NEUROSCIENCE
RESEARCH ARTICLE
W. Wang et al. / Neuroscience 505 (2022) 10–20

A Novel End-to-end Network Based on a bidirectional GRU and a Self-Attention Mechanism for Denoising of Electroencephalography Signals

Wenlong Wang, Baojiang Li * and Haiyan Wang

The School of Electrical Engineering, Shanghai DianJi University, Shanghai, China

Intelligent Decision and Control Technology Institute, Shanghai DianJi University, Shanghai, China

Abstract—Electroencephalography (EEG) signals are nonlinear and non-stationary sequences that carry much information. However, physiological signals from other body regions may readily interfere with EEG signal capture, having a significant unfavorable influence on subsequent analysis. Therefore, signal denoising is a crucial step in EEG signal processing. This paper proposes a bidirectional gated recurrent unit (GRU) network based on a self-attention mechanism (BG-Attention) for extracting pure EEG signals from noise-contaminated EEG signals. The bidirectional GRU network can simultaneously capture past and future information while processing continuous time sequence. And by paying different levels of attention to the content of varying importance, the model can learn more significant feature of EEG signal sequences, highlighting the contribution of essential samples to denoising. The proposed model is evaluated on the EEGdenoiseNet data set. We compared the proposed model with a fully connected network (FCNN), the one-dimensional residual convolutional neural network (1D-ResCNN), and a recurrent neural network (RNN). The experimental results show that the proposed model can reconstruct a clear EEG waveform with a decent signal-to-noise ratio (SNR) and the relative root mean squared error (RRMSE) value. This study demonstrates the potential of BG-Attention in the pre-processing phase of EEG experiments, which has significant implications for medical technology and brain-computer interface (BCI) applications.

Key words: Electroencephalography (EEG), EEG signal denoising, Self-attention mechanism, Bidirectional gated recurrent unit (GRU), Brain-computer interface (BCI).

INTRODUCTION

Electroencephalography (EEG), functional near-infrared spectroscopy (fNIRS), magnetoencephalography (MEG), and other related techniques can record the brain's stimulus signals to objects and the surrounding environment (Van et al., 2019). We may gain rich psychological and pathological information as well as the health condition of the body system by evaluating the signals captured by these approaches (Muhammad et al., 2020). EEG is a non-invasive, affordable, high-time resolution, and portable neuroimaging method that indirectly detects neural activity by putting electrodes on the scalp to reflect the spontaneous bioelectrical potential in the brain (Fernández-Rodríguez et al., 2020). EEG science research is mainly focused on emotion categorization, the epilepsy analysis, brain-computer interface (BCI), and other significant areas (Lazzarotto et al., 2021; Maimaiti et al., 2021).

EEG has a high temporal resolution and is a highly unpredictable nonlinear nonstationary signal. Besides the brain activity, EEG signals also comprise other artifacts, such as electrooculogram (EOG) artifacts (Croft and Barry, 2000), electromyography (EMG) artifacts (Song et al., 2016; Mathe et al., 2022), and cardiac abnormalities in rare circumstances, resulting in a variety of distortions. These distortions and noise impeded the further investigation of EEG signals (Molla et al., 2012). As a result, it is critical to develop effective methods for removing artifacts from EEG signals on a theoretical and practical level.

EEG denoising methods have been researched for a long time. The primary focus of these methods is artifact removal, that is, identifying the artifact in the original signal and subsequently removing it. The regression-based technique first derives the noise signal using a noise template, then subtracts the estimated noise from

*Correspondence to: Baojiang Li, The School of Electrical Engineering, Shanghai DianJi University, Shanghai, China.
E-mail address: libj@sdju.edu.cn (B. Li).
Abbreviations: BCI, brain-computer interface; BG-Attention, a bidirectional GRU network based on a self-attention mechanism; EEG, Electroencephalography; FCNN, fully connected network; GRU, gated recurrent unit; RNN, recurrent neural network; RRMSE, relative root mean squared error; SNR, signal-to-noise ratio.

https://doi.org/10.1016/j.neuroscience.2022.10.006
0306-4522/

the EEG data to remove artifacts (McMenamin et al., 2009). Even though the regression-based technique has been improved several times, it still needs one or more excellent regression reference channels and is not entirely suited for denoising EMG artifacts. Methods based on adaptive filters filter out noise by dynamically calculating filter coefficients depending on the input EEG signal itself (He et al., 2004). This approach has high computational costs and requires extra sensors to give reference inputs (He et al., 2007). Blind source separation (BSS)-based approaches divide the EEG signal into many components, assigning them to neural and artifactual sources before recombining the neural components to recover a clean EEG sequence. (Cichocki et al., 2005). However, the BSS approach can only be employed when many electrodes are available, making it unsuitable for single-channel denoising (Chen et al., 2019a, 2019b). Independent component analysis (ICA)-based techniques disassemble EEG signals into fundamental signal components, from which experts may identify and eliminate artifact components (Delorme and Makeig, 2004). These methods were collectively regarded as traditional denoising methods, which has made significant progress in EEG denoising (Jiang et al., 2019; Xie et al., 2021). However, some of these approaches are designed to target certain artifacts, while others will harm critical information in EEG signals.

Over the last several years, deep learning (DL) has gotten much attention (Wang et al., 2020; Wu et al., 2022). EEG-based categorization (Craik et al., 2019), EEG reconstruction (Zhang et al., 2021), and EEG recognition (Miao et al., 2021; Wang and Yuan, 2021) are all examples of the DL approach being used in the area of EEG signal analysis. For EEG denoising, DL has lately been applied with superior results to the traditional denoising approaches (Egambaram et al., 2016; Chen et al., 2019a, 2019b). Sun et al. proposed the one-dimensional residual convolutional neural network (1D-ResCNN), using the trained 1D-ResCNN model as a filter to automatically remove artifacts from contaminated EEG signals (Sun et al., 2020). Pion-Tonachini et al. trained a CNN classifier to distinguish between artifacts and signal components (Pion-Tonachini et al., 2019). Zhang et al. used four neural networks, including a fully connected network (FCNN), a simple convolutional network, a complex convolutional network, and a recurrent neural network (RNN), to carry out EEG denoising (Zhang et al., 2021). These DL-based network designs offer an end-to-end model. Moreover, they are appropriate for multi-artifacts and low signal-to-noise ratio (SNR) situations, fully retaining the nonlinear information of the EEG signal and, to a large extent, restoring the benchmark signal. In many instances, it is too expensive to obtain high-quality training data, and consequently, studying the performance of DL models on limited data is beneficial.

