DOMAIN SPECIFIC DENOISING DIFFUSION PROBABILISTIC MODELS FOR BRAIN DYNAMICS

Anonymous authors
Paper under double-blind review

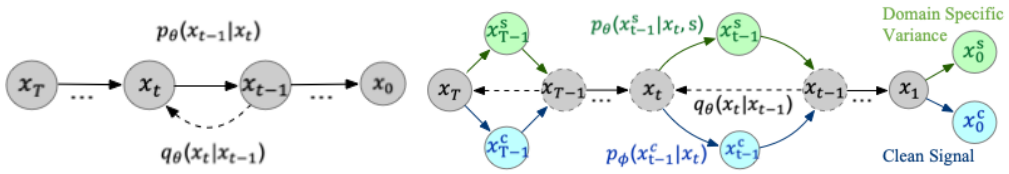
ABSTRACT

The distribution differences in brain dynamics according to different human subjects, which is a kind of human-subject noise, as referred to human artifacts, have severely limited the generalized ability of brain dynamics recognition. Previous human artifacts removal methods normally utilize traditional spectrum filtering or blind Source Separation techniques, based on a simple assumption of prior distributions, which limit the capacity of learning domain variance of each subject. We propose a new approach to model the human artifacts removal as generative denoising process, which can generate and learn subject-specific domain variance and the invariant brain signals, simultaneously. We propose Domain Specific Denoising Diffusion Probabilistic Model (DS-DDPM) to decompose the denoising process into the subject domain variance and invariant content at each step. Subtle constraints and probabilistic design are proposed to formulate domain variance and invariant content into orthogonal spaces and further supervise the domain variance with the subject classifier. This method is the first work to explicitly separate human subject-specific variance through generative denoising processes, which outperforms previous methods in two aspects, 1) DS-DDPM could learn more accurate subject-specific domain variance by domain generative learning rather than previous filtering methods 2) DS-DDPM is the first work could explicitly generate subject noise distribution. Comprehensive experimental results suggest that DS-DDPM could help alleviate domain distribution bias for cross-domain brain dynamics signal recognition. Code is available in https://anonymous.4open.science/r/DS-DDPM-2EA6

1 INTRODUCTION

The recognition of human brain dynamics signals such as electroencephalogram (EEG) (Nunez et al., 2006) and Event-related Potential (ERP) (Picton et al., 1995) is of vital importance for the non-invasive brain-computer interface. However, the brain signal distribution from different human subjects exhibits severe distributional differences (Henry, 2006; Jiang et al., 2019), which means the recognition model (Lawhern et al., 2018; Duan & Feng, 2019) trained for a set of human subjects might not be efficient for other unknown human subjects. This weakens the generalized ability of deep learning-based models.

Previous human artifacts removal (Jiang et al., 2019) efforts are introduced to alleviate the above problem, which could be categorized into four-fold 1) Regression (Al-Nuaimi et al., 2018), 2) Blind Source Separation (BSS) (Sweeney et al., 2012; Somers & Bertrand, 2016), Empirical-mode Decomposition (EMD), and Wavelet Transform algorithm to their hybrid methods (James & Hesse, 2004). Among these methods, Regression and BSS are two widely used approaches. Classic regression methods (Sweeney et al., 2012) are applied under the assumption that each channel is the cumulative sum of clean EEG data and a proportion of artifact given by known reference signals (Hillyard & Galambos, 1970; Wallstrom et al., 2002) through a set of regressed transmission factors. Yet, the previous methods only learn transmission factors rather than directly learn the domain variance. For BSS methods, Independent Component Analysis (ICA) decompose observed signal into independent components (ICs) (Somers & Bertrand, 2016) from linear mixtures of cerebral and artifactual sources. However, current ICA methods still require reference signal information to perform artifacts separation, which have not considered the domain variance between different human subjects. To be concluded, previous human artifacts removal methods normally assumes domain



(a) DDPM (b) DS-DDPM

Figure 1: Illustration of probabilistic model graphs, where Figures 1(a) and 1(b) respectively denote the original model DDPM and the proposed DS-DDPM. DS-DDPM separate the subject-specific domain variance $\mathbf{x}_t^s$ (human artifacts), and the clean signal $\mathbf{x}_t^c$ at each denoising time step t. We constrain the summation of the separated $\mathbf{x}_t^s$ and $\mathbf{x}_t^c$ equal to diffusion result $\mathbf{x}_t$ at time step t, which denotes the two separated variance shares the same diffusion process. Here, the dotted line denotes the diffusion process, while the solid line denotes the denoising process.

variance as simple prior distributions. It limited the capacity of modeling domain variance of each subject. Our intuition is to introduce learned subject specific domain variance rather than simple prior distribution. We propose to modeling the human artifacts removal as domain specific denoising process, and perform denoising by generate subject-specific domain variance and the invariant brain signals through the denoising process.

In this paper, we propose to revisit the human artifacts removal for EEG signals by modeling the removal as domain-specific denoising diffusion probabilistic models (DS-DDPM). We follow the widely accepted assumption that the recorded EEG signals are mixtures of clean samples and *subject-specific human artifact noises*. We propose a novel probabilistic model based on the diffusion-denoising generative process (Nichol & Dhariwal, 2021; Choi et al., 2021), which has been the recent research spotlight, especially for image generation tasks (Ramesh et al., 2021). The proposed DS-DDPM formulate the artifact removal into every denoising step in the original model as shown in Figure 1, where the diffusion result $\mathbf{x}_t$ at time step t is separated by two denoising models respectively into domain variance $\mathbf{x}_t^s$ (human artifacts), and the clean signal $\mathbf{x}_t^c$. We formulate the diffusion result is the summation of two separated content as $\mathbf{x}_t = \mathbf{x}_t^c + \mathbf{x}_t^s$, which means the two stream shares the same diffusion process. This constraint naturally stands as our assumption is that the noisy EEG signal could be decomposed into clean signals and domain-specific noise, where Section 2.1 give a detailed problem definition mathematically. We decompose the domain variance and clean signal into two mutually orthogonal spaces, as described in Section 2.2. Moreover, an auxiliary cosine classifier is applied to disentangling variance spaces for subject variance noises. As none of the previous methods has explored the DDPM models for EEG signal, Section 2.3 give technical details of the efficient structure of DDPM for EEG signals. Comprehensive experimental results are provided to illustrate that 1) the domain-specific separation is significant in subject-domain-wise (Section 3.2), 2) domain-specific denoising could help improve the cross-subject classification performance (Section 3.3). We also provide an ablation study to discuss the contribution of each proposed component in Section 3.4. The contribution of this paper could be categorized threefold as below.

