

Learning Local-to-Global Spatial-Temporal Representation for Motor Imagery Classification

Shaozhe Liu, Leike An, Xinliang Zhou, Xiaojun Ning, and Ziyu Jia*

Abstract—Motor imagery (MI) is widely used in brain-computer interface (BCI) research as a classical paradigm. Significant progress has been made in applying deep learning methods to MI classification based on electroencephalogram (EEG) signals. However, previous studies have only focused on extracting spatial-temporal features from EEG signals. Neuropsychological research indicates that distinct brain regions contribute variably to specific behaviors, highlighting the significance of their collaborative activity. Additionally, the connectivity and activation patterns of the brain may vary across different time segments, even during the execution of a similar task. Furthermore, the variability among individuals adds to the complexity of the classification process. To tackle these challenges, we propose an innovative model named STL2G-DG, comprised of a spatial-temporal local feature extraction module (STLE), a local-to-global feature extraction module (L2G), and a domain generalization module (DG). Specifically, STLE can extract spatial and temporal local features from EEG signals, whereas L2G merges and extracts local spatial-temporal features to generate a local-to-global representation for MI classification. Additionally, DG is implemented to learn invariant representations across subjects, thereby improving the generalization capability of our approach. We compare our method with state-of-the-art approaches, including DeepConvNet, EEGNet, MMCNN, and STL-DG. The results indicate that in the experimental setup of cross-subject validation, STL2G-DG outperforms these methods.

Index Terms—Electroencephalography (EEG), Local to global, Motor imagery, Spatial-temporal representation.

I. INTRODUCTION

RESEARCH focused on motor imagery (MI) classification tasks provides valuable insights into the relationship between human cognition and motor control systems. By studying MI classification tasks, we can better understand the similarities and differences in brain activity during the imagination and execution of movements, thus uncovering the underlying mechanisms behind human cognition and behavior. This knowledge can be applied to improve movement rehabilitation training and the development of brain-machine interfaces (BCIs) by decoding brain movement signals [1], [2]. Electroencephalogram (EEG) signals, a non-invasive technique

Shaozhe Liu is with Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

Leike An is with China Mobile Information Technology Center, Beijing, China.

Xinliang Zhou is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore.

Xiaojun Ning is with the School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China.

Ziyu Jia is with Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

Corresponding author: Ziyu Jia.

The PyTorch implementation of this paper is available at <https://github.com/shaozheliu/STL2G-DG>.

with high temporal resolution, are commonly used as primary input signals in BCIs due to their ability to accurately capture millisecond-level brain activity.

Accurate extraction of EEG signals features plays a crucial role in MI classification tasks. Traditional approaches such as common spatial pattern (CSP) and filter bank common spatial pattern (FBCSP) [3]–[6] rely on prior knowledge to build feature extractors with robust machine learning classification models like support vector machine (SVM), multilayer perceptron (MLP), k-nearest neighbor (KNN), fall short in capturing the high-dimensional and complex patterns present in EEG signals [7]–[9]. In recent years, significant progress has been made in MI classification, with deep-learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), emerging as highly effective methodologies in BCI applications [10]–[12]. Through their proficiency in representation learning, CNNs and RNNs possess the ability to autonomously extract intricate and discriminative characteristics from raw EEG signals. This enables these models to demonstrate improved performance and robustness in real-life MI classification scenarios. Despite the plethora of methods proposed in recent years, the field of MI classification still faces significant challenges that demand thorough investigation and resolution.

Firstly, it is essential to acknowledge that cognition in the brain is formed through the transmission and integration of information between neural circuits. Different brain regions have distinct roles in this process, including perception, memory, decision-making, and more [13]–[15]. Communication and information integration occurs through neurotransmitters and synaptic connections among these regions. Researchers have discovered that localized cortical areas form a computational space with hierarchical structures. Despite their hierarchies, these regions can concurrently and distributively process specific information, such as perception, movement, and memory [16]. These regions are interconnected by excitatory neurons and axons with broadly distributed long-range connections. Through selective connectivity, they can regulate or inhibit the output of specific local region, contributing to overall information processing. This distributed population of neurons progressively receives and processes information in a bottom-up, layered manner. Importantly, the connections among brain regions are dynamic in both spatial and temporal dimensions, with their strength changing over time [17]. These understandings lead us to consider not only the transmission of information between different brain regions during feature extraction but also the distribution of connection states across different periods. Therefore, our challenge lies in designing

a logical network architecture capable of learning spatial-temporal, local-to-global representation from EEG signals.

Secondly, deep learning methods are suitable for end-to-end learning without prior knowledge of the target problem and demonstrate strong generalization on unseen datasets. However, when applied to EEG signals analysis, these methods encounter challenges in capturing local-to-global patterns across different subjects, thereby hindering the improvement of generalization performance. There are several reasons for this phenomenon. Firstly, EEG signals are often characterized by small sample sizes and high dimensionality. Due to the limited samples available, it is difficult to learn the features representing local-to-global patterns across subjects [18], [19]. Secondly, EEG signals exhibit significant intersubject variability. A subject-independent classifier trained directly on EEG signals from multiple subjects often lacks sufficient generalization capability when applied to a new subject. Various factors contribute to intersubject variability, such as individual differences in brain structure influenced by specific genetic factors and environmental effects, variations in data collection resulting from different skull shapes, and the influence of the subjects' mental state. **The second challenge is to learn local-to-global features across subjects by considering the characteristics of the EEG signals..**

This paper presents a novel methodology for MI classification that effectively tackles the aforementioned challenges. The proposed approach consists of three parts: the Spatial-Temporal Local feature extraction module (STLE), the Spatial-Temporal L2G feature extraction module (L2G), and the Domain Generalization module (DG). The STLE and L2G utilize a multi-level extraction approach to learn spatial-temporal local to global representation. Through this approach, the model can simulate the process of aggregating knowledge from individual neurons to local brain regions and, eventually, to the global level. The DG module is specifically designed to tackle the issue of cross-subject variability commonly encountered in EEG analysis. By incorporating a domain generalization method, we learn the subject-invariant representation of EEG signals. The primary contributions of this research can be summarized as follows:

- To the best of our knowledge, this is the first attempt to integrate Local-to-global Spatial-Temporal attention and Domain Generalization techniques within a unified framework for enhancing MI classification.
- We propose the STL2G module to simulate the integration process of information from local to global among neural circuits and learn spatial-temporal representation.
- In our study, we design the technique of Domain Generalization to learn subject-invariant representation, thereby enhancing the generalization capabilities of the proposed model.
- The proposed STL2G-DG achieves the state-of-the-art performance on BCI-2A and OpenBMI datasets in cross-subject validation.

Compared to the Spatial-Temporal Transformer based on Domain Generalization (ST-DG) published in our preliminary work [20], STL2G-DG has the following improvements: 1)

A brain network spanning from local to global scales, incorporating both spatial and temporal dimensions, is established. This network offers comprehensive spatiotemporal topological insights for the MI classification. 2) A spatial and temporal fusion module is incorporated to effectively combine spatial and temporal features. 3) A t-SNE projection is utilized on the extracted feature maps to assess the efficacy of the local-to-global feature extraction. Additionally, ablation experiments are conducted to evaluate the impact of each component of STL2G-DG on performance.