EEG signals are often lengthy, one-dimensional, complicated, and nonlinear time sequence signals. The memory unit of RNN may impart the network with a particular memory and can better combine the sequence information to model the input data (Cho et al., 2014). However, gradient explosion or gradient disappearance may occur during training, causing the training to halt. Based on the traditional RNN, long short-term memory (LSTM) and gated recurrent unit (GRU) introduce the gate structure, which is preferable in overcoming this problem (Yu et al., 2019). GRU is a famous variant of LSTM (Chen et al., 2019a, 2019b), which synthesizes the forgetting gate and input gate to a single update gate and also mixes cell state and hidden state. So, the final GRU model is simpler and faster than the standard LSTM model (Dey and Salem, 2017). The bidirectional GRU can process both forward and backward information and connect its output to the same output layer to capture the bidirectional context information of the sequence (Lynn et al., 2019). It is more effective to denoise the EEG signal by simultaneously analyzing past and future data. However, as the duration of the sequence increases, this capacity to capture information dependencies diminishes as information is lost with each recurrence. In other words, the sequential model fails to express hierarchical details effectively. The self-attention mechanism is a variant of the attention mechanism. Self-attention mechanism decreases reliance on external information and is more effective at collecting internal data correlation; calculating the interaction between time sequences allows it to tackle the issue of long-term dependence (Vaswani et al., 2017). So, to understand the subtle relationships among the data while processing lengthy time sequences, we analyze EEG signals using the bidirectional GRU based on the self-attention mechanism.

This paper proposes a BG-Attention algorithm for denoising EEG signals based on the aforementioned ideas. The main contributions of this paper are as follows:

(1) This paper introduces a novel EEG signal denoising model based on the BG-Attention algorithm. It is the first time the RNN and the self-attention network have been used to denoise EEG signals.
(2) The network is an end-to-end structure. Without preprocessing or feature extraction, the incoming EEG data is processed directly.
(3) Strong self-learning capability significantly enables the network to recover EEG signals and preserves their nonlinearity. We put it to the test in the EEGdenoiseNet data set. The experimental results show that the proposed model can reconstruct a clear EEG waveform with a decent SNR and the relative root mean squared error (RRMSE) value.

EXPERIMENTAL PROCEDURES

This BG-Attention algorithm is shown in Fig. 1. It primarily consists of the bidirectional GRU layer, the self-attention layer, and the dense layer. The input of the network is the noise-contaminated EEG signal segment. Then the signal segment is processed through the encoded and decoded network framework, and finally, the network output is the reconstruction of the contaminated signal. The encoding and decoding layers have a similar structure; the decoding layer is depicted in the figure. The encoding layer consists of a bidirectional GRU layer

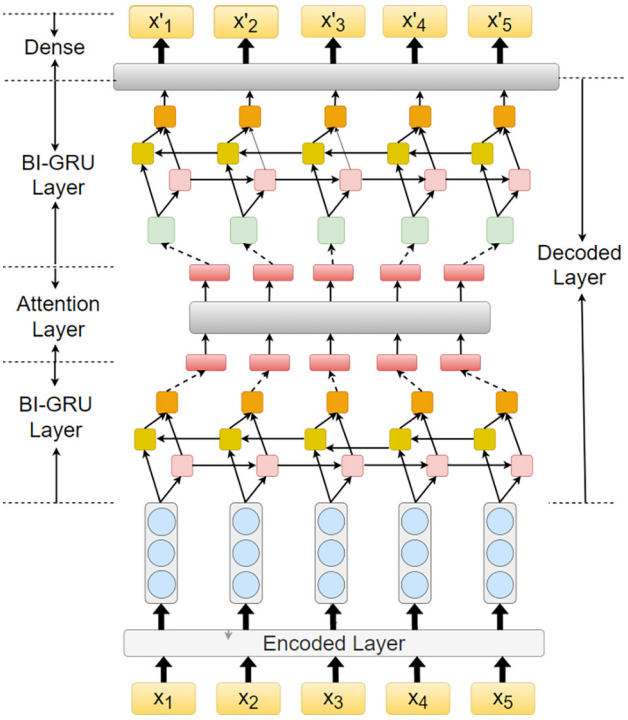


Fig. 1. The BG-Attention network structure.

and a self-attention mechanism layer. Details of this BG-Attention algorithm are described below.

Bidirectional GRU network

In addition to resolving the gradient explosion and disappearance issues that arise when RNN analyzes sequence data, the bidirectional GRU network can also collect simultaneously the past and future characteristic information of EEG signals, which is crucial for processing complicated EEG signals. Bidirectional GRUs are composed of forwarding and backward GRUs, so we first introduce the working process of GRU. The structure of GRU is shown in Fig. 2.

The new state in GRU is updated at time t as:

$$H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \widetilde{H}_t \tag{1}$$

where Z_t, H_{t-1} and $\widetilde{H}_t$ represent the update gate, hidden state and candidate hidden state, respectively. This is done to compute the linear interpolation between the hidden state H_{t-1} and the candidate hidden state $\widetilde{H}_t$ with the new sequence information. Update Gate Z_t determines how much past information to keep and how much new information to add. The larger the value of Z_t, the more information about the previous state is saved. Z_t is updated as:

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_r) \tag{2}$$

where X_t is the sample vector at time t, and the candidate states $\widetilde{H}_t$ are computed in the same way as the hidden layers of a conventional RNN network:

$$\widetilde{H}_t = tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h) \tag{3}$$

where R_t represents the reset gate, which controls the contribution of the previous state to the current candidate state. The smaller the R_t values, the smaller the contribution to the previous state. If $R_t = 0$, it will forget the previous state. The Reset gate R_t update formula is:

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \tag{4}$$

Bidirectional GRU encodes the input signal by utilizing the entire time sequence, including both positive and negative direction GRUs. Each bidirectional GRU structure consists of three distinct layers. The first layer of the GRU contains the current time series information, the second layer has the reverse reading of the same sequence, and the third layer is the dense layer. Bidirectional GRU networks extend the unidirectional GRU networks by introducing a second layer, where the hidden layers to hidden layers connections flow in opposite temporal order. This structure offers the output layer with the input sequence's entire past and future information. During training, the network's weights are modified by forward and backward propagation to the output neurons. The structure of bidirectional GRU unit is shown in Fig. 3. As can be seen from the picture, x_{t-1} and h_{t-1} indicate the state input and hidden layer output of previous moment, x and h represent the state input

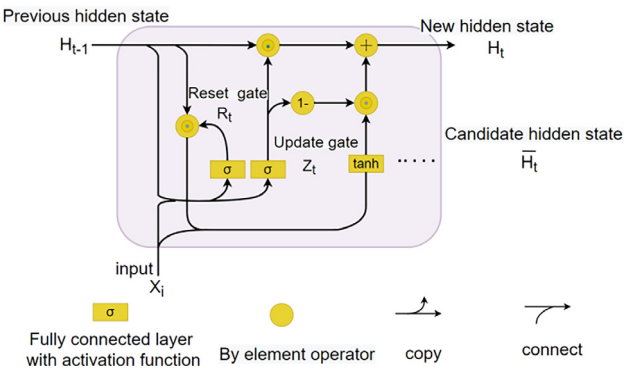


Fig. 2. Internal structure diagram of the GRU network.

