

EEGdenoiseNet: A benchmark dataset for deep learning solutions of EEG denoising

Haoming Zhang^{1,†}, Mingqi Zhao^{1,2,†}, Chen Wei¹,
Dante Mantini^{2,3}, Zherui Li¹, Quanying Liu^{1,*}

September 29, 2020

¹ Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen 518055, P.R. China

² Movement Control and Neuroplasticity Research Group, KU Leuven, Leuven 3001, Belgium

³ Brain Imaging and Neural Dynamics Research Group, IRCCS San Camillo Hospital, Venice 30126, Italy

† Equal contribution

* Corresponding author liuqy@sustech.edu.cn to Q.L.

Abstract

Deep learning networks are increasingly attracting attention in several fields. Among other applications, deep learning models have been used for denoising of electroencephalography (EEG) data. These models provided comparable performance with that of traditional techniques. At present, however, the lack of well-structured, standardized datasets with specific benchmark limits the development of deep learning solutions for EEG denoising. Here, we present EEGdenoiseNet, a benchmark EEG dataset that is suited for training and testing deep learning-based denoising models. This permits making valid comparisons across different models. EEGdenoiseNet contains 4514 clean EEG epochs, 3400 ocular artifact epochs and 5598 muscular artifact epochs. This allow the user to produce a large number of noisy EEG epochs with ground truth for model training and testing. We used EEGdenoiseNet to evaluate the performance of four classical deep learning networks (a fully-connected network, a simple convolution network, a complex convolution network and a recurrent neural network). Our analysis suggested that deep learning methods have great potential for EEG denoising even under high noise contamination. Through EEGdenoiseNet, we hope to accelerate the development of the emerging field of deep learning-based EEG denoising. The dataset and the code for benchmarking deep learning networks are publicly available on github (<https://github.com/ncclabsustech/EEGdenoiseNet>).

Keywords: Deep learning network, Electroencephalography dataset, Benchmark dataset, EEG artifact removal, EEG denoising

1 Introduction

Electroencephalography (EEG) solutions permit sensing variations of electrical potential over the scalp, which are generated by neurons in the gray matter. EEG has proven to be one of the most important direct and noninvasive approaches to study brain activity during task and rest conditions. It is nowadays widely used in psychological, neurological and psychiatric studies, as well as for brain-computer interface research [1, 2, 3, 4, 5, 6].

EEG signals contain not only brain activity, but also several kinds of noise and artefacts, including ocular [7], myogenic artefacts [8, 9], and in rare cases cardiac artefacts. Accordingly, a fundamental step to enable the study of neural activity using EEG data is denoising, or artifact attenuation [10]. Ocular and myogenic artefacts contaminate EEG signals in different ways: the former are often visible as pulses with relatively large amplitude over frontal regions [11], whereas the latter frequently appear over temporal and occipital regions, and have broad spectral content [9, 12].

Various traditional denoising techniques have been developed for the removal of artefacts from EEG data. Among them, regression-based methods cancel the artifacts by subtracting estimated noise template signals from EEG data [12, 13, 14, 15]. Adaptive filter-based approaches rely on the dynamic estimation of filtering coefficients according to the input signal [16, 17]. Blind source separation (BSS) methods decompose EEG signals into components [18, 19, 20], assign them to neural and artifactual sources, and reconstruct the clean signal by recombining the neural components only [9, 12, 21]. BSS methods, however, can only be used when a large number of electrodes is available.

Deep learning (DL) networks [22, 23, 24, 25] have attracted increasingly attention in the last years. Thanks to increasing computing resources, boosting data size, and the availability of new network architectures and learning algorithms, DL networks could be successfully applied to address various technological problems, as for instance image [22, 23, 26, 27] and natural language processing [24, 25, 28]. More recently, DL methods have been used for EEG signal analysis [29], such as classification [30, 31, 32], reconstruction [33, 34] and generation [35, 36]. They have also been applied to EEG denoising, providing comparable performance with that of traditional denoising techniques [37, 38, 39]. Deep neural networks can learn the hidden nature of neural oscillations in EEG, and therefore remove the fluctuations that are not sourced from authentic neural activities but from biological artefacts. The performance of deep neural networks relies fundamentally on the size of training and testing datasets, or, in other words, require big data [40, 41, 42]. The availability of EEG datasets for benchmark DL-based denoising methods would be crucial to assess newly developed DL algorithms as compared to existing approaches. To the best of our knowledge, very few studies have made publicly available raw datasets that may be suitable for deep learning-based EEG denoising studies. Those datasets were collected in participants during resting state [43, 44], or during performance of cognitive tasks [45, 46, 47], motor-related tasks [48, 49, 50, 51]. However, the lack of preprocessed EEG data and of benchmark algorithms has largely limited the development of DL methods for EEG denoising.

In this study, we present a publicly-available structured dataset, named EEGdenoiseNet [52], that is particularly suitable for deep network-based EEG noise attenuation studies. Specifically,

the dataset contains 4514 trials of clean EEG epochs, 3400 trials of ocular artifact epochs, and 5598 trials of myogenic artifact epochs. Each trial was extracted from signals in online public repositories, and had a duration of 2 seconds. In addition, we also included four typical deep neural networks to be used as benchmark: a fully-connected neural network (FCNN), a simple convolution neural network (CNN), a complex CNN, and a recurrent neural network (RNN). In summary, EEGdenoiseNet can be used to easily generate simulated signals with specific properties for training and testing DL-based denoising models, without requiring extensive knowledge on EEG data collection and processing.

2 EEGdenoiseNet Dataset

2.1 Data acquisition and preprocessing

Our main goal is to provide a dataset that is suited for deep learning network-based EEG denoising study. In this respect, we acquired EEG, EOG and EMG data from several publicly available data repositories [53, 54, 55, 56, 57, 58, 59, 60]. Then, we preprocessed the data, and segmented them into epochs of 2 seconds; afterwards, we scaled the epochs to the same variance and visually checked them to ensure they were clean and ready for use; lastly, we uploaded the epochs of each category to a publicly available repository named EEGdenoiseNet . All the data used in this study were acquired from open access archives, and were used in previous studies. Those studies were ethically approved by their respective local ethical committees, and followed the Helsinki Declaration of 1975, revised in 2000.

For the EEG epochs (Figure 1a), we used a EEG datasets for motor imaginary brain computer interface [53]. The EEG dataset contained 64-channel (with international 10-10 system, and sampled at 512 Hz) EEG both for imaginary and real movement conditions of left and right hand. The data were band-pass filtered between 1 to 80 Hz and notched at powerline frequency, followed by re-sampling to 256 Hz; then the 64-channel signals were processed using ICLabel [9] to attenuate artefacts. Then the EEG signals were segmented into epochs of 2 seconds.

For the ocular artifact epochs (Figure 1b), multiple dataset were collected from open-access EEG data [54, 55, 56, 57, 58, 59]. The horizontal and vertical electroculagraphy (EOG) signals of the datasets were band-pass filtered between 0.3 and 10 Hz, followed by re-sampling to 256 Hz; then the signals were segmented into epochs of 2 seconds.

For the myogenic artifact epochs (Figure 1c), a facial surface electromyography (EMG) dataset was used [60]. We choose facial EMG because they are the main sources of myogenic artefacts. The EMG signals were band-pass filtered between 1 to 120 Hz and notched at powerline frequency, followed by re-sampling to 512 Hz; then the signals were segmented into epochs of 2 seconds.

For all the categories, the epochs were standardized by subtracting their mean and dividing by their standard deviation, and then were visually checked by an expert. We finally acquired 4514 EEG epochs, 3400 ocular artifact epochs and 5598 myogenic artifact epochs. The epochs for each category were saved separately as Matlab matrix files and Python numpy matrix

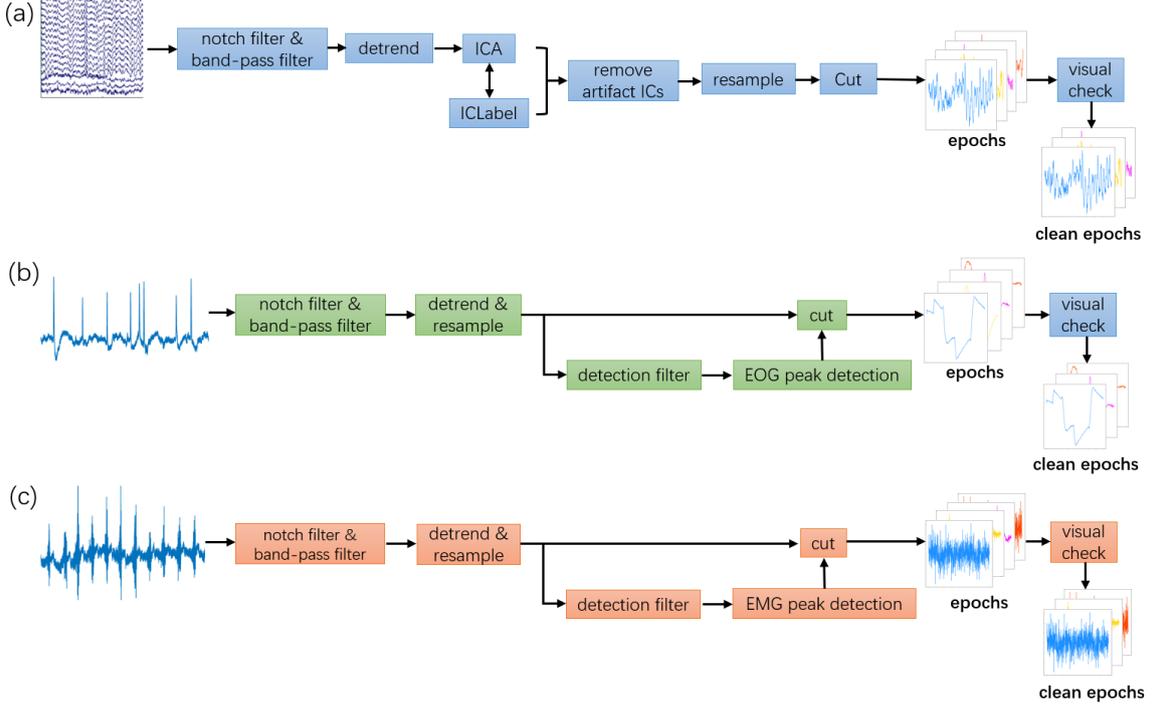


Figure 1: The pipeline for obtaining clean EEG, EOG and MEG.

files to a public data repository. Figure 2 displays a example of clean EEG, horizontal EOG, vertical EOG and EMG.

