Short-to-Long Functional Connectivity Transfer via Structure-Aware Latent Diffusion

Benjamin Torres[1], Yufeng Liu[2], and Zhengwu Zhang[1]

[1] University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
ibta@unc.edu, zhengwu_zhang@unc.edu
[2] University of Michigan, Ann Arbor, MI 48109, USA
yufliu@umich.edu

Abstract. Resting-state functional connectivity (FC) is widely used to model brain organization and predict phenotypes, yet short fMRI acquisitions often yield unreliable estimates. We propose a Structure-Aware Latent Diffusion framework, SALD, to synthesize long-scan-like FC from limited observations. Our model conditions the generation process on short-scan FC, structural connectivity (SC) as a lightweight anatomical prior, and subject-level covariates. Crucially, in the Adolescent Brain Cognitive Development Study (ABCD) cohort, our results suggest a fidelity–utility tension in FC synthesis: likelihood-based diffusion training may improve distributional fidelity while attenuating subtle inter-subject variation relevant to phenotype prediction. To address this, we introduce a reward-guided Low-Rank Adaptation (LoRA) strategy that distills a guidance signal isolating the FC-congruent portion of phenotype variance, steering generation toward preserving this signal while maintaining sample realism. Experiments on the ABCD cohort show that our method better preserves phenotype-relevant signal in short-to-long FC synthesis, with the most pronounced improvements over short-scan baselines occurring in the low-scan-time regime across the evaluated phenotypes. These findings suggest a promising direction for phenotype-aware short-to-long FC synthesis.

Keywords: Functional Connectivity · Phenotype Prediction · Latent Diffusion Models · Reward-Guided Generation · Low-Rank Adaptation (LoRA)

1 Introduction

Learning stable representations from limited observations is a fundamental challenge in modern machine learning. In high-dimensional settings, representations are often estimated from short or noisy measurement horizons, leading to instability that propagates to downstream tasks. This issue becomes particularly acute when the representation itself is a structured object, such as a covariance or connectivity matrix, whose estimation error scales unfavorably with dimensionality [5, 12].

A compelling instance arises in functional connectivity (FC) analysis in computational neuroscience. FC is widely used to characterize individual brain organization and predict cognitive traits [7, 10]. It is typically computed as pairwise correlations between blood-oxygen-level-dependent (BOLD) time series measured via functional MRI (fMRI) [3]. Because these correlations are estimated from finite time series, their reliability depends strongly on scan duration: short scans yield noisy and unstable connectivity estimates, whereas longer acquisitions produce substantially more reproducible patterns [2, 20]. Importantly, FC reliability directly constrains phenotype prediction performance [21]. Despite this, scan duration is often limited by cost, participant burden, and clinical feasibility in real studies.

This setting exposes a broader question: *Can we recover long-horizon, reliable structured representations from short observations by leveraging auxiliary information?* In neuroimaging, complementary modalities are often available, including structural connectivity (SC) from diffusion MRI and demographic covariates such as age and sex. We investigate whether short-scan FC can be transformed into *long-scan-like* FC by conditioning on these auxiliary signals.

This problem naturally forms a *conditional distribution transfer task under partial observability*. Importantly, the goal of this generation task is twofold. First, **fidelity**: how much effective scan duration can be recovered from short acquisitions? Second, **utility**: can the generated connectivity preserve the subtle, phenotype-discriminative structure required for downstream neuroscience analysis? In practice, FC is rarely used purely for reconstruction; its primary value lies in predicting individual differences in cognition and behavior [28]. The second objective is substantially more challenging. Phenotype prediction from FC is intrinsically noisy: even high-quality long-scan FC explains only a modest fraction of variance, with predictive correlations often saturating around 0.5 [7]. This weakly explainable regime implies that only a small portion of connectivity variability is task-relevant, while the remainder reflects noise or idiosyncratic fluctuations. Under such ambiguity, likelihood-based generative models optimize distributional fidelity rather than discriminative structure. When multiple long-scan reconstructions are consistent with the same short observation, maximum-likelihood objectives favor conditional averages of plausible solutions. As a result, subject-specific variability is attenuated, precisely the variability that carries predictive signal. Consequently, a diffusion model may generate visually realistic connectivity patterns while inadvertently degrading downstream utility.

In this work, we study the *fidelity–utility tension* in short-to-long FC synthesis. We first show empirically that conditional diffusion trained with likelihood-based objectives improves distributional similarity to long-scan FC, but can under-recover phenotype-relevant subject-level variation. We then introduce a controlled distribution adaptation framework designed to restore *explainable discriminative structure* while maintaining generative fidelity to long-scan FC.

Our approach proceeds in two stages. We learn a structure-conditioned latent diffusion model that maps short-scan FC to the distribution of long-scan FC,

conditioning on SC and demographic covariates. SC serves as a soft anatomical prior that provides structural guidance grounded in evidence that white-matter architecture shapes functional coupling [16, 26, 37], thereby reducing reconstruction ambiguity. To counteract attenuation of predictive signal, we introduce a reward-aligned adaptation mechanism regularized by one-step denoising anchoring, a lightweight proxy for KL regularization adapted from DPOK [8, 11]. A frozen predictor trained on high-quality long-scan FC provides a differentiable distillation-style reward, encouraging the generator to preserve *explainable phenotype variance* rather than overfit noise. Adaptation is implemented efficiently in latent space, modifying only the components necessary to restore discriminative structure while remaining close to the base generative model.

We evaluate our framework on the Adolescent Brain Cognitive Development Study (ABCD) cohort [6,33], synthesizing long-scan-like FC from multiple short-scan durations. Results demonstrate that generative reconstruction alone is insufficient for downstream prediction, and that explicit reward alignment is necessary to recover subject-specific variability critical for phenotype modeling. Our contributions are as follows:

- We show that, in ambiguous short-to-long FC reconstruction settings, likelihood-based conditional diffusion can attenuate subject-specific variability, thereby reducing downstream predictive power.
- We develop SALD, a structure-conditioned latent diffusion for short-to-long connectivity transfer with structural conditioning.
- We introduce phenotype-reward adaptation for short-to-long FC synthesis, improving phenotype-relevant variation while maintaining fidelity to long-scan FC.

2 Related Work

Functional Connectivity Reliability and Scan Duration. Functional connectivity reliability increases with scan duration, and this reliability directly affects downstream phenotype prediction performance [2, 20]. Large-scale studies have therefore emphasized collecting more data per subject to stabilize estimates rather than relying on computational corrections alone, highlighting practical limits on time, cost, and scalability [19, 21]. Our work proposes a computational alternative to longer acquisitions by synthesizing long-scan-like estimates from short observations.

