for the disabled, and intention recognition. Motor Imagery (MI), as one of the core research directions of BCI, realizes device control by decoding EEG signals (MI-EEG) induced when individuals imagine limb movements. It has the advantages of non-invasiveness and convenient operation, making it a preferred paradigm for clinical transformation and practical application. However, the inherent low signal-to-noise ratio, deep coupling of time-space-frequency features, and significant individual differences of MI-EEG signals make it extremely difficult to extract effective features, directly restricting the improvement of decoding accuracy, which has become a core challenge to be solved in current MI-EEG research [1].

Existing MI-EEG decoding methods can be divided into two categories: traditional machine learning and deep learning. Traditional methods, represented by Common Spatial Pattern (CSP) [2], rely on manually designed features and manually selected channels, which are difficult to adapt to the dynamic and complex characteristics of EEG signals, resulting in limited generalization ability[3]. Early deep learning models such as EEGNet adopt a single convolutional branch structure Although they realize end-to-end feature learning, they fail to fully separate and model the complementary information of spatiotemporal dimensions [4], leading to insufficient feature representation ability. In recent years, dual-branch networks have become a research hotspot due to the advantage of parallel feature extraction. Li et al. proposed P-3DCNN, which constructs inputs by stacking spatial-spectral feature maps [5], but its manual feature mapping process is prone to temporal information compression and distortion. Mo et al. designed a weight fusion feature recalibration network [6], which learns global and local information through dual branches respectively, verifying the effectiveness of multi-branch structures in modeling dependencies between feature groups, but does not involve decoupled extraction of spatiotemporal dimensions.

Despite the demonstrated advantage of dual-branch structures in feature mining, existing methods still have obvious shortcomings. Firstly, they fail to adaptively distinguish the importance of feature channels, and redundant channels and noise information are likely to interfere with decision-making. Secondly, the fusion of attention mechanisms and backbone networks lacks flexibility, and most adopt fixed weight fusion modes, which may lead to excessive suppression of key features or loss of basic information. Thirdly, some models rely on complex convolution operations, resulting in high computational costs, which is not conducive to deployment on resource-constrained terminal devices. Therefore, this paper proposes a spatiotemporal dual-branch MI-EEG decoding method based on hierarchical channel attention and learnable residual connections. The main contributions include: (1) Constructing a parallel spatiotemporal dual-branch structure to achieve decoupled extraction and complementary fusion of spatiotemporal features; (2)

Designing a hierarchical channel attention mechanism to adaptively enhance key electrode channel features, and proposing a learnable residual connection strategy to balance feature enhancement and original information retention, avoiding the loss of basic information; (3) Building a lightweight backbone network based on depthwise separable convolution splitting modules to balance model performance and computational efficiency.

2 DATASET AND METHODS

2.1 Dataset

This experiment selects the BCI Competition IV 2A dataset as the benchmark dataset for model training and performance evaluation [7]. The dataset contains four types of motor imagery EEG signals from 9 healthy subjects, corresponding to four motor imagery tasks: left hand, right hand, feet, and tongue. The signal acquisition adopts a 25-channel acquisition mode, among which 22 channels are used to record EEG signals and 3 channels are used to record electrooculogram (EOG) signals. The sampling frequency is set to 250Hz, and the spatial distribution of sampling electrodes strictly follows the international 10-20 electrode system. The sample distribution of the dataset is balanced, with each subject completing 288 experimental trials, and 72 trials corresponding to each of the four motor imagery tasks. The motor imagery paradigm used in this experiment is shown in Figure 1. In each experimental trial, a direction arrow lasting about 1.25s is presented at t=2s as a motor imagery cue. Subjects need to continuously perform the specified motor imagery for about 3s from the appearance of the cue. After the trial ends, they enter the next experimental trial after a short rest.

