EEGDiR: Electroencephalogram denoising network for temporal information storage and global modeling through Retentive Network

Bin Wang[a,1], Fei Deng[a,*,1] and Peifan Jiang[b,2]

[a] *College of Computer Science and Cyber Security, Chengdu University of Technology, ChengDu, 629100, SiChuan, China*
[b] *College of Geophysics, Chengdu University of Technology, ChengDu, 629100, SiChuan, China*

ARTICLE INFO

Keywords:
Electroencephalogram (EEG) Denoising
Retentive Network (Retnet)
Deep Learning
Signal Embedding

ABSTRACT

Electroencephalogram (EEG) signals play a pivotal role in clinical medicine, brain research, and neurological disease studies. However, susceptibility to various physiological and environmental artifacts introduces noise in recorded EEG data, impeding accurate analysis of underlying brain activity. Denoising techniques are crucial to mitigate this challenge. Recent advancements in deep learning-based approaches exhibit substantial potential for enhancing the signal-to-noise ratio of EEG data compared to traditional methods. In the realm of large-scale language models (LLMs), the Retentive Network (Retnet) infrastructure, prevalent for some models, demonstrates robust feature extraction and global modeling capabilities. Recognizing the temporal similarities between EEG signals and natural language, we introduce the Retnet from natural language processing to EEG denoising. This integration presents a novel approach to EEG denoising, opening avenues for a profound understanding of brain activities and accurate diagnosis of neurological diseases. Nonetheless, direct application of Retnet to EEG denoising is unfeasible due to the one-dimensional nature of EEG signals, while natural language processing deals with two-dimensional data. To facilitate Retnet application to EEG denoising, we propose the signal embedding method, transforming one-dimensional EEG signals into two dimensions for use as network inputs. Experimental results validate the substantial improvement in denoising effectiveness achieved by the proposed method.

1. Introduction

Electroencephalography (EEG) records potential changes on the scalp, originating from neurons in the gray matter [27]. These potential changes are typically captured by an electrode collection system positioned on the brain's surface [21]. Analysis of EEG provides a comprehensive spectrum of physiological, psychological, and pathological insights [18]. However, owing to the high temporal resolution of EEG signals, they are susceptible to various complex noises, including cardiac artifacts, ocular artifacts, muscle artifacts, and external interference [10]. The persistent presence of these disruptive noises during data acquisition poses a substantial challenge in extracting pure EEG signals, imposing significant constraints on subsequent EEG research and applications [15]. Hence, an effective EEG denoising method is urgently required to mitigate the noise in acquired EEG signals while preserving essential information for further EEG research.

Numerous traditional denoising methods have been proposed for mitigating noise in EEG signals, including regression-based and adaptive filter-based approaches. Specifically, regression-based methods derive the noise signal using a predefined noise template. The computed noise signal is then employed as a suppression signal to eliminate noise from the EEG signal, yielding the denoised result [13] [5]. In contrast, the adaptive filter method embraces a distinct concept, suppressing noise by predicting filter coefficients directly from the input EEG signals [7] [11]. Nonetheless, conventional methods possess certain drawbacks. Notably, hyperparameter configuration in traditional methods significantly influences the efficacy of EEG noise removal, necessitating researchers' empirical awareness to set reasonable parameters. Additionally, the traditional approach risks suppressing crucial EEG information while eliminating noise, potentially impacting subsequent research endeavors.

The EEG signal is a complex waveform characterized by nonlinear features crucial for its analysis. Therefore, denoising methods must preserve these nonlinear features while eliminating noise [18]. The advancement in computer processing power and the expansion of EEG datasets [27] have spurred recent research endeavors to leverage deep learning for EEG signal denoising. Commonly employed architectures for EEG denoising networks include feedforward neural networks (FNN) [3] [25], convolutional neural networks (CNN) [1] [18], and recurrent neural networks (RNN) [26], along with their variations, such as long and short-term memory networks (LSTM) [14] [27]. As EEG is collected in the time dimension, establishing a temporal relationship between sampling points, basic network architectures have demonstrated significant improvement in denoising performance compared to traditional methods. However, they face challenges either in retaining temporal information or lacking global modeling capability while preserving temporal information. To address this, some studies have explored integrating the Transformer model [22] into EEG denoising

*Corresponding author. College of Computer Science and Cyber Security, Chengdu University of Technology, ChengDu, 629100, China
woldier@foxmail.com (B. Wang); dengfei@cdut.edu.cn (F. Deng); jpeifan@qq.com (P. Jiang)
ORCID(s):
[1]Bin Wang and Fei Deng are with the College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu 610059, China(e-mail: woldier@foxmail.com, e-mail: dengfei@cdut.edu.cn).
[2]Peifan Jiang are with the College of Geophysics, Chengdu University of Technology, Chengdu 610059, China (e-mail: jpeifan@qq.com).

tasks, as it [16] effectively preserves temporal information and enables efficient data parallel computation, yielding notable results.

In recent years, with the rapid advancement of large-scale language modeling (LLM), a novel network called Retentive Network [19] has emerged. Retentive Network exhibits a favorable disposition towards temporal information, boasts robust global modeling capabilities for nonlinear features, and demonstrates commendable performance. However, when applied directly to denoise EEG signals, a challenge arises. This stems from the fact that EEG signals, akin to natural language, possess temporal characteristics and encompass global nonlinear features. Unfortunately, using Retentive Network directly for EEG denoising is unfeasible due to a misalignment in the dimensional requirements. Retentive Network, designed for two-dimensional input, conflicts with the one-dimensional nature of EEG signals. Unlike the approach in [16], reshaping a 1D signal into a 2D format results in a fixed sum of input dimensions after reshaping, compromising subsequent network feature extraction. To address this issue, we propose a signal embedding method capable of transforming the signal into a sequence of arbitrary length and embedding dimensions, enhancing network flexibility. Additionally, while EEGdenoiseNet [27] introduces a standard deep learning EEG dataset, expediting the development of EEG denoising methods, the dataset remains unprocessed, lacking sample pairs. This necessitates mixing various noise types (muscle artifacts and eye artifacts) during preprocessing. Divergent data preprocessing approaches may yield disparate network results, impeding the comparison of methodologies. To mitigate this, we curated an open-source dataset using the huggingface datasets library [12] from preprocessed data. [1] [2]

The main contributions of this paper can be summarized as follows:

(1) We introduce the Retentive Network, an advanced infrastructure in natural language processing, to the domain of EEG signal denoising. This integration not only offers a novel avenue for exploring this intersection but also broadens the scope of related research. Through this incorporation, we successfully establish an innovative approach for EEG signal denoising, thereby expanding the possibilities for further research in this domain.

(2) To effectively integrate 1D EEG signals with our proposed network, we introduce a method called signal embedding. Unlike a simple reshaping of the signal into 2D, this method achieves an organic fusion of the 1D EEG signal and the network through a sophisticated embedding strategy. This innovative approach not only enhances the network's adaptability but also contributes to an overall improvement in system performance.

(3) When examining the standard deep learning EEG dataset provided by EEGDenoiseNet, we observe that the raw nature of the dataset and the absence of pairs of training samples hinder the comparison of different methods. To address this limitation, we create an open-source dataset using preprocessed data. This not only eliminates challenges related to noise and ensures data consistency but also facilitates the exploration of deep learning-based denoising methods for EEG signals.

[1] https://github.com/woldier/EEGDiR

[2] https://huggingface.co/datasets/woldier/EEGDiRDataset

2. Related work

The regression model serves as a pivotal tool for data smoothing and analysis in EEG signal processing [11]. The primary goal is to establish relationships, be they linear or nonlinear, between EEG signal channels and reference signals like EOG, EMG, or ECG channels. The overarching objective is to formulate mathematical equations through regression analysis to efficiently eliminate noise from EEG data [25]. Fundamentally, the method relies on a widely used regression-based approach, presupposing that each channel comprises a cumulative sum of pure EEG data and a fraction of artifact. This strategy is successful for channels with available reference data, utilizing exogenous reference channels such as ECG to eliminate various artifacts and improve the overall signal quality. However, it's crucial to highlight a notable limitation: the regression-based method encounters challenges when handling electrooculogram (EOG) and electromyogram (EMG) signals due to the lack of an appropriate exogenous reference channel for these specific artifacts.

The Wavelet Transform method [23] is employed to convert time domain signals into both time and frequency domains. This approach is preferred over Fourier transform due to its superior tunable time–frequency tradeoff and capabilities for analyzing non-stationary signals. The Wavelet Transform maps the signal to the wavelet domain, utilizing distinct properties and mechanisms of wavelet coefficients generated by signal and noise across different scales [4]. The primary objective of the method is to eliminate noise-generated wavelet coefficients while maximizing the retention of coefficients originating from real signals.

Traditional methods in EEG signal processing encompass regression modeling and wavelet transform methods. However, these methods exhibit drawbacks, particularly when processing complex EEG signals. The regression model may face limitations due to the absence of specific exogenous reference channels, and the wavelet transform method may lack sensitivity to the time-frequency characteristics of noise. To address these limitations, deep learning methods have emerged as an appealing alternative, leveraging their robust feature learning and representation capabilities, resulting in notable achievements in EEG denoising tasks.

