IFNet: An Interactive Frequency Convolutional Neural Network for Enhancing Motor Imagery Decoding From EEG

Jiaheng Wang[ID], Lin Yao[ID], and Yueming Wang[ID]

Abstract—**Objective: The key principle of motor imagery (MI) decoding for electroencephalogram (EEG)-based Brain-Computer Interface (BCI) is to extract task-discriminative features from spectral, spatial, and temporal domains jointly and efficiently, whereas limited, noisy, and non-stationary EEG samples challenge the advanced design of decoding algorithms. Methods: Inspired by the concept of cross-frequency coupling and its correlation with different behavioral tasks, this paper proposes a lightweight Interactive Frequency Convolutional Neural Network (IFNet) to explore cross-frequency interactions for enhancing representation of MI characteristics. IFNet first extracts spectro-spatial features in low and high-frequency bands, respectively. Then the interplay between the two bands is learned using an element-wise addition operation followed by temporal average pooling. Combined with repeated trial augmentation as a regularizer, IFNet yields spectro-spatio-temporally robust features for the final MI classification. We conduct extensive experiments on two benchmark datasets: the BCI competition IV 2a (BCIC-IV-2a) dataset and the OpenBMI dataset. Results: Compared with state-of-the-art MI decoding algorithms, IFNet achieves significantly superior classification performance on both datasets while improving the winner's result in BCIC-IV-2a by 11%. Moreover, by conducting sensitivity analysis on decision windows, we show IFNet attains the best trade-off between decoding speed and accuracy. Detailed analysis and visualization verify IFNet can capture the coupling across frequency bands along with the known MI signatures. Conclusion: We demonstrate the effectiveness and superiority of the proposed IFNet for MI decoding. Significance: This study suggests IFNet holds promise for rapid response and accurate control in MI-BCI applications.**

Index Terms— **Brain–computer interface, motor imagery, cross-frequency interactions, convolutional neural networks, data augmentation.**

I. INTRODUCTION

BRAIN-computer interface (BCI) enables direct communication and control between a brain and a machine [1], wherein electroencephalogram (EEG) is the mostly used neurophysiological signal modality in noninvasive BCI systems. There are varieties of EEG-based BCI paradigms wherein motor imagery (MI) representing a typical type of self-induced mental activity has been widely investigated for various purposes [2], [3], [4]. With the fast development of BCI technology referring to signal acquisition, signal processing, and machine learning techniques, MI-BCI has shown feasibility and effectiveness for real-world applications such as post-stroke motor rehabilitation [5], 2D continues control [6], [7], 3D Quadcopter control [8], and so on. To prop up these applications, accurate and fast MI decoding plays a vital role in MI-BCI.

As a common practice in the BCI community, accurate algorithms for decoding intrinsic brain states are principally guided by neurophysiological priors along with advanced machine learning techniques. Considering MI classification, it is well known that task-specific MI will evoke event-related desynchronization (ERD) and event-related synchronization (ERS) of sensory-motor rhythms in distinct frequency bands and brain areas [9]. Targeting at those neural signatures, traditional hand-crafted MI features are well designed using diverse machine learning approaches. For example, common spatial pattern (CSP) [10] is widely adopted for 2-class MI feature extraction. Filter-bank common spatial pattern (FBCSP) [11] further harnesses CSP features in multi-bands together with mutual information-based feature selection methods. In addition, Riemannian geometry-based methods provide new tools

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institute for Knowledge Discovery, Graz University of Technology (for approving the BCIC-IV-2a dataset), and the Korea University Institutional Review Board (for approving the OpenBMI dataset).

Jiaheng Wang is with the MOE Frontiers Science Center for Brain and Brain–Machine Integration and the College of Computer Science, Zhejiang University, Hangzhou 310027, China.

Lin Yao is with the Department of Neurobiology, Affiliated Mental Health Center and Hangzhou Seventh People's Hospital, Zhejiang University School of Medicine, Hangzhou 310030, China, and also with the MOE Frontiers Science Center for Brain and Brain–Machine Integration and the College of Computer Science, Zhejiang University, Hangzhou 310027, China (e-mail: lin.yao@zju.edu.cn).

Yueming Wang is with the Qiushi Academy for Advanced Studies (QAAS), College of Computer Science, Zhejiang University, Hangzhou 310027, China (e-mail: ymingwang@zju.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2023.3257319, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3257319

for building simple and accurate classification models [12]. Nevertheless, most traditional approaches have limited capability to extract spectro-spatio-temporally informative features and they are mainly optimized for binary classification. Besides, they suffer from the high susceptibility to inter-trial and inter-subject variability. To enhance the utility of decoding models, transfer learning-based approaches are typically investigated in previous studies [13], [14].

Apart from spectral power variations in response to different MI tasks, previous studies have suggested cross-frequency coupling (CFC) might play a functional role in sensory, motor, and cognitive tasks [15], [16]. CFC represents the association of multiple frequency neural oscillations, that is, one frequency band modulates the activity of a different frequency band [17]. Recent studies have observed CFC characteristics during movement-related mental tasks [18]. Moreover, strong Phase-Amplitude Coupling between alpha and high gamma was observed during motor planning and this feature vanished during motor execution [19]. In another study, [20] demonstrated the effectiveness of motor-related CFC features for a 4-class MI-BCI system, suggesting interactions among multi-bands provide an additional way for discriminative representation of MI tasks. To summarize, we refer to cross-frequency interactions representing complex interplay across different frequency bands from different electrodes. Although cross-frequency interactions are widely investigated to account for neural mechanisms, they have been rarely considered for MI decoding.

Deep learning-based approaches alleviate the necessity of hand-crafted features in an end-to-end learning manner [21]. On the one hand, it is of great interest to explore spectro-spatio-temporal representation in an effective and efficient manner. On the other hand, limited, noisy, and non-stationary EEG samples obscure the effective learning of deep learning models. To that effect, state-of-the-art neural network models principally share a similar spirit with traditional approaches, that is, they are inspired by neurophysiological priors and enjoy the advancement of neural networks [22], [23]. For example, EEG activities are transformed into a sequence of topology-preserving multi-spectral images and fed into a deep recurrent CNN [24]. Due to the local receptive field of the convolution operator, their network is difficult in capturing spectral interactions among long-range electrodes. In another study, [25] proposed a compact CNN as a general-purpose EEG decoder, namely EEGNet. EEGNet learns frequency filters entirely by backpropagation, which omits hand-crafted spectral bands that might be beneficial for mining mutual spectral information. More recently, channel group attention was introduced in [26] to deal with filter bank inputs. Different from the perspective of band interaction, their inter-channel attention targets scaling the feature channels to improve the expression ability of representative features in all bands. Yet, most deep learning models learn cross-frequency interactions implicitly, resulting in increased susceptibility to the inter-trial variance of MI tasks.

Lastly, fast MI decoding also plays an important role in real-world applications. Specifically, a practical MI-BCI system should respond accurately to users' mental intentions while as fast as possible. On the other side of the coin, a shorter window size commonly results in lower decoding accuracy due to high intra-trial variance of MI response. Thus, it is necessary to investigate models' sensitivity to decision window length. Apart from accuracy evaluation on different window sizes, we leverage information transfer rate which incorporates both speed and accuracy in a single value and is widely used for measuring communication and control systems [27].

With the purpose of accurate and fast MI decoding, in this paper, we propose a lightweight Interactive Frequency Convolutional Neural Network (IFNet) by combing neurophysiological priors with efficient convolutional architectures. The main contributions of this paper are as follows:

- We propose a lightweight Interactive Frequency Convolutional Neural Network (IFNet) to explore cross-frequency interactions for enhancing spectro-spatio-temporally robust representation of MI tasks.
- We introduce repeated trial augmentation as a regularizer to ease overfitting caused by limited and noisy EEG samples during network training.
- We conduct extensive experiments against state-of-the-art MI decoding algorithms on two benchmark datasets, demonstrating the effectiveness and superiority of the proposed IFNet. Moreover, it achieves the best trade-off between decoding speed and accuracy. Last but not least, we verify IFNet learns neuro-physiologically interpretable features.