- This work is the first to introduce denoising diffusion probabilistic models into EEG signals, where UNet-EEG is first proposed for long time-sequence sample generation.
- DS-DDPM give a novel approach to explicitly separate long existing domain variance related to human subject difference through domain specific denoising process.
- Comprehensive Experimental results suggest that the proposed DS-DDPM is efficient in domain-specific denoising for both relevance analysis and classification performance.

2 METHODOLOGY

Section 2.1 provide the mathematical description of how we revisit the human artifact removal by formulating it as a domain-specific denoising diffusion probabilistic process. Section 2.2 provides technical details about how we constrain and separate the two variables through novel training loss. Section 2.3 illustrate how we design the model structure to fit the properties of EEG signals.

2.1 DEFINITION OF DOMIAN-SEPCIFIC DENOISING

The separation of domain-specific variance is decomposed into denoising process of a denoising diffusion probabilistic model (Nichol & Dhariwal, 2021). We define $\mathbf{x}_0$ as the original recorded EEG signals and $\mathbf{x}_t$ as the variable by adding Gaussian noise distribution by sequentially t times iteration. Thus, we could continuously add noise into original $\mathbf{x}_0$ through a Markov process sampling variables $\{\mathbf{x}_0, \mathbf{x}_1, ...\mathbf{x}_{t-1}, \mathbf{x}_t, ..., \mathbf{x}_T\}$ until $\mathbf{x}_T$ becomes a normal noise distribution $p(\mathbf{x}_T) \sim \mathcal{N}(\mathbf{x}_T; 0, I)$ as shown Figure 1. Here, the transition is also called *diffusion process or forward process* as below.

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \quad q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t I) \tag{1}$$

where $\beta_1, \beta_2, ...\beta_T$ is a fixed variance coefficient schedule. Follow the Gaussian distribution assumption of DDPM, $\mathbf{x}_t$ could further represented as combination of $\mathbf{x}_0$ and sampled variance $\boldsymbol{\varepsilon}$.

$$q(\mathbf{x}_t|\mathbf{x}_0) := \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}), \quad \alpha_t := 1 - \beta_t, \quad \bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i \tag{2}$$

where α_t is also a fixed variance coefficient schedule corresponding to β_t. In practice, the representation of $\mathbf{x}_t$ could be obtained by extending the diffusion process defined in Equation 2 as below.

$$\mathbf{x}_t = \alpha_t\mathbf{x}_{t-1} + \beta_t\boldsymbol{\varepsilon}_t = \alpha_t\left(\bar{\alpha}_{t-1}\mathbf{x}_0 + \bar{\beta}_{t-1}\bar{\boldsymbol{\varepsilon}}_{t-1}\right) + \beta_t\boldsymbol{\varepsilon}_t = \bar{\alpha}_t\mathbf{x}_0 + \alpha_t\bar{\beta}_{t-1}\bar{\boldsymbol{\varepsilon}}_{t-1} + \beta_t\boldsymbol{\varepsilon}_t \tag{3}$$

where $\boldsymbol{\varepsilon_t} \sim \mathcal{N}(0, \mathbf{I})$ is a gaussian distribution which represent the stochastic property to the diffusion process. It also give description of how to represent the diffusion result $\mathbf{x}_t$ by real sample $\mathbf{x}_0$ and given fixed variance scheduler α_t and β_t.

Different from the classical DDPMs which model a direct *reverse process or denoising process* as $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$, we decompose the denoising process subject to different subject domain s by separate the process into to two variables, $q(\mathbf{x}^s_{t-1}, \mathbf{x}^c_{t-1}|\mathbf{x}_t)$. Here, $\mathbf{x}^s_{t-1}$ denotes the separated domain variance of subject s and $\mathbf{x}^c_{t-1}$ denotes the compared clean signal generated at step t, where $\mathbf{x}_{t-1} = \mathbf{x}^s_{t-1} + \mathbf{x}^c_{t-1}$ as we assume the signal is the mixture of subject noise and the clean signal. As the direct reverse of the diffusion process $q(\mathbf{x}^s_{t-1}, \mathbf{x}^c_{t-1}|\mathbf{x}_t)$ is intractable (Sohl-Dickstein et al., 2015), we use two separate function to express the denoising process as below.

$$p_\theta(\mathbf{x}^s_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}^s_{t-1}; \mu_\theta(\mathbf{x}_t, t, s), \sigma_t^2\mathbf{I}), \quad p_\phi(\mathbf{x}^c_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}^c_{t-1}; \mu_\phi(\mathbf{x}_t, t), \sigma_t^2\mathbf{I}) \tag{4}$$

where $p_\theta(\mathbf{x}^s_{t-1}|\mathbf{x}_t)$ and $p_\phi(\mathbf{x}^c_{t-1}|\mathbf{x}_t)$ are the two seperated denoising functions decomposed from $p_(\mathbf{x}_{t-1}|\mathbf{x}_t) = p_\theta(\mathbf{x}^s_{t-1}|\mathbf{x}_t) + p_\phi(\mathbf{x}^c_{t-1}|\mathbf{x}_t)$. Here, σ_t^2 denote the variance in transition. The core transition $\mu_\theta(\mathbf{x}_t, t, s)$ and $\mu_\phi(\mathbf{x}_t, t)$ is learned by deep neural networks. We follow previous experimental settings (Nichol & Dhariwal, 2021; Choi et al., 2021) that σ_t^2 is directly set as β_t or $\frac{1-\bar{\alpha}_t}{1-\alpha_t}\beta_t$, which suggest similar results in previous experiments. Thus, the variable $\mathbf{x}_t$ at time step t could be expressed as the summation of domain variance, and the clean data follow the original DDPM conduction (Nichol & Dhariwal, 2021).

$$\mathbf{x}_{t-1} = \frac{1}{1-\alpha_t}(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}(\mu_\theta(\mathbf{x}_t, t, s) + \mu_\phi(\mathbf{x}_t, t)) + \sigma_t\mathbf{z}, \tag{5}$$

where the two learned transition deep model $\mu_\theta(\mathbf{x}_t, t, s)$ and $\mu_\phi(\mathbf{x}_t, t)$ share a same variance coefficient schedule corresponding to β_t. Given relation we defined above $\mathbf{x}_{t-1} = \mathbf{x}^s_{t-1} + \mathbf{x}^c_{t-1}$, we could further approximate the separated domain variance of subject s as $\mathbf{x}^s_{t-1} = \frac{1}{1-\alpha_t}(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\mu_\theta(\mathbf{x}_t, t, s)) + \sigma_t\mathbf{z}$, and the separated clean signal as $\mathbf{x}^c_{t-1} = \frac{1}{1-\alpha_t}(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\mu_\phi(\mathbf{x}_t, t)) + \sigma_t\mathbf{z}$. It is noted that, after the denoising process sampled to $\mathbf{x}_1$, we could directly calculate the desired domain variance $\mathbf{x}^s_0$ of human subjected s and clean data $\mathbf{x}^c_0$ according to the equations defined above.

