

Domain Specific Denoising Diffusion Probabilistic Models for Brain Dynamics

Yiqun Duan, *Member, IEEE*, Jinzhao Zhou, Zhen Wang, Yu-Cheng Chang *Member, IEEE*, Yu-Kai Wang *Senior Member, IEEE*, Chin-Teng Lin *Fellow, IEEE*

Abstract—The differences in brain dynamics across human subjects, commonly referred to as human artifacts, have long been a challenge in the field, severely limiting the generalizability of brain dynamics recognition models. Traditional methods for human artifact removal typically employ spectrum filtering or blind source separation, based on simple prior distribution assumptions, which ultimately constrain the capacity to model each subject's domain variance. In this paper, we propose a novel approach to model human artifact removal as a generative denoising process, capable of simultaneously generating and learning subject-specific domain variance and invariant brain signals. We introduce the Domain Specific Denoising Diffusion Probabilistic Model (DS-DDPM), which decomposes the denoising process into subject domain variance and invariant content at each step. By incorporating subtle constraints and probabilistic design, we formulate domain variance and invariant content into orthogonal spaces and further supervise the domain variance with a subject classifier. This method is the first to explicitly separate human subject-specific variance through generative denoising processes, outperforming previous methods in two aspects: 1) DS-DDPM can learn more accurate subject-specific domain variance through domain generative learning compared to traditional filtering methods, and 2) DS-DDPM is the first approach capable of explicitly generating subject noise distribution. Comprehensive experimental results indicate that DS-DDPM effectively alleviates domain distribution bias for cross-domain brain dynamics signal recognition.

Index Terms—Diffusion Denoising Model, Brain Computer Interface, EEG, Domain Generation

I. INTRODUCTION

THE recognition of human brain dynamics signals such as electroencephalogram (EEG) [1] and Event-related Potential (ERP) [2] is of vital importance for the non-invasive brain-computer interface. However, the brain signal distribution from different human subjects exhibits severe distributional differences [3], [4], which means the recognition model [5]–[8] 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 efforts for human artifacts removal [4] have sought to alleviate the aforementioned problem and can be categorized into four main approaches: 1) Regression [9], 2) Blind Source Separation (BSS) [10], [11], 3) Empirical-mode Decomposition (EMD), and 4) Wavelet Transform algorithm combined

with their hybrid methods [12]. Among these methods, Regression, and BSS are the most commonly used. Classic regression methods [10] operate under the assumption that each channel is a sum of clean EEG data and a proportion of artifacts given by known reference signals [13], [14] through a set of regressed transmission factors. However, these methods only learn transmission factors instead of directly learning domain variance. BSS methods, such as Independent Component Analysis (ICA), decompose observed signals into independent components (ICs) [11] from linear mixtures of cerebral and artifactual sources. Current ICA methods still require reference signal information for artifact separation and do not consider the domain variance between different human subjects.

Previous human artifact removal methods generally assume domain variance as simple prior distributions, limiting their capacity to model domain variance for each subject. We propose to model human artifact removal as a domain-specific denoising process and perform denoising by generating subject-specific domain variance and invariant brain signals through the denoising process. Our intuition is to introduce learned subject-specific domain variance instead of a simple prior distribution. To achieve this, we propose modeling human artifact removal as a domain-specific denoising process and performing denoising by generating subject-specific domain variance and invariant brain signals through the denoising process. This approach enables us to learn to generate domain-specific variance according to each subject, rather than separating subject noise through simple filtering or analysis.

In this paper, we propose to revisit human artifact removal for EEG signals by modeling the removal process using domain-specific denoising diffusion probabilistic models (DS-DDPM). We adhere to the widely accepted assumption that recorded EEG signals comprise mixtures of clean samples and subject-specific human artifact noises.

We introduce a novel probabilistic model based on the diffusion-denoising generative process [15]–[17], which has recently gained prominence, particularly for image generation tasks [18]. The proposed DS-DDPM incorporates artifact removal into every denoising step in the original model, as illustrated in Figure 1. At time step t , the diffusion result \mathbf{x}_t is separated by two denoising models into domain variance \mathbf{x}^{st} (human artifacts) and the clean signal \mathbf{x}^{ct} . We represent the diffusion result as the summation of the two separated contents: $\mathbf{x}_t = \mathbf{x}^{ct} + \mathbf{x}^{st}$, indicating that the two streams share the same diffusion process. This constraint naturally corresponds to our assumption that the noisy EEG signal can be decomposed into clean signals and domain-specific noise.

Yiqun Duan, Jinzhao Zhou, Yu-Cheng Chang, Yu-Kai Wang and Chin-Teng Lin are with Human-centric Artificial Intelligence Centre, Australia Artificial Intelligence Institute, School of Computer Science, the University of Technology Sydney, Zhen Wang is with the School of Computer Science, University of Sydney

Manuscript submitted May 15, 2023;

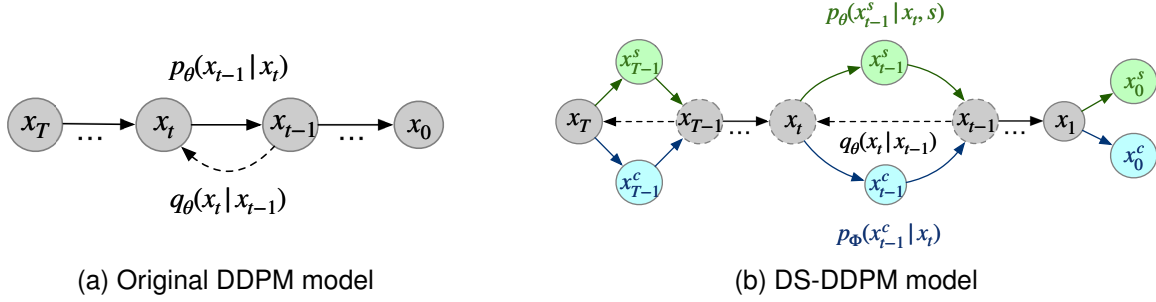


Fig. 1. Illustration of probabilistic model graphs, where Figures 1a and 1b respectively denote the original model DDPM and the proposed DS-DDPM. DS-DDPM separate the subject-specific domain variance \mathbf{x}_t^s (human artifacts), and the clean signal \mathbf{x}_t^c at each denoising time step t . 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.

A detailed mathematical problem definition is provided in Section III-A.

We decompose the domain variance and clean signal into two mutually orthogonal spaces, as described in Section III-B. Additionally, an auxiliary cosine classifier is applied to disentangle variance spaces for subject-specific noises. Since no previous methods have explored DDPM models for EEG signals, Section III-C offers technical details of the efficient structure of DDPM for EEG signals.

Comprehensive experimental results demonstrate that 1) the domain-specific separation is significant in a subject-domain-wise manner (Section IV-B), and 2) domain-specific denoising can help enhance cross-subject classification performance (Section IV-C). We also present an ablation study discussing the contribution of each proposed component in Section V. The contributions of this paper can be summarized into three main aspects:

- This work is the first to introduce denoising diffusion probabilistic models into EEG signals.
- DS-DDPM gives a novel approach to explicitly separate long-existing domain variance related to human subject difference through a 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.