In the training phase of the deep learning method, we utilize the framework illustrated in Figure 1, where Noisy EEG signal and Noise-free EEG signal form the pairs for training. The objective is to model the EEG signal with noise, aiming to produce an output that closely resembles the Noise-free state through the learning of network weight

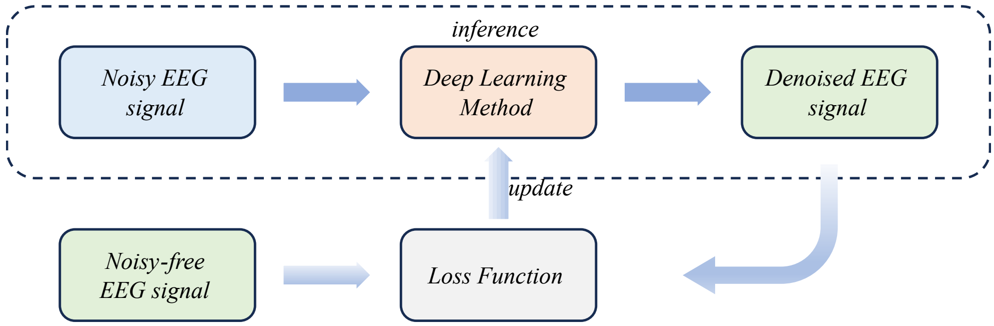


Figure 1: The diagram illustrates the training procedure of the deep learning method. The dataset comprises sample pairs, namely Noisy EEG signal and Noise-free EEG signal, representing EEG signals with and without noise, respectively. Throughout the training process, the Noisy EEG signal serves as the network input. The network, in turn, produces the Denoised EEG signal, which is utilized as the input for the subsequent training steps. The Loss Function calculates the disparity between the Denoised EEG signal and the Noise-free EEG signal, facilitating the optimization of the network. The black dashed box delineates the inference phase of the network, omitting the optimization component. This inference stage can be likened to the end-to-end output where the Noisy EEG signal is input into the network, yielding the Denoised EEG signal as the output.

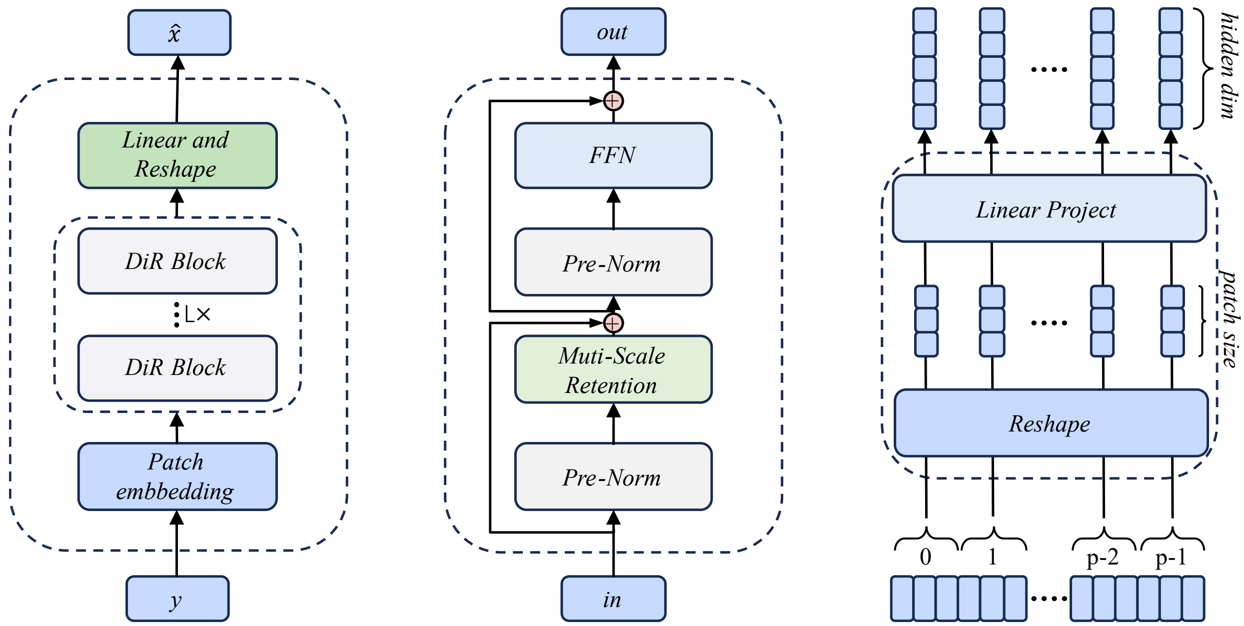
parameters. The training initiates with the input of the Noisy EEG signal into the network, generating the corresponding Denoised EEG signal. The disparity between this Denoised EEG signal and the actual Noise-free EEG signal is quantified as a loss, computed by the Loss Function. The black dashed box in the figure emphasizes the distinction between the signal and the Noise-free EEG signal.

The black dashed box in the figure delineates the network's inference process. During inference, the network directly takes the noisy EEG signal as input and produces the denoised EEG signal as output, bypassing the optimization of weighting parameters. This end-to-end inference capability empowers the network to denoise new and unknown EEG signals in practical applications. The overarching objective of the entire training process is to achieve noise suppression by optimizing the network parameters, enabling it to extract genuine signal features amidst noise interference. Such a design facilitates the network in learning a more efficient representation, subsequently enhancing its denoising performance and providing robust support for real EEG signal processing.

Given the shared structure between the training and inference processes, researchers can concentrate their efforts on investigating the network's architecture. This facilitates a more profound exploration of the application of deep learning in EEG signal processing. In recent years, deep learning methods have made significant strides in domains such as natural language processing [22] [20] and computer vision [6] [17]. Notably, their efficacy extends to signal processing, demonstrating impressive performance in signal denoising [18] [25] [27] [16].

To address ocular artifacts, and muscle artifacts in EEG signals, Yang et al. [25] introduced DLN, a straightforward and efficient fully-connected neural network that surpasses traditional EEG denoising methods in processing efficiency, requiring no human intervention. Sun et al. [18] proposed a one-dimensional residual CNN (1D-ResCNN) model based on convolutional neural network CNN, showcasing superior denoising performance compared to DLN by adeptly employing various convolutional kernel sizes (1×3, 1×5, 1×7) and integrating a residual layer [6]. Zhang et al. [27] presented a comprehensive EEG dataset, reducing the dataset collection challenge, and outlined four fundamental network models utilizing fully connected neural networks (FCNN), convolutional neural networks (CNN), and recurrent neural networks (RNN) for the removal of ocular and muscle artifacts. Additionally, Pu et al. [16] introduced EEGDnet, leveraging the Transformer model, which outperforms prior networks in both nonlocal and local self-similarity within the model architecture. On Zhang et al.'s benchmark EEG dataset, EEGDnet surpasses previous networks in eliminating ocular artifacts, and muscle artifacts. This body of work provides valuable references and innovations to propel the advancement of EEG deep learning.

The temporal information in EEG signals is inherently long-term and characterized by numerous temporal correlations. Traditional methods often encounter difficulties in handling extensive time-series data. However, the integration of deep learning methods proves advantageous in accommodating the temporal intricacies of EEG signals. As EEG signals emanate from the entire brain, comprehensive global modeling becomes imperative for enhanced comprehension and processing. Despite the simplicity and efficiency of the DLN model, its fully-connected structure may exhibit limitations when dealing with prolonged time-series information and global modeling. While the 1D-ResCNN model surpasses DLN in denoising performance, its dependence on a single convolutional kernel size might present constraints. The model could face challenges in addressing multi-scale features and intricate temporal information. In the case of EEGDnet, its incorporation of the Transformer model demonstrates superior architectural performance concerning nonlocal and local self-similarity. However, given the diverse frequencies present in EEG signals, effective feature capture across different scales becomes crucial.



(a) EGGDiR (b) DiR Block (c) Signal Embbedding

Figure 2: (a) illustrates the architecture of the EEGDiR network. This network generates hidden dimensions through Patch Embedding and obtains the output via linear projection and transformation following multi-level DiR Block processing. EEGDiR operates as an end-to-end model, taking a noisy signal as input and producing a noise-free signal, denoted as x_hat. The DiR Block, depicted in (b), comprises Pre-Norm, Multi-scale Retention, and Residual Connection, with Pre-Norm utilizing Layer Normalization. The SignalEmbedding structure, outlined in (c), involves segmenting the input sequence into new sequences based on the patch size. The hidden dimension after reshaping aligns with the patch size, and after linear projection, it matches the final hidden dimension.