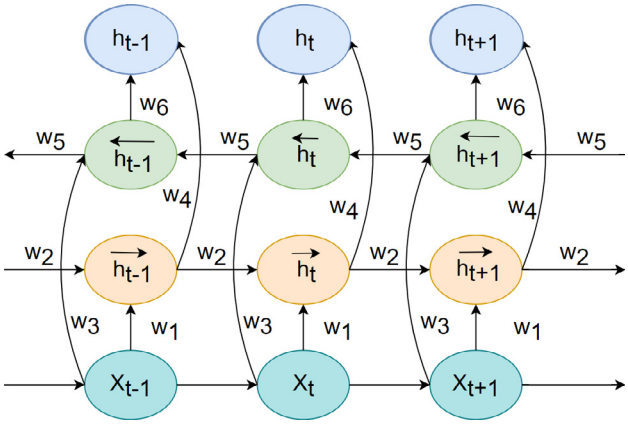


Fig. 3. Bidirectional GRU network structure.

and hidden layer output of current moment, and x_{t+1} and h_{t+1} represent the state input and hidden layer output of next moment.

Self-attention structure

Multiple attention layers encode the EEG signals behind the RNN so that they may be correlated with one another and extract global characteristics. In other words, we introduced the self-attention mechanism at the end of each bidirectional GRU layer to provide varying degrees of attention to the content of different relevance. The dot product, the cosine similarity, and the splicing similarity are used in the attention mechanism (Gao et al., 2021). The scaled dot product (Vaswani et al., 2017) is utilized to assess the similarity between two points in this paper. The attention structure is shown in Fig. 4, and the expression is as follows:

$$sim(Q,K) = \frac{Q * K^T}{\sqrt{d_k}} \tag{5}$$

where Q and K in the formula are query and key respectively. From the figure, we can see that both Q and K are the output vector of the previous process. d_k is the length of the K vector.

$$X\prime = \sum_{j=1}^{n} w_{ij} * softmax(sim(Q,K)) * V \tag{6}$$

where$X\prime$ is the attention mechanism's output vector, w_{ij} is the attention mechanism's weight coefficient, and V is the value. In this research, $V, Q, andK$ are considered to be identical.

The self-attention mechanism comprises two sub-layers: a multi-head attention mechanism (Yang et al., 2016) and a fully connected feed-forward network. Its structure is shown in Fig. 4. Transformer linearly transforms the matrices Q, K, and V into h components so that they may execute scaled dot product attention computation concurrently to facilitate parallel computing of the attention mechanism. Finally, the output values of each part are concatenated and mapped again. The multi-head self-attention calculation can be expressed as:

$$MultiHead(Q,K,V) = Concat(head_1, head_2, ..., head_h)W^0$$

$$wherehead_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{7}$$

where parameter matrix $W_i^Q \in R^{d_{model} \times d_k}$, $W_i^K \in R^{d_{model} \times d_k}$, $W_i^V \in R^{d_{model} \times d_k}$, $W^0 \in R^{d_{model} \times hd_v}$, h represents the number of parallel attention layers, whereas d_k and d_v represent the dimension values of K and V, respectively. In this experiment, $h = 2$ (according to the experience (Vaswani et al., 2017)),$d_k = d_v = d_{model} =$ 512/1024 (the length of the EEG signal segment. The EOG and EMG artifacts conditions are different).

The self-attention approach enables distinct representation subspaces of the model to be cooperatively focused on information in different places. It can model long-distance relationships without raising the computational load via parallel computing (Li et al., 2018). This paper introduces the self-attention mechanism to capture the global time dependence of EEG signal characteristics, reduce the number of model parameters, and recover the EEG's abnormal state.

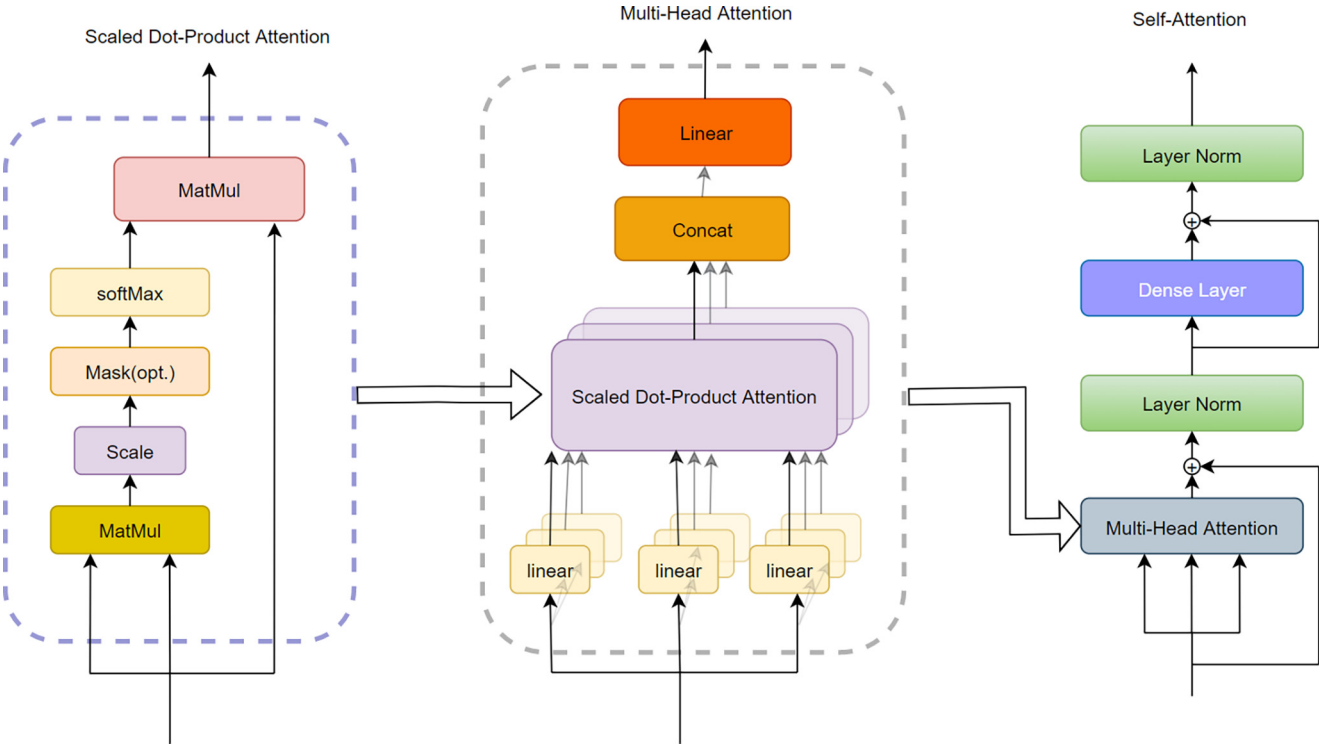


Fig. 4. Self-attention structure.

Algorithm 1. Proposed method

1. Normalize the noisy EEG according to Eq. (8);
2. Construct a BG-Attention network;
3. Train the network model in which the parameters of the BG-Attention model are adjusted to minimize the loss function in Eq. (10);
4. Input the test data to the trained BG-Attention model and output the denoised EEG signal.

Training process

In order to facilitate the learning procedure, we normalized the input contaminated EEG segment and the benchmark EEG segment by dividing the standard deviation of the contaminated EEG segment according to equations:

$$\hat{x} = \frac{x}{\sigma_y}, \qquad \hat{y} = \frac{y}{\sigma_y}, \tag{8}$$

where y denotes the mixed one-dimensional signal of EEG and artifacts; x denotes the pure EEG signal as the benchmark; σ_y is the standard deviation of y. We store the standard deviation of each noise segment in advance. The network output is then multiplied by the corresponding standard deviation to restore the EEG signal.