II. RELATED WORKS

A. Denoising Methods for EEG

Artifact removal from EEG signals is a crucial step before extracting neural information for subsequent analyses. Artifacts, which may arise from imprecise recording systems, casual recording procedures, or human subjects themselves, can be categorized into intrinsic or extrinsic artifacts based on whether they originate from the human subject. Regardless of their source, these artifacts can lead to misleading results in brain-computer interface (BCI) applications [19], [20]. In particular, physiological artifacts caused by human subjects cannot be eliminated by merely applying bandwidth filtering or strict experimental procedures. Mainstream methods for physiological artifact removal include signal decomposition

and artifact signal estimation, incorporating techniques such as regression [9], [21], Blind Source Separation (BSS) [10], [11], [22], [23], Empirical-mode Decomposition (EMD), and spectrum filtering methods [12].

Spectrum filtering methods, which involve transforming temporal signals into frequency components, enable the separation or extraction of signals. In the frequency domain, signals can be decomposed into distinct combinations of individual frequencies and amplitudes, thereby separating useful signals from artifacts. Common methods for spectrum filtering in EEG denoising include Wavelet Transform [24], [25] and Fourier Transform [26], [27]. However, these approaches typically require some prior knowledge about the characteristics of useful components and artifacts in the EEG signal, which may be limiting when applied to a specific EEG-related application without sufficient exploration and understanding of its EEG signal.

In contrast, BSS methods are unsupervised learning algorithms that learn an underlying linear transformation, transforming the source EEG signal to the observed noisy EEG signal without needing prior information and reference channels. Principal Component Analysis (PCA) [28], Independent Component Analysis (ICA) [29]–[32], and Canonical Correlation Analysis (CCA) [33]–[35] are some major BSS methods. Although these methods achieve high accuracy in artifact removal, they strongly assume that the source signals can be approximated by linear transformation from the noisy observation. We argue that this oversimplified transformation may not adequately represent the complex nature of EEG artifacts.

B. Domain Separation

EEG signals are well-known for their substantial inter-subject variance. Our method resembles domain separation and adaptation methods, as we aim to capture both the similarities in high-level EEG patterns and the differences in subject-related patterns using distinct models. Specifically, existing domain separation methods assume that signals can be decomposed into domain-specific and domain-invariant features. In our approach, the domain-invariant feature directly refers to the underlying mental state derived from EEG signals

collected from human subjects, while the domain-specific feature corresponds to subject-specific features that manifest as unwanted artifacts in EEG analysis.

Early work on domain separation, such as Domain Separation Networks (DSN) [36], proposed an architecture to learn and separate both domain-specific and domain-invariant features by introducing additional network branches for each feature. By reducing the discrepancy between the domain-specific subspace and the class-specific feature space, their method can learn separable domain-specific information that is orthogonal to the domain-invariant feature space. Later, Adversarial Discriminative Domain Adaptation (ADDA) [37] employed a domain discriminator [38] to allocate features obtained from different domains into a shared space. In this way, domain-specific information is not captured by the feature extractor for different domains.

More recent approaches proposed further separation of domain-specific and domain-invariant features by utilizing additional regularization terms that eliminate the dependency between domain-specific features and domain-invariant features. These terms include Entropy regularization [39] and Variance Penalty [40], [41]. Although these methods vary in network architectures and regularizations, the core idea behind domain separation can be summarized as eliminating the discrepancy between domain-specific features and domain-invariant features. To the best of our knowledge, our work is the first to establish a connection between domain separation and subject modeling in EEG processing.

III. METHODOLOGY

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

A. Definition of Domain-Specific Denoising

The separation of domain-specific variance is decomposed into the denoising process of a denoising diffusion probabilistic model [15]. 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 $\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, \mathbf{x}_t, \dots, \mathbf{x}_T\}$ 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.

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1}),$$

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

where $\beta_1, \beta_2, \dots, \beta_T$ is a fixed variance coefficient schedule. Following the Gaussian distribution assumption of DDPM, \mathbf{x}_t

could be further represented as the combination of \mathbf{x}_0 and sampled variance ε .

$$q(\mathbf{x}_t|\mathbf{x}_0) := \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I}),$$

$$\alpha_t := 1 - \beta_t, \quad \bar{\alpha}_t = \prod_{i=1}^t \alpha_i \quad (2)$$

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.

$$\begin{aligned} \mathbf{x}_t &= \alpha_t \mathbf{x}_{t-1} + \beta_t \varepsilon_t = \alpha_t (\bar{\alpha}_{t-1} \mathbf{x}_0 + \bar{\beta}_{t-1} \bar{\varepsilon}_{t-1}) + \beta_t \varepsilon_t \\ &= \bar{\alpha}_t \mathbf{x}_0 + \alpha_t \bar{\beta}_{t-1} \bar{\varepsilon}_{t-1} + \beta_t \varepsilon_t \end{aligned} \quad (3)$$

where $\varepsilon_t \sim \mathcal{N}(0, \mathbf{I})$ is a Gaussian distribution that represents the stochastic property of the diffusion process. It also gives a 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 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 for different subject domains s by separating the process into two variables: $q(\mathbf{x}_{t-1}^s, \mathbf{x}_{t-1}^c|\mathbf{x}_t)$. Here, \mathbf{x}_{t-1}^s represents the separated domain variance for subject s , and \mathbf{x}_{t-1}^c represents the corresponding clean signal generated at step t . We assume the signal is a mixture of subject noise and the clean signal, such that $\mathbf{x}_{t-1} = \mathbf{x}_{t-1}^s + \mathbf{x}_{t-1}^c$. As the direct reversal of the diffusion process $q(\mathbf{x}_{t-1}^s, \mathbf{x}_{t-1}^c|\mathbf{x}_t)$ is intractable [42], we employ two separate functions to express the denoising process as follows.

$$\begin{aligned} p_\theta(\mathbf{x}_{t-1}^s|\mathbf{x}_t) &:= \mathcal{N}(\mathbf{x}_{t-1}^s; \mu_\theta(\mathbf{x}_t, t, s), \sigma_t^2 \mathbf{I}), \\ p_\phi(\mathbf{x}_{t-1}^c|\mathbf{x}_t) &:= \mathcal{N}(\mathbf{x}_{t-1}^c; \mu_\phi(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I}) \end{aligned} \quad (4)$$

Where $p_\theta(\mathbf{x}_{t-1}^s|\mathbf{x}_t)$ and $p_\phi(\mathbf{x}_{t-1}^c|\mathbf{x}_t)$ are the two separated denoising functions decomposed from $p(\mathbf{x}_{t-1}|\mathbf{x}_t) = p_\theta(\mathbf{x}_{t-1}^s|\mathbf{x}_t) + p_\phi(\mathbf{x}_{t-1}^c|\mathbf{x}_t)$. Here, σ_t^2 denotes the variance in transition. The core transitions $\mu_\theta(\mathbf{x}_t, t, s)$ and $\mu_\phi(\mathbf{x}_t, t)$ are learned by deep neural networks. We follow previous experimental settings [15], [16], where σ_t^2 is directly set as β_t or $\frac{1-\bar{\alpha}_t}{1-\alpha_t}\beta_t$, which have shown 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 [15].

$$\mathbf{x}_{t-1} = \frac{1}{1 - \alpha_t} (\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \mu) + \sigma_t \mathbf{z}, \quad (5)$$

$$\mu = \mu_\theta(\mathbf{x}_t, t, s) + \mu_\phi(\mathbf{x}_t, t) \quad (6)$$

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 $\mathbf{x}_{t-1} = \mathbf{x}_{t-1}^s + \mathbf{x}_{t-1}^c$, we could further approximate the separated domain variance of subject s as $\mathbf{x}_{t-1}^s = \frac{1}{1-\alpha_t} (\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \mu_\theta(\mathbf{x}_t, t, s)) + \sigma_t \mathbf{z}$, and the separated clean signal as $\mathbf{x}_{t-1}^c = \frac{1}{1-\alpha_t} (\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \mu_\phi(\mathbf{x}_t, t)) + \sigma_t \mathbf{z}$. It is noted that, after the denoising process sampled to \mathbf{x}_1 , we could directly calculate the desired domain variance \mathbf{x}_0^s of human subjected s and clean data \mathbf{x}_0^c according to the equations defined above.

