EasyChair Preprint

№ 10275

EEGWave: a Denoising Diffusion Probabilistic Approach for EEG Signal Generation

Szabolcs Torma and Luca Szegletes

May 26, 2023

EEGWave: a Denoising Diffusion Probablistic Approach for EEG Signal Generation

Szabolcs Torma Dr. Luca Szegletes
Department of Automation and Applied Informatics
Budapest University of Technology and Economics
{Torma.Szabolcs, Szegletes.Luca}@aut.bme.hu

Abstract. The importance of brain-computer interface (BCI) systems in modern-day healthcare and robotics is indisputable. BCIs are commonly based on signals recorded by electroencephalography (EEG) due to the ease of use and relatively low cost of the measurement technology. Evoked potentials (EP) are well-measurable with EEG and can be utilized to control BCIs. The processing of these signals is a very complex task, because of the low signal-to-noise ratio, the inter-subject, and inter-measurement variability. In recent years, deep learning (DL) methods have proven to be powerful algorithms for signal processing and decoding. However, a large amount of data with good quality and variety is needed in order to train DL models properly and use them as generally as possible. Finding such a publicly available data set or even creating one is a resource-intensive task. Denoising Diffusion Probabilistic Models (DDPM) got more attention over generative adversarial networks (GAN) only recently and in image processing tasks they achieved state-of-the-art results. In this work, We propose a DDPM, called EEGWave, for brain signal generation and show that DDPMs are promising options for augmenting EEG data sets. We present our results on a data set which contains EPs to present the performance of the model in the signal synthesis task.

Keywords: EEGWave; Deep Learning; Diffusion Probabilistic Model; Brain-Computer Interface (BCI); Electroenchephalography (EEG); Signal Generation;

1 Introduction

Brain-computer interfaces (BCIs) give the opportunity to communicate or control external devices only with brain signals. These systems are used in many fields from healthcare to entertainment, but most importantly, BCIs are useful associate devices for people with limited communication or movement abilities in their everyday life. BCIs record brain activity, detect temporal and frequency patterns and decode these data into actions. Low limitation, high safety and comfort are examples for requirements of the interfaces. BCIs must be usable in real-time with high speed and accuracy.

Although many brain recording technologies exist (e.g. MRI, CT), electroencephalography (EEG) is one of the procedures with the highest temporal resolution and lowest costs. EEGs can be placed both on the scalp and intracranial

(often called electrocorticography), but non-invasive EEGs mean much fewer complications. Event-related potentials (ERPs) and motor imagery (MI) signals can be used for controlling BCIs, and these signals are well-measurable by EEGs, thus making EEG-based BCIs highly common. BCIs translate the measured signals into actions through classification, detection, or similar techniques. Non-invasive EEG signals have a very low signal-to-noise ratio (SNR) and the recordings are contaminated with so-called artifacts that originate from different parts of not just the cortex but the whole body. Furthermore, the intra-subject and intra-measurement variability are high, so the signal decoding algorithm has to be highly robust to be applicable as generally as possible.

Deep learning (DL) is more and more popular in EEG signal processing. [1] The major advantage of DL algorithms is automatic feature extraction from minimally pre-processed or even raw data, whereas traditional algorithms (e.g. independent component analysis, support vector machines) are highly dependent on proper manual feature engineering. DL approaches are becoming more efficient in complex problem solving, however, a lot of data is inevitable for the proper training of these models. [2]

In the medical field, finding sufficient publicly available data to develop a well-functioning DL algorithm is a challenge. Recording a data set is a time and resource-consuming process. Professionals are often needed to check and exclude invalid data samples in order to obtain a set with samples of high quality. Furthermore, the handling of personal information must be appropriate and has to be done with great care. [3, 4] Because of the mentioned issues, it would be beneficial to find other ways to obtain EEG data.