2.2 SEPARATE DOMAIN SPECIFIC VARIANCE BY CONSTRAINT

Under the domain specific denoising process defined in Section 2.1, we separate the domain specific variance by the combination of three constraint. 1) Section 2.2.1 introduce how we constraint the summation separated clean signal and domain specific variance could reconstruct the EEG signals by reverse process loss $\mathcal{L}_r$. 2) Section 2.2.2 introduce orthogonal constraint between clean signal and domain specific variance by orthogonal loss $\mathcal{L}_o$. 3) Section 2.2.3 further formulate the domain variance space seperable according to different human subjects by Arc-Marging loss $\mathcal{L}_{arc}$.

2.2.1 REVERSE PROCESS

Basically, we modeling the diffusion results $\mathbf{x}_t$ at time step t as the mixture of domain variance $\mathbf{x}_t^s$ and clean data $\mathbf{x}_t^c$ as defined in Section 2.1 and Figure 1(b). We first discuss how to ensure the effectiveness of diffusion-denoising training. Given the decomposition the of denoising results as $\mathbf{x}_t = \mathbf{x}_t^s + \mathbf{x}_t^c$, generative training loss is granted by minimizing the distance $\|\mathbf{x}_{t-1} - (\mu_\theta(\mathbf{x}_t, t, s) + \mu_\phi(\mathbf{x}_t, t))\|^2$ between diffusion results $\mathbf{x}_t$ and denoising results. Follow the original conduction of DDPM (Nichol & Dhariwal, 2021), Considering the definition in Section 2.1 that the seperated two denoising models share the same variance coefficient scheduler α and β we could approximate the summation of the denoising process as.

$$\mu_\theta(\mathbf{x}_t, t, s) + \mu_\phi(\mathbf{x}_t, t) = \frac{1}{\alpha_t}(\mathbf{x}_t - \beta_t \varepsilon_\theta(x_t, t, s) - \beta_t \varepsilon_\phi(x_t, t)) \tag{6}$$

where θ and ϕ are respectively the training parameter for domain specific denoising and content denoising. ε_θ and ε_ϕ denote the generated variance from these models. Thus by introducing Equation 6, we could rewrite the distance to be minimized as in Equation 7.

$$\|\mathbf{x}_{t-1} - (\mu_\theta(\mathbf{x}_t, t, s) + \mu_\phi(\mathbf{x}_t, t))\|^2 = \frac{\beta_t^2}{\alpha_t^2}\|\varepsilon_t - \varepsilon_\theta(x_t, t, s) - \varepsilon_\phi(x_t, t))\|^2 \tag{7}$$

where $\frac{\beta_t^2}{\alpha_t^2}$ is a loss coefficient which we use hyper parameter λ_r to represent. Also by introducing the diffusion process to represent $\mathbf{x}_t$ defined in Equation 3, we could give the training loss $\mathcal{L}_r$ for the reverse process in Equation 8.

$$\mathcal{L}_r = \lambda_r\|\varepsilon_t - \varepsilon_\theta(\bar{\alpha}_t\mathbf{x}_0 + \alpha_t\bar{\beta}_{t-1}\bar{\varepsilon}_{t-1} + \beta_t\varepsilon_t, t, s) - \varepsilon_\phi(\bar{\alpha}_t\mathbf{x}_0 + \alpha_t\bar{\beta}_{t-1}\bar{\varepsilon}_{t-1} + \beta_t\varepsilon_t, t))\|^2 \tag{8}$$

The reverse process training loss $\mathcal{L}_r$ could be minimized by given recorded EEG signals $\mathbf{x}_0$, fixed variance scheduler $\{\alpha_{1:T}, \beta_{1:T}\}$ and standard Gaussian distribution ε_t sampled at each time step.

2.2.2 SEPARATE SUBJECT DOMAIN VARIANCE APART

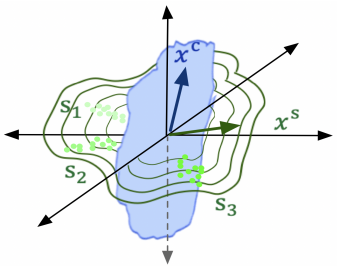


Figure 2: The orthogonal domain variance separation of DS-DDPM.

The training of the reverse process only ensures the summation of the generated $\mathbf{x}_0^s$ and $\mathbf{x}_0^c$ is the recorded EEG signal $\mathbf{x}_0$. Yet, we impose two constraints to separate the domain variance. First, we formulate the generated $\mathbf{x}_0^s$ and $\mathbf{x}_0^c$ at each time step into two orthogonal spaces as shown in Figure 2, where the blue space contains the signal distribution of the clean EEG signals, green space contains the distribution of the subject-specific domain variance. The orthogonal property is granted by introducing a constraint loss defined in Equation 9.

$$\mathcal{L}_o = \lambda_o \left\| \left(\mu_\theta(\mathbf{x}_t, t, s)^\top \mu_\phi(\mathbf{x}_t, t) - \boldsymbol{I}\right) \otimes (1 - \boldsymbol{I}) \right\|^2 \tag{9}$$

where $\mu_\theta(\mathbf{x}_t, t, s)$ and $\mu_\phi(\mathbf{x}_t, t)$ are the denoising process defined in Equation 6. According to this representation, we make an approximation to directly optimize the model output $\varepsilon_\theta(x_t, t, s)^\top \varepsilon_\phi(x_t, t)$ instead, the accordingly adjustment of the coefficient λ_o regularize the value scale of $\mathcal{L}_o$.

2.2.3 FORMULATE SUBJECT DOMAIN VARIANCE TO HUMAN SUBJECTS

In order to perform domain specific denoising related to given human subject s, DS-DDPM also constraint the domain variance separable in subject wise. The constraint is constructed by introducing a extra subject classifier to supervise conditional denoising process $\mu_\theta(\mathbf{x}_t, t, s)$. Here, the generated domain variance of subject s is expected to predict the subject label s given the domain variance by introducing a conditional probabilistic model $p(s|\mathbf{x}_t^s)$. Also, under our assumption, we design the clean data variance totally independent to the subject information. Thus, we could eliminate all items in Equations defined above for representation of the classification model $p(s|\mathbf{x}_t^s)$. We use a simple EEGNet (Lawhern et al., 2018) classifier to extract feature $\bar{\mathbf{x}}_t^{s_i}$ over denoisig output $\mathbf{x}_t^s$, where the transition is represented as $\mathbf{W}_\theta$. We propose to formulate the decision boundary of predicting subject s into cosine space, where the traditional SoftMax (Liu et al., 2016) loss is replaced by introducing

