Cognitive Computation (2025) 17:84
https://doi.org/10.1007/s12559-025-10442-0

RESEARCH

A^2DM: Enhancing EEG Artifact Removal by Fusing Artifact Representation into the Time-Frequency Domain

Haoran Li[1] · Fan Feng[2] · Jiarong Kang[1] · Jin Zhang[1] · Xiaoli Gong[1,3] · Tingjuan Lu[4] · Shuang Li[5] · Zhe Sun[6] · Jordi Solé-Casals[7,8,9]

Received: 12 August 2024 / Accepted: 11 March 2025

Abstract
The electroencephalogram (EEG) provides essential data for analyzing brain activities. However, artifacts such as electrooculography (EOG) and electromyography (EMG) often interleave with the EEG signals, significantly affecting the quality of EEG signal analysis. The heterogeneous distribution of these artifacts in the time-frequency domain makes it challenging to remove multiple artifacts using a unified model. In this paper, we propose an *a*rtifact-*a*ware EEG *d*enoising *m*odel, referred to as A^2DM, to effectively remove various types of artifacts in a unified manner. We first obtain an artifact representation that indicates the type of artifact from a pre-trained artifact classification model. This artifact representation is then used as prior knowledge, which is fused into the denoising model in the time-frequency domain. This enables the model to become aware of the artifact type and precisely remove artifacts based on their type. Due to the heterogeneous distributions of artifacts in the frequency domain, we introduce a frequency enhancement module that can identify specific types of artifacts based on their representation and remove them using a hard attention mechanism. Additionally, we design a time-domain compensation module to enhance the denoising capability of A^2DM by compensating for potential losses of global information. Comprehensive experiments demonstrate that A^2DM significantly outperforms the novel CNN in denoising EEG signals, showing a notable 12% improvement in correlation coefficient (CC) metrics. This work demonstrates that artifact representation can be used in artifact removal models to effectively remove multiple types of artifacts.

Keywords Electroencephalography · Artifact representation · EEG denoising · Frequency domain representation

Introduction

Electroencephalography (EEG) is a fundamental tool for analyzing brain activity[1–6]. During EEG signal acquisition, artifacts introduced by the subjects themselves significantly contaminate the EEG data, adversely impacting the quality of subsequent signal analysis. Therefore, removing artifacts is a crucial step in practical applications [7–11].

EEG signals often contain multiple types of artifacts. For example, ocular artifacts and muscle artifacts are two major artifacts [12]. Ocular artifacts, primarily caused by eye movements or blinking, propagate over the scalp and are captured during EEG recording. Such ocular artifacts can be recorded using the electrooculogram (EOG). Muscle artifacts are primarily induced by muscle movements in regions such as the head or limbs [13]. Such muscle artifacts can be recorded using the electromyogram (EMG). In practical scenarios, artifact generation exhibits random characteristics, with multiple artifact types concurrently interleaved within the EEG signal. Figure 1a shows a pure EEG segment, while Fig. 1b depicts the pure EEG contaminated by both EOG and EMG. Interleaved artifacts interfere with the data distribution of the EEG and degrade the performance of EEG signal analysis, while it is challenging to remove the interleaved artifacts simultaneously.

Previous works [14, 15] generally focus on removing a single type of artifact. For example, the EEGANet [15] is used to remove the EOG. Although NovelCNN [16] and EEGdenoiseNet [1] can remove either EOG or EMG artifacts, they lack the capability to handle interleaved artifacts. However, these models are trained on single artifact types, limiting their ability to remove interleaved artifacts. Thus, the denoise model for EOG may have limited performance in EMG removal. As seen in Fig. 1b, the EOG introduces a large amplitude to the signal in the time domain [17], while EMG includes high-frequency noise [18, 19]. Moreover, Fig. 1c

Extended author information available on the last page of the article

Published online: 21 March 2025

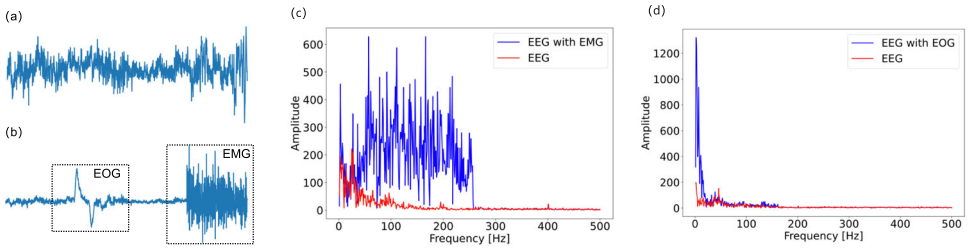

Fig. 1 Analyzing EOG and EMG difference characteristics in time and frequency domains. **a** The pure EEG segment. **b** The pure EEG segment interleaved the EOG and EMG artifacts. **c** and **d** Spectral analyses of noisy EEG sub-segment (**b**) contaminated with EMG and EOG artifacts, respectively

and d showcase spectral analyses of EEG segments contaminated with EOG or EMG artifacts. Notably, EOG artifacts are primarily concentrated in the low-frequency spectrum, while EMG artifacts are broadly distributed across the mid-to-high frequency range, spanning from 20 Hz to over 200 Hz. It is challenging to remove multiple types of artifacts simultaneously with neural network models since it is difficult to integrate the features of artifact types into a unified model. Similarly, the denoise model for EMG has limited performance in EOG removal due to the heterogeneity of the distribution in the time-frequency domain.

We tackle the challenge by developing an artifact-aware denoising framework for removing interleaved artifacts within EEG. We utilized an artifact awareness module to acquire the type of artifacts and then adaptively remove EOG and EMG artifacts using hard attention-based frequency filtering techniques. Specifically, we train an artifact-aware module (*AAM*) designed to identify the artifact type based on these distinctions. Concurrently, the *AAM* is capable of acquiring an artifact representation (AR) obtained from the high-level feature within the artifact, serving as prior knowledge for the denoising module. This prior information makes the denoise model adaptive to the variance of artifacts and then removes them in a unified model.

Using artifact representation, we can infer the type of artifact. Furthermore, we aim to develop more targeted denoising strategies for different artifacts within EEG data. The distinct distribution characteristics of EOG and EMG in the frequency domain provide opportunities to generate tailored denoising strategies. As illustrated in Fig. 1, EMG is characterized by high-frequency noise information in the mid-to-high-frequency range [18, 19], whereas EOG contributes amplitude variations to the signal [17] in the low-frequency range. Thus, we introduce frequency domain analysis to identify the frequency components of artifacts based on artifact representation. We propose a frequency enhancement module (*FEM*). The FEM utilizes hard attention [20], tailored to selectively remove specific frequency components according to the various distributions of EOG and EMG. It is worth noting that hard attention is more suitable for artifact removal tasks compared to soft attention. The hard attention mechanism can directly eliminate the modes containing artifact features, rather than dynamically adjusting the distribution of frequency domain components. We validate our conclusion in Section "The Effectiveness of Hard Attention in FEM." Moreover, we find that aggressive removal of the frequency component through hard attention may bring about a loss of EEG information. Then, we design the time-domain compensation module (*TCM*) to compensate for the information lost. More detail can be found in Section "Time-Domain Compensation Module."

In this paper, we present the design and implementation of a novel model for the removal of multiple types of EEG artifacts, which we have named the artifact-aware denoising model (A^2DM). The A^2DM leverages an artifact-aware model (AAM) to generate an artifact representation (AR), thereby sensitizing a frequency enhancement module (*FEM*) to remove the specific type of artifact present. It is important to highlight that the denoise block leverages features of the artifact extracted by the awareness module, enabling it to efficiently address two types of artifacts within a cohesive framework. This approach not only streamlines the multiple artifact denoising process but also enhances the overall performance of the framework by ensuring a more targeted and effective artifact removal. After that, a time-domain compensation module (*TCM*) is employed within the time domain to mitigate potential information loss and enhance overall performance. We conducted a comprehensive evaluation of the A^2DM on the EEGdenoiseNet benchmark [1]. The results demonstrate that our method outperforms the state-of-the-art method, i.e., novel CNN [16] in multi-type EEG denoising tasks. Notably, the correlation coefficient has seen an improvement of 12%, underscoring the efficacy and advancement of our proposed model in the field of EEG signal processing. The main contributions of this paper are as follows:

(1) We propose an (A^2DM). In the A^2DM, the model adaptive removal of both ocular artifacts and muscle artifacts in a unified model according to the artifact type.

(2) We design a hard-attention-based mode enhancement module for the EEG denoise task, which can effectively enhance critical mode components in the frequency domain. We demonstrate that hard attention is more effective in FEM compared to soft attention in the removing artifact task. To our best known, *FEM* is the first module based on the convolution network to represent artifacts feature from the view of the frequency domain.

(3) We demonstrate the effectiveness of the proposed A^2DM on the EEGdenoiseNet dataset with comprehensive experiments, the results of which show that A^2DM exhibiting a notable 12% improvement in correlation coefficient metrics compared to advanced methods, such as novel CNN.

The paper is organized as follows: Section "Methods" illustrates our framework for A^2DM and dataset. Section "Results" demonstrates the effectiveness of the A^2DM and analyzes each module of the A^2DM. Section "Discussion" discusses the A^2DM. Section "Conclusion and Future Work" concludes the paper.