II. RELATED WORK

A. Motor Imagery Classification

In MI tasks, the extraction of spatial-temporal features is a crucial step. Deep learning methods have been widely applied in MI tasks and can effectively learn abstract representations of spatial-temporal features from raw signals. CNN-based approaches [21]–[23], utilize convolutional and pooling layers to enable the extraction of spatially invariant and localized features. A representative classic network, EEGNet [24], employs deep and separable convolutions to construct a framework for spatial-temporal feature extraction from EEG signals. Schirrmeister [25] introduced DeepConvNet and ShallowConvNet as deep and shallow convolutional neural networks, respectively, for processing EEG signals. SalientSleepNet [26], based on U2-Net, utilizes a time-convolutional network to learn multi-scale temporal representation. RNN-based methods [27], [28], are proficient in capturing the temporal information inherent in EEG trials. Some methods [29]–[31], employ LSTM networks to extract spatial-temporal features separately. In recent years, researchers have started to focus on extracting local features within each electrode neighborhood and constructing methods for local-to-global feature representation. This approach has the advantage of capturing the interaction information between electrodes. Song [32] designed dynamic graph convolutional neural networks for EEG emotion recognition with a trainable adjacency matrix. Zhong [33] defined the adjacency matrix based on spatial distance and added global connections based on asymmetry in neuronal activities. However, the current methods initiate a local-to-global framework from a spatial perspective. Conducting urgent research on feature extraction methods is necessary, focusing on extracting local to global features from a spatial-temporal perspective.

B. Domain Generalization

Transfer learning-based approaches have become prominent solutions for enhancing the generalization capability of models. By utilizing pre-trained models or pre-processing techniques, transfer learning allows researchers to transfer knowledge between different subjects and tasks, thereby improving accuracy and robustness in MI classification. This approach is expected to remain a significant area of research in the field of BCI.

Domain Adaptation (DA) is a transfer learning framework that utilizes multi-domain source data, to either adapt the classifier from the source domain to the target domain or extract

common features that are robust for different domains [34], [35]. Sakhavi and Guan [36] trained a CNN model on the FBCSP features from multiple subjects and then transferred the model parameters by fine-tuning on the new subject's data. Raza et al. [37] presented a covariate shift-detection (CSD) method and retrain the classifier once the covariate shifts are detected. However, the need for target domain data in DA approaches presents a challenge in the context of MI classification, as obtaining data from new subjects is time-consuming and expensive.

Conversely, Domain Generalization (DG) aims to create a model capable of generalizing to unseen data from multiple domains, without relying on specific information from a target domain [38], [39]. This makes DG a suitable choice for addressing challenges related to cross-subject and cross-label variations in MI classification, where a model should perform well across numerous subjects and extract invariant features. Through exclusive reliance on annotated data from heterogeneous subjects, DG approaches facilitate improved classification accuracy and a more robust framework for EEG feature representation. Therefore, DG-based methods have promising applications for addressing challenges related to cross-subject variability in MI classification and can make valuable contributions to the field of BCI research.

III. PRELIMINARIES

Given an EEG signals instance $x \in \mathbb{R}^{T \times C}$ with T sample points and C EEG channels, MI classification task aims to learn a mapping function $\mathcal{F}: x^{T \times C} \mapsto y$ which maps EEG signals x into the corresponding imagery class y , where $y = \{y_1, y_2, \dots, y_L\}$ and L is the number of MI class.

Definition 1 (Domain and Domain Label). We denote \mathcal{X} as a nonempty input space, \mathcal{Y} as label output space, and \mathcal{D} as domain output space. Domain i consists of EEG signals that are sampled from the distribution of subject i . Consequently, we assign the same domain label to samples belonging to the same domain. We represent it as $S^i = \{(x_i, y_i, d_i)\}_{i=1}^n \sim P_{XYD}^i$, where $x_i \in \mathcal{X}$, $y_i \in \mathcal{Y}$, $d_i \in \mathcal{D}$, and P_{XYD}^i denotes the joint distribution of the input sample, output ground truth label and domain label.

Definition 2 (Domain generalization). As shown in Fig. 1, assume that our dataset contains EEG signals from M subjects. Then, in domain generalization, we are given $M - 1$ training (source) domains $S_{train} = \{S^i | i = 1, \dots, M - 1\}$. The left is to test (target) domain $S_{test} = \{S^M | i = M\}$. The joint distribution between each pair of domains is different: $P_{XYD}^i \neq P_{XYD}^j$, $1 \leq i \neq j \leq M - 1$. The goal of domain generalization is to help the classifier to learn a robust and generalizable mapping function \mathcal{F} from the $M - 1$ training domains to achieve a minimum classification error on an unseen test domain S_{test} (i.e., S_{test} cannot be accessed in training procedure and $P_{XYD}^{test} \neq P_{XYD}^i$ for $i \in \{1, \dots, M - 1\}$)

$$\min_h \mathbb{E}_{(x,y) \in \mathcal{S}_{test}} [\ell(h(x), y)] \quad (1)$$

where \mathbb{E} is the expectation, h is the classifier function, and ℓ is the loss function.

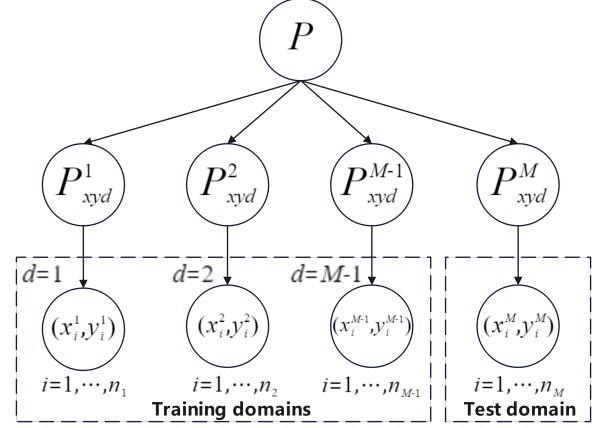


Fig. 1. Illustration of cross-subject multi-source domain generalization. Adapt from [40]. In the training domain, data sampled from the distribution of subject i constitutes domain i . There are a total of $M-1$ subjects forming $M-1$ domains. Domain generalization aims to enhance the model's generalization performance in the test domain by learning domain-invariant features representation from different domains.

IV. METHODOLOGY

The diagram depicting the overall architecture of the STL2G-DG model can be found in Fig. 2. The process of feature learning in this model is outlined as follows: 1) We define the Spatial-Temporal Local Feature Extraction module (STLE), which mimics the information integration mechanism within individual cerebral regions during MI tasks and captures crucial local temporal features within the MI task cycle. 2) We introduce the Spatial-Temporal Local-to-Global Feature Extraction module (L2G) to merge local temporal features with local spatial features, thereby replicating the exchange of information between cerebral regions during the MI process and enhancing it through feature fusion. 3) We develop the Domain Generalization module (DG) to extract domain-invariant features across subjects.

A. Spatial-Temporal Local Feature Extraction Module

Local spatial-temporal features play a critical role in MI classification. Existing approaches indirectly derive local spatial features by transforming raw signals into time-frequency images, which may lead to information loss and necessitate prior information. To capture the spatial-temporal local features within EEG sequences, we introduce a Spatial-Temporal Local Feature Extraction module (STLE), as illustrated in Fig. 3.