The code is publicly available at `https://github.com/Jiaheng-Wang/IFNet`.

II. RELATED WORKS

A. MI Decoding Algorithms

MI decoding remains a hotspot in the EEG-based BCI field in past decades. A thorough survey on traditional machine learning algorithms for MI decoding from EEG can be found in [28]. Among them, FBCSP was adopted by the winning entry of BCIC-IV-2a and is widely used for comparison with deep learning models. Deep learning-based approaches have shown promising results and consistent progress in MI-BCI. Among them, EEGNet was first proposed as a general-purpose lightweight network for a wide range of EEG-based BCI paradigms. More recently, a novel Filter-Bank Convolutional Network (FBCNet) [29] achieved the highest classification accuracy on the BCIC-IV-2a dataset. It employs a filter bank data representation followed by spatial filtering to extract spectro-spatially discriminative features. In contrast to EEGNet, FBCNet abandons temporal filters and completely extracts spectrally localized signals manually. Both of them have provided open-source codes for better reproducibility. Therefore, these two models serve as baseline methods in our experiments. Additionally, in terms of recent advancements in network architecture design, multi-path representation and cross-feature interactions have gained much attention in the computer vision community [30], [31]. With the above considerations, we propose IFNet targeting cross-frequency interactions in an effective and efficient manner.
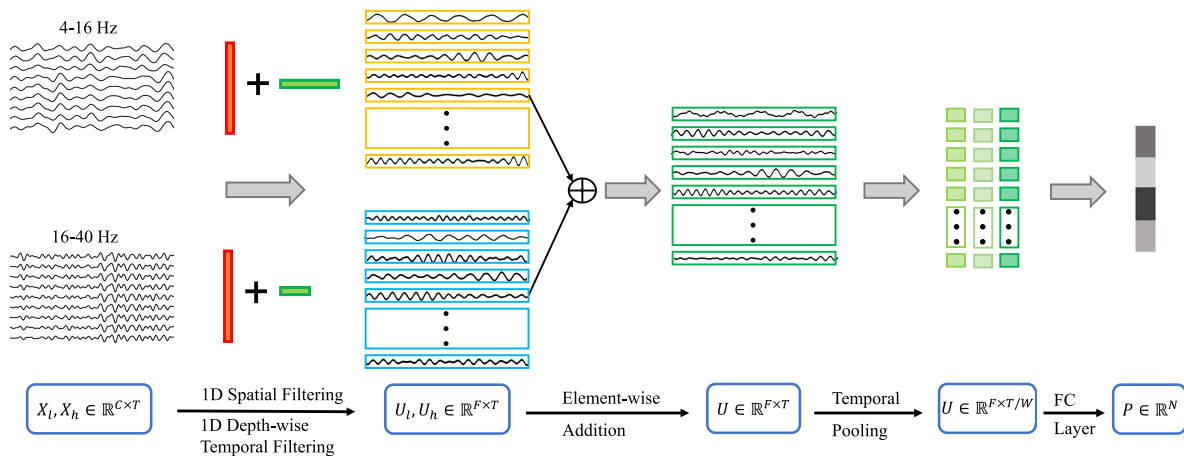
Fig. 1. Architecture of our proposed IFNet implemented with 1D convolution (X: input EEG signals, U: feature maps, C: number of EEG channels, T: number of time points, F: number of spatial filters, W: temporal pooling size, N: number of classes).

B. Data Augmentation and Regularization

Regularization is a common practice aimed at preventing overfitting in the training of neural networks. From the point of view of input data, data augmentation is an explicit form of regularization by means of artificially enlarging the training dataset from existing data using various translations. As for MI-BCI, high intra-class variability in MI signatures, as well as limited, noisy, and non-stationary EEG samples result in unstable training and low generalization ability of deep learning models. Meanwhile, it is difficult to design an effective data augmentation strategy due to indiscernible task-specific information incorporated in raw EEG signals. While noise adding is commonly utilized with neural networks in the BCI field, the performance is marginal owing to low signal-to-noise ratio of generated EEG samples. Following the idea of batch augmentation proposed for image recognition [32], we introduce repeated trial augmentation consisting of random cropping and random erasing to strengthen the learning ability of CNNs for EEG decoding.

III. METHODS

A. Interactive Frequency Convolutional Neural Network

To build an efficient network architecture while incorporating neurophysiological priors of MI-EEG, we first conduct a preliminary experiment on a representative network named EEGNet to guide the compact yet powerful design of IFNet. Detailed analysis is presented in Appendix A. We empirically demonstrate the first block is the major determinant of EEGNet and its learning capability can be strengthened by width scaling. More importantly, in the case of $D = 1$ in the first block of EEGNet, we find it can be implemented more efficiently with 1D convolution by reversing the sequence of temporal and spatial filtering. Altogether, inspired by the concept of CFC as well as the efficient design of EEGNet, we propose IFNet to capture spectro-spatial-temporally discriminative features effectively and efficiently. IFNet is composed of three stages: spectro-spatial feature representation, cross-frequency interactions, and classification. The architecture of IFNet is

illustrated in Fig. 1 and summarized in Appendix C. In the following subsections, we will introduce each stage in detail.

1) Spectro-Spatial Feature Representation: Firstly, a single-trial raw EEG sample can be represented as a 2D map $X \in \mathbb{R}^{C \times T}$, where C represents the number of EEG channels, T represents time points. With the aim of explicit band interaction, we divide EEG signals into two characteristic frequency bands. Motivated by the fact that brain oscillations are typically categorized into specific frequency bands (delta: <4 Hz, theta: 4-7 Hz, alpha: 8-12 Hz, beta: 12-30 Hz, gamma: >30 Hz), EEG signals are first filtered into low (4-16 Hz) and high-frequency (16-40 Hz) bands, respectively. The choice of these two bands covers mu and beta rhythms most relevant to MI signatures. Other reasonable band segmentation options are discussed in ablation studies. We denote $X_l \in \mathbb{R}^{C \times T}$ to represent EEG data filtered in the low-frequency band, and $X_h \in \mathbb{R}^{C \times T}$ to represent EEG data filtered in the high-frequency band. As discussed in Appendix A, we adopt spatial filtering followed by temporal filtering to learn spectro-spatially discriminative patterns. Specifically, in our implementation, X is regarded as a 1D image along the temporal dimension with multi-channels. Then, spectro-spatial features are produced by:

$$U_l = \widehat{\mathcal{F}}_t \left(\widehat{\mathcal{F}}_s (X_l) \right)$$
$$U_h = \widetilde{\mathcal{F}}_t \left(\widetilde{\mathcal{F}}_s (X_h) \right) \tag{1}$$

where $\mathcal{F}_s$ and $\mathcal{F}_t$ are 1D point-wise spatial convolution and 1D depthwise temporal convolution, respectively. Both of them are followed by a Batch Normalization (BN) layer [33]. In the sequel, $U_l \in \mathbb{R}^{F \times T}$ and $U_h \in \mathbb{R}^{F \times T}$ are the output representing spectro-spatial features in each band, where F is the number of spatial filters per band. Note operations in each band are mathematically equivalent to the first block of EEGNet with $D = 1$ except for batch normalization layers, but computational complexity is significantly reduced and can be implemented more efficiently using 1D convolution operators. In the default settings, F is set to 64. The kernel size of temporal filters k is set to 63 for low-band input to capture a whole cycle of sinusoidal signal down to 4 Hz at 250 Hz sampling rate, and we halve the kernel size for the high band.

2) Cross-Frequency Interactions: To enhance representation ability of spectro-spatial features, we model the interplay between different frequency bands. Concretely, we investigate varieties of interaction operators.