The denoising network is used to learn a nonlinear function f, which maps the contaminated EEG signals to the benchmark EEG signals:

$$\tilde{x} = f(\hat{y}, \theta) \tag{9}$$

The learning process was implemented by gradient descent to minimize the error between the contaminated and pure EEG signal segments. We use the mean squared error (MSE) as the loss function (Christoffersen and Jacobs, 2004):

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N} ||\tilde{x}_i - \hat{x}_i||_2^2 \tag{10}$$

where N is the number of one section of the EEG samples; $\tilde{x}_i$ is the i-th sample of the neural network output; $\hat{x}_i$ is the i-th sample of the benchmark signal x. The training process of the BG-Attention model in practice is depicted in Algorithm 1.

DATASET

The EEGdenoisenet dataset (Zhang et al., 2021), proposed by Zhang in 2020, was utilized in this paper. The data set comprised 52 people who completed both actual and imagined left- and right-hand motor activities. EEG signals were captured at 512 Hz sampling frequency, 256 Hz sampling frequency for EOG, and 512 Hz sampling frequency for EMG artifact. Moreover, it contains 4514 pure EEG segments, 3400 EOG artifact segments, and 5598 EMG artifact segments. The data set is available at https://github.com/ncclabsustech/EEG- denoiseNet. This data set allows the user to synthesize a variety of SNR signals for use, as detailed below.

Overview

The artifacts in EEG signals mainly come from environmental factors, experimental errors, and physiological artifacts. This paper focuses primarily on the impact of physiological artifacts on EEG signals and the related solutions. Physiological artifacts mainly include EOG artifacts, EMG artifacts, and ECG artifacts (Jiang et al., 2019). It is worth pointing out that ECG artifacts only appear in a few cases, so this paper will not do too much research on them. EOG artifacts are widely present in the collected EEG signals. They originate from eye movements and blinking and can travel across the scalp to be recorded along with the EEG signals. Dealing with EEG signals' EMG artifacts is difficult since they originate from diverse muscle groups. These artifacts may be created by muscle contractions and stretch anywhere near where the signal is collected. Theoretically, the muscle artifact frequency distribution ranges from 0 Hz to 200 Hz (Chen et al., 2017). Therefore, it is particularly challenging to eliminate EMG artifacts. In this paper, the experimental performance of the proposed network is tested for EMG and EOG artifacts, respectively.

Semi-synthetic data

We merged pure EEG, pure EMG, and pure EOG data to create simulated contaminated EEG signals using Eq. (11).

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

where n indicates the (EMG or EOG) artifact; λ is the hyper-parameter that controls the SNR of contaminated EEG signal. Specifically, the SNR of the contaminated segment can be adjusted according to Eq. (12).

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

The RMS is the root mean squared. Notably, the higher λ level represents lower SNR, representing more EOG or EMG artifacts added to the EEG signal. From previous studies, the SNR of EEG contaminated by EOG artifacts is usually ranging from $-$ 7 to 2 dB (Wang et al., 2015), while the SNR of EEG contaminated by EMG artifacts are between $-$ 7 and 4 dB (Chen et al., 2017).

In this experiment, the benchmark signal was the pure EEG signal, and the contaminated EEG signal was the mixed segment corresponding to it. To create EEG signals containing EOG artifacts, we utilized 3400 EEG segments and 3400 EOG segments. Similarly, EEG signals containing EMG artifacts were generated using 4514 EEG and 5598 EMG segments. We repeated randomly selected EEG segments to match the amount of EEG segments with EMG artifact segments. Then according to Eq. (9), each set was generated by randomly mixing EEG segments and artifact segments linearly, and the SNR values are made up of ten different levels (-7dB, $-$6dB, $-$5dB, $-$4dB, $-$3dB, $-$2dB, $-$1dB, 0 dB, 1 dB, 2 dB). So, the data size is tenfold increased. 80 % of the created data sets are utilized to

create training sets, 10 % to create verification sets, and 10 % to create test sets.

EXPERIMENTS

The denoising performance of the proposed network is demonstrated by reconstructing the EEG signals contaminated by EMG and EOG artifacts, two types of the most challenging artifacts in EEG signal analysis. 1D-ResCNN, FCNN, and RNN are adopted here as comparisons. 1D-ResCNN has been used to eliminate EOG, EMG, and ECG artifacts with better results than traditional denoising techniques. RNN is capable of evaluating the performance gap between GRU and bidirectional GRU. FCNN is the most basic deep neural network, demonstrating DL's potential for EEG signal denoising.

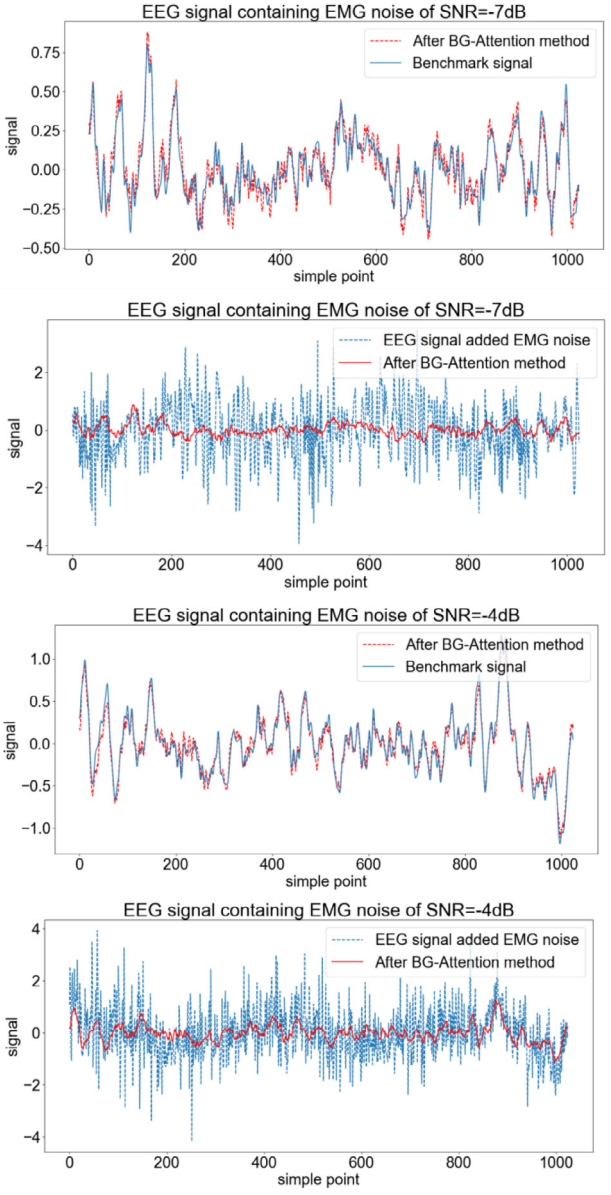


Fig. 5. Denoised results under different SNR situations.