1. In order to efficiently extract local features, it is crucial to first segment regions in both temporal and spatial dimensions.

$$SR_i, TR_j = \text{Split}(X) \quad (2)$$

$$SR_i = [x_{.1}, x_{.2}, \dots, x_{.n_i}] \quad (3)$$

$$TR_j = [x_{1.}, x_{2.}, \dots, x_{m_j.}] \quad (4)$$

where Split refers to the operation of partitioning the spatial and temporal dimensions, and i represents the

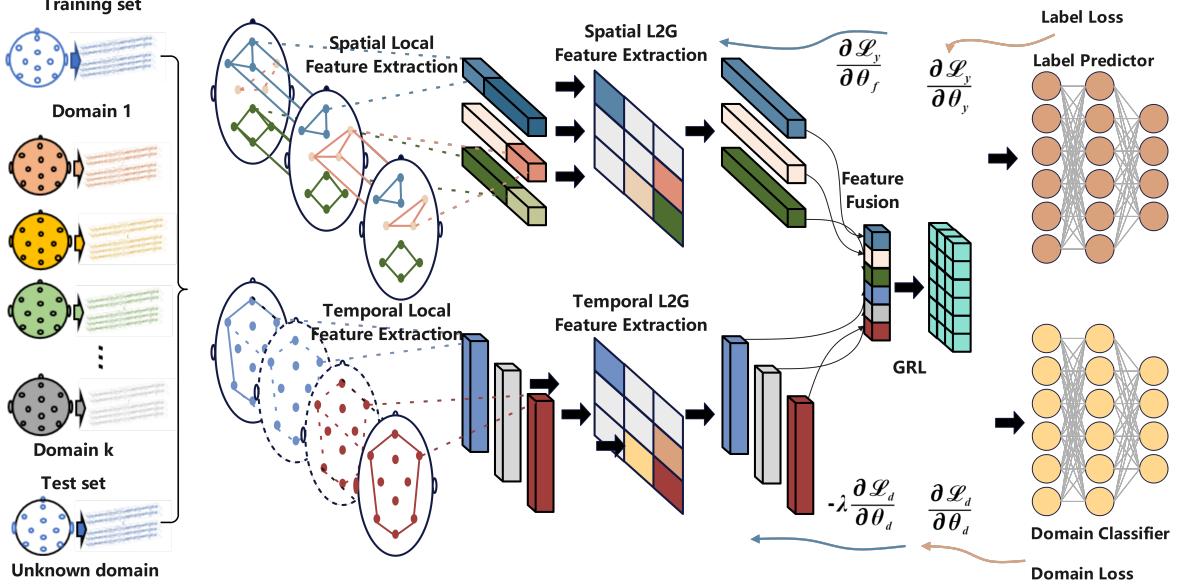


Fig. 2. Overview of our methods. Firstly, we partition the EEG signals into spatial and temporal regions. In each region, we use multi-layer convolution for local feature learning. Moreover, the attention mechanism captures activation features at varying brain regions and time intervals, transitioning from local to global representation. Finally, we integrate the obtained discriminative features with domain generalization techniques and feed them into the classifier, thereby completing the end-to-end network architecture.

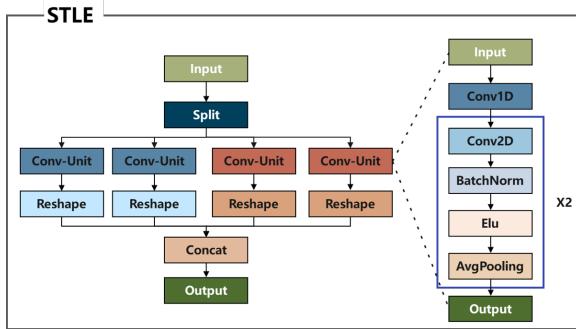


Fig. 3. The Spatial-Temporal Local Feature Extraction module (STLE) is designed to learn local spatial and temporal feature representation. The initial sequence is partitioned into various patches. Each spatiotemporal patch undergoes dimension reduction through Conv1D processing, followed by spatiotemporal local feature extraction using stacked convolution operations.

spatial regions, j signifies the temporal regions. n_i denotes the number of EEG channels within spatial region i , and m_j indicates the number of EEG sampling points within temporal region j . $SR_i \in \mathbb{R}^{T \times n_i}$ denotes the spatial patch of the sequence X , and $TR_j \in \mathbb{R}^{m_j \times C}$ represents the temporal patch of X . We segment multiple electrodes in the dataset according to distinct functional areas of the brain. The association between EEG electrodes and functional regions is delineated in Table I, while the partitioning of the spatial regions in two distinct datasets is illustrated in Fig.4 and Fig.5.

2. The spatial and temporal patches SR_i and TR_j are fed into the convolution unit to extract local features. Specifically, we employ convolutions on different patches. Then, the feature maps learned from different patches

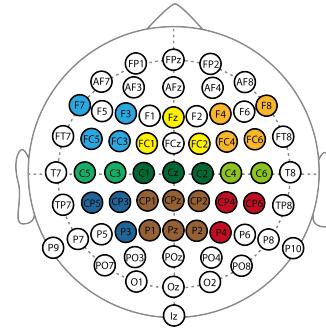


Fig. 4. A depiction of the partitioning of spatial regions in the OpenBMI dataset.

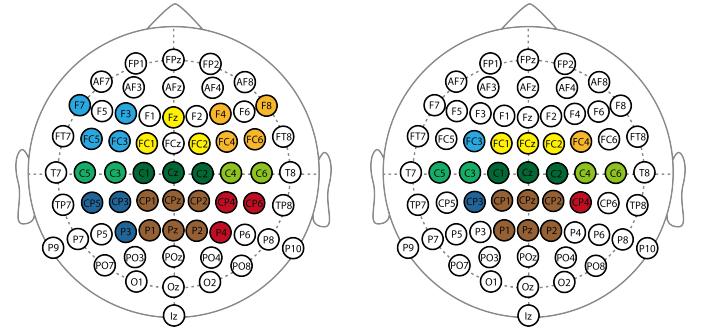


Fig. 5. A depiction of the partitioning of spatial regions in the BCIV-2A dataset.

are concatenated to obtain a local feature map, which is defined as:

$$SR'_i = \text{Reshape}(\text{ConvUnit}(SR_i)) \quad (5)$$

$$TR'_j = \text{Reshape}(\text{ConvUnit}(TR_j)) \quad (6)$$

$$X_{S-LOCAL} = \text{Concat}(SR'_1, \dots, SR'_i) \quad (7)$$

$$X_{T-LOCAL} = \text{Concat}(TR'_1, \dots, TR'_j) \quad (8)$$

where ConvUnit represents the multi-convolution operation, Reshape signifies the channel-reshape operation, and SR'_i and TR'_j refer to the local spatial and temporal feature maps.