2.2 Data Usage

Simulated noisy signals can be generated by linearly mixing the EEG epochs with EOG or EMG epochs according to formula 1 (see Figure 4):

$$y = x + \lambda \cdot n \quad (1)$$

where y denotes the mixed signal of EEG and artifacts (ocular or myogenic), x denotes a clean EEG signal, n denotes artifacts, and λ denotes the relative artifact's contribution. In this way, a signal-to-noise ratio (SNR) of the noisy epoch can be adjusted by changing the parameter λ according to formula 2:

$$SNR = 10 \log \frac{RMS(x)}{RMS(\lambda \cdot n)} \quad (2)$$

in which the Root Mean Squared (RMS) value is defined as formula 3:

$$RMS(g) = \sqrt{\frac{1}{N} \sum_{i=1}^N \|g_i\|_2^2} \quad (3)$$

where N denotes the number of samples of an epoch g , and g_i denotes the i^{th} sample of a

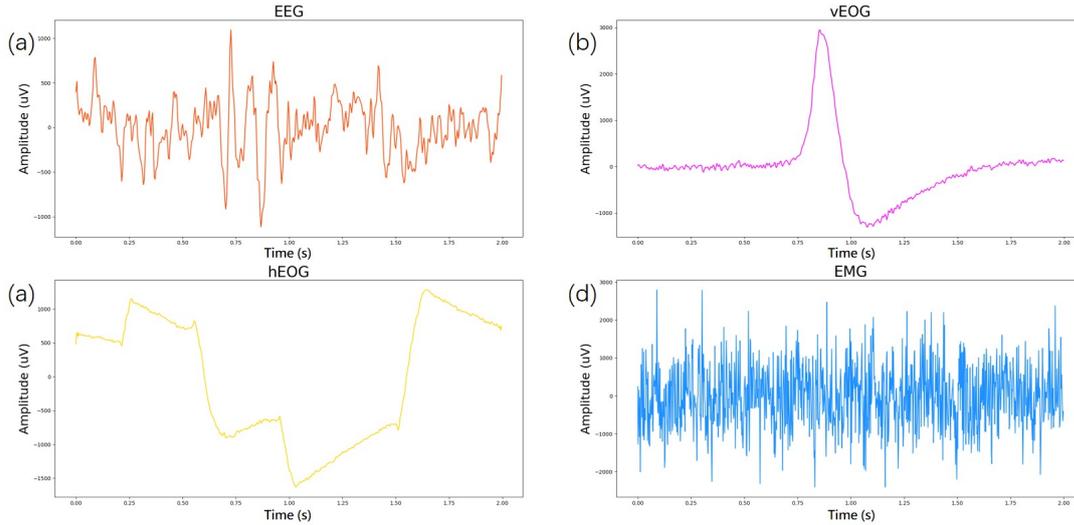


Figure 2: Examples of epochs in EEGdenoiseNet dataset. (a) An EEG epoch. (b) A vertical EOG (vEOG) epoch. (c) A horizontal EOG (hEOG) epoch. (d) An EMG epoch.

epoch g . According to previous studies, the SNR for EEG epochs contaminated by ocular artifacts should range from -7dB to 2dB[61], and the SNR for those contaminated by myogenic artifacts should range from -7dB to 4dB [62, 63]. In this way, the clean EEG epochs can be considered as ground truth, and the mixed epochs as contaminated EEG.

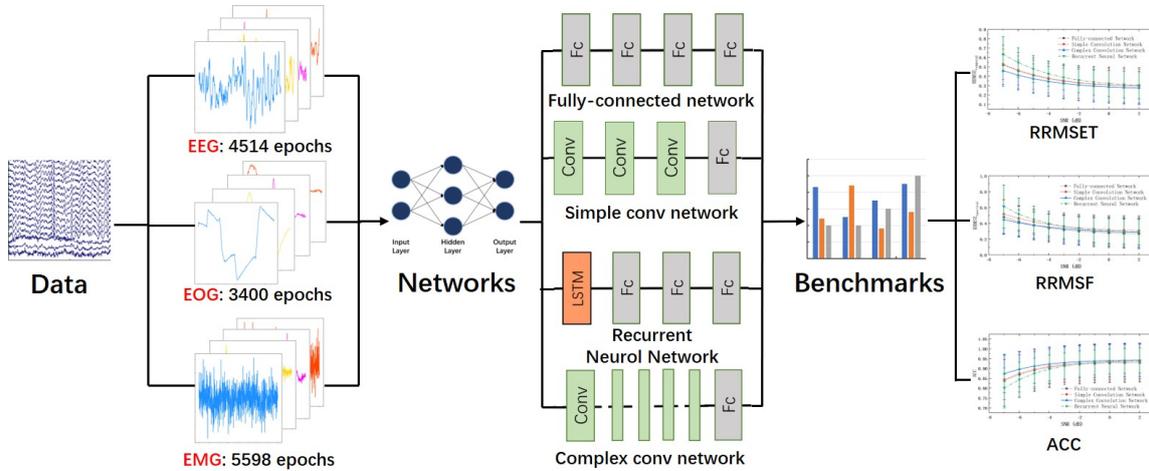


Figure 3: The structure of EEGdenoiseNet

3 Benchmarking deep learning algorithms

Another goal of this study is to provide a set of benchmark algorithms together with the dataset. In this respect, we first prepared two sets of simulated noisy EEG signals contaminated by ocular artifacts and myogenic artifacts individually for training and testing the four aforementioned networks, then evaluated the the networks and defined the evaluation parameters as benchmarks.

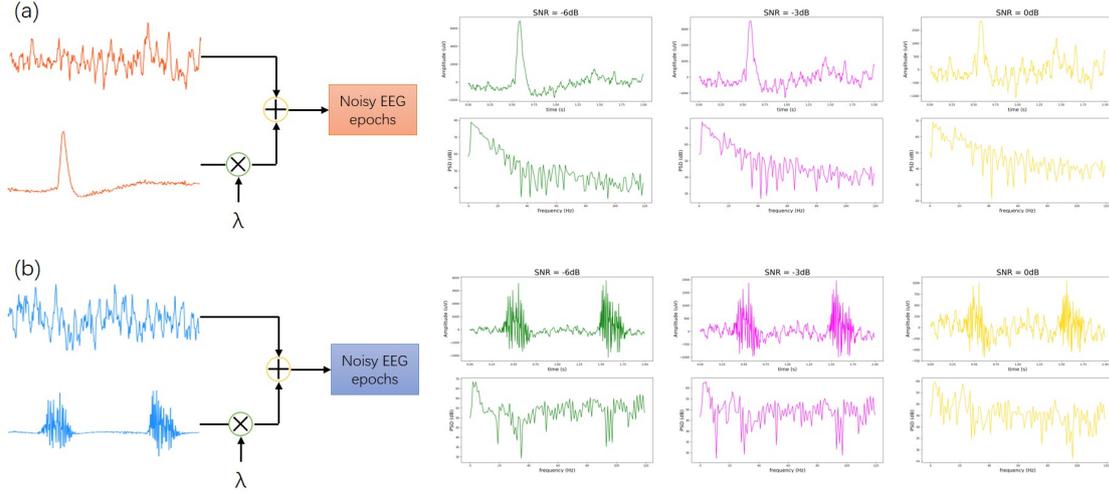


Figure 4: The pipeline for mixing noises to clean EEG signals. (a) Addition of ocular artifacts. (b) Addition of myogenic artifacts.

3.1 Preparation of simulated signals

For the ocular artifacts contaminated signals, we used 3400 pairs of EEG and EOG epochs in EEGdeoiseNet, in which 3000 pairs for generating the training set, and 400 pairs for generating the test set. For the training set, we randomly combined 3000 pairs of EEG and ocular artifact data ten times according to section 2.2 with the value of the SNR followed a uniform distribution from -7dB to 2dB. In this way, for each of the four networks, the training set expanded to 30000 epochs. Likewise, for the testing set, we also randomly combined 400 pair of epochs with ten different SNR levels (-7dB, -6dB, -5dB, -4dB, -3dB, -2dB, -1dB, 0dB, 1dB, 2dB), and expanded the testing set to 4000 epochs.

For the myogenic artifacts contaminated signals, we used 4514 EEG epochs and 5598 myogenic artifact epochs from our dataset. To match the sampling frequency between EEG epochs and EMG epochs, we upsampled the EEG epochs to 512Hz. To match the number of EEG epochs with EMG, we randomly reused some of the data to increase the number of EEG epochs to 5598 and got 5598 pairs of EEG and myogenic artifact epochs. A total number of 5000 pairs was used for generating the training set, and 598 pairs for generating the test set. Similar to the procedure for ocular artifact contaminated signals, we expanded the size of the training and test sets ten times larger, setting the SNR from -7 to 2 dB. We did not use 4 dB as described in previous section because we posited that the denoising performance would not show any noticeable change when SNR was larger than 2 dB.

3.2 Network architectures

3.2.1 Fully-connected Neural Network

We first trained a fully-connected network (Figure 5a) of four hidden layers with the ReLu activation function. The neurons number in each layer is equal to the sampling number of the input signal (512 for ocular artifacts reduction, 1024 for myogenic artifacts reduction).