B. Separate Domain Specific Variance by Constraint

Under the domain-specific denoising process defined in Section III-A, we separate the domain-specific variance by the combination of three constraints. 1) Section III-B1 introduces 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 III-B2 introduces orthogonal constraint between clean signal and domain-specific variance by orthogonal loss \mathcal{L}_o . 3) Section III-B3 further formulate the domain variance space separable according to different human subjects by Arc-Margin loss \mathcal{L}_{arc} .

1) *Reverse Process*: Basically, we model 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 III-A and Figure 1b. 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 [15], Considering the definition in Section III-A that the separated two denoising models share the same variance coefficient scheduler α and β we could approximate the summation of the denoising process as.

$$\mu_\theta(\mathbf{x}_t, t, s) + \mu_\phi(\mathbf{x}_t, t) = \frac{1}{\alpha_t}(\mathbf{x}_t - \beta_t \varepsilon_\theta(x_t, t, s) - \beta_t \varepsilon_\phi(x_t, t)) \quad (7)$$

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 7, we could rewrite the distance to be minimized as in Equation 8.

$$\|\mathbf{x}_{t-1} - (\mu_\theta(\mathbf{x}_t, t, s) + \mu_\phi(\mathbf{x}_t, t))\|^2 = \frac{\beta_t^2}{\alpha_t^2} \|\varepsilon_t - \varepsilon_\theta(x_t, t, s) - \varepsilon_\phi(x_t, t)\|^2 \quad (8)$$

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 9.

$$\mathcal{L}_r = \lambda_r \|\varepsilon_t - \varepsilon_\theta(\bar{\alpha}_t \mathbf{x}_0 + \alpha_t \bar{\beta}_{t-1} \bar{\varepsilon}_{t-1} + \beta_t \varepsilon_t, t, s) - \varepsilon_\phi(\bar{\alpha}_t \mathbf{x}_0 + \alpha_t \bar{\beta}_{t-1} \bar{\varepsilon}_{t-1} + \beta_t \varepsilon_t, t)\|^2 \quad (9)$$

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) *Separate Subject Domain Variance Apart*: 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.

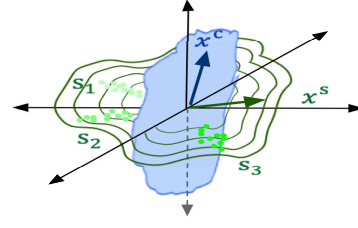


Fig. 2. The orthogonal domain variance separation of DS-DDPM.

The orthogonal property is granted by introducing a constraint loss defined in Equation 10.

$$\mathcal{L}_o = \lambda_o \left\| (\mu_\theta(\mathbf{x}_t, t, s)^\top \mu_\phi(\mathbf{x}_t, t) - \mathbf{I}) \otimes (\mathbf{I} - \mathbf{I}) \right\|^2 \quad (10)$$

where $\mu_\theta(\mathbf{x}_t, t, s)$ and $\mu_\phi(\mathbf{x}_t, t)$ are the denoising process defined in Equation 7. 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 .

3) *Formulate Subject Domain Variance to Human Subjects*: In order to carry out domain-specific denoising pertaining to a particular human subject s , the DS-DDPM model also enforces subject-wise separability of the domain variance. This constraint is achieved by introducing an additional subject classifier to supervise the conditional denoising process $\mu_\theta(\mathbf{x}_t, t, s)$. The domain variance generated for subject s is expected to predict the subject label s based on the domain variance, by introducing a conditional probabilistic model $p(s|\mathbf{x}_t^s)$.

Under our assumption, we design the clean data variance to be entirely independent of the subject information. Consequently, we can eliminate all terms in the equations previously defined for representing the classification model $p(s|\mathbf{x}_t^s)$. We employ a simple EEGNet [5] classifier to extract the feature $\bar{\mathbf{x}}_t^{s_i}$ from the denoising output \mathbf{x}_t^s , with the transition represented as $\mathbf{W}\theta$.

We propose formulating the decision boundary for predicting subject s in the cosine space, where the traditional SoftMax [43] loss is replaced by the Additive Angular Margin classification (Arc-Margin) loss [44]. The Arc-Margin 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 θ_{s_j} is the angle between the weight \mathbf{W}_{θ, 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 signifies that the weights make predictions solely dependent on the angle between the feature and the weight. We also let $\|\bar{\mathbf{x}}_t^{s_i}\| = r$, representing that the learned embedding features are distributed on a hypersphere with a radius of r . As the feature $\bar{\mathbf{x}}_t^{s_i}$ is calculated by $\mu_\theta(\mathbf{x}_t, t, s)$ through a series of classification transformations ω , $\bar{\mathbf{x}}_t^{s_i}$ can be approximated by eliminating irrelevant terms in Equation 7 as follows.

$$\cos \theta_{s_j} = \frac{\mathbf{W}_{\theta, s_j}^\top \omega(\varepsilon_\theta(\bar{\alpha}_t \mathbf{x}_0 + \alpha_t \bar{\beta}_{t-1} \bar{\varepsilon}_{t-1} + \beta_t \varepsilon_t, t, s))}{\|\mathbf{W}_{\theta, s_j}\| \|\omega(\varepsilon_\theta(\bar{\alpha}_t \mathbf{x}_0 + \alpha_t \bar{\beta}_{t-1} \bar{\varepsilon}_{t-1} + \beta_t \varepsilon_t, t, s))\|} \quad (11)$$

An additive angular margin penalty m is added between $\bar{\mathbf{x}}_t^{s_i}$ and \mathbf{W}_{θ, s_j} to simultaneously enhance the intra-class com-

pactness and inter-class discrepancy [44]. Also by giving the coefficient λ_{arc} for subject classification, we could formulate the Arc-Margin loss \mathcal{L}_{arc} could be defined in Equation 12.

$$\mathcal{L}_{arc} = -\lambda_{arc} \frac{1}{N} \sum_{i=1}^N \log \frac{e^{r(\cos(\theta_{y_i} + m))}}{e^{r(\cos(\theta_{y_i} + m))} + \sum_{s_j=1, s_j \neq y_i}^n e^{r \cos \theta_{s_j}}} \quad (12)$$

where y_i^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 increases the interoperability of DS-DDPM.

4) *Temporal Difference Constraint*: As described in the methodology, we create the network inputs by stacking segments with overlapping regions in both temporal and spatial dimensions, as illustrated in Figure 4. Consequently, the results generated by the EEG-UNet model also contain overlaps. When performing the denoising process, we expect the same time points to have deterministic outputs, which means we aim to minimize the temporal difference in overlaps. To address this, we constrain the temporal difference by minimizing \mathcal{L}_{td} , as defined in Equation 13.

$$\mathcal{L}_{td} = \frac{1}{n} \sum_i i = 1^n \text{Overlap}(\mathbf{y}_{t-1} - \mathbf{y}_t)^2, \quad \mathbf{y}_t \in \mathbf{x}_0 \quad (13)$$

In this equation, we calculate the Mean Squared Error (MSE) between the sequence of sampled segments \mathbf{y}_t generated by dividing \mathbf{x}_0 into stackings. The constraint is achieved by relaxing it into a regularization term and incorporating it into the main loss function.