Several works have been published, demonstrating the promising capabilities of general adversarial networks (GAN) [5–7] and, more recently, denoising diffusion probabilistic models (DDPM) [8–10] in data generation tasks. In these setups, life-like samples are generated by algorithms, which have similar statistical properties to the original data samples. GANs and variational autoencoders (VAEs) can generate images with so high fidelity that even a person could not tell that the synthetic image is fake. The training of GANs is ill-posed due to stability issues and VAEs need auxiliary objective functions in lots of cases. DDPMs have gained more attention in the last couple of years as they require neither special losses nor specialized training to be trained well. DDPMs achieved state-of-the-art results in both image and audio processing tasks. However, in EEG signal processing the capabilities of this framework have not been explored, yet.

Our work aims to synthesize EEG data with ERP signals. In our approach, we generate new data samples from the underlying distribution of the real samples by using a DL architecture in a DDPM framework. The current study has the following contributions:

- Introduction of EEGWave, a novel DDPM model with bi-directional dilated convolutional layers for high-resolution EEG signal processing in a DDPM framework.

- Synthesis of EEG recorded ERP data with a DDPM. A comparison of the performance of EEGWave and a GAN shows that our model is capable of brain signal generation with high fidelity and diversity without producing many artifacts in time or frequency domains.

The structure of this work is as follows: in 2 the basics of EEG technology and DDPM models are given. Our EEGWave architecture is presented in 3, the description of the experiment with the used data set, procedures and metrics is given in 4 and finally, we present the performance of EEGWave in 5, followed by closing thoughts in 6.

2 Background

2.1 Electroencephalography

Many of brain signals that are measurable via EEG have been useful in understanding brain activity, diagnosing mental and physical illnesses, and regulating BCIs. EEG can record the electrical activity of the brain, which is the result of the firing of neurons. The measured signals have several characteristics from which the most common ones are the amplitude alterations in time domain, the spectral components and the phase. For example, based on frequency, signals can be grouped into 5 major frequency bands: the delta, theta, alpha, beta and gamma.

EEG has several advantages, but also challenges. [11] Brain activity is measured in a non-invasive manner using electrodes placed on the scalp. The use of an EEG device is relatively inexpensive and less complicated compared to other brain imaging technologies, e.g. fMRI, PET, CT, although fMRI and NIRS are often used as complementary methods to obtain more information from the brain. [12], [13] The temporal resolution of the technology is determined by the sampling frequency therefore samples can be obtained with a difference of only a couple of milliseconds. Most EEGs record signals with a sampling frequency from 128-1024 Hz. [14]

The major challenge of the use of EEG is that as the electrodes are placed on the scalp, they can record only the superposition of different signals from the different areas of the brain. The spatial resolution of the technology is considered low, (especially compared to e.g. fMRI or CT) and it is the indication of the number of electrodes used during the procedure. With more electrodes, the origin areas of the signals can be specified better, although the number of electrodes is limited to 256 at most in most cases. The 10-20 and 10-10 international standards define the position of the electrodes. Another issue with non-invasive electrodes is that they can only measure neural activity in the order of microVolts. Therefore these signals are often heavily contaminated with noise and artifacts, thus being the signal-to-noise ratio (SNR) of the measured data low. [14]

ERP signals are well-explored and commonly used in BCI applications. ERPs are elicited by both external and internal stimuli. A subcategory of ERPs

consists of evoked potentials (EPs) from which the P300 is highly popular. P300 is characterized by its positive amplitude at $300ms$ after the onset of an external visual or auditory stimulus. The detection of P300 components can be utilized in control and communication tasks (e.g. P300 speller).

2.2 Denoising Diffusion Probabilistic Models

Diffusion probabilistic models (DPMs) were introduced by *Sohl-Dickstein, Maheswaranathan & Ganguli* in [15]. DPMs are generative models that tend to be more stable during training than GANs, also in contrast to other generative a simple L1 or L2 objective function was enough in recent works to train the model properly. Diffusion models learn the underlying log-likelihood of the original data set, thus being capable of data synthesis with higher diversity than any other generative models

A DPM is based on a fixed Markov chain that consists of a forward and reverses the process, namely forward diffusion and reverse sampling. At one end of the chain, there is a data sample from the real distribution of the data set, while at the other end there is a noise sample from simple distribution (e.g. Gaussian). The Markov chain consists of T number or states that in between the transition are defined by a priori and a posteriori log-likelihoods.