Dropout regularization was used for reducing overfitting [64]. Noisy EEG was sent in from the first layer, and then denoised EEG was obtained from the last layer.

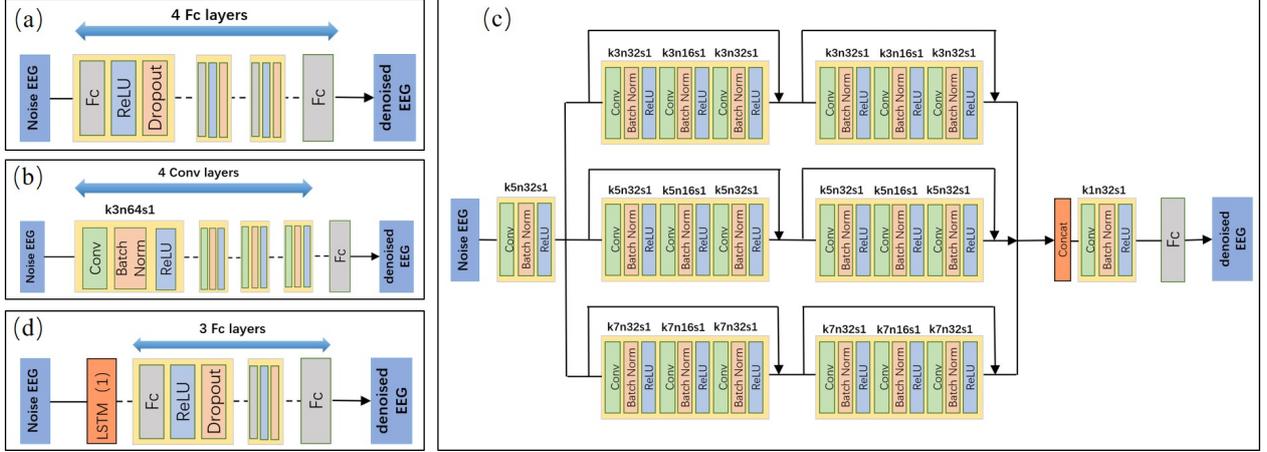


Figure 5: The structures of the four DL-based methods for benchmarking. (a)FCNN ; (b) Simple CNN; (c) Complex CNN; (d) RNN.

3.2.2 Simple Convolution Neural Network

The second Network we used was a simple convolution network (Figure 5b). The network contained four 1D-convolution layers with small 1×3 kernels, 1 stride, and 64 feature maps (k3n64s1). Each 1D-convolution layer was followed by a batch-normalization layer (BN) [65] and a ReLU activation function. To reconstruct the signal, the last convolution layer was followed by a flatten layer and a dense layer which has 512 or 1024 outputs, the same as the input.

3.2.3 Complex Convolution Neural Network

The third network for benchmark was an one-dimensional residual convolutional Neural Networks (1D-ResCNN) model adapted from [38] (Figure 5c). This model has a more complex structure than the simple convolution network. The difference is that a ResNet with skip-layer connections is introduced to avoid gradient explosion [23], so that we can train a deeper network to obtain better capability of feature extraction. To extract multi-scale features, we repeatedly stacked the residual blocks which use 1×3 , 1×5 , 1×7 convolution kernel of multiple scales twice and arranged three sets of residual blocks branches in parallel [27, 66].

3.2.4 Recurrent Neural network (Long-Short Term Memory)

A Long Short-Term Memory (LSTM) network [67](Figure 5d) was considered as the benchmark of recurrent neural network (RNN). LSTM is capable of learning long-term dependencies, which might be useful to distinguish the long-term features in noise and EEG signals. Each EEG sample was sequentially input to LSTM cells, and the output was obtained from the state of

the each cell through a fully-connected network. This RNN model were initialised to have 1 hidden states, and output network was a three-layer fully-connected network with the ReLu activation function, dropout regularization and 512 or 1024 neurons in each layer.

3.3 Learning process

The goal of a denoising network is to learn a nonlinear function f that maps noisy EEG y to pure EEG x (see formula 4):

$$f : y \rightarrow x \quad (4)$$

where $y \in \mathbb{R}^{1 \times T}$ denotes the noisy EEG signal, $x \in \mathbb{R}^{1 \times T}$ denotes the clean EEG signal. If we consider x as a sample from the clean EEG signal distribution $P(x)$ and y as a sample from the noisy EEG signal distribution $P(y)$, the denoising network can also be described as a function f that maps samples from $P(y)$ to a distribution $P(z)$, in which the difference between $P(z)$ and $P(x)$ is minimized. That means a denoising network is actually a method to move one data distribution to another distribution.

In order to facilitate the training and application of neural networks, we normalized the input noisy EEG signal and the ground-truth EEG signal by dividing the standard deviation of noisy EEG signal according to formula 5 and 6:

$$\hat{x} = \frac{x}{\sigma_y} \quad (5)$$

$$\hat{y} = \frac{y}{\sigma_y} \quad (6)$$

where σ_y is the standard deviation of noisy signal y . The standard deviation value of each noise signal was recorded, and finally the magnitude of the denoised EEG signals was restored by multiplying the denoised output by the corresponding standard deviation value.

For all the four networks, we used the mean squared error (MSE) as loss function $L_{MSE}(f)$ (see formula 7). The learning process was implemented with gradient descent to minimize the error between the denoised signal and the ground-truth clean signal.

$$L_{MSE}(f) = \frac{1}{N} \sum_{i=1}^N \left\| f_i(y) - x_i \right\|_2^2 \quad (7)$$

where N denotes the number of samples of an epoch, $f_i(y)$ denotes i^{th} sample of the output epoch of the neural network, and x_i denotes the i^{th} sample of the ground truth epoch x .

For ocular artifact removal, we trained the fully-connected network and recurrent neural network with 60 epochs, while the simple CNN and complex CNN models were trained over 30 epochs. For myogenic artifact removal, we trained the fully—connected network with 40 epochs, recurrent neural network with 60 epochs, while the simple CNN and complex CNN models were trained over 10 epochs. In the optimization, the Adam algorithm was used to optimize the loss function, and the parameter were set to $\alpha = 5e^{-5}$, $\beta_1 = 0.5$, $\beta_2 = 0.9$.

All the four networks were implemented in Python 3.7 with Tensorflow 2.2 library, running on a computer with two NVIDIA Tesla V100 GPUs. The codes for the benchmarking algorithms are publicly available online at <https://github.com/ncclabsustech/EEGdenoiseNet>.

3.4 Performance Evaluation as Benchmark

We firstly employed network convergence as a qualitative evaluation across the four networks, which in a way provide rich information about the learning, testing and diagnose of the networks. The convergence curve of both training and testing processes were acquired by calculating the loss function value (see formula 7) with respect to the number of epochs.

We then quantitatively examined the performance of the networks by applying three objective measures on the denoised data [62], including Relative Root Mean Square Error (RRMSE) in the temporal domain ($RRMSE_{temporal}$, see formula 8), RRMSE in the spectral domain ($RRMSE_{spectral}$, see formula 9) and the average correlation coefficient (ACC see formula 10).

$$RRMSE_{temporal} = \frac{RMS(f(y) - x)}{RMS(x)} \quad (8)$$

$$RRMSE_{spectral} = \frac{RMS(PSD(f(y)) - PSD(x))}{RMS(PSD(x))} \quad (9)$$

where the function $PSD()$ denotes to the power spectral density of an input signal.

$$ACC = \frac{Cov(f(y), x)}{\sqrt{Var(f(y))Var(x)}} \quad (10)$$

4 Results

To give a general overview of the denoising outcome, we first display some example epochs in the test procedure in both time and frequency domains for ocular artifact removal (see Figure 6), and for myogenic artifact removal (see Figure 7), respectively. For each network and type of artifact, we displayed two examples: one of the best result (in left column) and one of the worst result (in right column). Generally, both in ocular and myogenic artifact removal, the artefacts were largely attenuated and the EEG samples without noise were well reconstructed. The frequency domain diagram show that the artefacts in low frequency bands were well detected and removed, but the high frequency were affected by some residual noise.

Importantly, we evaluated the four algorithms to be used as benchmark in two manner: qualitative evaluation using loss versus the number of epochs during training and testing, and quantitative evaluation using $RRMSE_{temporal}$, $RRMSE_{spectral}$ and ACC .