TABLE I
DESCRIPTION OF BRAIN REGION DIVISION

Dataset	Brain Region	Electrodes Name
BCI-2A	Frontal Central	FC1, FC2, FCz
	Left Frontal	FC3
	Right Frontal	FC4
	Central	C1, Cz, C2
	Left Central	C3, C5
	Right Central	C4, C6
	Central Parietal	CP1, CPz, CP2, P1, Pz, P2
	Left Parietal	CP3
	Right Parietal	CP4
OpenBMI	Frontal Central	Fz, FC1, FC2
	Left Frontal	F3, F7, FC3, FC5
	Right Frontal	F4, F8, FC4, FC6
	Central	C1, Cz, C2
	Left Central	C3, C5
	Right Central	C4, C6
	Central Parietal	CP1, CPz, CP2, P1, Pz, P2
	Left Parietal	CP3, CP5, P3
	Right Parietal	CP4, CP6, P4

B. Spatial-Temporal Local-to-Global Feature Extraction Module

“No neuron is an island.” The emergence of behavior and cognition is not solely caused by signal transmission between different brain regions but also involves the interactions between cortical areas [41]. Hence, it is crucial to model the interconnectivity among brain regions. Local features reflect the specific brain electrical activity of distinct areas, whereas global features showcase the activity patterns of the entire brain network. Learning the feature representation from a local to global scale enables the identification of the temporal and spatial evolution of brain electrical activity. To effectively capture the local-to-global features in both temporal and spatial domains, we introduce a Spatio-Temporal Local-to-Global Feature Extraction module (L2G). The L2G module, as shown in Fig. 6, utilizes region-based attention to transform the local feature maps. This process mimics the formation of activation patterns in the human brain.

- As stated in Section. IV-A, $X_{S-LOCAL}$, and $X_{T-LOCAL}$ represent the intermediate feature maps that contain local spatial and temporal information. The process of modeling local-to-global spatial-temporal information can be formulated as follows:

$$Q_s, K_s, V_s = \text{Linear}(X_{S-LOCAL}) \quad (9)$$

$$Q_t, K_t, V_t = \text{Linear}(X_{T-LOCAL}) \quad (10)$$

$$Att_s = \text{softmax}\left(\frac{Q_s K_s^T}{\sqrt{d_{X_{S-LOCAL}}}}\right) V_s \quad (11)$$

$$Att_t = \text{softmax}\left(\frac{Q_t K_t^T}{\sqrt{d_{X_{T-LOCAL}}}}\right) V_t \quad (12)$$

$$X_{S-att} = \text{LayerNorm}(Att_s + FC_s(Att_s)) \quad (13)$$

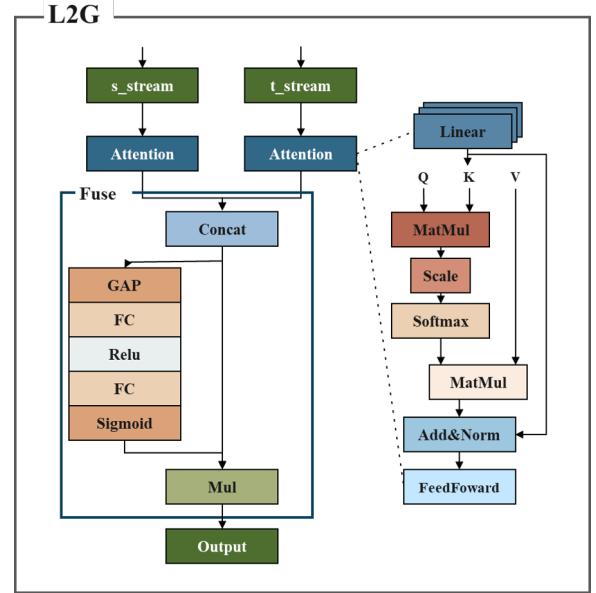


Fig. 6. The Spatial-Temporal Local-to-Global Feature Extractor (L2G) is developed to capture spatial-temporal features ranging from local to global. To comprehensively analyze the activation characteristics of each region from a local to a global perspective, we employ the GAP and MLP layers. Lastly, we combine the feature vectors of spatial-temporal dimensions using the addition operation, resulting in the formation of the final feature map.

$$X_{T-att} = \text{LayerNorm}(Att_t + FC_s(Att_t)) \quad (14)$$

where Q_s, K_s, V_s represent the transformed queries, keys, and values of the spatial feature map, while Q_t, K_t, V_t denote the transformed queries, keys, and values of the temporal feature map. Att_s and Att_t refer to the embedded features that are weighted by the attention score. X_{S-att} and X_{T-att} indicate the feature maps that have been encoded with the local to global information.

- Inspired by the fusion method of [42]:

$$X_{att} = \text{Concat}(X_{S-att}, X_{T-att}) \quad (15)$$

$$X_{fuse} = \sigma(FC_2(\theta(FC_1(GAP(X_{att})))) \quad (16)$$

$$X_{L2g} = X_{att} \odot X_{fuse} \quad (17)$$

where GAP denotes the global average pooling operation, FC_i denotes i -th fully connected layers, θ is the ReLU activation function, σ is the sigmoid activation function.

C. Domain Generalization

To alleviate the influence of inter-individual, an adversarial domain generalization approach is adopted to improve the resilience of our model. Specifically, DG aims to train the model in a manner that makes it impossible to differentiate the source domain of the sample data. Simultaneously, it strives to improve the classification performance of MI as much as possible. For instance, the model cannot discern whether the samples from domain i correspond to its own domain, yet it can still accurately recognize the motor imagery. Thus, the model has not learned personalized features specific to each

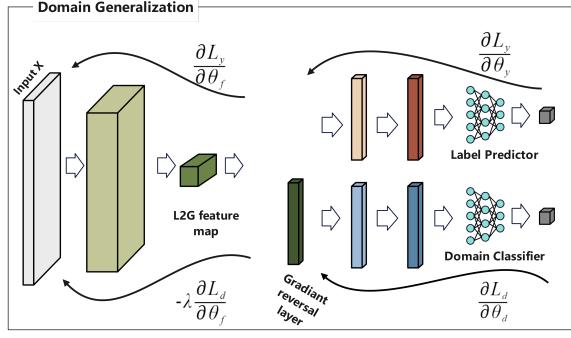


Fig. 7. The architecture of domain generalization.

domain, but rather subject-invariant features associated with MI classification. As shown in Fig. 7. Domain generalization comprises three main components: a feature extractor \mathcal{G}_f , a label predictor \mathcal{G}_l , and a domain classifier \mathcal{G}_d .

1. S/TLE and L2G module serve as the feature extractor \mathcal{G}_f that transforms the input data into a feature space that is invariant to domain shifts.

$$X_{L2G_i} = \mathcal{G}_f(X_i; \theta_f) \quad (18)$$

where X refers to the i -th sample of raw EEG data input, θ_f represents the model's trainable parameters, and X_{L2G} corresponds to the feature vector that has been embedded with L2G spatial and temporal information.

2. The extracted sequence is subsequently inputted into both the label predictor, denoted as \mathcal{G}_l , and the domain classifier, denoted as \mathcal{G}_d . A softmax function is employed for both of these stages:

$$\hat{y}_i = \text{softmax}(\mathcal{G}_l(X_{L2G_i}; \theta_y)) \quad (19)$$

$$\hat{d}_i = \text{softmax}(\mathcal{G}_d(X_{L2G_i}; \theta_d)) \quad (20)$$

where \mathbb{X}_i refers to the embedded sequence originating from the i^{th} sample.