The forward process takes the sample x_0 from the original data distribution $q(x_0)$ and adds a small amount of Gaussian noise to it in every step. The procedure is called diffusion and the resulting data sample at the sufficiently large step T, x_T can be viewed as a noise sample from the isotropic Gaussian distribution $\mathcal{N}(0, \mathbf{I})$. Each resulting element at the different time steps $x_1, x_2, ..., x_{T-1}$ is the perturbed version of the original sample. The whole diffusion process, which converts the complex data distribution into a tractable latent one, is defined as:

$$q(x_1, x_2, ..., x_T|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}) \tag{1}$$

, where the Markov transition is:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}) \tag{2}$$

Although, $\beta_t \in (0, 1)$ could be learned, in earlier works [9, 10, 16], they were hyper-parameters given in the form of a so-called variance schedule, which defines the βs for every step $t = 1, 2, ..., T$. Notice that the βs in the schedule are time-dependent and $\beta_{t-1} < \beta_t$ for every $t > 1$. As *Ho et al.* pointed out in [9], efficient training of DDPMs is possible through sampling x_t in a closed form, which is given as:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\overline{\alpha}_t}x_0, (1-\overline{\alpha}_t)\mathbf{I}) \tag{3}$$

, where $\alpha_t = 1 - \beta_t$ and $\overline{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ and t is from the uniform distribution $U[1, T]$.

A reverse diffusion process can be defined for data synthesis. If β_t is small enough, the inverse transition function have the same functional form as the

forward one. [15] An $x_T \sim \mathcal{N}(0, \mathbf{I})$ sample is taken and then denoised recursively, while going through the Markov chain in reverse and applying the conditional transitions $q(x_{t-1}|x_t)$, until the an $x_0 \sim q(x_0)$ sample is achieved. The task of the neural network in a DDPM is to approximate the $q(x_{t-1}|x_t)$ by $p_\theta(x_{t-1}|x_t)$. The chain for the transition from x_T to x_0 is given by

$$p_\theta(x_0, x_1, ..., x_{T-1}|x_T) = \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t) \tag{4}$$

,where the transition is given as:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)\mathbf{I}) \tag{5}$$

μ_θ and Σ_θ can be predicted by the same model that takes a diffused x_t sample and a step t as its inputs. Through μ_θ and Σ_θ, the model predicts the Gaussian noise with which the sample was perturbed. *Ho et al.* presented a re-parameterization of the problem in which the variance schedule is used instead of a Σ_θ, and the model predicts the noise ϵ_θ which should be subtracted from the diffused sample. This re-parameterization is given as:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)) \tag{6}$$

A DPM can be trained by maximizing the variational lower bound, however, based on the re-parametrization and the connection with Langevin dynamics under certain circumstances, *Ho et al.* showed that it is possible to use the simple L2 (or even L1) loss as objective functions. The simpler loss takes the form of:

$$L(\theta) = \mathbb{E}_{t,x_0,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|_2^2\right] \tag{7}$$

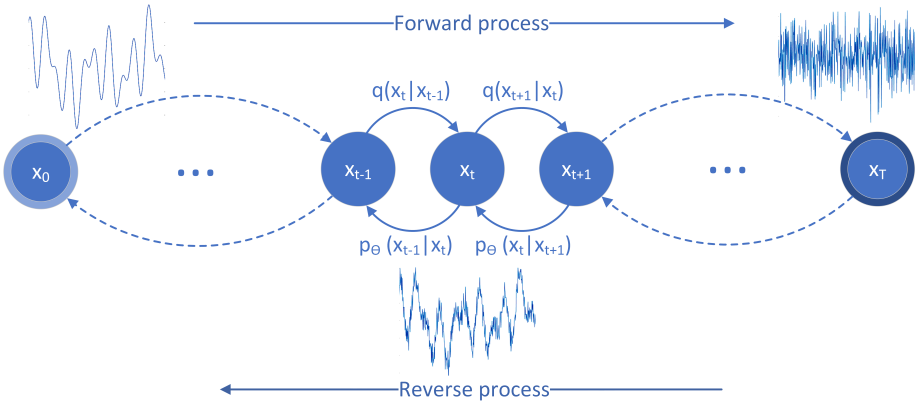

Figure 1: Markov chain of the diffusion process.