Figure 8 shows the loss versus the number of epochs for the four networks for ocular artifact removal. In general, both training and test loss of all the networks showed a decreasing trend with an increase of the epoch number. The training loss of each network was always lower than the test loss. RNN showed the highest initial loss value both in training loss and test loss

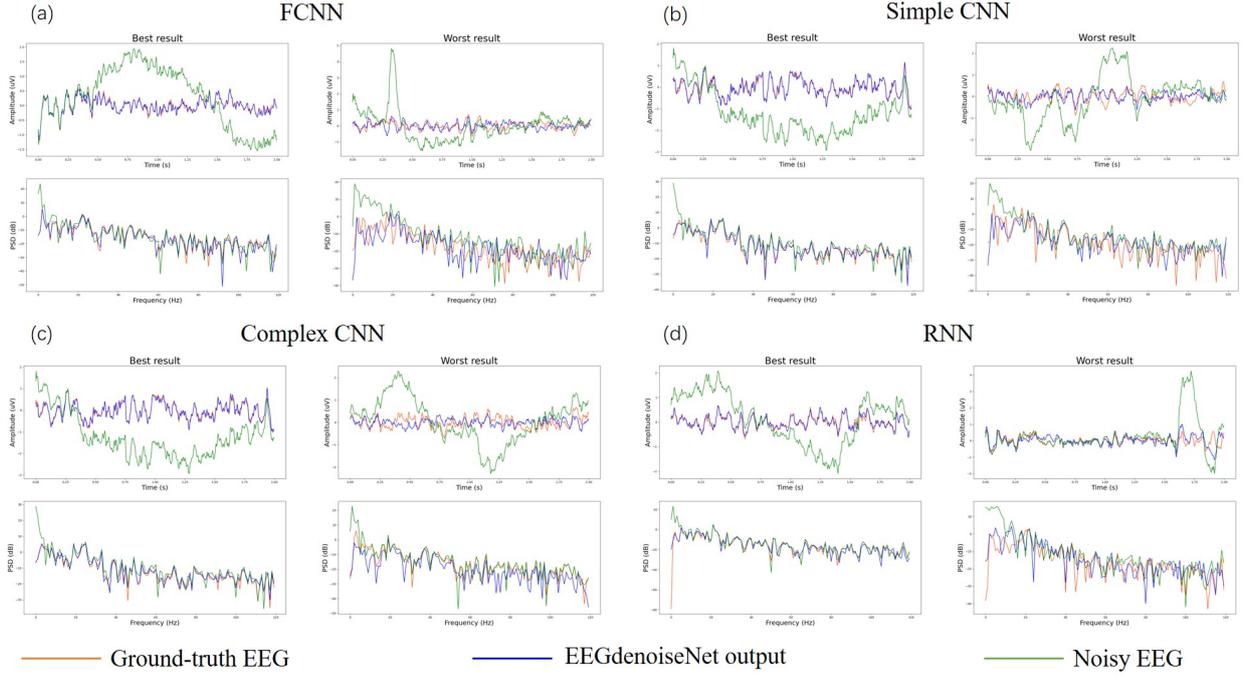


Figure 6: Some exemplary epochs of the performance in temporal domain (upper) and spectral domain (bottom) for ocular artifact removal. (a) FCNN. (b) Simple CNN. (c) Complex CNN. (d) RNN. (left) The examples with the best denoising performance; (right) the examples with the worst denoising performance. The green, red and blue line are the noisy EEG, the ground-truth EEG and the cleaned EEG by EEGdenoiseNet, respectively.

(see Figure 8d), followed by FCNN (see Figure 8a) and simple CNN (see Figure 8b), whereas complex CNN (see Figure 8c) showed the lowest training and test loss. The simple CNN and complex CNN showed larger dropping speed in training and test loss compared to the FCNN and RNN. The drop of training loss of all the networks was still visible after 25 epochs, whereas the drop of test loss in FCNN simple CNN and complex CNN was barely noticeable after 25 epochs. Only the test loss of RNN still showed a slight decreasing trend after 25 epochs.

Figure 9 shows the loss versus the number of epochs for the four networks for myogenic artifact removal. In general, FCNN and RNN showed decreasing trend both in training loss and test loss, whereas simple CNN and complex CNN showed decreasing trend only in training loss. Simple CNN and complex CNN presented a very serious phenomenon of over-fitting and reached their minimum values after one epoch and two epochs respectively. In order to eliminate the influence of over-fitting phenomenon in the results, for the two convolution networks, we selected the network with the minimum test loss as the trained network and tested its noise reduction ability. Comparing FCNN and RNN that have reasonable loss curves, we noticed that FCNN had lower training loss than RNN, but similar test loss, meaning that the difference between training loss and test loss of FCNN were larger than that of RNN.

Figure 10 displays the $RRMSE_{temporal}$, $RRMSE_{spectral}$ and ACC of the four networks in different noise levels for ocular artifact removal. Generally, for all the networks, the

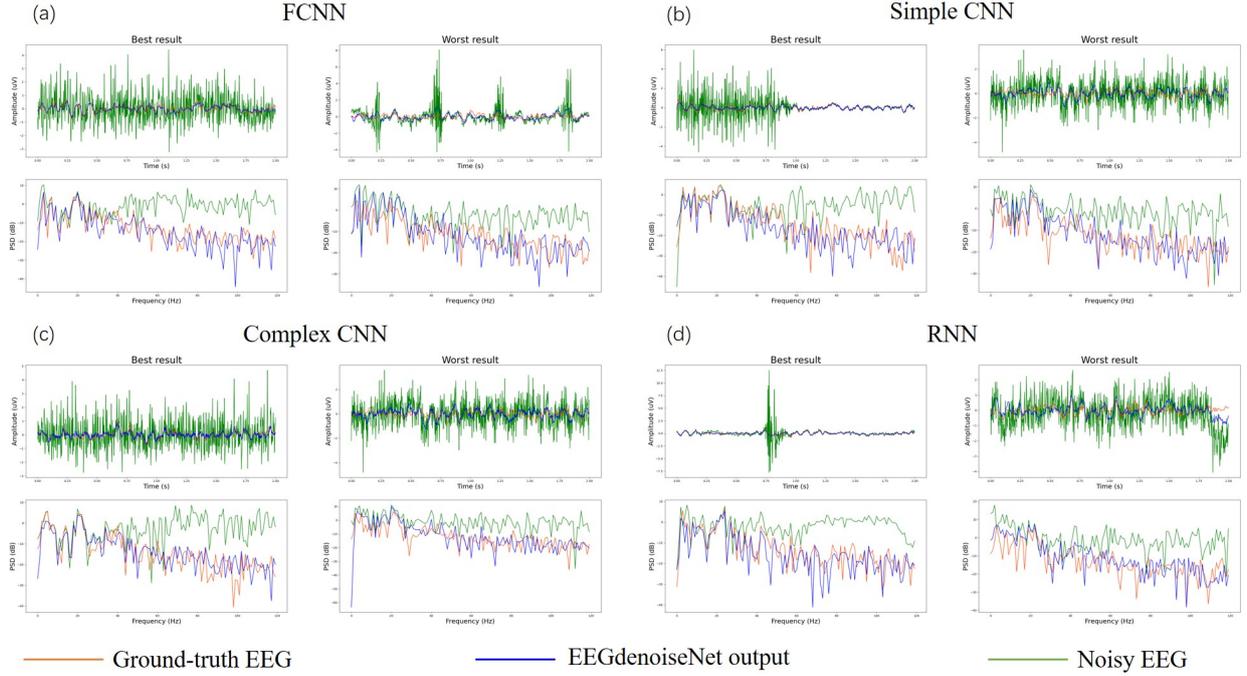


Figure 7: Some exemplary epochs showing the performance in temporal (upper) and spectral (bottom) domains for myogenic artifact removal. (a) FCNN. (b) Simple CNN. (c) Complex CNN. (d) RNN. (left) The examples with the best denoising performance; (right) the examples with the worst denoising performance. The green, red and blue line are the noisy EEG, the ground-truth EEG and the cleaned EEG by EEGdenoiseNet, respectively.

$RRMSE_{temporal}$ and $RRMSE_{spectral}$ values decreased and the ACC increased with decreasing noise level. As expected, all the networks showed a relatively higher $RRMSE_{temporal}$ and $RRMSE_{spectral}$, and lower ACC in high noise level compared to low noise level. In detail, in the high noise level part, complex CNN showed the lowest $RRMSE_{temporal}$ and $RRMSE_{spectral}$ among the four networks, followed by FCNN and simple CNN. Conversely, complex CNN had the largest ACC value in high noise level compared to FCNN and simple CNN. RNN showed the highest $RRMSE_{temporal}$ and $RRMSE_{spectral}$ and the lowest ACC value in high noise level. The values in the low noise level part among the networks are more or less similar, and barely change after the noise low to a certain level (eg. $SNR > 0$).

Figure 11 displays the three evaluation parameters for myogenic artifact removal in different noise levels. Similar to the case of ocular artifact removal, the $RRMSE_{temporal}$ and $RRMSE_{spectral}$ of the four networks showed a decreasing trend with increasing SNR, whereas the ACC showed an increasing trend. The parameters barely change when the SNR was larger than 0 dB. The simple CNN showed higher $RRMSE$ in low SNR than other networks. RNN had the lowest $RRMSE_{temporal}$ and $RRMSE_{spectral}$ and the highest ACC value in high noise level, whereas the values of the FCNN and RNN showed little difference in the low noise level.

We further summarized the quantitative benchmarks by averaging the values over all SNRs for ocular artifact removal (see Table 1) and myogenic artifact removal (see Table 2). From

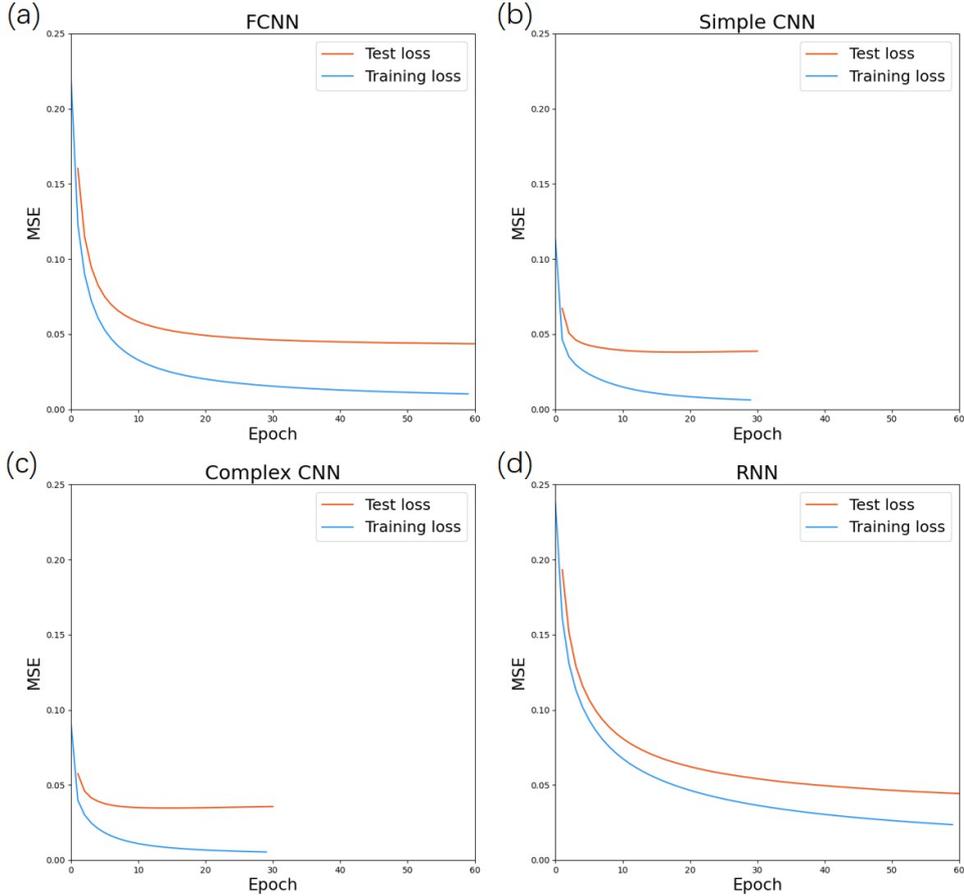


Figure 8: MSE loss of the training set and test set versus the number of epochs in ocular artifact removal. (a) FCNN; (b) Simple CNN; (c) Complex CNN; (d) RNN.