3. Method

3.1. Overall structure of the EEGDiR network

In this paper, we present EEGDiR, a novel network model tailored for 1D-EEG signal denoising. Our model incorporates Retnet into the realm of EEG signal denoising, introducing innovative perspectives to signal processing tasks. The overall network structure (see Fig. 2(a)) involves processing the noisy signal y through Signal Embedding, elaborated later, to augment the embedding dimension. Subsequently, the noise signal y undergoes processing via stacked DiR Blocks at multiple levels, with the final output linearly projected to match the input dimension. EEGDiR operates as an end-to-end model, taking a noisy input signal y and generating the corresponding noiseless signal x_hat. Fig. 2(b) depicts the structure of the DiR Block, which begins with pre-Norm, followed by multi-scale Retention and a Residual Connection. The output of the residual join undergoes pre-Norm once more before serving as the input for the Fully Functional Network (FFN). The term "pre-Norm" refers to Layer Normalization, employed for normalization before each submodule. It is crucial to note that an FFN typically includes a fully-connected module with a hidden layer that doubles the hidden dimension. The FFN's output layer reduces the hidden dimension to align with the input. The dimensions of input and output vectors remain constant across DiR and its submodules. Figure 2(c) illustrates the Patch Embedding structure: the input sequence is segmented into new sequences of length $|s|$ // patch_size based on the patch size. The initial hidden dimension of these sequences is equal to the patch size after reshaping. Through linear projection, the hidden dimension of the projected sequences matches the final hidden dimension. Given that EEG signals may encompass multiple frequencies and require effective feature capture at different scales, patch embedding is introduced to intelligently handle temporal information at varying scales. It enhances context preservation and temporal relationship retention by merging consecutive samples into a patch.

- Layer Normalization

  Layer Normalization (LN) [2] is another normalization technique that serves as an alternative to traditional Batch Normalization (BN) [8]. While BN can face challenges in performance when dealing with small batch sizes, Layer Normalization aims to address these issues by normalizing at the layer level. In contrast to BN, which computes mean and variance based on the entire batch, LN calculates normalization statistics independently for each layer. Specifically, the mean and variance are computed over all channels within a single layer, decoupling the normalization process from the batch size. This allows Layer Normalization to better adapt to varying batch sizes and perform effectively in scenarios where BN might struggle. By normalizing at the layer level, LN provides a robust estimation of statistics and can be advantageous in situations where small batch sizes or imbalanced data distribution pose challenges. This approach contributes to improved network performance and helps overcome limitations associated with traditional normalization methods under certain conditions.

- Residual learning

  The use of residual learning is an effective technique to further enhance feature learning in deep convolutional neural networks (CNNs). Deeper CNN architectures have the capability to capture more abstract features and improve performance in tasks such as classification and regression. However, simply increasing the depth of a network does not always lead to better performance, as it can suffer from issues like vanishing or exploding gradients [6]. Residual learning, introduced in the context of the ResNet architecture, addresses these problems by incorporating residual blocks that allow for the direct propagation of information from earlier layers to later layers. By using skip connections, residual learning enables the network to learn residual mappings, which can help mitigate the degradation problem associated with increasing network depth. These residual connections provide shortcuts for gradient flow, making it easier for the network to optimize the deeper layers and improve overall performance. In addition to addressing gradient related issues, residual learning has been observed to enhance network performance. The skip connections allow for the preservation of important information from earlier layers, enabling the network to effectively learn complex representations. This approach has demonstrated improved convergence and achieved state-of-the-art results in various deep learning tasks.

3.2. Muti-Scale Retention

In this study, we explore the integration of the Retentive Network model, originally designed for Natural Language Processing, into the realm of EEG signal denoising. It is important to emphasize that while the Retentive Network has demonstrated remarkable success in natural language processing, our investigation centers on its applicability to EEG denoising. The Retentive Network (Retnet) is comprised of L identical modules arranged in a stacked fashion, featuring residual connectivity and pre-LayerNorm akin to the Transformer architecture. Each Retnet module comprises two sub-modules: the Multi-scale Retention (MSR) module and the Feedforward Network (FFN) module. For a given input sequence $s = s_1 s_2 s_3 \dots s_{|s|}$ (where $|s|$ denotes the length of the sequence), the input vector is initially transformed to $X^0 = \left[x_1, x_2, \dots, x_{|x|}\right] \in \mathbb{R}^{|x| \times d_{model}}$, where d_{model} is the hidden dimension. Subsequently, the Retnet Block can be computed for each layer, denoted as $X^l = Retnet(X^{l-1}), l \in [1, L]$. The Simple Retention layer is defined as follows [19]:

$$Q = (XW_Q) \odot \Theta, \quad K = (XW_K) \odot \bar{\Theta}, \quad V = XW_V \tag{1}$$

$$\Theta_n = e^{in\theta}, \quad D_{mn} = \begin{cases} \gamma^{n-m}, & n \geq m \\ 0, & n < m \end{cases} \tag{2}$$

$$Retention(X) = (QK^T \odot D)V \tag{3}$$

where $\bar{\Theta}$ is the complex conjugate of Θ, $W_Q, W_K, W_V \in \mathbb{R}^{|x| \times |x|}$, $D \in \mathbb{R}^{|x| \times |x|}$ combines causal masking and exponential decay of relative distances into one matrix.

To achieve a multichannel-like effect, input sequences can be projected to lower dimensions d times, akin to the multiple-header mechanism in Transformer. This method is employed in each Retention layer with multiple headers $h = \frac{d_{model}}{d}$, where d represents the length of the sequences in each header. Each header utilizes distinct $W_Q, W_K, W_V \in \mathbb{R}^{|d| \times |d|}$, constituting the Muti-Scale Retention (MSR) block. Different γ hyperparameters are assigned to various heads in MSR, maintaining simplicity with the same γ across different layers. Additionally, a swish [9] gate is incorporated to enhance nonlinear features in each layer. Given an input X, the mathematical representation of the Muti-Scale Retention is provided as follows:

$$\gamma = 1 - 2^{-5 - arrange(0,h)} \in R^h \tag{4}$$

$$head_i = Retention(X, \gamma_i) \tag{5}$$

$$Y = GroupNorm_n(Concat(head_1, \dots, head_h)) \tag{6}$$

$$MSR(X) = (swish(XW_G) \odot Y)W_o \tag{7}$$

Here, $W_G, W_O \in \mathbb{R}^{d_{model} \times d_{model}}$ are learnable parameters, and GroupNorm normalizes the output of each head.

- Group Normalization

  Group Normalization (GN) [24] is an alternative to traditional Batch Normalization (BN) [8] that addresses the issue of poor performance with small batch sizes. When using small batch sizes, BN may not effectively estimate the mean and variance of the entire data distribution. Additionally, BN may yield suboptimal results for tasks with highly imbalanced binary classification. To overcome these limitations, GN divides the channels into groups and performs normalization within each group. Instead of computing the mean and variance based on the batch size, GN computes the mean over $(\frac{C}{G}) \times H \times W$, where C is the number of channels, G is the number of groups, H is the height, and W is the width. By doing so, GN decouples the normalization from the batch size and allows the network to achieve better performance. By normalizing within groups, GN provides a more robust estimation of statistics and helps mitigate the negative impact of small batch sizes or imbalanced data distribution.

3.3. Signal Embedding

In our extended investigation, it was observed that when the input sequence $s = s_1 s_2 s_3 \dots s_{|s|} \in \mathbb{R}^{|s| \times 1}$ is relatively short, direct embedding is feasible. However, in general scenarios, where the time-series information of the signal is usually lengthy, direct embedding incurs high computational complexity, hindering effective network training. To address this, we propose the introduction of a concept termed "patch", involving the amalgamation of a series of consecutive samples into a single input feature. This concept is inspired by speech signal processing, where a solitary sample may inadequately represent the current word, while a segment of samples offers more semantic expressiveness. It is noteworthy that EEG signals frequently encompass extensive temporal information, and signal embedding intelligently captures this temporal data. By grouping consecutive samples into patches, the network better retains context and temporal relationships in the signal, enhancing denoising effectiveness. This approach aligns with speech signal processing, where context is pivotal for accurate speech comprehension. Consequently, this paper introduces signal embedding, a more efficient process tailored to the characteristics of EEG signals.

The complete signal embedding process is illustrated in Figure 2(c). Assuming a given patch size, the original sequence is divided accordingly, reducing the sequence length to $|s|//patch\_size$. Subsequently, each patch undergoes reshaping and linear projection to attain the desired hidden dimension. This process not only mitigates computational complexity but also preserves timing information more effectively. It's crucial to note that the signal embedding used here does not employ positional encoding. This is because the Retention mechanism already incorporates positional encoding considerations, obviating the need for additional positional encoding. Mathematically, it can be expressed as follows:

$$s' = Pathchfiy(s) = s'_1 s'_2 s'_3 \dots s'_{\frac{|s|}{p}} \in \mathbb{R}^{\frac{|s|}{p} \times p} \tag{8}$$

$$X^0 = Embedding(s'; \omega) = [x_1, x_2, \dots, x_{|x|}] \in \mathbb{R}^{\frac{|s|}{p} \times d_{model}} \tag{9}$$

$$X^0 = SignalEmbedding(s; \omega) \tag{10}$$

Equation (8) delineates the patching process, wherein the original sequence s is segmented into smaller sequences through patching, with each patch serving as an input feature. This operation not only preserves the feature information of the input (EEG signal) but also significantly truncates the length of the input sequence, thereby diminishing the computational complexity of subsequent operations. Here, s' denotes the sequence post the patch operation. In Equation (9), we illustrate the feature embedding, signifying that after the patch sequence s', linear projection of the feature size results in the generation of the larger feature size X^0. This dispersion of signal features is conducive to the subsequent network's extraction of diverse features. This process allows the signal features to be spread out, facilitating the network in extracting distinct feature types. Here, ω denotes the learnable parameter, and X^0 remains consistent with the preceding section. If we conceptualize patching and embedding as an end-to-end operation, it can be expressed as (10). In other words, the input s' can be derived from the signal embedding module to yield X^0, with ω serving as the learnable parameter, akin to (9).

