Dynamic Weighted Filter Bank Domain Adaptation for Motor Imagery Brain-Computer Interfaces

Yukun Zhang, Shuang Qiu, Wei Wei, Xuelin Ma, Huiguang He, *Senior Member, IEEE*

***Abstract*—A motor imagery (MI)-based brain–computer interface (BCI) is a promising system that can help neuromuscular injury patients recover or replace their motor abilities. Currently, before one uses MI-BCI, we need to collect a large amount of training data to train the decoding model, and this process is time consuming. When trained with a small amount of data, existing decoding methods generally do not perform well in MI decoding tasks. Therefore, it is important to improve the decoding performance with short calibration data. In this study, we propose a dynamic weighted filter bank domain adaptation framework that uses data from an existing subject to reduce the requirement of data from the new subject. A filter bank is used to explore information from different frequency subbands. A feature extractor with two 1-D convolutional layers is designed to extract EEG features. The class-specific Wasserstein generative adversarial network (WGAN)-based domain adaptation network aligns the distribution of each class between the data from the new subject and the data from the existing subject. Additionally, we apply an attention network to dynamically allocate different weights for different frequency bands. We evaluate our method on a public MI dataset and a self-collected dataset. The experimental results show that the proposed method achieves the best decoding accuracy among the compared methods with different amounts of training data. On the public dataset, our method achieves 8.88% and 7.16% higher decoding accuracy than the best comparing method with on block of training data on the two sessions, respectively. This indicates that our method can enhance MI decoding accuracy with a small amount of training data.**

***Index Terms*—brain-computer interface, motor imagery, domain adaptation, filter bank, attention network**

I. INTRODUCTION

A brain–computer interface (BCI) is a system that translates the brain's signal into commands for devices[1]. Neuroimaging methods used in BCI systems include electroencephalography (EEG), electrocorticography (ECoG), functional near infrared spectroscopy (fNIRS), functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG). EEG is most extensively studied in BCI research owing to its easy access, high temporal resolution and high safety[2]. Therefore, EEG-based BCI systems have gained great attention in recent years and include several main paradigms, such as P300, steady-state visual evoked potentials (SSVEPs) and motor imagery (MI). Compared with other paradigms, MI decodes spontaneous human motor intention without external stimuli. MI-based BCI systems can be used to help neuromuscular injury patients recover or replace their motor abilities[3-6]. Furthermore, it is applicable for smart home applications, education, and entertainment[7-14].

There have been many studies on improving EEG decoding performance. Traditional EEG decoding methods generally follow the pipeline of feature extraction and classification[15]. The most widely used feature extraction methods are common spatial patterns (CSPs) [16]and their variants[17-21], which are followed by support vector machine (SVM) or linear discriminant analysis (LDA). Recently, deep learning-based algorithms have been introduced to MI decoding and reach equal or better decoding accuracy compared with traditional machine learning methods, such as shallow CNN, EEGNet and cascade convolutional recurrent neural networks[22-26].

Although MI decoding methods have made great achievements, the performance of these methods depends on a large amount of training data. In other words, to achieve better MI decoding accuracy, more data are needed to train the decoding model. Due to the nonstationary property of MI-EEG signals, directly using data from other subjects or data from the same subject collected before would result in poor decoding performance[15, 27]. Thus, before one uses MI-BCI, we have to collect many new data to train the decoding methods. It is inconvenient for subjects to use BCI systems. Additionally, the long data-collecting procedure may cause subject fatigue and distraction. Therefore, it is important to develop an MI decoding algorithm that performs well given a few training data.

Domain adaptation is the process of adapting one or more source domains to transfer information to improve the performance of a target learner[28]. It attempts to bring the distribution of the source closer to that of the target. Thus, in the case of MI decoding, EEG data of existing subjects can be utilized to help a new target subject train a decoding model and reduce his/her calibration time through domain adaptation, which is capable of transferring knowledge from existing EEG data to a new subject and enhancing the MI decoding accuracy

This work was supported in part by the National Natural Science Foundation of China under Grant 61976209, Grant 81701785, and Grant 61906188, and in part by the Strategic Priority Research Program of CAS under GrantXDB32040200. (Corresponding author: Huiguang He).

Yukun Zhang, Shuang Qiu and Huiguang He are with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China. (e-mail: zhangyukun2019@ia.ac.cn).

Yukun Zhang and Shuang Qiu contribute equally to this work.

Yukun Zhang, Shuang Qiu, Huiguang He, Wei Wei and Xuelin Ma is with Research Center for Brain-Inspired Intelligence, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 (e-mail: shuang.qiu@ia.ac.cn, huiguang.he@ia.ac.cn, weiwei2018@ia.ac.cn, maxuelin2@gmail.com).

Huiguang He is also with Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100190, China. Xuelin Ma is also with JD.com.

when the new subject has limited training data.

Several domain adaptation methods for MI tasks have been proposed, such as composite common spatial pattern (CCSP)[29] and Riemannian procrustes analysis (RPA)[30]. However, these methods mainly aim at enhancing decoding accuracy when a large amount of training data from new user subjects is available. The decoding accuracy of these methods with a short amount of training data is still limited. Therefore, the effective use of EEG data from existing subjects in domain adaptation methods is still a challenge and needs further study.

In this study, we proposed a dynamic weighted filter bank domain adaptation (DWFBDA) framework to enhance the classification accuracy for MI-BCIs with a short calibration time. A small amount of data collected from a new target subject and already collected data from a source subject is used for the input of our framework. Here, an existing subject with already collected data is called a source, and the new subject is called a target. Our DWFBDA framework includes four parts. First, EEG signals during MI tasks have various frequency bands that contain important information for MI decoding[31]. We adopt a filter bank to explore information from different frequency subbands. Second, we construct a CNN-based feature extractor for learning time and spatial information from EEG samples. A classifier with two fully connected layers is employed to predict the class label. Third, we introduced a Wasserstein generative adversarial network (WGAN) to align the distribution of two domains separately for each class in situations where the target has a small number of training samples. Finally, we designed a dynamic weight model based on an attention mechanism to dynamically assign different weights to each subband to improve the performance of decoding MI tasks.

The main contributions of this paper are summarized as follows:

- We design a class-specific WGAN-based domain adaptation network to deal with domain adaptation with a small number of samples and align the marginal and conditional distribution of the two domains.
- We propose to adopt a filter bank in the MI decoding model. To the best of our knowledge, this is the first work that uses the filter bank method in a deep learning-based MI domain adaptation model.
- We propose a dynamic weight model based on an attention network to dynamically integrate the predicted results from all subbands.
- Experiments on a public dataset and a self-collected dataset are conducted to evaluate our method. The results show that the proposed method achieves the best decoding performance.

II. Related Work

A. Non-domain adaptation MI decoding methods

In 1999, common spatial pattern (CSP)[32], one of the most widely used MI feature extraction methods, was introduced into MI classification by Müller-Gerking et al. An appropriate subject-specific frequency band can improve the classification accuracy of CSP. Thus, the filter bank common spatial pattern (FBCSP)[31] algorithm was proposed to automatically select subband features and has achieved encouraging results. In 2012, FBCSP combined with the Naïve Bayesian Parzen window classifier won BCI competition IV with a mean kappa value of 0.569 for four-class MI classification[33]. The Riemannian minimum distance to mean (RMDM)[34] has been successfully applied in many BCI paradigms, including MI, P300 and SSVEP, due to its simplicity and robustness. In four-class MI tasks, RMDM achieved a classification accuracy of 63.2%[34]. Recently, some studies have proposed some CSP-based methods to improve the decoding performance in MI tasks. Miao et al. proposed common time-frequency-spatial patterns (CTFSP)[35] to extract sparse CSP features from multiband filtered EEG data in multiple time windows and achieved a decoding accuracy of 75% and 85% in two three-class public datasets.

With the development of deep learning, many neural network-based MI decoding models have been proposed. In 2017, Schirrmeister et al. proposed a shallow CNN[22], achieving classification accuracies of 71.9% for a four-class MI task. In 2018, a compact CNN-based model, named EEGNet[23], was proposed, achieving similar accuracy to shallow CNN with fewer parameters. In 2019, Zhang et al. transformed 1-D EEG sequences into 2-D meshes according to the electrode distribution and then applied cascade and parallel convolutional recurrent neural networks to recognize MI intention[24]. In the same year, Sakhavi and colleagues used an FBCSP filtered signal envelope as input and proposed a channel-wise convolution with channel mixing (C2CM) network[36], which reached an accuracy of 74.46% for four-class MI tasks. More recently, some multiview and multitask methods have been proposed to extract better deep representations of EEG features. In 2021, Li et al. proposed to parallelly extract temporal and spectral EEG features and designed a squeeze-and-excitation feature fusion block[37]. This work achieved 74.71% decoding accuracy in a four-class public dataset. Liu et al. proposed a space-time-frequency EEG representation and designed a multitask learning framework[38]. In the same year, Mane et al. proposed FBCNet, which takes multiple EEG frequency bands as different views and designed a multiview convolutional neural network, which achieves 76.2% decoding accuracy in the four-class MI task[39].