C. Model Structure

This section introduces how we design domain-specific generative model structure and how we train the denoising model by combining the loss constrained proposed in Section III-B. For model structure, we follow the common UNet [45] 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 [45] structure into two generative streams.

The UNet-EEG model receives diffusion results \mathbf{x}_t and respectively generates domain variance \mathbf{x}_{t-1}^s and the clean content \mathbf{x}_{t-1}^c . Here for the 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 model's ability to maintain time sequential relations. Similar to the 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 the key and value for the attention layer.

Training and Sampling Procedure: In light of the structure defined above, we outline a comprehensive training procedure as presented in Algorithm 1. To expedite convergence, 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 a group of human subjects. As EEG signals inherently contain subject labels (indicating the human source of the EEG signal), training of ω is feasible. This approach helps the optimizer focus on the parameters θ and ϕ for the generative model.

For the entire training procedure, the time step t is sampled uniformly from $1, 2, \dots, T$, and the variance scheduler α, β is sampled and fixed. At each iteration, the algorithm samples $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ at every time step. During each iteration, the generative model for clean data and subject-specific domain variance is computed, taking into account the subject label s , time step t , real EEG signal \mathbf{x}_0 , and the sampled variables mentioned above. The parameters θ and ϕ of the generative model are optimized by minimizing the weighted combination of the loss functions proposed in Section III-B.

After training the generative model, we can generate subject-specific domain variances based on the probabilistic model described in Section III-A. Algorithm 2 presents a comprehensive procedure for sampling domain variances according to subject s and generating unconditional EEG signals from pure noise.

A straightforward approach for denoising a given signal involves sampling a subject-specific domain variance \mathbf{x}_0^s and directly computing $\mathbf{x}_0 - \mathbf{x}_0^s$ to obtain a cleaner signal. However, this method does not consider the noise present in the given signal. Furthermore, the iterative denoising process can be time-consuming. In real-life scenarios, it is more practical to perform denoising based not only on subject s but also on the signal itself.

Taking into account the denoising iterative process $\mathbf{x}_T, \dots, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_0$, the denoising model incrementally removes noise at each step. Since recorded EEG signals inherently contain noise, it is reasonable to directly perform denoising by assuming the recorded signal as $\mathbf{x}_{1/2}$, an intermediate state in the denoising process. As a result, given a recorded raw signal \mathbf{x}_{raw} , we can directly separate the domain variance and clean content by computing: $\mathbf{x}_0^s = \frac{1}{\sqrt{\alpha_1}} \left(\mathbf{x}_{raw} - \frac{1-\alpha_1}{\sqrt{1-\alpha_1}} \epsilon \theta(\mathbf{x}_{raw}, 1, s) \right) + \sigma_1 \mathbf{z}$ and $\mathbf{x}_0^c = \frac{1}{\sqrt{\alpha_1}} \left(\mathbf{x}_{raw} - \frac{1-\alpha_1}{\sqrt{1-\alpha_1}} \epsilon \theta(\mathbf{x}_{raw}, 1) \right) + \sigma_1 \mathbf{z}$.

This assumption allows for replacing the iterative sampling process with a single inference step. Experimental results demonstrate that this approach effectively improves the cross-subject classification task.

IV. EXPERIMENTS

In this section, we present comprehensive experiments to demonstrate the efficiency of the proposed Domain-Specific Denoising. We analyze the correlation in Section IV-B by 1) performing correlation coefficient calculations. Furthermore, we conduct experiments on the cross-subject classification task in Section IV-C to illustrate our method's efficiency in classification tasks. An additional ablation study is conducted in Section V to discuss the effectiveness of each component.

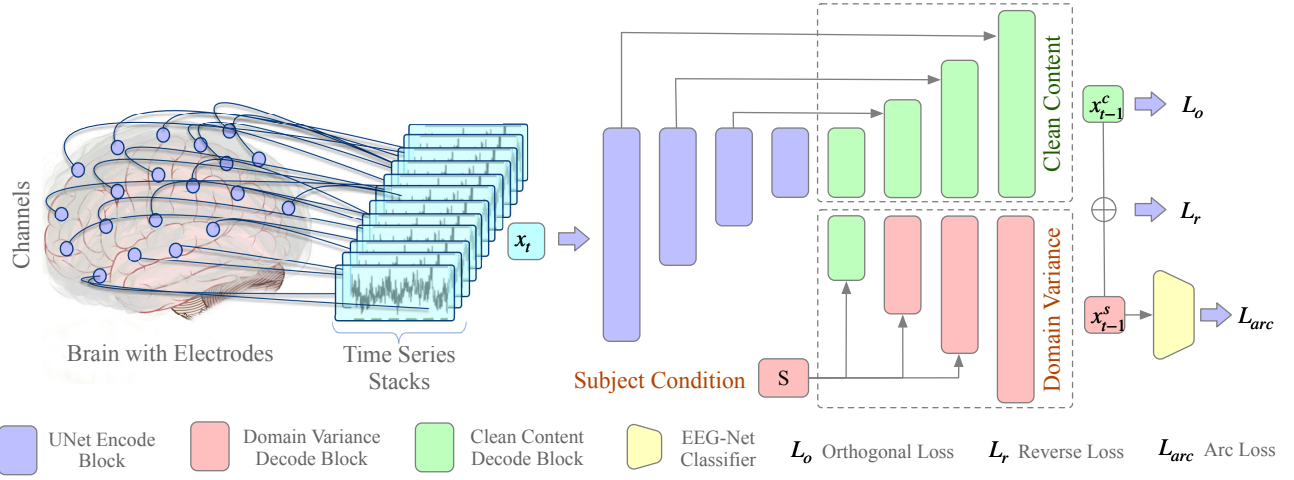


Fig. 3. The overall model structure of DS-DDPM, where the recorded EEG signals are sliced by time window 224, step 75 and stacked into shape $bs \times channel \times window \times stacks$ as \mathbf{x}_0 . The modified UNet-EEG structure receives diffusion results \mathbf{x}_t and respectively generates 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 .

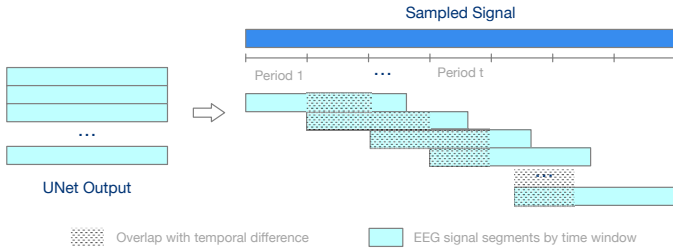


Fig. 4. Illustration of generative temporal difference in generative process.

A. Experimental Setup

We primarily conduct our experiments on the BCI-Competition-IV dataset [46], which is widely used for validating Motor Imagery classification tasks. The dataset is collected under the widely-used 10-20 system, comprising 22 EEG channels and 3 EOG channels. The dataset contains EEG and EOG signals with a sampling frequency of 250 Hz from nine subjects. Each subject was required to perform four classes of motor imagery (left hand, right hand, feet, and tongue) while recording brain dynamics.

We use only the 22-channel EEG signals, preprocessing them into a shape of 22×750 , where 750 is the time sequence length with a 250 Hz sampling rate for 3 seconds. As mentioned in Section III-C, we use a time window to slide along the time sequence and slice the time sequence into segments with overlaps. By using a time window size of 224 and stride size of 75, the data is processed into a shape of $bs \times 22 \times 224 \times 8$ and fed into the UNet-EEG model. For the UNet-EEG model, we sequentially perform three down-sampling and up-sampling blocks for the clean content stream and one down-sampling and up-sampling block for the domain variance stream. This slim structure for the domain variance stream prevents severe overfitting, making the training process more stable.