4. Experiments and results

4.1. Preliminary

Signals disturbed by noise are acquired through the linear combination of the electrooculogram (EOG) or electromyogram (EMG) with the pristine electroencephalogram (EEG). This procedure can be mathematically represented as Equation (11) [27]. The mixed EEG noise signal is denoted as $y \in \mathbb{R}^{|y|}$, where y represents the sequence length. The noise-free EEG signal, denoted as $x \in \mathbb{R}^{|x|}$, serves as the ground truth, and $n \in \mathbb{R}^{|n|}$ represents oculomotor or electromyographic artifacts (myogenic artifacts). It is important to note that the lengths of each sequence $|x|, |y|, |n|$ are equal. To control the noise level during mixing, we introduce the hyperparameter λ, regulating the signal-to-noise ratio (SNR) of the noisy signal. Adjustment of different λ values enables effective control of SNR magnitude to adapt to various noise environments. The SNR is calculated using Equation (12), while λ is determined by Equation (13), where $RMS(\cdot)$ denotes the root mean square of the sample, $RMS(x)$ is the root mean square of the noiseless EEG signal x, and RMS $RMS(\lambda \cdot n)$ is the root mean square of the mixed noise $\lambda \cdot n$. These formulas provide flexible adjustment of the signal-to-noise balance to meet diverse signal quality requirements in specific application scenarios.

$$y = x + \lambda \cdot n \tag{11}$$

$$SNR = 10log(\frac{RMS(x)}{RMS(\lambda \cdot n)}) \tag{12}$$

$$\lambda = \frac{RMS(x)}{RMS(n) \cdot (\frac{SNR}{10})^{10}} \tag{13}$$

In the context of deep learning applied to EEG signal denoising, the denoising process can be conceptualized as a nonlinear mapping function. This function, denoted as $\hat{x} = F(y;\theta)$, maps the EEG signal y with noise to the corresponding noise-free signal $\hat{x}$. Here, $F(\cdot)$ represents the nonlinear mapping function, our neural network model, and θ is the model's learnable parameter. To facilitate better parameter learning, we employ the mean square error (MSE) as the loss function. The MSE is defined by calculating the squared difference between the predicted value $\hat{x}_i$ and the true value x_i for each sample point i of the signal, summing these differences, and dividing by the number of samples n. Mathematically, this is expressed as Equation (14).

$$\mathbb{L}(x, \hat{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{14}$$

4.2. Experiments Detail

4.2.1. Datasets

To assess the denoising efficacy of the proposed EEGDiR model, we utilized the EEGDenoiseNet dataset [5], a widely adopted dataset in deep learning for EEG signal denoising, for both training and testing. The dataset encompasses various signal categories, including 4515 pristine EEG signals, 3400 ocular artifacts, and 5598 muscle artifacts. Each sample has a sampling time of 2 seconds at a rate of 256 samples per second. Pure EEG signals are denoted as x in Equation (11), while ocular artifacts or muscle artifacts are denoted as n in Equation (11).

For signals contaminated with ocular artifacts, 3400 samples were randomly chosen from pure EEG signals and all 3400 ocular artifact signals. Subsequently, the training and test sets were constructed in an 8:2 ratio, respectively. At specified signal-to-noise ratio (SNR) levels (-7 dB to 2 dB), pure EEG signals were linearly combined with ocular artifacts to generate ocular artifact-contaminated signals y. Notably, the parameter λ for superimposing eye movement artifacts in Equation (13) was directly calculated based on the given SNR value to obtain y. This dataset is referred to as the EOG dataset.

The signals contaminated with muscle artifacts were derived from pure EEG signals, and all 4515 samples were utilized along with the 5598 samples from the EOG artifact signals. To maintain consistency in the number of samples from pure EEG signals and EMG artifact signals, some samples from the pure EEG signal were reused. The resulting dataset was partitioned into training and test sets in an 8:2 ratio. Similarly, using specified SNR levels, pure EEG signals were randomly combined with EMG artifacts to generate EMG artifact-contaminated signals y. This dataset is denoted as the EMG dataset.

In order to facilitate the learning procedure, we normalized the input contaminated EEG segment and the ground-truth EEG segment by dividing the standard deviation of contaminated EEG segment according to Equation (15), where σ_y is the standard deviation of y (artifact contaminated signal).

$$\hat{x} = \frac{x}{\sigma_y}, \hat{y} = \frac{y}{\sigma_y} \tag{15}$$

However, it is important to note that the EEGDenoiseNet dataset solely provides raw data, necessitating researchers to conduct their own processing. This procedure is relatively intricate, posing inconvenience for the exploration of EEG signal denoising through deep learning methodologies. To address this challenge, this paper undertakes the preprocessing of the dataset and subsequently shares the processed dataset as an open-source resource, now accessible on the Hugging Face Hub. The primary objective of this endeavor is to facilitate researchers' access to and utilization of processed data, enabling them to concentrate more on the investigation of EEG deep learning denoising methods without being encumbered by the intricacies of data processing. By providing this dataset as an open-source entity, our aim is to stimulate increased research in EEG signal processing and offer a more streamlined resource for the academic community.

4.2.2. Train details

In this investigation, we opted to implement the EEGDiR model utilizing the PyTorch deep learning framework, renowned for its widespread usage and adaptability. To enhance the training efficiency of the network, we employed the AdamW [28] optimizer, a proficient choice for managing large-scale deep learning models. The learning rate was set to $5e-4$, and the betas parameter ranged from $(0.5, 0.9)$, with meticulous adjustment to optimize the network training process.

Throughout the training phase, the network underwent 5000 epochs to ensure comprehensive learning of dataset features. Furthermore, we configured the batch size to 1000, a standard choice that balances memory utilization and training efficacy. Notably, for accelerated training, we harnessed the computational capabilities of an NVIDIA GeForce RTX 3090 Graphics Processing Unit (GPU). Leveraging the parallel computing power of GPUs significantly augmented the speed of deep learning model training, facilitating rapid experimentation and tuning for researchers.

4.3. Results

4.3.1. Comparing method

We conducted comparative experiments to assess the efficacy of our proposed EEGDiR model against established state-of-the-art deep learning EEG denoising networks, encompassing the following models:

(1) Simple Convolutional Neural Networks (SCNN) [27]:

- Network Structure: Four 1D convolutional layers with 1×3 convolutional kernels, a 1-step size, and 64 channels.
- Interlayer Structure: Batch Normalization and ReLU activation functions follow each convolutional layer.

- Output: Features linearly projected through fully connected layers to match input dimensions.

(2) One-dimensional Residual Convolutional Neural Networks (1D-ResCNN) [18]:

- Network Structure: Utilizes three distinct convolutional kernels (1×3, 1×5, 1×7) ResBlocks for parallel feature extraction.
- ResBlock Structure: Each ResBlock comprises four 1D convolutional layers, with every two forming a residual block.
- Interlayer Structure: Activation through Batch Normalization and ReLU functions for each residual block.
- Output: Concatenation of the three ResBlocks' outputs, linearly projected through fully connected layers to maintain input dimensions.

(3) EEG Denoise Network (EEGDnet) [16]:

- Network Structure: Incorporates a Transformer infrastructure with four Transformer layers, featuring an Attention module and a FeedForward Network (FFN) module in each layer.
- Module Structure: Layer Normalization applied to the input of each module.
- Output: Linear projection to maintain input dimensions.

These benchmark networks represent diverse EEG denoising methodologies, serving as benchmarks to validate the superiority of our proposed EEGDiR model in denoising performance. These comparisons aim to offer readers a comprehensive understanding of the EEGDiR model's performance.