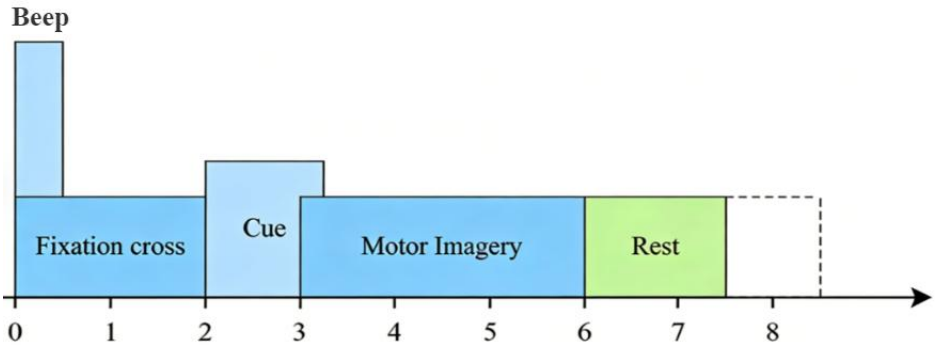

Figure 1 BCI Competition IV 2A Experimental Paradigm

2.2 Data Preprocessing

In the data preprocessing stage, this study sequentially performs baseline drift removal, z-score standardization, and sliding window slicing operations, aiming to stabilize the signal feature distribution, unify data dimensions, and expand the effective sample size, thereby improving data utilization efficiency and model generalization performance.

The baseline drift removal operation is specifically as follows: extract the resting period signal segment 1 second before the event trigger as the baseline reference, and subtract the average potential value of this segment channel by channel to eliminate potential offset artifacts caused by non-physiological factors and achieve zero potential alignment of all channel signals. After completing baseline correction, z-score standardization is performed. In this process, the amplitude of the original signal is first converted from volts (V) to microvolts (μV) to make the signal amplitude range suitable for network input requirements and avoid computational instability under extremely small value ranges. Then z -score standardization is performed to unify the data distribution of each channel. After completing z-score standardization, a sliding window with a window length of 1.5 seconds and a step size of 0.2 seconds is used to continuously slice the EEG signals. Each sliced signal segment retains the category label of its original trial to effectively expand the number of training samples.

To ensure the stability of the model training process and the reliability of performance evaluation, the experimental data are stratified randomly divided into a training set, a validation set, and a test set at the trial level in a ratio of 7:1:2. During the division process, it is strictly ensured that the same original trial and all its derived sliding window segments belong to only one dataset subset, avoiding evaluation bias caused by sample leakage.

2.3 3D EEG Tensor

This paper constructs the preprocessed EEG signals into a 7×7×T 3D tensor as the network input. According to the international 10-20 electrode layout, the multi-channel electrode potentials are mapped to a 7×7 rectangular grid, with zero-value filling for positions without electrodes, and stacked along the time dimension to form a tensor. T is the length of the time series, corresponding to 375-dimensional time features. The 3D EEG tensor not only ensures a regular input structure but also retains scalp spatial topology and signal temporal dynamic features, adapting to the feature extraction logic of deep convolution and attention mechanisms, and providing a reasonable input representation for model construction. Construction method of 3D EEG tensor based on EEG signals is shown in Figure 2.

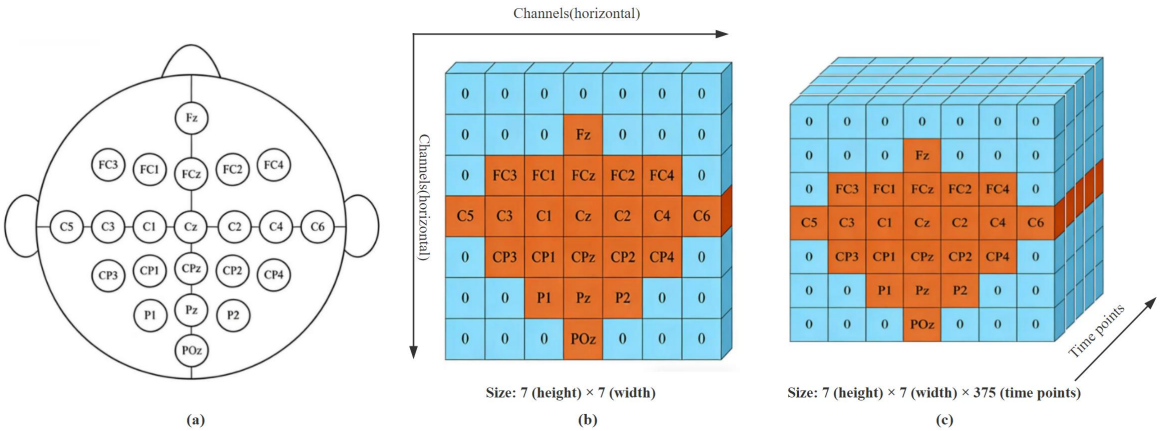

Figure 2 Construction Method of 3D EEG Tensor Based on EEG Signals: (a) EEG Electrode Layout Based on International 10-20 System; (b) 2D Electrode Topological Map; (c) 3D EEG Tensor Obtained by Stacking 2D Electrode Topological Maps along the Time Dimension