For the subject classifier, we use the original EEGNet

Algorithm 1 Training

```

1: while Not Converge do
2:    $\mathbf{x}_0, s \sim q(\mathbf{x}_0, s)$ , Sample  $\mathbf{x}_0$  from Subjects  $\{s\}$ 
3:   Pretrain subject classifier  $\omega$  given sample label pairs  $\{\mathbf{x}_0, s\}$ 
4:    $t \sim \text{Uniform}(\{1, \dots, T\})$ 
5:    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
6:   Calculate,  $\epsilon_\theta(x_t, t, s)$ ,  $\epsilon_\phi(x_t, t)$ ,  $\omega(\epsilon_\theta(x_t, t, s))$  with model parameters  $\theta$  and  $\phi$ .
7:   Take gradient descent step optimize  $\theta$  and  $\phi$  on
8:   Loss  $\mathcal{L} = \lambda_r \mathcal{L}_r + \lambda_o \mathcal{L}_o + \lambda_{arc} \mathcal{L}_{arc} + \lambda_{td} \mathcal{L}_{td}$ 
9: end while

```

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:    $\mathbf{x}_{t-1}^s = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \epsilon_\theta(\mathbf{x}_t, t, s) \right) + \sigma_t \mathbf{z}$ 
5:    $\mathbf{x}_{t-1}^c = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \epsilon_\phi(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$ 
6: end for
7: return  $\mathbf{x}_0^s, \mathbf{x}_0^c$ 

```

structure with a modified input size of $bs \times 22 \times 224 \times 8$. For the Arc-Margin \mathcal{L}_{arc} hyperparameters, we follow previous explorations on human face recognition, where we set the radius $r = 30$ and the margin as $m = 0.5$. We use the ADAM optimizer [47] with default hyperparameters. According to our experiments, we found that a batch size of 64 stabilizes the training process more than smaller numbers. The DS-DDPM is implemented based on PyTorch, and we open-source our codes¹ to the community.

¹<https://github.com/duanyiqun/DS-DDPM>.

TABLE I

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

TABLE II

SUBJECT-WISE CORRELATION ANALYSIS BETWEEN REAL EEG SAMPLES AND DS-DDPM GENERATIVE SAMPLES.

	Subject Correlation Coefficient of EEG BCI-IV EEG Signals								
	s1	s2	s3	s4	s5	s6	s7	s8	s9
s1	0.101	0.048	0.055	0.059	0.045	0.051	0.050	0.042	0.043
s2	0.048	0.074	0.040	0.038	0.033	0.041	0.045	0.037	0.038
s3	0.055	0.040	0.080	0.048	0.037	0.043	0.048	0.037	0.039
s4	0.059	0.038	0.048	0.088	0.041	0.042	0.053	0.043	0.044
s5	0.045	0.033	0.037	0.041	0.058	0.034	0.039	0.031	0.034
s6	0.051	0.041	0.043	0.042	0.034	0.071	0.044	0.034	0.038
s7	0.050	0.045	0.548	0.053	0.039	0.044	0.093	0.042	0.050
s8	0.042	0.037	0.037	0.043	0.031	0.054	0.042	0.069	0.032
s9	0.050	0.038	0.039	0.041	0.034	0.038	0.044	0.032	0.066
	Subject Correlation Coefficient Between DS-DDPM Sampled Signal and BCI-IV								
	s1	s2	s3	s4	s5	s6	s7	s8	s9
s1	0.109	0.071	0.077	0.080	0.067	0.073	0.072	0.063	0.070
s2	0.072	0.090	0.067	0.065	0.058	0.066	0.068	0.060	0.061
s3	0.076	0.065	0.096	0.072	0.061	0.068	0.070	0.060	0.063
s4	0.079	0.064	0.073	0.100	0.064	0.068	0.074	0.064	0.064
s5	0.069	0.060	0.065	0.067	0.076	0.062	0.063	0.055	0.059
s6	0.083	0.066	0.069	0.067	0.059	0.088	0.068	0.057	0.061
s7	0.073	0.069	0.079	0.075	0.070	0.069	0.103	0.066	0.066
s8	0.067	0.064	0.065	0.069	0.057	0.062	0.066	0.082	0.057
s9	0.072	0.064	0.067	0.067	0.059	0.065	0.067	0.056	0.082

B. Domain Specific Variance Analysis

1) *Correlation analysis*: In order to analyze the generative quality of the domain variance generative tasks, we conduct correlation analysis on both real signals and generative signals, as shown in Table II. 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 than the real signal distribution of the same human subject s , we think is significant to illustrate our efficiency. The results are reported in the lower part of Table II, where the correlation coefficient properties are 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 $0.031 \sim 0.059$. In comparison, the value of generated

signals varies between $0.069 \sim 0.083$. We argue that this phenomenon is rational as the real sample is actually the upper bound of the current regression-based domain separation methods.

C. Cross-Subject Classification Performance

To further verify the effectiveness of DS-DDPM in removing human artifacts, we conduct a cross-subject classification task and compare its performance with the widely used ICA method [23]. In practice, we train MI classifiers for each subject individually using only their EEG signals and then test the classification accuracy on EEG signals from all other subjects. The results are reported in Table I.

We observe that by introducing DS-DDPM and utilizing the separated invariant features for classification tasks, the proposed model outperforms the previous method for subjects 1, 3, 4, 5, 7, 8, and 9. This indicates the efficiency of our model in separating domain-invariant content signals. By using domain variance separation, the cross-subject classification performance is significantly improved.

D. Visualization of Domain Variance Distribution

We also visualize the domain variance distribution of each subject. We sample domain variance by giving different subject labels to each subject. Then the domain variance is visualized by introducing the t-Distributed Stochastic Neighbor Embedding (t-SNE) [48] algorithm. We compare the visualization of the original EEG signal \mathbf{x}_0 , the separated domain variance \mathbf{x}_0^s , and the invariance content \mathbf{x}_0^c in Figure 5.

As the raw signals of all channels are too large $bs \times channels \times 224 \times 75$ for t-SNE to perform significant clustering, we actually use an EEGNet pre-trained 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 downsampling does not introduce unfair bias. It could be observed that the invariance distribution of different subjects is distributed evenly in the whole space, which illustrates 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

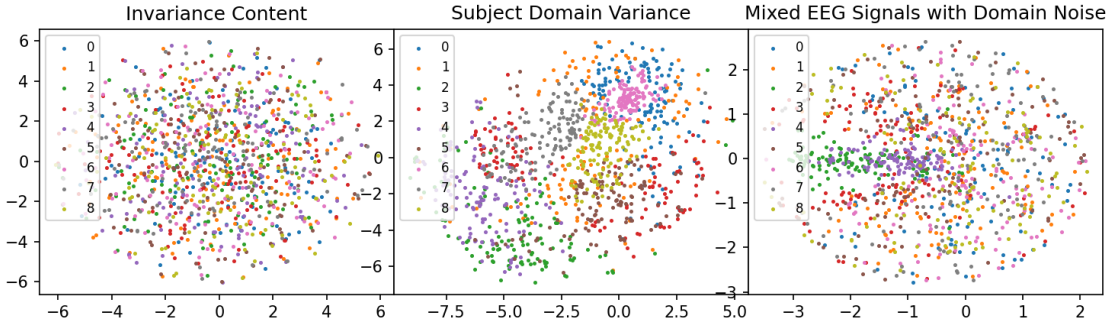


Fig. 5. T-SNE visualization of invariance content \mathbf{x}_0^c , subject domain variance \mathbf{x}_0^s , and mixed signal \mathbf{x}_0 . Different color denotes different human subjects, where the invariance distribution is irrelevant to different subjects while the domain variance is clearly separable according to subject wise.

variance is strongly correlated with subjects. These results support the efficiency of the proposed DS-DDPM method.

V. DISCUSSION

A. Generative Temporal Difference

As described in the methodology, we create the network inputs by stacking segments with overlapping regions in both temporal and spatial dimensions. Consequently, the results generated by the EEG-UNet model also contain overlaps. However, for the denoising process, it is essential to have deterministic outcomes in the time domain. To tackle this issue, we incorporate a temporal difference constraint to enhance the generative consistency across the time axis. This section presents an ablation study that compares the performance of the model with and without the temporal difference constraint.