Additive Angular Margin classification (Arc-Margin) loss (Deng et al., 2019). The Arc-Marging loss transforms the classification logits as $\mathbf{W}_{\theta,s_j}^{\top}\bar{\mathbf{x}}_t^{s_i} = \|\mathbf{W}_{\theta,s_j}\|\|\bar{\mathbf{x}}_t^{s_i}\|\cos\theta_{s_j}$, where the θ_{s_j} is the angle between the weight $\mathbf{W}_{\theta,s_j}$ and the feature $\bar{\mathbf{x}}_t^{s_i}$ representing subject s_i. The individual weight and features are fixed by L_2 normalization. Here we let $\|\mathbf{W}_{\theta,s_j}\| = 1$, which denotes the weights make the predictions only depend on the angle between the feature and the weight We also let $\|\bar{\mathbf{x}}_t^{s_i}\| = r$, which denotes the learned embedding features are thus distributed on a hyper-sphere with a radius of r. As the feature $\bar{\mathbf{x}}_t^{s_i}$ is calculate by $\mu_\theta(\mathbf{x}_t, t, s)$ through bunch of classification transformation ω. Thus, $\bar{\mathbf{x}}_t^{s_i}$ could be approximated by eliminate irrelevant items in Equation 6 as below.

$$\cos\theta_{s_j} = \frac{\mathbf{W}_{\theta,s_j}^{\top}\omega(\boldsymbol{\epsilon}_\theta(\bar{\alpha}_t\mathbf{x}_0 + \alpha_t\bar{\beta}_{t-1}\bar{\boldsymbol{\epsilon}}_{t-1} + \beta_t\boldsymbol{\epsilon}_t, t, s))}{\|\mathbf{W}_{\theta,s_j}^{\top}\|\|\omega(\boldsymbol{\epsilon}_\theta(\bar{\alpha}_t\mathbf{x}_0 + \alpha_t\bar{\beta}_{t-1}\bar{\boldsymbol{\epsilon}}_{t-1} + \beta_t\boldsymbol{\epsilon}_t, t, s))\|} \tag{10}$$

An additive angular margin penalty m is added between $\bar{\mathbf{x}}_t^{s_i}$ and $\mathbf{W}_{\theta,s_j}$ to simultaneously enhance the intra-class compactness and inter-class discrepancy (Deng et al., 2019). Also by given the coefficient λ_{arc} for subject classification, we could formulate the Arc-Margin loss $\mathcal{L}_{arc}$ coulde be defined in Equation 11.

$$\mathcal{L}_{arc} = -\lambda_{arc}\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{r\left(\cos\left(\theta_{y_s^i}+m\right)\right)}}{e^{r\left(\cos\left(\theta_{y_s^i}+m\right)\right)} + \sum_{s_j=1, s_j\neq y_s^i}^{n} e^{r\cos\theta_{s_j}}} \tag{11}$$

where y_s^i denotes the target subject label. By introducing $\mathcal{L}_{arc}$, we could regulate the domain variance space separable by angle according to different subject s_i. Since we already regulate the clean data space and the subject space orthogonal, the mixture of different subjects will be distributed evenly according to different subjects s_i as well, which increase the interoperability of DS-DDPM.

2.3 Denoising Model for EEG Signals

Model Structure: This section introduce how we design domain specific generative model structure and how we train the denoising model by combining the loss constrained proposed in Section 2.2. For model structure, we follow the common UNet (Huang et al., 2020) structure and modified it to fit long-time series signals (UNet-EEG). The overall model structure is shown in Figure 3, where we modified the UNet (Huang et al., 2020) structure into two generative streams. The UNet-EEG model

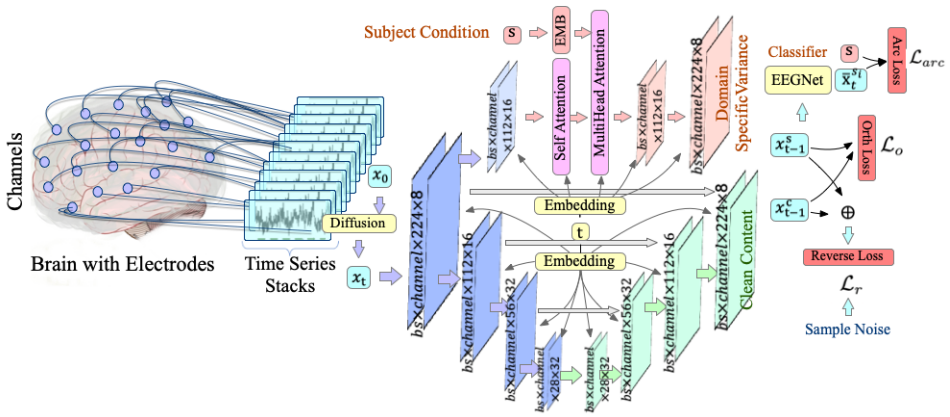


Figure 3: The overall model structure of DS-DDPM, where the recorded EEG signals is sliced by time window 224, step 75 and stacked into shape $bs \times channel \times window \times stacks$ as $\mathbf{x}_0$. The modified UNet-EEG structure receives diffusion results $\mathbf{x}_t$ and respectively generate domain variance $\mathbf{x}_{t-1}^s$ and the clean data $\mathbf{x}_{t-1}^c$ by two streams. The generated $\mathbf{x}_{t-1}^s$ and $\mathbf{x}_{t-1}^c$ are simultaneously supervised by reverse loss $\mathcal{L}_r$, orthogonal loss $\mathcal{L}_o$ and Arc-Margin loss $\mathcal{L}_r$.

receives diffusion results $\mathbf{x}_t$ and respectively generate domain variance $\mathbf{x}_{t-1}^s$ and the clean content $\mathbf{x}_{t-1}^c$. Here for clean content stream, the UNet-EEG sequentially perform three times down-sample and up-sample operation with the residual fusion within each feature map scale, which could enhance the ability of the model of maintaining time sequential relations. Similar to original DDPM, time label

t is tokenized into time embedding and fed into each layer of the UNet-EEG to improve the generative ability according to each time step t. For the domain variance separation stream, the subject condition s is tokenized into embedding and fused with UNet-EEG at the mid layer by multi-head attention layers. Here, we take the subject token as the query and the original feature map as key and value for the attention layer.