4.3.2. Evaluation measures

We assess the denoising outcomes using three methods: Relative Root Mean Squared Error in the temporal domain ($RRMSE_{temporal}$), RRMSE in the spectral domain ($RRMSE_{spectral}$), and the correlation coefficient (CC) [27]. This selection is grounded in a profound understanding of EEG signal characteristics. Firstly, considering the temporal significance of the EEG signal, we employ $RRMSE_{temporal}$ to quantify the relative error between the denoised and original signals in the time domain. This method exhibits sensitivity to denoising techniques preserving temporal information. Secondly, as EEG signals encapsulate rich spectral information with research often focusing on specific frequency ranges, $RRMSE_{spectral}$ is employed to ensure the preservation of features in the frequency domain. Lastly, acknowledging the synergistic activities between different brain regions in EEG signals, CC serves as an evaluation metric. CC reflects the linear relationship between the denoised and original signals, crucial for maintaining vital information about interregional correlation. The combined use of these methods facilitates a comprehensive evaluation of denoising performance across time and frequency domains, considering their adaptability to EEG signal characteristics. The mathematical expressions for $RRMSE_{temporal}$, $RRMSE_{spectral}$, and CC are represented in Equation (16), (17), and (18) respectively, where $RMS(\cdot)$ denotes root mean square, $PSD(\cdot)$ denotes power spectral density, and $Cov(\cdot)$, $Var(\cdot)$ represent covariance and variance, respectively.

$$RRMSE_{temporal} = \frac{RMS(F(y) - x)}{RMS(x)} = \frac{RMS(\hat{x} - x)}{RMS(x)} \tag{16}$$

$$RRMSE_{spectral} = \frac{RMS(PSD(F(y)) - PSD(x))}{RMS(PSD(x))} = \frac{Cov(\hat{x}, x)}{Var(\hat{x})Var(x)} \tag{17}$$

$$CC = \frac{Cov(F(y), x)}{\sqrt{Var(F(y))Var(x)}} = \frac{Cov(\hat{x}, x)}{\sqrt{Var(\hat{x})Var(x)}} \tag{18}$$

4.3.3. Ablation study

In this study, we pioneer the application of the Retnet network to 1D signal denoising and introduce the innovative concept of signal embedding. To assess the impact of each hyperparameter on denoising performance, we conduct ablation experiments, marking the inaugural exploration of these techniques. Our initial focus is on the influence of patch size and hidden dimension, investigating their effects on network performance. Table 1 presents quantized results for the network applied to EOG and EMG under various configurations of patch size and hidden dimension hyperparameters. Notably, we observe an enhancement in denoising performance with decreasing patch size while maintaining a consistent hidden dimension. This improvement is attributed to the impact of patched sequence length on the network's feature extraction capability—smaller patch sizes result in larger mini sequence lengths, preserving more information and thereby improving denoising effectiveness. However, a cautious approach is essential as blindly reducing patch size escalates computational complexity due to increased sequence length. Thus, a delicate balance between denoising performance and computational efficiency is imperative. Furthermore, maintaining the same patch size, an increased hidden dimension corresponds to improved denoising performance, aligning with the intuitive understanding that higher dimensionality facilitates enhanced feature extraction.

Subsequently, we assess the impact of varying the number of block layers L on network performance. Table 2 displays the denoising quantization results for the model applied to EOG and EMG datasets, with fixed parameters patch size 16, hidden dimension 512, and heads 8. The observations indicate a gradual enhancement in denoising performance with an increasing number of layers. This improvement is ascribed to the benefits of residual connections,

Table 1
The effect of patch size and hidden dim on the noise reduction performance of EEGDiR. Note mini sequence length must be $|s|//patchsize$, where $|s| = 512$, with 8 Heads and 4 DiRBlock layers.

Patch size	Mini sequence length	Hidden dim	Ocular artifact			Muscle artifact		
			$RRMSE_{temporal}$	$RRMSE_{spectral}$	CC	$RRMSE_{temporal}$	$RRMSE_{spectral}$	CC
32	16	512	0.339	0.367	0.928	0.556	0.561	0.793
32	16	256	0.353	0.377	0.911	0.569	0.575	0.791
32	16	64	0.382	0.371	0.909	0.572	0.577	0.790
16	32	512	0.327	0.361	0.932	0.532	0.501	0.807
16	32	256	0.348	0.371	0.925	0.598	0.573	0.776
16	32	64	0.374	0.378	0.912	0.654	0.593	0.701

Table 2
The effect of the layers of EEGDiR on the noise reduction performance. Note that patch size ,hidden dim and N Heads equal to 16, 512 and 8 respectively.

Layers	Ocular artifact			Muscle Artifact		
	$RRMSE_{temporal}$	$RRMSE_{spectral}$	CC	$RRMSE_{temporal}$	$RRMSE_{spectral}$	CC
4	0.327	0.361	0.932	0.532	0.501	0.807
3	0.356	0.380	0.925	0.578	0.583	0.789
2	0.372	0.383	0.917	0.591	0.596	0.781
1	0.394	0.429	0.908	0.613	0.594	0.766

whereby a higher number of layers does not lead to overfitting. The increased network depth contributes to superior feature extraction capabilities.

Following the ablation study, optimal performance is achieved when employing a patch size of 16, hidden dimension of 512, 8 heads, and 4 layers. Consequently, this configuration is chosen as the benchmark for subsequent comparisons with other networks.

4.3.4. Denoising effect of each method at all noise levels

Table 3 illustrates the denoising efficacy of various methods on EOG and EMG datasets. The outcomes in this table lead to the following: insights:

(1) Due to its relatively simplistic structure comprising only four convolutional layers and lacking residual connections, SCNN exhibits suboptimal denoising effects, potentially prone to overfitting.
(2) Featuring a more intricate architecture incorporating diverse convolutional kernels for multi-scale feature extraction and alleviating overfitting through the introduction of residual connections, 1D-ResCNN surpasses SCNN, significantly enhancing denoising outcomes.
(3) Leveraging the transformer architecture, EEGDnet excels in denoising, benefitting from the global modeling prowess of the attention mechanism, complemented by residual connections and layer normalization. This results in substantial denoising improvements compared to SCNN and 1D-ResCNN.
(4) Capitalizing on the Retnet framework, EEGDiR achieves superior denoising performance by comprehensively understanding input temporal information and exhibiting robust global modeling capabilities. The incorporation of residuals and multiple normalizations (layer norm, group norm) further distinguishes EEGDiR, outperforming other networks. Moreover, guided by our proposed signal embedding, EEGDiR intelligently processes temporal information, forming patches from consecutive samples. This strategy adeptly captures the contextual and temporal relationships within EEG signals, aligning with their prolonged temporal characteristics. The judicious embedding strategy contributes to optimized denoising performance, reinforcing EEGDiR's exceptional superiority over alternative networks.

4.3.5. Denoising effect of each method at different noise levels

In the subsequent section, we present the quantitative benchmarking results ($RRMSE_{temporal}$, $RRMSE_{spectral}$, CC) of diverse methods across varying SNR levels in the test set. Figures 3 and 4 showcase the test outcomes on the EOG and EMG test sets, pivotal for evaluating the denoising efficacy of the methods.

(1) Primarily, the performance of all methods exhibits a decline as the SNR level decreases. This negative correlation arises due to the gradual increase in noise level, posing a greater challenge for the methods in noise removal.
(2) Among the methods, SCNN displays the highest $RRMSE_{temporal}$ and $RRMSE_{spectral}$, along with the lowest CC. This indicates SCNN's inferior denoising performance, attributed to its relatively simple network structure hindering effective input feature extraction. In contrast, the more intricate 1D-ResCNN yields significantly improved denoising outcomes. However, compared to EEGDnet with a Transformer model and global modeling capability, there are discernible performance gaps. The EEGDiR model, incorporating Retnet, achieves the

Table 3
Average performances of all SNRs (from −7 dB to 2 dB). The smaller $RRMSE_{temporal}$ and $RRMSE_{spectral}$, and the larger CC, the better denoising effect. Note that all the models are trained and tested on the same data set. The baseline of EEGDiR consists of 4 layers and 8 Heads with patch size 16 and hidden dim 512. For $RRMSE_{temporal}$, $RRMSE_{spectral}$, the lower the better. For CC, the higher the better. The best result is shown in bold.

Model	Ocular artifact			Muscle artifact		
	$RRMSE_{temporal}$	$RRMSE_{spectral}$	CC	$RRMSE_{temporal}$	$RRMSE_{spectral}$	CC
SCNN	0.6176	0.5905	0.7938	0.7342	0.7977	0.7364
1D-ResCNN	0.5409	0.5900	0.8503	0.6921	0.6848	0.7434
EEFDnet	0.4819	0.4647	0.8725	0.6200	0.5565	0.7711
EEGDiR(ours)	**0.3279**	**0.3616**	**0.9329**	**0.5322**	**0.5004**	**0.8072**

lowest $RRMSE_{temporal}$ and $RRMSE_{spectral}$, coupled with the highest CC. It excels in denoising tasks across varying noise levels.

(3) Analyzing the $RRMSE_{temporal}$ results on the EOG dataset, denoising performance improves with decreasing noise levels and increasing SNR levels across all methods. However, the performance gap between methods persists, potentially due to EOG noise being more easily removed than EMG noise. On the EMG dataset, the performance gap diminishes as noise levels decrease (SNR levels increase), particularly evident for SCNN, 1D-ResCNN, and EEGDnet. Nevertheless, EEGDiR maintains superior denoising performance.

(4) Evaluation of the $RRMSE_{spectral}$ results on the EOG dataset indicates weaker denoising performance for SCNN and 1D-ResCNN, possibly due to limited global modeling capability. Conversely, EEGDnet and EEGDiR exhibit superior denoising performance owing to their robust global modeling ability. On the EMG dataset, despite decreasing differences in performance as noise levels decrease, EEGDnet and EEGDiR consistently outperform SCNN and 1D-ResCNN.