the tables, we could see that the complex CNN had the lowest average $RRMSE_{temporal}$ and highest average ACC , whereas the FCNN had the lowest average $RRMSE_{spectral}$ for ocular artifact removal. We also noticed that RNN had the lowest average $RRMSE_{temporal}$ and $RRMSE_{spectral}$ and highest average ACC in myogenic artifact removal.

To additionally evaluate the performance of the networks in different frequency bands, we calculated the average power ratio of each frequency band to whole band for ocular artifact removal (see Table 3) and myogenic artifact removal (see Table 4). The frequency bands under consideration were the delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-80 Hz) bands, whereas the whole band was between 1 and 80 Hz. For the ocular artifact removal, the power ratio of delta and theta bands, respectively, increased with the mixing of ocular artifacts, whereas the ratio of the other bands decreased. Comparing the ratio of ground truth with the ratio of the output from each model in ocular artifact removal, we noticed that in general, the DL networks permitted to recover the power ratio of each band from the noisy signal. In detail, the simple CNN showed the closest delta power ratio to the delta power ratio of ground truth; the RNN had the closest ratio to the ratio of ground truth in theta, alpha and gamma band; the complex CNN showed the closes ratio to the ratio of ground truth in beta band. In myogenic artifact removal, the power ratio of delta and gamma band noticeably

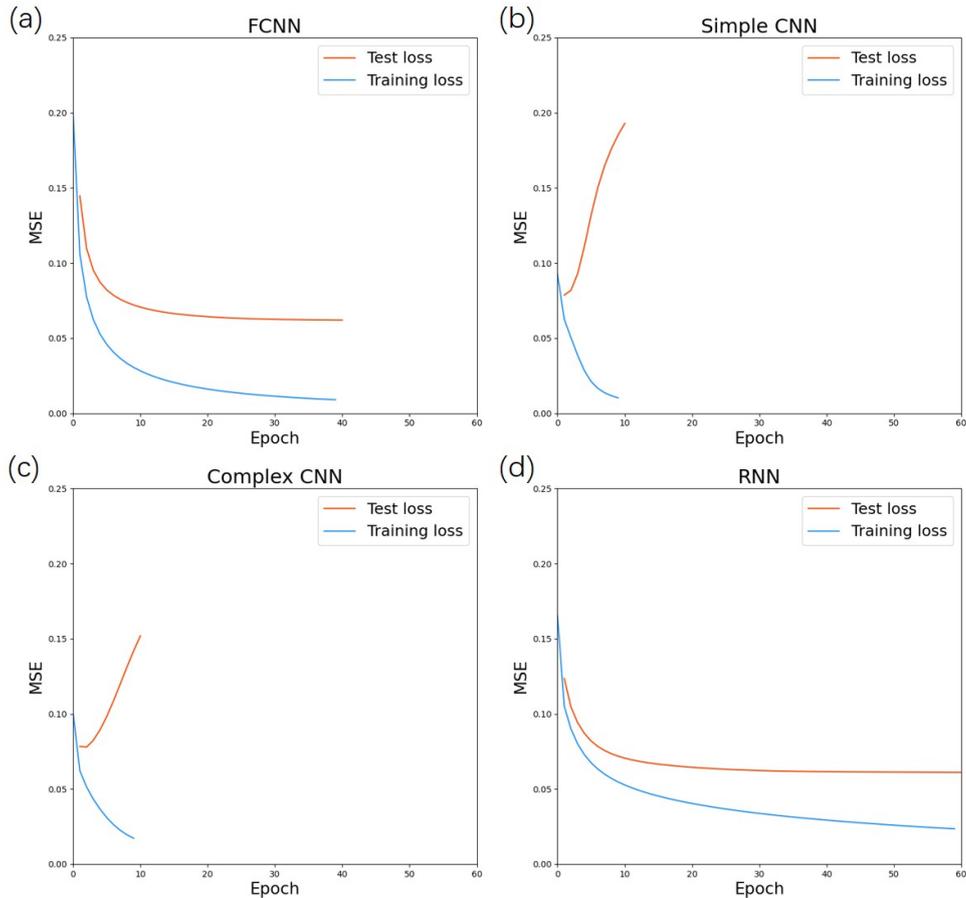


Figure 9: MSE loss of the training set and test set versus the number of epochs in myogenic artifact removal. (a) FCNN; (b) Simple CNN; (c) Complex CNN; (d) RNN.

increased after adding the artifact to the ground truth EEG, whereas the ratio of the other bands decreased. Comparing the ratio of ground truth with the ratio of the output from each model in myogenic artifact removal, we found that RNN showed the closest ratio to ground truth in theta, alpha, beta and gamma bands, while FCNN also showed the closest ratio to ground truth in alpha bands; for delta band, simple CNN showed the closest ratio to the ratio of ground truth.

5 Discussion

In this study, we provided a dataset, called EEGdenoiseNet, for training and testing DL-based denoising algorithms. To best fit the needs of DL-based denoising applications, the dataset was not only cleaned by using the ICLLabel [9] for EEG and then manual inspections for both EEG and noise, but also standardized and categorized to cleaned EEG, EOG and EMG epochs. Whereas in other publicly available dataset repositories, they provide datasets in resting state [43, 44], or related to specific tasks such as psychological tasks [45, 46, 47], motor imaginary or motor tasks [48, 49, 50, 51]. Another previous study offered semi-simulated dataset for EOG artifact removal, but they did not contain EMG signals and performance benchmarks [68]. The effective use of these datasets on DL-based denoising study requires extensive background

Table 1: Average performances of all SNRs in ocular artifact removal.

| Model | $RRMSE_{temporal}$ | $RRMSE_{spectral}$ | ACC |
|-------------|--------------------|--------------------|--------------|
| FCNN | 0.367 | 0.337 | 0.906 |
| Simple CNN | 0.359 | 0.361 | 0.916 |
| Complex CNN | 0.336 | 0.343 | 0.923 |
| RNN | 0.411 | 0.389 | 0.900 |

Table 2: Average performances of all SNRs in myogenic artifact removal.

| Model | $RRMSE_{temporal}$ | $RRMSE_{spectral}$ | ACC |
|-------------|--------------------|--------------------|--------------|
| FCNN | 0.569 | 0.552 | 0.804 |
| Simple CNN | 0.639 | 0.643 | 0.787 |
| Complex CNN | 0.632 | 0.604 | 0.786 |
| RNN | 0.561 | 0.521 | 0.816 |

Table 3: Power ratios of different frequency bands before and after ocular artifact removal

| Model | delta | theta | alpha | beta | gamma |
|--------------|--------------|--------------|--------------|--------------|--------------|
| FCNN | 0.122 | 0.124 | 0.087 | 0.496 | 0.171 |
| Simple CNN | 0.127 | 0.126 | 0.084 | 0.491 | 0.169 |
| Complex CNN | 0.126 | 0.126 | 0.088 | 0.490 | 0.171 |
| RNN | 0.104 | 0.128 | 0.101 | 0.499 | 0.167 |
| ground truth | 0.140 | 0.144 | 0.095 | 0.462 | 0.159 |
| noisy signal | 0.648 | 0.237 | 0.062 | 0.042 | 0.011 |

Table 4: Power ratios of different frequency bands before and after myogenic artifact removal

| Model | delta | theta | alpha | beta | gamma |
|--------------|--------------|--------------|--------------|--------------|--------------|
| FCNN | 0.152 | 0.162 | 0.092 | 0.458 | 0.135 |
| Simple CNN | 0.143 | 0.159 | 0.104 | 0.416 | 0.177 |
| Complex CNN | 0.132 | 0.163 | 0.104 | 0.419 | 0.181 |
| RNN | 0.126 | 0.138 | 0.094 | 0.467 | 0.174 |
| ground truth | 0.139 | 0.139 | 0.093 | 0.466 | 0.163 |
| noisy signal | 0.243 | 0.138 | 0.064 | 0.174 | 0.381 |