2.4 Network Structure

The core design of this network is a parallel spatiotemporal dual-branch architecture combined with a channel attention mechanism and learnable residual connections. The network takes a 375×7×7 dimensional EEG tensor as input, models the spatial topological associations of scalp electrodes and the multi-scale temporal dynamic features of EEG signals through two branches (spatial and temporal) respectively. Both branches embed a hierarchical channel attention mechanism with learnable residual connections, and finally fuse the features extracted by the dual branches to complete the classification and discrimination of four types of motor imagery tasks. The entire network uses depthwise separable convolution as the basic convolution unit, which is split into two independent modules: depthwise convolution and pointwise convolution. All convolution modules are equipped with batch normalization and LeakyReLU activation functions, and Dropout layers are added to suppress overfitting and enhance the robustness and nonlinear representation ability of feature extraction. Spatiotemporal dual-branch network structure diagram with 3D EEG tensor as input is shown in Figure 3.

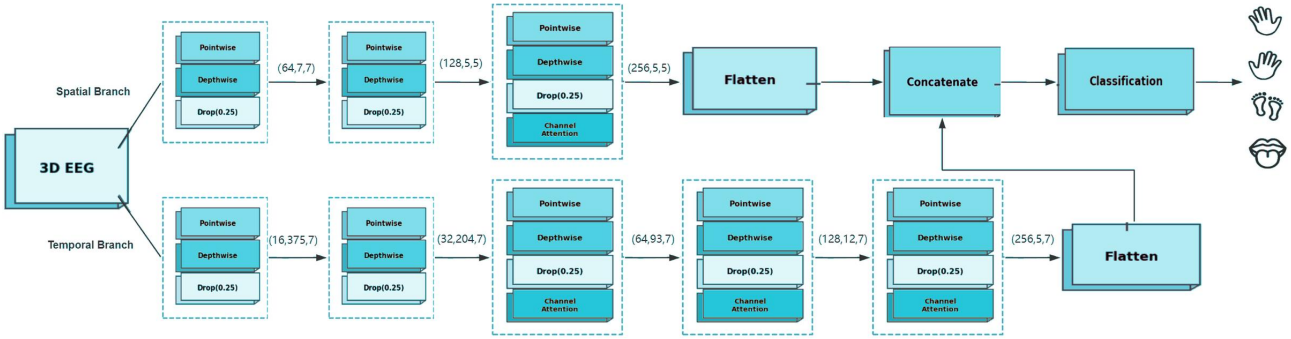

Figure 3 Spatiotemporal Dual-Branch Network Structure Diagram with 3D EEG Tensor as Input

(1) Channel Attention

The channel attention mechanism is designed with hierarchical intensity, which adapts to differentiated attention intensity coefficients according to the channel dimensions of different depths of the network. Smaller attention intensity is adopted for shallow features, and larger attention intensity is adopted for deep features. When the number of channels is less than or equal to 32, the coefficient is 0.3; when the number of channels is greater than 32 and less than or equal to 64, the coefficient is 0.5; when the number of channels is greater than 64 and less than or equal to 128, the coefficient is 0.8; when the number of channels is 256, the coefficient is 1; thereby accurately enhancing key channel features at different levels. The mechanism extracts global channel information through two paths: global average pooling and global max pooling. After learning channel weights through two shared fully connected layers, it outputs weight values in the range of 0 to 1, thereby realizing adaptive weighting of channel features. At the same time, a learnable residual connection is paired with the channel attention, and a learnable weight parameter with an initial value of 0.7 is set. Through this parameter, the fusion ratio of attention-enhanced features and original features is balanced, which strengthens effective features while avoiding the loss of basic information, allowing the model to independently learn the optimal feature fusion method. Schematic diagram of channel attention with learnable residual connection is shown in Figure 4.