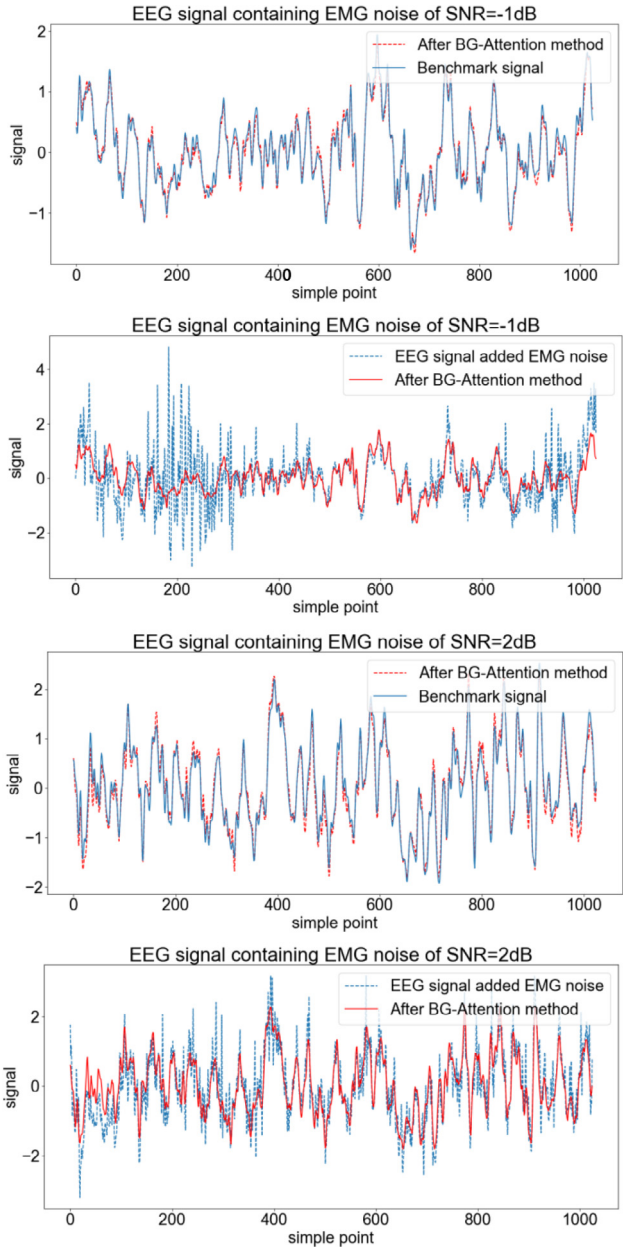


Fig. 5 (*continued*)

Performance indicators

We quantitatively examine the performance of the network by applying three objective indexes to the denoising data. Including the time-domain RRMSE (see Eq.13), the SNR (see Eq. (12)), and the correlation coefficient (CC, see Eq. (14)):

$$RRMSE = \frac{RMS(f(y) - x)}{RMS(x)} \tag{13}$$

$$CC = \frac{Cov(f(y), x)}{\sqrt{Var(f(y))Var(x)}} \tag{14}$$

where *cov* represents covariance and *var* represents variance. The RRMSE index evaluates the average difference between the benchmark signal and the

processed signal to determine the recovery degree of the processed signal. The SNR index can reflect the degree of artifacts in the processed signal. The CC index demonstrates the correlation between the benchmark signal and the processed signal; the more significant the correlation, the more nonlinearity is preserved in the EEG signal.

Electromyography (EMG) artifact elimination effect

We split the SNR range $-7 \sim 2$ dB into four equal portions to demonstrate the effect of the noise reduction. Fig. 5 shows eight pictures, two pictures in one group. Each group includes a denoised EEG signal and a benchmark signal, a noise-contaminated EEG signal and a denoised EEG signal. The signal comparison before and after restoration and the correlation between the benchmark signal and the signal processed by our network is thoroughly shown in each set of images. The figures show that the proposed network has strong signal reconstruction capabilities under different SNR situations and can preserve nonlinear information to a great degree.

Electrooculogram (EOG) artifact elimination effect

We split the SNR range $-7 \sim 2$ dB into four equal portions to demonstrate the noise reduction. Fig. 6 shows eight pictures, two pictures in one group. Each group includes a denoised EEG signal and a benchmark signal, a noise-contaminated EEG signal and a denoised EEG signal. The signal comparison before and after restoration and the correlation between the benchmark signal and the signal processed by our network is thoroughly shown in each set of images. The figures show that the proposed network has strong signal reconstruction capabilities under different SNR situations and can preserve nonlinear information to a great degree.

Technical performance analysis

We analyzed the performance of the proposed network based on SNR, RRMSE, and CC. Different SNR signals are used to display the processing effect of each network. Each test was repeated ten times and averaged.

Performance analysis based on the SNR

The SNR values for eliminating EMG and EOG artifacts using FCNN, 1D-ResCNN, RNN, and the proposed network are shown in Fig. 7(**A**,**B**). The graph shows that RNN has the worst performance, implying that although RNN has certain advantages in processing sequence signals, it is powerless to process complex sequence signals. Our proposed network has achieved the best experimental results, with the most excellent SNR value reaching 35.13 for EMG artifacts and 39.9 for EOG artifacts. However, it is easy to observe that when the SNR is -7 under EMG artifact, the outcome is not as excellent as 1D-ResCNN. In addition, we discovered a significant difference among the outputs of all networks in low SNRs and high SNRs, indicating that the SNR processing range of deep neural networks is limited and

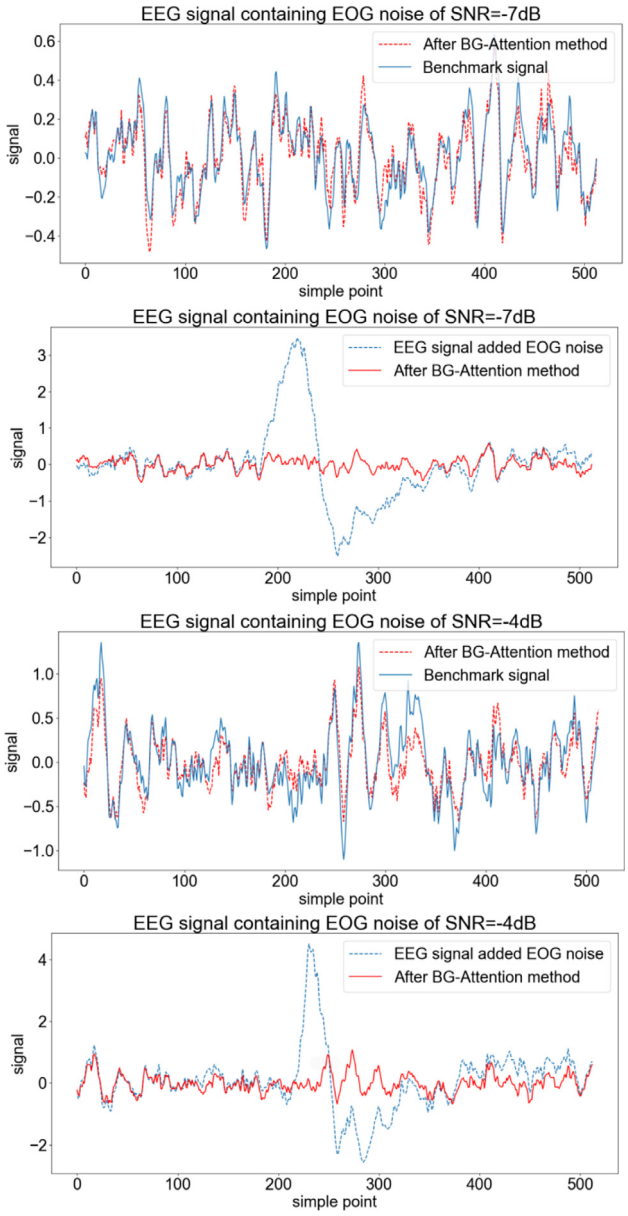


Fig. 6. Denoised results under different SNR situations.

that specific strategies may be used to improve it.The detailed results are shown in Table 1 and Table 2.