3 Methods

EEGWave is a novel denoising diffusion probabilistic model. The dimensions of the input and output data are the same and we omit sampling layers to avoid the production of temporal and spectral artifacts during data synthesis.

The model contains a group of residual layers, where each residual is connected to an input and output stream. In the residuals, the encoded and embedded diffusion noise at step t is added to the input temporal features (similar to [16]), then the data is passed through a bi-directional dilated convolution layer with a size 3 kernel. The resulted features are divided into two parts along the channel dimension. Sigmoid and Tanh activations are applied to the chunks and they are multiplied by element-wise multiplication. The data are then passed through a point-wise convolution layer and the output of the convolution is divided into two chunks along the channel dimension, again. One chunk goes into the next residual and the other is a direct output, called the skip output.

The skip outputs are added to the other skips of the remaining residual blocks. This design helps maintain proper gradient flow in the earlier layers as well. Aggregated skip outputs are normalized. The output of the model is produced by the final series of 1D convolution, ReLU, and 1D convolution layers. There is no activation used on the output sample and the final output is reshaped from (N x C x 1 x L) to (N x 1 x C x L), where L is the length of the series and C is the number of EEG channels.

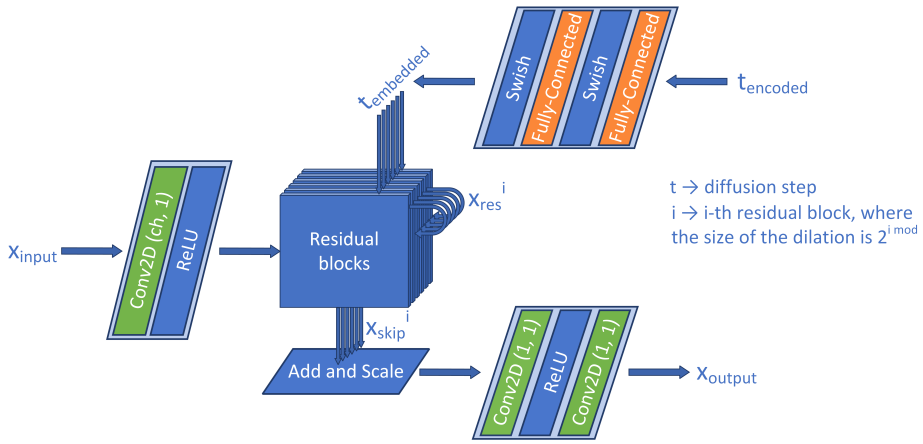

Figure 2: EEGWave architecture

An EEG data segment contains several frequency components from lower to higher, so the distance between related points in the signals can be large over the spatial-temporal dimension. The bi-directional dilated convolutions are responsible for learning global spatial-temporal features. Dilated convolutions

can operate relatively inexpensively while learning distant relationships with fixed kernels.

The diffusion steps are embedded by encoding the step into a vector and passing through a sequence of 2 linear layers, both followed by Swish activations. Each residual block has a third, linear layer that allows the residual to learn local characteristics from the global embedding. Coding is done with sine and cosine functions, as in [17]. From each diffusion step t, a 128-dimensional vector is created as follows:

$$t_{encoded} = [\sin(10^{\frac{ix4}{63}}t), \cos(10^{\frac{ix4}{63}}t)],\ i = 0:63 \tag{8}$$

The locally embedded noise features are produced by linear layers inside each residual block.