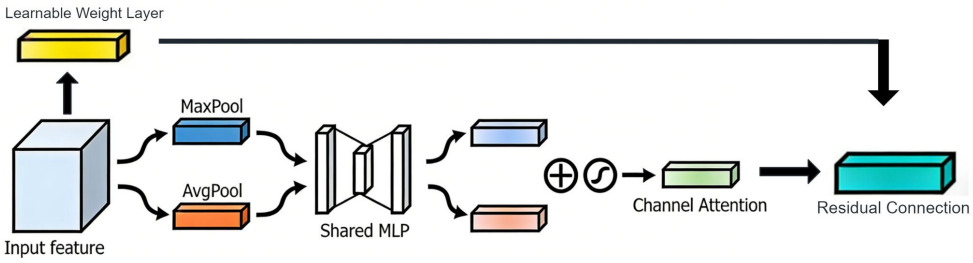

Figure 4 Schematic Diagram of Channel Attention with Learnable Residual Connection

(2) Spatial Branch

The spatial branch of the network focuses on extracting the spatial topological features of scalp electrodes. The input EEG tensor first compresses the number of channels from 375 to 64 through pointwise convolution to achieve feature dimensionality reduction and redundant information suppression. Then, it extracts long-range spatial associations across vertical electrodes through 1×5 depthwise convolution. After adding a Dropout layer, it expands the number of channels to 128 through pointwise convolution to improve feature representation ability. Then, it refines the cooperative activity features of electrodes in local brain regions through 3×3 depthwise convolution and continues to add a Dropout layer. After that, the number of channels is further expanded to 256 through pointwise convolution, paired with 5×1 depthwise convolution to extract the spatial distribution features of horizontal electrodes. After adding a Dropout layer, the full-intensity channel attention mechanism is applied in the deep dimension of 256 channels, and feature fusion is completed by combining with learnable residual connections. Finally, the processed feature map is flattened into a one-dimensional vector to prepare for subsequent feature fusion.

(3) Temporal Branch

The temporal branch focuses on capturing the multi-scale temporal dynamic features of EEG signals. It first performs dimension permutation on the input tensor to allow convolution operations to slide along the time axis, and then integrates spatial dimension features through pointwise convolution and outputs 16-channel features. On this basis, multi-scale temporal features are extracted step by step through a stacked structure of multi-scale depthwise convolution and pointwise convolution. First, 50×1 depthwise convolution is used to model global temporal dynamics under a wide receptive field. After adding a Dropout layer, the number of channels is expanded to 32 through pointwise convolution. Then, 172×1, 112×1, 82×1, and 8×1 depthwise convolutions are sequentially paired to gradually compress the time resolution and capture cross-scale long-short-range temporal dependencies from 32 milliseconds to 688 milliseconds. A Dropout layer is added after each convolution step, and the number of channels is gradually expanded to 64, 128, and 256 through pointwise convolution. The channel attention mechanism sequentially applies medium to full-intensity weighting in the 64, 128, and 256 channel dimensions. Each layer combines with learnable residual connections to balance the feature fusion ratio, and finally flattens the temporal feature map into a one-dimensional vector.

Finally, the one-dimensional feature vectors output by the spatial branch and the temporal branch are concatenated along the feature dimension to form a unified spatiotemporal joint representation with both spatial topology and temporal dynamics. The concatenated features are input into the fully connected classification head, and sequentially pass through 512-dimensional and 64-dimensional fully connected layers. Both fully connected layers are equipped with batch normalization and ReLU activation functions to realize further dimensionality reduction and nonlinear transformation of features. Finally, the discriminative mapping of four types of motor imagery tasks is completed through a 4-dimensional fully connected layer, and the classification result is output.