3. The predicted results of \mathcal{G}_l and \mathcal{G}_d denoted as \hat{y}_i and \hat{d}_i respectively are both multi-class classifiers. The loss functions can be defined as follows:

$$\mathcal{L}_y = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{C_y} y_{i,j} \log \hat{y}_{i,j} \quad (21)$$

$$\mathcal{L}_d = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{C_d} d_{i,j} \log \hat{d}_{i,j} \quad (22)$$

where \mathcal{L}_y represents the cross entropy loss function applied to the multi-classification problem, N refers to the total number of samples within the dataset, whereas C_y and C_d signify the number of classes and domains respectively. The variable y denotes the true label associated with a sample, while d represents the true domain of that sample.

To establish an adversarial relationship between the label predictor and the domain classifier, we introduce a Gradient Reversal Layer (GRL) [43] between the feature extractor \mathcal{G}_f

Algorithm 1 Training Procedure of STL2G-DG

Input: EEG signals X from M training domains $S_{train} = \{S^i | i = 1, \dots, M\}$; Ground truth label y ; domain label d

Output: $pred$, the prediction of STL2G-DG;

- 1: Initialization: Initialize model parameters: θ_f , θ_y , θ_d ;
- 2: **while** θ_f , θ_y , θ_d is not converged **do**
- 3: Sample data (x_i, y_i) , with domain label d_j and ground truth label y_i ;
- 4: Calculate the label predictor loss \mathcal{L}_y ;
- 5: Optimize the parameters of the feature extractor and label predictor by: $\theta_f \leftarrow \theta_f - \mu \frac{\partial \mathcal{L}_y}{\partial \theta_f}$, $\theta_y \leftarrow \theta_y - \mu \frac{\partial \mathcal{L}_y}{\partial \theta_y}$;
- 6: Calculate the domain loss \mathcal{L}_d ;
- 7: Optimize the parameters of the feature extractor and label predictor by: $\theta_f \leftarrow \theta_f + \mu \frac{\partial \mathcal{L}_d}{\partial \theta_f}$, $\theta_d \leftarrow \theta_d - \mu \frac{\partial \mathcal{L}_d}{\partial \theta_d}$;
- 8: **end while**

and the domain classifier \mathcal{G}_d . By incorporating GRL, we can effectively merge feature learning and domain generalization into a cohesive framework that enables seamless execution of backpropagation algorithms. The overall loss function and optimization procedure for our proposed model are formulated as follows:

$$\mathcal{L}_{DG} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{C_y} y_{i,j} \log \hat{y}_{i,j} + \lambda \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{C_d} d_{i,j} \log \hat{d}_{i,j} \quad (23)$$

$$(\hat{\theta}_f, \hat{\theta}_y) = \arg \min_{\theta_f, \theta_y} \mathcal{L}(\theta_f, \theta_y, \hat{\theta}_d) \quad (24)$$

$$(\hat{\theta}_d) = \arg \max_{\theta_d} \mathcal{L}(\hat{\theta}_f, \hat{\theta}_y, \theta_d) \quad (25)$$

Optimizing the loss function can enable the feature extractor \mathcal{G}_f to realize the objective of identifying the subject-invariant feature space. To achieve this, we utilize θ_y and θ_d as parameters for minimizing the loss of \mathcal{G}_y and \mathcal{G}_d , respectively. As shown in Algorithm.1, we perform optimization on the parameter θ_y in a manner that entails minimizing the loss associated with \mathcal{G}_y while concurrently maximizing the loss associated with \mathcal{G}_d .

V. EXPERIMENT

A. Datasets

We use two publicly available EEG datasets for MI classification to demonstrate the validity of our proposed method.

1) *BCI Competition IV-2A* [44]: The dataset consists of EEG data from 9 subjects who are instructed to perform four distinct motor imagery tasks: imagining the movement of the left hand, right hand, both feet and tongue. Each subject participated in two sessions on separate days, with each session comprising 6 runs interspersed with short breaks. Each run consisted of 48 trials, with 12 trials for each of the four potential classes. As a result, a total of 288 trials were conducted per session. We chose 22 EEG channels and extracted data with a duration of [2,6] seconds to conduct the experiments.

2) *OpenBMI* [45]: This dataset comprises 64-electrode EEG motor-imagery signals collected from 54 participants engaging

TABLE II
HYPERPARAMETER SETTINGS FOR STL2G-DG

Parameters	Value	Parameters	Value
Learning Rate	0.001	Batch Size	200
Optimizer	Adam	Epochs	200
Dropout Rate	0.001	λ	$2(1 + e^{-10})^{-1}$
Attention layer	2	Fully Connected layer	2

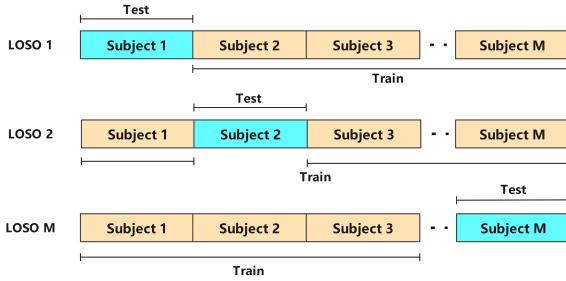


Fig. 8. Illustrations for experimental settings of Leave-one-subject-out (LOSO) validation.

in 2 sessions of imagined movements involving either the left or right hand. We chose 30 EEG channels and extracted data with a duration of [3,7] seconds to conduct the experiments.

B. Experiments Settings

Our approach is coded using Python 3.7 and the PyTorch library, running on a Geforce 2080Ti GPU. For both datasets, all the EEG channels are employed for classification. We train our model through backpropagation from scratch. The specific settings of the hyperparameters of the STL2G-DG are shown in Table II.

To demonstrate the generalizability of our MI classification model, we perform the experiments in a subject-independent manner. The leave-one-subject-out (LOSO) validation [46], [47] is applied to all algorithms used in our experimental setup, as shown in Fig. 8. In a specific dataset with M subjects, each fold of LOSO validation involves using a single subject's data as the testing set, while the data from the remaining $M - 1$ subjects serve as the training set. Finally, we calculate the average classification accuracy from the M evaluations to assess the overall performance of STL2G-DG and the baseline models.

C. Baselines

We evaluate the proposed method STL2G-DG against the following competing methods.

- EEGNET [24]: A classical compact fully convolutional network that incorporates depth-wise and separable convolutions for EEG classification via CNNs.
- ConvNet [48]: The ConvNet consists of four convolution-pooling blocks, which greatly enhance its capability to extract high-level features from raw EEG signals.
- MMCNN [49]: An end-to-end deep learning model encompasses a composition of five parallel EEG Inception

Networks, each comprised of an EEG Inception block, a Residual block, and a Squeeze and Excitation block.

- ST-DG [20]: A spatial-temporal attention within a domain generalization framework. The spatial and temporal attention serves as feature extractors for extracting spatial and temporal features. The domain generalization technique aids in extracting domain-invariant features for the model.

D. Evaluation

We adopt accuracy (ACC) and area under the curve (AUC) to evaluate the overall classification performance. The formulas can be expressed as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (26)$$

$$FPR = \frac{FP}{FP + TN} \quad (27)$$

$$TPR = \frac{TP}{TP + FN} \quad (28)$$

where TP is the true positive, TN is the true negative, and FP is the false positive, and FN is the false negative. AUC is defined as the area under the Receiver Operating Characteristic (ROC) curve. The x-axis of the ROC curve represents the False Positive Rate (FPR), while the y-axis represents the True Positive Rate (TPR).