Methods

The Framework of A^2DM

The framework of the Artifact-Aware EEG Denoising Model (A^2DM) is shown in Fig. 2. The A^2DM framework consists of the *AAM* and 6 denoise blocks, and then there is a fully connected layer (FC) following the sixth denoise block. We defined the EEG segment containing artifact as noisy EEG segments (EEG_{noisy}). The EEG_{noisy} is fed into the *AAM* (Fig. 2a) to produce the artifact representation (AR). Then, the noisy EEG segments and the artifact representation information are calculated through denoise blocks. The output of FC layer is the denoised EEG segments ($EEG_{denoise}$), which can be formulated as

$$EEG_{denoise} = A^2DM(EEG_{noisy}, AAM(EEG_{noisy})). \quad (1)$$

In each denoise block, there are two 1D-convolution (1D-Conv) layers followed by a ReLU layer. This layer increases the channels of EEG from 32 to 1024, providing more features for the subsequent learning procedure, as shown in the shadowed blocks in the left-bottom of Fig. 2. The *f*requency *e*nhancement module (*FEM*) (details shown in Fig. 2c) and the *t*ime-domain *c*ompensation module *f*usion module (*TCM*) (details shown in Fig. 2d) are followed by the 1D-Conv layers. The details of these two modules are explained in Sections "Frequency Enhancement Module" and "Time-Domain Compensation Module." The residual information [21] is involved in the network of the denoise block. At the tail of the denoise block is a 1D-Conv layer with 2 steps to reduce the dimension of the feature map, which is similar to VGGNet [22]. The architecture of the network is discussed in Section "Computational Resources Analysis."

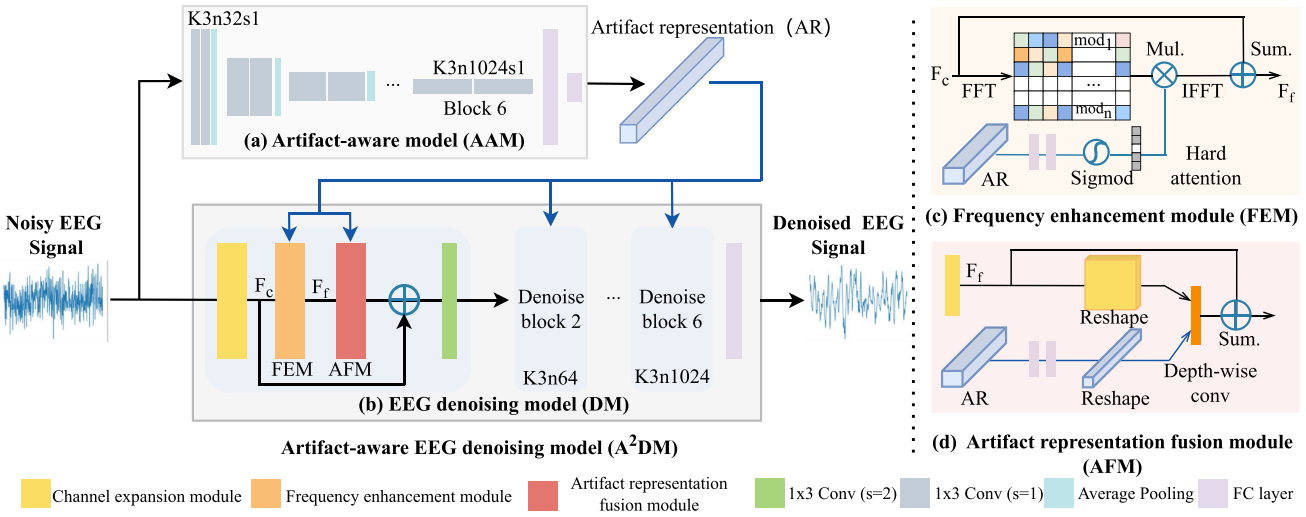

Fig. 2 The framework of A^2DM. The model has mainly two networks: an artifact-aware module and an EEG denoising model. The *AAM* extracts the artifact representations from EEG. In *DM*, the frequency enhancement module (*FEM*) and the time-domain compensation module (*TCM*) are used to fuse the artifact representation and EEG feature map from the frequency-time domain. K3n32s1 means that the size of the convolution kernel (k) is 1×3, the output channel (n) is 32, and the convolution stride (s) is 1

Artifact-Aware Module

EOG and EMG in the EEG signal are different types of artifacts with different statistical feature information in the time and frequency domain space [17, 18]. To this end, we introduce the concept of artifact representation, which captures artifact characteristics in EEG signals and serves as prior knowledge to guide the training of a denoising model. It is important to note that artifact representation is distinct from artifacts themselves; instead, it encapsulates high-level features derived from artifacts. In this way, the noisy EEG signals can be adapted to specific artifact spaces during the denoising process, leading to improved model performance.

The A^2DM aims to distinguish artifacts from other information in the representation space of the EEG signal by learning the artifact representation with the artifact-aware module (*AAM*). As the patterns of artifacts in EEG signals are variable and heterogeneous, we trained the *AAM* with supervised methods to make it robust. Specifically, we utilized the artifact classification task to pretrain *AAM*. The EEG signals containing multiple artifact types are employed as input for *AAM* training, and we consider two essential features of EEG noise: the artifacts type and the signal-to-noise ratio (SNR) level. More details are provided in Section "Dataset." In the pretrain process, the output of the *AAM* is the result of artifact classification. In this way, we make the *AAM* learn the difference between types of artifacts in the space of abstract representation. Finally, the artifact representation is the output embedding of the last FC layer of the *AAM*.

The backbone of the artifact-aware module is shown in Fig. 2a. *AAM* contains 6 blocks used for feature extraction of artifact information. In each of the first five blocks, two 1D-Conv layers with small 1×3 kernels, 1 stride, and a ReLU activation function are followed by a 1D-Global Average Pooling (GAP) with a pool size equal to 2. In the 6th block, two 1D-Conv layers are followed by a flattened layer. The number of channels of the convolutional layer in each block follows base 2 exponential functions from 32 to 1024.

The backbone of the *AAM* is designed based on the structure of VGGNet [22]. This model without the multi-branch design will reduce memory utilization and accelerate the inference [23]. It is notable that there is no restriction for the network of *AAM*, and other classical networks can be used, such as ResNet [21]. The architecture of the network is shown in Table 1. We train the *AAM* with a cross-entropy loss function as follows:

$$L = \frac{1}{N}\sum_{i} L_i = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic}\log\left(p_{ic}\right). \tag{2}$$

where the M indicated the number of classes, and p_{ic} is the predicted probability.

Frequency Enhancement Module

Different types of artifacts tend to be distributed across different frequency components (mode). Frequency-domain analysis is a robust technique for effectively discriminating

Table 1 Network architecture of AAM

Layer	Type	Input shape	Output shape
Input	-	(1, 1000)	-
Block1	[conv-(1×3,32, s=1)]×2	(1, 512)	(32, 512)
Block1	average pool	(32, 512)	(32, 256)
Block2	[conv-(1×3,64, s=1)]×2	(32, 256)	(64, 256)
Block2	average pool	(64, 256)	(64, 128)
Block3	[conv-(1×3,128, s=1)]×2	(64, 128)	(128, 128)
Block3	average pool	(128, 128)	(128, 64)
Block4	[conv-(1×3,256, s=1)]×2	(128, 64)	(256, 64)
Block4	dropout	(256, 64)	(256, 64)
Block4	average pool	(256, 64)	(256, 32)
Block5	[conv-(1×3,512, s=1)]×2	(256, 32)	(512, 32)
Block5	dropout	(1024, 16)	(1024, 16)
Block5	average pool	(512, 32)	(512, 16)
Block6	[conv-(1×3,1024, s=1)]×2	(512, 16)	(1024, 16)
Block6	dropout	(1024, 16)	(1024, 16)
Block7	Flatten	(1024, 16)	(1,16384)
Block7	FC	(1, 16384)	(1,512)
Block7	FC	(1, 512)	(1,10)

and removing artifact distributions. The *key point* of *FEM* lies in adaptively removing artifacts and preserving pure EEG features as much as possible, depending on the type of artifact. Generating mode selection strategies tailored to specific artifact types is challenging.

To address this issue, a hard attention mechanism is integrated into *FEM*, utilizing a learnable binary mask to facilitate the selection of frequency modes containing pure EEG features. The noisy EEG features are transformed into the frequency domain, where each frequency mode is multiplied by the binary mask sequence, composed of 1 s and 0 s. A value of 1 indicates that the mode should be preserved, while a value of 0 signifies that the mode should be suppressed. Notably, the binary mask is learned directly from artifact representations, producing distinct masks for different artifact types. This adaptability enables the model to effectively and simultaneously remove both EOG and EMG artifacts, enhancing its versatility in handling mixed noise scenarios. Unlike traditional filtering methods, our approach leverages the A^2DM neural network to transform EEG signals into frequency-domain feature maps, enabling precise removal of artifacts. The *FEM* module adaptively selects and retains layer-specific information that is essential for reconstructing artifact-free EEG signals. This process is guided by the loss function, which iteratively optimizes the model until convergence.

Additionally, CNNs exhibit general limitations on modeling explicit long-range relations [24]. It has been proven recently that the global features of data can be captured in the frequency domain [25–28]. By incorporating frequency representation, our model effectively captures long-term interactions in EEG features.

Specifically, for a given EEG feature map in *FEM* $F_c = [X_1, X_2, ..., X_l] \times C$, C is the number of channels. First, *FEM* splits the EEG features along the channel dimension and converts them into the frequency domain using FFT $\mathcal{F}$, as formulated below:

$$\tilde{F_f} = \mathcal{F}(F_c). \tag{3}$$

where $\tilde{F_f} = [mod_1, mod_2, ..., mod_{[l/2]+1}] \times C$, mod_i denotes the different modes in the frequency domain. It is also worth mentioning that the half of the FFT $\{mod[l] : 0 \leq l \leq \lceil l/2 \rceil\}$ contains the full information about the frequency characteristics of $x[l]$, due to the conjugacy of frequency domain.