Training and Sampling Procedure: Given the structure defined above, we provide detailed training procedure as shown in Algorithm 1. For fast converging, we pre-train the subject classifier ω (EEGNEt) by sampling $\{\mathbf{x}_0, s\} \sim q(\mathbf{x}_0, s)$ pairs from the recorded EEG signals of set of human subjects. Since EEG signals naturally contains the subject label (it means from which human the EEG signal is collected), the training of ω is practical. It could help the optimizer concentrate on the parameter θ, ϕ for generative model. For the whole training procedure, the time step t is sampled evenly between $\{1, 2, .., T\}$ and variance scheduler $\{\alpha, \beta\}$ is sampled and fixed. At each iteration, the algorithm sampling $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ at each time step. During each iteration, the generative model for clean data and subject-specific domain variance is calculated by giving the subject label s, time step t, real EEG signal $\mathbf{x}_0$ and variables sampled mentioned above. The parameters θ, ϕ of the generative model is optimized by minimizing the weighted combination of the loss function proposed in Section 2.2. After the generative model is trained, we could generate subject-specific

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0, s \sim q(\mathbf{x}_0, s)$, Sample $\mathbf{x}_0$ from Subjects $\{s\}$
3: Pretrain subject classifier ω given sample label pairs $\{\mathbf{x}_0, s\}$
4: $t \sim \text{Uniform}(\{1, \dots, T\})$
5: $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
6: Calculate, $\varepsilon_\theta(x_t, t, s)$, $\varepsilon_\phi(x_t, t)$, $\omega(\varepsilon_\theta(x_t, t, s))$ with model parameters θ and ϕ.
7: Take gradient descent step optimize θ and ϕ on
8: Combination Loss $\mathcal{L} = \lambda_r \mathcal{L}_r + \lambda_o \mathcal{L}_o + \lambda_{arc} \mathcal{L}_{arc}$
9: **until** converged

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \dots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}^s_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \varepsilon_\theta(\mathbf{x}_t, t, s) \right) + \sigma_t \mathbf{z}$
5: $\mathbf{x}^c_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \varepsilon_\phi(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
6: **end for**
7: **return** $\mathbf{x}^s_0, \mathbf{x}^c_0$

domain variance according to the probabilistic model discussed in Section 2.1. Algorithm 2 denotes a complete procedure of sampling domain variance according to subject s and generated unconditional EEG signals from pure noise. One natural process of perform denoising on give signal is that, sampling a subject specific domain variance $\mathbf{x}^s_0$, and direct perform $\mathbf{x}_0 - \mathbf{x}^s_0$ to get cleaner signal. However, this approach does not link the noise according to the given signal. Meanwhile, the iterative denoising process is time-consuming. For real-life scenarios, it's more convenient to perform denoising not only depending on subject s but also the signal itself. Revisiting the denoising iterative process $\mathbf{x}_T, ..., \mathbf{x}_t, \mathbf{x}_{t-1}, ...\mathbf{x}_0$, the denoising model remove the noise by small steps each time. As recorded EEG signal naturally contains noise, it's also reasonable that we direct perform denoising by assume recorded signal as $\mathbf{x}_{1/2}$, a mediate state in the denoising process. Thus, given recorded raw signal is $\mathbf{x}_{raw}$ we could directly separate the domain variance and the clean content by performing $\mathbf{x}^s_0 = \frac{1}{\sqrt{\alpha_1}} \left(\mathbf{x}_{raw} - \frac{1-\alpha_1}{\sqrt{1-\bar{\alpha}_1}} \varepsilon_\theta(\mathbf{x}_{raw}, 1, s) \right) + \sigma_1 \mathbf{z}$ and $\mathbf{x}^c_0 = \frac{1}{\sqrt{\alpha_1}} \left(\mathbf{x}_{raw} - \frac{1-\alpha_1}{\sqrt{1-\bar{\alpha}_1}} \varepsilon_\theta(\mathbf{x}_{raw}, 1) \right) + \sigma_1 \mathbf{z}$. This assumption could replace the iterative sampling process with inference only once. Experimental results suggest that this assumption is efficient in improving the cross subject classification task.

3 Experiments

In this section, we conduct comprehensive experiments to illustrate the efficiency of the proposed Domain-Specific Denoising, where Section 3.2 analysis the correlation by 1) performing correlation coefficient Moreover, we conduct experiments on cross subject classification task in Section 3.3 to illustrate our efficiency on classification tasks. Additional ablation study is conducted in Section 3.4 to discuss the effectiveness of each component.

3.1 Experimental Setup

We mainly conduct our experiments on BCI-Competition-IV dataset (Tangermann et al., 2012), which is widely used to validate Motor Imaginary classification tasks. The dataset is collected under a widely-used 10-20 system, where 22 EEG channels and 3 EOG channels are provided. The dataset contains EEG and EOG signals with a sampling frequency of 250 Hz from nine subjects. Each human subject was required to perform four classes (left hand, right hand, feet, and tongue) motor imaginary

while recording the brain dynamics. We only use the 22-channel EEG signals by pre-processing them into shape 22×750, where 750 is the time sequence length with a 250 Hz sampling rate for 3 seconds. As mentioned in Section 2.3, we use a time window to slide along the time sequence and slice time sequence to segments with overlaps. Here by using time window size 224 and stride size 75, the data is processed into shape $bs \times 22 \times 224 \times 8$ and fed into the UNet-EEG model. For the UNet-EEG model, we sequentially perform 3 times down-sampling and up-sampling blocks for the clean content stream and perform 1 times down-sampling and up-sampling for the domain variance stream. This compared slim structure for domain variance stream prevents severe over-fitting, thus making the training process more stable. For the subjects classifier, we use a original EEGNet structure with modification of input size $bs \times 22 \times 224 \times 8$ as well. For the hyper-parameters of Arc-Margin $\mathcal{L}_{arc}$, we follow the previous explorations on human face recognition where we respectively set the radius $r = 30$ and the margin as $m = 0.5$. The ADAM (Zhang, 2018) with default hyper-parameters is utilized as our optimizer. According to our experiments, we found batch size 64 could stabilize the training process than smaller numbers. The DS-DDPM is implemented based on PyTorch, where we release our code through anonymous link https://anonymous.4open.science/r/DS-DDPM-2EA6.

Table 1: Subject-wise correlation analysis inside between real EEG samples and DS-DDPM generative samples. It shows that the correlation coefficient of the same subject between real and generated sample is significantly higher than un-related subjects, which indicates the efficiency of DS-DDPM.