E. Subject-Independent Experiment Results

Table III presents the overall performance of our STL2G-DG and four baseline methods across all subjects for subject-independent settings. It is observed that in a subject-independent manner, STL2G-DG achieved the highest performance in terms of accuracy on the BCI-2A dataset and OpenBMI dataset.

TABLE III
CLASSIFICATION PERFORMANCE (ACCURACY (ACC) \pm SD AND AREA UNDER CURVE (AUC) \pm SD) FOR THE SUBJECT-INDEPENDENT SCHEMES ON BCIIV-2A AND OPENBMI COMPARED TO FIVE DIFFERENT METHODS

Dataset	Methods	Accuracy(%)	AUC
BCI-2A	EEGNet	51.639 \pm 1.7	0.777 \pm 0.013
	ConvNet	53.003 \pm 1.7	0.799 \pm 0.014
	MMCNN	52.930 \pm 1.5	0.789 \pm 0.013
	ST-DG	56.567 \pm 1.4	0.798 \pm 0.014
	STL2G-DG	59.743 \pm 1.4	0.819 \pm 0.014
OpenBMI	EEGNet	68.200 \pm 9.94	0.704 \pm 0.133
	ConvNet	67.325 \pm 12.04	0.675 \pm 0.130
	MMCNN	68.975 \pm 10.98	0.708 \pm 0.141
	ST-DG	68.938 \pm 11.81	0.715 \pm 0.132
	STL2G-DG	70.422 \pm 11.71	0.726 \pm 0.140

CNN-based methods (EEGNet, ConvNet, and MMCNN) leverage convolutional operations to generate temporal and spatial filters. Nonetheless, the translation invariance [50] characteristic of convolutional operations limits their effectiveness in capturing the distinctive attributes of EEG signals across diverse spatial and temporal regions. Additionally, although the

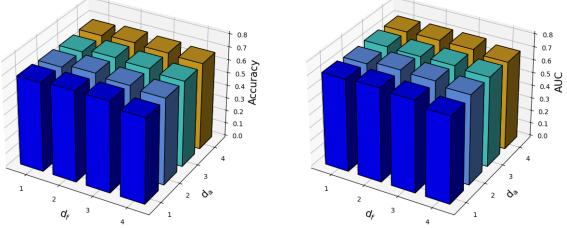


Fig. 9. The sensitivity test results. Two figures are provided, each illustrating the impact of different configurations on the MI classification performance on the OpenBMI dataset. These configurations include the number of spatial-temporal attention layers (d_a) and the number of feedforward layers (d_f).

ST-DG architecture integrates spatial-temporal attention and domain generalization, it treats brain regions in a comprehensive manner, thereby neglecting the complex spatial-temporal activation relationships within and among different regions. EEG signals exhibit significant variations across distinct brain regions, each associated with diverse functions and activity patterns. The distribution of these unique characteristics in the spatial-temporal domain is crucial as it provides essential insights into brain function and activation states. The proposed method partitions time and space, effectively capturing local-to-global spatial-temporal features through an attention mechanism. The result confirms that our proposed approach excels in acquiring discriminative features for MI classification.

In Fig. 9, all hyperparameters are fixed except for d_a and d_f . We investigate the classification performance of the STL2G-DG model on the OpenBMI dataset. Our observations indicate that the model performs better when d_a is set to 2 compared to other values. Additionally, we notice that the influence of d_a on model performance is more pronounced than that of d_f .

Fig. 10 illustrates the t-SNE projection of the learned embedding features of all datasets in a subject-independent manner. Specifically, the embedded features of distinct categories within STL2G-DG exhibit a higher degree of clustering. Conversely, the embedded features of baseline methods across diverse categories appear to be more widely spread and less densely packed within the projection space. This finding highlights the effectiveness of our proposed method, ultimately enriching and refining the discriminative pattern among various classes of EEG data.

F. Visualization

To further reveal the spatial and temporal activation patterns for MI learned by the proposed model, the visualization of multiple feature maps is conducted on the OpenBMI datasets. For a clear illustration, the feature maps of different subjects with different labels are obtained and mapped over the brain surface according to the electrode positions. In all plots, red regions mark channel relevance for the classification task whereas blue regions mark irrelevant ones. The Fig. 11 and Fig. 12 indicate the following conclusions.

- The STL2G module specifically targets small, specific spatial and temporal regions, as they are essential for extracting spatial-temporal features from EEG signals. The

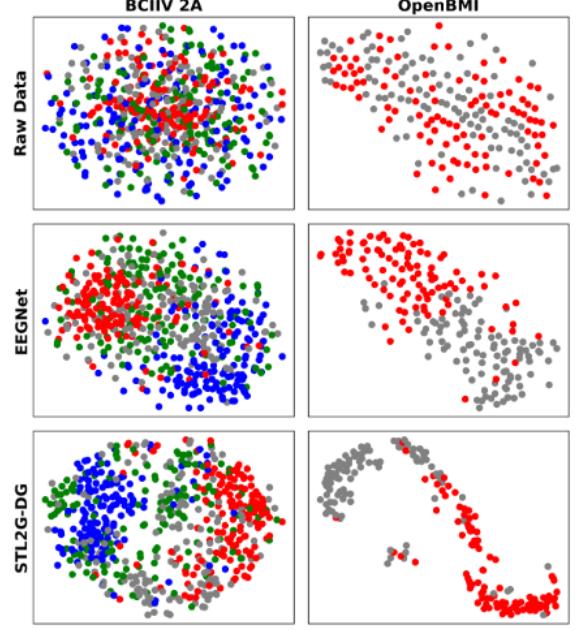


Fig. 10. Visualization of raw data and intermediate features maps of used approaches in the subject-independent experiment using t-SNE projection. The picture presents a comparison of two-dimensional t-SNE projections on both datasets.

model exhibits a pronounced spatial preference towards certain brain regions, with attention levels significantly skewed in their favor over other regions. Similarly, from a temporal perspective, the model displays a notable preference for specific time intervals, with attention levels markedly higher compared to other periods.

- The DG module is instrumental in facilitating the model to capture shared spatial-temporal activation features. Our model assigns comparable weights to corresponding brain regions across various subjects during the MI task. It is noteworthy that across almost all subjects, the model tends to allocate higher weights during the early experiment, with a noticeable decrease in weight allocation during the latter part of the task.

G. Ablation Experiment

1) Effectiveness of different modules: To thoroughly assess the influence of different components in the STL2G-DG framework, we dissect the modules of STL2G-DG and carry out experiments using the OpenBMI dataset.

- SL2G-DG:** To investigate the advantages of incorporating temporal information from EEG signals, the T-L2G module is removed.
- TL2G-DG:** To investigate the advantages of incorporating spatial information from EEG signals, the S-L2G module is removed.
- STL2G:** To investigate the advantages of incorporating subject-invariant information from EEG signals, the DG architecture is removed.