Then, the artifact representation is fed to two FC layers, and a sigmoid layer σ is employed to produce attention weights $w \in \mathbb{R}^{1\times i}, i \in \{1, 2, ..., [l/2]+1\}$. A selector matrix S is constructed where $S \in \{0, 1\}^{1\times[l/2]+1}$,

$$S_i = \begin{cases} S_i = 1, & w_i > t \\ S_i = 0, & otherwise. \end{cases} \tag{4}$$

where t is a threshold based on empirical data and will be determined in the implementation details in Section "Performance Evaluation on Semi-Synthetic EEG." the select operator is formalized as:

$$S = selector(\sigma(FC(AR))). \tag{5}$$

The hard attention mechanism is employed to guide the model to remove artifact features and preserve pure EEG information. Finally, S is used to select different modes components in $\tilde{F_f}$. $\mathcal{F}^{-1}$ is used to convert from the frequency domain to the time domain, which can be formalized as

$$F_f = \mathcal{F}^{-1}(\tilde{F_f} * S) + F_c. \tag{6}$$

where F_f is the enhancement result by *FEM* and $\mathcal{F}^{-1}$ is the inverse Fourier transform. The shortcut [21] is introduced into the *FEM* to avoid the loss of effective features. The pseudo-code of FEM is shown in Alg. 1.

Algorithm 1 Pseudo-code of FEM

Require:
F_c: EEG feature, the size of feature is $b \times c \times l$.
AR: Artifact representation embedding.
t: The threshold for selecting the frequency mode.
Ensure:
$F_f = DFT(F_c, dim = -1)$
$w_0 = MLP(AR)$
$w_i = \mathbb{I}(w_i \in w_0 > t)$
Update the w_o to get selector S
$F_f = IDFT(F_c * S, \text{dim} = -1) + F_c$

Time-Domain Compensation Module

When computing the selection matrix in *FEM*, if the selection factor is set to zero, the module will assign a zero amplitude to the corresponding mode. However, the same mode has EEG and artifact information in the low-frequency region. The aggressive removal of the mode through hard attention may bring about a loss of EEG information. We aim to compensate for the EEG information. CNNs excel at learning local features directly in the time domain [21]. As a result, we reconstruct this information in the time domain instead of the frequency domain. Consequently, we employ a serial network structure for our A^2DM, in which the *TCM* is executed after the *FEM*. In this way, the *FEM* compresses the EEG information in a specific artifact representation space, while the *TCM* increases the capacity of the representation space by reconstructing the time domain and enriching the feature information. The complementary properties between the two modules lead to improved denoising results. More detail will

be discussed in Section "The Effectiveness of Time-Domain Compensation Module."

Specifically, we built the *TCM* to learn how to predict the kernel of a depth-wise convolution [29] conditioned on artifact representation, as shown in Fig. 2b. Specifically, the one-dimensional artifact representation is fed to two FC layers and reshaped as a two-dimensional convolutional kernel $k \in \mathbb{R}^{c\times 1\times 3}$. The input F_f is processed with the depth-wise convolution and a 1×3 convolution to produce the reconstructed feature. The EEG feature within *TCM* is reconstructed by the guide of artifact representation. This means the *TCM* can also be adaptive to multiple artifacts.

Finally, we achieved adaptive artifact removal through the interaction of AR with *FEM* and *TCM*. Specifically, within *FEM*, AR is employed to generate a selecting mask. FEM leverages FFT techniques to transform EEG features into the frequency domain and uses the mask multiply to the EEG features across the channel. The features computed by *FEM* are then fed into *TCM*. Within *TCM*, AR is utilized to perform depth-wise convolution learning, which is then calculated with the previously obtained features to reconstruct and compensate for the missing EEG information.

Dataset

We use a semi-synthetic EEGdenoiseNet dataset [1] and a real EEG dataset (i.e., BCI Competition IV 2a) in this paper. EEGdenoiseNet contains three classes of data: 4514 clean EEG segments, 3400 ocular artifact segments, and 5598 muscular artifact segments. Those data are real-world EEG derived from different data sets. Moreover, BCI Competition IV 2a provides EEG data collected in experiments for motor imagery classification. It can be used to verify the effectiveness of different denoise models in real scenarios.

To simulate scenarios in which multiple types of artifacts are interleaved in the EEG, our dataset is mixed with both EOG and EMG noise. We used data from EEGdenoiseNet [1] to generate pairs of pure and noisy EEG signals for training and testing the A^2DM. The noisy signals are generated by linearly mixing the pure EEG segments with EOG and EMG artifact segments according to (7):

$$y = x + \lambda \cdot n. \tag{7}$$

where y denotes the mixed 1D signal of EEG and artifacts; x denotes the pure EEG signal as the ground truth; n denotes EOG and EMG artifacts. The SNR of the contaminated segment can be adjusted by changing the parameter λ according to $SNR = 10 \log \frac{RMS(x)}{RMS(\lambda \cdot n)}$. The root mean square (RMS) is defined as $RMS(x) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$ where N denotes the number of temporal samples in the segment x, and x_i denotes the i th sample of segment x.

Artifact Representation Dataset

We build an artifact classification dataset for pretraining the *AAM*. To ensure a fair comparison, we set the SNR levels between -7 dB and 2 dB and select EOG and EMG artifacts, following the parameter settings from previous work [1, 30, 31]. EEG segments containing both EOG and EMG noise segments to perform a 10-classification task. For two types of artifacts, the SNR of noise can be divided into 5 categories, varying from -7dB to 2dB with 2 dB increments. It is because too many classes can lead to fine-grained artifact representations, causing overfitting and negatively impacting the model's generalization ability. Conversely, too few classes cannot distinguish between different artifact types effectively. To ensure consistency in the number of EEG, EOG, and EMG segments during signal and artifact mixing, we randomly reused some segments to increase the number of EEG and EOG to 5000. We reuse this data mixing procedure 10 times, resulting in a training set of 500,000 samples ($5000 \times 2 \times 5 \times 10$).

Semi-Synthetic EEG for Denoise Dataset

As for the *denoise dataset* used in A^2DM, we randomly combined 3000 pairs of pure EEG segments with EOG and EMG artifact segments by linearly mixing, using the SNR sampled from a uniform distribution from -7dB to 2dB with 1 dB increments and employ them as the training set. In this way, we finally generate 60,000 ($3000 \times 2 \times 10$) segments as the trainset and 8000 segments as the test set. The length of the data is 512 as the horizontal coordinate. Figure 3 shows the samples of pure EEG, EOG, EMG, and the noisy EEG segment with different levels of SNRs. It is worth noting that our training set consists of EEG segments contaminated with a mixture of EOG and EMG artifacts. In contrast, the previous training set only included EEG segments with a single type of artifact.

Real EEG

We use a real EEG dataset to test the effectiveness of the A^2DM. The BCI Contest IV 2a (BCI-2a) contains data of 9 subjects performing motor imagery classification (left-hand, right-hand, tongue, and foot movements) [32, 33]. We use pre-processed data from MMCNN [34], which is three channels, and the length of the data is 1000. For each EEG segment, we added noise to the head and tail separately and finally processed all channels. We then randomly selected 3 subjects of data to generate the denoised training dataset. In this way, we also finally generate 60,000 ($3000 \times 2 \times 10$) segments for training.

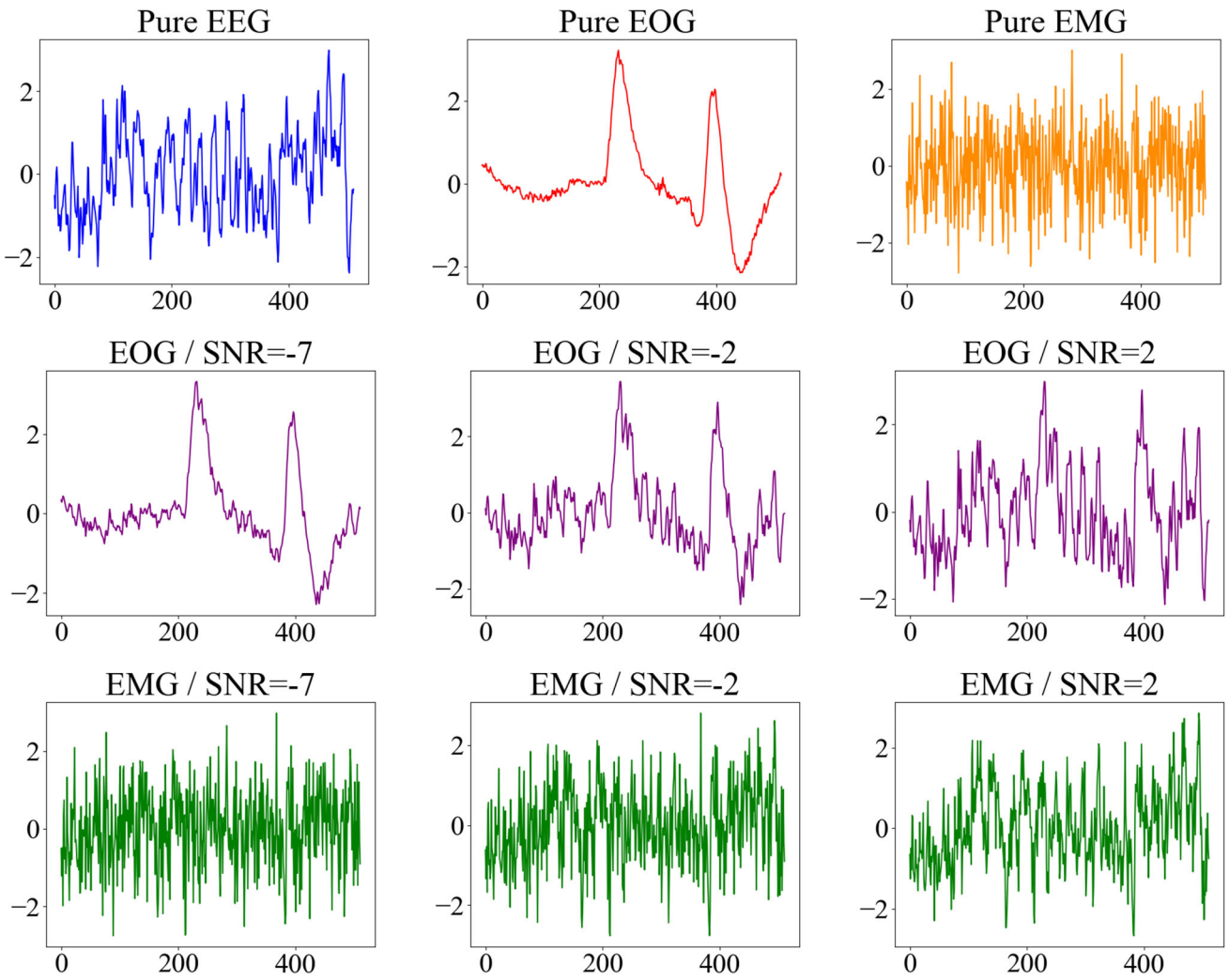

Fig. 3 The synthetic noisy EEG with different types of artifacts and levels of SNRs. The *EOG / SNR = -1* indicates the result of mixing pure EEG and pure EOG at a setting of SNR = -1, according to (7)

