

An Approach for EEG Denoising Based on Wasserstein Generative Adversarial Network

Yuanzhe Dong, Xi Tang, Qingge Li, Yingying Wang¹, Naifu Jiang¹, *Member, IEEE*, Lan Tian, Yue Zheng, Xiangxin Li¹, Shaofeng Zhao, Guanglin Li¹, *Senior Member, IEEE*, and Peng Fang¹, *Senior Member, IEEE*

Abstract—Electroencephalogram (EEG) recordings often contain artifacts that would lower signal quality. Many efforts have been made to eliminate or at least minimize the artifacts, and most of them rely on visual inspection and manual operations, which is time/labor-consuming, subjective, and incompatible to filter massive EEG data in real-time. In this paper, we proposed a deep learning framework named Artifact Removal Wasserstein Generative Adversarial Network (AR-WGAN), where the well-trained model can decompose input EEG, detect and delete artifacts, and then reconstruct denoised signals within a short time. The proposed approach was systematically compared with commonly used denoising methods including AutoEncoder, Wiener Filter, and Empirical Mode Decomposition, with both public and self-collected datasets. The experimental results proved the promising performance of AR-WGAN on automatic artifact removal for massive data across subjects, with correlation coefficient up to 0.726 ± 0.033 , and temporal and spatial relative root-mean-square error as low as 0.176 ± 0.046 and 0.761 ± 0.046 , respectively. This work

may demonstrate the proposed AR-WGAN as a high-performance end-to-end method for EEG denoising, with many on-line applications in clinical EEG monitoring and brain-computer interfaces.

Index Terms—Electroencephalogram, generative adversarial network, denoising, artifacts removal, BCI.

I. INTRODUCTION

ELECTROENCEPHALOGRAPHY (EEG) can directly reflect brain activities by recording scalp potential changes, and thus it is widely used in cognitive psychology, neurology, psychiatry, and brain-computer interface (BCI) as well [1], [2], [3]. In clinical applications, the recorded EEG contains not only useful information of brain activities but also unexpected artifacts like eye blinks (ocular artifacts), head muscle signals (myogenic artifacts), heartbeat signals, etc., which usually contaminate EEG in different ways and generate high energy in multiple frequency bands, e.g., myogenic and ocular artifacts can bring EEG with large amplitude spikes and significant drifts, respectively [4]. These interferences strongly lower the quality of EEG recordings and therefore denoising is very critical for EEG processing and analysis [5], [6], [7]. Traditionally, these artifacts were removed by manual operations and visual inspection [8], which is time/labor-consuming and inappropriate for online and real-time EEG data processing [9].

Several conventional machine-learning (ML) approaches, including regression-based methods, adaptive filter-based methods, blind source separation (BSS) methods, etc., have been developed to remove artifacts in raw EEG data. The regression-based methods firstly estimate and then subtract artifacts from raw data to gain pure EEG, but these methods are sensitive to outliers which may yield misleading results [10]. The adaptive filter-based methods dynamically estimate filter parameters based on EEG input and then remove artifacts, while they are linear least-squared estimators for stochastic processes, searching for different schemes to do the estimation, which is relatively slow to reach convergence and computationally expensive [11]. The BSS methods, especially the independent-component-analysis (ICA), decompose EEG into different components corresponding to brain activities and artifacts, respectively, and thereafter the components of brain activities are manually selected and combined to reconstruct pure EEG. Therefore, the BSS method cannot identify artifact-related components automatically, and a pre-trained classifier [12], [13], [14], [15] or manual selection is needed

Manuscript received 3 April 2023; revised 28 July 2023; accepted 24 August 2023. Date of publication 29 August 2023; date of current version 7 September 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFA0701103, in part by the National Natural Science Foundation of China under Grant U21A20479 and Grant 81927804, in part by the Chinese Academy of Sciences (CAS) Youth Innovation Promotion Association under Grant Y2022094, in part by the Shenzhen Fundamental Research Program under Grant JCYJ20200109114805984 and Grant JCYJ20190808102001750, and in part by the Futian Healthcare Research Project under Grant FTWS2022015. (Corresponding authors: Guanglin Li; Peng Fang.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, under Approval No. SIAT-IRB-190315-H0325.

Yuanzhe Dong, Xi Tang, Qingge Li, Yingying Wang, Naifu Jiang, Xiangxin Li, Guanglin Li, and Peng Fang are with the CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Shenzhen 518055, China, also with the Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Shenzhen Engineering Laboratory of Neural Rehabilitation Technology, Shenzhen 518055, China (e-mail: gl.li@siat.ac.cn; peng.fang@siat.ac.cn).

Lan Tian and Yue Zheng are with the CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Shenzhen 518055, China, and also with the Shenzhen Engineering Laboratory of Neural Rehabilitation Technology, Shenzhen 518055, China.

Shaofeng Zhao is with the Department of Rehabilitation Medicine, The Eighth Affiliated Hospital, Sun Yat-sen University, Shenzhen 518000, China (e-mail: zhaoshf6@mail.sysu.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3309815

to identify and reject artefactual components. Moreover, the ICA-based methods cannot cope with single-channel data and need auxiliary channels. In addition, they are commonly less effective in processing non-biological artifacts such as sudden impedance changes caused by headset movements [9]. Some studies investigated the artifact subspace reconstruction (ASR) [16], [17] for EEG denoising, which is similar to the principal-component-analysis (PCA) that large-variance components are rejected, and purified data are reconstructed by the remained components. Similarly, the ASR also needs auxiliary channels and has relatively high time complexity since it contains eigenvalue decomposition of the sample covariance. To address these issues, a fully automated and efficient artifact-removal algorithm that can process large amounts of data in real time is needed.

The deep learning (DL) methods [18] boomed with the coming out of Alexnet in 2015 [19] and have made a huge success in multiple fields, especially for computer vision and natural language processing, and also gained much attention in neural engineering [20], [21], [22]. Considering the boosting data size and advancing hardware support, more and more researchers preferred to apply the DL methods for EEG applications, including motor imagery classification [23], [24], [25], emotion recognition [26], [27], [28], data augmentation [29], [30], etc. Recently, the use of DL methods in EEG artifact removal was introduced, which gained more favorable performance compared with the conventional ML-based methods. Zhang et al. [31] provided a publicly available structured dataset and tested four deep neural networks as benchmarks, including a fully-connected neural network, a simple convolution neural network (CNN), a complex CNN, and a recurrent neural network (RNN). Sun et al. [32] proposed a one-dimensional residual CNN (1D-ResCNN) to remove various susceptible physiological signals from EEG. Inspired by the U-Net architecture in image segmentation, Chuang et al. [33] developed a novel model to remove pervasive EEG artifacts and reconstruct brain signals, which implemented the ensemble loss function to model complex signal fluctuations in EEG recordings. Leite et al. [34] presented a 2D deep convolutional autoencoder to filter noises of eye blinking and jaw clenching. These DL methods avoid the time-consuming preprocessing and feature extraction on raw EEG data to directly learn meaningful information, which can capture underlying and high-level features. However, they are vulnerable to adversarial samples, where real-world EEG data are prone to be corrupted by small perturbations [35], [36]. These perturbations may mislead the model and lead to dramatic degradation in the denoising performance.