By examining the results of the ablation experiments presented in Table IV, it becomes apparent that the overall classi-

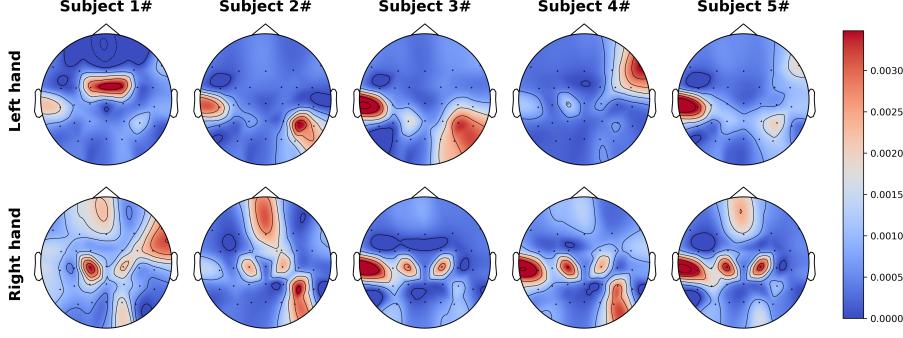


Fig. 11. The topographic maps of the channel-wise weights learned by the ST-L2G module. The row denotes different labels, while the column denotes different subjects from the OpenBMI dataset. A large value is indicated by the color red, indicating a major impact on MI classification, while a small value is represented by the color blue, indicating a minor impact.

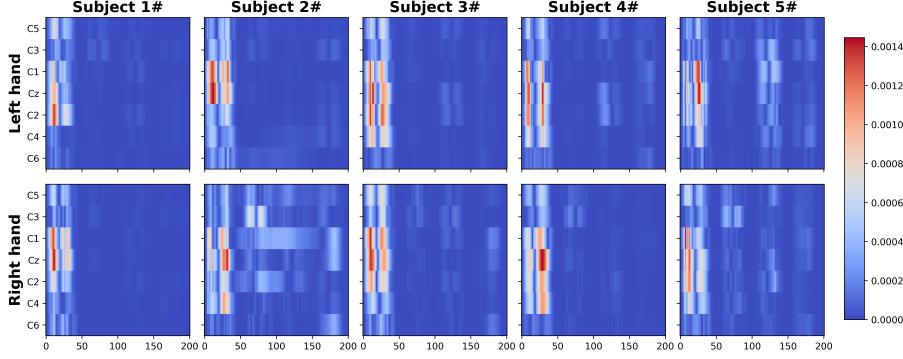


Fig. 12. Visualization of weights assigned to various temporal and spatial regions. We can observe that the weight assignment illustrates the traits of regional convergence.

TABLE IV

ABLATION STUDY RESULTS ON THE CLASSIFICATION PERFORMANCE OF OPENBMI DATASETS TO ASSESS THE INFLUENCE OF DIFFERENT COMPONENTS.

Dataset	Comparision Model	Accuracy(%)	AUC
OpenBMI	SL2G-DG	67.525 ± 9.16	0.703 ± 0.139
	TL2G-DG	68.500 ± 11.73	0.715 ± 0.141
	STL2G	67.931 ± 10.63	0.711 ± 0.132
	STL2G-DG	70.422 ± 11.71	0.726 ± 0.140

fication performance of STL2G-DG exceeds that of SL2G-DG and TL2G-DG in isolation.

2) *Impact of Region division:* To validate the effectiveness of the Local-to-Global approach, we thoroughly analyze the impact of different region division methods on the OpenBMI dataset. Our paper uses the spatial and temporal region setting based on some neurological findings. We compared our default method with several variants.

- **Variant 1:** Model without spatial-temporal region division.
- **Variant 2:** Random division, where channels and time points are divided into 6 groups without considering spatial and temporal relationships.
- **Variant 3:** Neighbor division, where channels and time points are divided according to space and time proximity.

TABLE V

ABLATION EXPERIMENTS OF DIFFERENT REGION DIVISION METHODS.

Variant model	Split method	Accuracy(%)
Variant 1	Without region division	68.938
Variant 2	Random division	67.103
Variant 3	Neighbor division	68.134
STL2G-DG	Default	70.422

From the results presented in Table V, we can conclude that a proper choice of region setting contributes to a better result. The selection of the division method helps the model learn spatial-temporal information and aggregate local information of a larger range. Therefore, in our experiments, the random region division method can only gather coarse-grained spatial information, and more regions may fail to aggregate local information properly, which both perform worse than our default setting.

Additionally, when comparing our default setting with the random setting, it becomes evident that the aggregation of spatial-temporal information is crucial for achieving optimal performance. This observation leads to two conclusions: 1) The consideration of locality functionality is essential when dividing regions. 2) The utilization of prior knowledge to construct a local-to-global pattern effectively extracts spatial-temporal information features in brain regions.

VI. CONCLUSION

In this article, we propose a novel STL2G-DG architecture to learn spatial-temporal local-to-global and more discriminative feature representation for MI classification. Firstly, the STLE module is presented, which leverages convolutional operations to capture information within local regions. Subsequently, the L2G module is utilized to acquire a feature representation that transitions from local to global through spatial-temporal attention. Then, the extracted feature map is fed into the domain generalization module. Especially, a domain discriminator is introduced for adversarial domain generalization on the extracted features, aiming at aligning different distributions across different subjects within the training set. As a result, we can learn the subject-invariant features and train a robust classifier with better generalization for the unseen test domain. The extensive experiments demonstrate that compared to other state-of-the-art methods, the proposed method achieves superior performance. Additionally, we visualize the feature map from the feature extractor, and we confirm that our proposed method can extract local-to-global information related to MI tasks.

ACKNOWLEDGMENTS

This project is funded by National Natural Science Foundation of China (Grant No.62306317) and China Postdoctoral Science Foundation (Grant No.GZC20232992 and 2023M733738)