Subject Correlation Coefficient of EEG BCI-IV EEG Signals									
	s1	s2	s3	s4	s5	s6	s7	s8	s9
s1	**0.1013**	0.0481	0.0546	0.0593	0.0450	0.0508	0.0502	0.0424	0.0431
s2	0.0481	**0.0744**	0.0401	0.0384	0.0333	0.0406	0.0451	0.0373	0.0378
s3	0.0546	0.0401	**0.0804**	0.0483	0.0365	0.0430	0.0483	0.0372	0.0394
s4	0.0593	0.0384	0.0483	**0.0876**	0.0409	0.0423	0.0529	0.0433	0.0438
s5	0.0450	0.0333	0.0365	0.0409	**0.0576**	0.0343	0.0387	0.0312	0.0341
s6	0.0508	0.0406	0.0430	0.0423	0.0343	**0.0713**	0.0444	0.0339	0.0376
s7	0.0502	0.0451	0.5483	0.0529	0.0387	0.0444	**0.0929**	0.0424	0.0498
s8	0.0424	0.0373	0.0372	0.0433	0.0312	0.0539	0.0424	**0.0687**	0.0321
s9	0.0498	0.0378	0.0394	0.0413	0.0341	0.0376	0.0438	0.0321	**0.0660**
Subject Correlation Coefficient Between DS-DDPM Sampled Signal and BCI-IV									
s1	**0.1089**	0.0711	0.0772	0.0797	0.0668	0.0731	0.0715	0.0634	0.0695
s2	0.0716	**0.0897**	0.0670	0.0645	0.0580	0.0660	0.0682	0.0595	0.0614
s3	0.0759	0.0652	**0.0960**	0.0721	0.0610	0.0680	0.0702	0.0596	0.0626
s4	0.0793	0.0637	0.0729	**0.1000**	0.0638	0.0678	0.0739	0.0637	0.0635
s5	0.0690	0.0603	0.0646	0.0666	**0.0757**	0.0617	0.0633	0.0553	0.0586
s6	0.0832	0.0655	0.0689	0.0673	0.0588	**0.0883**	0.0675	0.0571	0.0609
s7	0.0725	0.0689	0.0789	0.0753	0.0696	0.0691	**0.1025**	0.0659	0.0656
s8	0.0673	0.0636	0.0651	0.0686	0.0574	0.0619	0.0664	**0.0821**	0.0574
s9	0.0724	0.0637	0.0666	0.0670	0.0589	0.0647	0.0671	0.0559	**0.0816**

3.2 Domain Specific Variance Analysis

Correlation analysis: In order to analyse the generative quality of the domain variance generative tasks, we conduct correlation analysis on both real signals and generative signals, as shown in Table 1. The upper part reports the correlation coefficient matrix on the real EEG signal distribution from BCI-Competition-IV 2a datasets. It shows that the coefficient between the same subjects is significantly higher than cross subjects' coefficient, where the diagonal values of the matrix are significantly larger than others. By giving the subject label to the DS-DDPM model, we could also generate real EEG signals of each subject. In order to illustrate the domain variance separation according to subject s, we analysis the coefficient between real samples and DS-DDPM generated samples. If the generated signal distribution of subject s has higher to the real signal distribution of the same human subject s, we think is significant to illustrate our efficiency. The results are reported on lower part of Table 1, where the correlation coefficient properties is similar to the real samples. These results indicate the efficiency of DS-DDPM. We could also observe that the generated signals have higher correlation coefficients with unrelated subjects compared to real signals, where the value of generated signals varies between $0.0312 \sim 0.0593$. In comparison, the value of generated signals varies between $0.0695 \sim 0.0832$. We argue that this phenomenon is rational as the real sample is actually the upper bound of the current regression-based domain separation methods.

Visualization of Domain Variance Distribution: We also visualize the domain variance distribution of each subject. We sample domain variance by given different subject labels to each subject. Then the domain variance is visualized by introducing t-Distributed Stochastic Neighbor Embedding (t-SNE) (Belkina et al., 2019) algorithm. We compare the visualization of original EEG signal $\mathbf{x}_0$, the separated domain variance $\mathbf{x}_0^s$, and the invariance content $\mathbf{x}_0^c$ in Figure 4. As the raw signals of

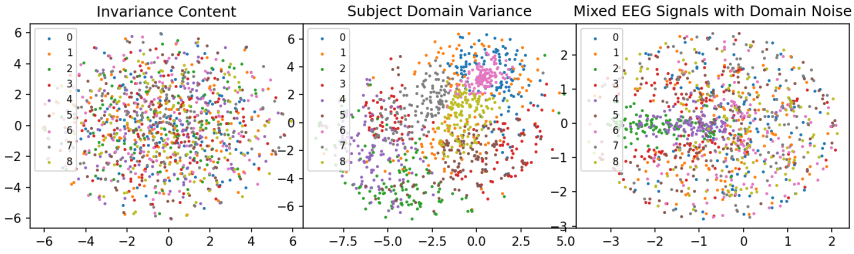
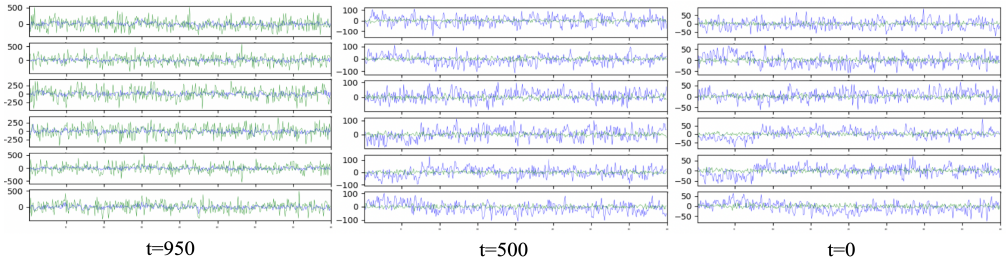
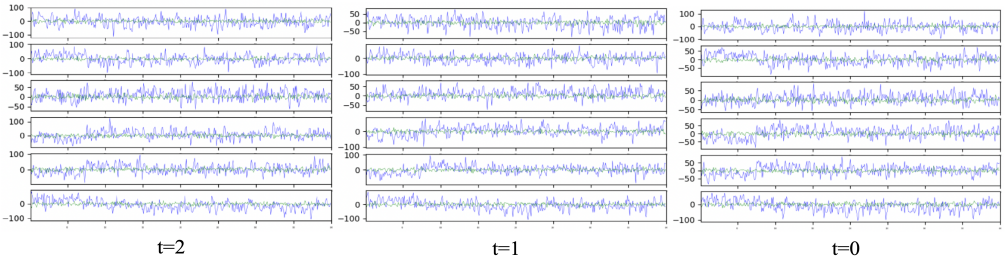


Figure 4: T-SNE visualization of invariance content $\mathbf{x}_0^c$, subject domain variance $\mathbf{x}_0^s$, and mixed signal $\mathbf{x}_0$. Different color denotes different human subjects, where the invariance distribution is irrelevant to different subjects while the domain variance clearly sperable accroding to subject wise.

all channels is too large $bs \times channels \times 224 \times 75$ for T-SNE to perform significant clustering, we actually use a EEGNet pretrained on the 4-class MI classification task, and take the mediate feature map to perform down-sampling on the raw signal. As the 4-class MI classification task is irrelevant to the subject information, so this down sampling does not introduce unfair bias. It could be observed that the invariance distribution of different subjects distributed evenly in the whole space, which illustrate the efficiency of our methods to separate "invariance" features from the original signal. For domain variance, although the degree of clustering between different classes is differentiated, the experimental results still show that the generated domain variance strongly corrlated with subjects. This results support the efficiency of the proposed DS-DDPM method.