B. Domain adaptation-based MI decoding methods

Some domain adaptation methods have been proposed to alleviate the need for training data from the target subject. In 2009, Hyohyeong et al. proposed CCSP[29], a CSP-based domain adaptation method, which works better than CSP with a small number of training samples. In 2019, He et al proposed an unsupervised Euclidean space data alignment method and achieved a classification accuracy of 79.79% on a two-class MI task[40]. In the same year, Rodrigues et al. proposed RPA, a Riemannian geometry-based domain adaptation method, which achieved over 60% accuracy on four-class MI recognition[30].

In 2017, Sakhavi et al. proposed a pretrain fine-tune domain adaptation paradigm based on deep learning for MI

classification [41]. They applied FBCSP to find spatial filters for each subject and then took the envelope power of the spatially filtered data as input for a convolutional neural network (CNN). This neural network is trained on data from multiple subjects and then fine-tuned on the new subject. In 2018, Dose et al. proposed training a global model on a large number of subjects and then fine-tune the global model on target subjects for a few training epochs [42]. In 2019, Jeon [43] et al. developed an MI domain adaptation framework with source selection. They first proposed a subject selection method with power spectral density. A gradient reversal layer was then used to extract common features for the new subject and the source subject.

More recently, Zhao et al. proposed a deep representation-based domain adaptation (DRBDA) framework for MI tasks to improve the classification accuracy of the target subject [44]. To further enhance the domain adaptation performance, a center loss was employed to minimize intraclass differences. Their model achieved an accuracy of 74.75% on a four-class MI decoding task. In 2022, Peterson et al. proposed a domain adaptation method called OTDA based on optimal transport theory[45]. They first trained a CSP-LDA decoding model on existing data. Then, they used the proposed domain adaptation pipeline to calibrate the model on new EEG data. Their method achieves 90.23% decoding accuracy in a two-class MI task. Overall, the existing domain adaptation-based MI decoding method still takes a large amount of training data from the target subject. They may not perform very well with limited training data from the target subject.

III. METHOD

A. Notations

We first introduce the notations and definitions for later use. Let $x \in \mathbb{R}^{C\times T}$ represent one sample of a multichannel EEG signal, where C is the number of channels and T is the time point. $y \in \{1, \ldots, N_{cls}\}$ is the class label, and N_{cls} is the number of MI classes. $X = \{x_i\}_{i=1}^{N}$ is a set of EEG samples, and $Y = \{y_i\}_{i=1}^{N}$ is a set of labels. $P(X)$ is the marginal distribution of X, and $P(X|Y)$ is the conditional distribution of X. A new subject for whom we are going to train a decoding model is called the target subject. A subject with already collected EEG data is referred to as a source subject. Let x_i^t and x_j^s denote the i^{th} and j^{th} EEG samples from the target and source subjects, respectively. y_i^t, y_j^s are the corresponding labels. Then, the target domain with N_t labeled samples from the target subject is defined as $\mathcal{D}^t = \{(x_i^t, y_i^t)\}_{i=1}^{N_t}$. Similarly, $\mathcal{D}^s = \{(x_j^s, y_j^s)\}_{j=1}^{N_s}$ is the source domain that contains N_s labeled samples from the source subject. The marginal and conditional distributions of the two domains are different: $P(X^t) \neq P(X^s)$, $P(X^t|Y^t) \neq P(X^s|Y^s)$.

Our motivation is to align the distribution of two domains in feature space such that $P(Y^t|F^t)$ is close to $P(Y^s|F^s)$ where $F^t = \{f_i^t\}_{i=1}^{N_t}$, $F^s = \{f_j^s\}_{j=1}^{N_s}$. f_i^t, f_j^s is the feature of x_i^t and x_j^s. The classifier trained on the aligned feature space would work well for both target and source domains.

B. Network architecture

The total framework of our method is illustrated in Fig. 1. First, N_b finite impulse response (FIR) bandpass filters are adopted to filter the EEG samples into multiple subbands. The passbands include 4-10, 8-14, …, and 32-38 Hz. These bands cover the frequency range where the main response of MI is[16, 22, 33]. In addition, the full band of 4-38 Hz, which is widely used in the case of only one single band[22], is also included.

A feature extractor with two 1-D convolutional layers is used to extract EEG features from different frequency bands. A 1-D temporal convolution kernel is first used to extract temporal features along the time dimension from each channel. The size of the temporal convolutional kernel is 25. A large kernel allows us to capture relatively long-range information. Next, a 1-D spatial convolutional kernel is adopted to extract the spatial feature of the multiband EEG signal along the channel dimension. The spatial convolution works as a spatial filter that takes the linear combination of EEG channels. The kernel shape of the spatial filter is $C \times 1$. Compared with a 2-D temporal-spatial convolutional kernel, cascaded 1-D kernels extract independent spatial and temporal patterns. The spatial filter takes the same channel combinations regardless of the temporal point. Furthermore, cascaded 1-D kernels reduce the number of parameters compared with one 2-D kernel. This would enhance the model robustness and relieve the overfitting problem. The features are squared to obtain power information. Finally, a temporal average pooling layer with a kernel length of 75 and stride of 15 is adopted to reduce the feature dimension.

Inspired by adversarial training, we further design a discriminator to predict which domain the current input belongs to. If the distribution between F^t and F^s is very different, then the discriminator can easily output the right answer. While the discriminator keeps learning how to predict the domain, the feature extractor tries to cheat the discriminator by extracting features that cannot be distinguished. Ideally, the competition between the feature extractor and discriminator would finally reach a balance point where the distribution of extracted features $P(F^t)$ and $P(F^s)$ is exactly the same and the discriminator is not able to predict the domain. This would result in a common feature space for both domains. We can further train classifiers on this common feature space that works well for both the target and source domains.

The traditional discriminator is trained by cross entropy loss. However, cross entropy cannot well measure the distribution difference when there is little or no overlap[46]. Considering that the amount of target data is limited and the distribution difference of EEG across subjects is large, we use the Wasserstein distance in our discriminator. The Wasserstein distance can measure the distribution difference even when the two distributions have no overlap[46]. Our discriminator contains four fully connected (FC) layers. The activation function is leaky rectified linear units (ReLU). To further improve the prediction performance of our model, we align the conditional distribution between two domains by creating different discriminators for each class separately.

Our classifier contains two FC layers. The activation

functions are leaky ReLU and LogSoftMax. To enhance the stability and performance of the classifier, batch normalization and dropout techniques are adopted.

For the dynamic weight model, we construct an attention network. For each band, the classifier outputs a set of probabilities that a sample belongs to each class. To combine the probabilities from all bands, an attention network is designed to output weights for each band dynamically. Compared with taking the mean average of band predictions, the attention network can give greater weights for bands with more information, hence improving the model performance. The attention network can also output different weights for different samples, which makes the decoding model more flexible. The attention network consists of one fully connected layer and a softmax output layer. Batch normalization and dropout techniques are also adopted. A regular term is used to prevent the output weights from being too far from $1/N_b$, which makes the attention network more stable. In addition, we adopted a multihead attention technique that allows the model to attend to subbands from different views and has been found to be beneficial in many studies[47].

C. Optimization objective

In our framework, we first train the feature extractor, classifier and discriminator on each frequency band. Then, we train the attention network to generate band weights for each band.

For predicting class labels on the ith frequency band, the optimization objective is:

$$\min_{F,C} \mathcal{L}_{cls}^{i} = -\mathbb{E}_{x,y\sim\mathcal{D}^t\cup\mathcal{D}^s}\, CEloss\left(y, C_i\left(F_i(x_i)\right)\right) \quad (1)$$

$$x_i = FB_i(x) \quad (2)$$

where $FB_i(\cdot)$, $F_i(\cdot)$, and $C_i(\cdot)$ are the ith bandpass filter, feature extractor and classifier, respectively, and $CEloss(\cdot)$ is the cross-entropy loss. The output of $C_i(\cdot)$ is a column vector that represents the probability of the current sample belonging to each class.

For adversarial domain adaptation, we first estimate the Wasserstein distance between the target and source domains by maximizing the following loss function:

$$\max_{D} \mathcal{L}_{D}^{i} = \mathbb{E}_{x^t\sim\mathcal{D}^t}\left[D_i\left(F_i(x_i^t)\right)\right] - \mathbb{E}_{x^s\sim\mathcal{D}^s}\left[D_i\left(F_i(x_i^s)\right)\right] \quad (3)$$

where $D_i(\cdot)$ is the discriminator of the ith frequency band. The feature extractor minimizes the Wasserstein distance between two domains against the discriminator:

$$\max_{F} \mathcal{L}_{adv}^{i} = \mathbb{E}_{x^t\sim\mathcal{D}^t}\left[D_i\left(F_i(x_i^t)\right)\right] - \mathbb{E}_{x^s\sim\mathcal{D}^s}\left[D_i\left(F_i(x_i^s)\right)\right] \quad (4)$$

The overall optimization target for the feature extractor and classifier is:

$$\min_{F,C} \mathcal{L}_{F,C}^{i} = w_C \mathcal{L}_{cls}^{i} + w_{adv} \mathcal{L}_{adv}^{i} \quad (5)$$

where w_C and w_{adv} are hyperparameters for weights of classification and domain adaptation loss. For each frequency band, we alternatively optimize $\mathcal{L}_D^i$ and $\mathcal{L}_{F,C}^i$. This objective is optimized on all frequency bands.