Structure-to-Function Modeling and Connectivity Synthesis. SC→FC has been studied extensively, with methods spanning regression-based predictors, fully connected neural networks, and graph neural networks that infer FC from anatomical connectivity [16, 26, 37]. Critiques have questioned the usefulness of SC→FC mappings at the individual level [29], while subsequent reviews emphasize that the mapping is non-trivial but modest [26, 37]. We adopt SC as helper context rather than primary guidance, and focus on generating a distribution of FC conditioned on both SC and short-duration FC, explicitly preserving phenotype-predictive structure in a way prior work has not addressed.

Diffusion and Generative Models for Connectome Synthesis. Diffusion models, including latent diffusion architectures, have achieved high-fidelity synthesis and scalable conditional generation [23]. For connectomes, BrainNet-Diff leverages multimodal neuroimaging with a latent diffusion backbone for brain network generation and includes an auxiliary classification loss to encourage downstream utility [39]. DiffGAN-F2S uses a diffusion GAN to map fMRI-derived signals to structural connectivity [40]. SFC-GAN adopts a CycleGAN-style bidirectional translation between SC and FC, reporting that translated connectomes remain useful for classification [32]. These methods focus on SC–FC translation or disease classification, whereas we condition on short-duration FC and SC to generate a posterior distribution of FC while preserving its phenotype-predictive structure.

Reward-Guided and Task-Aligned Diffusion Fine-Tuning. When a clear gradient signal is available, direct fine-tuning of diffusion models has been studied. DOODL optimizes latents to find favorable noise directions for classifier guidance, improving quality at the cost of longer sampling [34]. ReFL approaches update model weights by denoising from random noise and differentiating through a one-step x_0 prediction in later stages, but this can introduce distribution shift because the training denoising step differs from inference [36]. DRaFT instead evaluates gradients on the same output distribution as inference, using memory-efficient backpropagation through the last K steps with checkpointing and LoRA [8, 17]. We follow this line of reward-guided diffusion fine-tuning because our adaptation uses a task-specific MSE reward, while deviations from the pretrained model are constrained through one-step denoising anchoring, analogous to the stabilizing role of KL regularization in DPOK [11].

3 Methodology

3.1 Problem Formulation

Our goal is to recover high-fidelity FC from limited-duration observations. Let F_{long} denote FC recovered from a long (*e.g.*, 20-minute) scan, and let

$$\mathbf{h}_{\text{obs}} = [F_{\text{short}},\ \text{SC},\ e_{\text{cov}}], \tag{1}$$

where, with a slight abuse of notation, $\mathbf{h}_{\text{obs}}$ collects the conditioning inputs rather than denoting a single concatenated vector: $F_{\text{short}} \in \mathbb{R}^{100\times100}$ is short-scan FC for a short duration $\tau \in (0, \text{long})$ minutes, $\text{SC} \in \mathbb{R}^{100\times100}$ is structural connectivity, and e_{cov} contains subject covariates (e.g., age and sex). We aim to model

$$\hat{F}_{\text{long}} \sim p_\theta(F_{\text{long}} \mid \mathbf{h}_{\text{obs}})\,. \tag{2}$$

In practice, as described in the next subsection, this conditional distribution is learned in a latent space for computational efficiency. Figure 1 provides an overview of the SALD steering pipeline.

Because FC matrices are empirical correlation estimates from finite BOLD time series, they are noisy observations of latent connectivity structure. This

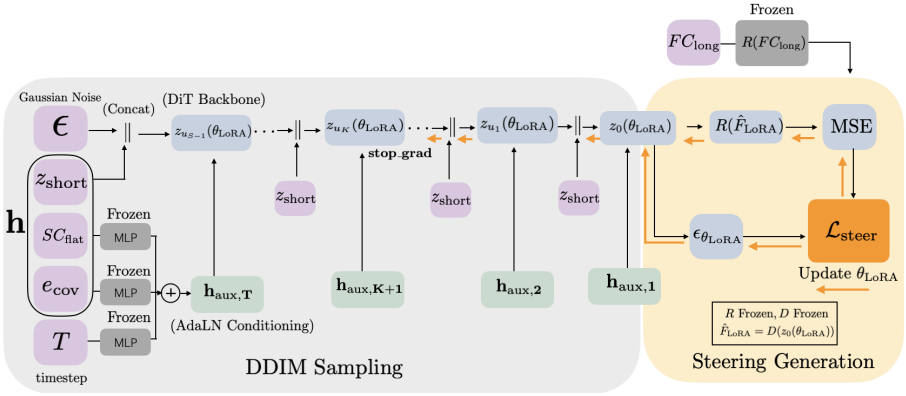

Fig. 1: SALD sampling steering. A pretrained conditional diffusion model generates connectivity from short-scan FC, structural connectivity (SC), and covariates. Following the denoising-time optimization strategy of DRaFT [8], gradients are propagated through the final K DDIM denoising steps while only LoRA parameters θ_{LoRA} are updated and the backbone remains frozen. In SALD, the decoded FC is evaluated by a frozen phenotype predictor, and the steering loss $\mathcal{L}_{\text{steer}}$ biases generation toward phenotype-discriminative connectivity.

motivates a probabilistic formulation of short-to-long FC synthesis, allowing uncertainty to be modeled rather than assumed away.

3.2 Latent Conditional Diffusion Framework

We model the conditional distribution $p_\theta(F_{\text{long}} \mid \mathbf{h}_{\text{obs}})$ using denoising diffusion probabilistic models (DDPMs) [14,18]. To improve computational efficiency while preserving fidelity we adopt a latent diffusion formulation [23]. This design is especially important because FC is typically high-dimensional (often larger than 100×100, depending on the brain atlas), making direct diffusion in data space computationally expensive and harder to scale.

To compress the FC data, a pretrained variational autoencoder (VAE) is trained to encode both short-duration and long-duration FC matrices into a shared latent space. A shared latent space is employed because F_{short} and F_{long} represent the same underlying biological object, the brain's functional network architecture, differing primarily in measurement noise rather than semantic content. By training a single VAE on both short- and long-duration connectivity matrices from each scan pair, we force the encoder to learn a unified manifold of biologically plausible connectivity patterns. This ensures that the noisy representation z_{short} and the high-fidelity target z_{long} lie in the same coordinate system, reducing the generative task from a complex cross-domain translation to a more tractable trajectory refinement within a shared feature space.