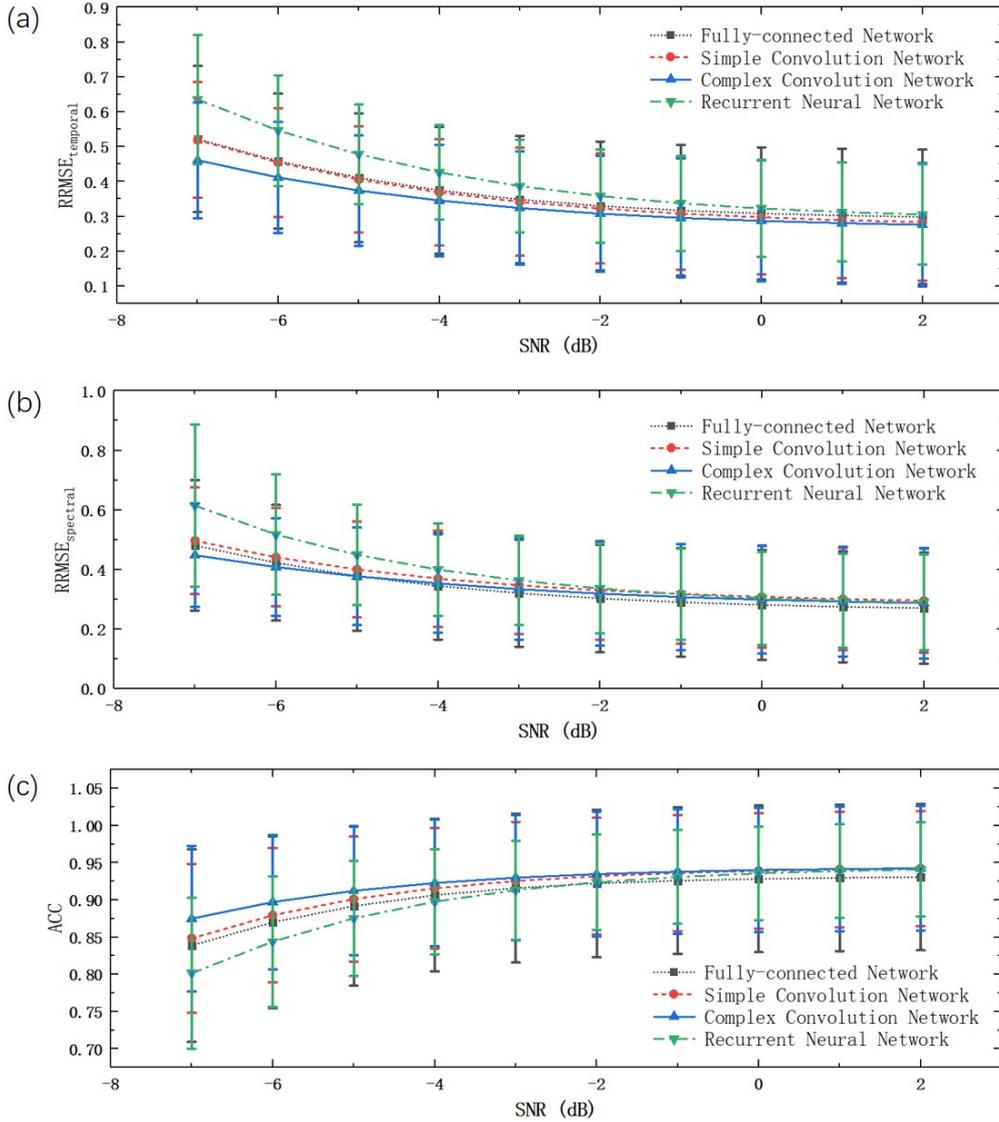


Figure 10: Performances of four networks in ocular artifact removal with different SNRs. (a) $RRMSE_{temporal}$. (b) $RRMSE_{spectral}$. (c) ACC.

knowledge on EEG, concerning the nature of the signals, EEG data format conversion, as well as EEG signal processing. In contrast, the epochs in our dataset have been already pre-processed such that the users can immediately create a large set of simulated noisy epochs with ground truth for their DL-based denoising study without being distracted by the detailed electrophysiological knowledge. With this advantage, our well-structured dataset would greatly contribute to the development of DL-based EEG denoising field.

A major issue in evaluating the performance of EEG artifact removal approaches based on DL is the lack of specific benchmarks. To fill this gap, we provided a set of benchmark algorithms along with a standardized EEG dataset. We chose well-known and relatively basic networks, i.e. a fully-connected network, a simple convolution network, a RNN and a complex convolution network, as benchmark algorithms. Performance of these DL mod-

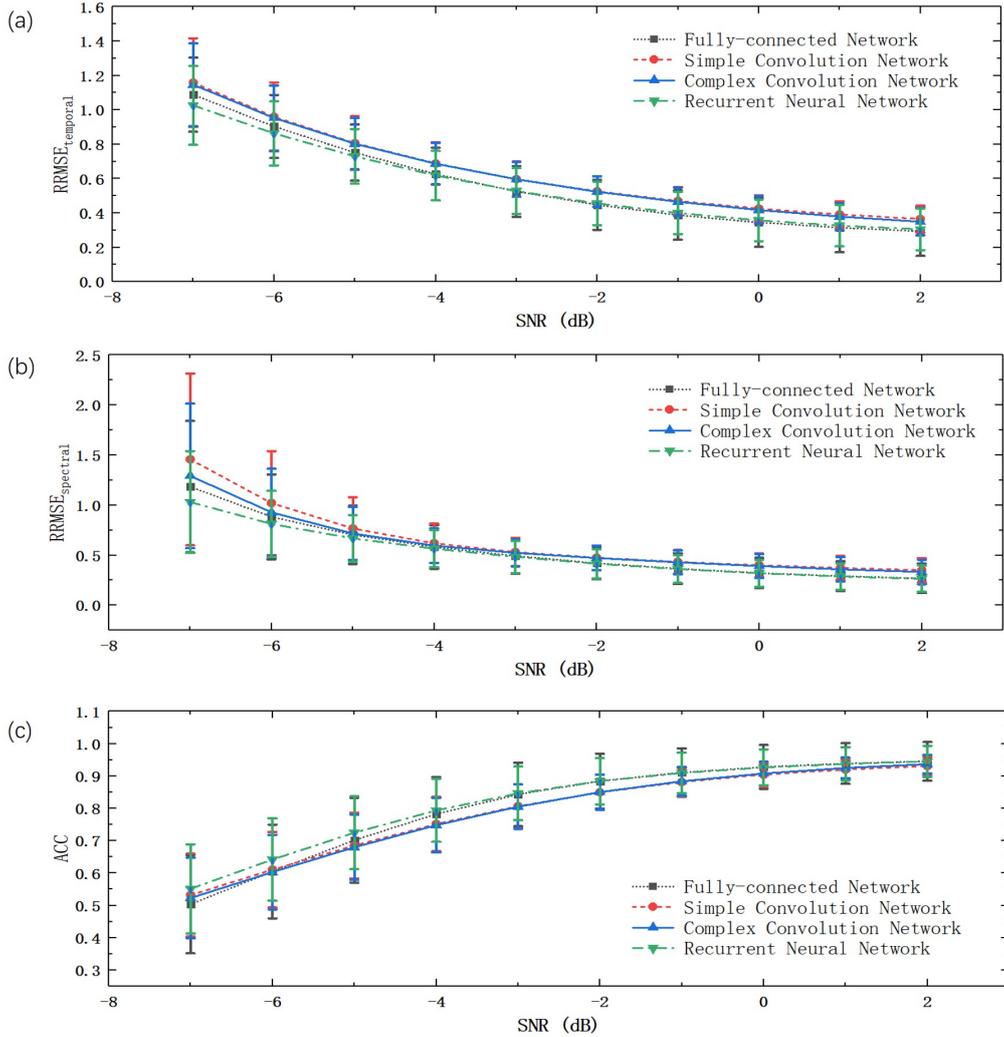


Figure 11: Performances of four networks in myogenic artifact removal with different SNRs. (a) $RRMSE_{temporal}$. (b) $RRMSE_{spectral}$. (c) ACC.

els in providing artifact-corrected EEG data was measured using standard measures such as RRMSE, PSD and ACC. Furthermore, we also defined network convergence, expressed by loss as a function of epoch number. We expect our work to contribute to the DL-based EEG denoising field, in particular because it standardizes the evaluation of method performance.

Comparisons of performance among our basic networks, traditional approaches and other DL-based methods would be interesting. Generally, the four networks were able to learn the characteristics of neural activity, and could remove ocular activity and to some extent attenuate myogenic activity from the noisy signals. In some epochs that has low level of ocular artifacts, the four networks had similar performance in ocular artifact removal, while in some cases with high level of ocular artifacts, the complex CNN outperformed other networks. We also found that the artifacts in low frequency bands were well detected and removed, but at high frequencies artifact removal was more complicated. The reason could be that deep neural networks often learn to fit the low-frequencies of target functions at the early stage of training and then fit the high-frequencies as iteration steps of training increase, which is

called *F-Principle* [69]. This also explains the fact that ocular artifacts can be more effectively attenuated than myogenic artifacts by the four networks; since ocular activity mainly lies in low frequencies while myogenic activity mostly compose of high frequency components. Interestingly, for the ocular artifact removal, the ACC of the four networks was comparable to previously reported results of traditional regression-based methods, and slightly higher than basic offline ICA-based methods [70], in turn, it was slightly lower than a previously reported BSS-REG method [70]. Previously reported DL-based ocular artifact removal method [37] also showed comparable results to traditional methods but they did not use standardized metrics for performance evaluation, which makes it hard to compare with other networks; the same problem is present when comparing with other traditional methods [71, 72, 73, 74]. Compared to ocular artifacts, myogenic artifacts were shown to be more difficult to remove using either traditional methods [75, 20], or DL-based methods (Table 1 & 2). Our results are in line with those findings. The DL methods in our study can to some extent recover neural activity from EEG contaminated by myogenic artifacts when the artifact level is relatively low. In high SNR (>0 dB) conditions, we found that the networks in our study provided comparable performance as compared to simple low-pass filters [63] or independent components analysis-based methods [63], but underperformed canonical correlation analysis-based methods [63], an independent vector analysis-based methods [76], ensemble empirical mode decomposition-joint blind source separation methods [77, 78, 79] and multivariate empirical mode decomposition-blind source separation methods [80, 81]. Remarkably, although most of the studies did not provide evaluations with low SNR (< 0 dB), we posit the DL networks may perform slightly better in this condition than most of the traditional methods. In our study we just tested four basic DL networks. We expect that more advanced networks would yield better performance in EEG artifact removal than traditional methods, especially in the case of low SNR (< 0 dB).