In this work, the number of residual layers is set to 40 with a dilation cycle of $2^{0:7}$. The kernel size is set to 3. The channel dimension is set to 128 in the residual blocks. Diffusion noise is embedded into a 512-dimensional feature space by the first two linears and into 128 by the third local linear layer in each residual. The noise schedule is a 50-step linear schedule with β_ts in the range $[1e-4, 0.05]$. The training process of the model is shown in Algorithm 1. The EEG signals were generated according to Algorithm 2 that shows how a white noise sample is taken at the beginning of the process, which is gradually denoised to a synthetic ERP epoch.

We used PyTorch 1.11.0 to implement, train, and validate the model in our framework written in Python 3.8.12. Training and testing ran on an NVidia GeForce GTX 1660-Ti Max-Q (6 GB) GPU.

Algorithm 1 Training process

repeat
  $x_0 \sim q(x_0)$
  $t \sim \mathcal{U}[1, T]$
  $\epsilon \sim \mathcal{N}(0, \mathbf{I})$
  $x_t = \sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon$
  Optimizing step on $\nabla_\theta \|\epsilon - \epsilon_\theta(x_t, t)\|_2$
until converged

Algorithm 2 Sampling process

$y_T \sim \mathcal{N}(0, \mathbf{I})$
for $t = T, T-1, ..., 0$ **do**
  $y_{t-1} = \frac{1}{\sqrt{\alpha_t}}(y_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(y_t, t))$
  if $t > 0$ **then**
    $z \sim \mathcal{N}(0, \mathbf{I})$
    $y_{t-1} = y_{t-1} + \tilde{\beta}_t z$
  end if
end for
return y_0

4 Experiment

4.1 Data set

The **BCI Competition III data set II** was recorded with a BCI2000 system. Two subjects took part in the measurements, where their task was to write particular characters by using a P300 speller. The P300 speller introduced by consists of a 6x6 letter and number matrix. The rows and columns of the matrix were intensified randomly at 5.7 Hz, so the target character was brightened up twice during an intensification period. A total of 180 intensifications occurred for each character. 64 electrodes were placed on the scalp and the recorded signals were digitized at 240 Hz, followed by band-pass filtering between 0.1 and 60Hz. [18] In this work, we use data from the C3, Cz, C4, F3, Fz, F4, P3, and P4 electrodes. Samples were extracted as epochs from the data set in the range of $[0, 1]$ seconds from the onset of the intensifications. The extracted samples were down-sampled to 128 Hz and normalized by channel-wise mean subtraction and variance division.

4.2 Training

In the signal generation task, the performance of EEGWave is compared to the GAN model from [19], called WGAN in this paper for brevity. Both models must generate EEG signals with P300 components. The generated and reference signals are ERP signals triggered by visual stimuli. The models were trained unconditionally without label embedding. Both models were trained for 1000 epochs. We used Adam optimizer for EEGWave with 2e-4 learning rate and (0.99, 0.999) βs. We trained the WGAN as defined by the authors in [19]. For inference we use the exponential moving averaged weights in the case of EEGWave.

The goal is to produce synthesized signals similar to real ones. The amplitude and waveform, furthermore the frequency components of the generated and real signals must be similar. The generated signals should contain as few temporal and spectral artifacts as possible or none at all.

4.3 Metrics

The quality of the generated samples is measured by metrics that are commonly used in studies regarding generative modeling.

Inception Score (IS) was introduced in [20], who observed that the metric was correlated with human judgment in image annotation tasks. IS is defined as:

$$IS = \exp(\mathbb{E}_{x \sim p_g} D_{KL}(p(x) \| \mathbb{E}_{x' \sim p_g} p(x'))) \tag{9}$$

, where p_g is the distribution over the generated samples and $\mathbb{E}_{x' \sim p_g}$ is the marginal label distribution.

Frechet Inception Distance (FID), which was defined as a metric for generative modeling in [21], can measure both the fidelity and diversity of the

generated samples. The metric is given as:

$$FID = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{\frac{1}{2}}) \tag{10}$$

, where μ_r and μ_g are the mean vectors corresponding to real and generated data features, also Σ_r and Σ_g are the correlation matrices of the real and fake data features, respectively.