1) Summation. $\mathcal{I}(U_l, U_h) = U_l + U_h$.
2) Concatenation. $\mathcal{I}(U_l, U_h) = [U_l, U_h]$.
3) Hadamard product. $\mathcal{I}(U_l, U_h) = U_l \odot U_h$.
4) Linear projection. $\mathcal{I}(U_l, U_h) = W_l U_l + W_h U_h$.
5) Split attention. We adopt *Fuse* and *Select* operators proposed in [34].

Here $\mathcal{I}(U_l, U_h)$ denotes the interaction function among multi-band features. The experiment results for each interaction operator are discussed in the ablation study. We empirically demonstrate using an element-wise addition operation yields the best performance for MI decoding from EEG. Not only does the summation operator couple features among different bands, but also preserves distinct characteristics in each band with the help of learnable affine parameters in BN layers before band interaction. Consequently, IFNet deals with cross-frequency interactions effectively and efficiently, requiring no extra parameters and a few more floating-point operations. Then, a GELU [35] activation function is applied after $\mathcal{I}(U_l + U_h)$.

3) Classification: The spectro-spatial features yielded from the first two stages remain a high-dimensional temporal representation. It is necessary to integrate temporal features to prevent overfitting while retaining characteristics of neural dynamics. While traditional approaches for MI decoding commonly employ variances as temporal characteristics, pooling is widely used for information aggregation in CNNs. It is known that pooling mechanisms are effective and efficient for dimensionality reduction and regularizing neural networks. Hence, we adopt temporal average pooling with a non-overlapping window size of W to extract robust temporal representation. By applying temporal aggregation, the output of second stage $U \in \mathbb{R}^{F \times T}$ is transformed to $U \in \mathbb{R}^{F \times T/W}$. Since oscillation rhythms of EEG can be assumed as stationary signals in a short moment, in this work, W is set to 125 under 250 Hz sampling rate, representing 0.5-s-long EEG characteristics. In the end, the flattened spectro-spatial-temporal features are fed to a fully-connected (FC) layer followed by the softmax operation, producing the output probabilities of each class. To help regularize our model, we use the dropout technique before the FC layer [36]. The dropout probability is set to a default value of 0.5.

B. Repeated Trial Augmentation

We introduce repeated trial augmentation consisting of random cropping and random erasing as a regularizer to stabilize training as well as enhance the generalization of neural networks in EEG decoding. Briefly speaking, it produces multiple instances of a sample in a mini-batch with several data transformations. An illustration of repeated trial augmentation is shown in Fig. 2.

Consider a mini-batch $\mathcal{B}$ with size $N \times C \times T$, where N denotes the number of samples, C and T are number of channels and time points. For each sample in a batch,
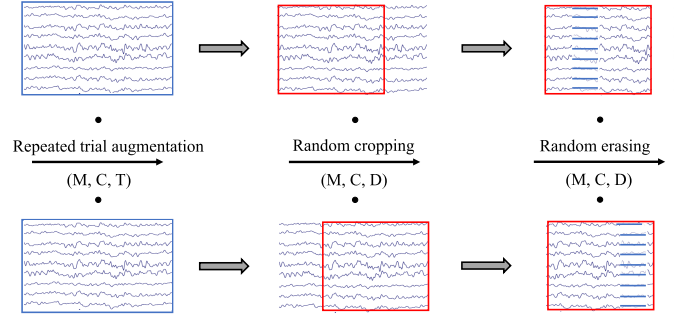

Fig. 2. Illustration of repeated trial augmentation (*M*: number of repeated trials, *D*: length of cropping window.) It produces multiple instances of a trial in a mini-batch using random cropping followed by random erasing.

we generate M multiple instances of it by applying similar data transformations. In particular, random cropping and random erasing are leveraged to produce multiple views of a selected sample. Firstly, we perform random cropping to stochastically crop the desired widow length from task trials during training, resulting in multiple temporal views of EEG signals. Concretely, a randomly initialized time point t is used to crop EEG signals with W window length, yielding $S_c = S[1 : C, \ t : t + W]$. As for the test, we simply use the fixed time segment for evaluation. Secondly, random erasing is performed along the temporal dimension acting like disconnecting abnormality which increases discrimination difficulty during network training. Moreover, by generating EEG instances with various levels of occlusion, it enforces networks to focus on task-specific stationary characteristics while ignoring artifacts appearing transiently during MI trials. To conduct random erasing, we randomly initialize the duration of erasing region to D, wherein $\frac{D}{W}$ is in the range specified by minimum D_l and maximum D_h. Then, a time point p is randomly initialized and the localized signals are erased with zero values, i.e., $S_c[1 : C, \ p : p + D] = 0$. Finally, an augmented batch with size $M \cdot N \times C \times W$ is produced and used for the training of neural networks in a step. In basic settings, W is set to 3-s-long time points; D_h is set to $\frac{1}{3}$; and M is set to 5.

IV. EXPERIMENTS

A. Datasets and Evaluation Protocols

Two publicly available datasets are utilized in this paper, which are denoted as the BCIC-IV-2a [37] and the OpenBMI [38] datasets. A brief description of each dataset is as follows.

The BCIC-IV-2a dataset is originally used as the official 2a dataset in BCI Competition IV. It aims at improving the cross-session performance of decoding algorithms for 4-class MI classification. Specifically, there are four MI tasks, namely the imagination of movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). The EEG data were recorded from 9 healthy subjects with 22 electrodes sampled at 250 Hz. The training and test data are from two sessions on different days for each subject. Each session contains 288 trials with 72 trials per class.

The OpenBMI dataset is a large benchmark containing 2 sessions of 2-class MI-EEG data from 54 healthy subjects. Each session consists of training and test phases, and each phase has 100 trials with balanced right and left-hand MI tasks. Note the test phase is conducted with online feedback using a CSP decoding model so that extra inter-phase variability is brought in the same session. Following the practice in the original paper, we select 20 electrodes located in the sensory-motor region and resample the time series to 250 Hz for compatibility with the BCIC-IV-2a dataset. No additional preprocessing is applied to both datasets.

We conduct two types of intra-subject performance evaluations, i.e., within-session evaluation and cross-session evaluation. The former is conducted on session 1 data of the OpenBMI dataset. For each subject, MI trials from the training phase and test phase are served as training data and test data, respectively. The latter is carried on whole sessions for both datasets. Training data and test data are from session 1 and session2, respectively. Meanwhile, since repeated trial augmentation employs random cropping to perform data transformation in the network training stage, 0-4 s post-cue data are used for training, while 0.5-3.5 s post-cue data are used for validation and test (corresponding to a 3-s-long cropping window implemented in random cropping).

To validate the superior performance of IFNet regarding the above evaluation settings, we compare IFNet with three baseline methods, i.e., FBCSP, EEGNet, and FBCNet. In particular, FBCNet has reported the best classification accuracy so far on the BCIC-IV-2a dataset in the cross-session setting. We re-implement these methods according to their open-source codes and retain key architectures as suggested by the respective authors. Furthermore, we perform statistical testing using Wilcoxon signed-rank test for the BCIC-IV-2a dataset (small sample size) and paired t-test for the OpenBMI dataset.

B. Training Procedure

We employ cross-entropy loss together with the AdamW [39] optimizer to update network parameters during training. To reduce performance variability and overfitting caused by multiple hyperparameters selection, we just take learning rates into consideration. Since a unified learning rate might result in suboptimal performance for different network architectures, to mitigate optimization inefficiency caused by inadequate learning rates, we perform a grid search of 2^{-8}– 2^{-15} before fine-grained experiments for each model on each dataset, respectively. The basic learning rate is selected by yielding superior cross-session accuracy achieved without repeated trial augmentation. Detailed settings of learning rates for each model are provided in Appendix B. To note, the optimal model-specific learning rates are relatively stable across datasets yet differ with each other. Other hyper-parameters regarding the AdamW optimizer and repeated trial augmentation are used in default settings. The batch size is set to a constant value of 32.

