

A Demographic-Conditioned Variational Autoencoder for fMRI Distribution Sampling and Generation of Unbiased Latents



Anton Orlichenko¹

Gang Qu¹ Ziyu Zhou² Anqi Liu³ Hong-Wen Deng³ Yu-Ping Wang^{1,2}

Tulane University, Department of Biomedical Engineering¹, Department of Computer Science², School of Medicine, Center for Biomedical Informatics and Genomics³



Scan me!

Introduction

Challenges

- fMRI is a non-invasive technique for measuring neural activity via the blood oxygen level-dependent (BOLD) signal
- fMRI and genomic data may be hard to acquire and large public repositories may only be accessible to authorized researchers
- It has been shown [2][4] that demographic data can confound fMRI features and taint the interpretability of predictive models

Goals

- Generation of Synthetic Data** Synthetic data generated by our model should be indistinguishable from real data and capture the full spectrum of individual variability
- Conditioning on Demographics** Data generation should be based on user-specified demographic features, recapitulating group differences in the population
- Decorrelation of Latent Features from Demographics** Latent features produced by the encoder $\mathbf{z} = E_\phi(\mathbf{x})$ should be uncorrelated with respect to demographic features \mathbf{y} . All demographic-related features in the reconstructed data $\hat{\mathbf{x}} = D_\theta(\mathbf{z}, \mathbf{y})$ should come from \mathbf{y} , not the latent features \mathbf{z} .

Validation

- The Philadelphia Neurodevelopmental Cohort (PNC) is a dataset of 1,500+ healthy adolescents and the Bipolar and Schizophrenia Network for Intermediate Phenotypes (BSNIP) is a dataset of 405 schizophrenia patients (SZ) and healthy controls

Variational Autoencoder (VAE)

The variational autoencoder is a machine learning model for generating synthetic data according to some preexisting distribution of real data.

A VAE is trained by maximizing the ELBO objective [1]:

$$\begin{aligned} L_{\theta, \phi} &= \mathbb{E}_{z \sim q_\phi(\cdot | x)} \left[\ln p_\theta(x, z) \right] \\ &= \mathbb{E}_{z \sim q_\phi(\cdot | x)} [\ln p_\theta(x | z) - D_{KL}(q_\phi(z | x) \| p_\theta(z))] \end{aligned} \quad (1)$$

When not considering interdependencies between latent features, the training loss takes the following form:

$$L_{\theta, \phi} = \|x - D_\theta(z)\|_2^2 + N\sigma_z^2 + \|\mu_z\|_2^2 - N\ln\sigma_z^2, \quad (2)$$

In multidimensional latent case, the KL divergence D_{KL} part is more complicated [3]:

$$D_{KL}(\mathcal{N}(\mu_z, \Sigma_z) \| \mathcal{N}(\mathbf{0}, \mathbf{I})) = \frac{1}{2} [\text{tr}(\Sigma_z) + \mu_z^\top \mu_z + \log(\det(\Sigma_z))] \quad (3)$$

The calculation of and backpropagation through the log determinant presents issues in practice which we remedy with a modified loss function.

References

- Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings, 2014.
- Jingwei Li, Danilo Bzdok, Jianzhong Chen, Angela Tam, Leon Qi Rong Ooi, Avram J Holmes, Tian Ge, Kaustubh R Patil, Mbemba Jabbi, Simon B Eickhoff, B T Thomas Yeo, and Sarah Genon. Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity. *Sci. Adv.*, 8(11):eabj1812, March 2022.
- Kevin P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023.
- Anton Orlichenko, Grant Daly, Ziyu Zhou, Anqi Liu, Hui Shen, Hong-Wen Deng, and Yu-Ping Wang. ImageNomer: Description of a functional connectivity and omics analysis tool and case study identifying a race confound. *Neuroimage Rep.*, 3(4):100191, December 2023.