3 EXPERIMENTS AND RESULTS

3.1 Experimental Settings and Evaluation Metrics

The motor imagery EEG signal decoding method proposed in this paper uses the Adam optimizer for parameter update during the training process, and selects multi-class cross-entropy with label smoothing as the loss function. The batch size and initial learning rate are 64 and 0.01 respectively, and the maximum number of training epochs is 100. To enhance convergence stability and alleviate overfitting, the ReduceLROnPlateau learning rate scheduling strategy is introduced. When the validation set loss does not decrease for 8 consecutive epochs, the learning rate is adaptively reduced by a factor of 0.6. At the same time, gradient clipping is performed with a maximum norm of 1.0 to prevent gradient explosion.

To comprehensively evaluate model performance, this paper adopts a differentiated evaluation index system according to different experimental purposes. In comparative experiments, average accuracy (Avg. Accuracy) and average Kappa coefficient (Avg. Kappa) are used as core indicators to verify the effectiveness of the channel attention mechanism. In ablation experiments, in addition to accuracy and Kappa coefficient, accuracy standard deviation (Std) is introduced as a key indicator to measure the cross-subject stability of the model.

Classification accuracy is used to measure the proportion of correct discrimination of the model on all samples. Its value range is [0,1], and the value closer to 1 indicates better overall classification performance. The calculation formula is:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

where TP represents the number of positive samples correctly identified as positive, TN represents the number of negative samples correctly identified as negative, FP represents the number of negative samples incorrectly identified as positive, and FN represents the number of positive samples incorrectly identified as negative.

The Kappa coefficient is used to quantify the consistency between model predictions and true labels. Its value range is [-1,1]. A value closer to 1 indicates higher consistency, a value around 0 indicates near-random level, and a negative value indicates lower than random consistency. The calculation formula is:

$$\text{Kappa}=\frac{P_o-P_e}{1-P_e} \tag{2}$$

where P_o represents the experimental accuracy, and P_e represents the random guess accuracy.

The accuracy standard deviation is used to evaluate the performance fluctuation of the model among different subjects. The calculation formula is:

$$\text{Std}=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\text{Acc}_i-\overline{\text{Acc}})^2} \tag{3}$$

where N is the total number of subjects, Acc_i is the classification accuracy of the i-th subject, and $\overline{\text{Acc}}$ is the average classification accuracy of all subjects. A smaller standard deviation indicates better cross-subject stability of the model.

3.2 Model Test Experiment

To intuitively present the classification performance and confusion pattern of the proposed method, Figure 5 shows the confusion matrix heatmap of 9 subjects. The number of samples is displayed above each cell, the recall rate is displayed below, and the color depth represents the sample size. Experimental results show that the proposed method performs excellently on most subjects: Subjects A03 and A07 with the highest accuracy have significant diagonal features. The recall rate of the right hand of A03 is 97.5%, and the recall rates of the left hand and tongue of A07 reach 93.3% and 94.2% respectively, verifying the effective feature extraction ability of the model. Subjects A02 and A05 with weaker performance show specific confusion patterns: A02 has severe cross-confusion between the right hand and the tongue, with a recall rate of the right hand of only 50.8%; the recall rates of the four types of tasks of A05 fluctuate between 51.7% and 64.2%. This phenomenon is consistent with neurophysiological mechanisms: the cortical areas of the hand and tongue are adjacent, the features of the left and right hands overlap, and the signal-to-noise ratio of the tongue task is low. Overall, the accuracy of all subjects is significantly higher than the 25% random level, proving that the model has basic discriminative ability.

From the perspective of misclassification patterns, it is mainly manifested as cross-confusion between the right hand and the tongue, and between the left hand and the right hand, reflecting the spatial proximity of cortical activation areas of different tasks. Overall, the confusion matrix intuitively shows the individual differences and confusion bottlenecks of model performance, verifies the effectiveness and stability of the proposed method in multi-class motor imagery EEG decoding tasks, and provides an analysis basis for subsequent ablation experiments. Confusion matrices of 9 subjects using the proposed method on the BCI Competition IV 2A dataset are shown in Figure 5.