The application of DL-based methods on EEG denoising is beneficial under several aspects, such as its flexibility and generalizability. DL-based denoising methods require a large amount of noisy EEG and ground truth EEG signals in the training stage. On the other hand, once the model is trained, it can be easily applied in many cases such as single-channel EEG, multi-channel EEG and high-density EEG without constraints on the number of the electrodes; or can be generalized to EEG collected during different tasks, regardless of the corresponding reference channels for artifact removal. Another benefit is that DL networks have complex neural network structures that can learn the true nature of neural activities from nonlinear and non-stationary EEG signals, whereas traditional methods often linearly attenuate the artifacts. Accordingly, DL-based methods hold the promise to offer better performance than traditional approaches.

A number of limitations of this study should also be mentioned. An important potential problem is the size of the dataset. We provided several thousands of epochs for EEG, ocular and myogenic artifact categories, but it is possible that more complex networks will need larger data size for training and testing. Another drawback is the diversity of the EEG type and artifact type. EEG data may be collected in resting state or in different task conditions; furthermore, artifacts in EEG recordings are not only limited to ocular and myogenic. For example, the removal of movement artifacts is important for EEG mobile applications. Implementing a submission, review and approval mechanism of additional EEG data would be helpful. Such a continuously evolving dataset would contribute to the generalization of the DL-based EEG

denoising networks to different brain states. Moreover, we focused in this study on the denoising of 2s-long EEG epochs. Notably, the length of trials with some tasks might be longer than 2 seconds, not to mention the case of resting EEG, for which no trial can be defined. Extensions of EEGdenoiseNet for continuous removal of artifacts will be necessary in the future. The continuous artifact removal problem can be solved by defining pseudo-epochs in continuous EEG recordings, and extracting hidden relationships between continuous epochs, such that the previous epoch can be used in the training stage as input to constrain the denoising process of the current epoch.

6 Conclusion

In this study, we provided a dataset containing thousands of clean EEG, ocular artifact and muscular artifact epochs, which is suited for benchmarking DL-based EEG denoising methods. The dataset is well-structured and publicly available online in different formats. In addition, we included a set of benchmark tools, to facilitate the assessment of newly-developed DL-based EEG denoising models. Our benchmarking results suggested that DL methods have great potential for the removal of both ocular and myogenic artifacts from EEG data, even at high noise levels. Our study might accelerate developments in the field of DL-based EEG denoising.

Acknowledgement

This research was supported by Guangdong Natural Science Foundation Joint Fund (No. 2019A1515111038), High-level University Fund (No. G02386301, G02386401), Research Foundation Flanders (No. G0F76.16N, G0936.16N and EOS.30446199) and Chinese Scholarship Council (No. 201708620182).

Competing interests

The authors declare no competing interests.

References

- [1] Hanshu Cai, Zhidiao Qu, Zhe Li, Yi Zhang, Xiping Hu, and Bin Hu. Feature-level fusion approaches based on multimodal eeg data for depression recognition. *Information Fusion*, 59:127–138, 2020.
- [2] Lindsay M Oberman, Edward M Hubbard, Joseph P Mccleery, Eric Lewin Altschuler, Vilayanur S Ramachandran, and Jaime A Pineda. Eeg evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, 24(2):190–198, 2005.
- [3] Jonathan R Wolpaw, Dennis J Mcfarland, Gregory W Neat, and Catherine Forneris. An eeg-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78(3):252–259, 1991.

- [4] Christoph Herrmann and Tamer Demiralp. Human eeg gamma oscillations in neuropsychiatric disorders. *Clinical Neurophysiology*, 116(12):2719–2733, 2005.
- [5] Quanying Liu, Seyedehrezvan Farahibozorg, Camillo Porcaro, Nicole Wenderoth, and Dante Mantini. Detecting large-scale networks in the human brain using high-density electroencephalography. *Human Brain Mapping*, 38(9):4631–4643, 2017.
- [6] Mingqi Zhao, Marco Marino, Jessica Samogin, Stephan P Swinnen, and Dante Mantini. Hand, foot and lip representations in primary sensorimotor cortex: a high-density electroencephalography study. *Scientific Reports*, 9(1):1–12, 2019.
- [7] Rodney J Croft and Robert J Barry. Removal of ocular artifact from the eeg : a review. *Neurophysiologie Clinique-clinical Neurophysiology*, 30(1):5–19, 2000.
- [8] Suresh D Muthukumaraswamy. High-frequency brain activity and muscle artifacts in meg/eeg: a review and recommendations. *Frontiers in Human Neuroscience*, 7:138–138, 2013.
- [9] Luca Piontonachini, Kenneth Kreutzdelgado, and Scott Makeig. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, 2019.
- [10] Mike X Cohen. *Analyzing neural time series data: theory and practice*. MIT press, 2014.
- [11] Maximilien Chaumon, Dorothy V M Bishop, and Niko A Busch. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, 250:47–63, 2015.
- [12] Brenton W Mcmenamin, Alexander J Shackman, Jeffrey S Maxwell, David R W Bachhuber, Adam M Koppenhaver, Lawrence L Greischar, and Richard J Davidson. Validation of ica-based myogenic artifact correction for scalp and source-localized eeg. *NeuroImage*, 49(3):2416–2432, 2010.
- [13] Gabriele Gratton, Michael GH Coles, and Emanuel Donchin. A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, 55(4):468–484, 1983.
- [14] Rodney J Croft and Robert J Barry. Eog correction : Which regression should we use? *Psychophysiology*, 37(1):123–125, 2000.
- [15] BRENTON W McMENAMIN, Alexander J Shackman, Jeffrey S Maxwell, Lawrence L Greischar, and Richard J Davidson. Validation of regression-based myogenic correction techniques for scalp and source-localized eeg. *Psychophysiology*, 46(3):578–592, 2009.
- [16] Ping He, G Wilson, and C Russell. Removal of ocular artifacts from electro-encephalogram by adaptive filtering. *Medical and biological engineering and computing*, 42(3):407–412, 2004.
- [17] Ping He, Glenn Wilson, Christopher Russell, and Maria Gerschutz. Removal of ocular artifacts from the eeg: a comparison between time-domain regression method and adaptive filtering method using simulated data. *Medical & biological engineering & computing*, 45(5):495–503, 2007.

- [18] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.
- [19] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.
- [20] Xun Chen, Xueyuan Xu, Aiping Liu, Soojin Lee, Xiang Chen, Xu Zhang, Martin J McKeown, and Z Jane Wang. Removal of muscle artifacts from the eeg: a review and recommendations. *IEEE Sensors Journal*, 19(14):5353–5368, 2019.
- [21] Tzyy-Ping Jung, Scott Makeig, Marissa Westerfield, Jeanne Townsend, Eric Courchesne, and Terrence J Sejnowski. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111(10):1745–1758, 2000.
- [22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [24] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.
- [26] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [27] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.
- [28] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.
- [29] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5):051001, aug 2019.
- [30] Siavash Sakhavi, Cuntai Guan, and Shuicheng Yan. Parallel convolutional-linear neural network for motor imagery classification. In *Signal Processing Conference*, 2015.
- [31] Yousef, Rezaei, Tabar, Ugur, and Halici. A novel deep learning approach for classification of eeg motor imagery signals. *Journal of Neural Engineering*, 2017.

- [32] Zhichuan Tang, Chao Li, and Shouqian Sun. Single-trial eeg classification of motor imagery using deep convolutional neural networks. *Optik - International Journal for Light and Electron Optics*, 2017.
- [33] Tian-jian Luo, Yachao Fan, Lifei Chen, Gongde Guo, and Changle Zhou. Eeg signal reconstruction using a generative adversarial network with wasserstein distance and temporal-spatial-frequency loss. *Frontiers in Neuroinformatics*, 14:15, 2020.
- [34] Isaac A. Corley and Yufei Huang. Deep eeg super-resolution: Upsampling eeg spatial resolution with generative adversarial networks. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 2018.
- [35] Kay Hartmann, Robin Schirrmester, and Tonio Ball. Eeg-gan: Generative adversarial networks for electroencephalographic (eeg) brain signals, 06 2018.
- [36] Fatemeh Fahimi, Strahinja Dosen, Kai Keng Ang, Natalie Mrachacz-Kersting, and Cuntai Guan. Generative adversarial networks-based data augmentation for brain-computer interface. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
- [37] Banghua Yang, Kaiwen Duan, Chengcheng Fan, Chenxiao Hu, and Jinlong Wang. Automatic ocular artifacts removal in eeg using deep learning. *Biomedical Signal Processing and Control*, 43:148–158, 2018.
- [38] Weitong Sun, Yuping Su, Xia Wu, and Xiaojun Wu. A novel end-to-end 1d-rescnn model to remove artifact from eeg signals. *Neurocomputing*, 2020.
- [39] Conor Hanrahan. Noise reduction in eeg signals using convolutional autoencoding techniques. 2019.
- [40] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [41] Michaël Defferrard, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. Fma: A dataset for music analysis. *arXiv preprint arXiv:1612.01840*, 2016.
- [42] Alexander Panchenko, Eugen Ruppert, Stefano Faralli, Simone Paolo Ponzetto, and Chris Biemann. Building a web-scale dependency-parsed corpus from commoncrawl. *arXiv preprint arXiv:1710.01779*, 2017.
- [43] Logan T Trujillo, Candice T Stanfield, and Ruben D Vela. The effect of electroencephalogram (eeg) reference choice on information-theoretic measures of the complexity and integration of eeg signals. *Frontiers in Neuroscience*, 11:425, 2017.
- [44] Mastaneh Torkamani-Azar, Sumeyra Demir Kanik, Serap Aydin, and Mujdat Cetin. Prediction of reaction time and vigilance variability from spatio-spectral features of resting-state eeg in a long sustained attention task. *IEEE journal of biomedical and health informatics*, 2020.