Let $z_{\text{long}} = E(F_{\text{long}})$ and $z_{\text{short}} = E(F_{\text{short}})$ denote the latent representations of long- and short-duration FC. Our objective is to learn the conditional distri-

bution $p_\theta(z_{\text{long}} \mid \mathbf{h})$ where $\mathbf{h} = [z_{\text{short}}, \text{SC}_{\text{flat}}, e_{\text{cov}}]$ (See Figure 1) contains the latent short-scan FC, flattened structural connectivity, and subject covariates. We use a denoising diffusion model [14] to learn $p_\theta(z_{\text{long}} \mid \mathbf{h})$, which contains forward and reverse diffusion processes.

Forward diffusion process. Diffusion models learn this distribution by reversing a stochastic corruption process applied to the target latent variable. Starting from $z_0 = z_{\text{long}}$, Gaussian noise is progressively injected according to

$$q(z_t|z_{t-1}) = \mathcal{N}\big(z_t; \sqrt{\alpha_t}\, z_{t-1}, (1-\alpha_t)I\big), \tag{3}$$

where $\{\alpha_t\}_{t=1}^T$ is a predefined noise schedule controlling the signal decay at each step and I is the identity matrix. Define $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$. This process admits the closed form perturbation

$$z_t = \sqrt{\bar{\alpha}_t}\, z_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \tag{4}$$

where ϵ denotes the Gaussian noise injected during the forward process. As t increases the signal progressively vanishes and for sufficiently large T we obtain $z_T \sim \mathcal{N}(0, I)$.

Reverse diffusion process. The generative model learns to invert this corruption process by predicting the noise component that produced z_t. The reverse Markov chain is parameterized as

$$p_\theta(z_{t-1}|z_t, \mathbf{h}) = \mathcal{N}\big(z_{t-1}; \mu_\theta(z_t, t, \mathbf{h}), \Sigma_q(t)\big), \tag{5}$$

where the covariance $\Sigma_q(t)$ is analytically determined from the forward diffusion process and therefore tractable. Following [14], the mean can be written in terms of a neural network that predicts the noise term,

$$\mu_\theta(z_t, t, \mathbf{h}) = \frac{1}{\sqrt{\alpha_t}} z_t - \frac{1-\alpha_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}} \hat{\epsilon}_\theta(z_t, t, \mathbf{h}), \tag{6}$$

where $\hat{\epsilon}_\theta$ estimates the Gaussian noise ϵ introduced by the forward process.

Likelihood-based training. The DDPM framework is trained by minimizing a variational bound on the negative log-likelihood of the data distribution. This objective reduces to the following noise-prediction loss

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{t,\epsilon}\left[\|\epsilon - \epsilon_\theta(z_t, t, \mathbf{h})\|^2\right], \tag{7}$$

where $t \sim U(1, T)$ and $\epsilon \sim \mathcal{N}(0, I)$. This training objective ensures that the reverse process matches the distribution of long-duration FC in latent space. However, likelihood-based diffusion training only guarantees accurate modeling of the data distribution; it does not explicitly enforce preservation of phenotype-discriminative structure. Consequently, samples that faithfully match the FC distribution may still fail to preserve subject-specific information that can be used to predict behavioral phenotypes. Addressing this limitation motivates the *steering objective* introduced in the following section.

Conditional generation mechanism. While many diffusion models employ classifier-free guidance by randomly dropping conditioning variables during

training to trade off mode coverage for sample fidelity [15], we treat FC synthesis as a conditional restoration task rather than unconditional generation. In this setting, the short-scan representation $\mathbf{h}$ provides necessary structural constraints for subject-specific reconstruction. Consequently, following image-to-image translation frameworks [24, 25], the conditioning signal $\mathbf{h}$ is injected at each denoising step, enabling the model to use the input constraints throughout reconstruction.

3.3 Architecture and Conditioning

We adopt a Diffusion Transformer (DiT) backbone [22] as the denoising network for our diffusion model, as illustrated in Figure 2. Because fidelity transfer resembles an image-to-image translation problem, we follow conditioning strategies similar to Palette [24]. The model input is constructed by concatenating the noisy target latent with the short-scan latent, i.e., $[z_t, z_{\text{short}}] \in \mathbb{R}^{8\times 618}$, where each latent representation has dimensionality 4×618 in our setting. Since z_{short} represents the same type of latent connectivity representation as the target z_{long}, it provides a natural fine-grained structural guide during denoising. Concatenating z_{short} with z_t allows the transformer to directly attend to the short-scan connectivity while predicting the denoised latent.

Conditioning information is incorporated through the vector $\mathbf{h}$ with three elements: z_{short}, SC_{flat}, and e_{cov}. Here, z_{short} denotes the latent representation of short-scan FC, and SC_{flat} denotes the flattened structural connectivity matrix. Structural connectivity provides an anatomical prior that guides generation toward plausible connectivity patterns, although it plays a secondary role relative to the short-scan latent representation. Demographic covariates are encoded as a 128-dimensional embedding composed of a one-hot encoding of sex and a sinusoidal embedding of age (in months), denoted by e_{cov}.

These components are fused with the timestep embedding and projected through MLP layers to produce a conditioning representation h_{aux}, which modulates each transformer block via adaptive layer normalization (AdaLN-Zero), as shown in Figure 2. Through this mechanism, covariates and structural connectivity influence the denoising process at a coarse level by modulating the transformer activations, while the concatenated short-scan latent z_{short} provides direct fine-grained guidance through the token stream.

3.4 Phenotype-Aware Loss Steering

Likelihood-based diffusion training optimizes reconstruction fidelity but does not explicitly preserve phenotype-discriminative variation. When multiple plausible connectivity structures are consistent with the conditioning signal, likelihood-based training can emphasize population-level regularities, potentially attenuating subject-specific variability in practice. For connectome generation, however, structural fidelity alone is insufficient: generated FC must retain the variance that supports downstream phenotype prediction. In other words, generated samples should remain both *structurally plausible* and *phenotype-informative*.