After the feature extractor and classifier were well trained, we began to train the attention network.

$$\min_{A} \mathcal{L}_A = -\mathbb{E}_{x,y\sim\mathcal{D}^t\cup\mathcal{D}^s}\, CEloss(y, p_{weighted}) + w_R \cdot \sum_{i=1}^{N_b} \left(A\left(\left(F_i(x_i)\right)\right) - 1/N_b\right) \quad (6)$$

$$p_{weighted} = \left[A\left(\left(F_i(x_i)\right)\right)\right] \times \left[C_i\left(F_i(x_i)\right)\right]^T \quad (7)$$

where $A(\cdot)$ is the attention network and outputs a scalar for each frequency band and $[\cdot]$ represents the rowwise concatenation operation. w_R is the hyperparameter for the weight of our regular term. The optimization method for training the discriminator is RMSProp. For training the feature extractor and classifier, Adam is used.

D. Training pipeline

In the training stage, the feature extractor, classifier and discriminator are alternatively optimized. For each training

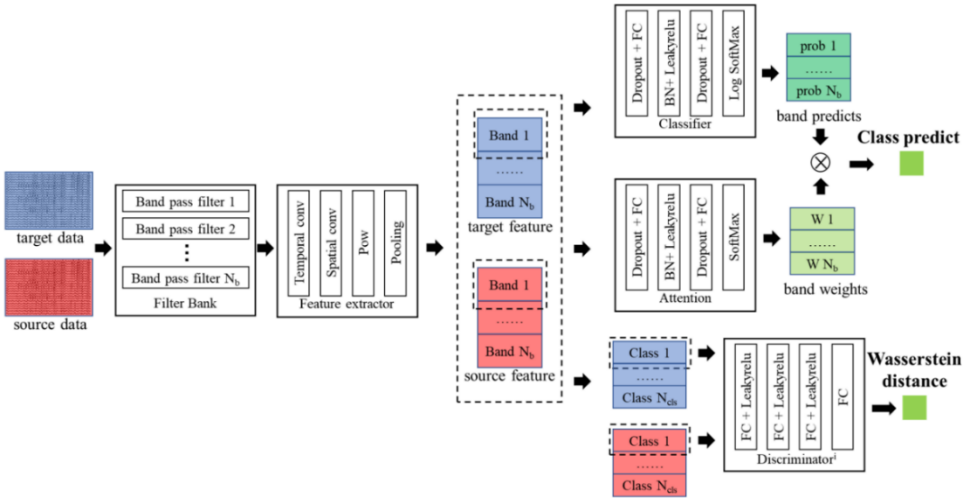

Fig. 1. Proposed filter bank Wasserstein adversarial domain adaptation framework. Target data and source data are separately input into the filter bank. Band passed data are then feed into CNN based feature extractor to get feature maps. The discriminator estimates the Wasserstein distance between target feature maps and source feature maps for each class separately. Feature extractor in the contrary tries to extract features that could not be distinguished by the discriminator. By adversarial training, feature extractor learns to extract common feature between target domain and source domain. The common feature maps are then input into classifier and attention network. The classifier output class predicts for each frequency bands separately and the attention network output the weights for each frequency band. The weighted band predicts are taken as final output.

epoch, we first train D for n_d iterations and then train F and C for n_c iterations. The training pipeline for each frequency band is summarized in Algorithm 1. For convenience, we omit the frequency band index in algorithm 1. Note that the attention network is trained after F, C, D is well trained and is not included in Algorithm 1.

Algorithm 1. Training pipeline of our domain adaptation framework

Input: training data from target and source domain $\mathcal{D}^s$, $\mathcal{D}^t$, maximum training epoch n_{epoch}, number of iterations for training classifier and feature extractor per epoch n_c, number of iterations for training discriminator per epoch n_d, batch size m

Output: feature extractor F, classifier C, discriminator D_1,…, $D_{N_{class}}$

initialize F, C, D_1,…, $D_{N_{class}}$, with parameters θ_F, θ_C, θ_{D_1},…$\theta_{D_{N_{class}}}$ randomly

for t = 1, …, n_{epoch}:

  for i_{class} = 1,…, N_{cls}:

    for t_d = 1, …, n_d:

      sample a batch $\{(x_i^t, y_i^t)\}_{i=1}^m$ from target class i_{class}

      sample a batch $\{(x_j^s, y_j^s)\}_{j=1}^m$ from source class i_{class}

$g_{\theta_{D_{i_{class}}}} \leftarrow \nabla_{\theta_{D_{i_{class}}}}[\frac{1}{m}\sum_{i=1}^m D_{i_{class}}(F(x_i^t)) - \frac{1}{m}\sum_{j=1}^m D_{i_{class}}(F(x_j^s))]$

$\theta_{D_{i_{class}}} \leftarrow \theta_{D_{i_{class}}} + w_D \cdot \mathrm{RMSProp}(\theta_{D_{i_{class}}}, g_{\theta_{D_{i_{class}}}})$

    end for

  end for

  for t_c = 1, …, n_c:

    sample a batch $\{(x_i^t, y_i^t)\}_{i=1}^m$ from target

    sample a batch $\{(x_j^s, y_j^s)\}_{j=1}^m$ from source

$g_{\theta_F} \leftarrow -\nabla_{\theta_F}[\frac{1}{m}\sum_{i=1}^m \log\left(\Sigma_k^{N_{cls}} C(F(x_i^t))_k \cdot \delta(k = y_i^t)\right) + \frac{1}{m}\sum_{j=1}^m log(\Sigma_k^{N_{cls}} C\left(F(x_j^s)\right)_k \cdot \delta(k = y_j^s))]$

    for i_{class} = 1,…, N_{class}:

$g_{\theta_F} \leftarrow g_{\theta_F} - \nabla_{\theta_F}[\frac{1}{m}\sum_{i=1}^m D_{i_{class}}(F(x_i^t)) - \frac{1}{m}\sum_{j=1}^m D_{i_{class}}(F(x_j^s))]$

    end for

$g_{\theta_C} \leftarrow -\nabla_{\theta_C}[\frac{1}{m}\sum_{i=1}^m \log\left(\Sigma_k^{N_{cls}} C(F(x_i^t))_k \cdot \delta(k = y_i^t)\right) + \frac{1}{m}\sum_{j=1}^m log(\Sigma_k^{N_{cls}} C\left(F(x_j^s)\right)_k \cdot \delta(k = y_j^s))]$

  $\theta_F \leftarrow \theta_F + w_F \cdot \mathrm{Adam}(\theta_F, g_{\theta_F})$, $\theta_C \leftarrow \theta_C + w_C \cdot \mathrm{Adam}(\theta_C, g_{\theta_C})$

  end for

end for

w_D, w_F, w_C are hyperparameters, and $\delta(a == b)$ is the indicator function, which equals 1 if $a == b$; otherwise, it equals 0.

E. Source selection

The selection of source subjects is important in the MI domain adaptation task. An appropriate source subject would improve the decoding accuracy for the target subject, while an inappropriate source subject would result in negative transfer and harm the prediction accuracy.

Our source selection method has two principles. First, the distribution between the target domain and source domain should be close. In this case, the distribution alignment process would be easier. Second, the source domain should provide helpful information for classification. The distribution difference between MI tasks in the source domain should be large. In other words, the source domain itself should have fine decoding accuracy.

For each subject from one dataset, we perform sixfold-fold cross validation on its own data. The subjects are then ranked according to their accuracy. The top five subjects are taken as optional source subjects for this dataset. Then, we use the training set of the target subject to test the models of all optional source subjects. The corresponding subject of the trained model that achieves the highest classification accuracy is finally selected as the source subject for this target subject.

F. Cropped training

We adopt a cropped training strategy to effectively train the proposed neural network. A sliding window with a length of 500 time points and a stride of 10 time points is first used to generate cropped samples. Our network is then trained on those cropped samples. For an original EEG sample $x \in R^{C \times T}$, 63 cropped samples are generated. The predicted results of the 63 cropped samples are averaged to obtain the predicted result for the original EEG sample. Note that we first divide the data into a training set and test set and then crop the samples. The accuracy is calculated based on the original EEG samples rather than the cropped samples.

For faster training and inference, in the feature extractor, we first apply temporal and spatial convolution on the original sample and then crop the sample with a sliding window. By doing so, we save the repeated convolution calculations.