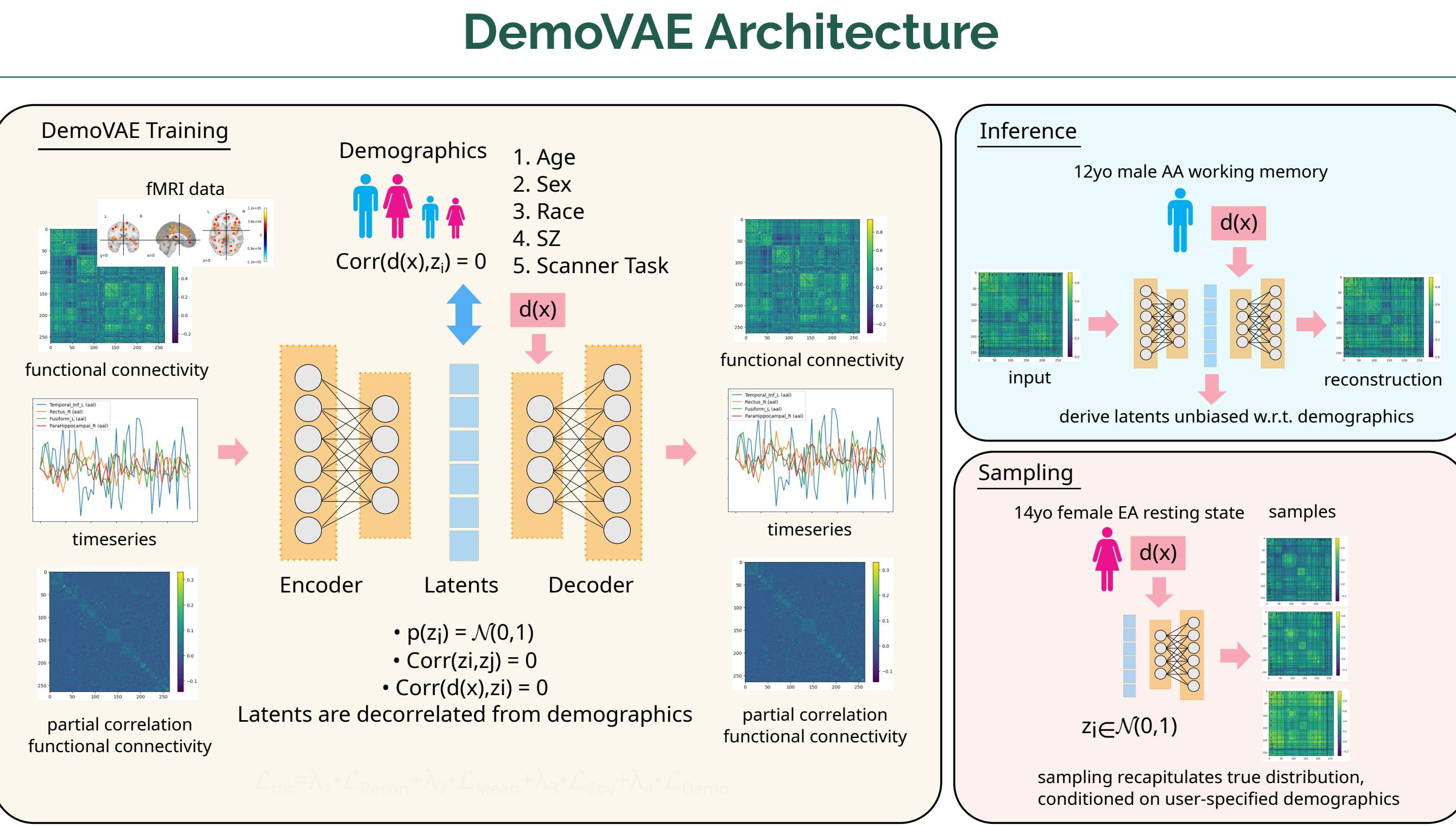


Figure 1. Overview of the DemoVAE model. DemoVAE learns to decorrelate latents from demographic information during training. Inference and sampling represent the two use cases for the trained model.