Adversarial training [37] can solve this problem by enhancing the robustness of the model intrinsically, and Generative Adversarial Network (GAN) is one of the most representative works. GAN is a novel technology for unsupervised learning, which was first used in the area of computer vision to implement multiple tasks like generating non-existent images, transferring styles of images, and improving the resolution of images. The initial GAN was proposed in 2014 by Goodfellow et al. [38], who synthesized real-like images by implicitly modeling data into high-dimensional

spaces. Usually, a GAN contains two opposite components, i.e., a generator and a discriminator, and both are trained in the learning stage and compete with each other. The generator tries to generate real-like data with random input to fool the discriminator, whereas the discriminator receives both generated and realistic data and aims to judge reality. After multiple rounds of competition, the generator can synthesize real-like data that the discriminator cannot tell apart, reaching a so-called Nash equilibrium. As a simple analogy, the generator is a counterfeiter making fake bank note without detection and the discriminator is a policeman trying to distinguish between fake and real money. Different types of GAN have been proposed in the past few years. The Deep Convolutional GAN (DCGAN) replaces the multi-perceptrons with convolutional layers [39]; the Conditional GAN (CGAN) provides additional information which helps the discriminator in finding conditional probability instead of joint probability [40]; the CycleGAN provides a way for image-to-image translations, like translating horse images into zebra images [41]; the Wasserstein GAN (WGAN) minimizes the approximation of Wasserstein distance rather than the Jensen-Shannon divergence as in the original GAN formulation, which provides more stability in training and could avoid mode collapse [42].

Many applications of GAN have been proposed and realized. Some studies kept tracking GAN-based works and tried to implement GANs for sequence and time-series data generation, imputation, and augmentation [43]; Some other employed GANs in EEG data augmentation to solve the problem of data scarcity and imbalance, improving accuracies in various classification tasks [29], [30]; A few works have implemented GANs in EEG denoising, e.g., Brophy et al. [44] used GAN to denoise real-world EEG signals by mapping noisy signals to clean signals according to the nature of respective artifacts. An et al. [45] proposed a GAN-based denoising method to automatically filter multichannel EEG signals and defined a new loss function to ensure that the filtered signals could preserve effective original information and energy as much as possible. Nevertheless, most current GAN-based models only remove EMG and EOG individually from EEG but did not remove the combined artifacts. Although some pioneering studies, e.g., Sumiya et al. [46], have proposed a GAN-based noise-reduction model to remove the combined artifacts in EEG, more solid evidence of the improvements in filtered data should be explored and provided.

To address these issues, we presented an EEG denoising approach based on the Wasserstein GAN (WGAN), named Artifact Removal WGAN (AR-WGAN), and evaluated its performance by using a public dataset and a self-collected dataset. This study systematically overviews the WGAN model and demonstrates the algorithm of AR-WGAN. The proposed model was compared with some commonly used filtering algorithms, including Denoised AutoEncoder, Wiener Filter, and Empirical Mode Decomposition, in both time and time-frequency domains, with evaluation metrics such as power ratios, relative root mean squared error (RRMSE), and correlation coefficient (CC). The robustness and generalization capacity in real implementations of this work was discussed and some outlook was provided. The main contribution of this

study is that we provided an end-to-end denoising model based on WGAN, which can filter massive raw EEG data within a short time. The approach can embody enough robustness and cope with real-world EEG data with inevitable perturbations, and the fast denoising demonstrates the AR-WGAN as an automatic and online-capable artifact removal approach.

II. DATASET AND METHODS

A. Methods

1) *Generative Adversarial Network*: In a basic GAN, the discriminator can be approximately characterized as a function to map the data into a probability (0 to 1) where the input belongs to the real data. The generator is fixed when the discriminator is being trained to classify the real data (output close to 1) and fake data (output close to 0) from the generator. Similarly, the discriminator will be frozen when it is optimized, and the generator continues to learn the real data distribution to lower the accuracy of the discriminator.

The critical point of GANs is the probability density or probability mass function of observation data. GANs are trained by implicit computing and maximizing the similarity of probability distribution between the real data and generated data from the candidate model. The training of GANs contains the optimization of discriminator parameters that can maximize the classification accuracy, and also the finding of generator parameters that can maximally generate fake data to confuse the discriminator. The cost of training can be formularized as

$$\min_{\theta_g} \max_{\theta_d} V(G, D) = E_{x_r \sim p_r} [\log(D(x_r))] + E_{x_g \sim p_g} [\log(1 - D(x_g))], \quad (1)$$

where θ_d , θ_g , x_r , and x_g are the discriminator parameters, generator parameters, real data, and synthetic data, respectively, $D(x)$ returns the probability (0 to 1) of x belonging to the real data (p_r) or generated data (p_g) distributions. Specifically, the input data to the generator are random noises, which can be stated as

$$x_g = G(z), \quad (2)$$

where variables G and D denote the generator and discriminator, respectively.

In the training procedure, the parameters in one model are updated, whereas the parameters of the other model are fixed. Goodfellow et al. [28] demonstrated that for a fixed generator, the optimal discriminator is $D(x) = p_{\text{data}}(x)/(p_{\text{data}}(x) + p_g(x))$. They also proved that the generator is optimal when $p_{\text{data}}(x) = p_g(x)$, which is equivalent to the situation that the probability for any sample predicted by an optimal discriminator is 0.5. In other words, the desired state is that the generator can maximally fool the discriminator to prevent any accurate distinguishing between the real and fake samples.

Initially, the fully connected layers were used in GAN architectures for both generator and discriminator, and were applied to three simple image data sets of MNIST, CIFAR-10, and Toronto Face Data Set [38]. Arjovsky et al. [42] proposed the WGAN with an alternative cost function which is an approximation of the Wasserstein distance (also known

as the Earth-Mover distance). Considering its stability and fast convergence in training, the WGAN was implemented as the basic model in this study.

The adversarial training of conventional GANs can be formularized by minimizing the Jensen-Shannon divergence between the probability of real and generated data. However, the discontinuity of the Jensen-Shannon divergence makes it hard to provide useful gradients to the generator for optimally updated parameters, which is one of the main reasons for the instability of GANs. To solve this problem, the Jensen-Shannon divergence is replaced with the Wasserstein distance in WGAN as

$$W(P_r, P_g) = \inf_{\gamma \sim \Pi(P_r, P_g)} E_{(P_r, P_g) \sim \gamma} [\|x_r - x_g\|], \quad (3)$$

where $\Pi(P_r, P_g)$ denotes all the possible joint distributions of real distribution P_r and generated data distribution P_g . The Wasserstein distance is continuous and differentiable almost everywhere, and thus can provide useful reasoning gradients to optimize generator. Since it is difficult to implement (3), the early studies on WGAN normally provided the Kantorovich-Rubinstein duality to measure the Wasserstein distance as

$$W(P_r, P_g) = \frac{1}{K} \sup_{\|f\|_L \leq K} E_{x \sim p_r} [f(x_r)] - E_{x \sim p_g} [f(x_g)], \quad (4)$$

where f denotes the set of 1-Lipschitz function. In real implementations, f is replaced by discriminator D and $\|f\|_L \leq K$ is replaced by $\|D\|_L \leq 1$, and the final loss function can be formularized as

$$\min_{\theta_G} \max_{D_G} L(P_r, P_g) = E_{x_r \sim P_r} [D(x_r)] - E_{x_g \sim P_g} [D(x_g)]. \quad (5)$$

In GAN applications, given the fact that the convolutional layers are powerful to extract features and with relatively less computational burden compared with the fully connected layers, many EEG-relevant works concentrate more on CNNs. Generally, most studies utilized 2D convolutional layers in GAN-based models to extract EEG features. Some of them treated 64 EEG channels like a map of 8×8 , however, the whole model could not contain too many convolutional layers and failed to extract deep latent features in EEG. Some other studies directly reshaped the EEG time series into a format of image, but this inappropriate operation may cause irreversible damage to local features. Considering that the 1D convolutional layer is extremely well suited for EEG time series, in this work, the 1D convolutional and transposed convolutional layers were implemented in the generator, and the 1D convolutional layers were used in the discriminator. The basic blocks and architecture for both the generator and discriminator are illustrated in Fig. 1, and the detailed input/output data are shown in TABLES I and II. The training process can be seen in Fig. 2, where the contaminated EEG data y is input to the generator, and then decomposed and reconstructed by the generator, with output of the generated data $G(y)$. Thereafter, the generated $G(y)$ and the real data together are input into the discriminator, which gives the probability of the input belonging to real data that ranges