IV. Experiment and Results

A. Dataset

Our method is evaluated on two datasets. The first dataset we use is dataset 2a of BCI competition IV[48] (referred to as dataset 2a for convenience). Nine subjects were instructed to perform four classes of MI tasks (left hand, right hand, both feet, and tongue). For each subject, EEG data on different days were collected. For EEG data from one day, there were 72 samples for each MI task (288 EEG samples in total), which was called one session. Twenty-two channels of EEG signals were collected. The sampling rate is 250 Hz.

The second dataset, MI-2, was collected by our self [49] and contains data from twenty-five subjects on two different MI tasks from the same limb (right hand and right elbow). For each subject, there were 100 samples of MI data for each MI task. To evaluate our method in situations where available data are short, only the first 60 samples from each task are used. EEG data were acquired using a Neuroscan 64-channel amplifier. The left mastoid was set as the reference channel. The sampling rate is 1000 Hz. 0.5-100 Hz bandpass filter and a 50 Hz notch filter were used for data acquisition. To remain consistent with dataset 2a, we downsample the EEG data to 250 Hz.

For both datasets, the raw EEG signal is bandpassed to 4-38 Hz. Although the frequency range of MI cortical oscillations mainly lies in the alpha and beta bands, some studies found that

EEG signals in gamma and theta bands contain MI-related information[50-52]. In this study, we chose 4-38 Hz following the study of Schirrmeister et al. [22] to extract information from a wide frequency band.

B. Compared methods

1) ***Baseline methods:*** *Baseline methods include CSP-based methods and a Reimann geometry-based method.*

- CSP[16] and FBCSP[31] followed by LDA are compared. They are realized by the MNE toolbox[53] https://mne.tools/stable/index.html.
- Riemannian geometry-based method RMDM[34] extracts covariance matrices and makes predictions according to the nearest centroid. It is adopted from https://github.com/pyRiemann/pyRiemann.

2) Deep learning-based methods: Deep learning-based methods include EEGNet, shallow and deep CNN, C2CM and FBCNet.

- EEGNet[23] is a compact CNN-based MI decoding model. we adopt the source code from https://github.com/vlawhern/arl-eegmodels.
- Shallow CNN and deep CNN[22] use temporal and spatial convolution layers to extract MI features and output class predictions with a dense layer. Codes adopted from https://github.com/TNTLFreiburg/braindecode.
- C2CM[36] takes the envelope of FBCSP filtered signals as input and utilizes convolutional layers to learn temporal and spatial information. As we do not have access to the source code, we reimplemented this method following the original paper.
- FBCNet[39] employs a multiview data representation followed by spatial filtering to extract spectra-spatially discriminative features. We adopted their codes from https://github.com/ravikiran-mane/FBCNet.

3) Domain adaptation ***methods:*** domain adaptation methods include OTDA, CCSP, RPA and DRBDA

- OTDA[45] first builds a CSP-based decoding model on source data and then transports the new training data with their optimal transport pipeline to calibrate the model. The codes are adopted from https://github.com/vpeterson/otda-mibci.
- CCSP[29] extracts a spatial filter with weighted summed covariance metrics from different subjects. It is adopted from the MNE toolbox.
- RPA[30] applies recentering, stretching and rotation to the covariance metrics from different subjects and makes predictions with minimum distance to mean classifiers. We adopt the source code from https://github.com/plcrodrigues/RPA.
- DRBDA[44] extracts features with convolutional layers and utilizes GAN-based adversarial domain adaptation to push the feature extractor to learn common features. A center loss is adopted to further align the feature distribution. We reimplemented this method following the original paper.

For traditional methods, including SPD, CSP, FBCSP, RPA, CCSP and OTDA, a time segment of 0.5 s to 2.5 s after the onset of the visual cue is used according to Ang's study [31]. For deep learning methods, including EEGNet, shallow CNN, deep CNN, FBCNet, DRBDA and our method, an exponential moving average is adopted to reduce the nonstationarity of the EEG signal. The decay factor is 0.999. The time segment from 0.5 s before to 4 s after the onset of the MI task is used according to Schirrmeister's study [22]. For C2CM, which uses CSP filtered signals and convolutional layers, a time segment from 0 s to 4 s after the onset of the visual cue is used according to the original paper [36].

TABLE I CLASSIFICATION ACCURACY OF DIFFERENT ALGORITHMS WITH DIFFERENT NUMBER OF TRAINING BLOCKS ON SESSION 1 OF DATASET 2A OF BCI COMPETITION IV (IN PERCENTAGE %, '*', '**' REPRESENTS COMPARED WITH OUR METHOD P<0.05 AND P<0.01 RESPECTIVELY)

method	Training data amounts (blocks)				
	1	2	3	4	5
SPD	55.57 **	58.80 **	60.52 **	61.42 **	61.38 **
CSP	54.50 **	57.54 **	58.95 **	59.03 **	60.11 **
FBCSP	53.24 **	63.74 **	66.86 **	70.62 *	71.84 *
EEGNet	31.99 **	42.26 **	51.65 **	57.18 **	58.02 **
Shallow CNN	51.87 **	60.23 **	66.41 **	70.31 **	70.99 **
Deep CNN	42.85 **	52.57 **	60.01 **	62.75 **	65.93 **
C2CM	52.46 **	60.10 **	66.80 *	69.00 *	71.76 *
FBCNet	50.91**	62.22**	67.19*	70.99	71.95*
RPA	52.84 **	54.95 **	57.52 **	58.41 **	59.57 **
CCSP	55.20 **	58.15 **	59.41 **	60.57 **	60.69 **
DRBDA	55.96 **	58.76 **	61.91 *	67.09 **	69.06 **
OTDA	45.51 **	47.13 **	46.91 **	46.76 **	47.05 **
ours	**64.84**	**69.92**	**73.01**	**74.61**	**76.62**

TABLE II CLASSIFICATION ACCURACY OF DIFFERENT ALGORITHMS WITH DIFFERENT NUMBER OF TRAINING BLOCKS ON SESSION 2 OF DATASET 2A OF BCI COMPETITION IV (IN PERCENTAGE %, '*', '**' REPRESENTS COMPARED WITH OUR METHOD P<0.05 AND P<0.01 RESPECTIVELY)

method	Training data amounts (blocks)				
	1	2	3	4	5
SPD	59.03 **	62.55 **	63.72 **	64.66 **	64.89 **
CSP	56.18 **	59.36 **	61.01 **	61.00 **	61.19 **
FBCSP	55.73 **	64.79 **	67.90 *	69.68 **	71.99 *
EEGNet	33.02 **	47.29 **	54.58 **	59.63 **	64.04 **
Shallow CNN	53.24 **	62.12 **	67.57 **	70.93 **	73.26 **
Deep CNN	45.46 **	57.15 **	62.41 **	67.80 **	67.55 *
C2CM	55.98 **	65.37 **	69.71 **	71.55 **	72.03 **
FBCNet	52.65 **	64.44 **	69.16 *	72.22 *	74.42
RPA	55.43 **	57.89 **	59.39 **	61.00 **	61.07 **
CCSP	56.67 **	60.17 **	61.59 **	62.25 **	61.73 **
DRBDA	54.55 **	59.18 **	61.86 **	67.23 **	69.64 **
OTDA	47.44 **	47.19 **	48.08 **	48.90 **	48.63 **
ours	**66.19**	**70.98**	**74.11**	**76.29**	**77.70**

C. Comparison experiment

We evaluated the decoding accuracy of the proposed method

TABLE III CLASSIFICATION ACCURACY OF DIFFERENT ALGORITHMS WITH DIFFERENT NUMBER OF TRAINING BLOCKS ON MI-2 DATASET (IN PERCENTAGE %, '*', '**' REPRESENTS COMPARED WITH OUR METHOD P<0.05 AND P<0.01 RESPECTIVELY)

method	Training data amounts (blocks)				
	1	2	3	4	5
SPD	53.89	54.52 **	55.18 *	55.07 **	57.03 *
CSP	54.56	55.23 **	57.57	57.40 *	58.97 *
FBCSP	52.25 **	53.62 **	55.58 *	56.03 *	56.53 *
EEGNet	50.63 **	52.13 **	54.07 **	55.20 **	56.00 **
Shallow CNN	55.09	57.91	58.59	59.73	60.97
Deep CNN	52.26 **	53.78 **	55.21 **	55.42 **	58.20 **
C2CM	52.45 *	54.94 **	56.33 *	55.10 *	56.12 *
FBCNet	51.31 **	54.47 **	55.52 **	57.70 *	57.57 *
RPA	54.35	54.67 **	54.76 **	54.75 **	57.10 *
CCSP	54.51	55.10 **	56.86 *	57.15 *	59.20 *
DRBDA	55.37	57.41 **	58.23	59.17 *	60.13
OTDA	51.31 **	51.72 **	51.60 **	53.13 **	52.63 **
ours	**56.58**	**59.23**	**59.77**	**60.82**	**62.07**

by comparing existing methods on each session (each day) from dataset 2a and the MI-2 dataset. For each session of dataset 2a, we use one, two, …, five blocks of data from the target subject as the training set. The remaining data are taken as the test set. Data from the source subject are also used for training our models. For each amount of training data, we perform six runs of experiments using repeated K-folder cross validation. For each folder, a part of the data in the training set are used as validation data.