Sliced Wasserstein Distance (SWD) decomposes the Wasserstein distance (also called Earth-Mover Distance) into multiple 1D distributions and calculates the metric based on the projections to save high computational burden. The SWD is given as:

$$SWD_p = \big(\int_{\mathbb{S}^{n-1}} W_p(\pi_\theta^* P_x, \pi_\theta^* P_y)^p d\theta \big)^{\frac{1}{p}} \tag{11}$$

, where P_x and P_y are the probability distributions, $\theta \in \mathbb{S}^{n-1}$ is a unit vector that has an inner product $\pi_\theta(x) = \theta^T x$ and a marginal distribution $\pi_\theta^* P_x = P_x \cdot \pi_\theta^{-1}$. ([22])

To be able to calculate IS and FID scores, the EEGNet ([23]) classification model was used. EEGNet was trained to detect P300 components. EEGNet was chosen for this task as it achieved better results in the binary classification problem and was less prone to over-fitting the training data. The classificator was trained for 100 epochs and stopped the process with early-stop, based on the validation loss. IS was calculated based on the softmaxed outputs of EEGNet, while for FID determination the first two blocks of the model were used, the last softmax activation and the linear layer were cut off from the network.

5 Results

We present qualitative and quantitative results. Table 1 compares the performance of the models based on the synthesized signals by using the metrics given in 4.3. EEGWave could outperform WGAN in all three metrics, indicating that EEGWave could generate signals with better fidelity and higher diversity.

Table 1: Performance of WGAN and EEGWave in the EEG signal synthesis task.

Model/Source	IS↑	FID↓	SWD↓
WGAN	1.0046	1.4855	4.8696
EEGWave	**1.0050**	**1.2232**	**4.6730**

Figure 3 shows the comparison of the averaged generated and real epochs in time domain. The amplitudes are the means of the 8 channels. Based on the figure both models could capture the main temporal dynamics of the P300. The epochs from WGAN have higher amplitudes, while the epochs from EEGWave are more compressed. In our experiments, we observed that spatial convolutions

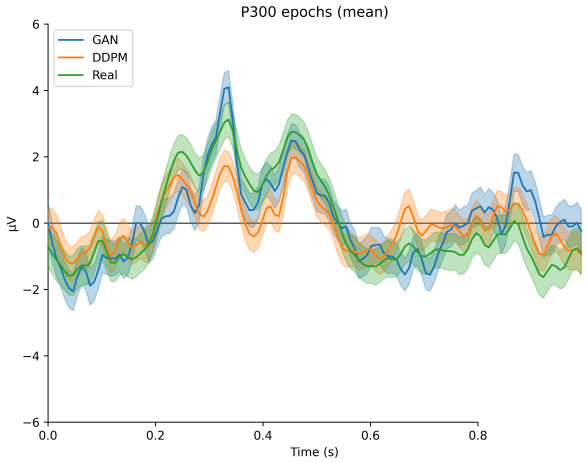

Figure 3: Amplitude of the averaged epochs.

were more effective in keeping the temporal characteristics regarding amplitude magnitude than point-wise convolutions.

Figure 4 presents the mean spectral amplitudes of the averaged epochs over 8 channels. WGAN introduced more frequency artifacts into the epochs than EEGWave. We hypothesize that these artifacts are results of the interpolations and transposed convolutions as the authors observed similar effects in their work (as well as in [6]). Epochs from EEGwave have much more similar spectral characteristics to the real ones.

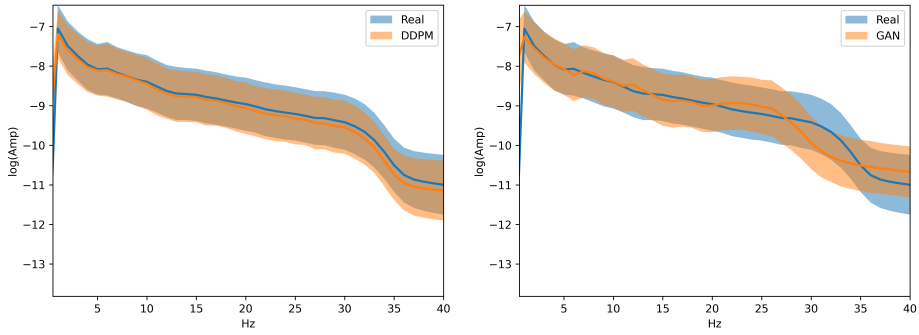

Figure 4: Spectral amplitudes of the averaged epochs.