As it is done in [40] and [29], we employ a two-stage training strategy wherein the training data is further divided into a training set and a validation set. In the first stage, the

TABLE I
CROSS-SESSION CLASSIFICATION ACCURACY FOR EACH MODEL ON THE BCIC-IV-2A DATASET

Subject	FBCSP-SVM	EEGNet	FBCNet	IFNet
1	81.60	85.62	85.80	**88.47**
2	52.78	49.58	56.01	**56.35**
3	84.38	90.66	89.67	**91.77**
4	65.28	67.12	70.87	**73.78**
5	56.25	62.01	65.70	**69.72**
6	44.44	53.72	57.17	**60.42**
7	89.24	82.29	**90.70**	89.24
8	81.94	81.67	84.48	**85.42**
9	72.92	84.06	83.06	**88.72**
Avg	69.87*	72.97**	75.94*	**78.21**
Std	15.90	15.13	13.73	13.51

The * and ** represent the statistically significant difference between the performance of IFNet and the given model with *: $p < 0.05$, and **: $p < 0.01$.

TABLE II
CLASSIFICATION ACCURACY (MEAN±STD) COMPARISON OF DIFFERENT METHODS ON THE OPENBMI DATASET

Test mode	FBCSP-SVM	EEGNet	FBCNet	IFNet
Within-session	66.04 ±19.17**	68.38 ±18.53**	67.84 ±14.75	**71.22 ±18.85**
Cross-session	65.67 ±16.53**	69.45 ±16.19**	67.80 ±13.47*	**71.23 ±16.05**

The * and ** represent the statistically significant difference between the performance of IFNet and the given model with *: $p < 0.05$, and **: $p < 0.01$.

network is trained on the training set, and the model which produces the lowest validation loss is saved. In the second stage, the entire training data are used for network fine-tuning, while the optimizer is resumed from the checkpoint selected in the first stage. We stop stage 2 training when the training loss reduces below the stage 1 training loss. The maximum training epochs are set to 1000 and 500 for stage 1 and stage 2, respectively.

To this end, the training set is split into 5 folds in a sequential, class-balanced manner wherein each fold serves as a validation set alternately. In this manner, we perform network training twice resulting in 2×5 evaluation folds for each subject under specified evaluation settings. The final classification accuracy is averaged over all folds and all subjects on respective unseen test data.

C. Comparison With State-of-the-Art Approaches

We compare the performance of IFNet against state-of-the-art approaches using various evaluation protocols. The same training procedure and repeated trial augmentation are applied to all deep learning models.

Firstly, Table I shows the classification results of all methods on the BCIC-IV-2a dataset. IFNet consistently outperforms the other methods, and the discrepancy in average classification accuracy is statistically significant between IFNet and all the other methods. Moreover, the highest accuracy for each subject is achieved by IFNet except a slightly higher accuracy of FBCNet for subject 7. In particular, IFNet improves the
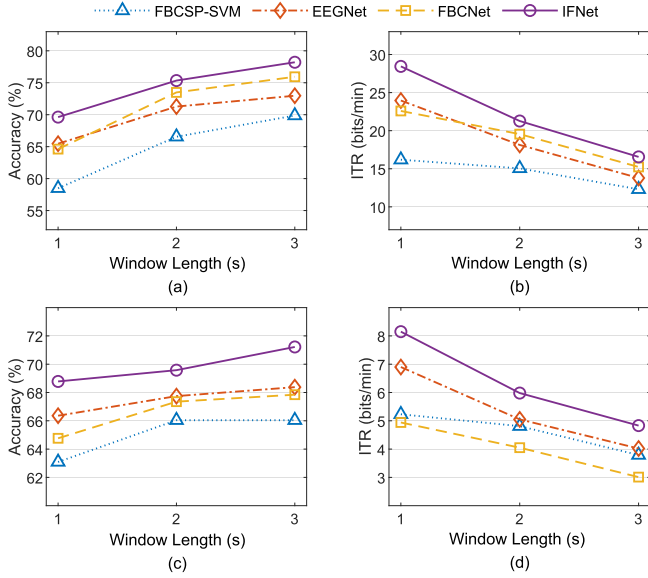
Fig. 3. Average classification accuracy and ITR with three window sizes (1 s, 2 s, and 3 s) for each method on both datasets. (a) Average cross-session accuracy on the BCIC-IV-2a dataset. (b) Average cross-session ITR on the BCIC-IV-2a dataset. (c) Average within-session accuracy on the OpenBMI dataset. (d) Average within-session ITR on the OpenBMI dataset. We see that, IFNet consistently outperforms the other methods for different window sizes across extensive performance evaluations.

average accuracy of FBCSP-SVM by 8.34% in the cross-session evaluation, demonstrating the complex feature learning ability of deep learning models while keeping robust to deal with inter-session variability.

Secondly, Table II presents cross-session and within-session performance of all methods on the OpenBMI dataset. Similarly, IFNet achieves the highest average classification accuracy in both evaluation settings. While FBCNet yields the second-highest average accuracy in the former dataset, it performs worse than EEGNet in this large 2-class classification dataset. However, IFNet exhibits consistent superiority in different datasets and evaluation settings. Notably, although cross-session decoding is considered to be much more difficult than within-session decoding, all methods achieve comparable performance with respect to evaluation protocols. This is partially explained by fewer training samples in the within-session evaluation (100 versus 200), indicating a higher amount of training data might be beneficial for data-hungry methods like deep learning models.

Lastly, we scale the network widths of IFNet F ranging from 16 to 256 to explore the effect of model capacity on classification accuracy in the case of limited and noisy EEG samples. The results with different network widths are shown in Fig. 9 in Appendix E. IFNet effectively makes use of additional feature channels without serious overfitting problems. Remarkably, IFNet-256 achieves the highest 78.74% accuracy for 4-class classification on the BCIC-IV-2a dataset.

D. Sensitivity Analysis on Decision Windows

To further demonstrate the effectiveness of IFNet with shorter decision window lengths, we conduct sensitivity anal-

ysis on decision windows for all methods on both datasets. Concretely, window lengths of 1, 2, and 3 s are investigated while using the same training procedure and repeated trial augmentation as in previous experiments. In order to train networks with different window lengths, we adjust the cropping window size according to the targeted window length, and the dimensions of input channels of the FC layer are scaled proportionally to window size. As for performance evaluation, 0.5-1.5 s and 0.5-2.5 s post-cue data are evaluated under the 1-s-long window and 2-s-long window settings, respectively. In addition to accuracy evaluation, we leverage information transfer rate (ITR) as a composite index incorporating both speed and accuracy. The ITR representing number of transfer bits per minute is given by,

$$ITR = \frac{60}{D}\left(\log_2 N + P\log_2 P + (1-P)\log_2\frac{1-P}{N-1}\right) \quad (2)$$

where N denotes number of classes, P denotes classification accuracy for a subject, and D denotes the duration of a sample in seconds. Results of this analysis are presented in Fig. 3. We can draw the following conclusions. First, IFNet consistently outperforms the other methods for all window sizes on both datasets in terms of average accuracy and ITR. In addition, IFNet obtains the same average accuracy for a 1-s-long window as that of FBCSP for a 3-s-long window on the BCIC-IV-2a dataset, and it also attains competitive average accuracy for a 1-s-long window as that of state-of-the-art methods for a 3-s-long window on the OpenBMI dataset. Second, we observe longer window lengths result in higher classification accuracy whereas the reverse applies to ITR, and differences among these windows in terms of average accuracy and ITR are statistically significant (all $p < 0.05$) for all methods on both datasets. It suggests practitioners should take both accuracy and response speed into consideration depending on specific circumstances. Third, compared to FBCSP-SVM, deep learning models achieve superior or comparable average accuracy and ITR for all window sizes on both datasets, further manifesting a deep learning-based MI-BCI system is promising and within reach. Of note, we also conduct the same analysis without using repeated trial augmentation, where we gain the similar conclusions except overall lower accuracy owing to the absence of data augmentation strategies. In sum, IFNet achieves the best trade-off between decoding speed and accuracy.