Performance analysis based on RRMES values

Fig. 8(**A**,**B**) show the RRMSE values after denoising by FCNN, 1D-ResCNN, RNN, and the proposed algorithm at different SNR values. Only in the situation of EMG artifacts can it be seen that the RRMSE value of 1D-ResCNN steadily decreases with increasing SNR, which indicates that other networks exhibit the overfitting phenomena; however, for our proposed network, the RRMSE growth trend is not evident; hence the effect is minimal. We can also notice that the RRMSE value of the RNN network is substantially higher than the other three networks while processing the EOG artifact, indicating its weakness in handling EOG artifacts.

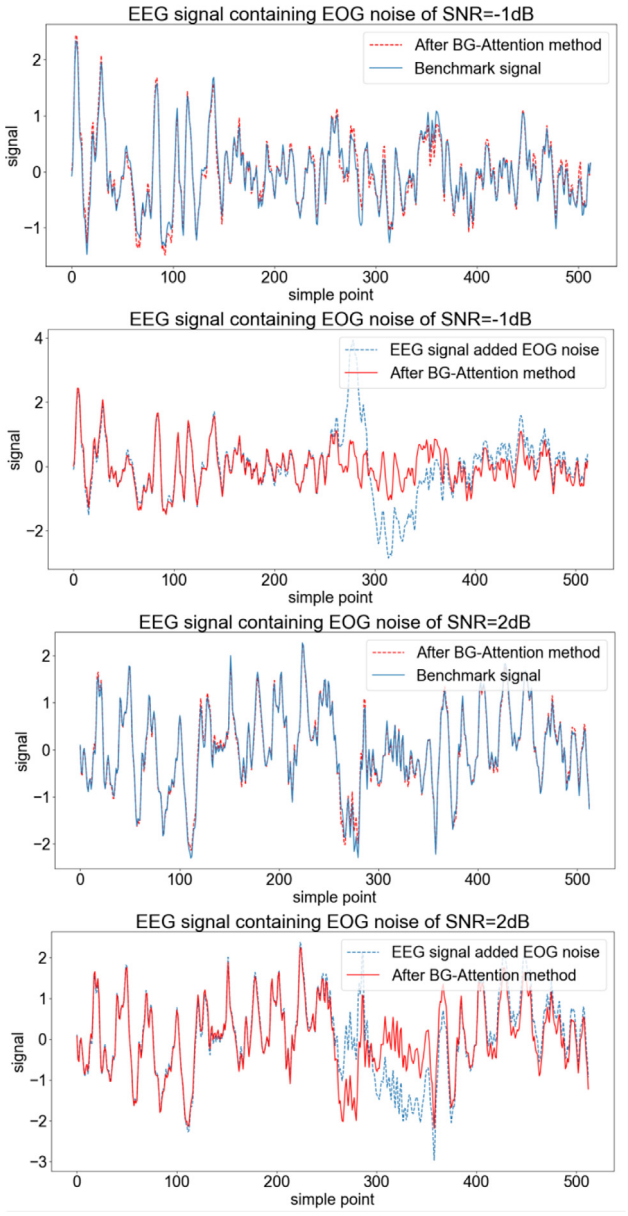


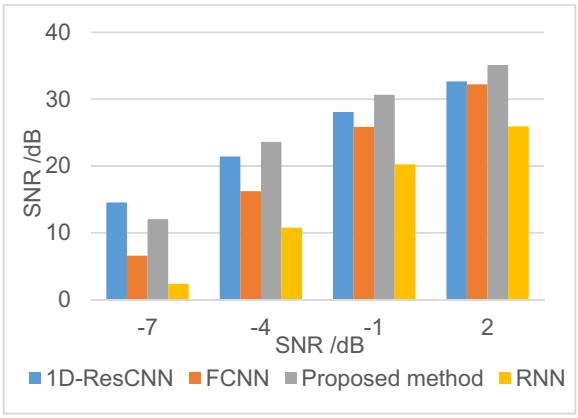
Fig. 6 (*continued*)

Compared to the FCNN, 1D-ResCNN, and RNN networks, the proposed network has the minimum RRMSE. Therefore, the proposed network is better at reconstructing EEG signals. The detailed results are shown in Table 3 and Table 4.

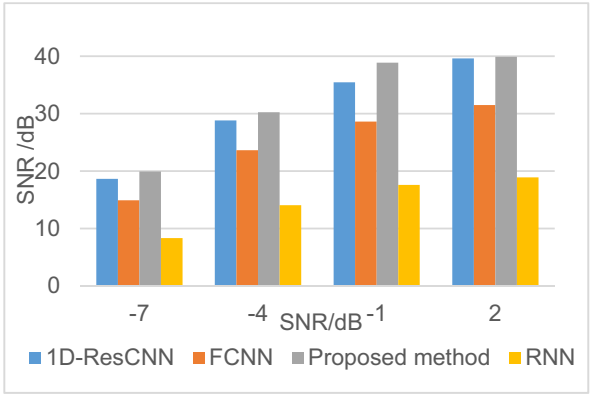
Influence of denoising on nonlinear characteristics of signals

Fig. 9 illustrates the power spectral density (PSD) of EEG signals processed by FCNN, 1D-ResCNN, RNN, and the network proposed in this paper with an SNR of −7dB. The smaller the value of the PSD, the better the network's performance (Sun et al., 2020). As can be seen from the figure, the PSD value of the EEG signal after the noise reduction by the four methods decreases in the whole frequency range of the noise. Moreover, our proposed network consistently outperforms the other three networks.

Fig. 10(**A**,**B**) show the CC indicator between the benchmark signal and contaminated EEG signal after four kinds of network processing, respectively. And we can see that our proposed network produces the highest



(a) SNR results for removing EMG artifact from EEG signals



(b) SNR results for removing EOG artifact from EEG signals

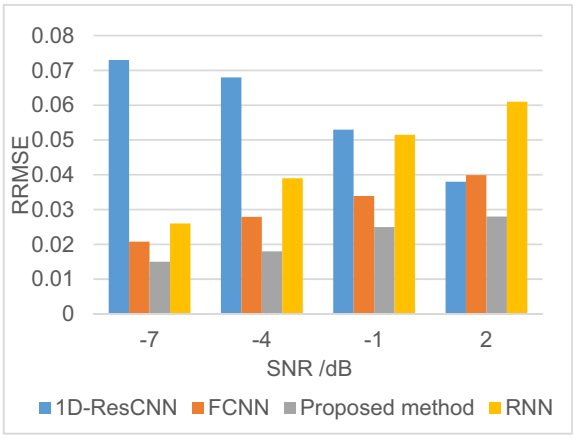
Fig. 7. SNR results of removing artifacts from EEG signal.