REFERENCES

- [1] V. Gandhi, G. Prasad, D. Coyle, L. Behera, and T. M. McGinnity, “Eeg-based mobile robot control through an adaptive brain–robot interface,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 9, pp. 1278–1285, 2014.
- [2] J. Zhuang, K. Geng, and G. Yin, “Ensemble learning based brain–computer interface system for ground vehicle control,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 9, pp. 5392–5404, 2021.
- [3] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-r. Muller, “Optimizing spatial filters for robust eeg single-trial analysis,” *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.
- [4] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, “Filter bank common spatial pattern (fbcsp) in brain-computer interface,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 2390–2397.
- [5] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, “Temporally constrained sparse group spatial patterns for motor imagery bci,” *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3322–3332, 2019.
- [6] B. Chen, Y. Li, J. Dong, N. Lu, and J. Qin, “Common spatial patterns based on the quantized minimum error entropy criterion,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 11, pp. 4557–4568, 2020.
- [7] K.-R. Muller, C. Anderson, and G. Birch, “Linear and nonlinear methods for brain-computer interfaces,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 165–169, 2003.
- [8] D. Garrett, D. Peterson, C. Anderson, and M. Thaut, “Comparison of linear, nonlinear, and feature selection methods for eeg signal classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 141–144, 2003.
- [9] J. Fumanal-Idocin, Y.-K. Wang, C.-T. Lin, J. Fernández, J. A. Sanz, and H. Bustince, “Motor-imagery-based brain–computer interface using signal derivation and aggregation functions,” *IEEE Transactions on Cybernetics*, vol. 52, no. 8, pp. 7944–7955, 2022.
- [10] S. Pancholi, A. Giri, A. Jain, L. Kumar, and S. Roy, “Source aware deep learning framework for hand kinematic reconstruction using eeg signal,” *IEEE Transactions on Cybernetics*, vol. 53, no. 7, pp. 4094–4106, 2023.
- [11] J.-H. Cho, J.-H. Jeong, and S.-W. Lee, “Neurograsp: Real-time eeg classification of high-level motor imagery tasks using a dual-stage deep learning framework,” *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 13 279–13 292, 2022.
- [12] W. Dang, D. Lv, M. Tang, X. Sun, Y. Liu, C. Grebogi, and Z. Gao, “Flashlight-net: A modular convolutional neural network for motor imagery eeg classification,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–10, 2024.
- [13] E. K. Miller and J. D. Cohen, “An integrative theory of prefrontal cortex function,” *Annual review of neuroscience*, vol. 24, no. 1, pp. 167–202, 2001.
- [14] T. Allison, A. Puce, and G. McCarthy, “Social perception from visual cues: role of the sts region,” pp. 267–278, 2000.
- [15] A. I. Luppi, P. A. Mediano, F. E. Rosas, N. Holland, T. D. Fryer, J. T. O’Brien, J. B. Rowe, D. K. Menon, D. Bor, and E. A. Stamatakis, “A synergistic core for human brain evolution and cognition,” *Nature Neuroscience*, vol. 25, no. 6, pp. 771–782, 2022.
- [16] D. J. Felleman and D. C. Van Essen, “Distributed hierarchical processing in the primate cerebral cortex,” *Cerebral cortex (New York, N.Y. 1991)*, vol. 1, no. 1, pp. 1–47, 1991.
- [17] C. Chang and G. H. Glover, “Time-frequency dynamics of resting-state brain connectivity measured with fmri,” *NeuroImage (Orlando, Fla.)*, vol. 50, no. 1, pp. 81–98, 2010.
- [18] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, “A review of classification algorithms for eeg-based brain–computer interfaces,” *Journal of neural engineering*, vol. 4, no. 2, pp. R1–R13, 2007.
- [19] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, “A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update,” *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018.
- [20] S. Liu, L. An, C. Zhang, and Z. Jia, “A spatial-temporal transformer based on domain generalization for motor imagery classification,” in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2023, pp. 20–25.
- [21] S. Sakhavi, C. Guan, and S. Yan, “Learning temporal information for brain-computer interface using convolutional neural networks,” *IEEE transaction on neural networks and learning systems*, vol. 29, no. 11, pp. 5619–5629, 2018.
- [22] D. Huang, S. Chen, C. Liu, L. Zheng, Z. Tian, and D. Jiang, “Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for eeg emotion recognition,” *Neurocomputing (Amsterdam)*, vol. 448, pp. 140–151, 2021.
- [23] Y. Liu and Z. Jia, “Bsst: A bayesian spatial-temporal transformer for sleep staging,” in *The Eleventh International Conference on Learning Representations*, 2023.
- [24] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: A compact convolutional network for eeg-based brain-computer interfaces,” *Journal of Neural Engineering*, vol. 15, no. 5, pp. 056013.1–056013.17, 2018.
- [25] R. T. Schirrmeister, J. T. Springenberg, L. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, “Deep learning with convolutional neural networks for EEG decoding and visualization,” 2017.
- [26] Z. Jia, Y. Lin, J. Wang, X. Wang, P. Xie, and Y. Zhang, “Salientsleepnet: Multimodal salient wave detection network for sleep staging,” 2021, pp. 2614–2620.
- [27] C. Wei, L.-l. Chen, Z.-z. Song, X.-g. Lou, and D.-d. Li, “EEG-based emotion recognition using simple recurrent units network and ensemble learning,” *Biomedical signal processing and control*, vol. 58, p. 101756, 2020.
- [28] L. Xiang, D. Song, Z. Peng, G. Yu, and B. Hu, “Emotion recognition from multi-channel EEG data through convolutional recurrent neural network,” in *IEEE International Conference on Bioinformatics and Biomedicine*, 2017.
- [29] W. Ping, A. Jiang, X. Liu, S. Jing, and Z. Li, “LSTM-based EEG classification in motor imagery tasks,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, no. 11, pp. 1–1, 2018.
- [30] X. Ma, Q. Shuang, C. Du, J. Xing, and H. He, “Improving EEG-based motor imagery classification via spatial and temporal recurrent neural networks,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018.
- [31] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, “Making sense of spatio-temporal preserving representations for eeg-based human intention recognition,” *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3033–3044, 2020.

- [32] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE transactions on affective computing*, vol. 11, no. 3, pp. 532–541, 2020.
- [33] P. Zhong, D. Wang, and C. Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE transactions on affective computing*, vol. 13, no. 3, pp. 1290–1301, 2022.
- [34] E. Jeon, W. Ko, and H. I. Suk, "Domain adaptation with source selection for motor-imagery based BCI," in *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, 2019.
- [35] A. Rozantsev, M. Salzmann, and P. Fua, "Beyond sharing weights for deep domain adaptation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 4, pp. 801–814, 2018.
- [36] S. Sakhavi and C. Guan, "Convolutional neural network-based transfer learning and knowledge distillation using multi-subject data in motor imagery BCI," in *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2017, pp. 588–591.
- [37] H. Raza, H. Cecotti, Y. Li, and G. Prasad, "Adaptive learning with covariate shift-detection for motor imagery-based brain-computer interface," *Soft Computing*, vol. 20, pp. 3085–3096, 2016.
- [38] Y. Cui, Y. Xu, and D. Wu, "EEG-based driver drowsiness estimation using feature weighted episodic training," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 27, no. 11, pp. 2263–2273, 2019.
- [39] Z. Jia, Y. Lin, J. Wang, X. Ning, Y. He, R. Zhou, Y. Zhou, and H. L. Li-wei, "Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1977–1986, 2021.
- [40] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and P. Yu, "Generalizing to unseen domains: A survey on domain generalization," *IEEE Transactions on Knowledge and Data Engineering*, 2022.
- [41] M. Thiebaut de Schotten and S. J. Forkel, "The emergent properties of the connected brain," *Science*, vol. 378, no. 6619, pp. 505–510, 2022.
- [42] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 3559–3568.
- [43] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.
- [44] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the bci competition iv," *Frontiers in Neuroscience*, vol. 6, 2012.
- [45] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, and S.-W. Lee, "Eeg dataset and openbmi toolbox for three bci paradigms: An investigation into bci illiteracy," *GigaScience*, vol. 8, no. 5, p. giz002, 2019.
- [46] S. Saeb, L. Lonini, A. Jayaraman, D. Mohr, and K. Kording, "Voodoo machine learning for clinical predictions," *BioRxiv*, 06 2016.
- [47] S. Kunjan, T. S. Grummert, K. J. Pope, D. M. W. Powers, S. P. Fitzgibbon, T. J. Bastiampillai, M. Battersby, and T. W. Lewis, "The necessity of leave one subject out (losos) cross validation for eeg disease diagnosis," in *BI*, 2021.
- [48] R. T. Schirrmeister, J. T. Springenberg, L. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," 2017.
- [49] Z. Jia, Y. Lin, J. Wang, K. Yang, T. Liu, and X. Zhang, "MMCNN: A multi-branch multi-scale convolutional neural network for motor imagery classification," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part III*. Springer, 2021, pp. 736–751.
- [50] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, and R. Faulkner, "Relational inductive biases, deep learning, and graph networks," 2018.