E. Ablation Studies

In this section, to validate the effectiveness of data augmentation, band segmentation, and band interaction, we ablate important design elements in the proposed IFNet using cross-session analysis on the BCIC IV-2a dataset. We adopt the same training procedure for all ablation models.

1) Ablation on Data Augmentation: To evaluate the effect of repeated trial augmentation on enhancing the generalization of neural networks, we conduct an ablation study on data augmentation methods utilized in this paper. Table III summarizes the average classification accuracy achieved with different combinations of data augmentation methods for all investigated CNN models. Note in this context, repeated trial

TABLE III

AVERAGE ACCURACY OF DEEP LEARNING MODELS WITH DIFFERENT COMBINATIONS OF DATA AUGMENTATION STRATEGIES ON THE BCIC-IV-2A DATASET

Random cropping	Random erasing	Repeated trial aug.	EEGNet	FBCNet	IFNet
–	–	–	70.75	74.31	76.06
✓	✓	✓	72.97	**75.94**	**78.21**
✓	–	–	71.12	75.42	76.69
✓	✓	–	**73.28**	75.63	77.77
✓	–	✓	70.66	75.82	76.93
–	✓	–	72.39	74.65	76.94
–	✓	✓	72.07	74.97	77.45

TABLE IV

RESULTS OF IFNET WITH DIFFERENT BAND SEGMENTATION OPTIONS ON THE BCIC-IV-2A DATASET

Band segmentation	Accuracy	Params
{(4, 40)}	76.36	7236
{(4, 16), (16, 40)}	**78.21**	10884
{(4, 8), (8, 16), (16, 40)}	76.94	14532
{(4, 8), (8, 16), (16, 30), (30, 40)}	76.79	20228

TABLE V

RESULTS OF IFNET WITH DIFFERENT INTERACTION OPERATORS ON THE BCIC-IV-2A DATASET

Interaction operator	Accuracy	Params
Summation	**78.21**	10884
Concatenation	77.12	12420
Hadamard product	75.10	10884
Linear projection	76.31	19204
Channel-wise attention	75.31	14100

augmentation refers to generating $M = 5$ multiple instances of a sample with any predefined data transformations. From Table III, we have the following observations: (1) All data augmentation methods consistently improve performance against a naive training mode across all models, indicating the effectiveness of proposed data augmentation in the context of MI-EEG decoding. (2) Combinations of different methods outperform a single one in most cases. Particularly, IFNet and FBCNet achieve the highest accuracy with repeated trial augmentation compromising random cropping and random erasing. EEGNet attains the best accuracy in the specific case that $M = 1$ in single instance repeated trial augmentation. (3) Random erasing improves accuracy marginally for FBCNet, which can attribute to no temporal filters to be regularized in FBCNet. (4) While repeated trial augmentation moderately strengthens the generalization of IFNet and FBCNet, it is detrimental to the performance of EEGNet. Considering more diverse training instances produced in a batch as well as a relatively low capacity of EEGNet, it would be better to use multiple instances repeated trial augmentation for models with an ample capacity.

2) Ablation on Band Segmentation: We investigate cross-frequency interactions explicitly by means of multi-band inputs of EEG signals. While there exist various band segmentation means, we ablate on neuro-physiologically significant spectral bands, following the common design philosophy regarding MI decoding. Accordingly, four band segmentation options are investigated in Table IV. The option without band segmentation represents raw EEG input which is the same as the input fashion of EEGNet. The option corresponding to the maximum number of band segments is similar to the division of theta, alpha, beta, and gamma frequency bands. It can be observed that the network without band segmentation is inferior to the other multi-band network architectures, demonstrating the effectiveness of band segmentation in our IFNet. Besides that, compared to IFNet proposed in this paper, further division of frequency bands decreases the classification performance significantly. On the one hand, more narrow frequency bands result in a linearly increasing number of parameters, hence the network tends to overfit due to the scarcity of training data. On the other hand, the first stage of IFNet incorporates temporal filters acting like frequency filters, hence IFNet is capable of learning

spectrally localized features from a broad frequency band. In conclusion, low and high-band segmentation is preferable to achieve high performance and efficiency.

3) Ablation on Interaction Operators: The second key component of IFNet lies on the efficient band interaction operation. As described in Section III-A.2, we employ five diverse interaction functions to explore cross-frequency interactions. The results in Table V indicate a simple yet effective element-wise addition operation is preferable to MI decoding from EEG. It outperforms the second-highest operator with an improvement of 1% on the BCIC-IV-2a dataset. In particular, the concatenation operator exerts the least interaction between two band features while the corresponding model still outperforms FBCNet with an improvement of 1%. More significantly, by applying a decent interaction operation, extra discriminative interactions can be mined to gain statistically significant improvement ($p < 0.05$ between summation operator and concatenation operator). It is noted that the summation operation is one of the solutions covered by linear projection. However, the large margin between these two methods indicates overfitting problems faced by linear projection due to limited and noisy EEG samples. Apart from parameter-free operators, we also manage to exploit a channel-wise attention operator. Although it has shown success in image recognition, it does not transfer well in MI decoding. Further investigation on attention mechanism might help facilitate EEG feature representation [26]. In conclusion, the summation operator deals with cross-frequency interactions effectively and efficiently.

F. Interpretability Analysis

To gain insight into how IFNet extracts neuro-physiologically sound features, we present interpretability analysis from two perspectives on the BCIC-IV-2a dataset.

1) Relation With CFC: To understand how IFNet works with cross-frequency interactions, we analyze the output produced
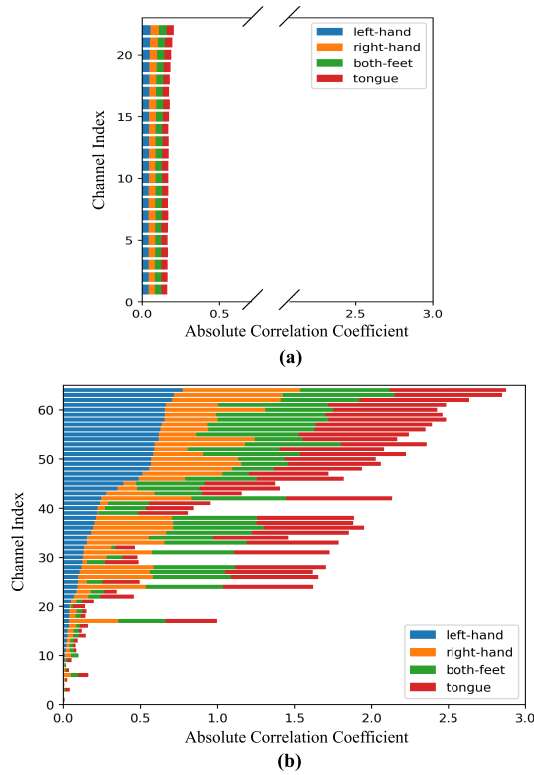
Fig. 4. Subject-level channel-wise absolute correlation coefficient between low and high-frequency bands averaged over all trials of each MI task for subject 3 on the BCIC-IV-2a dataset. (a) Channel-wise absolute correlation coefficient of input signals. (b) Channel-wise absolute correlation coefficient of output features produced by the first stage in IFNet. We can observe large discrepancies of correlation coefficients between input signals and extracted features, verifying IFNet captures coupling across frequency bands for complementary and discriminative feature representation.

by the first stage in IFNet. Since a CFC signature is typically described as a high correlation of features between two different frequency bands, we inspect the channel-wise correlation of features between low and high-frequency bands in IFNet. Specifically, given an input sample $[X_l, X_h]$, IFNet produces $[U_l, U_h]$ through the first stage, where $U_l \in \mathbb{R}^{F \times T}$, $U_h \in \mathbb{R}^{F \times T}$, and F is the number of output channels. We compute the Pearson correlation coefficient for each channel, yielding $C \in \mathbb{R}^F$. Then subject-level C is averaged over all trials of each MI task for a subject. For comparison, C is also calculated between input band signals of samples. Fig. 4 shows the average channel-wise absolute correlation coefficients for subject 3. High correlation can be observed between a portion of output features from low and high-frequency bands, whereas nearly zero correlation is observed for input signals. This demonstrates IFNet holds similar characteristics of CFC and verifies IFNet is capable of capturing coupling across frequency bands from raw EEG signals. Furthermore, as for a specific channel, there exists a correlation discrepancy across different MI tasks. This is consistent with observations that different behavioral tasks evoke different CFC signatures [41]. We assume such correlation discrepancies contribute to complementary and discriminative feature representation.