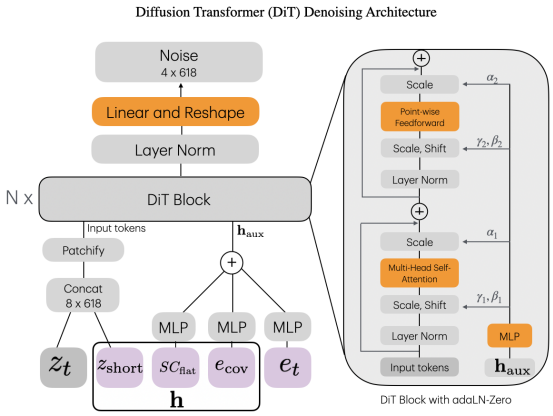

Fig. 2: Diffusion Transformer (DiT) denoising architecture [22]. The denoiser operates on latent tokens obtained from functional connectivity representations. Input tokens are formed by concatenating the noisy latent z_t with the short-scan representation z_{short}. Auxiliary information $\mathbf{h} = [z_{\text{short}}, SC_{\text{flat}}, e_{\text{cov}}]$ combines the short-scan latent with embedding vectors e_{cov} derived from subject covariates and the flattened structural connectivity SC_{flat}. These inputs are projected through MLPs to produce a conditioning embedding $\mathbf{h}_{\text{aux}}$, which modulates each transformer block through adaptive layer normalization (AdaLN-Zero). During fine-tuning, LoRA adapters (orange) are applied to the attention and feedforward layers.

This creates a fidelity–utility tension in short-to-long FC synthesis: improving similarity to long-scan FC does not necessarily preserve phenotype-relevant predictive structure.

To address this issue, we adapt Differential Reward Fine-Tuning (DRaFT) [8] to steer the generative process toward phenotype-preserving solutions, as illustrated in Figure 1. DRaFT operates on sampled outputs rather than intermediate activations, allowing the loss to act directly on reconstructed FC that will ultimately be used at inference. This property is important in our setting because the phenotype predictor operates in FC space, and steering must therefore influence the final generated matrices rather than only latent representations.

Fine-tuning is performed through parameter-efficient LoRA adapters [17], applied to the transformer linear layers responsible for attention projections and output transformations (highlighted in orange in Figure 2). In particular, LoRA modules are inserted in the query, key, and value projections as well as the attention output projections, the MLP projections, and the final output projection. These layers control how the model attends to conditioning signals and how latent features are combined to produce the denoised representation. Adapting these components allows the model to modify where attention is allocated and how predictive information is integrated, while keeping the pretrained generative backbone largely unchanged. This design reflects the assumption that phenotype-

aware adaptation requires a moderate adjustment of feature interactions rather than full retraining of the diffusion model.

Let

$$\hat{F}_{\mathrm{LoRA}} = D(z_{\theta_{\mathrm{LoRA}}}(\mathbf{h})) \tag{8}$$

denote a sampled FC matrix conditioned on $\mathbf{h}$ after fine-tuning, where the VAE decoder D remains frozen. Following DRaFT [8], we propagate gradients through the last K sampling steps using DDIM [31] so that the loss acts on the generated FC itself. Let $\{u_s\}_{s=0}^{S}$ denote the reduced DDIM timestep sequence (with $S \ll T$), where $u_0 = 0$ and u_1 is the first nonzero DDIM timestep. We match the denoising prediction at u_1:

$$\epsilon_{\mathrm{LoRA}} = \epsilon_{\theta_{\mathrm{LoRA}}}(z_{u_1}, u_1, \mathbf{h}), \tag{9}$$

$$\epsilon_{\mathrm{base}} = \epsilon_{\theta_{\mathrm{base}}}(z_{u_1}, u_1, \mathbf{h}). \tag{10}$$

Here, θ_{LoRA} denotes only the small set of trainable low-rank adapter parameters, whereas θ_{base} denotes the pretrained diffusion model parameters before fine-tuning, which remain completely frozen during adaptation.

The steering objective is defined as

$$\mathcal{L}_{\mathrm{steer}} = \gamma \left\| \epsilon_{\mathrm{LoRA}} - \epsilon_{\mathrm{base}} \right\|_2^2 + \lambda \left\| R(\hat{F}_{\mathrm{LoRA}}) - R(F_{\mathrm{long}}) \right\|_2^2. \tag{11}$$

The first term is a one-step noise-prediction anchor whose form is motivated by the per-step KL decomposition used in DPOK [11] and whose stabilizing role during reward-based diffusion fine-tuning follows the empirical strategy explored in DRaFT [8]. The second term encourages generated FC to preserve phenotype-relevant structure as measured by the frozen predictor $R(\cdot)$, whose parameters are not updated during adaptation.

Phenotypic traits are inherently noisy and only partially explained by FC. Directly optimizing prediction error against the observed phenotype can therefore push the generator toward unrealistic connectomes that attempt to fit irreducible noise. To mitigate this effect, we employ a distillation strategy in which the reward matches the predictions produced by high-quality long-scan FC rather than the raw phenotype target. This approach encourages the model to preserve the phenotype-discriminative component of connectivity while avoiding overfitting to unexplained variance.

Noise anchoring [11] is incorporated during fine-tuning to maintain alignment with the pretrained diffusion model. This term balances the fidelity–utility trade-off: stronger regularization favors broader generative coverage, whereas weaker regularization allows greater emphasis on phenotype-discriminative utility. In practice, this regularization stabilizes training and prevents distributional drift during reward optimization.

An additional practical advantage of the distillation formulation is that it allows fine-tuning on the full training set. While phenotype labels may be missing for a subset of subjects, distilled targets derived from long-scan FC provide a

reward signal for every training example, enabling efficient adaptation even in limited-label regimes.

Together, DRaFT-based reward steering, LoRA parameter-efficient adaptation, and noise anchoring regularization enable phenotype-aware connectome generation that preserves predictive structure while remaining faithful to the pretrained generative model.

4 Experiments

4.1 Experimental Setup

Dataset. We use the Adolescent Brain Cognitive Development (ABCD) Study Release 4.0 data [33], one of the largest publicly available neuroimaging datasets, containing children aged 9–10 at baseline with longitudinal follow-up. We consider resting-state fMRI scans [3], which have been shown to capture connectivity patterns predictive of behavioral and cognitive phenotypes [7, 10].