Demographic-Conditioned and Decorrelated Variational Autoencoder (DemoVAE)

We simplify the reconstruction and KL divergence loss terms, and add terms penalizing demographic-latent space correlations as well as providing classifier guidance for generated data.

$$\begin{aligned} \mathcal{L}_{\text{Recon}} &= \frac{1}{N} \sum_i^N \|\mathbf{x}_i - D_\theta(\mathbf{z}_i, \mathbf{y}_i)\|_2^2 \\ \mathcal{L}_{\text{Cov}} &= \frac{1}{N} \|\mathbf{Z}\mathbf{Z}^\top - N\mathbf{I}\|_F^2 \\ \mathcal{L}_{\text{Mean}} &= \frac{1}{NN_z} \sum_i^N \|\mu_{\mathbf{z}_i}\|_2^2, \\ \mathcal{L}_{\text{Demo}} &= \frac{1}{N_z N_y} \sum_j^N \sum_k^y \|\rho_{\mathbf{z}_j, \mathbf{y}_k}\|_2^2 \\ \mathcal{L}_{\text{Guide}} &= \frac{1}{NN_y} \sum_i^N \left\{ \begin{array}{ll} \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_2^2, & \mathbf{y}_i \text{ continuous} \\ -\sum_c \mathbf{y}_{i,c} \log(\mathbf{p}_{i,c}) & \mathbf{y}_i \text{ categorical} \end{array} \right. \end{aligned} \quad (4)$$

The total loss function for training the DemoVAE is:

$$\mathcal{L} = \mathcal{L}_{\text{Recon}} + \lambda_1 \mathcal{L}_{\text{Cov}} + \lambda_2 \mathcal{L}_{\text{Mean}} + \lambda_3 \mathcal{L}_{\text{Demo}} + \lambda_4 \mathcal{L}_{\text{Guide}} \quad (5)$$

Commercialization Potential

- Generating latents from fMRI data with DemoVAE removes demographics-related bias.
- Drug trials may find significant correlations between drug and patient outcomes that are partly or mostly due to demographics differences.
- Use of the DemoVAE tool may save drug companies money by not investing in drugs or treatments where results were biased by demographics.

Funding sources: NIH (grants R01 GM109068, R01 MH104680, R01 MH107354, P20 GM103472, R01 EB020407, R01 EB006841, R56 MH124925, 5U19 AG055373), NSF (#1539067).