2) Attribution Analysis on Input Signals: It is of great interest to understand what EEG features that IFNet learns to pay attention to during diverse MI tasks. Meanwhile, to ensure that the classification performance is driven by task-specific features as opposed to noise or artifacts in the data, we adopt a gradient-based method called Integrated Gradients [42] to attribute predictions of IFNet to its input signals. Briefly, in an EEG decoding network, it could tell us which sample points of the EEG signals are responsible for a certain label (task) being picked. On implementation, Integrated gradients aggregate the gradients along the straightline between the baseline (usually zero scores) and the input, and can be computed easily using a few calls to the gradient operation. The analyses are performed on IFNet models for subject 3 and subject 7 on the BCIC-IV-2a dataset. Concretely, integrated gradients are first calculated on each input sample, yielding $G_l \in \mathbb{R}^{C \times T}$, and $G_h \in \mathbb{R}^{C \times T}$. Next, we average absolute integrated gradients along the temporal dimension, yielding channel attributions $G_l \in \mathbb{R}^C$, and $G_h \in \mathbb{R}^C$. Then subject-level attributions are calculated by averaging channel attributions for each band respectively over all trials of each MI task for a subject. Finally, normalized spectro-spatial attributions are mapped to the corresponding electrode locations, resulting in attribution patterns that associate with brain regions and frequency bands for each MI task. Attribution patterns for subject 3 and subject 7 are shown in Fig. 5 (a) and (b).

As for subject 3, large attributions mainly lie in the high-frequency band. Meanwhile, channel attributions mostly concentrate on the right, left, and middle sensorimotor areas for left-hand, right-hand, and both-feet MI, respectively. These characteristics are closely associated with the well-known MI-related brain activation patterns. On the other hand, large attributions are observed on the low-frequency band for subject 7. This could be explained by the fact that discriminative features lie on subject-specific frequency bands. Also, the most relevant features focus on contralateral brain regions with regard to MI tasks. Apart from spectro-spatially localized activation patterns, as explored in this paper, cross-frequency interactions can also be utilized and complement discriminative MI features. In short, our IFNet learns neuro-physiologically sound features and can in turn provide complementary insights on neurophysiological bases of mental tasks by explainable AI techniques.

V. DISCUSSION

In this paper, a lightweight IFNet architecture is designed to extract spectro-spatio-temporally robust features for MI decoding from EEG. Guided by neurophysiological priors along with efficient convolution operations, IFNet achieves fast and accurate MI decoding in the presence of limited, noisy, and non-stationary EEG samples. Moreover, we validate the effectiveness of repeated trial augmentation as a regularizer for better generalization of neural networks. The extensive experiment results and ablation studies suggest the inclusion of neurophysiological priors while designing an efficient network can lead to nontrivial improvements in decoding performance, which corroborates with findings in previous studies [29], [43]. On the other hand, previous studies rarely report effect
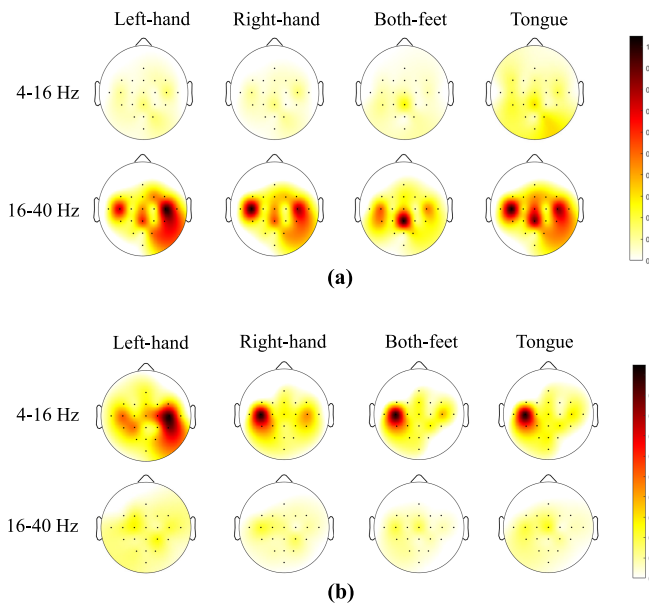
Fig. 5. Subject-level spectro-spatial attribution patterns averaged over all trials of each MI task for a subject on the BCIC-IV-2a dataset. (a) Attribution patterns for subject 3. (b) Attribution patterns for subject 7. We see that large attributions lie in task-specific brain regions and distinctive frequency bands, which are closely associated with the known MI signatures.

of decision windows on performance of MI decoding algorithms whereas it is not trivial in practical MI-BCI utilization. Through sensitivity analysis on different decision windows, we manifest the competitive ability of IFNet for a 1-s-long window as that state-of-art models for a 3-s-long window, indicating the superiority of decoding speed of IFNet while providing comparable accuracy. Consequently, IFNet increases ITR significantly and has shown a promising tool for online MI-BCI systems. As a final point, we show neurophysiological signatures of MI are effectively captured by IFNet, providing insights of knowledge learned by neural networks.

Concerning model size and training complexity, as shown in Table VI, IFNet has moderate size while demanding less training time compared to other deep learning models. As stressed in Section III-A.1, IFNet is implemented with 1D convolution and the computation complexity is reduced substantially as contrasted with EEGNet. In particular, using a NVIDIA GTX1080Ti graphic card, the training time of IFNet is 1.23 and 1.20 times faster than EEGNet and FBCNet, respectively. The results indicate potential advantage of IFNet for fast deployment in online MI decoding. In inference mode, i.e. predicting a single trial, the model requires less than 10 ms running on a CPU device, which is suitable for computation-intensive continuous control tasks.

There remain several limitations to be further explored in future work. Firstly, all evaluations performed in this work are offline analyses following the common practice in previous studies. As the aim of MI-BCI is to establish real-time direct commutation and control between a human brain and a machine, the proposed method necessitates further validation in online settings [44]. Moreover, the online co-adaptation between a user and algorithm will exert an additional effect on

TABLE VI
MODEL SIZE AND TRAINING TIME PER FOLD

Model	Params	Training time (s) per fold
IFNet	10884	295
EEGNet	1820	362
FBCNet	10660	355

the online performance of MI-BCI systems. Secondly, since IFNet is a general neural network architecture targeted at efficient processing of EEG signals, it can be investigated on other EEG measurements, such as emotion recognition and sleep staging. Thirdly, although repeated trial augmentation is utilized to prevent overfitting, IFNet is still data-hungry and its capacity can be easily enlarged by width scaling. We consider transfer learning as a potential solution to perform fast calibration using fewer targeted samples [45], [46]. Finally, while recent studies have leveraged attention mechanisms in other mental tasks [47], [48], we observe deteriorated performance of IFNet using different attention operators in our unreported experiments. We conjecture that limited EEG samples obscure the generalization of learned attention, and further investigation on attention mechanism might guide the effective utilization of attention for MI-EEG decoding.