Table I shows the averaged decoding accuracies of different methods with different amounts of training data from the target subject in the first session of dataset 2a. The two-way repeated-measures ANOVA showed significant main effects of different methods and training data amounts on decoding performance as well as a significant interaction effect between the two factors (all: $p<0.001$). With each amount of training data, our method achieves the best performance among all compared methods. With one block of training data, the Riemannian geometry-based method achieves the best accuracy among traditional methods, and C2CM gains the best accuracy among deep learning-based methods. CCSP, a CSP-based TL method, is higher than CSP. DRBDA, a deep domain adaptation method, obtains higher decoding accuracy than all compared methods. This indicates that domain adaptation can improve MI decoding performance. Our method achieves a decoding accuracy of 64.84%, which is 8.88% higher than that of DRBDA ($p<0.01$). Post hoc tests show that our method significantly outperforms each compared method ($p<0.01$). With two blocks of training data, the decoding accuracy of FBCSP is 63.74%, which is better than that of the other compared methods. This suggests that FBCSP is still a very competitive MI decoding method. With each kind of data amount, our method obtains significantly higher results than the best compared methods ($p < 0.05$). Except for FBCNet with four blocks of training data ($p = 0.052$). These results indicate that our method improves the MI classification performance. Table II summarizes the comparison results in session two of dataset 2a. It shows similar results to Table I. Our method achieves better performance than the best compared method ($p \leq 0.05$) with each data amount. Except for FBCNet with five blocks of training data ($p = 0.0501$). Specifically, with one block of training data, our method improves the accuracy by up to 7.16% compared with

TABLE IV ABLATION STUDY RESULTS ON THE FIRST SESSION OF DATASET 2A OF BCI COMPETITION IV (IN PERCENTAGE %, '*', '**' REPRESENTS COMPARED WITH MODEL 7 P<0.05 AND P<0.01 RESPECTIVELY)

Model	**ADV**	**FB**	**ATT**	**SRC**	**t1**	**t2**	**t3**	**t4**	**t5**	**t6**	**t7**	**t8**	**t9**	**mean**
1					58.82	36.39	68.82	39.79	32.43	31.81	74.51	69.44	58.19	52.25 **
2				√	66.74	38.54	75.63	42.71	37.01	41.25	75.42	73.54	62.57	57.04 **
3	√			√	64.65	37.36	74.86	45.83	42.57	36.53	74.24	73.68	65.14	57.21 **
4		√			69.51	45.49	70.35	50.97	50.35	41.88	75.28	75.49	64.58	60.43 **
5		√		√	75.00	41.53	73.96	49.03	47.92	42.78	77.99	80.07	67.85	61.79 **
6	√	√		√	75.35	45.28	77.22	54.17	53.82	44.24	78.96	79.72	69.58	64.26 *
7	√	√	√	√	75.83	45.42	78.96	55.07	54.51	44.44	79.86	79.03	70.42	64.84

TABLE V ABLATION STUDY RESULTS ON THE SECOND SESSION OF DATASET 2A OF BCI COMPETITION IV (IN PERCENTAGE %, '*', '**' REPRESENTS COMPARED WITH MODEL 7 P<0.05 AND P<0.01 RESPECTIVELY)

Model	**ADV**	**FB**	**ATT**	**SRC**	**t1**	**t2**	**t3**	**t4**	**t5**	**t6**	**t7**	**t8**	**t9**	**mean**
1					66.39	38.06	69.51	52.22	32.99	37.57	69.44	67.78	73.89	56.43 **
2				√	73.54	35.35	71.74	51.39	31.25	41.94	68.89	67.08	75.42	57.40 **
3	√			√	75.00	38.75	75.69	53.96	35.83	41.60	67.92	70.83	78.26	59.76 **
4		√			73.06	43.89	68.61	58.26	42.85	44.17	74.58	71.46	76.94	61.54 **
5		√		√	78.68	44.03	75.90	58.13	40.00	43.19	79.10	73.68	79.65	63.60 **
6	√	√		√	79.24	47.85	76.18	62.71	47.22	43.82	81.11	74.86	80.76	65.97
7	√	√	√	√	79.03	47.15	77.57	64.03	45.42	44.93	82.15	74.24	81.18	66.19

TABLE VI ABLATION STUDY RESULTS ON MI-2 DATASET (IN PERCENTAGE %, '*', '**' REPRESENTS COMPARED WITH MODEL 7 P<0.05 AND P<0.01 RESPECTIVELY)

Model	ADV	FB	ATT	SRC	mean
1					55.58
2				√	55.22 **
3	√			√	56.03
4		√			56.28
5		√		√	56.04
6	√	√		√	56.38
7	√	√	√	√	56.58

the best comparing method. The comparison results on the MI-2 dataset show that our method obtains the best performance (Table III). Therefore, the proposed method outperformed non-domain adaptation methods and domain adaptation methods on datasets 2a and MI-2.

Bland–Altman (BA) analysis shows that the proposed method's classification accuracies with one and two blocks of training data from the target subject are similar to those of the best compared method with one more block of training data (one block vs. two blocks: ours vs. FBCSP (Table I), p=0.250; ours vs. C2CM (Table II), p=0.294; ours vs. shallow CNN (Table III), p=0.068; two block vs. three blocks: ours vs. FBCNet (Table I), p=0.091; ours vs. C2CM (Table II), p=0.212; ours vs. shallow CNN (Table III), p=0.285). Moreover, our model trained on three blocks of data achieves similar classification results with the best compared method trained with five blocks (ours vs. FBCNet (Table I), p= 0.276; ours vs. FBCNet (Table II), p=0.423; ours vs. shallow CNN (Table III), p=0.222). Therefore, our method can reduce the need for training data for at least one block on both dataset 2a of the BCI Competition IV and MI-2 datasets.

D. *Ablation study*

We conduct an ablation study to evaluate the effectiveness of utilizing source data, domain adaptation framework, filter bank and attention network. Table IV shows the comparison results on one block of training data from session one of dataset 2a. Six folder cross validation is conducted. Model 1 is our basic feature extractor-classification network trained with one target subject's data. Model 2 is our basic network trained with data from both target and source subjects.

One-way repeated-measures ANOVA revealed a significant effect of different models ($p<0.001$) on the decoding accuracy. The classification accuracy of our basic network trained with both target and source data (model 2) is significantly higher than that of the basic network trained with only target data (model 1) ($p < 0.05$). After adopting the filter bank strategy, the performance of the network trained with both target and source data (model 5) is not different from that trained with target data (model 4) ($p=0.136$). This indicates that data from source subjects are not helpful for the classification performance of our filter bank network.

With the filter bank strategy, our basic network trained with target data (model 4), our basic network trained with target and source data (model 5) and our basic network with adversarial domain adaptation (model 6) perform better than those without the filter bank (models 1, 2, and 3) ($p<0.01$). This indicates that the filter bank can effectively increase the model performance. Moreover, our domain adaptation method can enhance decoding performance (model 6 vs. model 5, $p<0.01$). Furthermore, the network after adding the attention network improves decoding accuracy (model 7 vs. model 6, $p<0.05$) and achieves the best accuracy. In summary, each part of our model has a significant impact on the decoding performance.

Table V and Table VI show the ablation study results on session two of dataset 2a and the MI-2 dataset. For MI-2 dataset, only the mean results are presented. Accuracy for each subject is presented in the Supplementary Information.

E. *Visualization results*

We applied t-distributed stochastic neighbor embedding (t-SNE) to explore the effect of our adversarial training method. Fig. 2 shows an example from one target subject and the corresponding source subject with one block of training data from the target subject. All data of the target subject and all data of the source subject are plotted. Fig. 2 (a) is the distribution of raw data. Either the source domain and target domain or different classes have maximum overlap. Fig. 2 (b) shows that features extracted by our basic feature extractor network trained without adversarial training are separated into clusters. However, the distribution between the two domains does not overlap well. There exists an obvious distribution gap. Fig. 2 (c) shows the distribution of features extracted by the model trained with our domain adaptation method. The distribution between the two domains is more consistent. The distribution of samples from the same class between different domains is closer. We

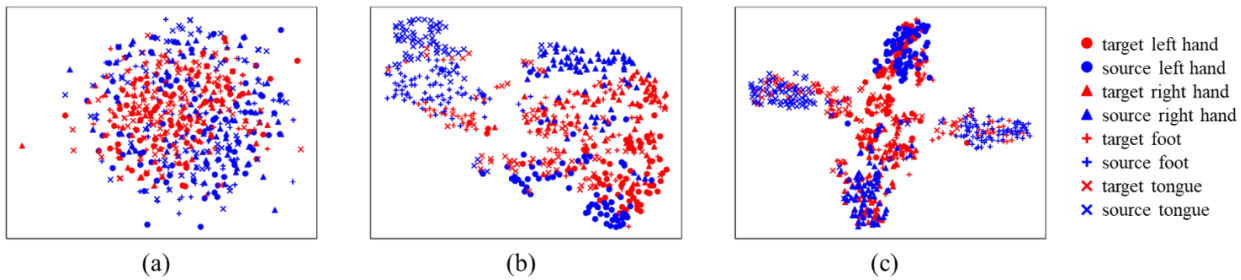

Fig. 2. Visualization of data distribution. Figure (a) is the distribution of raw EEG data. Figure (b) is the distribution of feature extracted by our model without domain adaptation. Figure (c) is the feature extracted by our model with domain adaptation.