We discuss the impact of incorporating temporal differences on four key metrics to assess the model's performance and report the results in Figure 6. 1) Reverse loss (\mathcal{L}_r), as defined in Section III-B of the main paper, measures how accurately the generative process can reconstruct the real EEG signals following the diffusion process. 2) Subject classification loss (\mathcal{L}_{arc}) in arc-space [44], which reflects the degree of separability among individual human subjects in the extracted noise. 3) Orthogonal loss (\mathcal{L}_o) assesses the level of separability between the clean content space and noise space. 4) Temporal difference, measured using mean square error, evaluates the effectiveness of the temporal constraint in ensuring consistency across the time axis.

Figure 6 demonstrates that incorporating the temporal constraint effectively enables the model to learn and minimize temporal differences. In contrast, without the temporal constraint, the model generates substantial temporal differences, which are not suitable for practical applications.

Additionally, it can be observed that while the convergence speed of reverse loss is slower with the temporal constraint, it eventually outperforms the model without the constraint after 8,000 training steps, further validating the efficacy of our approach. Similar observations also stand for both classification loss and orthogonal loss. Notably, implementing the temporal constraint is particularly beneficial for the classification model, as real data exhibits minimal temporal differences due to its unique and deterministic nature.

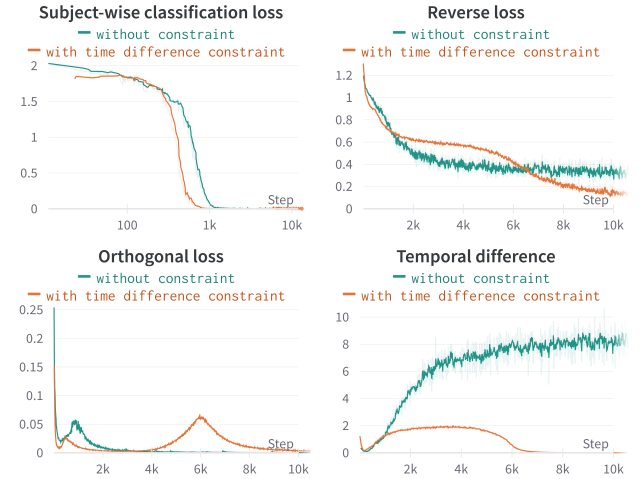


Fig. 6. The comparison between training metrics with or without temporal difference constraint.

B. Single Channel Denoising

One significant advantage of DS-DDPM is its ability to perform denoising without the need for reference channels. This feature makes DS-DDPM suitable for single-channel denoising tasks. To showcase this capability, we conduct extensive experiments comparing DS-DDPM with previous methods in the context of single-channel denoising.

TABLE III
COMPARING DS-DDPM WITH TRADITIONAL DENOISING METHODS

Sub.	SCCA	ECCL	MsetCCA	TRCA	ETRCAL	SSCOR	DS-DDPM	Avg.
s1	0.50	0.67	0.50	0.00	0.00	0.17	0.29	0.25
s2	0.67	0.83	0.33	0.67	0.83	0.67	0.83	0.71
s3	1.00	0.83	1.00	0.83	0.83	0.83	0.89	0.88
s4	0.83	1.00	0.83	0.67	0.83	0.67	0.68	0.80
s5	0.83	0.83	0.83	0.67	0.67	0.67	0.67	0.74
s6	0.17	0.50	0.33	0.33	0.00	0.67	0.29	0.27
s7	1.00	1.00	1.00	1.00	0.83	0.83	0.68	0.87
s8	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.82
s9	0.33	0.50	0.33	0.17	0.17	0.17	0.50	0.36
s10	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.04
s11	0.50	0.83	0.67	0.83	1.00	0.67	1.00	0.82
s12	0.83	0.83	0.83	0.67	0.67	0.67	0.68	0.73
s13	0.67	0.83	0.33	0.67	0.67	0.50	0.83	0.64
s14	0.17	0.33	0.33	0.17	0.33	0.00	0.33	0.26
s15	1.00	1.00	1.00	1.00	0.83	0.83	0.83	0.88
Avg.	0.63	0.73	0.61	0.57	0.57	0.54	0.63	

We utilize a dataset²(BRI: Brain Robotic Interaction) consisting of single-channel EEG signals collected from human subjects controlling robot dogs using eight specific commands. This results in an eight-class classification task. Data were collected from 15 participants, each providing 45 minutes of EEG signals. The signals are first processed using different denoising methods and subsequently classified through a standard SSEVP classification process. The comparison methods include widely used techniques such as SCCA [49], ECCA [50], MsetCCA [50], TRCA [51], ETRCA [52], and SSCOR [53]. The results are presented in Table III. It should be noted that for CCA-based methods, we use an additional reference channel as denoising references. However, for DS-DDPM, denoising is performed solely using single-channel data. DS-DDPM achieves the second-highest average accuracy across the 15 subjects without utilizing any reference signal. This observation highlights the effectiveness of DS-DDPM in single-channel denoising tasks.

C. Limitations

The application of DS-DDPM presents a novel approach to explicitly address subject domain variance in EEG signals. However, it still faces several limitations, which we discuss in this section.

1) *Computational Complexity*: The DDPM method involves iterative processes and can be computationally expensive, particularly when dealing with high-dimensional EEG signals. This may result in longer training and inference times, as well as increased memory consumption, which might not be suitable for real-time or on-chip deployment in wearable devices. However, we note that this problem can be partially mitigated by introducing improved diffusion processes, such as DDIM and consistency models.

2) *Model Generalization*: The performance of the DDPM-based method largely depends on the quality of the training data. The DS-DDPM model alleviates the generalization problem by introducing human-subject-wise classifier guidance. However, this mechanism relies on human subjects' labels, which somewhat resembles semi-blind source separation rather than traditional blind source separation techniques used for denoising (e.g., CCA). This means that calibration data must be acquired before the model can be applied for denoising.

Currently, available datasets only contain EEG signals from tens of subjects. Drawing from experience in the computer vision and natural language processing domains, this issue could be further alleviated if the number of human subjects in EEG signal datasets was to increase on a larger scale.

VI. CONCLUSION

In conclusion, this paper presents DS-DDPM, a novel conditional diffusion-denoising probabilistic model designed for domain separation in brain dynamics. By modifying the standard denoising process into two streams, we generate content signals and domain variance tailored to individual

subjects. To create a feature space with desirable properties, we introduce three subtle constraints: 1) the combination of the separated streams can reconstruct the signal at each diffusion step; 2) the content feature and domain variance are orthogonal in the feature space; and 3) the domain variances are further divided according to the subjects. Additionally, we propose UNet-EEG, a customized generative model structure specifically designed for handling long sequential samples such as EEG signals.

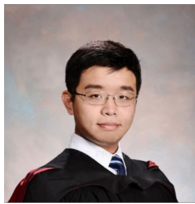
The proposed DS-DDPM is not only capable of generating domain variance distributions for EEG signals from pure noise but also effectively removes human artifacts by using real EEG signals as intermediate diffusion states. Experimental results demonstrate the efficiency of DS-DDPM in both generating domain variance explicitly for specific subjects and enhancing cross-subject classification performance. Overall, DS-DDPM represents a new approach to the field of brain dynamics denoising and domain separation. Future work may focus on further refining the model to improve its applicability across various EEG-based tasks and exploring the potential of DS-DDPM in other neuroimaging modalities.

REFERENCES

- [1] P. L. Nunez, R. Srinivasan *et al.*, *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006.
- [2] T. W. Picton, O. G. Lins, and M. Scherg, "The recording and analysis of event-related potentials," *Handbook of neuropsychology*, vol. 10, pp. 3–3, 1995.