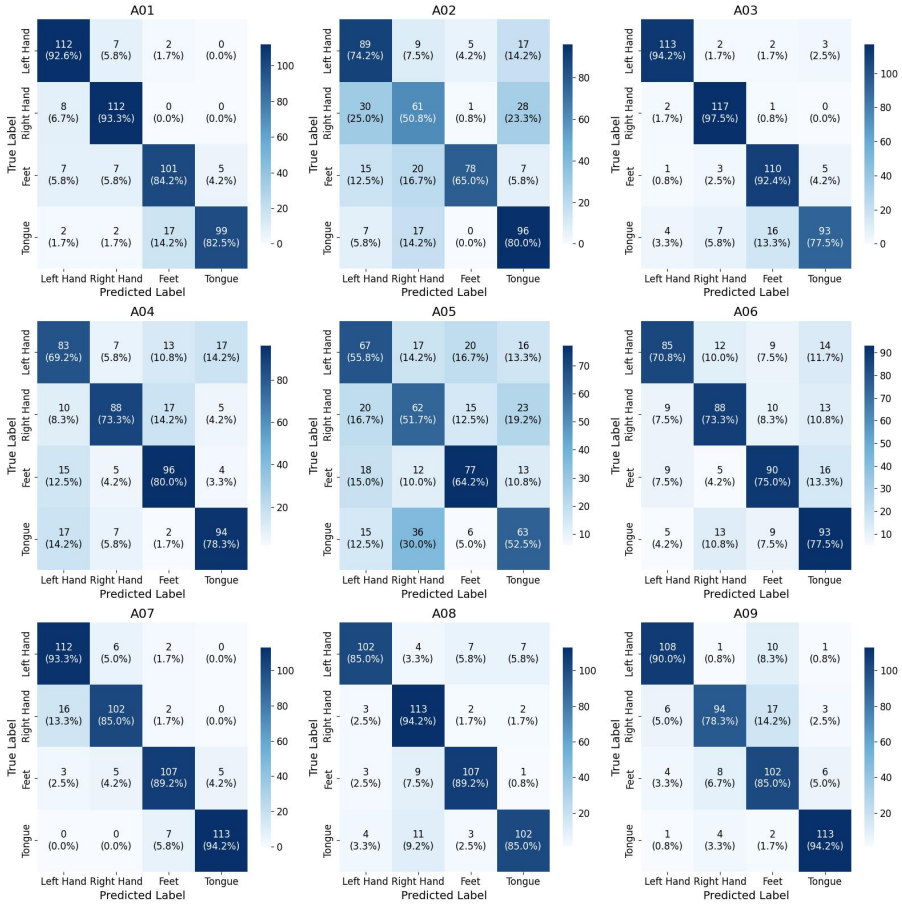

Figure 5 Confusion Matrices of 9 Subjects Using the Proposed Method on the BCI Competition IV 2A Dataset

3.3 Model Ablation Experiment

On the four-class motor imagery tasks of the BCI Competition IV 2A dataset, the classification accuracy, Kappa coefficient, and accuracy standard deviation of the proposed complete model and branch ablation are compared as shown in Table 1. Among them, Avg. Accuracy represents the average accuracy, Avg. Kappa represents the average Kappa coefficient, and Std represents the accuracy standard deviation, which is used to measure the stability of the model among different subjects. The temporal branch is an ablation model that retains the temporal feature extraction branch, the spatial branch is an ablation model that retains the spatial feature extraction branch, and the spatiotemporal dual-branch is the complete model proposed in this study.

The ablation experiment completes the comparative verification on the four types of motor imagery tasks of the BCI Competition IV 2A dataset. The experimental results show the significant advantages of complementary fusion of spatiotemporal dual branches and the differentiated values of single branches: the dual-branch fusion model achieves an average classification accuracy of 79.63% and an average Kappa coefficient of 0.7284, which are 0.60% and 0.0080 higher than those of the temporal branch, and 4.35% and 0.0580 higher than those of the spatial branch respectively. This fully verifies the effectiveness of decoupled feature extraction by spatiotemporal dual branches. A single branch can only capture single-dimensional features of EEG signals, while the dual-branch structure realizes complementary fusion of feature dimensions by parallel modeling of spatial topological associations of scalp electrodes and temporal dynamic dependencies of EEG signals, effectively improving the overall discriminative ability of the model. From the perspective of single-branch characteristics, the temporal branch becomes the model with the highest accuracy among single branches with an average accuracy of 79.03%, indicating that the temporal dynamic features of motor imagery EEG signals are the core discriminative information for decoding tasks, which is consistent with the physiological mechanism of MI-EEG. However, the accuracy standard deviation of this branch reaches 0.1107, and the cross-subject generalization stability is poor. The reason is that there are significant individual differences in the EEG temporal rhythms of different subjects, and a single temporal feature is difficult to adapt to all subjects. Although the average accuracy of the spatial branch is only 75.28%, its accuracy standard deviation is only 0.0956, which is the most stable among the three groups of models. It proves that the spatial topological features of scalp electrodes have stronger cross-subject consistency and can effectively make up for the insufficient stability of the temporal branch. At the same time, from the perspective of individual subjects, the dual-branch model has a particularly significant improvement effect on subjects with poor performance. For example, on subject A02 with low accuracy, the accuracy of the dual-branch model reaches 67.50%, which is 10.83% and 9.37% higher than that of the temporal branch and the spatial branch respectively. This indicates that the dual-branch fusion architecture can effectively alleviate the performance