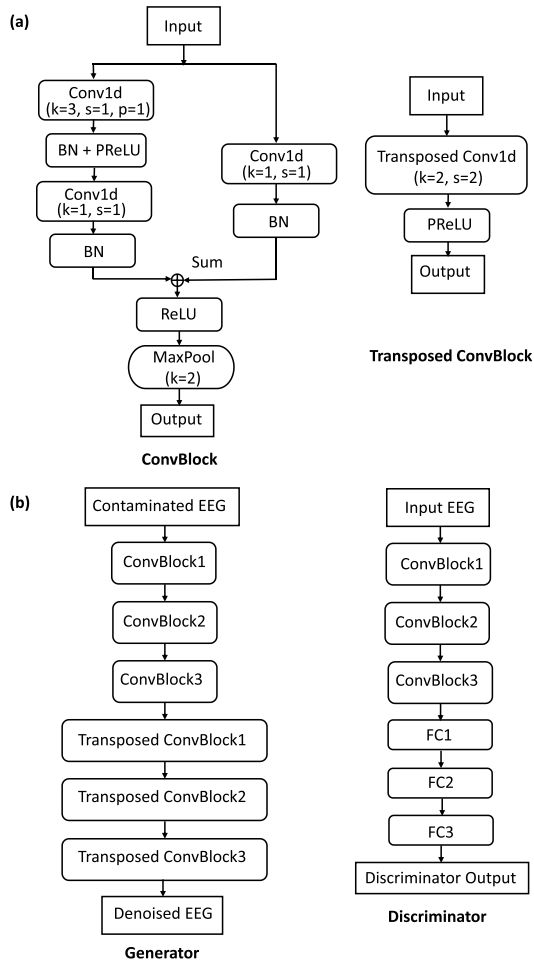


Fig. 1. (a) The basic blocks and (b) architecture for both the generator and discriminator. The BN in ConvBlocks denotes the batch normalization and FC in Discriminator denotes the fully connected layers.

TABLE I
PARAMETERS FOR THE GENERATOR

Block	Generator	
	Input	Output
ConvBlock1	(b,1,3400)	(b,8,1700)
ConvBlock2	(b,8,1700)	(b,32,850)
ConvBlock3	(b,32,850)	(b,128,425)
Transposed ConvBlock1	(b,128,425)	(b,32,850)
Transposed ConvBlock2	(b,32,850)	(b,8,1700)
Transposed ConvBlock3	(b,8,1700)	(b,16,3400)

from 0 to 1. The discriminator can be seen as a binary classifier, and training of the discriminator strengthens its ability to distinguish contaminated signals, enforcing the generator to remove artifacts and generate filtered data as perfectly as possible. It is worth noting that the output $G(y)$ from the generator is the input to the discriminator, and the outputs of the denoised data y and contaminated data n from the discriminator are $D(G(y))$ and $D(n)$. Algorithm 1 summarizes the procedure and default parameters for AR-WGAN.

2) *Denoised AutoEncoder*: The Autoencoder (AE) is one of the generative and unsupervised learning models in the deep learning field with encoder-decoder architecture and is commonly used for the task of representation learning. Deep AEs can learn high-order statistical information from

TABLE II
PARAMETERS FOR THE DISCRIMINATOR

Block	Discriminator	
	Input	Output
ConvBlock1	(b,1,3400)	(b,8,1700)
ConvBlock2	(b,8,1700)	(b,32,850)
ConvBlock3	(b,32,850)	(b,128,425)
FC1	(b,54400)	(b,5000)
FC2	(b,5000)	(b,200)
FC3	(b,200)	(b,1)

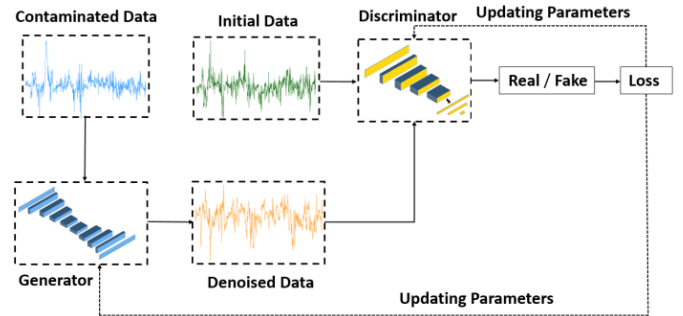


Fig. 2. The training process of AR-WGAN for denoising, where the generator received contaminated data and output the denoised data, and the discriminator received both initial and denoised data. After multiple rounds of adversarial training, the generator could remove artifacts in the contaminated data quite well.

the input data and are usually built symmetrically concerning their dimensionality. Normal AEs mainly contain three parts, i.e., encoder, bottleneck, and decoder. The encoder is a set of blocks to compress input data to the bottleneck. The responsibility for the bottleneck is restricting the flow of information from encoder to decoder, allowing the most important information to pass through. The bottleneck can also be seen as a compressed representation of the input data in higher dimensional space. Finally, the decoder serves as a decompressor to reconstruct the bottleneck's output.

Inspired by the work of Leite et al. [34], we designed a Denoised AutoEncoder (DAE) to remove artifacts from EEG data and compared the results with the cleaned EEG achieved by the AR-WGAN. We implemented 1D convolutional layers with 1D max-pooling layers in the encoder, and 1D convolutional layers followed by 1D upsampling layers in the decoder. The 1D max-pooling layers concentrate neurons and use the maximum value to represent these concentrated neurons, reducing the dimensionality after each convolutional operation. The 1D upsampling layers in the decoder restore the dimensionality to compensate for the effect caused by the 1D max-pooling. To prevent possible overfit of the discriminator, we implemented dropout regularization in every convolutional block.

3) *Wiener Filter*: The Wiener filter (WF) is a noise removal algorithm based on statistical approaches, which also served as a comparison for the method proposed by this work. The main idea of WF is to minimize the overall mean square error or the average squared error between the denoised and original signals, which means the difference between the pure and filtered signals should be minimized [47]. The wiener filter produces an estimate \hat{d} of the filtered signal, i.e., $\hat{d} = W^T Y$.

The final optimization goal can be represented by a minimum mean squared error (MMSE) as

$$\min_{\mathbf{W}} E\{\|d - \mathbf{W}^T y\|^2\}. \quad (6)$$

4) *Empirical Mode Decomposition*: The Empirical Mode Decomposition (EMD) method is a data-driven and adaptive method, which breaks down signals into several components without leaving the time domain. Unlike the Fourier transform, the EMD method does not require signals to be stationary over time. The main idea of EMD method is to decompose the nonstationary and nonlinear signals $x(t)$ into intrinsic mode functions (IMFs). Each IMF is assumed to capture a meaningful local frequency and different IMFs do not exhibit the same frequency at the same time [48]. Considering the frequency characteristics of EEG and the aim of artifact removal, in practical applications, high-frequency IMFs would be deleted, and the rest of IMFs and the residue are summed to reconstruct denoised EEG data. In this paper, we adopted the EMD method proposed by Bono et al. [49] to filter EEG and compared its performance with our proposed method.

B. Datasets

In this paper, we used a public dataset (*EEGdenoiseNet Benchmark Dataset*) and a self-collected dataset to evaluate the performance of proposed approach. The *EEGdenoiseNet* is a semi-synthetic EEG dataset released by Zhang et al [8], which contains 4514 clean EEG segments, 3400 ocular artifact segments, and 5598 muscular artifact segments. In the *EEGdenoiseNet*, the EEG, EOG, and EMG data are acquired from several public and open-access datasets, and then pre-processed including segmentation, filtering, and resampling with the frequency of 256 Hz. For the self-collected dataset, totally four healthy subjects were recruited, where 64-channel raw EEG data were acquired from three of them (*Neuroscan EEG Acquisition System, USA*), and EMG (*Acquisition System of Neuromuscular Electrophysiological Signals, NES-64B01, SIAT, CAS, China*) and EOG (*Neuroscan EEG Acquisition System, USA*) artifacts were from the rest one. Before the superimposition of artifacts into EEG, the self-collected raw EEG data were filtered, and their initial artifacts were manually removed. Health examination before experiments showed that all subjects were in a good mental state and qualified to participate in the whole experimental procedure. The experimental protocol of this study was approved by the Institutional Review Board of Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (IRB Number: SIAT-IRB-190315-H0325). All subjects agreed to participate in the study and signed informed consent permission for the publication of data for scientific and educational purposes.