- [3] J. C. Henry, "Electroencephalography: basic principles, clinical applications, and related fields," *Neurology*, vol. 67, no. 11, pp. 2092–2092, 2006.
- [4] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from eeg signals: a review," *Sensors*, vol. 19, no. 5, p. 987, 2019.
- [5] 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 brain-computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.
- [6] Z. Wang, L. Liu, Y. Duan, and D. Tao, "Continual learning through retrieval and imagination," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, pp. 8594–8602, 2022.
- [7] Y. Duan and C. Feng, "Learning internal dense but external sparse structures of deep convolutional neural network," in *International Conference on Artificial Neural Networks*. Springer, 2019, pp. 247–262.
- [8] Y. Duan, Z. Wang, Y. Li, J. Tang, Y.-K. Wang, and C.-T. Lin, "Cross task neural architecture search for eeg signal classifications," *arXiv preprint arXiv:2210.06298*, 2022.
- [9] A. H. H. Al-Nuaimi, E. Jammeh, L. Sun, and E. Ifeachor, "Complexity measures for quantifying changes in electroencephalogram in alzheimer's disease," *Complexity*, vol. 2018, 2018.
- [10] K. T. Sweeney, T. E. Ward, and S. F. McLoone, "Artifact removal in physiological signals—practices and possibilities," *IEEE transactions on information technology in biomedicine*, vol. 16, no. 3, pp. 488–500, 2012.
- [11] B. Somers and A. Bertrand, "Removal of eye blink artifacts in wireless eeg sensor networks using reduced-bandwidth canonical correlation analysis," *Journal of neural engineering*, vol. 13, no. 6, p. 066008, 2016.
- [12] C. J. James and C. W. Hesse, "Independent component analysis for biomedical signals," *Physiological measurement*, vol. 26, no. 1, p. R15, 2004.
- [13] S. A. Hillyard and R. Galambos, "Eye movement artifact in the cnv," *Electroencephalography and clinical neurophysiology*, vol. 28, no. 2, pp. 173–182, 1970.
- [14] G. L. Wallstrom, R. E. Kass, A. Miller, J. F. Cohn, and N. A. Fox, "Correction of ocular artifacts in the eeg using bayesian adaptive regression splines," in *Case studies in Bayesian statistics*. Springer, 2002, pp. 351–365.
- [15] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8162–8171.

²The dataset is not publicly available due to the ongoing review process. However, the data can be provided for further verification during the reviewing process.

- [16] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, "Ilvr: Conditioning method for denoising diffusion probabilistic models," *arXiv preprint arXiv:2108.02938*, 2021.
- [17] Y. Duan, X. Guo, and Z. Zhu, "Diffusiondepth: Diffusion denoising approach for monocular depth estimation," *arXiv preprint arXiv:2303.05021*, 2023.
- [18] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8821–8831.
- [19] M. M. N. Mannan, M. A. Kamran, S. Kang, and M. Y. Jeong, "Effect of eeg signal filtering on the removal of ocular artifacts and eeg-based brain-computer interface: A comprehensive study," *Complexity*, vol. 2018, 2018.
- [20] J. Zhou, Y. Duan, Y. Zou, Y.-C. Chang, Y.-K. Wang, and C.-T. Lin, "Speech2eeg: Leveraging pretrained speech model for eeg signal recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023.
- [21] M. A. Klados, C. Papadelis, C. Braun, and P. D. Bamidis, "Reg-ica: a hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts," *Biomedical Signal Processing and Control*, vol. 6, no. 3, pp. 291–300, 2011.
- [22] A. R. Teixeira, A. M. Tomé, E. W. Lang, P. Gruber, and A. M. Da Silva, "Automatic removal of high-amplitude artefacts from single-channel electroencephalograms," *Computer methods and programs in biomedicine*, vol. 83, no. 2, pp. 125–138, 2006.
- [23] A. Subasi and M. I. Gursay, "Eeg signal classification using pca, ica, lda and support vector machines," *Expert systems with applications*, vol. 37, no. 12, pp. 8659–8666, 2010.
- [24] P. S. Kumar, R. Arumuganathan, K. Sivakumar, and C. Vimal, "Removal of ocular artifacts in the eeg through wavelet transform without using an eeg reference channel," *Int. J. Open Problems Compt. Math.*, vol. 1, no. 3, pp. 188–200, 2008.
- [25] D. Safieddine, A. Kachenoura, L. Albera, G. Birot, A. Karfoul, A. Pasnicu, A. Biraben, F. Wendling, L. Senhadji, and I. 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*, vol. 2012, no. 1, pp. 1–15, 2012.
- [26] H. Behnam, A. Sheikhan, M. R. Mohammadi, M. Noroozian, and P. 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*. IEEE, 2007, pp. 1240–1244.
- [27] M. Murugappan and S. 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*. IEEE, 2013, pp. 289–294.
- [28] 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*, vol. 12, no. A, p. 49, 1991.
- [29] T.-P. Jung, S. Makeig, A. J. Bell, and T. J. Sejnowski, "Independent component analysis of electroencephalographic and event-related potential data," in *Central auditory processing and neural modeling*. Springer, 1998, pp. 189–197.
- [30] R. N. Vigário, "Extraction of ocular artefacts from eeg using independent component analysis," *Electroencephalography and clinical neurophysiology*, vol. 103, no. 3, pp. 395–404, 1997.
- [31] R. Vigário, J. Sarela, V. Jousmiki, M. Hamalainen, and E. Oja, "Independent component approach to the analysis of eeg and meg recordings," *IEEE transactions on biomedical engineering*, vol. 47, no. 5, pp. 589–593, 2000.
- [32] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects," *Clinical Neurophysiology*, vol. 111, no. 10, pp. 1745–1758, 2000.
- [33] M. Borga, O. Friman, P. Lundberg, and H. Knutsson, "A canonical correlation approach to exploratory data analysis in fmri," in *Proceedings of the ISMRM Annual Meeting, Honolulu, Hawaii*. Citeseer, 2002.
- [34] L. Dong, Y. Zhang, R. Zhang, X. Zhang, D. Gong, P. A. Valdes-Sosa, P. Xu, C. Luo, and D. Yao, "Characterizing nonlinear relationships in functional imaging data using eigenspace maximal information canonical correlation analysis (emicca)," *NeuroImage*, vol. 109, pp. 388–401, 2015.
- [35] W. 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*, vol. 53, no. 12, pp. 2583–2587, 2006.
- [36] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain separation networks," *Advances in neural information processing systems*, vol. 29, 2016.
- [37] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7167–7176.
- [38] Y. Duan, Z. Wang, Y. Li, and J. Wang, "Cross-domain multi-style merge for image captioning," *Computer Vision and Image Understanding*, vol. 228, p. 103617, 2023.
- [39] S. Zhao, M. Gong, T. Liu, H. Fu, and D. Tao, "Domain generalization via entropy regularization," *Advances in Neural Information Processing Systems*, vol. 33, pp. 16096–16107, 2020.
- [40] C. Heinze-Deml and N. Meinshausen, "Conditional variance penalties and domain shift robustness," *Machine Learning*, vol. 110, no. 2, pp. 303–348, 2021.
- [41] P. Wang, Z. Lu, and S. Xiao, "A generalized separation for the variance contributions of input variables and their distribution parameters," *Applied Mathematical Modelling*, vol. 47, pp. 381–399, 2017.
- [42] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International Conference on Machine Learning*. PMLR, 2015, pp. 2256–2265.