degradation of the single temporal branch on subjects with low signal-to-noise ratio and high individual differences through the supplement of spatial features. This ablation experiment shows that the temporal branch provides core classification accuracy support for the model, and the spatial branch provides cross-subject stability guarantee for the model. After combining with the hierarchical channel attention mechanism, it not only retains the high-precision advantage of the temporal branch but also absorbs the strong stability characteristics of the spatial branch. The complementarity of the two makes the model achieve the optimal performance.

Table 1 Comparison of Classification Accuracy (%), Kappa Coefficient, and Accuracy Standard Deviation of the Proposed Method and Branch Ablation

Model	Subjects	Subjects	Subjects	Subjects	Subjects	Subjects	Subjects	Subjects	Subjects	Avg. Accuracy	Avg. Kappa	Std
	A01	A02	A03	A04	A05	A06	A07	A08	A09			
Spatial Branch	75.21	58.13	82.08	80.21	57.71	71.46	86.67	88.96	77.08	75.28	0.6704	0.0956
Temporal Branch	86.67	56.67	91.46	85.00	60.83	73.33	86.88	87.92	82.50	79.03	0.7204	0.1107
Spatiotemporal Dual-Branch (Proposed Method)	88.12	67.50	90.00	75.21	56.04	74.17	90.42	88.33	86.88	79.63	0.7284	0.1084

3.4 Model Comparison Experiment

To systematically evaluate the classification performance of the proposed method, this paper selects classic models and recently proposed models as controls and conducts comparisons on the public dataset BCI Competition IV 2A. The results are shown in Table 2, where Avg. Accuracy represents the average accuracy and Avg. Kappa represents the average Kappa coefficient.

It can be seen from Table 2 that the proposed method achieves an average accuracy of 79.63% and an average Kappa coefficient of 0.7284 in the four-class tasks of BCI Competition IV 2A, which is overall better than the comparison models. Compared with the classic baseline model EEGNet, the average accuracy of the proposed method is increased by 5.09%, and the average Kappa coefficient is increased by 0.0650. Compared with the recently proposed FBCNet, the average accuracy is increased by 3.43%, and the average Kappa coefficient is increased by 0.0458. Compared with MBCNN-TCN-Net, the average accuracy is increased by 4.55%, and the average Kappa coefficient is increased by 0.0609. The main reason why the performance of the proposed method is significantly better than that of the comparison models is that the hierarchical channel attention mechanism effectively suppresses the channel redundant noise in the EEG signals of these subjects and strengthens the key electrode channel features related to motor imagery. At the same time, the lightweight structure of depthwise separable convolution reduces the risk of model overfitting and improves the classification performance on low signal-to-noise ratio samples. The above results verify the effectiveness of the proposed method and indicate that it is feasible and advantageous to take the 3D EEG temporal tensor as input and jointly model its temporal and spatial features.