(a) Sampling EEG signals with domain variance from pure noise distribution



(b) Perform denoising by assign real EEG signals as mediate state.

Figure 5: Visualization of the EEG signals after performing domain variance separation given subject $s = 3$, where green curves denotes the domain variance and blue curves denotes clean EEG signal.

Visualization of Denoised Signals: Here, we also visualize the EEG signal waves after performing domain specific variance separation. For the visualization, we show both 1) utilizing real EEG signal as the mediate state of the diffusion process and 2) sampling EEG signals of subject s from pure noise in Figure 5. For the generative process from the pure noise, it is observed that the domain variance is very large at the beginning. However, along with the demonising process, the proportion domain variance gradually converged into a rational ration. According to our observation, at half time steps ($t = 500$), the denoising process could generate EEG signals with quality, which support the assumption that we do not need to perform a complete denoising process for domain separation.

For denoising on real EEG signal, setting x_3 as the recorded EEG signal, we could also got rational results by only performing three steps of the denoising process. This assumption could largely save the computational cost as well as the sampling time for real life scennarios.

3.3 Cross Subject Classification Performance

For further verifying the effectiveness of DS-DDPM on human artifacts removal, we conduct cross subject classification task and compare it with widely used ICA (Subasi & Gursoy, 2010) method. In practical, we respectively train MI classifier by only given EEG signals from one single subject, and test the classification accuracy on EEG signals from all other human subjects and reported in Table 2. It is observed that by introducing DS-DDPM and use the separated invariant feature to perform classification tasks, the DS-DDPM outperforms previous model on subject $1, 3, 4, 5, 7, 8, 9$. It suggests the efficiency of our model to separate domain invariant content signals, where by using the domain variance separation, the cross-subject classification performance is significantly improved.

Table 2: Cross-subject classification performance on BCI-IV dataset, where each column denotes a single model trained from one single subject and single denoising method, each row denotes which subject is selected for the training set. **M** denotes the mean accuracy of each model.

	Train with ICA Denoising (Acc. %)									Train with DS-DDPM Denoising (Acc. %)								
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s1	s2	s3	s4	s5	s6	s7	s8	s9
s1	89.29	39.29	25.93	50.00	40.74	27.27	62.96	29.63	36.00	85.59	46.43	46.43	52.00	37.04	31.82	68.22	37.04	38.46
s2	82.61	92.86	33.33	41.67	33.33	31.82	33.33	37.04	36.00	85.71	90.01	34.28	44.00	44.44	37.27	32.14	44.18	35.77
s3	53.57	85.71	81.48	58.33	33.33	54.55	44.44	33.33	48.00	60.71	92.86	82.11	58.00	58.15	55.00	49.29	51.85	44.62
s4	53.57	46.43	81.48	91.67	33.33	54.55	37.04	37.04	44.00	52.86	47.21	78.57	88.45	48.15	50.00	38.57	40.74	46.15
s5	42.86	35.71	37.04	91.67	88.89	54.55	40.74	33.33	40.00	45.86	35.71	37.04	91.67	85.47	52.00	42.81	36.12	46.15
s6	50.00	39.29	37.04	50.00	77.78	90.91	44.44	33.33	44.00	52.14	35.00	46.43	44.44	79.12	88.00	49.15	33.33	50.00
s7	57.14	50.00	37.04	41.67	55.56	90.91	74.07	33.33	40.00	55.71	50.00	42.16	52.00	55.56	90.91	76.12	43.10	47.28
s8	39.29	39.29	44.44	50.00	59.26	45.45	66.67	74.07	44.00	39.29	39.29	44.44	47.25	59.26	45.45	70.37	78.88	44.00
s9	32.14	32.14	37.04	45.83	44.44	45.45	33.33	74.07	84.00	25.15	35.71	32.14	38.20	44.44	45.45	34.07	73.08	85.15
M	55.61	51.19	46.09	**57.87**	51.85	**55.05**	48.56	42.80	46.22	**55.89**	**52.47**	**49.29**	57.33	**56.85**	54.48	**51.19**	**48.70**	**48.62**

3.4 Ablation Study

We evaluate each constraint to give a detailed analysis of how each component contributes to the final result. Limited to paper length, please refer to Appendix B for details.

4 Related Works

In this section, we respectively give a comprehensive literature review of both 1) the previous progress on human artifacts removal for brain dynamics and 2) the current advances in domain separation. Mainstream methods for human artifact removal could be categorized into fourfold 1), signal regression (Al-Nuaimi et al., 2018; Klados et al., 2011), Blink Source Separation (BSS) (Sweeney et al., 2012; Somers & Bertrand, 2016; Teixeira et al., 2006; Subasi & Gursoy, 2010), Empirical-mode Decomposition (EMD), and Spectrum filtering methods (James & Hesse, 2004). Meanwhile, we also give brief reviews of the current domain adaption and domain generation papers, which are similar in mathematical mechanism to our paper. Limited to paper length, please refer to Appendix A for details.

5 Conclusion

In this paper, we propose DS-DDPM, the first conditional diffusion-denoising probabilistic model for domain separation on brain dynamics. We modify the normal denoising process into two streams and respectively generate content signals and the domain variance according to different subjects. Three subtle constraints are introduced to formulate the feature space with good properties: 1) the mix of separated two streams could reconstruct the signal in each diffusion step; 2) content feature and domain variance are orthogonal in feature space; and 3) The domain variances are further divided in terms of subjects. Moreover, UNet-EEG is first proposed to customize proper generative model structure, especially for long sequential samples such as EEG signals. The proposed DS-DDPM could not only generate domain variance distribution for EEG signals from pure noise, but also perform human artifact removal by giving real EEG signals as mediate diffusion states. Experimental results suggest that DS-DDPM is efficient in both explicitly generating domain variance of certain subjects and improving the cross-subject classification performance.

REFERENCES

Ali H Husseen Al-Nuaimi, Emmanuel Jammeh, Lingfen Sun, and Emmanuel Ifeachor. Complexity measures for quantifying changes in electroencephalogram in alzheimer's disease. *Complexity*, 2018, 2018.

Hamid Behnam, Ali Sheikhani, Mohammad Reza Mohammadi, Maryam Noroozian, and Pari Golabi. Analyses of eeg background activity in autism disorders with fast fourier transform and short time fourier measure. In *2007 International Conference on Intelligent and Advanced Systems*, pp. 1240–1244. IEEE, 2007.

Anna C Belkina, Christopher O Ciccolella, Rina Anno, Richard Halpert, Josef Spidlen, and Jennifer E Snyder-Cappione. Automated optimized parameters for t-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nature communications*, 10(1): 1–12, 2019.