Resting-state fMRI data were acquired across 21 sites using 3T scanners from Siemens, General Electric, and Philips, following a harmonized acquisition protocol (TR = 800 ms, TE = 30 ms, flip angle = 90°, 2.4 mm isotropic resolution) [6]. The standard protocol includes four 5-minute runs, totaling 20 minutes of acquisition per subject. We processed the raw data using the Surface-Based Connectivity Integration (SBCI) pipeline [9]. This pipeline performs standard motion correction and nuisance regression before projecting the volumetric BOLD signal onto cortical surface vertices. To define functional nodes, we applied the Schaefer atlas with 100 parcels [27] to the surface data. For each region of interest (ROI), we computed a representative time series by averaging the BOLD signals of all vertices within the parcel. To ensure a uniform high-quality reference for training, we included only subjects with at least 20 minutes of data, truncating the concatenated time series to exactly 20 minutes (1,500 time points), resulting in a final cohort of 11,156 scans. FC matrices are computed as Pearson correlations between ROI time series, producing symmetric matrices $FC \in \mathbb{R}^{100\times100}$. For short scans of duration τ, correlations are computed using only the first τ minutes, while the full 20-minute matrix serves as the high-quality reference.

The SC matrices were also computed using the SBCI pipeline [9]. Unlike traditional voxel-based approaches, SBCI maps white matter tractography endpoints directly to the cortical surface to estimate a continuous representation of SC. This continuous connectivity was then discretized using the same Schaefer 100-parcel atlas used for the functional data. The resulting matrix $SC \in \mathbb{R}^{100\times100}$ quantifies the number of reconstructed streamlines connecting each pair of regions. To mitigate the heavy-tailed distribution of streamline counts, we applied a logarithmic transformation $\log(1 + x)$ to the final matrices.

Demographic covariates (age and sex) are obtained from curated ABCD tables distributed through the *NIMH Data Archive* (NDA). As the target phenotype we use the *Cognition Total Composite Score (Age-Corrected Standard Score)* from the NIH Toolbox, which aggregates multiple cognitive domains (*e.g.*, executive function, memory, attention, and processing speed).

Training and Evaluation Protocol. The dataset is split into training, validation, and test sets using an 80/10/10 scan-level partition. We consider input durations $\tau \in \{1, \ldots, 10\}$ minutes, targeting the 20-minute ground-truth FC. As generative-task baselines, we include BicycleGAN with an MLP backbone [38] and a conditional VAE with a U-Net backbone [30]; full architectural and training details are provided in the Supplementary Material. While a Kernel Ridge Regression (KRR) model with a correlation kernel [13] provides the reward signal for diffusion guidance, final evaluation uses a cross-architecture Linear Ridge Regression (LRR) trained on the 20-minute matrices. This tests whether synthesized utility gains transfer beyond the predictor used for reward optimization.

Final results are reported on the 613 test scans for whom the selected phenotypes are available. Because diffusion sampling is stochastic, evaluation metrics are reported as a Monte Carlo estimate of the expected performance of a single generated connectivity matrix. For each subject, we draw ten independent samples, evaluate each against the reference connectivity or phenotype target, and report the average of the ten metric values to reduce stochastic variance while estimating expected single-sample quality. Because ABCD is longitudinal, the test set may include follow-up scans of individuals seen at an earlier timepoint in training; the Supplementary Material isolates entirely new subjects and shows the same trends.

4.2 Experimental Results

Reconstruction Fidelity. We first assess whether SALD preserves realistic FC structure while improving phenotype prediction, since predictive gains alone may arise from spurious, biologically implausible connectivity patterns. Figure 3 quantifies reconstruction fidelity to the 20-minute reference (FC_{long}) using Frobenius distance and correlation across scan durations τ. SALD shows the strongest overall agreement with FC_{20} among the generative baselines and consistently outperforms Conditional DDPM, particularly at shorter, noisier scan durations. Because phenotype steering does not explicitly constrain the model to preserve realistic FC structure, noise anchoring keeps the fine-tuned denoising process close to the pretrained diffusion model and limits structural drift. We examine its contribution through ablation experiments later in this section.

Phenotype Prediction Performance. We next evaluate whether the generated FCs preserve information relevant to the NIH Toolbox Total Cognition Composite Score, the phenotype SALD was finetuned on. Figure 4 reports MSE, correlation, and recovered predictive variance relative to the 20-minute reference. For the latter, we average repeated predictions within each subject before computing variance, so it reflects stable between-subject differences rather than phenotype-irrelevant fluctuations. Two observations are notable. First, the Conditional DDPM baseline consistently underperforms across all durations and metrics, indicating that likelihood-based diffusion training alone fails to preserve the behaviorally discriminative structure of FC even when reconstruction fidelity is reasonable: matching the distribution of FC does not guarantee preservation of its predictive signal. Second, SALD substantially improves prediction across all

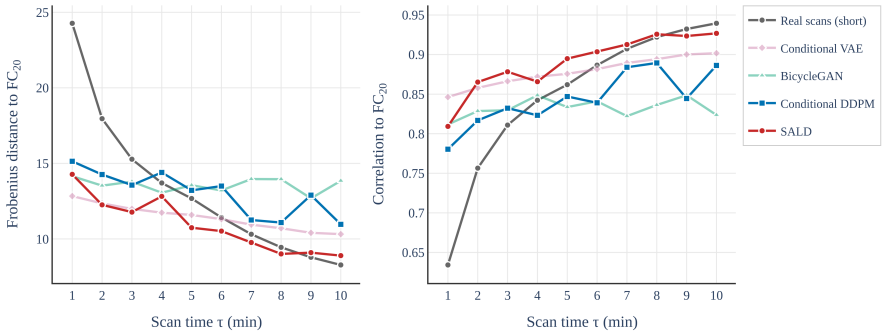

Fig. 3: Reconstruction fidelity to the 20-minute reference FC (FC_{20}) across input scan durations τ. Left: Frobenius distance (↓); right: Pearson correlation (↑). SALD achieves the strongest overall agreement with FC_{20} among the generative baselines and consistently improves over Conditional DDPM. Conditional VAE is competitive in Frobenius distance, consistent with its explicit reconstruction objective. Real short-scan FCs are included as a non-generative reference.

metrics and durations and recovers more long-scan predictive variance, showing its gains concentrate in meaningful subject differences. By steering generation toward phenotype-relevant structure, SALD recovers predictive utility largely lost under pure likelihood training. Importantly, this steering signal is itself noisy: the predictor explains only about 25% of phenotype variance using long-scan FC (dashed blue line in Fig. 4), reflecting the intrinsic difficulty of neuroimaging-based prediction. Despite this weak supervision, SALD consistently improves performance, indicating that it concentrates updates on the explainable structure of FC. These gains are particularly meaningful in the short-scan regime, where improved predictive utility can translate into reduced scanning time, and diminish as τ approaches longer durations (9–10 min), where short scans already approximate the reference.