VI. CONCLUSION

In this work, we propose IFNet to further advance MI decoding accuracy, which shows significantly improved performance on two benchmark MI datasets as compared with state-of-the-art methods. Besides, it achieves the best trade-off between decoding speed and accuracy. We also introduce a data augmentation strategy named repeated trial augmentation to improve the generalization of neural networks. IFNet is compact while powerful to extract spectro-spatio-temporally robust features, which is also neuro-physiologically interpretable. The proposed IFNet could be beneficial for MI-based BCI applications, and other BCI paradigms for feature-less decoding.

APPENDIX
A. Preliminary Experiment on EEGNet

EEGNet is powerful for various EEG decoding tasks while remaining compact as much as possible. To reveal the key to the success of EEGNet, we conduct a pilot experiment on EEGNet regarding model width and depth. Concretely, EEGNet consists of two blocks. The first block contains F temporal filters and each of which is followed by D spatial filters. The second block adopts depthwise-separable convolution to further process spectro-spatial features along with temporal dynamics. We question whether these two blocks contribute equally to effective feature extraction from EEG. Moreover, as a matter of fact, width and depth are two essential dimensions for network scaling. Therefore, we investigate width scaling and depth scaling on EEGNet.

With the above considerations, three baseline networks with different depths but the same width and capacity are

TABLE VII
IFNet Architecture

Block	Layer	# Kernels	Kernel Size	# Parameters	Output	Options
Band Segmentation	Input				$(2*C, T)$	
Spatial Convolution	Group Conv1D	$2*F$	$(1,)$	$2*C*F$	$(2*F, T)$	groups = 2
	BatchNorm1D			$4*F$	$(2*F, T)$	
Temporal Convolution	Depthwise Conv1D	$2*F$	$(63,)$ and $(31,)$	$63*F+31*F$	$(2*F, T)$	padding = same
	BatchNorm1D			$4*F$	$(2*F, T)$	
	Activation-GELU				$(2*F, T)$	
Band Interaction	Element-wise Addition				(F, T)	
Temporal Aggregation	1D Average Pooling		$(W,)$		$(F, T/W)$	
	Dropout				$(F, T/W)$	$p = 0.5$
Classifier	Flatten				$F*T/W$	
	Fully-connected Layer	N		$F*T/W*N+N$	N	

C: number of EEG channels, T: number of time points, F: number of convolution filters per band, W: temporal pooling size, N: number of classes

TABLE VIII
Average Subject-Specific Classification Accuracy

Dataset	Test mode	FBCSP-SVM	EEGNet	FBCNet	IFNet	IFNet V2[1]
BCIC-IV-2A	Cross-session	69.87±15.90*	72.97±15.13**	75.94±13.73*	78.21±13.51	**79.78**±12.23
OpenBMI	Within-session	66.04±19.17**	68.38±18.53**	67.84±14.75	71.22±18.85	**71.78**±18.32
	Cross-session	65.67±16.53**	69.45±16.19**	67.80±13.47*	71.23±16.05	**72.45**±15.67

The * and ** represent the statistically significant difference between the performance of IFNet and the given model with *: $p < 0.05$, and **: $p < 0.01$. [1]Up to the publication of this work, IFNet V2 has been implemented in online codes.
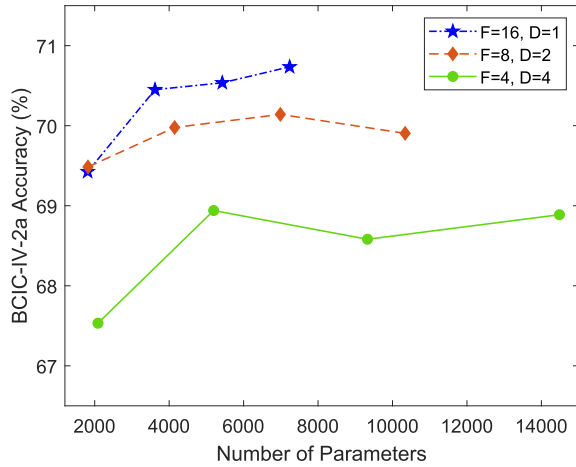

Fig. 6. Scaling network width with three EEGNet variants of different depths. Each dot in a line represents a model with a different width coefficient w, corresponding to 1, 2, 3, 4, respectively.
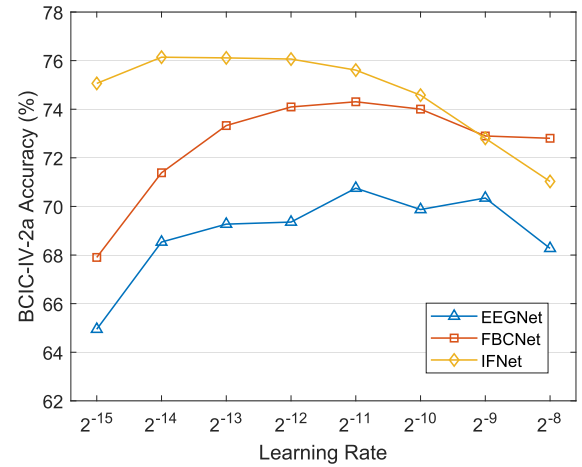

Fig. 7. Cross-session accuracy of each model with different learning rates on the BCIC-IV-2a dataset.

constructed. The first employs only the first block of EEGNet with $F = 16$, $D = 1$. The second is set to $F = 8$, $D = 2$ with two activation layers, which corresponds to the original EEGNet. The third employs one first block followed by two second blocks using $F = 4$, $D = 4$. All of them yield the same width of features fed to the fully-connected layer but differ in network depth. Next, as for each baseline network, we modify the width coefficient w denoting the multiplier of F filters. We scale w to 1, 2, 3, 4. The training and

evaluation procedures are elucidated in Section IV-B and IV-A. The experiment results on the BCIC-IV-2a dataset are shown in Fig. 6. On the one hand, shallow networks outperform deep networks in a large margin for the decoding of MI from EEG. On the other hand, all networks benefit from width scaling while the first is more computation-efficient. Consequently, we empirically demonstrate the first block is the major determinant of EEGNet and its learning capability can be strengthened by width scaling.
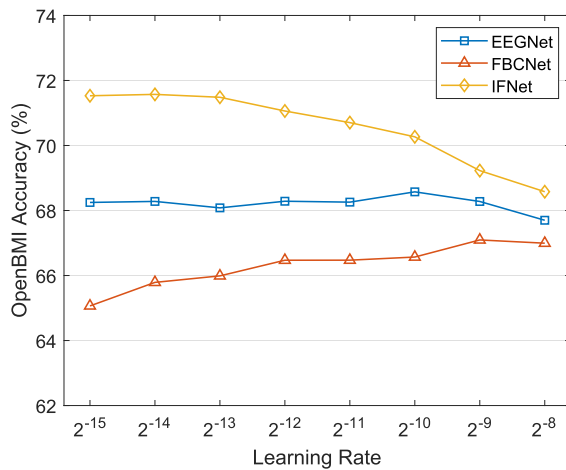
Fig. 8.   Cross-session accuracy of each model with different learning rates on the OpenBMI dataset.
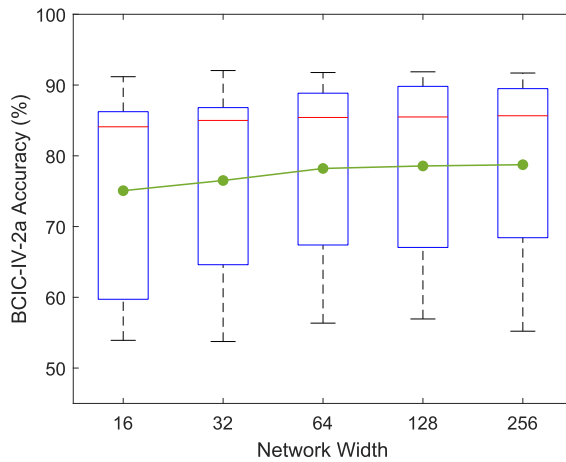

Fig. 9.   Cross-session accuracy of IFNet with different network widths on the BCIC-IV-2a dataset.