The contaminated EEG can be seen as a linear combination of pure EEG and bio-signal artifacts, which can be written as

$$y = n + \lambda \cdot d, \quad (7)$$

where y denotes the contaminated EEG, n denotes the pure EEG, d denotes the ocular or myogenic artifacts, and λ is the parameter to control signal-to-noise ratio (SNR) in the mixed

signal y . The SNR can be written as

$$SNR = 10 \times \log \frac{RMS(n)}{RMS(\lambda \cdot d)}, \quad (8)$$

where RMS is the root mean square defined as

$$RMS(d) = \sqrt{\frac{1}{N} \sum_{i=1}^N d_i^2}, \quad (9)$$

where N denotes the number of data segments and d_i denotes the i th data segments.

C. Preprocessing

Considering the public dataset has different data lengths for EEG, EMG, and EOG, we set the shortest length of 3400 among the three signals as the data length for processing. The self-collected data were resampled with a frequency of 256 Hz, which was the same as the frequency of the public data, and the data length was segmented as 3400, as well.

For the self-collected datasets, the EEG from three amputees was randomly mixed to evaluate the cross-subject performance of our proposed model, and the data size was kept consistent as that of public dataset, which was 512×3400 for all the EEG, EMG, and EOG. Besides, all the public and self-collected data were normalized to avoid possible gradient exploding and vanishing during training, and EMG and EOG artifacts were artificially added to pure EEG according to (7). Then, 448 and 64 samples were set as the training and testing data, respectively, to evaluate the generalization and robustness performances of the models. To better simulate the real situation, the original and artificial data with the sample amplitude were mixed, i.e., the amplitudes of the added EOG (A_{EOG}) and EMG (A_{EMG}) artifacts were the same as that of EEG (A_{EEG}). If only one type of artifact was superimposed, the amplitude (A_{EOG} or A_{EMG}) was doubled to ensure the total amplitude (A_{total}) equals to three times of A_{EEG} .

D. Evaluation Metrics

To quantitatively assess the performance of models, some metrics commonly used in the benchmark evaluation methods were also adopted in this work, including the power ratios, relative root mean squared error in temporal ($RRMSE_t$) and spectral ($RRMSE_s$) domains, and temporal CC, which are expressed as

$$RRMSE_t = \frac{RMS(g(y) - n)}{RMS(n)}, \quad (10)$$

$$RRMSE_s = \frac{RMS(PSD(g(y)) - PSD(n))}{RMS(PSD(n))}, \quad (11)$$

$$\text{and } CC = \frac{Cov(g(y), n)}{\sqrt{var(g(y))var(n)}}, \quad (12)$$

where y is the contaminated data, n is the Initial data, $g(y)$ denotes the denoised data passed through the model, RMS is the root mean square defined as (9), PSD is the power spectral density of the input data, and var and cov denote the variance and covariance, respectively.

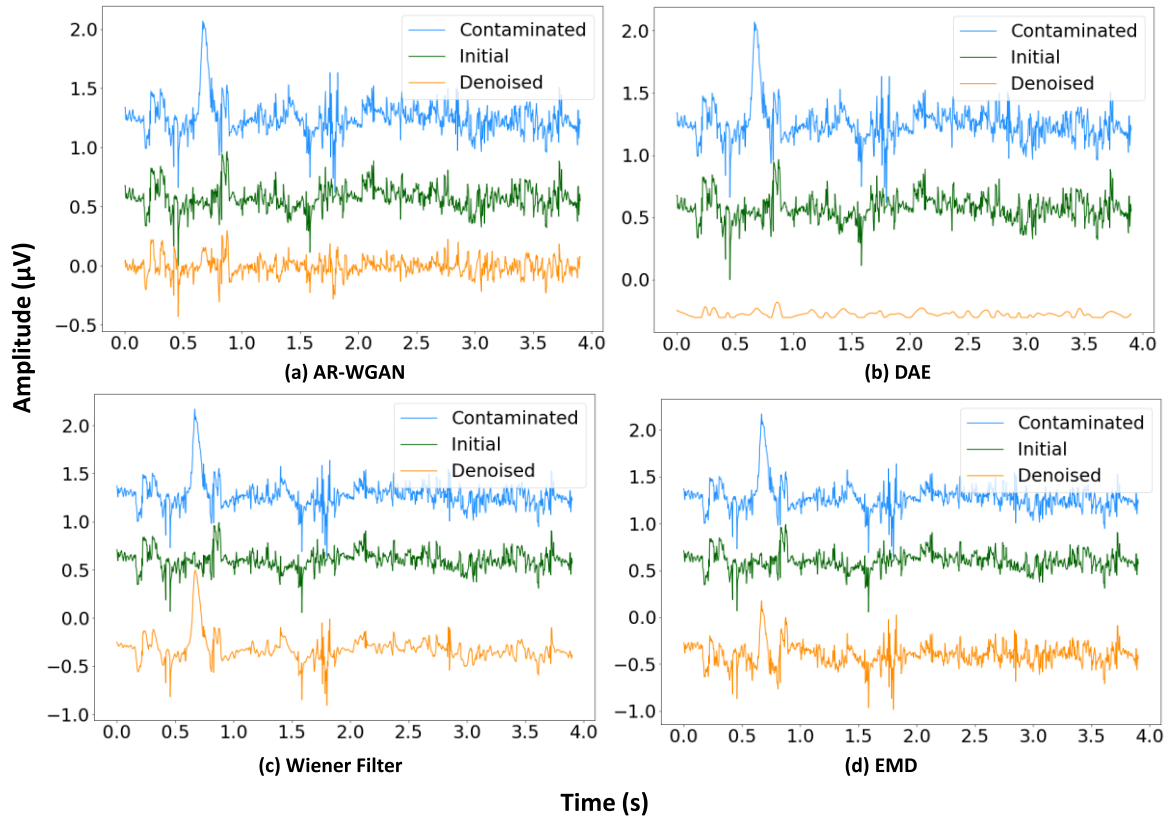


Fig. 3. Denoising performance of different methods on the public dataset in the time domain.

Algorithm 1 AR-WGAN. We Use Default Values of $n_{disc} = 1$, $\alpha = 0.001$, $\beta_1 = 0.5$, $\beta_2 = 0.999$, $c = 0.01$

Require: The number of discriminator iterations per generator iteration n_{disc} , the batch size = 5, Adam hyper-parameters α , β_1 , β_2 , the clipping parameter c .

Require: Initial discriminator parameters w_0 , initial generator parameters θ_0 .