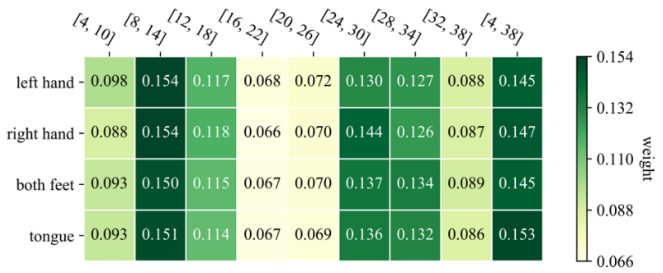

Fig. 3. Visualization of band weights that output by our proposed attention model

observed that samples from different classes were separable. These results demonstrate that our adversarial training framework can effectively align the target and source distributions.

We further visualize the band weights output by our proposed attention network to explore how the attention network works for MI decoding. Fig. 3 shows the average band weights of one run from the test set of a target subject. The model is trained with one block of target data and all data from the corresponding source subject. We can see large weights for alpha and high-beta bands, which is consistent with the EEG response during MI tasks. This reveals that EEG data in these frequency bands provide useful information for classification. Additionally, the whole frequency band (4-38 Hz) attains large weights. Moreover, the different classes are slightly different from each other. At 8-14 Hz, left- and right-hand MI tasks have larger weights than feet and tongue MI tasks. In 4-38 Hz, the feet and tongue tasks receive larger weights. These results indicate that the attention mechanism can adaptively assign different weights to each frequency band based on different input samples.

F. Source selection

In this study, one source subject is necessary to train our DWFBDA model. We select m_s best subjects as the source set ($m_s = 1,2,3,\ldots,8$) and then choose one best fitted subject from the source set as the final source subject. Fig. 4 displays the classification accuracy of our method for different source sets. Repeated-measures ANOVA shows a significant effect of different strategies on accuracy ($p<0.05$). The classification accuracy is the worst when the source subject is randomly selected. This indicates that the source selection strategy is vital for decoding performance. With the source selection strategy, the averaged decoding accuracy increased gradually from one to five best sources and became stable from the five best sources. Therefore, the source set with five optional source subjects is used for training the domain adaptation model.

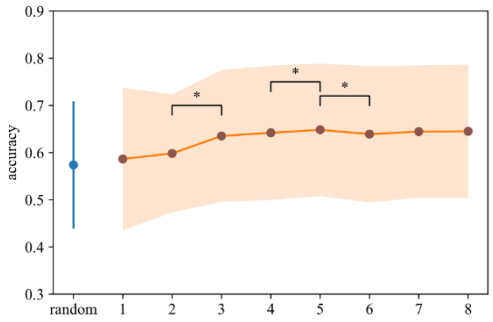

Fig. 4. Decoding accuracy of different source selection strategy. For horizontal axis, 'random' refers to case where source subject is randomly selected and numbers indicate how many optional source subjects is used. '*' indicate that in paired t-test $p<0.05$.

V. Discussion

In this study, we proposed a dynamic weighted filter bank domain adaptation framework to classify MI tasks through already collected data from the source subject and a small amount of data from the target subject. Our study focuses on a training model with less training data. In our experiment, we test the model performance with different amounts of training data. Our method outperforms the compared MI decoding methods on dataset 2a of the BCI competition IV and MI-2 datasets. Additionally, our method can achieve similar MI decoding performance with one fewer block of training data.

A. The proposed DWFBDA network

Our framework consists of three models: a filter bank strategy, a dynamic weight model based on an attention network and a WGAN-based adversarial training network. The filter bank strategy has been proven to be effective in many traditional EEG decoding methods[31, 54]. Particularly, in MI tasks, FBCSP, which combines CSP and filter bank strategy, shows a great improvement on MI classification tasks compared to CSP. In this study, the filter bank strategy is adopted in our MI decoding model. Ablation experiments show that the filter bank can effectively improve the decoding accuracy (Table IV). This may be because the filter bank can mine information from multiple frequency bands thoroughly and ensemble multiband results. It is worth mentioning that the filter bank strategy is also efficient for decoding SSVEP signals. Traditional CCA focuses on fundamental frequencies, while FBCCA can focus on not only fundamental frequencies but also high harmonic frequencies. FBCCA grasps more useful information for decoding. In addition, FBCCA benefits from the advantage of ensemble learning and fuses the prediction results from all frequency bands[55].

Generally, bandpass filters and 1-dimensional convolutional layers both conduct convolution operations on a temporal sequence. However, their parameters were obtained in different

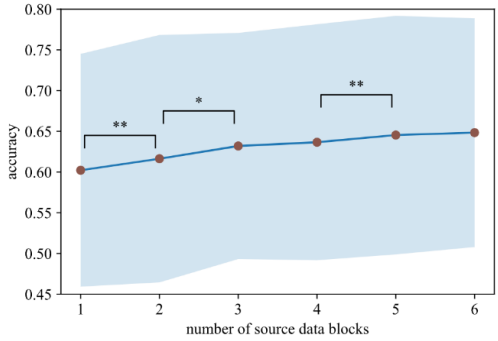

Fig. 5. Decoding accuracy with different amount of source data blocks. '*', '**' indicate that in paired t-test $p<0.05$ and $p<0.01$respectively.

ways. The coefficients of the bandpass filters are computed through specific methods, such as the Butterworth method[56] and Chebyshev[57] method, while the coefficients of the 1-D convolutional layers are learned from the training data in a data-driven manner. Some DL-based MI decoding methods, for example, shallow CNN and EEGNet, employ a temporal convolutional layer to simulate the performance of bandpass filters[22]. However, these temporal filters in CNN-based decoding models may extract time patterns and can hardly learn to mimic bandpass filters. Thus, they have difficult mining information from different frequency bands. In our framework, we explicitly deploy a filter bank and push the network to extract features from multiple frequency bands. Therefore, our framework is more capable of mining multiband information.

We proposed a WGAN-based domain adaptation network to utilize source data from the other subject to support the training of the decoding model for the target subject. Ablation study shows that our domain adaptation network significantly enhances the decoding performance. Furthermore, visualization of the feature distribution shows the effectiveness of our domain adaptation network. The advantage of our domain adaptation network depends on two aspects: the Wasserstein distance and domain adaptation separately applied for each class. In our case, MI EEG data have a large distribution gap between subjects. As shown in Fig 2 b, the distribution between two subjects could be very different and have no overlap for some classes. Additionally, we only use a small amount of data from one target subject, which further increases the risk that distribution mismatches. Compared with GAN, WGAN can measure distribution differences even when two distributions have no overlap by the Wasserstein distance[46]. Thus, WGAN is suitable for our case and better aligns the distributions. The second advantage of our domain adaptation network is that we align the distribution for each class separately. Even when the marginal distribution has been well aligned, it is difficult to train a classifier that can discriminate classes well for both subjects with a large conditional distribution difference. Thus, we draw the distribution between EEGs of two subjects for each class separately to align the conditional distribution between the target and source. Therefore, with domain adaptation for each class in our framework, source data can assist the training of the decoding model for the target subject effectively and enhance the decoding performance.

B. Difference between our method and compared methods

In our experiments, FBCSP and FBCNet achieved the best decoding accuracy among traditional methods and deep learning-based methods, respectively. Both methods adopted filter bank. This shows that the filter bank is a powerful tool in MI-BCI that helps the model to better explore the frequency information. However, both FBCSP and FBCNet did not achieve the best decoding accuracy with one block of training data. This may be because that many training data are needed to fully take advantage of the filter bank. Our domain adaptation method utilizes the existing training data from a source subject, thus allowing us to take advantage of filter bank with limited training data from the target subject.

We compared four domain adaptation methods, including RPA, CCSP, DRBDA, and OTDA. Thereinto, OTDA, originally proposed on the cross-session domain adaptation of MI, did not achieve a similar high decoding accuracy in our experimental setting. This may be caused by the large difference of the CSP matrix between different people. OTDA could work better in a cross-session domain adaptation paradigm rather than a cross-subject domain adaptation paradigm.