Table 1. SNR experimental results of EMG artifact under different SNR conditions.

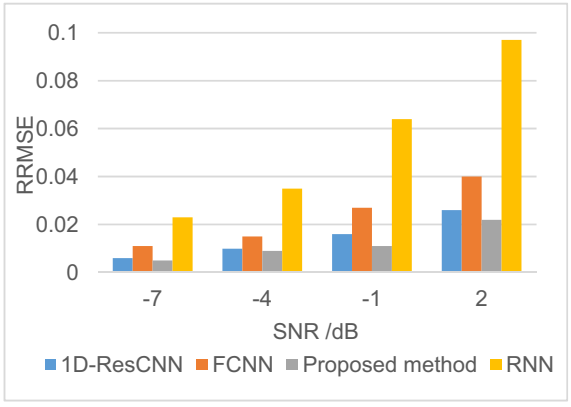
Evaluation index SNR	−7dB	−4dB	−1dB	2 dB
1D-ResCNN	15.54	22.45	28.11	32.67
FCNN	6.57	16.22	25.85	32.22
RNN	2.37	10.77	20.26	25.94
Proposed method	10.06	22.61	30.66	35.13

Table 2. SNR experimental results of EOG artifact under different SNR conditions.

Evaluation index SNR	−7dB	−4dB	−1dB	2 dB
1D-ResCNN	18.63	28.81	35.47	39.61
FCNN	14.91	23.61	28.64	31.51
RNN	8.34	14.04	17.58	18.92
Proposed method	19.92	30.25	38.88	39.93



(a)RRMSE results for removing EMG artifact from the EEG signals



(b)RRMSE results for removing EOG artifact from EEG signals

Fig. 8. RRMSE results of removing artifacts from EEG signal.

Table 3. Add RRMSE experimental results of EMG artifact.

Evaluation index RRMSE	−7dB	−4dB	−1dB	2 dB
1D-ResCNN	0.073	0.068	0.053	0.038
FCNN	0.021	0.028	0.034	0.039
RNN	0.026	0.039	0.052	0.061
Proposed method	0.015	0.018	0.025	0.028

Table 4. Add RRMSE experimental results of EOG artifact.

Evaluation index RRMSE	−7d	−4dB	−1dB	2 dB
1D-ResCNN	0.006	0.009	0.016	0.026
FCNN	0.011	0.015	0.027	0.042
RNN	0.023	0.035	0.064	0.097
Proposed method	0.005	0.009	0.011	0.022

intercorrelation between the processed EEG signal and the benchmark EEG signal. It dramatically minimizes the correlation between the artifacts and the EEG signal. In the situation of EOG artifacts, the performance of the proposed network is marginally superior to that of 1D-ResCNN, whereas RNN has the poorest processing

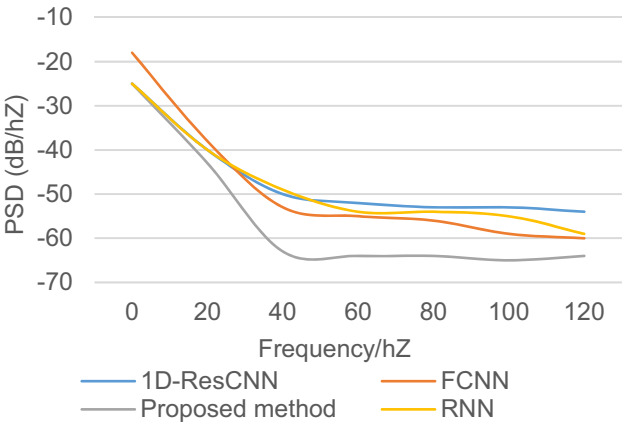
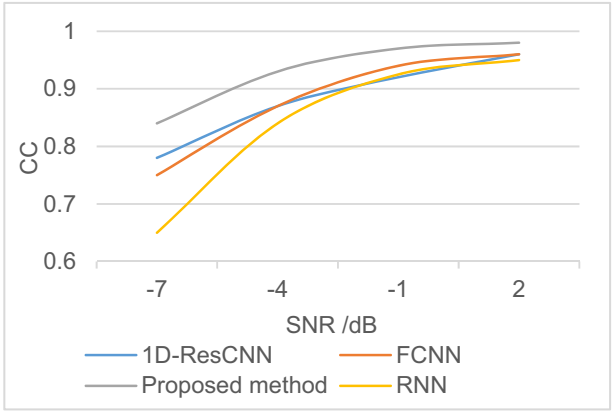
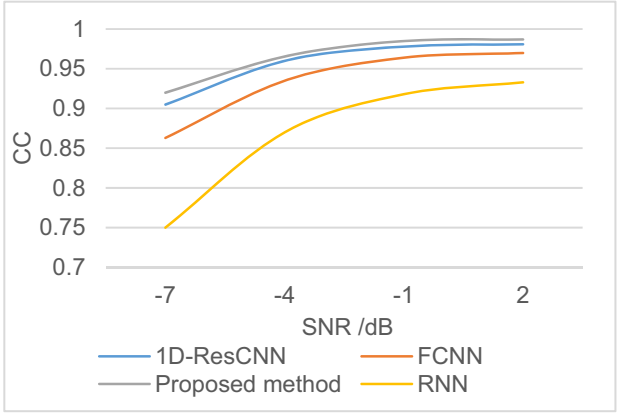


Fig. 9. PSD results for removing EOG artifact from EEG signals.



(a) CC results for removing EOG artifact from EEG signals



(b) CC results for removing EMG artifact from EEG signals

Fig. 10. CC results of removing artifacts from EEG signal.

effect based on numerous indexes. In the situation of EMG artifacts, our network's advantages are more prominent. FCNN and 1D-ResCNN have comparable CC values, followed by RNN. The detailed results are shown in Table 5 and Table 6.

According to the above indexes, BG-Attention achieves superior performance over the other three

Table 5. Correlation between processed EEG and benchmark EEG after adding EMG artifact.

Evaluation index CC	−7dB	−4dB	−1dB	2 dB
1D-ResCNN	0.791	0.883	0.942	0.971
FCNN	0.755	0.894	0.953	0.973
RNN	0.653	0.855	0.933	0.952
Proposed method	0.844	0.933	0.971	0.982

Table 6. Correlation between processed EEG and benchmark EEG after addition of EOG artifact.

Evaluation index CC	−7dB	−4dB	−1dB	2 dB
1D-ResCNN	0.915	0.963	0.977	0.981
FCNN	0.863	0.942	0.964	0.972
RNN	0.752	0.873	0.918	0.933
Proposed method	0.922	0.966	0.985	0.987

networks. In the situation of low SNR containing EMG artifacts, however, none of the four methods performs well. This is not difficult to understand, considering that processing a signal with low SNRs means the need to remove a great deal of noise from the signal, which is a challenging, if not impossible, task.