Input: Initial data n to the discriminator, contaminated data y to the generator.

```

1: while  $\theta$  has not converged do
2:   for  $t = 1, \dots, n_{disc}$  do
3:     sample  $\{x(i)\}_{i=1}^m \sim P_n$  a batch from initial data.
4:     sample  $\{z(i)\}_{i=1}^m \sim p(z)$  a batch of prior samples.
5:      $L^{(i)} \leftarrow \nabla_w [\frac{1}{m} \sum_{i=1}^m d_w(x^{(i)}) - \frac{1}{m} \sum_{i=1}^m d_w(g_\theta(z^{(i)}))]$ 
6:      $w \leftarrow Adam(w, L^{(i)}, \alpha, \beta_1, \beta_2)$ 
7:      $w \leftarrow clip(w, -c, c)$ 
8:   end for
9:   sample a batch of latent variables  $\{z^{(i)}\}_{i=1}^m \sim p(z)$ .
10:   $\theta \leftarrow Adam(\nabla_\theta [\frac{1}{m} \sum_{i=1}^m -D_w(G_\theta(z))], \theta, \alpha, \beta_1, \beta_2)$ 
11: end while

```

III. RESULTS

Figs. 3 and 4 illustrate a comparison of denoising performance among different methods on the public and self-collected datasets in the time domain, respectively, where “initial” denotes the original data from either the public database or the self-collected signals, “contaminated” means the data that were firstly denoised and then manually superimposed with EMG and EOG artifacts, and “denoised” represents the final data after denoising. It should be noted that for clearer visualization, the offsets were artificially introduced

into the denoised data, and thus the amplitudes are changed and do not reflect true values. As can be seen, the proposed AR-WGAN shows better denoising performance for both public and self-collected data than other three methods, which may properly remove the artifacts, and therefore the output can perfectly match the initial data. Even though the other three methods can also filter the artifacts, some important and detailed time domain features are not preserved. For instance, some spikes contained in the initial data are abandoned and the denoised data are quite smooth. In special, DAE fails to preserve most features and the output is almost untrustworthy, even as reported its loss curve shows that the model is converged; WF just smoothens the contaminated data and some obvious artifact spikes still exist; although EMD performs a little bit better than DAE and WF, some obvious artifacts are still contained in the denoised data.

Figs. 5 and 6 show the time-frequency analysis of denoising performance for different methods on the public and self-collected datasets, respectively, which could reflect energy changes after the removal of artifacts. By adding the artifacts artificially, the energy of contaminated data is much higher than that of initial data. As shown in the figures, AR-WGAN can effectively eliminate the artifacts and reduce the energy in high-frequency bands that contain most artifacts; the DAE and WF processed data lose much information in both time and frequency domains and could not be used in the following analysis; the data processed by EMD are much better in the time domain compared with those by DAE and WF, however, they are still noisier in contrast with the output by AR-WGAN.

TABLE III
THE POWER RATIOS IN DIFFERENT EEG FREQUENCY BANDS OF DIFFERENT METHODS ON THE PUBLIC DATASET

Denoising methods (under -4dB)	Delta	Theta	Alpha	Beta	Gamma
AR-WGAN	0.199	0.211	0.233	0.186	0.171
DAE	0.649	0.303	0.004	0.006	0
Wiener filter	0.724	0.100	0.006	0.006	0.005
EMD	0.538	0.116	0.008	0.115	0.151
Initial data	0.200	0.165	0.130	0.177	0.205
Contaminated data	0.656	0.009	0.059	0.084	0.114

TABLE IV
THE POWER RATIOS IN DIFFERENT EEG FREQUENCY BANDS OF DIFFERENT METHODS ON THE SELF-COLLECTED DATASET

Denoising methods (under -5dB)	Delta	Theta	Alpha	Beta	Gamma
AR-WGAN	0.124	0.402	0.323	0.112	0.033
DAE	0.493	0.439	0.059	0.008	0
Wiener filter	0.365	0.244	0.142	0.133	0.116
EMD	0.272	0.206	0.145	0.1255	0.122
Initial data	0.200	0.306	0.172	0.153	0.077
Contaminated data	0.201	0.138	0.089	0.151	0.420

Figs. 7 and 8 summarize the calculated temporal RRMSE, spectral RRMSE, and temporal CC to visualize the performance of different methods on the public and self-collected datasets, respectively. In general, AR-WGAN exhibits the best performance across all the methods. For the public dataset, AR-WGAN has the lowest temporal and spectral RRMSE and highest CC with limited standard deviation. For the self-collected dataset, AR-WGAN still owns the lowest temporal and spectral RRMSE. Although it shows similar CC as DAE and EMD, its standard deviation is less.

TABLES III and IV summarize the power ratios in different EEG frequency bands of different methods on the public and self-collected datasets, respectively, where EEG signals are divided into five frequency bands, i.e., delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-80 Hz). The mean SNRs across all the contaminated data from the public and self-collected datasets were calculated as -4 and -5 dB, respectively, and power ratios are used to evaluate the signal reconstruction quality under these computed mean SNRs. As can be observed from the tables, generally, the signals denoised by AR-WGAN in different frequency bands have power ratios relatively closer to the initial data, especially in the frequency bands of delta and theta which we are normally interested in. This result means the proposed method can reconstruct signal powers very well, while other methods cannot properly filter the contaminated data and alter the power ratios of processed data.

IV. DISCUSSION

EEG provides a non-invasive way to reflect brain waves and activities with high fidelity, which can be used in BCI for real-time control of external devices and interaction with

environments. However, EEG signals recorded on the scalp are inevitably contaminated by some other bio-signals such as EMG and EOG, and most of the existing filtering methods are usually time- and labor-consuming. Many traditional ML-based denoising methods have been studied and implemented, like the Wiener Filter, EMD-based methods, and ICA-based methods. These approaches need to set pre-defined parameters, and visual inspection and manual operations are required, which is time-consuming, laborious, subjective, and inappropriate for online and real-time implementations. Besides, the experiments in this work showed that transient, large-amplitude artifacts cannot be fully removed by these traditional methods. Many works tried to fix these problems with DL-based models, while these models are not robust enough when encountering small perturbations. Thus, an automatic, fast, and robust filtering algorithm that can support massive data processing is always needed. In this work, we proposed an artifact removal method called AR-WGAN, which can effectively remove artifacts and extract useful EEG features, and its denoising performance has been evaluated with comparisons to some commonly used methods including DAE, WF, and EMD. We used a dataset from the public databank and a dataset from our lab, to confirm the performance of proposed model. In addition, to simulate the most realistic scenarios, both EMG and EOG artifacts were mixed into EEG for signal denoising.

It is focused on how much information of initial data could be retained in the denoised signal, and hence in this work we analyzed the data in both time and frequency domains and calculated RRMSE and CC to evaluate the performance of different denoising methods, where RRMSE is used to measure the degree of information retention of initial data and CC reflects the similarity between the denoised and initial

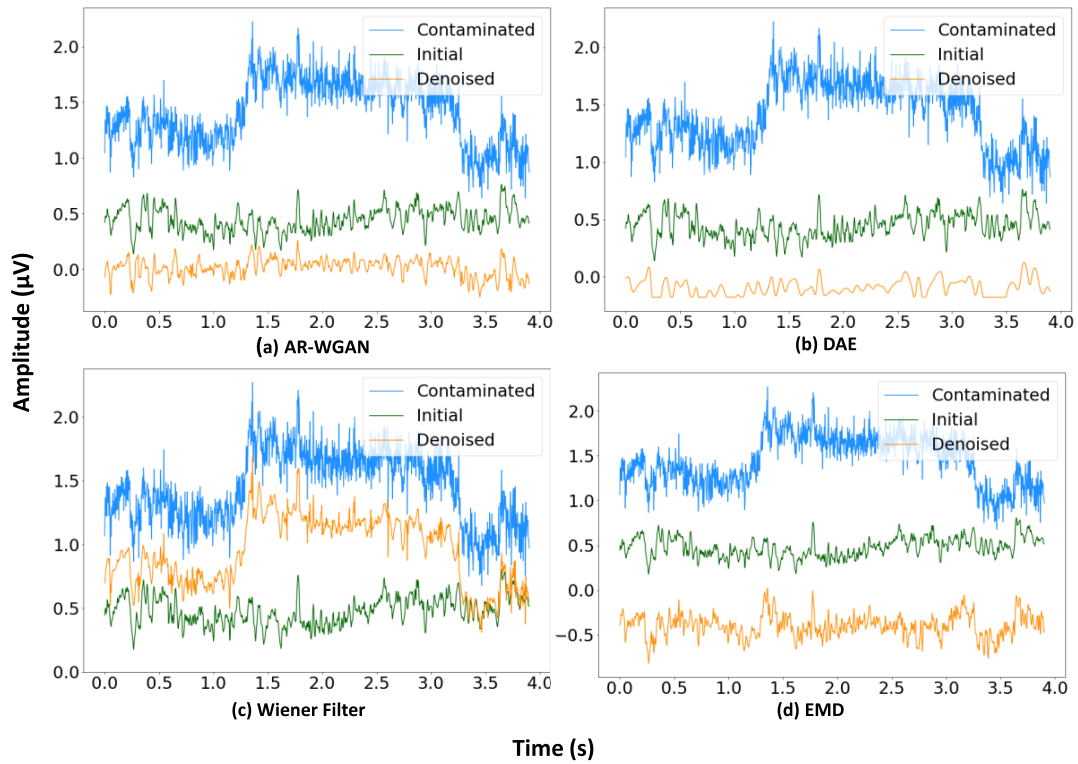


Fig. 4. Denoising performance of different methods on the self-collected dataset in the time domain.