CCSP, RPA and OTDA focus on the spatial features of EEG signals and ignore the temporal features. DRBDA, a deep learning-based domain adaptation network, extracts deep representations of EEG signals from both spatial and temporal perspectives. The experimental results show that DRBDA generally achieves a better decoding accuracy than CSP, RPA and OTDA. Compared with DRBDA, our method uses a class-specific domain adaptation that can align the conditional distribution of the target and source, while DRBDA only algins the marginal distribution. We use WGAN-based adversarial domain adaptation, which works better when training data are limited. In addition, we adopt a filter bank and a dynamic weight model to fully explore the frequency information that lies in the EEG signal. Therefore, our method obtains higher decoding accuracy, especially when the training data are short.

C. Effect of the amount of source data

In our experiments, we used all available data from the source subject to train our proposed model. To investigate the effect of the amount of data from the source subject on the performance of our model. We conduct the following experiment using one block of target data and different amounts of source data (one, two, three…six blocks) to train our model. The rest of the target data are used as test data. The decoding accuracy is shown in Fig. 5. Repeated-measures ANOVA revealed that the amount of source data had a significant main effect on the classification accuracy ($p<0.001$). Classification accuracy increases as source data increase from one block to three blocks (both $p<0.05$), increases but not significantly from three blocks to four blocks ($p=0.15$) and significantly increases from four to five blocks ($p<0.01$). When increasing the source data from five blocks to six blocks, the decoding accuracy only increases by 0.3% ($p=0.259$). These results show that decoding accuracy reaches a relatively stable level when source data reach five blocks. In our comparative experiments, using six blocks of source data is sufficient for training our domain adaptation model.

D. Simulated online experiment

We use cross validation in our experiments to fully exploit the dataset and evaluate the proposed method. Cross validation has also been adopted by many MI decoding studies [23, 33, 34]. However, the EEG signals of MI tasks also suffer from nonstationarity over time. EEG signals from the same block or session may share similar patterns. Cross validation will ignore

TABLE VII COMPARISON OF DIFFERENT FREQUENCY RANGE (IN PERCENTAGE %, '*', '**' REPRESENTS COMPARED WITH OUR METHOD P<0.05 AND P<0.01 RESPECTIVELY)

Band	Method	Training data amounts (blocks)				
		1	2	3	4	5
8-30 Hz	FBCNet	51.35**	60.53*	64.76*	68.40	69.83*
	DRBDA	56.40**	61.20**	63.34**	66.15**	67.36**
	ours	62.30	67.53	69.70	71.74	74.31
4-38 Hz	FBCNet	50.91**	62.22**	67.19**	70.99	71.95*
	DRBDA	55.96**	58.76**	61.91*	67.09**	69.06**
	ours	64.84	69.92	73.01	74.61	76.62

time nonstationary problems and may result in higher decoding accuracy. To test whether our proposed method works well in a real application situation, we also perform the following experiment. We select the first 48 samples of each target subject from session 1 of dataset 2a as target training data, and the rest of the samples are taken as test data. In this manner, we simulate the situation where a new target subject comes to use the BCI system, and we collect only a small amount of training data for this subject. The average decoding accuracy is 62.36%. Indeed, this is lower than the cross validation accuracy on session 1 of dataset 2a, but it is still significantly higher than the cross validation accuracy of all compared methods (paired t test, all $p < 0.05$) (Table I). This suggests that our proposed method may work well in real MI-BCI application scenes.

E. Effect of frequency range selection

The selection of frequency range is important in MI decoding. Although the main EEG response of MI lies in the alpha and beta bands, mostly from 8-30 Hz, frequency bands near 8-30 Hz also contain some valuable information for decoding[50-52]. We compare the performance of two frequency bands (8-30 Hz and 4-38 Hz) through our method, FBCNet (the best deep learning method among the compared methods), and DRBDA (the best transfer learning method among the compared methods) on session one of dataset 2a. The results are presented in table VII.

The three-way repeated-measures ANOVA showed significant main effects of method ($p<0.01$), data amounts ($p<0.01$) and frequency ($p<0.05$) on decoding performance. For FBCNet, the decoding accuracy between the two frequency bands is not significantly different with each amount of training data. For DRBDA, decoding accuracy between two frequency bands is only significantly different with four and five blocks of training data (4-38 Hz > 8-30 Hz, $p<0.05$). Our method obtains significantly better decoding accuracy in the frequency range of 4-38 Hz than 8-30 Hz with each data amount (all $p<0.05$). Therefore, it is feasible for our proposed method to use a filter range of 4-38 Hz. This is because we use a filter bank and an attention network to automatically assign weights for different frequency bands. Our network is more capable of capturing information from a wider frequency range.

In addition, when using the 8-30 Hz frequency band, our method still achieves significantly better decoding accuracy than DRBDA with each amount of training data (all $p<0.05$). Our method achieves significantly better decoding accuracy than FBCNet with one, two, three and five blocks of training data (all $p<0.05$). Our method tends to achieve better decoding accuracy than FBCNet ($p=0.08$) with four blocks of training data.

F. Future works

Although our proposed method enhanced the decoding accuracy of MI-BCI with a short amount of training data, the decoding accuracy was still relatively low. There is still much work to do to realize a convenient and practical MI-BCI. First, future work should propose an algorithm to adaptively update the model parameters with online EEG data to prevent the performance decline of the decoding model over a long period of time. Second, in future studies, we could employ multimodal recordings, such as EEG and fNIRS, in combination with deep learning methods to further improve the decoding performance of MI-BCI.

CONCLUSION

In this paper, we proposed a dynamic weighted filter bank domain adaptation framework to improve the classification accuracy for MI-BCIs. Experiments on both the public dataset and self-collected dataset demonstrate that our method could achieve the best performance with the same amount of training data compared with existing methods. In particular, our method enhances the decoding accuracy when the target subject has only one block of training data. This study indicates that our method can enhance MI decoding performance with a short calibration time.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 61976209, Grant 81701785, and Grant 61906188 and in part by the Strategic Priority Research Program of CAS under Grant XDB32040200. (Corresponding author: Huiguang He).

REFERENCES

[1] WOLPAW J R, BIRBAUMER N, HEETDERKS W J, et al. Brain-computer interface technology: a review of the first international meeting [J]. IEEE Trans Rehabil Eng, 2000, 8(2): 164-73.

[2] TIWARI N, EDLA D R, DODIA S, et al. Brain computer interface: A comprehensive survey [J]. Biologically Inspired Cognitive Architectures, 2018, 26(118-29.

[3] WOLPAW J R, BIRBAUMER N, MCFARLAND D J, et al. Brain–computer interfaces for communication and control [J]. Clinical neurophysiology, 2002, 113(6): 767-91.

[4] ZIMMERMANN-SCHLATTER A, SCHUSTER C, PUHAN M A, et al. Efficacy of motor imagery in post-stroke rehabilitation: a systematic review [J]. Journal of NeuroEngineering and Rehabilitation, 2008, 5(1): 8.

[5] MENG J, ZHANG S, BEKYO A, et al. Noninvasive Electroencephalogram Based Control of a Robotic Arm for Reach and Grasp Tasks [J]. Sci Rep, 2016, 6(38565.

[6] I BADIA S B, MORGADE A G, SAMAHA H, et al. Using a hybrid brain computer interface and virtual reality system to monitor and promote cortical reorganization through motor activity and motor imagery training [J]. IEEE Trans Neural Syst Rehabil Eng, 2012, 21(2): 174-81.

[7] ASENSIO-CUBERO J, GAN J Q, PALANIAPPAN R. Multiresolution analysis over graphs for a motor imagery based online BCI game [J]. Computers in biology and medicine, 2016, 68(21-6.

[8] KOSMYNA N, TARPIN-BERNARD F, BONNEFOND N, et al. Feasibility of BCI control in a realistic smart home environment [J]. Frontiers in human neuroscience, 2016, 10(416.

[9] JAIS A A, MANSOR W, LEE K Y, et al. Motor imagery EEG analysis for home appliance control; proceedings of the 2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA), F, 2017 [C]. IEEE.

[10] KIM H-J, LEE M-H, LEE M. A BCI based smart home system combined with event-related potentials and speech imagery task; proceedings of the 2020 8th International Winter Conference on Brain-Computer Interface (BCI), F, 2020 [C]. IEEE.

[11] MENG J, STREITZ T, GULACHEK N, et al. Three-dimensional brain–computer interface control through simultaneous overt spatial attentional and motor imagery tasks [J]. IEEE Transactions on Biomedical Engineering, 2018, 65(11): 2417-27.

[12] BHATTACHARYYA S, KONAR A, TIBAREWALA D. Motor imagery and error related potential induced position control of a robotic arm [J]. IEEE/CAA Journal of Automatica Sinica, 2017, 4(4): 639-50.

[13] RAKSHIT A, KONAR A, NAGAR A K. A hybrid brain-computer interface for closed-loop position control of a robot arm [J]. IEEE/CAA Journal of Automatica Sinica, 2020, 7(5): 1344-60.

[14] LIU Y, SU W, LI Z, et al. Motor-imagery-based teleoperation of a dual-arm robot performing manipulation tasks [J]. IEEE Transactions on Cognitive and Developmental Systems, 2018, 11(3): 414-24.