Limitations and room for improvement

Although we have achieved good outcomes, several aspects may still be improved. First, we mainly focused on removing ocular and muscular artifacts since they are available in many datasets. Muscular artifacts are widely considered the most complex artifacts to eliminate. Still, we must consider the coexistence of many artifacts, which will unquestionably raise the difficulty of the experiment but is necessary.

Secondly, Low SNR EEG signal processing is a critical and vital study area since we cannot just deal with simple high SNR situations in practical applications. It is thus crucial to increase the processing efficiency of DL networks under conditions of low SNR.

Thirdly, the data set used in this experiment is aimed at the single-channel processing problem, which compromises the spatial characteristics of EEG signals to a large extent. Next, we will concentrate on utilizing the spatial properties of EEG signals to solve the problem.

The BG-Attention method is proposed in this paper to deal with artifacts in EEG signals. The bidirectional GRU networks are used to process consecutive signals. They may gather past and future information at the same time. The attention mechanism is used to pay attention to details of varying relevance, emphasizing the contribution of crucial segments to denoising. The combination of the two networks can solve the denoising problem of EEG well. Experimental results show that the proposed network achieves lower RRMSE values, a higher SNR, and better noise suppression than FCNN, 1D-ResCNN, and RNN on the test set. Furthermore, the CC index is maintained reasonably high under different SNR values, indicating that the proposed algorithm maintains the nonlinear properties of the EEG signal well under different SNR settings. This algorithm can be applied to the pre-processing stage of the EEG signals, which is significant for medical diagnosis and BCI. Next, we will study multiple artifacts, high-density artifacts, and spatial characteristics of EEG signals to further improve the all-around performance of this algorithm. In the future, we hope proposed processing network can play a crucial role in EEG signal processing.

REFERENCES

Chen X et al (2017) Independent vector analysis applied to remove muscle artifacts in EEG data. IEEE Trans Instrum Meas 66 (7):1770–1779.

Chen X et al (2019b) Removal of muscle artifacts from the EEG: a review and recommendations. IEEE Sens J 19(14):5353–5368.

Chen JX, Jiang DM, Zhang YN (2019a) A hierarchical bidirectional GRU model with attention for EEG-based emotion classification. IEEE Access 7:118530–118540.

Cho, Kyunghyun, et al. (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Christoffersen P, Jacobs K (2004) The importance of the loss function in option valuation. J Financ Econ 72(2):291–318.

Cichocki A et al (2005) EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease. Clin Neurophysiol 116(3):729–737.

Craik A, He Y, Contreras-Vidal JL (2019) Deep learning for electroencephalogram (EEG) classification tasks: a review. J Neural Eng 16(3) 031001.

Croft RJ, Barry RJ (2000) Removal of ocular artifact from the EEG: a review. Neurophysiol Clin Clin Neurophysiol 30(1):5–19.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134(1):9–21.

Dey R, Salem FM (2017) Gate-variants of gated recurrent unit (GRU) neural networks. 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS), 2017.

Egambaram A et al (2016) Comparison of envelope interpolation techniques in Empirical Mode Decomposition (EMD) for eyeblink artifact removal from EEG. 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2016.

Fernández-Rodríguez Á et al (2020) Effects of spatial stimulus overlap in a visual P300-based brain-computer interface. Neuroscience 431:134–142.

Gao X-Y et al (2021) Multi-head self-attention for 3D point Cloud classification. IEEE Access 9:18137–18147.

He P et al (2007) Removal of ocular artifacts from the EEG: a comparison between time-domain regression method and adaptive filtering method using simulated data. Med Biol Eng Compu 45(5):495–503.

He P, Wilson G, Russell C (2004) Removal of ocular artifacts from electro-encephalogram by adaptive filtering. Med Biol Eng Comput 42(3):407–412.

Jiang X, Bian G-B, Tian Z (2019) Removal of artifacts from EEG signals: a review. Sensors 19(5):987.

Lazzarotto G et al (2021) Effect of Memantine on Pentylenetetrazol-induced Seizures and EEG Profile in Animal Model of Cortical Malformation. Neuroscience 457:114–124.

Li J et al (2018) Multi-head attention with disagreement regularization. arXiv preprint arXiv:1810.10183, 2018.

Lynn HM, Pan SB, Kim P (2019) A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks. IEEE Access 7:145395–145405.

Maimaiti B et al (2021) An Overview of EEG-based Machine Learning Methods in Seizure Prediction and Opportunities for Neurologists in this Field. Neuroscience.

Mathe M, Padmaja M, Battula TK (2022) Artifact Removal Methods in EEG Recordings: A Review. Proc Eng Technol Innov 20:35–56.
McMenamin BW et al (2009) Validation of regression-based myogenic correction techniques for scalp and source-localized EEG. Psychophysiology 46(3):578–592.
Miao M, Wenjun Hu, Zhang W (2021) A spatial-frequency-temporal 3D convolutional neural network for motor imagery EEG signal classification. SIViP 15(8):1797–1804.
Molla MKI et al (2012) Artifact suppression from EEG signals using data adaptive time domain filtering. Neurocomputing 97:297–308.
Muhammad G, Shamim Hossain M, Kumar N (2020) EEG-based pathology detection for home health monitoring. IEEE J Sel Areas Commun 39(2):603–610.
Pion-Tonachini L, Kreutz-Delgado K, Makeig S (2019) ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. NeuroImage 198:181–197.
Song J-L, Wenfeng Hu, Zhang R (2016) Automated detection of epileptic EEGs using a novel fusion feature and extreme learning machine. Neurocomputing 175:383–391.
Sun W et al (2020) A novel end-to-end 1D-ResCNN model to remove artifact from EEG signals. Neurocomputing 404:108–121.
Van SF et al (2019) Critical comments on EEG sensor space dynamical connectivity analysis. Brain Topogr 32(4):643–654.
Vaswani A et al (2017) Attention is all you need. Adv Neural Inform Process Syst 30.
Wang Ge, Ye JC, De Man B (2020) Deep learning for tomographic image reconstruction. Nat Mach Intell 2(12):737–748.
Wang M-Y, Yuan Z (2021) EEG Decoding of Dynamic Facial Expressions of Emotion: Evidence from SSVEP and Causal Cortical Network Dynamics. Neuroscience 459:50–58.
Wang G et al (2015) The removal of EOG artifacts from EEG signals using independent component analysis and multivariate empirical mode decomposition. IEEE J Biomed Health Inform 20 (5):1301–1308.
Wu F et al (2022) Learning Spatial-Spectral-Temporal EEG Representations with Deep Attentive-Recurrent-Convolutional Neural Networks for Pain Intensity Assessment. Neuroscience 481:144–155.
Xie J, Colonna JG, Zhang J (2021) Bioacoustic signal denoising: a review. Artif Intell Rev 54(5):3575–3597.
Yang Z et al (2016) Stacked attention networks for image question answering. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
Yu Y et al (2019) A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput 31(7):1235–1270.
Zhang H et al (2021) Eegdenoisenet: A benchmark dataset for deep learning solutions of eeg denoising. J Neural Eng 18(5) 056057..

(Received 20 April 2022, Accepted 4 October 2022)
(Available online 12 October 2022)