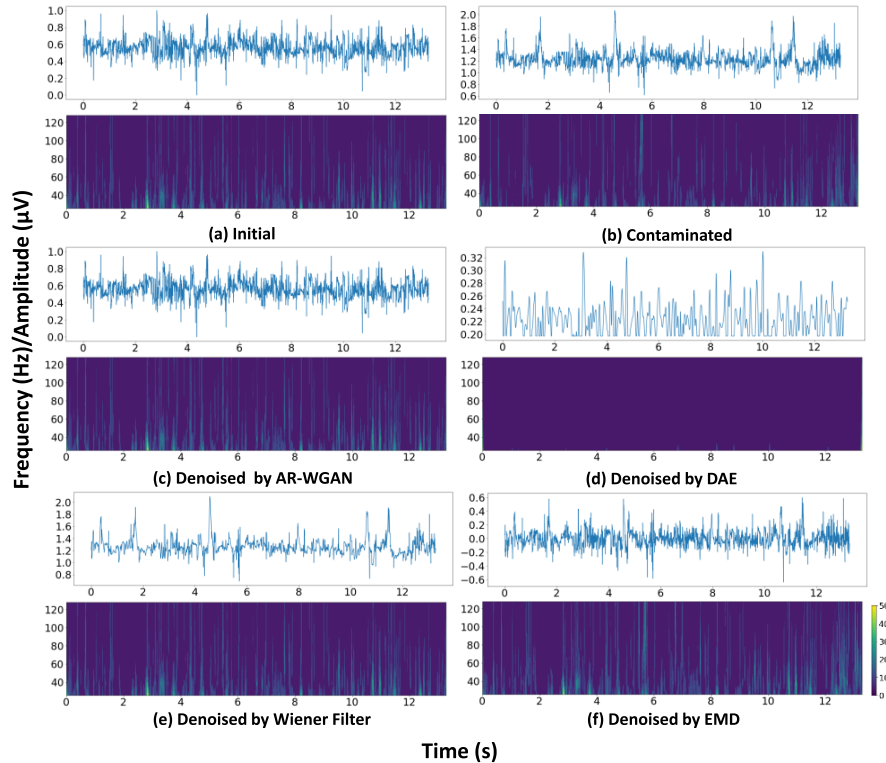


Fig. 5. Time-frequency analysis of denoising performance for different methods on the public dataset.

data. The results on both public and self-collected datasets suggest that our proposed AR-WGAN has superiority over the reference methods in removing artifacts while preserving EEG features as much as possible, which may serve as a

multipurpose and universal tool for extracting EEG from mixed signals.

The well-trained generator in AR-WGAN behaves like a non-linear function that can implement the decomposition

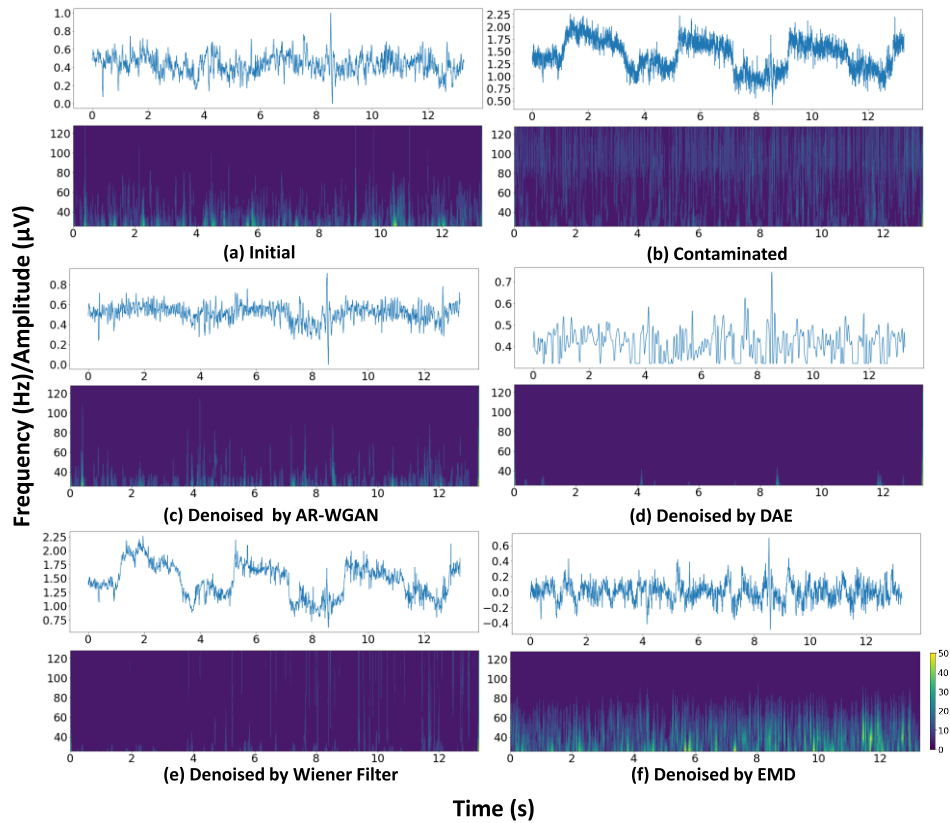


Fig. 6. Time-frequency analysis of denoising performance for different methods on the self-collected dataset.

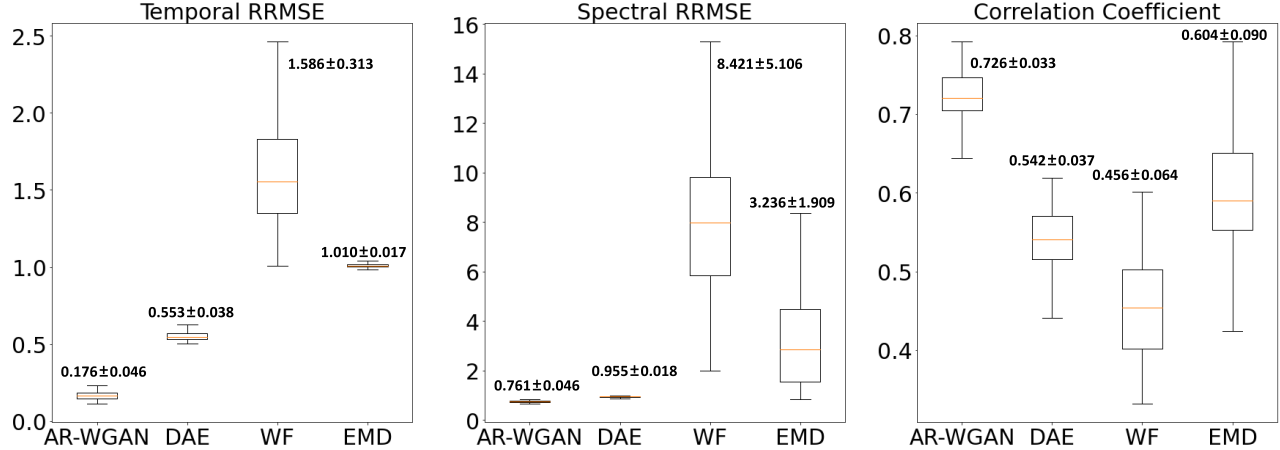


Fig. 7. The calculated temporal RRMSE, spectral RRMSE, and CC for different methods on the public dataset.

and the following reconstruction of contaminated signals. The 1D convolutional operations capture the latent features that contain artifactual components, and the 1D transposed convolutional operations reconstruct these components supervised by the discriminator, trying to minimize the difference between the original and reconstructed data. Considering that the deep convolutional layers were adopted by the model and some important features might deteriorate in the data flow, we implemented residual connections to ensure that the data quality would not be damaged before and after convolutions. We tried to go deeper with the transposed convolutional layers in both generator and discriminator, but the performance was

not significantly improved, which proves that the existing convolutional blocks are good enough to extract EEG features. Besides, we found that adding 1D max pooling layers could significantly improve the performance of generated data. The intuitive interpretation of max pooling is that the network concentrates on some particular features of initial EEG and the network parameters are decreased, and thus the computational load and probability of overfitting are reduced and the performance is improved.