Results

Validation Metrics

To quantitatively evaluate the performance of models, three metrics are used on the test dataset, including RRMSE in the temporal domain ($RRMSE_t$), RRMSE in the spectral domain ($RRMSE_f$), and the correlation coefficient (CC).

$$RRMSE_t = \frac{RMS(\hat{y} - y)}{RMS(y)}, \tag{8}$$

$$RRMSE_f = \frac{RMS(PSD(\hat{y}) - PSD(y))}{RMS(PSD(y))}, \tag{9}$$

$$\mathrm{CC} = \frac{Cov(\hat{y}, y)}{\sqrt{Var(\hat{y})Var(y)}}. \tag{10}$$

where RMS is defined as above in Section "Dataset," PSD denotes the power spectral density of the input data; Var and Cov denote the variance and covariance, respectively.

Performance Evaluation on Semi-Synthetic EEG

We employed classical artifact removal methods, specifically the adaptive filter and Hilbert-Huang transform (HHT), for evaluation on the same dataset. Our results indicate that deep learning-based techniques significantly outperform traditional methods in artifact removal, attributed to the robust data fitting capabilities of deep learning models.

To validate the performance of A^2DM, we compared it with four denoising approaches: DeepSeparator [31], novel CNN [16], FCNN [1], simple CNN [1], and complex CNN [1]. The experiment results are shown in Table 2. It can be seen that A^2DM achieves the best performance, with the $RRMSE_t$ and $RRMSE_f$ being 0.6869 and 0.5314, respec-

Table 2 Comparison of experiment results of the models

	EOG			EMG			EOG and EMG		
	$RRMSE_t$	$RRMSE_f$	CC	$RRMSE_t$	$RRMSE_f$	CC	$RRMSE_t$	$RRMSE_f$	CC
Adaptive Filter	1.365	1.213	0.551	2.134	2.310	0.378	1.750	1.762	0.465
HHT	2.067	2.123	0.398	1.635	1.784	0.498	1.851	1.954	0.448
FCNN [1]	0.8548	0.7408	0.5464	0.8525	0.7532	0.5161	0.8536	0.7470	0.5313
Simple CNN [1]	0.7015	0.6087	**0.7647**	0.7147	0.6320	0.7205	0.7081	0.6204	**0.7426**
Complex CNN [1]	0.6898	0.5908	0.7560	0.7358	0.6319	0.7133	0.7128	0.6113	0.7347
DeepSeparetor [31]	0.7598	1.0368	0.7340	0.7433	0.7477	**0.7459**	0.7515	0.8922	0.7399
Novel CNN [16]	0.8216	0.8207	0.5534	0.8303	0.8294	0.5756	0.8260	0.8250	0.5645
Novel CNN$^{\phi}$ [16]	0.7618	0.6093	0.6584	0.7597	0.6047	0.6358	0.7608	0.6070	0.6471
A^2DM(ours)	**0.6700**	**0.5086**	0.7409	**0.7039**	**0.5540**	0.7087	**0.6869**	**0.5314**	0.7248

Novel CNN$^{\phi}$ indicates that the block with 2048 channels is removed in the original novel CNN model
The bold entries indicate the best results

tively. This shows that our model can effectively remove multiple types of artifacts.

Simple CNN achieves the highest CC scores but performs poorly in $RRMSE$ metrics. Our analysis suggests that this outcome may be attributed to the fact that shallow neural networks exhibit greater sensitivity to the trend information in temporal data, allowing them to learn structural information more effectively. Consequently, the denoised signals produced by shallow networks closely resemble the ground truth. In Section "The Insight of Shallow CNN Model for Denoise Task," we conducted a further analysis of this phenomenon.

We find that A^2DM achieves 15% higher CC scores than novel CNN. The reason for this may be due to the inconsistent length of the model input signals. For the fair comparison, we removed the blocks with channel 2048 from the novel CNN so that we obtained the CNN$^{\phi}$ model with the same number of denoise blocks as the A^2DM. Our model also shows a significant improvement over the novel CNN$^{\phi}$ model.

We conduct a visual analysis of denoising results using various models further to validate the effectiveness of A^2DM. Specifically, two distinct EEG segments containing EOG and EMG are randomly selected from the test set to showcase the practical outcomes of the denoising process. The resulting denoise EEG is shown in Fig. 4. A^2DM can effectively remove both EOG and EMG artifacts. Furthermore, we observe that the denoised EEG (depicted by the green line) generated by A^2DM exhibits a similar pattern to the ground

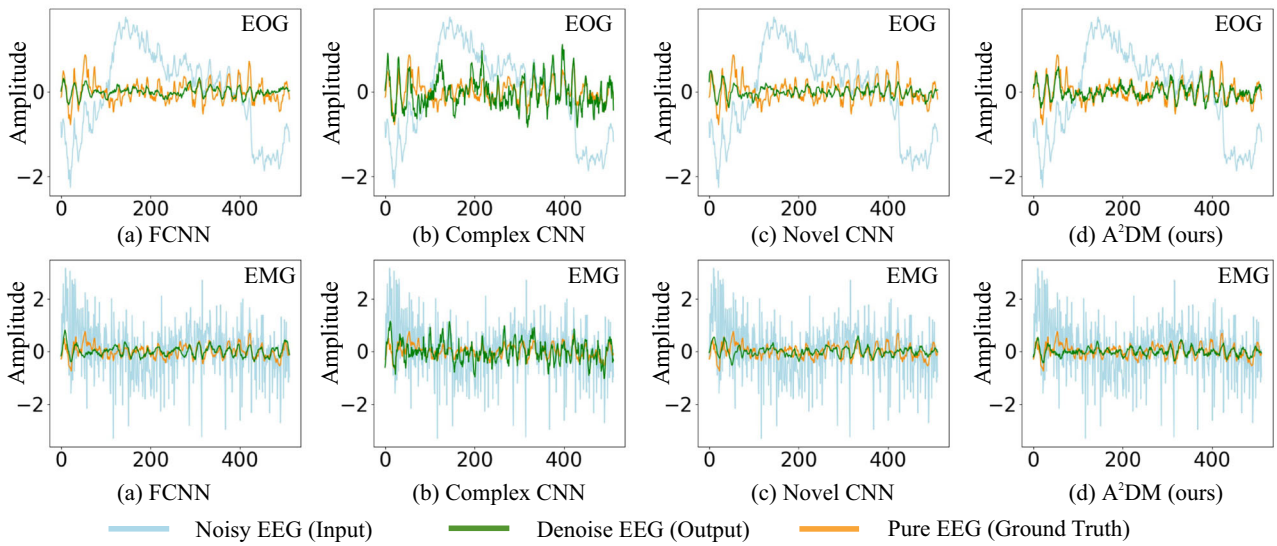

Fig. 4 Denoise performance comparison of different methods on different artifacts from testset; the first and second rows show the EOG and EMG denoising results, respectively. The noisy EEG (input), the denoised EEG (output), and the pure EEG (ground truth) are shown with the light blue line, green line, and orange line, respectively

truth (depicted by the orange line) in both amplitude and tendency, setting it apart from alternative models. This consistency distribution between the ground truth and denoised results is attributed to the A^2DM model's superior ability to preserve the global properties of the EEG data.

In addition, we conduct a time-frequency analysis to evaluate the efficacy of denoise performance. Spectrograms are generated from the input EEG segment, the pure EEG, and the denoised EEG signal using different models, as presented in Fig. 5. Our results demonstrate that the denoised EEG segment produced by the A^2DM exhibits the highest degree of similarity to the pure EEG. This observation suggests that A^2DM is effective in capturing the crucial mode of EEG data, enabling accurate reconstruction of the noisy data. We find that A^2DM tends to remove information in the 20–30 Hz frequency range. This could be attributed to our FEM, which led the model to more aggressively filter out noise while retaining important frequency-domain information. Our model still outperforms other methods in terms of artifact removal. Moreover, it preserves the semantic information of the original signal, especially in the low-frequency region, during downstream tasks. This is a key advantage of A^2DM, as it leverages the FEM module to retain task-relevant information while effectively removing artifacts selectively.

Ablation Studies

To evaluate the efficacy of each module in the proposed A^2DM, we conducted ablation studies. Owing to the inherent flexibility in the design of *FEM* and *TCM*, we could seamlessly remove them from A^2DM without any alteration to the model's overall architecture. In Table 3, we can see that the performance decreases by around 3% without adding *FEM* in model 2 and *TCM* in model 3. This result implies that incorporating *FEM* and *TCM* is beneficial for A^2DM.

Table 3 Ablation study on effects of modules for A^2DM

Model	FEM	TCM	$RRMSE_t$	$RRMSE_f$	CC
A^2DM	✓	✓	**0.6869**	**0.5314**	**0.7248**
Model 2	×	✓	0.7195	0.5783	0.6944
Model 3	✓	×	0.7192	0.5862	0.6947
Model 4	×	×	0.7365	0.6035	0.6752

The bold entries indicate the best results

Next, we analyze *FEM* and its impact on denoising results using $RRMSE_f$ as the evaluation metric. Compared to model 2, model 3 achieves a higher $RRMSE_f$ after incorporating the *FEM* model. This improvement is due to *FEM*'s ability to remove global information. Additionally, the inclusion of *TCM* enhances the denoising performance of A^2DM, highlighting the complementary roles of *FEM* and *TCM*.