(5) Examination of CC results on the EOG dataset reveals improved denoising performance for all methods as noise levels decrease, with relatively stable performance differences. In the EMG dataset, SCNN exhibits poorer performance due to the dataset's more complex noise. Conversely, the denoising performance of the remaining three networks improves as noise levels decrease, with consistent performance differences. Notably, EEGDiR maintains excellent denoising performance throughout.

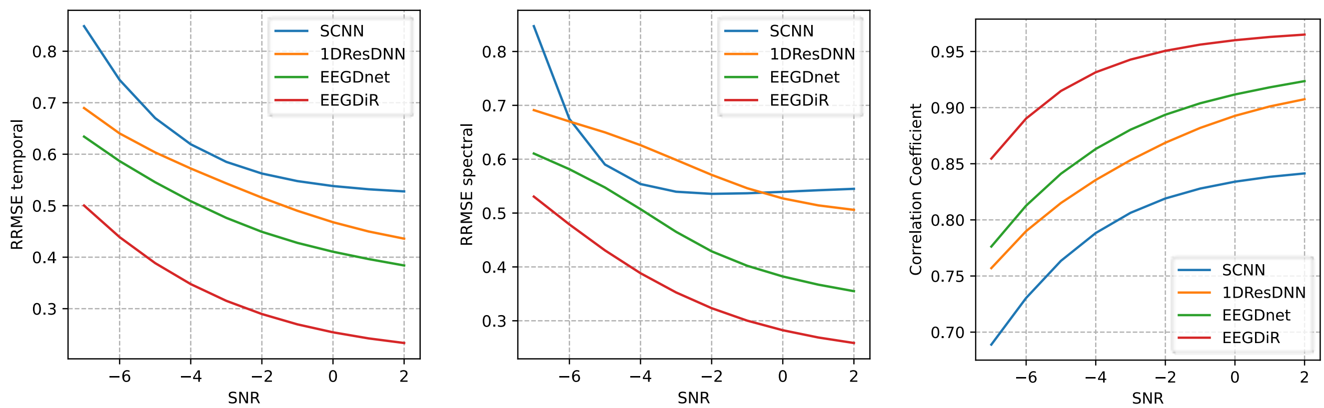
Figures 5 and 6 illustrate the ANOVA results for models evaluated on the EOG and EMG datasets. Drawing conclusions from the provided information and ANOVA analyses, the following observations emerge:

(1) $RRMSE_{temporal}$ Rancking: The denoising performance across the four methods is observed as follows: CNN < 1D-ResCNN < EEGDnet < EEGDiR. ANOVA analysis indicates significant differences in $RRMSE_{temporal}$, with marked distinctions in the EOG dataset and relatively modest differences in the EMG dataset. EEGDiR significantly outperforms other methods in time-domain denoising for both EOG and EMG datasets, followed by EEGDnet and 1D-ResCNN, while SCNN exhibits the least efficacy.

(2) $RRMSE_{spectral}$ Rancking: The denoising performance order for $RRMSE_{spectral}$ is 1D-ResCNN < SCNN < EEGDnet < EEGDiR. The occurrence of 1D-ResCNN < SCNN is attributed to 1D-ResCNN's superior extraction of features in the time domain, leading to diminished denoising performance in the spectral features. ANOVA results show a significant difference in $RRMSE_{spectral}$ among methods on the EOG dataset, while the difference is relatively weak on the EMG dataset. EEGDiR significantly outperforms other methods in spectral denoising, followed by EEGDnet and SCNN, while 1D-ResCNN is less effective.

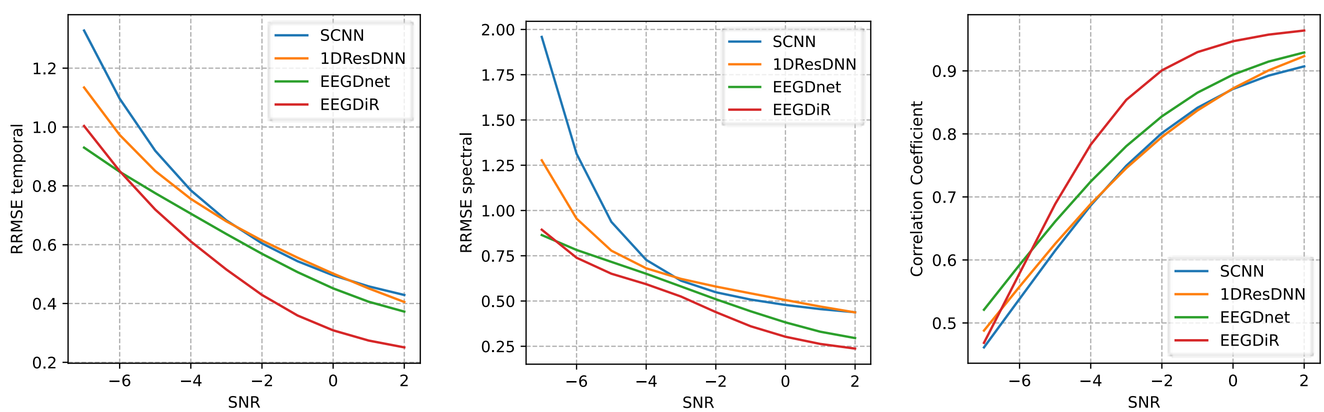
(3) CC Metric Ranking: The denoising performance sequence on the CC metric remains SCNN < 1D-ResCNN < EEGDnet < EEGDiR. ANOVA analysis reveals a significant difference in CC between methods for the EOG dataset, while the difference is relatively weak for the EMG dataset. Comparing mean values, EEGDiR excels in correlation, followed by EEGDnet and 1D-ResCNN, while SCNN exhibits poor CC performance.

In summary, the outstanding denoising performance of the EEGDiR method can be attributed to multiple factors. The Retnet architecture provides enhanced global modeling capability, enabling a more accurate restoration of input timing information. The proposed signal embedding method adeptly handles the prolonged temporal information of EEG signals, capturing context and temporal relationships intelligently through the combination of successive sampling points into patches. This advantage enables EEGDiR to achieve superior denoising effects in both the time domain and spectral characteristics. Additionally, the synergy of residual connectivity and multiple normalization methods (layer norm, group norm) enhances EEGDiR's denoising performance and robustness to noise. The advanced Retnet architecture, skillful embedding strategy, and enhanced network design collectively contribute to EEGDiR's exceptional performance in time-domain and spectral denoising, as well as correlation.



(a) $RRMSE_{temporal}$ (b) $RRMSE_{spectral}$ (c) CC

Figure 3: Performance of four deep-learning networks at different SNR levels with ocular artifact removal. The smaller $RRMSE_{temporal}$ and $RRMSE_{spectral}$, and the larger Correlation Coefficient(CC), the better denoising effect. The denoising performance increases as the SNR increases.



(a) $RRMSE_{temporal}$ (b) $RRMSE_{spectral}$ (c) CC

Figure 4: Performance of four deep-learning networks at different SNR levels with muscle artifact removal. The smaller $RRMSE_{temporal}$ and $RRMSE_{spectral}$, and the larger Correlation Coefficient(CC), the better denoising effect. The denoising performance increases as the SNR increases.

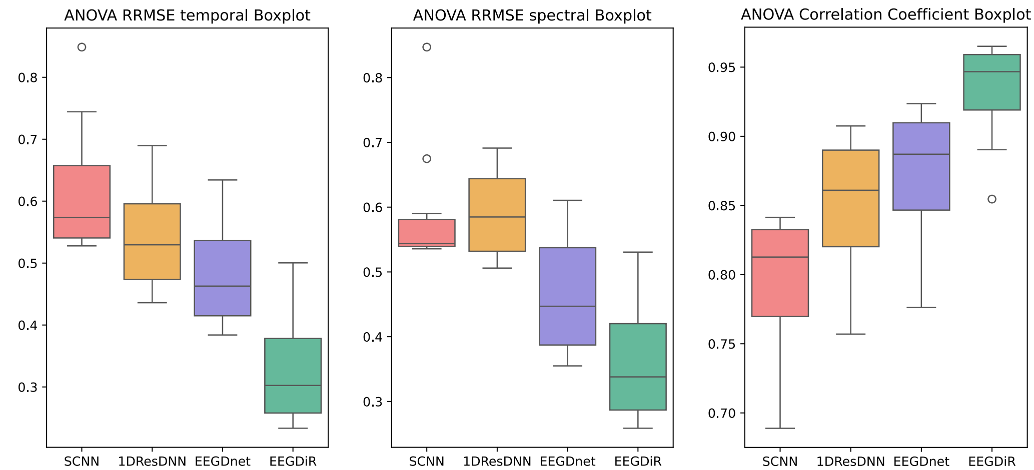
4.3.6. Visualization of denoising results for each method on EOG and EMG datasets

The visualization results depicting the impact of EOG and EMG noise on EEG signals are presented in Figure 7a and Figure 7b, yielding the following observations. It is noteworthy that the dataset has undergone variance normalization. When presenting the visualization results, Equation (15) is employed to scale down the results to the original data scale, enhancing the accuracy of showcasing the noise effect on EEG signals.