P Berg and M Scherg. Dipole modelling of eye activity and its application to the removal of eye artefacts from the eeg and meg. *Clinical Physics and Physiological Measurement*, 12(A):49, 1991.

Magnus Borga, Ola Friman, Peter Lundberg, and Hans Knutsson. A canonical correlation approach to exploratory data analysis in fmri. In *Proceedings of the ISMRM Annual Meeting, Honolulu, Hawaii*. Citeseer, 2002.

Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. *Advances in neural information processing systems*, 29, 2016.

Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021.

Wim De Clercq, Anneleen Vergult, Bart Vanrumste, Wim Van Paesschen, and Sabine Van Huffel. Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE transactions on Biomedical Engineering*, 53(12):2583–2587, 2006.

Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4690–4699, 2019.

Li Dong, Yangsong Zhang, Rui Zhang, Xingxing Zhang, Diankun Gong, Pedro A Valdes-Sosa, Peng Xu, Cheng Luo, and Dezhong Yao. Characterizing nonlinear relationships in functional imaging data using eigenspace maximal information canonical correlation analysis (emicca). *NeuroImage*, 109:388–401, 2015.

Yiqun Duan and Chen Feng. Learning internal dense but external sparse structures of deep convolutional neural network. In *International Conference on Artificial Neural Networks*, pp. 247–262. Springer, 2019.

Christina Heinze-Deml and Nicolai Meinshausen. Conditional variance penalties and domain shift robustness. *Machine Learning*, 110(2):303–348, 2021.

J Craig Henry. Electroencephalography: basic principles, clinical applications, and related fields. *Neurology*, 67(11):2092–2092, 2006.

Steven A Hillyard and Robert Galambos. Eye movement artifact in the cnv. *Electroencephalography and clinical neurophysiology*, 28(2):173–182, 1970.

Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1055–1059. IEEE, 2020.

Christopher J James and Christian W Hesse. Independent component analysis for biomedical signals. *Physiological measurement*, 26(1):R15, 2004.

Xiao Jiang, Gui-Bin Bian, and Zean Tian. Removal of artifacts from eeg signals: a review. *Sensors*, 19(5):987, 2019.

Tzyy-Ping Jung, Scott Makeig, Anthony J Bell, and Terrence J Sejnowski. Independent component analysis of electroencephalographic and event-related potential data. In *Central auditory processing and neural modeling*, pp. 189–197. Springer, 1998.

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.

Manousos A Klados, Christos Papadelis, Christoph Braun, and Panagiotis D Bamidis. Reg-ica: a hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts. *Biomedical Signal Processing and Control*, 6(3):291–300, 2011.

P Senthil Kumar, R Arumuganathan, K Sivakumar, and C Vimal. Removal of ocular artifacts in the eeg through wavelet transform without using an eog reference channel. *Int. J. Open Problems Compt. Math*, 1(3):188–200, 2008.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

Weiyang Liu, Yandong Wen, Zhiding Yu, and Meng Yang. Large-margin softmax loss for convolutional neural networks. *arXiv preprint arXiv:1612.02295*, 2016.

Malik M Naeem Mannan, M Ahmad Kamran, Shinil Kang, and Myung Yung Jeong. Effect of eog signal filtering on the removal of ocular artifacts and eeg-based brain-computer interface: A comprehensive study. *Complexity*, 2018, 2018.

M Murugappan and Subbulakshmi Murugappan. Human emotion recognition through short time electroencephalogram (eeg) signals using fast fourier transform (fft). In *2013 IEEE 9th International Colloquium on Signal Processing and its Applications*, pp. 289–294. IEEE, 2013.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pp. 8162–8171. PMLR, 2021.

Paul L Nunez, Ramesh Srinivasan, et al. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006.

Terence W Picton, Otavio G Lins, and Michael Scherg. The recording and analysis of event-related potentials. *Handbook of neuropsychology*, 10:3–3, 1995.

Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, pp. 8821–8831. PMLR, 2021.

Doha Safieddine, Amar Kachenoura, Laurent Albera, Gwénaël Birot, Ahmad Karfoul, Anca Pasnicu, Arnaud Biraben, Fabrice Wendling, Lotfi Senhadji, and Isabelle Merlet. Removal of muscle artifact from eeg data: comparison between stochastic (ica and cca) and deterministic (emd and wavelet-based) approaches. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–15, 2012.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pp. 2256–2265. PMLR, 2015.

Ben Somers and Alexander Bertrand. Removal of eye blink artifacts in wireless eeg sensor networks using reduced-bandwidth canonical correlation analysis. *Journal of neural engineering*, 13(6): 066008, 2016.

Abdulhamit Subasi and M Ismail Gursoy. Eeg signal classification using pca, ica, lda and support vector machines. *Expert systems with applications*, 37(12):8659–8666, 2010.

Kevin T Sweeney, Tomás E Ward, and Seán F McLoone. Artifact removal in physiological signals—practices and possibilities. *IEEE transactions on information technology in biomedicine*, 16 (3):488–500, 2012.

Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J Miller, Gernot Mueller-Putz, et al. Review of the bci competition iv. *Frontiers in neuroscience*, pp. 55, 2012.

Ana R Teixeira, Ana Maria Tomé, Elmar Wolfgang Lang, Peter Gruber, and A Martins Da Silva. Automatic removal of high-amplitude artefacts from single-channel electroencephalograms. *Computer methods and programs in biomedicine*, 83(2):125–138, 2006.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7167–7176, 2017.

Ricardo Vigário, Jaakko Sarela, Veikko Jousmiki, Matti Hamalainen, and Erkki Oja. Independent component approach to the analysis of eeg and meg recordings. *IEEE transactions on biomedical engineering*, 47(5):589–593, 2000.

Ricardo Nuno Vigário. Extraction of ocular artefacts from eeg using independent component analysis. *Electroencephalography and clinical neurophysiology*, 103(3):395–404, 1997.

Garrick L Wallstrom, Robert E Kass, Anita Miller, Jeffrey F Cohn, and Nathan A Fox. Correction of ocular artifacts in the eeg using bayesian adaptive regression splines. In *Case studies in Bayesian statistics*, pp. 351–365. Springer, 2002.

Pan Wang, Zhenzhou Lu, and Sinan Xiao. A generalized separation for the variance contributions of input variables and their distribution parameters. *Applied Mathematical Modelling*, 47:381–399, 2017.

Zijun Zhang. Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, pp. 1–2. Ieee, 2018.

Shanshan Zhao, Mingming Gong, Tongliang Liu, Huan Fu, and Dacheng Tao. Domain generalization via entropy regularization. *Advances in Neural Information Processing Systems*, 33:16096–16107, 2020.