- [43] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," *arXiv preprint arXiv:1612.02295*, 2016.
- [44] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.
- [45] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. 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)*. IEEE, 2020, pp. 1055–1059.
- [46] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz *et al.*, "Review of the bci competition iv," *Frontiers in neuroscience*, p. 55, 2012.
- [47] Z. Zhang, "Improved adam optimizer for deep neural networks," in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*. Ieee, 2018, pp. 1–2.
- [48] A. C. Belkina, C. O. Ciccolella, R. Anno, R. Halpert, J. Spidlen, and J. E. Snyder-Cappione, "Automated optimized parameters for t-distributed stochastic neighbor embedding improve visualization and analysis of large datasets," *Nature communications*, vol. 10, no. 1, pp. 1–12, 2019.
- [49] D. R. Hardoon and J. Shawe-Taylor, "Sparse canonical correlation analysis," *Machine Learning*, vol. 83, pp. 331–353, 2011.
- [50] L. Du, K. Liu, X. Yao, S. L. Risacher, J. Han, A. J. Saykin, L. Guo, and L. Shen, "Multi-task sparse canonical correlation analysis with application to multi-modal brain imaging genetics," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 18, no. 1, pp. 227–239, 2019.
- [51] H. Tanaka, T. Katura, and H. Sato, "Task-related component analysis for functional neuroimaging and application to near-infrared spectroscopy data," *NeuroImage*, vol. 64, pp. 308–327, 2013.
- [52] M. Nakanishi, Y. Wang, X. Chen, Y.-T. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of ssveps for a high-speed brain speller using task-related component analysis," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 1, pp. 104–112, 2017.
- [53] K. A. Couch and D. R. Lillard, "Sample selection rules and the intergenerational correlation of earnings," *Labour Economics*, vol. 5, no. 3, pp. 313–329, 1998.



Yiqun Duan received the B.Eng. degree in electrical engineering from Hohai University, Nanjing, China, and M.A.Sc. degree in electrical engineering from the University of British Columbia, Vancouver, Canada. He is currently pursuing a Ph.D. degree in CIBCI lab, Australian Artificial Intelligence Institute, School of Computer Science, University of Technology Sydney, Australia. His current research topic is vision & language scenery understanding for embodied robotic systems and brain-computer interfaces. He is currently serving as a reviewer for

top conferences in both machine learning and robotic systems. His research interests include natural language processing, reinforcement learning, and multi-model vision & language understanding.



Jinzhao Zhou received the B.Eng. degree from the School of Computer Science & Technology and the B.M.S. degree from the School of Electronic Commerce, both from the South China University of Technology (SCUT), Guangzhou, China, in 2018. He then received the M.S. degree from the School of Computer Science and Engineering from SCUT in 2021. He is currently pursuing a Ph.D. degree in CIBCI lab, Australian Artificial Intelligence Institute, School of Computer Science, University of Technology Sydney, Australia. His current research

interests include Brain-Computer Interface, multi-modal understanding, and reinforcement learning.



Zhen Wang received the B.Sc. degree from the Hunan University, Changsha, China, in 2016, and the M.Sc. degree from Tsinghua University, Beijing, China, in 2019. He is currently pursuing a Ph.D. degree with the School of Computer Science, The University of Sydney, Darlingtown, NSW, Australia. His current research interest includes streaming label learning, continual learning, and quantitative finance. He serves as a reviewer for IJCV, IEEE Trans. on Cybernetics, IEEE Intelligent Systems, ACM TKDD, NeurIPS, ICML, ICLR, CVPR, ICCV,

AAAI, etc.



Dr. Yu-Cheng Chang received the B.S. degree in vehicle engineering from the National Taipei University of Technology, Taipei, Taiwan, in 2008, the M.S. degree with a specialization in system and control from the Department of Electrical Engineering, National Chung-Hsing University, Taiwan, in 2010 and the Ph.D. degree in software engineering from University of Technology Sydney (UTS), Australia in 2021. He currently is a Postdoctoral Researcher with the CIBCI lab, UTS. His current research

interests include fuzzy systems, human performance modeling, and novel human-agent interaction.



Dr. Yu-Kai Wang (M'13) received the M.S. degree in biomedical engineering and the Ph.D. degree in computer science in 2009 and 2015, respectively, both from the National Chiao Tung University, Taiwan. He is currently a Senior Lecturer at the Faculty of Engineering and Information Technology at the University of Technology Sydney, Australia. He is the author of 35 published original articles in international journals and more than 40 contributions to international conferences. His current research

interests include computational neuroscience, human performance modeling, Brain-Computer Interface, and novel human-agent interaction.



Dr. Chin-Teng Lin received a Bachelor's of Science from National Chiao-Tung University (NCTU), Taiwan in 1986, and holds Master's and PhD degrees in Electrical Engineering from Purdue University, USA, received in 1989 and 1992, respectively. He is currently a distinguished professor and Co-Director of the Australian Artificial Intelligence Institute within the Faculty of Engineering and Information Technology at the University of Technology Sydney, Australia. He is also an Honorary Chair Professor of Electrical and Computer Engineering at NCTU.

For his contributions to biologically inspired information systems, Prof Lin was awarded Fellowship with the IEEE in 2005, and with the International Fuzzy Systems Association (IFSA) in 2012. He received the IEEE Fuzzy Systems Pioneer Award in 2017. He has held notable positions as editor-in-chief of IEEE Transactions on Fuzzy Systems from 2011 to 2016; seats on Board of Governors for the IEEE Circuits and Systems (CAS) Society (2005-2008), IEEE Systems, Man, Cybernetics (SMC) Society (2003-2005), IEEE Computational Intelligence Society (2008-2010); Chair of the IEEE Taipei Section (2009-2010); Distinguished Lecturer with the IEEE CAS Society (2003-2005) and the CIS Society (2015-2017); Chair of the IEEE CIS Distinguished Lecturer Program Committee (2018-2019); Deputy Editor-in-Chief of IEEE Transactions on Circuits and Systems-II (2006-2008); Program Chair of the IEEE International Conference on Systems, Man, and Cybernetics (2005); and General Chair of the 2011 IEEE International Conference on Fuzzy Systems. Prof Lin is the co-author of Neural Fuzzy Systems (Prentice-Hall) and the author of Neural Fuzzy Control Systems with Structure and Parameter Learning (World Scientific). He has published more than 390 journal papers including over 180 IEEE journal papers in the areas of neural networks, fuzzy systems, brain-computer interface, multimedia information processing, cognitive neuro-engineering, and human-machine teaming, that have been cited more than 28,000 times. Currently, his h-index is 77, and his i10-index is 349.