(1) All methods exhibit some degree of noise suppression in noisy signals, underscoring the variability among different denoising approaches. Particularly notable is the substantial difference between the denoising results of the SCNN method and the noisy signal. This divergence may be attributed to the relatively simplistic network structure of SCNN, hindering comprehensive feature extraction.

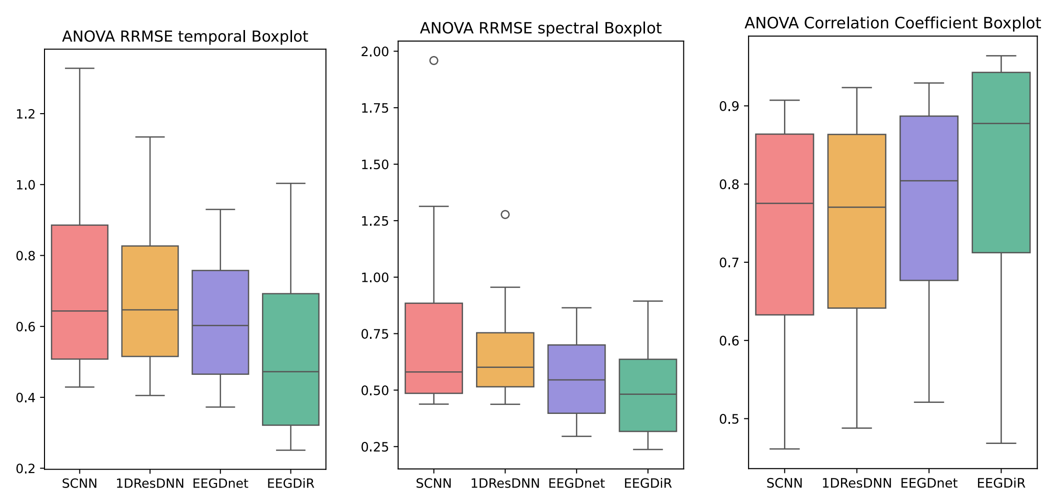
(2) The relatively complex structure and residual connectivity of 1D-ResCNN result in an improvement in denoising compared to SCNN, emphasizing the impact of network architecture on denoising performance. Leveraging the global modeling and feature extraction capabilities facilitated by the Transformer's attention mechanism, EEGDnet outperforms SCNN and 1D-ResCNN in denoising. This underscores the enhancement of network performance with the introduction of the attention mechanism.

(3) Overall, the denoising effect achieved by EEGDiR closely approaches that of a noise-free signal. This notable advantage can be attributed to the synergistic effect of the Retnet architecture and our proposed



(a) $RRMSE_{temporal}$ (b) $RRMSE_{spectral}$ (c) CC

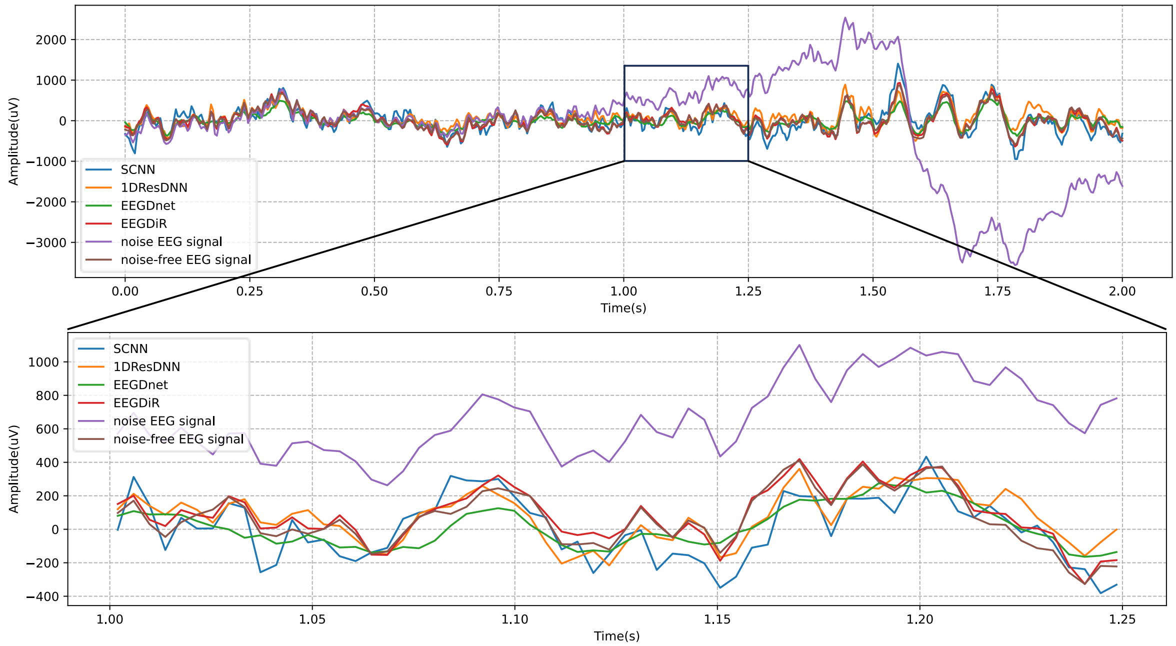
Figure 5: Performance of four DL networks (SCNN, 1D-ResDNN, EEGDnet, EEGDiR) in ocular artifact removal. The smaller $RRMSE_{temporal}$ and $RRMSE_{spectral}$, and the larger Correlation Coefficient(CC), the better denoising effect. EEGDiR models robustly outperform other model for EEG denosing.



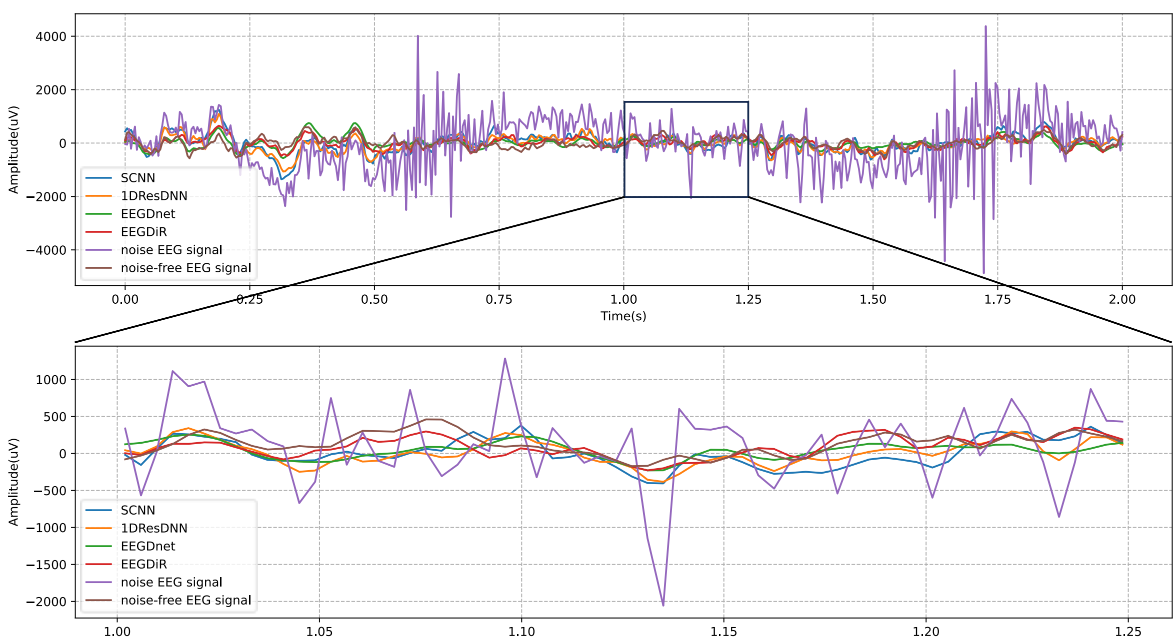
(a) $RRMSE_{temporal}$ (b) $RRMSE_{spectral}$ (c) CC

Figure 6: Performance of four DL networks (SCNN, 1D-ResDNN, EEGDnet, EEGDiR) in muscle artifact removal. The smaller $RRMSE_{temporal}$ and $RRMSE_{spectral}$, and the larger Correlation Coefficient(CC), the better denoising effect. EEGDiR models robustly outperform other model for EEG denosing.

signal embedding, tailored to match the characteristics of EEG signals. Firstly, the Retnet architecture enhances understanding of timing information in EEG signals through its robust global modeling capability. This enables the network to accurately capture the complex time-domain structure, thereby improving denoising performance. Secondly, our signal embedding method adeptly addresses the challenge of handling long temporal information in EEG signals. By intelligently grouping consecutive sampling points into patches, this method effectively preserves the context and temporal relationships of the signal, facilitating the network in learning and restoring features more efficiently. The sensitivity to long temporal information aligns with the characteristics of EEG signals, forming the basis for EEGDiR's outstanding denoising effect. Therefore, the performance of EEGDiR in approaching noise-free signals arises not only from the superior processing of temporal information by the Retention mechanism but also from the mutually reinforcing capabilities of Retnet and signal embedding. This synergy enables the network to better comprehend and process the intricate structure of EEG signals.