Generalization to additional phenotypes. To assess if SALD preserves phenotype-relevant information beyond the fine-tuning target, we evaluate UPPS negative urgency, Child Behavior Checklist (CBCL) Total Problems, and Sleep Disturbance Scale for Children (SDS) total sleep disturbance [1,4,35]. As shown in Figure 5, SALD generally recovers more of the 20-minute predictive variance and achieves comparable or stronger phenotype correlations than C-DDPM, indicating that its phenotype utility extends to outcomes unseen during fine-tuning.

Reliability via differential identifiability. Reconstruction and prediction metrics do not directly measure whether generated FCs preserve subject-specific patterns. We therefore compute differential identifiability, defined as the average similarity to the same subject's long-scan FC minus the average similarity to other subjects'. Higher values indicate stronger subject specificity. To isolate reward alignment from likelihood-based training, we compare SALD with its Conditional DDPM base model. As shown in Table 1, the two perform similarly at

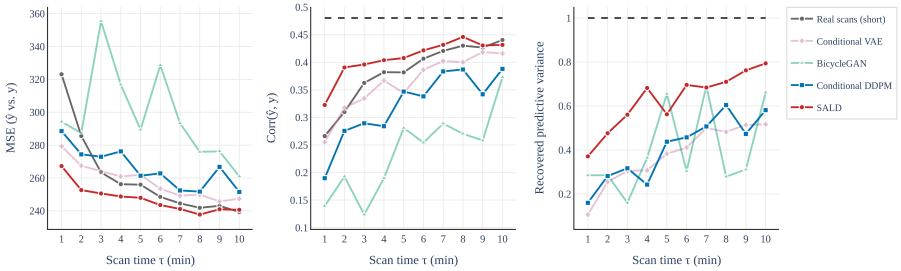

Fig. 4: Phenotype prediction across scan durations τ: MSE (↓), correlation (↑), and recovered predictive variance (↑). In the short-scan regime, SALD improves over real short-scan FCs and is the strongest generative baseline overall, recovering more long-scan predictive variance than Conditional DDPM; BicycleGAN shows unstable phenotype utility across durations despite stable reconstruction error. Dashed lines denote the 20-minute reference.

Table 1: Mean differential identifiability across stochastic FC samples (higher is better). The two methods are comparable at $\tau = 1$; SALD exceeds Conditional DDPM from $\tau = 2$ onward, with a widening margin.

Time (min)	1	2	3	4	5	6	7	8	9	10
Conditional DDPM	4.21	8.49	9.92	11.40	12.84	13.92	16.27	17.12	16.24	18.17
SALD	4.10	10.57	12.97	15.22	15.76	17.77	18.42	19.25	20.46	20.83

$\tau = 1$, while SALD consistently achieves higher identifiability from $\tau = 2$ onward, indicating that reward alignment better preserves subject-specific connectivity than likelihood training alone. Differential identifiability captures subject-level distinguishability, one facet of reliability; a more exhaustive assessment is needed to fully characterize the reliability of generated FCs.

Ablation Study. In Table 2, we isolate each SALD component: auxiliary conditioning (SC and covariates), the one-step denoising anchoring term, and the distillation objective. We focus on $\tau = 1$ to 3 minutes, where short-scan connectivity is noisiest and each design choice matters most.

Conditioning. Removing SC and covariates jointly ($-$(SC,cov)) degrades both reconstruction fidelity and phenotype prediction across all durations, largest at $\tau = 1$, and removing SC alone ($-$SC) produces a comparable drop, indicating SC is a substantial contributor to the conditioning signal. The sharpest degradation comes from removing the short-scan input ($-\mathrm{FC}_\tau$), which is constant across τ since the model loses its duration-dependent connectivity. Together, SC and demographics provide complementary guidance that partially offsets the information lost in short acquisitions, while short-scan drives performance.

Anchoring. One-step denoising anchoring governs an explicit trade-off between generative fidelity and reward optimization. Removing it allows the model

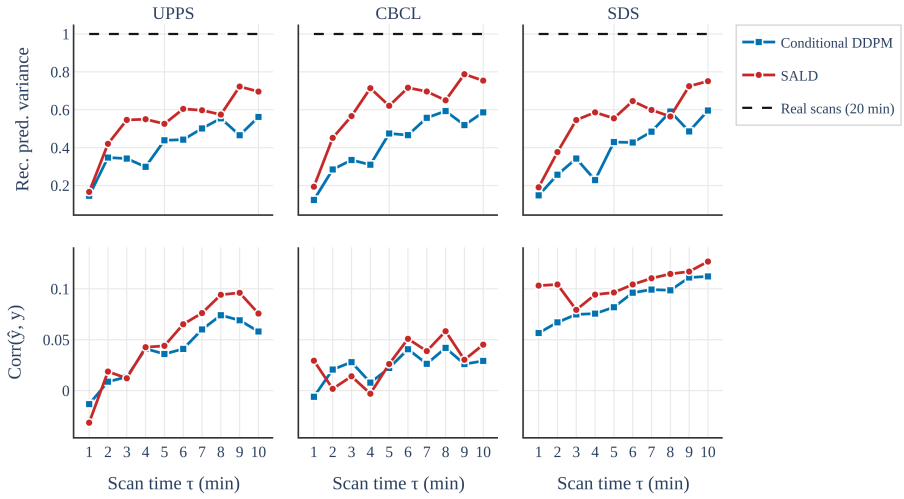

Fig. 5: Phenotype utility across scan duration. Top: recovered predictive variance relative to real 20-min scans; bottom: correlation between predicted and observed phenotypes. SALD generally recovers more predictive variance and achieves comparable or stronger phenotype prediction than C-DDPM across UPPS, CBCL, and SDS.

to pursue the reward signal more aggressively, which can reduce Frobenius distance at the cost of distributional stability. The mixed pattern across τ values reflects this trade-off: at very short durations where inputs are highly noisy, unconstrained reward optimization incidentally improves reconstruction, while at longer durations the regularization becomes more important for preventing drift from the pretrained prior. Rather than indicating instability, this pattern confirms that anchoring balances fidelity and utility.

Distillation vs. direct supervision. Replacing the distillation objective with direct phenotype supervision (Direct-y) produces the largest and most consistent performance drop across all τ and metrics. This result serves as a deliberate negative control: if raw phenotype labels provided a sufficient training signal, distillation would be unnecessary. The degradation demonstrates that phenotype labels are too noisy to supervise connectivity generation directly; the irreducible variance in behavioral scores pushes the generator toward unrealistic connectomes that overfit unexplained noise. Distillation through the frozen long-scan predictor smooths this signal, providing stable supervision that preserves connectivity structure while recovering phenotype-discriminative variation.