Study of Artifact Representation

To demonstrate the efficacy of the artifact representation generated by the *AAM*, we randomly selected 20 samples from the test set for each class of segments and extracted the corresponding artifact representations. We visualized the resulting artifact representations using t-SNE [35] and UMAP [36], as depicted in Fig. 7. The clustering of the same types of artifacts indicates that *AAM* can successfully capture the representation information of artifact in the EEG data.

Furthermore, we introduced a variant (model 4) without the artifact representation by removing the *FEM* and *TCM*. The performance of the model 4 decreased by 4.9% in CC when compared to A^2DM, which highlights the significance of artifact representations in enabling A^2DM to handle multiple artifacts effectively.

We present the denoise results from the baseline model and A^2DM at multiple SNR levels shown in Fig. 6. We

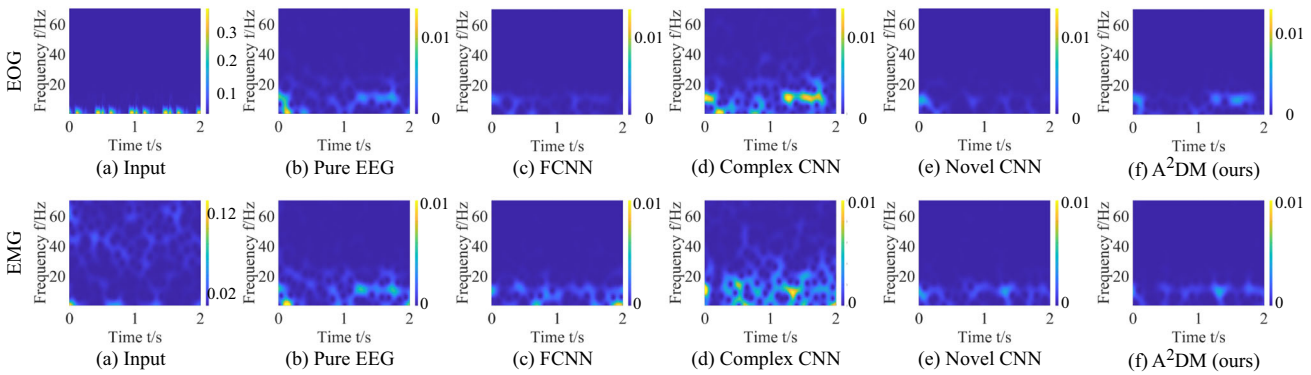

Fig. 5 The time-frequency analysis for different denoise models. We plot the spectrogram of the input EEG (**a**), the pure EEG as ground truth (**b**), and the output EEG denoised by FCNN (**c**), complex CNN (**d**), novel CNN (**e**), our A^2DM (**f**), respectively. The first and second rows show the EOG and EMG spectrogram analysis results, respectively

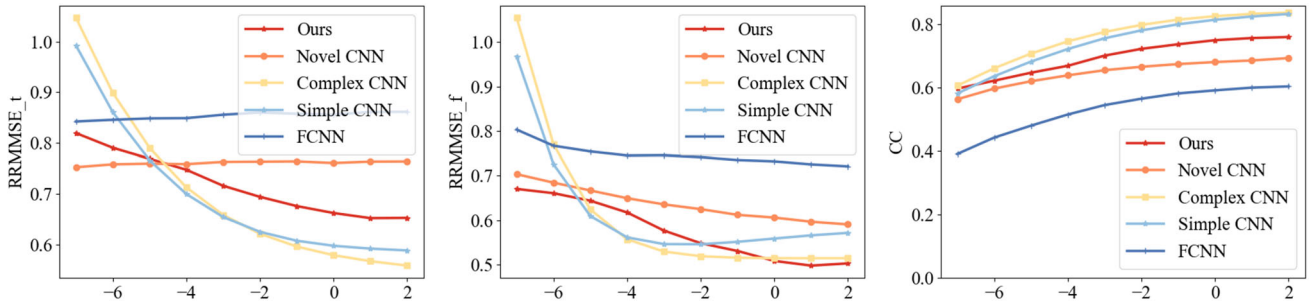

Fig. 6 The temporal RRMSE (RRMSE_t), spectral RRMSE (RRMSE_f), and correlation coefficient (CC) result for EEG denoise at multiple SNR levels. The SNR ranges from -7 dB to 2 dB

observe that as the SNR level increases, the denoising performance of A^2DM improves (RRMSEt: mean=0.6994, std=0.045), while the performance of novel CNN remains largely unchanged (RRMSEt: mean=0.7608, std=0.003). This suggests that A^2DM is adaptable to multiple artifact spaces. On the other hand, the novel CNN models share a single artifact space during the denoising process. Moreover, we observed that simple CNN and complex CNN exhibit higher sensitivity to SNR levels. In Section "The Insight of Shallow CNN Model for Denoise Task," we will discuss the mechanisms of shallow models for denoising tasks.

Analysis of Frequency Enhancement Module

In this section, we analyze the artifact removal process performed by the FEM on EOG and EMG signals separately, using the test dataset for evaluation. Specifically, we compute and visualize the probability density distribution of the selector matrix S, generated by the FEM component within the Denoise Block 1 of the A^2DM, as shown in Fig. 8a. The horizontal axis represents the mode of information retention, where 0 corresponds to complete removal and 1 corresponds to full preservation. A greater tendency towards 0 indicates that more mode information is removed during denoising, while a tendency towards 1 reflects the retention of more modes.

Fig. 7 t-SNE (**a**) and UMAP (**b**) visualization of representation for artifact representation. The dots are the artifact representation in 2D space. Ten types of artifact representation clustered by t-SNE and UMAP have different colors

As illustrated in Fig. 8a, the EMG test dataset exhibits a prominent peak around a mean value of 0.45, significantly higher than that of the EOG dataset. This indicates that the FEM removes more mode information when processing EMG artifacts, likely due to the broader frequency domain distribution of EMG artifacts compared to EOG artifacts.

Conversely, the FEM retains more mode information when addressing EOG artifacts, as indicated by a peak around a mean value of 1. This can be attributed to the narrower frequency range of disturbances associated with EOG artifacts relative to EMG artifacts. These findings highlight the FEM's capability to adaptively discriminate between different types of artifacts based on their frequency distribution characteristics.

To further evaluate the FEM's sensitivity to the removal of EOG and EMG artifacts in the low-frequency domain, we analyzed the distribution of removed modes. Figure 8b depicts the histogram of removed modes for EMG artifact processing, while Fig. 8c shows the corresponding histogram for EOG artifact processing. Here, the horizontal axis represents various modes, and the vertical axis denotes the average mask values across these modes.

Our analysis reveals that within the 5–12 Hz range, the FEM removes more low-frequency information for EOG artifacts than for EMG artifacts. This indicates that the FEM effectively identifies and removes the distribution range of EOG artifacts. Furthermore, this targeted removal approach, based on hard attention mechanisms, can be effectively complemented by the TCM, enabling a more refined denoising process.

The Effectiveness of Hard Attention in FEM

We validate the advantages of the hard attention mechanism in FEM for artifact removal tasks. We utilize a soft attention mechanism to rescale the frequency domain modes within the FEM. Specifically, the artifact representation is processed through the MLP layer and mapped via a sigmoid function to learn a set of modulation factors. This means that instead of discrete values within the selection matrix S, there is a con-

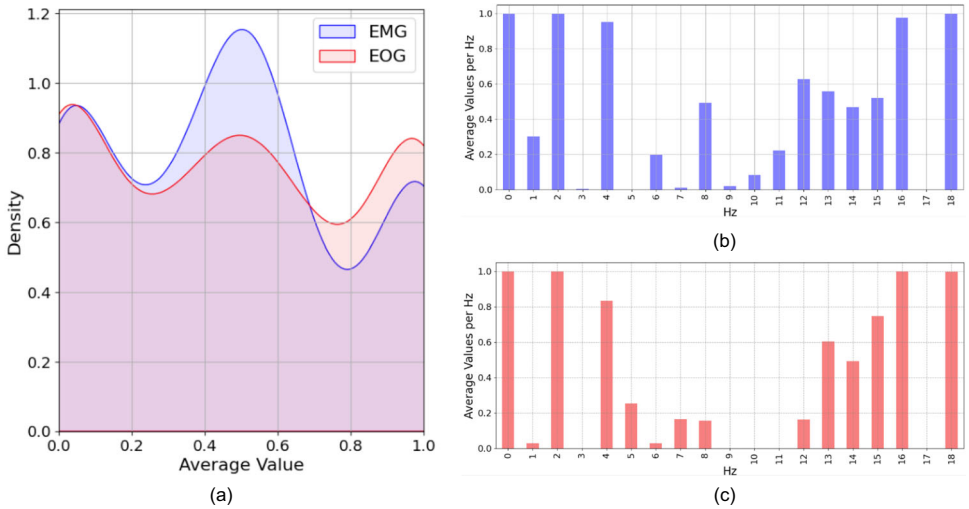

Fig. 8 Analysis of frequency enhancement module. **a** The probability density distribution of the selector matrix S generated by the FEM on the testset. **b** and **c** show histograms of the distribution of selector matrix S in the low-frequency domain, corresponding to the EMG and EOG testsets, respectively

tinuous sequence of matrices. These factors are used to scale the frequency domain modes while the rest of the network structure remains unchanged, resulting in the soft attention-A^2DM model.

As shown in Table 4, the soft attention-A^2DM model exhibits a decline in denoising performance across all three metrics compared to the A^2DM model. This phenomenon can be explained from a frequency domain perspective. Different artifacts and original data are distributed across distinct frequency modes. The *FEM* with a hard attention mechanism can adaptively remove the modes containing artifact features. However, the soft attention mechanism dynamically adjusts the distribution of data across different modes without directly eliminating artifact features, resulting in less effective artifact removal compared to hard attention.