[15] LOTTE F, BOUGRAIN L, CICHOCKI A, et al. A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update [J]. J Neural Eng, 2018, 15(3): 031005.

[16] RAMOSER H, MULLER-GERKING J, PFURTSCHELLER G. Optimal spatial filtering of single trial EEG during imagined hand movement [J]. IEEE Transactions on Rehabilitation Engineering, 2000, 8(4): 441-6.

[17] SONG X M, YOON S C, PERERA V, et al. Adaptive Common Spatial Pattern for Single-Trial EEG Classification in Multisubject BCI [M]. 2013 6th International Ieee/Embs Conference on Neural Engineering. 2013: 411-4.

[18] ZHAO Q B, ZHANG L Q, CICHOCKI A, et al. Incremental Common Spatial Pattern Algorithm for BCI [M]. 2008 Ieee International Joint Conference on Neural Networks, Vols 1-8. 2008: 2656-+.

[19] THOMAS K P, GUAN C, LAU C T, et al. A new discriminative common spatial pattern method for motor imagery brain–computer interfaces [J]. IEEE Transactions on Biomedical Engineering, 2009, 56(11): 2730-3.

[20] AGHAEI A S, MAHANTA M S, PLATANIOTIS K N. Separable common spatio-spectral patterns for motor imagery BCI systems [J]. IEEE Transactions on Biomedical Engineering, 2015, 63(1): 15-29.

[21] CHAKRABORTY B, GHOSH L, KONAR A. Designing phase-sensitive common spatial pattern filter to improve brain-computer interfacing [J]. IEEE Transactions on Biomedical Engineering, 2019, 67(7): 2064-72.

[22] SCHIRRMEISTER R T, SPRINGENBERG J T, FIEDERER L D J, et al. Deep learning with convolutional neural networks for EEG decoding and visualization [J]. Hum Brain Mapp, 2017, 38(11): 5391-420.

[23] LAWHERN V J, SOLON A J, WAYTOWICH N R, et al. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces [J]. J Neural Eng, 2018, 15(5): 056013.

[24] ZHANG D L, YAO L N, ZHANG X, et al. Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-Based Intention Recognition for Brain Computer Interface [J]. Thirty-Second Aaai Conference on Artificial Intelligence / Thirtieth Innovative Applications of Artificial Intelligence Conference / Eighth Aaai Symposium on Educational Advances in Artificial Intelligence, 2018, 1703-10.

[25] LU N, LI T, REN X, et al. A deep learning scheme for motor imagery classification based on restricted Boltzmann machines [J]. IEEE Trans Neural Syst Rehabil Eng, 2016, 25(6): 566-76.

[26] WANG P, JIANG A, LIU X, et al. LSTM-based EEG classification in motor imagery tasks [J]. IEEE Trans Neural Syst Rehabil Eng, 2018, 26(11): 2086-95.

[27] WU D, XU Y, LU B-L. Transfer learning for EEG-based brain-computer interfaces: a review of progress made since 2016 [J]. IEEE Transactions on Cognitive and Developmental Systems, 2020,

[28] WEISS K, KHOSHGOFTAAR T M, WANG D. A survey of transfer learning [J]. Journal of Big Data, 2016, 3(1):

[29] HYOHYEONG K, YUNJUN N, SEUNGJIN C. Composite Common Spatial Pattern for Subject-to-Subject Transfer [J]. IEEE Signal Processing Letters, 2009, 16(8): 683-6.

[30] PEDRO, LUIZ, COELHO, et al. Riemannian Procrustes Analysis: Transfer Learning for Brain-Computer Interfaces [J]. IEEE Transactions on Bio Medical Engineering, 2018,

[31] KAI KENG A, ZHANG YANG C, HAIHONG Z, et al. Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface, F 2008, 2008 [C]. IEEE.

[32] MüLLER-GERKING J, PFURTSCHELLER G, FLYVBJERG H. Designing optimal spatial filters for single-trial EEG classification in a movement task [J]. Clinical Neurophysiology, 1999, 110(5): 787-98.

[33] ANG K K, CHIN Z Y, WANG C, et al. Filter Bank Common Spatial Pattern Algorithm on BCI Competition IV Datasets 2a and 2b [J]. Front Neurosci, 2012, 6(39.

[34] BARACHANT A, BONNET S, CONGEDO M, et al. Multiclass Brain-Computer Interface Classification by Riemannian Geometry [J]. IEEE Transactions on Biomedical Engineering, 2012, 59(4): 920-8.

[35] MIAO Y, JIN J, DALY I, et al. Learning common time-frequency-spatial patterns for motor imagery classification [J]. IEEE Trans Neural Syst Rehabil Eng, 2021, 29(699-707.

[36] SAKHAVI S, GUAN C, YAN S. Learning Temporal Information for Brain-Computer Interface Using Convolutional Neural Networks [J]. IEEE Trans Neural Netw Learn Syst, 2018, 29(11): 5619-29.

[37] LI Y, GUO L, LIU Y, et al. A temporal-spectral-based squeeze-and-excitation feature fusion network for motor imagery EEG decoding [J]. IEEE Trans Neural Syst Rehabil Eng, 2021, 29(1534-45.

[38] LIU X, LV L, SHEN Y, et al. Multiscale space-time-frequency feature-guided multitask learning CNN for motor imagery EEG classification [J]. Journal of Neural Engineering, 2021, 18(2): 026003.

[39] MANE R, CHEW E, CHUA K, et al. FBCNet: A multi-view convolutional neural network for brain-computer interface [J]. arXiv preprint arXiv:210401233, 2021,

[40] HE H, WU D. Transfer learning for Brain–Computer interfaces: A Euclidean space data alignment approach [J]. IEEE Transactions on Biomedical Engineering, 2019, 67(2): 399-410.

[41] SAKHAVI S, GUAN C T, IEEE. Convolutional Neural Network-based Transfer Learning and Knowledge Distillation using Multi-Subject Data in Motor Imagery BCI [M]. 2017 8th International Ieee/Embs Conference on Neural Engineering. New York; Ieee. 2017: 588-91.

[42] DOSE H, MøLLER J S, IVERSEN H K, et al. An end-to-end deep learning approach to MI-EEG signal classification for BCIs [J]. Expert Systems with Applications, 2018, 114(532-42.

[43] JEON E, KO W, SUK H I, et al. Domain Adaptation with Source Selection for Motor-Imagery based BCI [M]. 2019 7th International Winter Conference on Brain-Computer Interface. New York; Ieee. 2019: 134-7.

[44] ZHAO H, ZHENG Q, MA K, et al. Deep Representation-Based Domain Adaptation for Nonstationary EEG Classification [J]. IEEE transactions on neural networks and learning systems, 2020, PP

[45] PETERSON V, NIETO N, WYSER D, et al. Transfer learning based on optimal transport for motor imagery brain-computer interfaces [J]. IEEE Transactions on Biomedical Engineering, 2022, 69(2): 807-17.

[46] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein gan [J]. arXiv preprint arXiv:170107875, 2017,

[47] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need [M]//GUYON I, LUXBURG U V, BENGIO S, et al. Advances in Neural Information Processing Systems 30. 2017.

[48] TANGERMANN M, MüLLER K-R, AERTSEN A, et al. Review of the BCI competition IV [J]. Frontiers in neuroscience, 2012, 6(55.

[49] MA X, QIU S, HE H. Multi-channel EEG recording during motor imagery of different joints from the same limb [J]. Scientific Data, 2020, 7(1):

[50] GROSSE-WENTRUP M, SCHöLKOPF B, HILL J. Causal influence of gamma oscillations on the sensorimotor rhythm [J]. NeuroImage, 2011, 56(2): 837-42.

[51] AHN M, AHN S, HONG J H, et al. Gamma band activity associated with BCI performance: simultaneous MEG/EEG study [J]. Frontiers in human neuroscience, 2013, 7(848.

[52] TRAMBAIOLLI L R, DEAN P J, CRAVO A M, et al. On-task theta power is correlated to motor imagery performance; proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), F, 2019 [C]. IEEE.

[53] GRAMFORT A, LUESSI M, LARSON E, et al. MEG and EEG data analysis with MNE-Python [J]. Frontiers in neuroscience, 2013, 7(267.

[54] NAKANISHI M, WANG Y, CHEN X, et al. Enhancing Detection of SSVEPs for a High-Speed Brain Speller Using Task-Related Component Analysis [J]. IEEE Trans Biomed Eng, 2018, 65(1): 104-12.
[55] CHEN X, WANG Y, NAKANISHI M, et al. High-speed spelling with a noninvasive brain-computer interface [J]. Proc Natl Acad Sci U S A, 2015, 112(44): E6058-67.
[56] SELESNICK I W, BURRUS C S. Generalized digital Butterworth filter design [J]. IEEE Transactions on signal processing, 1998, 46(6): 1688-94.
[57] KARAM L J, MCCLELLAN J H. Chebyshev digital FIR filter design [J]. Signal processing, 1999, 76(1): 17-36.