(a) Denoising outcomes on Ocular Artifact.



(b) Denoising outcomes on Muscle Artifact.

Figure 7: Visualization of denoising outcomes for various state-of-the-art models: (a) Denoising outcomes on Ocular Artifact. (b) Denoising outcomes on Muscle Artifact. A closer inspection can be facilitated by zooming in for a more detailed view. It is important to highlight that we have restored the network output's amplitude back to the original data scale through back-normalization. The temporal domain operates at a sampling rate of 256 SPS. From the provided results, it is evident that the denoising results achieved by the proposed EEGDiR model in this study closely approximate the true signal.

5. Discussions

In contrast to prevailing deep learning denoising methodologies, our approach involves incorporating the Retnet architecture into Long Short-Term Memory (LLM) for EEG signal denoising. The Retnet architecture exhibits notable strengths in comprehending temporal information and excelling in global modeling feature extraction. The decision to integrate Retnet stems from its proficiency in handling temporal information and its exceptional global modeling performance. Concurrently, we introduce signal embedding to intelligently manage the protracted time-series information inherent in EEG signals. This approach enhances context and temporal relationship preservation by aggregating consecutive sampling points into patches. Given the prolonged time-series nature of EEG signals, this method aligns well with their characteristics, contributing to enhanced denoising performance. Hence, the incorporation of Retnet and patch embedding serves not only to refine the comprehension and processing of EEG signals' temporal structure but also to augment denoising efficacy from diverse perspectives. This dual strategy of leveraging Retnet and patch embedding renders our method adaptable to the intricate characteristics of EEG signals, resulting in superior denoising outcomes.

Retnet's adept handling of temporal information in processing EEG signals extends its applicability to various temporal signals, including electromagnetic, seismic, and biomedical signals. Its affinity for temporal information bestows a distinct advantage when addressing signals necessitating comprehensive time-based modeling. This advantageous trait is particularly pertinent to other temporal signals like electromagnetic, seismic, and biomedical signals. The friendly temporal information processing of Retnet proves beneficial in scenarios requiring a global temporal perspective. The incorporation of signal embedding in EEG signals enhances the preservation of contextual and temporal relationships within the signals. This characteristic, valuable for signal processing across various domains, especially in cases with extensive temporal information or a requirement for comprehensive temporal modeling, extends beyond EEG signals. The network proposed in this paper achieves notable performance enhancements by amalgamating Retnet's global modeling and signal embedding's temporal information processing strategies. The broad applicability of this strategy suggests its potential effectiveness in diverse signal processing domains, particularly tasks demanding a synthesis of global and temporal information. The integrated strategy of Retnet and signal embedding may evolve into a generalized signal processing approach, particularly suitable for scenarios requiring consideration of both global features and temporal information. Subsequent research endeavors could explore the application of this approach to other signal types, validating its adaptability and performance across distinct domains.

6. Conclusion

By incorporating the Retnet architecture and employing signal embedding for processing 1D EEG signals, this study introduces an innovative methodology aiming to leverage Retnet comprehensively for EEG signal denoising. The integration of Retnet architecture enhances the understanding and processing of temporal information in EEG signals, while the utilization of signal embedding underscores the processing of prolonged temporal information and feature extraction. Experimental results showcase the outstanding denoising performance of our proposed EEGDiR network on EOG and EMG datasets. In comparison to traditional EEG denoising methods, EEGDiR demonstrates notable enhancements in temporal information processing and global modeling.

The global modeling prowess of Retnet, coupled with its favorable handling of temporal information, positions it as an optimal choice for processing EEG signals. The incorporation of signal embedding further refines the representation of EEG signals, preserving context and temporal relationships more effectively. The synergistic application of the Retnet and signal embedding strategy yields a substantial improvement in the denoising performance of the EEGDiR network.

This study holds significant implications as a guide for the integration of deep learning in neuroscience, offering valuable insights to enhance the efficacy and application potential of EEG signal processing. By amalgamating Retnet's global modeling and temporal information processing advantages with signal embedding's intelligent handling of prolonged temporal information, our approach introduces novel ideas and paradigms to advance EEG signal processing research.

CRediT authorship contribution statement

Bin Wang: Conceptualization of this study, Methodology, Software.

References

[1] Albawi, S., Mohammed, T.A., Al-Zawi, S., 2017. Understanding of a convolutional neural network, in: 2017 international conference on engineering and technology (ICET), Ieee. pp. 1–6.
[2] Ba, J.L., Kiros, J.R., Hinton, G.E., 2016. Layer normalization. arXiv preprint arXiv:1607.06450 .
[3] Bebis, G., Georgiopoulos, M., 1994. Feed-forward neural networks. Ieee Potentials 13, 27–31.
[4] Burger, C., Van Den Heever, D.J., 2015. Removal of eog artefacts by combining wavelet neural network and independent component analysis. Biomedical Signal Processing and Control 15, 67–79.
[5] Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artifact. Electroencephalography and clinical neurophysiology 55, 468–484.
[6] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
[7] He, P., Wilson, G., Russell, C., 2004. Removal of ocular artifacts from electro-encephalogram by adaptive filtering. Medical and biological engineering and computing 42, 407–412.
[8] Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International conference on machine learning, pmlr. pp. 448–456.
[9] Jahan, I., Ahmed, M.F., Ali, M.O., Jang, Y.M., 2023. Self-gated rectified linear unit for performance improvement of deep neural networks. ICT Express 9, 320–325.
[10] Jiang, X., Bian, G.B., Tian, Z., 2019. Removal of artifacts from eeg signals: a review. Sensors 19, 987.
[11] Klados, M.A., Papadelis, C., Braun, C., Bamidis, P.D., 2011. Reg-ica: a hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts. Biomedical Signal Processing and Control 6, 291–300.
[12] Lhoest, Q., del Moral, A.V., Jernite, Y., Thakur, A., von Platen, P., Patil, S., Chaumond, J., Drame, M., Plu, J., Tunstall, L., et al., 2021. Datasets: A community library for natural language processing. arXiv preprint arXiv:2109.02846 .

[13] McMenamin, B.W., Shackman, A.J., Maxwell, J.S., Bachhuber, D.R., Koppenhaver, A.M., Greischar, L.L., Davidson, R.J., 2010. Validation of ica-based myogenic artifact correction for scalp and source-localized eeg. Neuroimage 49, 2416–2432.

[14] Memory, L.S.T., 2010. Long short-term memory. Neural computation 9, 1735–1780.

[15] Molla, M.K.I., Islam, M.R., Tanaka, T., Rutkowski, T.M., 2012. Artifact suppression from eeg signals using data adaptive time domain filtering. Neurocomputing 97, 297–308.

[16] Pu, X., Yi, P., Chen, K., Ma, Z., Zhao, D., Ren, Y., 2022. Eegdnet: Fusing non-local and local self-similarity for eeg signal denoising with transformer. Computers in Biology and Medicine 151, 106248.

[17] Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

[18] Sun, W., Su, Y., Wu, X., Wu, X., 2020. A novel end-to-end 1d-rescnn model to remove artifact from eeg signals. Neurocomputing 404, 108–121.

[19] Sun, Y., Dong, L., Huang, S., Ma, S., Xia, Y., Xue, J., Wang, J., Wei, F., 2023. Retentive network: A successor to transformer for large language models. arXiv preprint arXiv:2307.08621 .

[20] Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to sequence learning with neural networks. Advances in neural information processing systems 27.

[21] Turnip, A., Junaidi, E., 2014. Removal artifacts from eeg signal using independent component analysis and principal component analysis, in: 2014 2nd International Conference on Technology, Informatics, Management, Engineering & Environment, IEEE. pp. 296–302.

[22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

[23] Weidong, Z., Yingyuan, L., 2001. Eeg multiresolution analysis using wavelet transform, in: 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE. pp. 1854–1856.

[24] Wu, Y., He, K., 2018. Group normalization, in: Proceedings of the European conference on computer vision (ECCV), pp. 3–19.

[25] Yang, B., Duan, K., Fan, C., Hu, C., Wang, J., 2018. Automatic ocular artifacts removal in eeg using deep learning. Biomedical Signal Processing and Control 43, 148–158.

[26] Zaremba, W., Sutskever, I., Vinyals, O., 2014. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329 .

[27] Zhang, H., Zhao, M., Wei, C., Mantini, D., Li, Z., Liu, Q., 2020. Eegdenoisenet: A benchmark dataset for end-to-end deep learning solutions of eeg denoising. arXiv preprint arXiv:2009.11662 .

[28] Zhuang, Z., Liu, M., Cutkosky, A., Orabona, F., 2022. Understanding adamw through proximal methods and scale-freeness. arXiv preprint arXiv:2202.00089 .