Figure 5 presents the averaged epochs with 8 channels. The top plots show that both WGAN and EEGWave could model similar dynamics compared to the real one, although the signals from EEGWave seem to be more accurate.

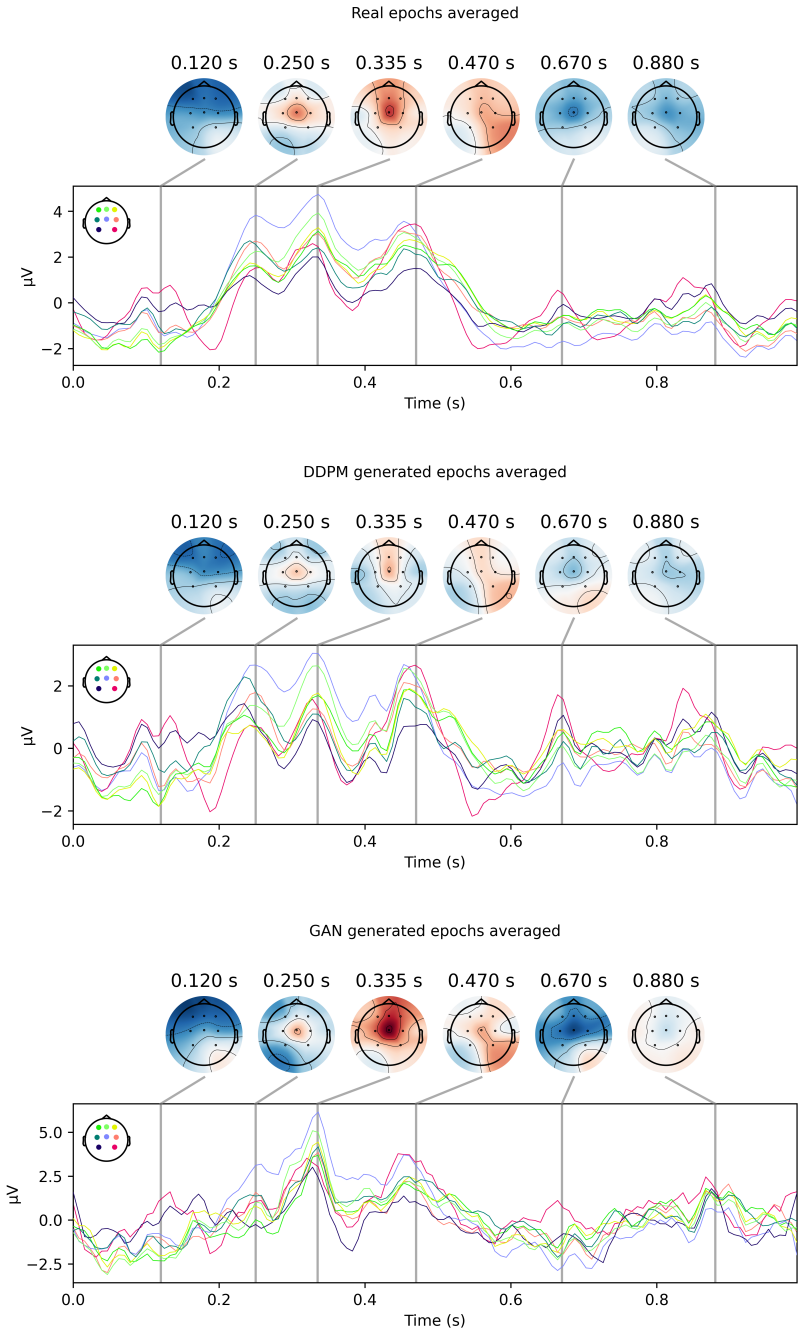

Figure 5: Averaged P300 containing signal epochs on 8 channels with top plots.