- [45] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.
- [46] Gijsbrecht Van Veen, Alexandre Barachant, Anton Andreev, Grégoire Cattan, Pedro Coelho Rodrigues, and Marco Congedo. Building brain invaders: Eeg data of an experimental validation. *arXiv preprint arXiv:1905.05182*, 2019.
- [47] Louis Korczowski, Martine Cederhout, Anton Andreev, Grégoire Cattan, Pedro C Rodrigues, Violette Gautheret, and Marco Congedo. Brain invaders calibration-less p300-based bci with modulation of flash duration dataset (bi2015a). *Archives Overtes*, 2019.
- [48] Matthew D Luciw, Ewa Jarocka, and Benoni B Edin. Multi-channel eeg recordings during 3,936 grasp and lift trials with varying weight and friction. *Scientific data*, 1(1):1–11, 2014.
- [49] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggenberger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, aug 2017.
- [50] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043, 2004.
- [51] Murat Kaya, Mustafa Kemal Binli, Erkan Ozbay, Hilmi Yanar, and Yuriy Mishchenko. A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Scientific data*, 5:180211, 2018.
- [52] Haoming Zhang, Mingqi Zhao, Chen Wei, Dante Mantini, Zherui Li, and Quanying Liu. Eegdenoisenet. <https://github.com/nccclabsustech/EEGdenoiseNet>.
- [53] Cho Hohyun, Ahn Minkyu, Ahn Sangtae, Kwon Moonyoung, and Jun Sung Chan. Eeg datasets for motor imagery brain computer interface. *Gigaence*, 6(7), 2017.
- [54] Suguru Kanoga, Masaki Nakanishi, and Yasue Mitsukura. Assessing the effects of voluntary and involuntary eyeblinks in independent components of electroencephalogram. *Neurocomputing*, 193(Jun.12):20–32, 2016.
- [55] Mehrdad Fatourech, Ali Bashashati, Rabab K. Ward, and Gary E. Birch. Emg and eeg artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, 118(3):480–494, 2007.
- [56] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller. Seperability of four-class motor imagery data using independent components analysis. *Journal of Neural Engineering*, 3(3):208–216, 2006.
- [57] Schlogl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, and Pfurtscheller and. A fully automated correction method of eeg artifacts in eeg recordings. *Clinical Neurophysiology Shannon*, 2007.

- [58] Clemens Brunner, Muhammad Naeem, Robert Leeb, Bernhard Graimann, and Gert Pfurtscheller. Spatial filtering and selection of optimized components in four class motor imagery eeg data using independent components analysis. *Pattern Recognition Letters*, 28(8):957–964, 2007.
- [59] Alois Schlögl, Julien Kronegg, Jane Huggins, and Steve Mason. Evaluation criteria for bci research. *Toward Brain-Computer Interfacing*, 01 2007.
- [60] Ville Rantanen, Mirja Ilves, Antti Vehkaoja, Anton Kontunen, Jani Lylykangas, Eeva Mäkelä, Markus Rautiainen, Veikko Surakka, and Jukka Lekkala. A survey on the feasibility of surface emg in facial pacing. volume 2016, pages 1688–1691, 08 2016.
- [61] Gang Wang, Chaolin Teng, Kuo Li, Zhonglin Zhang, and Xiangguo Yan. The removal of eeg artifacts from eeg signals using independent component analysis and multivariate empirical mode decomposition. *IEEE Journal of Biomedical & Health Informatics*, 20(5):1301–1308, 2016.
- [62] Xun Chen, Hu Peng, Fengqiong Yu, and Kai Wang. Independent vector analysis applied to remove muscle artifacts in eeg data. *IEEE Transactions on Instrumentation and Measurement*, pages 1–10, 2017.
- [63] Wim De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel. Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Transactions on Biomedical Engineering*, 53(12):2583–2587, 2006.
- [64] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- [65] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [66] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.
- [67] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [68] Manousos A Klados and Panagiotis D Bamidis. A semi-simulated eeg/eog dataset for the comparison of eeg artifact rejection techniques. *Data in brief*, 8:1004–1006, 2016.
- [69] Zhi-Qin John Xu, Yaoyu Zhang, Tao Luo, Yanyang Xiao, and Zheng Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523*, 2019.
- [70] Roberto Guarnieri, Marco Marino, Federico Barban, Marco Ganzetti, and Dante Mantini. Online eeg artifact removal for bci applications by adaptive spatial filtering. *Journal of neural engineering*, 15(5):056009, 2018.

- [71] Hoang-Anh T Nguyen, John Musson, Feng Li, Wei Wang, Guangfan Zhang, Roger Xu, Carl Richey, Tom Schnell, Frederic D McKenzie, and Jiang Li. Eog artifact removal using a wavelet neural network. *Neurocomputing*, 97:374–389, 2012.
- [72] JC Woestenburg, MN Verbaten, and JL Slangen. The removal of the eye-movement artifact from the eeg by regression analysis in the frequency domain. *Biological psychology*, 16(1-2):127–147, 1983.
- [73] Thomas Elbert, Werner Lutzenberger, Brigitte Rockstroh, and Niels Birbaumer. Removal of ocular artifacts from the eeg—a biophysical approach to the eog. *Electroencephalography and clinical neurophysiology*, 60(5):455–463, 1985.
- [74] Germán Gómez-Herrero, Wim De Clercq, Haroon Anwar, Olga Kara, Karen Egiazarian, Sabine Van Huffel, and Wim Van Paesschen. Automatic removal of ocular artifacts in the eeg without an eog reference channel. In *Proceedings of the 7th Nordic Signal Processing Symposium-NORSIG 2006*, pages 130–133. IEEE, 2006.
- [75] Brenton W McMenamin, Alexander J Shackman, Lawrence L Greischar, and Richard J Davidson. Electromyogenic artifacts and electroencephalographic inferences revisited. *Neuroimage*, 54(1):4–9, 2011.
- [76] Xun Chen, Hu Peng, Fengqiong Yu, and Kai Wang. Independent vector analysis applied to remove muscle artifacts in eeg data. *IEEE Transactions on Instrumentation and Measurement*, 66(7):1770–1779, 2017.
- [77] Xun Chen, Aiping Liu, Hu Peng, and Rabab K Ward. A preliminary study of muscular artifact cancellation in single-channel eeg. *Sensors*, 14(10):18370–18389, 2014.
- [78] Xun Chen, Chen He, and Hu Peng. Removal of muscle artifacts from single-channel eeg based on ensemble empirical mode decomposition and multiset canonical correlation analysis. *Journal of Applied Mathematics*, 2014, 2014.
- [79] Xun Chen, Qiang Chen, Yu Zhang, and Z Jane Wang. A novel eemd-cca approach to removing muscle artifacts for pervasive eeg. *IEEE Sensors Journal*, 19(19):8420–8431, 2018.
- [80] Xun Chen, Xueyuan Xu, Aiping Liu, Martin J McKeown, and Z Jane Wang. The use of multivariate emd and cca for denoising muscle artifacts from few-channel eeg recordings. *IEEE transactions on instrumentation and measurement*, 67(2):359–370, 2017.
- [81] Xueyuan Xu, Xun Chen, and Yu Zhang. Removal of muscle artefacts from few-channel eeg recordings based on multivariate empirical mode decomposition and independent vector analysis. *Electronics Letters*, 54(14):866–868, 2018.

Figures and Tables

Figure 1: The pipeline for obtaining clean EEG, EOG and MEG.

Figure 2: Examples of epochs in EEGdenoiseNet dataset. (a) An EEG epoch. (b) A vertical EOG (vEOG) epoch. (c) A horizontal EOG (hEOG) epoch. (d) An EMG epoch.

Figure 3: The structure of EEGdenoiseNet.

Figure 4: The pipelines for mixing noises to clean EEG signal. (a) Addition of ocular artifacts. (b) Addition of myogenic artifacts..

Figure 5: The structures of the four DL-based methods for benchmarking. (a) FCNN ; (b) Simple CNN; (c) Complex CNN; (d) RNN.

Figure 6: Some exemplary epochs of the performance in temporal domain (upper) and spectral domain (bottom) for ocular artifact removal. (a) FCNN. (b) Simple CNN. (c) Complex CNN. (d) RNN. (left) The examples with the best denoising performance; (right) the examples with the worst denosing performance. The green, red and blue line are the noisy EEG, the ground-truth EEG and the denoised EEG by EEGdenoiseNet, respectively.

Figure 7: Some exemplary epochs showing the performance in temporal (upper) and spectral (bottom) domains for myogenic artifact removal. (a) FCNN. (b) Simple CNN. (c) Complex CNN. (d) RNN. (left) The examples with the best denoising performance; (right) the examples with the worst denoising performance. The green, red and blue line are the noisy EEG, the ground-true EEG and the cleaned EEG by EEGdenoiseNet, respectively.

Figure 8: MSE loss of the training set and test set versus the number of epochs in ocular artifact removal.

Figure 9: MSE loss of the training set and test set versus the number of epochs in myogenic artifact removal.

Figure 10: Performances of four networks in ocular artifact removal with different SNRs.

Figure 11: Performances of four networks in myogenic artifact removal with different SNRs.

Table 1: Average performances of all SNRs in ocular artifact removal The $RRMSE_{temporal}$, $RRMSE_{spectral}$ and ACC of FCNN, simple CNN, complex CNN and RNN are listed.

Table 2: Average performances of all SNRs in myogenic artifact removal

Table 3: Power ratios of different frequency bands before and after ocular artifact removal.

Table 4: Power ratios of different frequency bands before and after myogenic artifact removal.