The Effectiveness of Time-Domain Compensation Module

In this subsection, we analyze the impact of *TCM* on the denoising model. In Fig. 9a, we show the architecture of the A^2DM reported in the paper where the *FEM* is executed before the *TCM* and the overall network structure is serial.

Table 4 Comparison of experiment results of the variant models

Model	$RRMSE_t$	$RRMSE_f$	CC
A^2DM (ours)	**0.6869**	**0.5314**	**0.7248**
Soft attention-A^2DM	0.6993	0.5893	0.7128
Dual-stream network	0.7382	0.5972	0.6783
Inverted denoise block	0.7082	0.5817	0.7049

The bold entries indicate the best results

We investigate how the positioning of *FEM* and *TCM* affects the denoising model's performance. The inverted denoising block is shown in Fig. 9b, where the *TCM* is executed before the *FEM*. As shown in Table 4, this structure performs worse than A^2DM. Placing *FEM* after *TCM* results in information loss due to the hard attention mechanism, thereby reducing denoising performance. This implies the *TCM* supplements the lost EEG information, thereby further enhancing the model's performance. The representation of the time and frequency domains is complementary to the representation of the EEG signal in our A^2DM.

Moreover, we analyze the relationship between the frequency and the time domain representation within the *FEM* and *TCM*.

Figure 9c depicts a dual-stream network architecture where *FEM* and *TCM* operate in parallel, with the output features averaged across the two branches. In the experiment, we observe the network converges in the *FEM* branch faster than in the *TCM* branch. In addition, the *FEM* extracts features in the frequency domain, while the *TCM* extracts features in the time domain. The data distribution gap between the two branches further reduces the performance of the network. As Table 4 shows, the model performance is reduced by 4.6% in CC compared to A^2DM. Finally, we use the serial setting in the A^2DM.

Performance Evaluation on Real EEG

To evaluate our method on real data, we conducted the following experiments. We selected the MMCNN [34] model, which contains five modules, for the classification task. To accelerate training, we only used the EIN-a module.

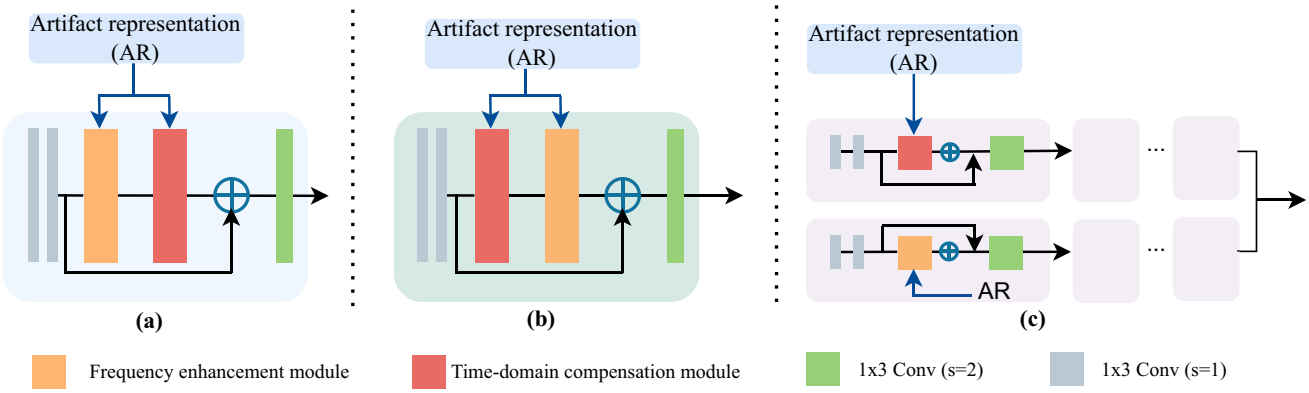

Fig. 9 Comparisons between the two versions of A^2DM. **a** depicts the framework of our A^2DM, **b** is a *dual-stream network* consisting FEM and TCM, and **c** is an *inverted denoise block* used in the variants for A^2DM

As seen in Fig. 10, we added noise to the BCI-2a data to obtain the noisy EEG data. We divide the noisy BCI-2a data into training and test sets. The training set is used to retrain the denoising model, which is then applied to the test set to produce denoised BCI-2a data. The denoised BCI-2a data are used to complete the motion imagery classification task in the ELN-a module. The accuracy of the model is 59.37%.

To evaluate the effectiveness of the A^2DM, we first trained the EIN-a module using the noisy BCI-2a data as the baseline model. Additionally, we generated denoised data using different denoising models for the motor imagery classification task. For comparison, we selected the complex CNN [1] and novel CNN [16] models. We employed cross-validation to further reduce subject bias.

As can be seen from Table 5, we find that the model trained by dataset denoising from A^2DM improved the classification accuracy by 3.36% over the baseline model and 2.47% over the novel CNN. Experiments demonstrate the effectiveness of model denoising.

The Generalization of Hyperparameter *t*

To select the hyperparameter and evaluate the generalization of t, we split the dataset into training, validation, and test sets with a ratio of 8:1:1. Based on validation set performance, we set t=0.4 and evaluated it on the test dataset, as shown in Fig. 11. This setting yielded the best denoising performance compared to other thresholds.

In this study, we present findings regarding the selection of the threshold parameter t to identify the number of modes to be enhanced in the EEG frequency domain. Our experiments show that approximate results can be achieved when t is set to 0.4 or 0.7. The comparable outcomes suggest that the model may have an inherent robustness to noise or redundant information, allowing it to perform well under different filtering conditions.

It is worth noting that a larger t value may lead to the discarding of important features, resulting in a loss of feature information. Conversely, a smaller t value may introduce noise from the high-frequency domain, thereby reducing denoising performance. In light of the above considerations and the need to optimize the generalization ability of the model, we choose t=0.4. It is worth mentioning that the choice of t can be reassessed based on the performance of the downstream task.

Discussion

Computational Resources Analysis

Table 6 lists the details of A^2DM. We used the *ptflops* tool [37] to calculate the parameters and floating point operations

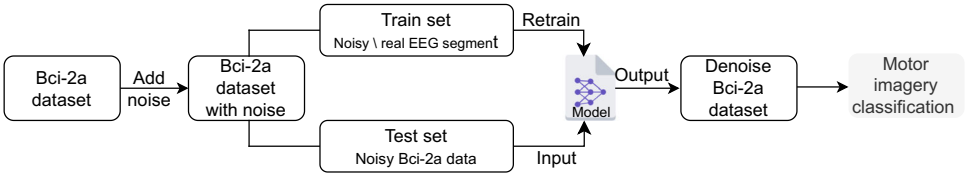

Fig. 10 Schema of the real-data experiment

Table 5 Performance comparison of the MMCNN model trained by denoising data generated by different denoising models

Model	Baseline	Complex CNN [1]	Novel CNN [16]	A^2DM
ACC	56.01±2.64	54.10±0.02	56.90±1.65	59.37±1.83
Recall	54.01±1.78	53.56±0.10	57.20±1.97	58.98±1.12
F1	54.99±1.63	53.83±0.08	57.05±1.24	59.17±1.65

(Flops) of A^2DM and the baseline method in this paper. The specific results are shown in Table 7. We can see that the parameter of A^2DM is 1.31 times that of novel CNN, but the denoising performance is significantly improved. Furthermore, we verified the inference efficiency of the model. In the same workstation environment with Intel Core i7-8700 CPU, 32GB memory, and an NVIDIA TITAN RTX GPU, the inference time of novel CNN on the test set is 45.59 s, and that of A^2DM is 67.97 s. Although the inference time of A^2DM is relatively large, the average inference time of a segment is 0.008 s, which can meet the requirements of real-time applications.

The Insight of Shallow CNN Model for Denoise Task

As shown in Table 2, A^2DM does not perform as well as the complex CNN and simple CNN in terms of CC metrics. Besides, this phenomenon is consistent with the denoising result. As shown in Fig. 4, the complex CNN model fits the structural features and trends of the pure EEG well, resulting in high CC values. However, it is not very sensitive to changes in values, leading to high $RRMSE$ values. In this section, we explore the reasons why the complex CNN has a higher CC value but a lower $RRMSE$.

Firstly, we introduce the structure of the complex CNN model, which is a shallow network with multiple branches. It utilizes 1×3, 1×5, and 1×7 convolutional kernels to extract features from EEG data. The features are fused through concatenation, and the denoised EEG data is generated through fully connected layers.

To explore the mechanisms behind generating denoised data, we visualized the channel feature maps before the fully connected output layer of the model. The input data, pure EEG, and channel feature maps are shown in Fig. 12. We considered different scenarios under various SNRs (high and low SNRs correspond to levels of 2 and -7, respectively).

As shown in Fig. 12, for the low SNR denoise task, the channel feature map exhibits significant differences from the pure EEG distribution but aligns well with the input EEG signal distribution, even displaying some highly similar segments. In contrast, in the high SNR denoise task, the channel feature map gradually approaches the pure EEG distribution and the occurrence of similar segments increases. It is worth noting that in the high SNR scenario, the input EEG and pure EEG have highly similar distributions, and the data segments have overlapping features. It seems that the complex CNN model simply learns the characteristics of the input signal and thus achieves good performance in CC value.