6 Conclusion

In this work, we examined a novel type of generative model, specifically the denoising diffusion probabilistic model, for the EEG signal synthesis task. A DDPM is based on a Markov chain that makes it possible to generate synthetic EEG signals from latent noise variables. We have shown that the P300 component containing EEG signals can be generated with DDPMs not only on single-channel but also on multi-channel. Furthermore, the quality of the generated signals proved to be better in our framework and setup than that of a GAN at the cost of inference time.

In future work, other architectural design options should be explored to achieve better amplitude magnitudes in the time domain. Attention modules are commonly used in other generative models, therefore these can be the starting points in addition to spatial convolutions. The optimization of the noise schedule can also result in EEG epochs of better quality. The inference speed is much slower compared to WGAN. DDIMs [24] and knowledge distillation are possible solutions for achieving better inference speeds. The application of these methods is left to our future work.

References

[1] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, p. 031005, apr 2018.

[2] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of Neural Engineering*, vol. 16, p. 031001, apr 2019.

[3] S. Haradal, H. Hayashi, and S. Uchida, "Biosignal data augmentation based on generative adversarial networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 368–371, 2018.

[4] G. M. A. S. M. Abdelfattah and M. Wang, "Augmenting the size of eeg datasets using generative adversarial networks," *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, 2018.

[5] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *International Conference on Learning Representations*, 2019.

[6] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, "Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals," *CoRR*, vol. abs/1806.01875, 2018.

[7] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *CoRR*, vol. abs/1812.04948, 2018.

[8] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *CoRR*, vol. abs/2105.05233, 2021.

[9] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, eds.), vol. 33, pp. 6840–6851, Curran Associates, Inc., 2020.

[10] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro, "Diffwave: A versatile diffusion model for audio synthesis," in *International Conference on Learning Representations*, 2021.

[11] M. K. Islam, A. Rastegarnia, and Z. Yang, "Methods for artifact detection and removal from scalp eeg: A review," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 46, no. 4, pp. 287–305, 2016.

[12] G. Mele, C. Cavaliere, V. Alfano, M. Orsini, M. Salvatore, and M. Aiello, "Simultaneous eeg-fmri for functional neurological assessment," *Frontiers in Neurology*, vol. 10, p. 848, 2019.

[13] T. Zama, Y. Takahashi, and S. Shimada, "Simultaneous eeg-nirs measurement of the inferior parietal lobule during a reaching task with delayed visual feedback," *Frontiers in Human Neuroscience*, vol. 13, p. 301, 2019.

[14] M. Teplan, "Fundamental of eeg measurement," *MEASUREMENT SCIENCE REVIEW*, vol. 2, 01 2002.

[15] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proceedings of the 32nd International Conference on Machine Learning* (F. Bach and D. Blei, eds.), vol. 37 of *Proceedings of Machine Learning Research*, (Lille, France), pp. 2256–2265, PMLR, 07–09 Jul 2015.

[16] J. Lee and S. Han, "Nu-wave: A diffusion probabilistic model for neural audio upsampling," *Interspeech 2021*, Aug 2021.

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[18] D. K. Jonathan R. Wolpaw, Gerwin Schalk, "Wadsworth bci dataset (p300 evoked potentials)." `https://www.bbci.de/competition/iii/`, 2004.

[19] S. Panwar, P. Rad, T.-P. Jung, and Y. Huang, "Modeling eeg data distribution with a wasserstein generative adversarial network to predict rsvp events," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 8, pp. 1720–1730, 2020.

[20] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems* (D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, eds.), vol. 29, Curran Associates, Inc., 2016.

[21] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[22] J. Wu, Z. Huang, W. Li, and L. V. Gool, "Generative autotransporters," *CoRR*, vol. abs/1706.02631, 2017.

[23] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brainâcomputer interfaces," *Journal of Neural Engineering*, vol. 15, p. 056013, Jul 2018.

[24] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *International Conference on Learning Representations*, 2021.