Theoretically, DAE contains an encoder and decoder to decompose data, extract features, and reconstruct data, which coincides with the generator of AR-WGAN. However,

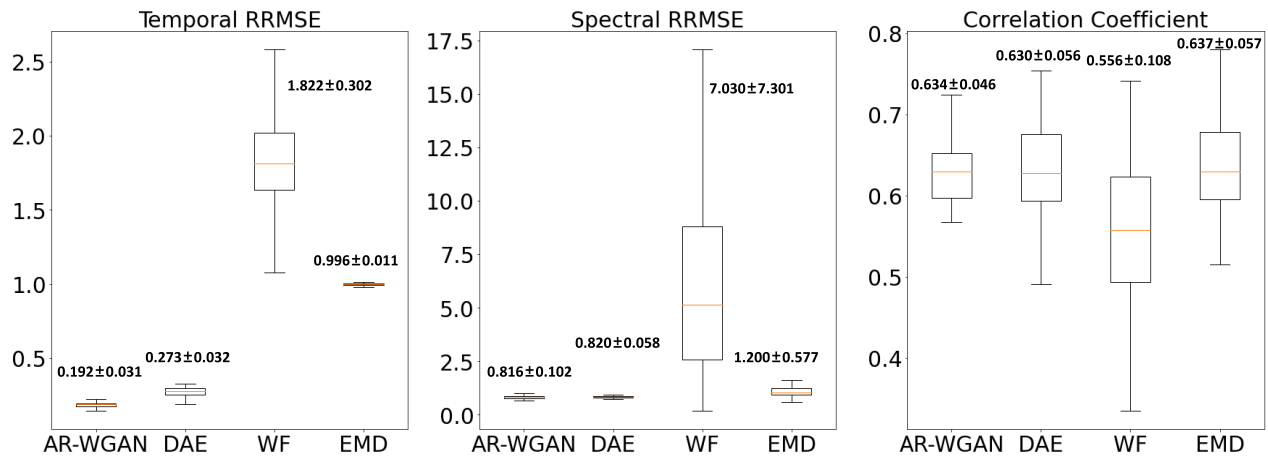


Fig. 8. The calculated temporal RRMSE, spectral RRMSE, and CC for different methods on the self-collected dataset.

DAE cannot learn to efficiently generalize the model and extract features without adversarial training, and the upsampling operation may not accurately reconstruct data, with low-quality output. WF aims to minimize the mean squared error between the expected and initial signals, which can preserve overall trends and smoothen signals in the time domain, but important features are usually lost and large artifact drifts are not removed. EMD can decompose data into several IMFs and a residual component, but it is also sensitive to noise and problematic in mode mixing.

Compared with other commonly used artifact removal algorithms which rely on prior knowledge and manual artifactual component removal, AR-WGAN can obtain better achievements with arbitrary EMG and EOG artifacts. Once the model is well-trained, the processing speed is fast and the model can filter massive data automatically. Besides, the quality of data filtered by AR-WGAN can be observed by the discriminator, which informs the generator whether the filtered data are still noisy. The training process would be terminated when the filtered data are good enough. By comparison, other DL-based filtering methods simply remove artifacts and the filtering model cannot be judged directly. Moreover, the results from the self-collected dataset show that AR-WGAN can cope with the data from different subjects and perform pretty well, meaning that the model may have a certain cross-subject ability and cope with EEG data that the model never sees. Most importantly, the EEG data in real applications are much more complicated, containing some perturbations and large-amplitude artifacts caused by headset motions, which are approximately insolvable by simple DL models but capable of AR-WGAN.

Even though the AR-WGAN shows great effectiveness in removing large-amplitude artifacts and handling with inevitable perturbations, it is worth noting that there exist several improvements for the proposed AR-WGAN method. The model tends to excessively lower the energy in the low-frequency band, despite it has strong capability to remove EMG and EOG artifacts that are mostly in high- and low-frequency bands, respectively. However, the initial EEG data still contain some energy in the band of 0-20 Hz, and it is

challenging for the current model to accurately remove extra noise without changing the original energy, which is a common issue as reported in other work [34]. Besides, this work only focuses on the information in time and frequency domains, while the spatial information between EEG channels should be considered in future work. Additionally, the proposed method still needs some manually-processed data to train the model. In applications, it needs to implement some classical methods like independent component analysis (ICA) to process EEG data in advance, and then input purified and raw data to the discriminator and generator, respectively, to train the model. Therefore, it is suggested to design an unsupervised filtering network that can improve itself during training without annotated data.

V. CONCLUSION

In this work, we proposed a denoising network called Artifact Removal Wasserstein Generative Adversarial Network (AR-WGAN) and compared its performance with some other commonly used denoising methods based on two datasets. The experimental results demonstrated that the proposed method has superiority over the conference methods on EEG denoising. Additionally, this framework can extract EEG features from raw data of any number of channels and also be extended to other signal-filtering tasks.

REFERENCES

- [1] S. Vaid, P. Singh, and C. Kaur, "EEG signal analysis for BCI interface: A review," in *Proc. 5th Int. Conf. Adv. Comput. Commun. Technol.*, Haryana, India, Feb. 2015, pp. 143–147.
- [2] M. Hamedi, S.-H. Salleh, and A. M. Noor, "Electroencephalographic motor imagery brain connectivity analysis for BCI: A review," *Neural Comput.*, vol. 28, no. 6, pp. 999–1041, Jun. 2016.
- [3] Z. Feng et al., "Design a novel BCI for neurorehabilitation using concurrent LFP and EEG features: A case study," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 5, pp. 1554–1563, May 2022.
- [4] A. Q. Hamal and A. W. B. A. Rehman, "Artifact processing of epileptic EEG signals: An overview of different types of artifacts," in *Proc. Int. Conf. Adv. Comput. Sci. Appl. Technol.*, Kuching, Malaysia, Dec. 2013, pp. 358–361.
- [5] D. Gorjan, K. Gramann, K. De Pauw, and U. Marusic, "Removal of movement-induced EEG artifacts: Current state of the art and guidelines," *J. Neural Eng.*, vol. 19, no. 1, Feb. 2022, Art. no. 011004.