Together these results confirm that auxiliary conditioning, anchoring, and distillation each contribute distinct and complementary functions: structural guidance for reconstruction, stability during reward optimization, and noise-robust phenotype alignment respectively.

Table 2: Ablation analysis of SALD at short scan durations ($\tau = 1$–3 min); Frob./MSE both lower is better. Removing SC and covariates ($-$(SC,cov)) or SC alone ($-$SC) degrades both; removing the short-scan input ($-\mathrm{FC}_\tau$, constant across τ) drops sharply. $-$Anchoring exposes the fidelity–utility trade-off, and Direct-y yields the largest drop.

τ	Model	Base	$-(\mathrm{SC}, \mathrm{cov})$	$-\mathrm{SC}$	$-\mathrm{FC}_\tau$	$-$Anchoring	Direct-y
1	SALD	14.27 267.25	14.65 285.06	14.77 285.15	14.36 272.92	13.32 265.95	14.27 266.84
1	C-DDPM	15.13 288.52	15.83 304.55	16.62 307.50	17.78 311.95	–	–
2	SALD	12.25 252.82	14.26 257.55	13.58 256.90	14.36 272.92	12.44 250.56	15.43 257.97
2	C-DDPM	14.26 274.23	15.83 282.21	15.01 283.95	17.78 311.95	–	–
3	SALD	11.77 250.73	12.99 255.30	13.25 253.00	14.36 272.92	12.21 248.85	14.36 256.04
3	C-DDPM	13.55 272.74	15.53 290.48	13.85 276.23	17.78 311.95	–	–

5 Limitations and Clinical Perspective

Our results rest on a single ABCD configuration (SBCI, Schaefer 100-parcel atlas), so generalization to other cohorts, protocols, populations, or processing choices awaits external validation. Our evaluation spans three complementary aspects of generation quality: reconstruction fidelity, phenotype utility, and differential identifiability, assessing whether generated FCs preserve long-scan-like structure, phenotype-relevant information, and subject-specific differences. Broader reliability analysis, including test-retest behavior, would require repeated long-scan acquisitions per subject, unavailable at scale in this cohort.

Clinical perspective. Although motivated partly by the burden of lengthy pediatric imaging protocols, SALD targets population-level research, not individual clinical decisions, recovering long-scan-like phenotype-relevant signal that is attenuated in short acquisitions and in the likelihood-based diffusion baseline. We do not evaluate diagnostic outcomes or patient-level reliability: clinical use would require prospective external validation, calibrated uncertainty, robustness across populations and acquisition settings, and safeguards against clinically misleading features. SALD is therefore not a substitute for acquiring sufficient data in individual diagnostic settings.

6 Conclusion

We introduced SALD, a structure-conditioned latent diffusion framework that synthesizes long-scan-like functional connectivity from short fMRI acquisitions. In the ABCD cohort, our results surface a fidelity-utility tension: likelihood-based diffusion can improve reconstruction fidelity to long-scan FC while attenuating the inter-subject variation relevant to phenotype prediction. We address this with reward-aligned LoRA adaptation combining one-step denoising anchoring and phenotype distillation, stabilizing guidance while focusing optimization on phenotype-discriminative connectivity. Across durations, SALD improves over task-matched baselines in both reconstruction fidelity and phenotype prediction, with the largest gains in the short-scan regime where connectivity is least reliable. Our findings highlight the value of aligning generative objectives with downstream scientific utility when synthesizing neuroimaging data.

References

1. Achenbach, T.M., Rescorla, L.A.: Manual for the ASEBA School-Age Forms & Profiles. University of Vermont, Research Center for Children, Youth, & Families, Burlington, VT (2001)
2. Birn, R.M., Molloy, E.K., Patriat, R., Parker, T., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V.: The effect of scan length on the reliability of resting-state fMRI connectivity estimates. NeuroImage **83**, 550–558 (2013). https://doi.org/10.1016/j.neuroimage.2013.05.099
3. Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic Resonance in Medicine **34**(4), 537–541 (1995). https://doi.org/10.1002/mrm.1910340409
4. Bruni, O., Ottaviano, S., Guidetti, V., Romoli, M., Innocenzi, M., Cortesi, F., Giannotti, F.: The sleep disturbance scale for children (SDSC): Construction and validation of an instrument to evaluate sleep disturbances in childhood and adolescence. Journal of Sleep Research **5**(4), 251–261 (1996). https://doi.org/10.1111/j.1365-2869.1996.00251.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2869.1996.00251.x
5. Cai, T.T., Ren, Z., Zhou, H.H.: Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation (2016)
6. Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., Orr, C.A., Wager, T.D., Banich, M.T., Speer, N.K., Sutherland, M.T., Riedel, M.C., Dick, A.S., Bjork, J.M., Thomas, K.M., Chaarani, B., Mejia, M.H., Hagler, D.J., Cornejo, M.D., Sicat, C.S., Harms, M.P., Dosenbach, N.U.F., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni, J.R., Kuperman, J.M., Fair, D.A., Dale, A.M.: The adolescent brain cognitive development (abcd) study: Imaging acquisition across 21 sites. Developmental Cognitive Neuroscience **32**, 43–54 (2018). https://doi.org/10.1016/j.dcn.2018.03.001
7. Chen, J., Tam, A., Kebets, V., Orban, C., Ooi, L.Q.R., Marek, S., Dosenbach, N.U.F., Bzdok, D., Holmes, A.J., Yeo, B.T.T.: Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. Nature Communications **13**, 2217 (2022). https://doi.org/10.1038/s41467-022-29766-8
8. Clark, K., Vicol, P., Swersky, K., Fleet, D.: Directly fine-tuning diffusion models on differentiable rewards. In: Kim, B., Yue, Y., Chaudhuri, S., Fragkiadaki, K., Khan, M., Sun, Y. (eds.) International Conference on Learning Representations. vol. 2024, pp. 4793–4822 (2024), https://proceedings.iclr.cc/paper_files/paper/2024/file/13e8be77982beb73d7ed0bbf122f9f3c-Paper-Conference.pdf
9. Cole, M., Murray, K., St-Onge, E., Risk, B., Zhong, J., Schifitto, G., Descoteaux, M., Zhang, Z.: Surface-based connectivity integration: An atlas-free approach to jointly study functional and structural connectivity. Human Brain Mapping **42**(11), 3481–3499 (2021)
10. Duffy, K.A., Helwig, N.E.: Resting-state functional connectivity predicts attention problems in children: Evidence from the ABCD study. NeuroSci **5**(4), 445–461 (2024). https://doi.org/10.3390/neurosci5040033
11. Fan, Y., Watkins, O., Du, Y., Liu, H., Ryu, M., Boutilier, C., Abbeel, P., Ghavamzadeh, M., Lee, K., Lee, K.: Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. In: Advances in Neural Information Processing Systems (2023)

12. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics **9**(3), 432–441 (2008)
13. He, T., Kong, R., Holmes, A.J., Sabuncu, M.R., Eickhoff, S.B., Bzdok, D., Yeo, B.T.T.: Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. NeuroImage **206**, 116276 (2020). https://doi.org/10.1016/j.neuroimage.2019.116276
14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS (2020), https://arxiv.org/abs/2006.11239
15. Ho, J., Salimans, T.: Classifier-free diffusion guidance. CoRR **abs/2207.12598** (2022), https://arxiv.org/abs/2207.12598
16. Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P.: Predicting human resting-state functional connectivity from structural connectivity. Proceedings of the National Academy of Sciences **106**(6), 2035–2040 (2009). https://doi.org/10.1073/pnas.0811168106
17. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Lu, W., Chen, W.: LoRA: Low-rank adaptation of large language models. In: ICLR (2022)
18. Luo, C.: Understanding diffusion models: A unified perspective. CoRR **abs/2208.11970** (2022), https://arxiv.org/abs/2208.11970
19. Marek, S., Tervo-Clemmens, B., Calabro, F.J., et al.: Reproducible brain-wide association studies require thousands of individuals. Nature **603**(7902), 654–660 (2022). https://doi.org/10.1038/s41586-022-04492-9
20. Mumford, J.A., Nichols, T.E.: Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. NeuroImage **39**(1), 261–268 (2008). https://doi.org/10.1016/j.neuroimage.2007.07.061
21. Ooi, L.Q.R., Orban, C., Zhang, S., Nichols, T.E., et al.: Longer scans boost prediction and cut costs in brain-wide association studies. Nature **644**(8077), 731–740 (2025). https://doi.org/10.1038/s41586-025-09250-1
22. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: ICCV (2023)
23. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10684–10695 (2022). https://doi.org/10.1109/CVPR52688.2022.01042
24. Saharia, C., Chan, W., Chang, H., Lee, C.A., Ho, J., Salimans, T., Fleet, D.J., Norouzi, M.: Palette: Image-to-image diffusion models. ACM Transactions on Graphics (TOG) **41**(4), 98:1–98:10 (2022). https://doi.org/10.1145/3528223.3530127
25. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE transactions on pattern analysis and machine intelligence **45**(4), 4713–4726 (2022)
26. Sarwar, T., Tian, Y., Yeo, B.T.T., Ramamohanarao, K., Zalesky, A.: Structure-function coupling in the human connectome: A machine learning approach. NeuroImage **226**, 117609 (2021). https://doi.org/10.1016/j.neuroimage.2020.117609
27. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T.: Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cerebral Cortex **28**(9), 3095–3114 (2018). https://doi.org/10.1093/cercor/bhx179
28. Shen, X., Finn, E.S., Scheinost, D., Rosenberg, M.D., Chun, M.M., Papademetris, X., Constable, R.T.: Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nature Protocols **12**(3), 506–518 (2017). https://doi.org/10.1038/nprot.2016.178

29. Smolders, L., De Baene, W., Rutten, G.J., van der Hofstad, R., Florack, L.: Can structure predict function at individual level in the human connectome? Brain Structure and Function **229**(5), 1209–1223 (2024). https://doi.org/10.1007/s00429-024-02796-2
30. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 28. Curran Associates, Inc. (2015), https://proceedings.neurips.cc/paper_files/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf
31. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: ICLR (2021)
32. Tan, Y.F.: Sfc-gan: A generative adversarial network for brain functional and structural connectome translation. arXiv preprint arXiv:2501.07055 (2025)
33. Volkow, N.D., Koob, G.F., Croyle, R.T., Bianchi, D.W., Gordon, J.A., Koroshetz, W.J., P'erez-Stable, E.J., Riley, W.T., Bloch, M.H., Conway, K., Deeds, B.G., Dowling, G.J., Grant, S., Howlett, K., Matochik, J.A., Morgan, G.D., Murray, M.M., Noronha, A., Spong, C.Y., Wargo, E.M., Warren, K.R., Weiss, S.R.B.: The conception of the abcd study: From substance use to a broad nih collaboration. Developmental Cognitive Neuroscience **32**, 4–7 (2018). https://doi.org/10.1016/j.dcn.2017.10.002
34. Wallace, B., Gokul, A., Ermon, S., Naik, N.: End-to-end diffusion latent optimization improves classifier guidance. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 7280–7290 (October 2023)
35. Watts, A.L., Smith, G.T., Barch, D.M., Sher, K.J.: Factor structure, measurement and structural invariance, and external validity of an abbreviated youth version of the UPPS-P Impulsive Behavior Scale. Psychological Assessment **32**(4), 336–347 (2020). https://doi.org/10.1037/pas0000791
36. Xu, J., Liu, X., Wu, Y., Tong, Y., Li, Q., Ding, M., Tang, J., Dong, Y.: Imagereward: Learning and evaluating human preferences for text-to-image generation (2023), https://arxiv.org/abs/2304.05977
37. Zalesky, A., Sarwar, T., Tian, Y., Liu, Y., Yeo, B.T.T., Ramamohanarao, K.: Predicting an individual's functional connectivity from their structural connectome: Evaluation of evidence, recommendations, and future prospects. Network Neuroscience **8**(4), 1291–1309 (2024). https://doi.org/10.1162/netn_a_00400
38. Zhu, J.Y., Zhang, R., Pathak, D., Darrell, T., Efros, A.A., Wang, O., Shechtman, E.: Toward multimodal image-to-image translation. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 465–476. NIPS'17, Curran Associates Inc., Red Hook, NY, USA (2017)
39. Zong, Y., Jing, C., Chan, J.H., Wang, S.: Brainnetdiff: Generative ai empowers brain network construction via multimodal diffusion. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 2093–2097. IEEE (2024)
40. Zuo, Q., Li, R.: Diffgan-f2s: Symmetric and efficient denoising diffusion gans for structural connectivity prediction from brain fmri. arXiv preprint arXiv:2309.16205 (2023)