Specifically, the complex CNN network, with its multi-branch architecture and strong ability to extract local features, is capable of accurately capturing the structural feature of the target, thereby achieving a high CC value. However, due to the limitations in learning and network expression brought about by its shallow architecture, the denoise signal output by the network still contains a large amount of structural information from the input signal. Additionally, we have derived the following corollary: The shallow model's

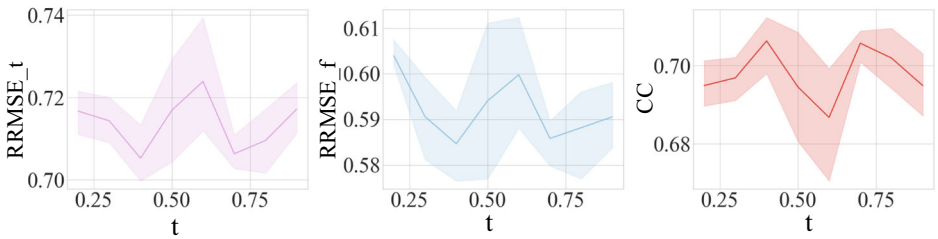

Fig. 11 Results of evaluation metrics using different selection thresholds on the testset

Table 6 The network architecture of A^2DM

Layer	Type	Input shape	Output shape
Input	Noisy EEG and AR	(1, 1000) and (1, 10)	-
Block1	[conv-(1×3,32, s=1)]×2	(1, 512)	(32, 512)
Block1	FEM	(32, 512)	(32, 512)
Block1	TCM	(32, 512)	(32, 512)
Block1	conv-(1×3,32, s=2)	(32, 512)	(32, 256)
Block2	[conv-(1×3,64, s=1)]×2	(32, 256)	(64, 256)
Block2	FEM	(64, 256)	(64, 256)
Block2	TCM	(64, 256)	(64, 256)
Block2	conv-(1×3,64, s=2)	(64, 256)	(64, 256)
Block3	[conv-(1×3,128, s=1)]×2	(64, 128)	(128, 128)
Block3	FEM	(128, 128)	(128, 128)
Block3	TCM	(128, 128)	(128, 128)
Block3	conv-(1×3,128, s=2)	(128, 128)	(128, 64)
Block4	[conv-(1×3,256, s=1)]×2	(128, 64)	(256, 64)
Block4	dropout	(256, 64)	(256, 64)
Block4	FEM	(256, 64)	(256, 64)
Block4	TCM	(256, 64)	(256, 64)
Block4	conv-(1×3,256, s=2)	(256, 64)	(256, 32)
Block5	[conv-(1×3,512, s=1)]×2	(256, 32)	(512, 32)
Block5	dropout	(512, 32)	(512, 32)
Block5	FEM	(512, 32)	(512, 32)
Block5	TCM	(512, 32)	(512, 32)
Block5	conv-(1×3,512, s=2)	(512, 32)	(512, 16)
Block6	[conv-(1×3,1024, s=1)]×2	(512, 16)	(1024, 16)
Block6	dropout	(1024, 16)	(1024, 16)
Block6	FEM	(1024, 16)	(1024, 16)
Block6	TCM	(1024, 16)	(1024, 16)
Block6	conv-(1×3,1024, s=1)	(1024, 16)	(1024, 8)
Output	Flatten	(1024, 8)	(1,8192)
Output	FC	(1, 8192)	(1,512)

$RRMSE$ metric exhibits high sensitivity to variations in SNR. In denoising tasks with low SNR, the input EEG's significant distributional differences from the pure EEG yield higher $RRMSE$ values, whereas, in high SNR denoising tasks, these differences are smaller and thus result in lower $RRMSE$ values. This trend is illustrated in Fig. 6, where the $RRMSE$ value experiences a sharp decrease as SNR increases.

Furthermore, it is challenging for the network to meet the training requirements at the task level. According to Table 5, the EEG signals denoised by the complex CNN model perform even worse in downstream tasks than the noisy EEG signals, indicating that the shallow model has compromised the representation information of the original EEG signals, resulting in poor downstream task performance.

Table 7 Comparison of experiment results of the variant models

Model	Parameters (M)	FLOPs (M)	Time (s)
FCNN	1.05	1.04	3.06
Simple CNN	16.82	36.27	8.72
Complex CNN	8.42	42.61	24.25
Novel CNN	33.56	307.33	45.59
A^2DM (ours)	44.08	658.42	67.97

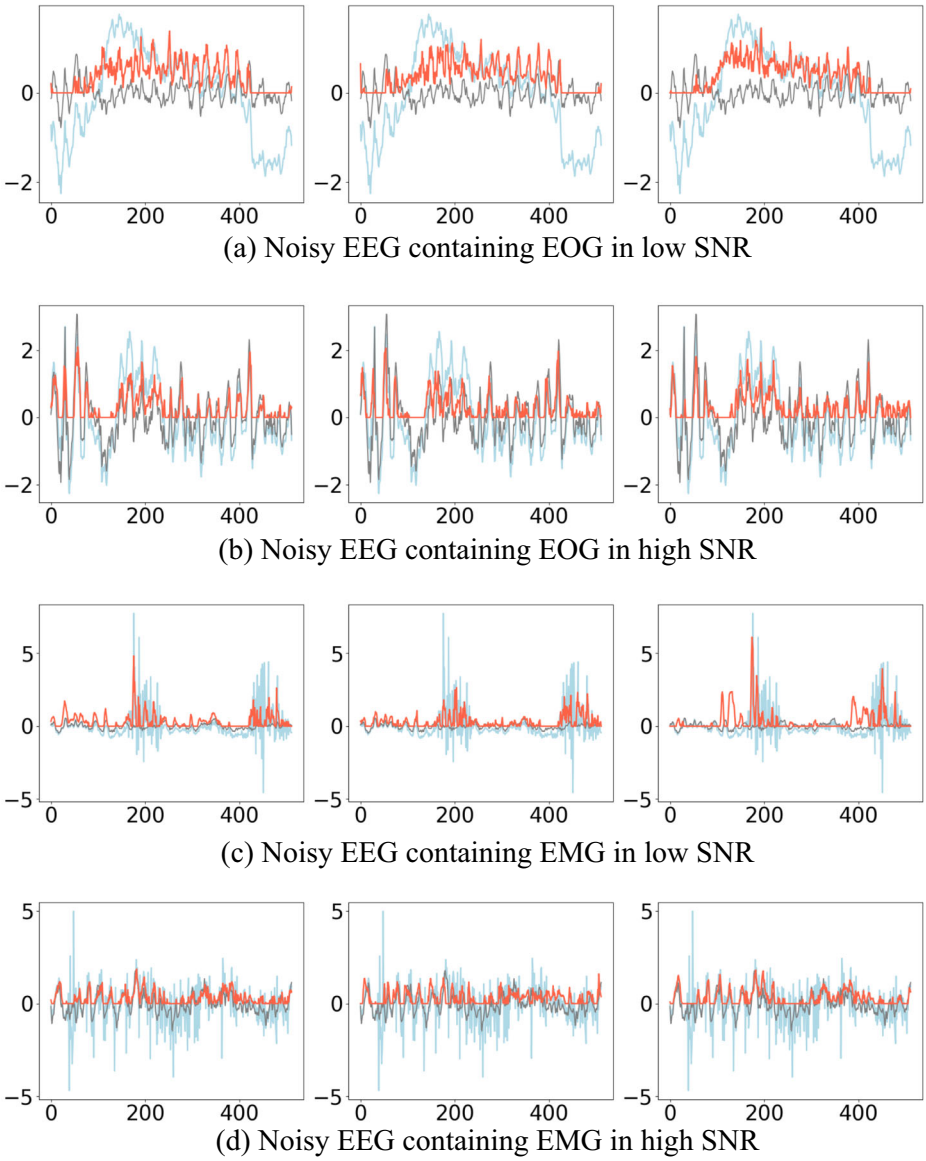

Fig. 12 Visual analysis of channel feature maps. We visualized the channel feature maps before the fully connected output layer of the model. We randomly select three feature maps. The noisy EEG (input), the feature map, and the pure EEG (ground truth) are shown with the light blue line, red line, and grey line, respectively

Conclusion and Future Work

In this paper, we propose an A^2DM for multi-type artifact denoising. This is the first denoising model to consider artifact representations. Experiments have demonstrated that artifact representations of EEG can improve the performance of the denoising network. Finally, we test our model on the EEGdenoiseNet benchmark; the results show that our method achieves state-of-the-art performance in multi-type EEG denoise tasks. The metrics for the correlation coefficient have improved by 12% to novel CNN.

There are several promising directions for future work. First, A^2DM currently focuses on two types of artifacts, EOG and EMG, while considering limited SNR levels. In future research, we aim to explore self-supervised representation learning methods for artifact signals, enabling more fine-grained and robust representations of artifacts within EEG data. Additionally, we plan to extend our model's application to event-related synchronization (ERS) and event-related desynchronization (ERD), further evaluating its effectiveness in these specific EEG analysis scenarios.

Acknowledgements In the paper, we utilized the semi-synthetic benchmark data, EEGdenoiseNET, which is available at (https://github.com/ncclabsustech/EEGdenoiseNet), and the real-world dataset for motor imagery provided by BCI competition IV is available at (http://www.bbci.de/competition/iv/#dataset2a). The data we used does not involve ethical issues.

Author Contributions H.L. contributed to method design, implementation, evaluation, and writing of the original draft. F.F. was involved in method design and writing of the original draft. J.K. conducted formal analysis and method evaluation. J.Z. also conducted formal analysis and

method evaluation. X.G. was responsible for formal analysis, method design, and writing of the original draft. T.L. handled method evaluation. S.L. took part in formal analysis and method evaluation. Z.S. focused on method evaluation and writing—review and editing. J. S-C. also focused on method evaluation and writing—review and editing.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was supported in part by the Natural Science Foundation of China (62372254, 62172239), Innovative Development Joint Fund Key Projects of Shandong NSF (ZR2022LZH009, ZR2023LZH003), and Open Fund of PDL (WDZC20245250102).