Based on these findings, we leverage the first baseline model to guide the efficient design of our networks. More importantly, thanks to a one-to-one correspondence between temporal filters and spatial filters, we put them reversely so that the computational complexity can be reduced significantly by roughly $\frac{k}{k+1} \approx 1$, where k is the kernel size of temporal filters.

B. Selection of Learning Rates

We conduct learning rate selection ranging from 2^{-15} to 2^{-8} for each model on both datasets. Notably, we do not apply repeated trial augmentation that exerts additional effect on network training. Cross-session evaluations on two datasets are shown in Fig. 7 and Fig. 8, respectively. Accordingly, as for the BCIC-IV-2a dataset, we use learning rates of 2^{-12}, 2^{-11}, 2^{-11} for IFNet, FBCNet, EEGNet, respectively. Also, we select learning rates yielding superior accuracy for each model on the OpenBMI dataset. 2^{-13}, 2^{-9}, 2^{-10} are selected for IFNet, FBCNet, EEGNet, respectively. The above learning rates are fixed through the entire experiments.

C. IFNet Architecture

Here we present implementation level specifications of IFNet in Table VII

D. Evaluation Results

The summary results of classification accuracy achieved by baseline methods and IFNet along with the statistical significance are provided in Table VIII

E. Effect of Network Widths

The results of IFNet with different network widths are shown in Fig. 9. It is observed that scaling up network width consistently improves classification accuracy.

REFERENCES

[1] D. J. McFarland and J. R. Wolpaw, "Brain-computer interfaces for communication and control," *Commun. ACM*, vol. 54, no. 5, pp. 60–66, 2011.

[2] B. He, B. Baxter, B. J. Edelman, C. C. Cline, and W. W. Ye, "Noninvasive brain-computer interfaces based on sensorimotor rhythms," *Proc. IEEE*, vol. 103, no. 6, pp. 907–925, Jun. 2015.

[3] D. Xu and Q. Wang, "Noninvasive human-prosthesis interfaces for locomotion intent recognition: A review," *Cyborg Bionic Syst.*, vol. 2021, pp. 1–14, Jan. 2021.

[4] Y. Wang, W. Li, S. Togo, H. Yokoi, and Y. Jiang, "Survey on main drive methods used in humanoid robotic upper limbs," *Cyborg Bionic Syst.*, vol. 2021, pp. 1–12, Jan. 2021.

[5] F. Pichiorri et al., "Brain-computer interface boosts motor imagery practice during stroke recovery," *Ann. Neurol.*, vol. 77, no. 5, pp. 851–865, 2015.

[6] Y. Li et al., "An EEG-based BCI system for 2-D cursor control by combining mu/beta rhythm and P300 potential," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 10, pp. 2495–2505, Oct. 2010.

[7] B. J. Edelman et al., "Noninvasive neuroimaging enhances continuous neural tracking for robotic device control," *Sci. Robot.*, vol. 4, no. 31, Jun. 2019, Art. no. eaaw6844.

[8] K. LaFleur, K. Cassady, A. Doud, K. Shades, E. Rogin, and B. He, "Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain-computer interface," *J. Neural Eng.*, vol. 10, no. 4, Aug. 2013, Art. no. 046003.

[9] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

[10] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[11] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012.

[12] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Comput. Interface*, vol. 4, no. 3, pp. 155–174, Mar. 2017.

[13] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1352–1359, Jul. 2019.

[14] W. Zhang and D. Wu, "Manifold embedded knowledge transfer for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1117–1127, May 2020.

[15] R. T. Canolty et al., "High gamma power is phase-locked to theta oscillations in human neocortex," *Science*, vol. 313, no. 5793, pp. 1626–1628, Sep. 2006.

[16] O. Jensen and L. L. Colgin, "Cross-frequency coupling between neuronal oscillations," *Trends Cognit. Sci.*, vol. 11, no. 7, pp. 267–269, Jul. 2007.

[17] A. Hyafil, A.-L. Giraud, L. Fontolan, and B. Gutkin, "Neural cross-frequency coupling: Connecting architectures, mechanisms, and functions," *Trends Neurosci.*, vol. 38, no. 11, pp. 725–740, Nov. 2015.

[18] D. Gwon and M. Ahn, "Alpha and high gamma phase amplitude coupling during motor imagery and weighted cross-frequency coupling to extract discriminative cross-frequency patterns," *NeuroImage*, vol. 240, Oct. 2021, Art. no. 118403.

[19] E. Combrisson et al., "From intentions to actions: Neural oscillations encode motor processes through phase, amplitude and phase-amplitude coupling," *NeuroImage*, vol. 147, pp. 473–487, Feb. 2017.

[20] N. Feng, F. Hu, H. Wang, and M. A. Gouda, "Decoding of voluntary and involuntary upper-limb motor imagery based on graph Fourier transform and cross-frequency coupling coefficients," *J. Neural Eng.*, vol. 17, no. 5, Oct. 2020, Art. no. 056043.

[21] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.

[22] Y. Hou, L. Zhou, S. Jia, and X. Lun, "A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN," *J. Neural Eng.*, vol. 17, no. 1, Feb. 2020, Art. no. 016048.

[23] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016025.

[24] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," in *Proc. ICLR*, 2016, pp. 1–15.

[25] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[26] J. Chen, W. Yi, D. Wang, J. Du, L. Fu, and T. Li, "FB-CGANet: Filter bank channel group attention network for multi-class motor imagery classification," *J. Neural Eng.*, vol. 19, no. 1, Feb. 2022, Art. no. 016011.

[27] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophys.*, vol. 113, no. 6, pp. 767–791, 2002.

[28] F. Lotte et al., "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.

[29] R. Mane et al., "FBCNet: A multi-view convolutional neural network for brain-computer interface," 2021, *arXiv:2104.01233*.

[30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.

[31] H. Zhang et al., "ResNeSt: Split-attention networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 2735–2745.

[32] E. Hoffer, T. Ben-Nun, I. Hubara, N. Giladi, T. Hoefler, and D. Soudry, "Augment your batch: Improving generalization through instance repetition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8126–8135.

[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37. Lille, France, 2015, pp. 448–456.

[34] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 510–519.

[35] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.

[36] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.

[37] M. Tangermann et al., "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, Jul. 2012.

[38] M.-H. Lee et al., "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, May 2019, Art. no. giz002.

[39] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. ICLR*, 2019, pp. 1–19.

[40] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, pp. 5391–5420, Nov. 2017.

[41] R. T. Canolty and R. T. Knight, "The functional role of cross-frequency coupling," *Trends Cognit. Sci.*, vol. 14, no. 11, pp. 506–515, Nov. 2010.

[42] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. ICML*, 2017, pp. 3319–3328.

[43] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.

[44] J. R. Stieger, S. A. Engel, D. Suma, and B. He, "Benefits of deep learning classification of continuous noninvasive brain-computer interface control," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 046082.

[45] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026007.

[46] K. Zhang, N. Robinson, S.-W. Lee, and C. Guan, "Adaptive transfer learning for EEG motor imagery classification with deep convolutional neural network," *Neural Netw.*, vol. 136, pp. 1–10, Apr. 2021.

[47] E. Su, S. Cai, L. Xie, H. Li, and T. Schultz, "STAnet: A spatiotemporal attention network for decoding auditory spatial attention from EEG," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 7, pp. 2233–2242, Jul. 2022.

[48] E. Eldele et al., "An attention-based deep learning approach for sleep stage classification with single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 809–818, 2021.