Table 2 Comparison of Classification Accuracy (%) and Kappa Coefficient of Different Classification Models on the BCI Competition IV 2A Dataset

Model	Subjects	Subjects	Subjects	Subjects	Subjects	Subjects	Subjects	Subjects	Subjects	Avg. Accuracy	Avg. Kappa
	A01	A02	A03	A04	A05	A06	A07	A08	A09		
EEGNet[4]	88.27	60.24	86.73	70.31	56.24	55.29	86.74	83.27	83.77	74.54	0.6634
FBCNet [8]	85.42	60.42	90.63	76.39	74.31	53.82	84.38	79.51	80.90	76.20	0.6826
MBCNN-TCN-Net [9-10]	82.56	67.14	89.01	64.91	74.28	59.07	86.64	76.75	75.38	75.08	0.6675
Proposed Method	88.12	67.50	90.00	75.21	56.04	74.17	90.42	88.33	86.88	79.63	0.7284

4 CONCLUSIONS

Aiming at the problem that relying on manual features or a single type of feature extraction is difficult to fully mine the complex spatiotemporal structure of EEG signals in motor imagery EEG decoding, this paper proposes a spatiotemporal dual-branch MI-EEG decoding method based on hierarchical channel attention and learnable residual connections. The

method retains the original spatiotemporal structure of EEG signals by constructing a 3D temporal tensor, introduces a hierarchical intensity channel attention module with learnable residual connections to adaptively enhance discriminative channel responses, and designs a spatiotemporal dual-branch network to extract temporal dynamic features and spatial topological features respectively, realizing effective fusion of multi-dimensional information.

Experimental results show that the method achieves an average accuracy of 79.63% and a Kappa coefficient of 0.7284 in the four-class tasks of the BCI Competition IV 2A dataset, which is better than the compared classic models and recently proposed methods. Compared with the single temporal branch model, the complete dual-branch model has an average accuracy increase of 0.60%, a Kappa coefficient increase of 0.008, and an accuracy standard deviation decrease of 0.0023, achieving dual improvements in classification performance and cross-subject stability. Compared with the single spatial branch model, the average accuracy is increased by 4.35%, and the Kappa coefficient is increased by 0.058, verifying the dominant role of the temporal branch in classification performance, the improvement effect of the spatial branch on generalization, and the core improvement effect of the hierarchical intensity channel attention mechanism on model performance.

Overall, the proposed method achieves good discriminative performance without complex manual features, verifies the effectiveness of the spatiotemporal dual-branch fusion strategy guided by channel attention, provides a new idea for the design of lightweight models for motor imagery EEG decoding, and the proposed hierarchical intensity channel attention and learnable residual connection strategies can also provide reference for other EEG signal processing tasks, with certain versatility.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Wang X, Liesaputra V, Liu Z, et al. An in-depth survey on Deep Learning-based Motor Imagery Electroencephalogram (EEG) classification. Artificial Intelligence in Medicine, 2024, 147: 102738.

[2] Tibrewal N, Leeuwis N, Alimardani M. Classification of motor imagery EEG using deep learning increases performance in inefficient BCI users. PLOS ONE, 2022, 17(7): e0268880.

[3] Altaheri H, Muhammad G, Alsulaiman M, et al. Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: a review. Neural Computing and Applications, 2023, 35(20): 14681–14722.

[4] Lawhern V J, Solon A J, Waytowich N R, et al. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. Journal of neural engineering, 2018, 15(5): 056013.

[5] Li X, Chu Y, Wu X. 3D convolutional neural network based on spatial-spectral feature pictures learning for decoding motor imagery EEG signal. Frontiers in Neurorobotics, 2024, 18: 1485640.

[6] Mo Y, Li Y, Zhang B X, et al. Motor imagery EEG classification based on weight fusion feature recalibration network. Journal of Electronic Measurement and Instrumentation, 2025, 39(1): 70-79.

[7] Brunner C, Leeb R, Müller-Putz G, et al. BCI Competition 2008–Graz data set A. Institute for knowledge discovery (laboratory of brain-computer interfaces), Graz University of Technology, 2008, 16(1-6): 1.

[8] Mane R, Chew E, Chua K, et al. FBCNet: A Multi-view Convolutional Neural Network for Brain-Computer Interface. 2021.

[9] Yu S, Wang Z, Wang F, et al. Multiclass classification of motor imagery tasks based on multi-branch convolutional neural network and temporal convolutional network model. Cerebral Cortex, 2024, 34(2): bhad511.

[10] Yu S, Wang Z, Wang F, et al. Multiclass motor imagery classification based on multi-branch CNN and TCN for BCI. Neurocomputing, 2023, 543: 126-138.