Data Availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. Zhang H, Zhao M, Wei C, Mantini D, Li Z, Liu Q. EEGdenoiseNet: a benchmark dataset for deep learning solutions of EEG denoising. J Neural Eng. 2021;18(5):056057. Available from: https://dx.doi.org/10.1088/1741-2552/ac2bf8.
2. Solé-Casals J, Vialatte FB. Towards semi-automatic artifact rejection for the improvement of Alzheimer's disease screening from EEG signals. Sensors. 2015;15(8):17963–76. Available from: https://www.mdpi.com/1424-8220/15/8/17963.
3. Abenna S, Nahid M, Bouyghf H, Ouacha B. EEG-based BCI: a novel improvement for EEG signals classification based on real-time preprocessing. Comput Biol Med. 2022;148:105931. Available from: https://www.sciencedirect.com/science/article/pii/S0010482522006643.
4. Zhu J, Jiang C, Chen J, Lin X, Yu R, Li X, et al. EEG based depression recognition using improved graph convolutional neural network. Comput Biol Med. 2022;148:105815. Available from: https://www.sciencedirect.com/science/article/pii/S0010482522005765.
5. Fatimah B, Singhal A, Singh P. A multi-modal assessment of sleep stages using adaptive Fourier decomposition and machine learning. Comput Biol Med. 2022;148:105877. Available from: https://www.sciencedirect.com/science/article/pii/S0010482522006278.
6. Dai Y, Duan F, Feng F, Sun Z, Zhang Y, Caiafa CF, et al. A fast approach to removing muscle artifacts for EEG with signal serialization based ensemble empirical mode decomposition. Entropy. 2021;23(9):1170. Available from: https://doi.org/10.3390/e23091170.
7. Rajabioun R, Akyürek AÖ, Sezer EA. Deep learning approach for EEG artifact identification and classification. IEEE; 2021. p. 320–325. Available from: https://ieeexplore.ieee.org/abstract/document/9558979.
8. Zhao M, Bonassi G, Guarnieri R, Pelosin E, Nieuwboer A, Avanzino L, et al. A multi-step blind source separation approach for the attenuation of artifacts in mobile high-density electroencephalography data. J Neural Eng. 2021;18(6):066041. Available from: https://dx.doi.org/10.1088/1741-2552/ac4084.
9. Pu X, Yi P, Chen K, Ma Z, Zhao D, Ren Y. EEGDnet: fusing non-local and local self-similarity for EEG signal denoising with transformer. Comput Biol Med. 2022;151(PA). Available from: https://doi.org/10.1016/j.compbiomed.2022.106248.
10. Wang G, Teng C, Li K, Zhang Z, Yan X. The removal of EOG artifacts from EEG signals using independent component analysis and multivariate empirical mode decomposition. IEEE J Biomed Health Inform. 2016;20(5):1301–8. Available from: https://ieeexplore.ieee.org/abstract/document/7134704.
11. Vialatte FB, Solé-Casals J, Cichocki A. EEG windowed statistical wavelet scoring for evaluation and discrimination of muscular artifacts. Physiol Meas. 2008;29(12):1435. Available from: https://doi.org/10.1088/0967-3334/29/12/007.
12. Jiang X, Bian GB, Tian Z. Removal of artifacts from EEG signals: a review. Sensors. 2019;19(5):987. Available from: https://www.mdpi.com/1424-8220/19/5/987.
13. Urigüen JA, Garcia-Zapirain B. EEG artifact removal—state-of-the-art and guidelines. J Neural Eng. 2015;12(3):031001. Available from: https://iopscience.iop.org/article/10.1088/1741-2560/12/3/031001/meta.
14. Zhang Z, Yu X, Rong X, et al. A novel multimodule neural network for EEG denoising. IEEE Access. 2022;10:49528–41. Available from: https://ieeexplore.ieee.org/abstract/document/9770811.
15. Sawangjai P, Trakulruangroj M, Boonnag C, Piriyajitakonkij M, Tripathy RK, Sudhawiyangkul T, et al. EEGANet: removal of ocular artifacts from the EEG signal using generative adversarial networks. IEEE J Biomed Health Inform. 2022;26(10):4913–24. Available from: https://ieeexplore.ieee.org/abstract/document/9627782.
16. Zhang H, Wei C, Zhao M, Wu H, Liu Q. A novel convolutional neural network model to remove muscle artifacts from EEG. In: 2021 IEEE international conference on acoustics, speech and signal processing (ICASSP 2021). IEEE; 2021. Available from: https://ieeexplore.ieee.org/abstract/document/9414228.
17. Chaumon M, Bishop DVM, Busch NA. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. J Neurosci Methods. 2015;250:47–63. Cutting-edge EEG Methods. Available from: https://www.sciencedirect.com/science/article/pii/S0165027015000928.
18. Pion-Tonachini L, Kreutz-Delgado K, Makeig S. ICLabel: an automated electroencephalographic independent component classifier, dataset, and website. NeuroImage. 2019;198:181–97. Available from: https://www.sciencedirect.com/science/article/pii/S1053811919304185.
19. McMenamin BW, Shackman AJ, Maxwell JS, Bachhuber DR, Koppenhaver AM, Greischar LL, et al. Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. Neuroimage. 2010;49(3):2416–32. Available from: https://www.sciencedirect.com/science/article/pii/S1053811909010817.
20. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR); 2018. Available from: https://openaccess.thecvf.com/content_cvpr_2018/papers/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.pdf.
21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on computer vision and pattern recognition (CVPR); 2016. Available

from: https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf.
22. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations; 2015. Available from: https://arxiv.org/abs/1409.1556.
23. Ding X, Zhang X, Ma N, Han J, Ding G, Sun J. RepVGG: making VGG-style ConvNets great again. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021. p. 13733–13742. Available from: https://openaccess.thecvf.com/content/CVPR2021/papers/Ding_RepVGG_Making_VGG-Style_ConvNets_Great_Again_CVPR_2021_paper.pdf.
24. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. ICLR. 2021. Available from: https://arxiv.org/abs/2010.11929.
25. Zhou T, Ma Z, Wen Q, Wang X, Sun L, Jin R. FEDformer: frequency enhanced decomposed transformer for long-term series forecasting. In: Proc. 39th international conference on machine learning (ICML 2022); 2022. Available from: https://proceedings.mlr.press/v162/zhou22g.html.
26. Rao Y, Zhao W, Zhu Z, Lu J, Zhou J. Global filter networks for image classification. In: Advances in neural information processing systems (NeurIPS); 2021. Available from: https://proceedings.neurips.cc/paper_files/paper/2021/file/07e87c2f4fc7f7c96116d8e2a92790f5-Paper.pdf.
27. Wu H, Xu J, Wang J, Long M. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. In: Advances in neural information processing systems; 2021. Available from: https://proceedings.neurips.cc/paper_files/paper/2021/file/bcc0d400288793e8bdcd7c19a8ac0c2b-Paper.pdf.
28. Wu H, Hu T, Liu Y, Zhou H, Wang J, Long M. TimesNet: temporal 2D-variation modeling for general time series analysis. In: International conference on learning representations; 2023. Available from: https://arxiv.org/pdf/2210.02186v2/1000.
29. Chollet F. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1251–1258. Available from: https://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf.
30. Pu X, Yi P, Chen K, Ma Z, Zhao D, Ren Y. EEGDnet: fusing non-local and local self-similarity for EEG signal denoising with transformer. Comput Biol Med. 2022;151(PA). Available from: https://doi.org/10.1016/j.compbiomed.2022.106248.
31. Yu J, Li C, Lou K, Wei C, Liu Q. Embedding decomposition for artifacts removal in EEG signals. J Neural Eng. 2022;19(2):026052. Available from: https://dx.doi.org/10.1088/1741-2552/ac63eb.
32. Sajda P, Gerson A, Muller KR, Blankertz B, Parra L. A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces. IEEE Trans Neural Syst Rehab Eng. 2003;11(2):184–5. Available from: https://ieeexplore.ieee.org/abstract/document/1214716.
33. Blankertz B, Muller KR, Curio G, Vaughan TM, Schalk G, Wolpaw JR, et al. The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. IEEE Trans Biomed Eng. 2004;51(6):1044–51. Available from: https://ieeexplore.ieee.org/abstract/document/1300800.
34. Jia Z, Lin Yea. MMCNN: a multi-branch multi-scale convolutional neural network for motor imagery classification. In: Machine learning and knowledge discovery in databases. Cham: Springer International Publishing; 2021. p. 736–751. Available from: https://link.springer.com/chapter/10.1007/978-3-030-67664-3_44.
35. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9(11). Available from: https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbcl.
36. McInnes L, Healy J, Saul N, Grossberger L. UMAP: uniform manifold approximation and projection. J Open Source Softw. 2018;3(29):861. Available from: https://arxiv.org/abs/1802.03426.
37. Sovrasov V. PTFLOPS: a flops counting tool for neural networks in Pytorch framework. 2018-2023. Available from: https://github.com/sovrasov/flops-counter.pytorch.

Authors and Affiliations

Haoran Li[1] · Fan Feng[2] · Jiarong Kang[1] · Jin Zhang[1] · Xiaoli Gong[1,3] · Tingjuan Lu[4] · Shuang Li[5] · Zhe Sun[6] · Jordi Solé-Casals[7,8,9]

✉ Jin Zhang

✉ Xiaoli Gong

✉ Jordi Solé-Casals
jordi.sole@uvic.cat

[1] College of Computer Science, Tianjin Key Laboratory of Brain Science and Intelligent Rehabilitation, Nankai University, Tianjin, China

[2] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China

[3] State Key Laboratory of High-end Server and Storage Technology, Beijing, China

[4] The 903Rd Hospital of PLA, Hangzhou 310000, China

[5] Weifang Hospital Of Traditional Chinese Medicine, Shandong, China

[6] Faculty of Medicine and Faculty of Health Data Science, Juntendo University, Tokyo, Japan

[7] Data and Signal Processing Group, University of Vic—Central University of Catalonia, 08500 Vic, Catalonia, Spain

[8] College of Artificial Intelligence, Nankai University, 300350 Tianjin, China

[9] Department of Psychiatry, University of Cambridge, Cambridge CB2 3EB, UK