- [6] K. Cao, Y. Guo, and S. W. Su, "A review of motion related EEG artifact removal techniques," in *Proc. 9th Int. Conf. Sens. Technol. (ICST)*, Auckland, New Zealand, Dec. 2015, pp. 600–604.
- [7] J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal-state-of-the-art and guidelines," *J. Neural Eng.*, vol. 19, no. 3, 2015, Art. no. 031001.
- [8] T.-P. Jung et al., "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, Mar. 2000.
- [9] X. Chen, X. Xu, A. Liu, M. J. McKeown, and Z. J. Wang, "The use of multivariate EMD and CCA for denoising muscle artifacts from few-channel EEG recordings," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 359–370, Feb. 2018.
- [10] M. A. Klados, C. Papadelis, C. Braun, and P. D. Bamidis, "REG-ICA: A hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts," *Biomed. Signal Process. Control*, vol. 6, no. 3, pp. 291–300, Jul. 2011.
- [11] P. He, M. Kahle, G. Wilson, and C. Russell, "Removal of ocular artifacts from EEG: A comparison of adaptive filtering method and regression method using simulated data," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, Shanghai, China, 2005, pp. 1110–1113.
- [12] T. Radüntz, J. Scouten, O. Hochmuth, and B. Meffert, "Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features," *J. Neural Eng.*, vol. 14, no. 4, Aug. 2017, Art. no. 046004.
- [13] L. Frölich, T. S. Andersen, and M. Mørup, "Classification of independent components of EEG into multiple artifact classes," *Psychophysiology*, vol. 52, no. 1, pp. 32–45, Jan. 2015.
- [14] I. Winkler, S. Brandl, F. Horn, E. Waldburger, C. Allefeld, and M. Tangermann, "Robust artifactual independent component classification for BCI practitioners," *J. Neural Eng.*, vol. 11, no. 3, Jun. 2014, Art. no. 035013.
- [15] N. Bigdely-Shamlo, K. Kreutz-Delgado, C. Kothe, and S. Makeig, "EyeCatch: Data-mining over half a million EEG independent components to construct a fully-automated eye-component detector," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 5845–5848.
- [16] C.-Y. Chang, S.-H. Hsu, L. Pion-Tonachini, and T.-P. Jung, "Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 4, pp. 1114–1121, Apr. 2020.
- [17] S. Blum, N. S. J. Jacobsen, M. G. Bleichner, and S. Debener, "A Riemannian modification of artifact subspace reconstruction for EEG artifact handling," *Frontiers Hum. Neurosci.*, vol. 13, p. 141, Apr. 2019.
- [18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, vol. 1, 2012, pp. 84–90.
- [20] O. P. Idowu, P. Fang, X. Li, Z. Xia, J. Xiong, and G. Li, "Towards control of EEG-based robotic arm using deep learning via stacked sparse autoencoder," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Kuala Lumpur, Malaysia, Dec. 2018, pp. 1053–1057.
- [21] O. P. Idowu et al., "Neuro-evolutionary approach for optimal selection of EEG channels in motor imagery based BCI application," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102621.
- [22] O. P. Idowu, A. E. Ilesanmi, X. Li, O. W. Samuel, P. Fang, and G. Li, "An integrated deep learning model for motor intention recognition of multi-class EEG signals in upper limb amputees," *Comput. Methods Programs Biomed.*, vol. 206, Jul. 2021, Art. no. 106121.
- [23] M. K. M. Rabby, R. B. Eshun, S. Belkasim, and A. K. M. K. Islam, "Time-frequency based EEG motor imagery signal classification with deep learning networks," in *Proc. IEEE 4th Int. Conf. Artif. Intell. Knowl. Eng. (AIKE)*, Laguna Hills, CA, USA, Dec. 2021, pp. 133–134.
- [24] X. Li et al., "EEG motor imagery classification based on multi-spatial convolutional neural network," in *Proc. 5th Int. Conf. Artif. Intell. Big Data (ICAIBD)*, Chengdu, China, May 2022, pp. 433–437.
- [25] H. K. Lee and Y.-S. Choi, "A convolution neural networks scheme for classification of motor imagery EEG based on wavelet time-frequency image," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Chiang Mai, Thailand, Jan. 2018, pp. 906–909.
- [26] Q. Li et al., "EEG signal processing and emotion recognition using convolutional neural network," in *Proc. Int. Conf. Electron. Inf. Eng. Comput. Sci. (EIECS)*, Changchun, China, Sep. 2021, pp. 81–84.
- [27] Y. Dai, E. Ji, and Y. Yao, "Stacked auto-encoder driven automatic feature extraction for web-enabled EEG emotion recognition," in *Proc. 7th Annu. Int. Conf. Netw. Syst. Comput. (ICNISC)*, Guiyang, China, Jul. 2021, pp. 991–997.
- [28] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, and H. He, "Multisource transfer learning for cross-subject EEG emotion recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3281–3293, Jul. 2020.
- [29] Y. Song, L. Yang, X. Jia, and L. Xie, "Common spatial generative adversarial networks based EEG data augmentation for cross-subject brain-computer interface," 2021, *arXiv:2102.04456*.
- [30] R. Zhang et al., "ERP-WGAN: A data augmentation method for EEG single-trial detection," *J. Neurosci. Methods*, vol. 376, Jul. 2022, Art. no. 109621.
- [31] H. Zhang, M. Zhao, C. Wei, D. Mantini, Z. Li, and Q. Liu, "EEG-denoiseNet: A benchmark dataset for deep learning solutions of EEG denoising," *J. Neural Eng.*, vol. 18, no. 5, Oct. 2021, Art. no. 056057.
- [32] W. Sun, Y. Su, X. Wu, and X. Wu, "A novel end-to-end 1D-ResCNN model to remove artifact from EEG signals," *Neurocomputing*, vol. 404, pp. 108–121, Sep. 2020.
- [33] C.-H. Chuang, K.-Y. Chang, C.-S. Huang, and T.-P. Jung, "IC-U-Net: A U-Net-based denoising autoencoder using mixtures of independent components for automatic EEG artifact removal," *NeuroImage*, vol. 263, Nov. 2022, Art. no. 119586.
- [34] N. M. N. Leite, E. T. Pereira, E. C. Gurjão, and L. R. Veloso, "Deep convolutional autoencoder for EEG noise filtering," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Madrid, Spain, Dec. 2018, pp. 2605–2612.
- [35] Z. Liu, L. Meng, X. Zhang, W. Fang, and D. Wu, "Universal adversarial perturbations for CNN classifiers in EEG-based BCIs," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 0460a4.
- [36] X. Zhang and D. Wu, "On the vulnerability of CNN classifiers in EEG-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 814–825, May 2019.
- [37] T. Bai, J. Luo, J. Zhao, B. Wen, and Q. Wang, "Recent advances in adversarial training for adversarial robustness," 2021, *arXiv:2102.01356*.
- [38] I. Goodfellow et al., "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, Montreal, QC, Canada, 2014, pp. 2672–2680.
- [39] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Juan, Puerto Rico, 2016, pp. 1–6.
- [40] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
- [41] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017, *arXiv:1703.10593*.
- [42] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*.
- [43] E. Brophy, Z. Wang, Q. She, and T. Ward, "Generative adversarial networks in time series: A systematic literature review," *ACM Comput. Surv.*, vol. 55, no. 10, pp. 1–31, Oct. 2023.
- [44] E. Brophy, P. Redmond, A. Fleury, M. De Vos, G. Boylan, and T. Ward, "Denoising EEG signals for real-world BCI applications using GANs," *Frontiers Neuroergonomics*, vol. 2, Jan. 2022, Art. no. 805573.
- [45] Y. An, H. K. Lam, and S. H. Ling, "Auto-denoising for EEG signals using generative adversarial network," *Sensors*, vol. 22, no. 5, p. 1750, Feb. 2022.
- [46] Y. Sumiya, K. Horie, H. Shiokawa, and H. Kitagawa, "NR-GAN: Noise reduction GAN for mice electroencephalogram signals," in *Proc. 4th Int. Conf. Biomed. Imag., Signal Process.*, Oct. 2019, pp. 94–101.
- [47] B. Somers, T. Francart, and A. Bertrand, "A generic EEG artifact removal algorithm based on the multi-channel Wiener filter," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 036007.
- [48] V. Bajaj and R. B. Pachori, "Classification of seizure and nonseizure EEG signals using empirical mode decomposition," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1135–1142, Nov. 2012.
- [49] V. Bono, S. Das, W. Jamal, and K. Maharatna, "Hybrid wavelet and EMD/ICA approach for artifact suppression in pervasive EEG," *J. Neurosci. Methods*, vol. 267, pp. 89–107, Jul. 2016.