DemoVAE Samples Accurately Capture Data Distribution

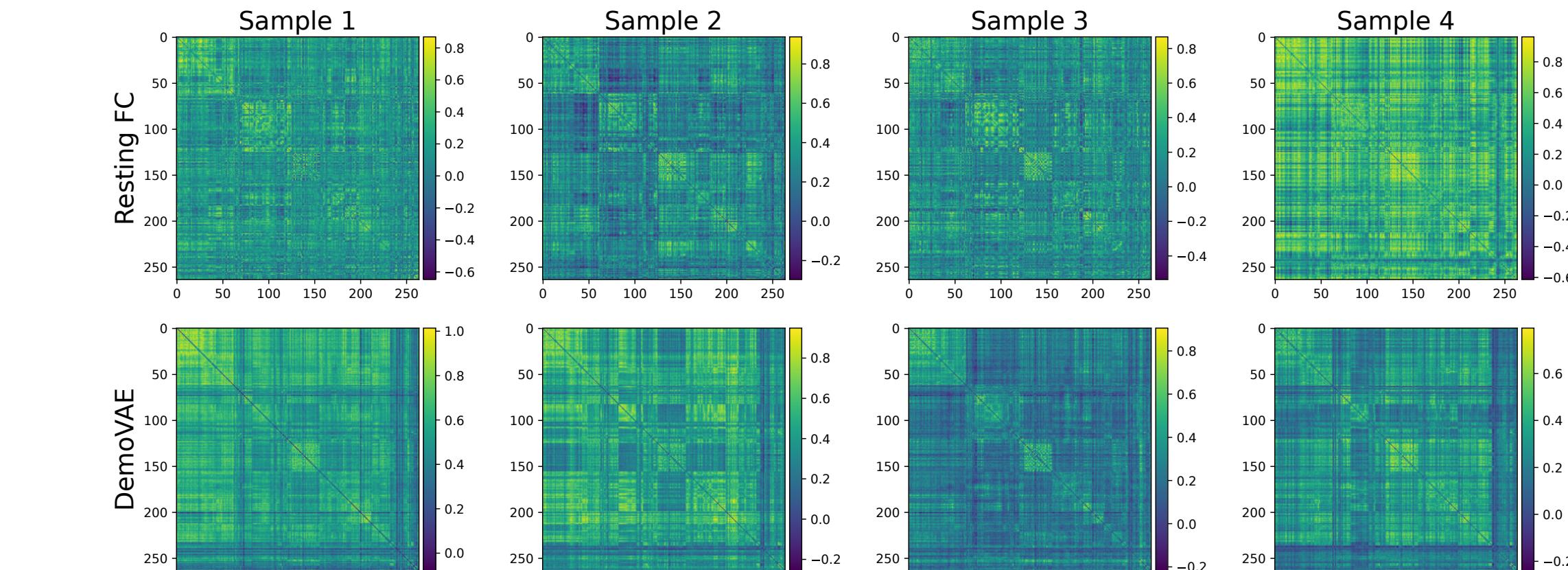


Figure 2. Sampled FC matrices for real PNC resting state scans compared to synthetic DemoVAE data.

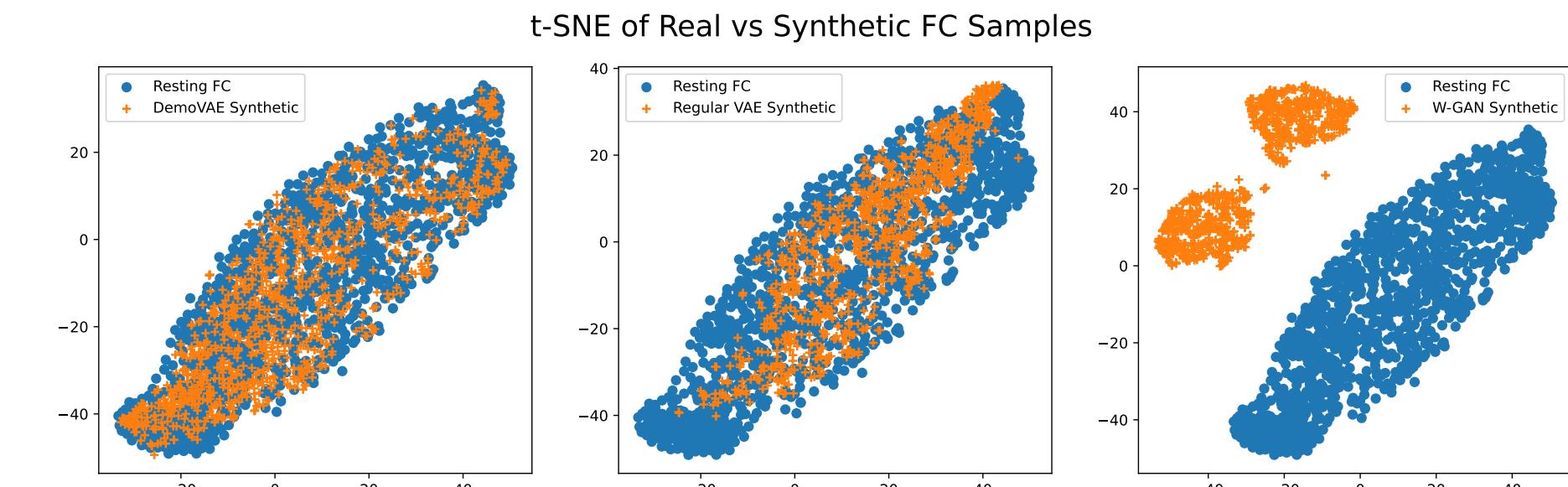


Figure 3. t-SNE embeddings of synthetic FC data from DemoVAE, traditional VAE, and W-GAN models overlaid on top of t-SNE embeddings of real resting state FC data from the PNC dataset. DemoVAE is best at capturing the full distribution of FC data.

Synthetic Samples Recapitulate Group Differences

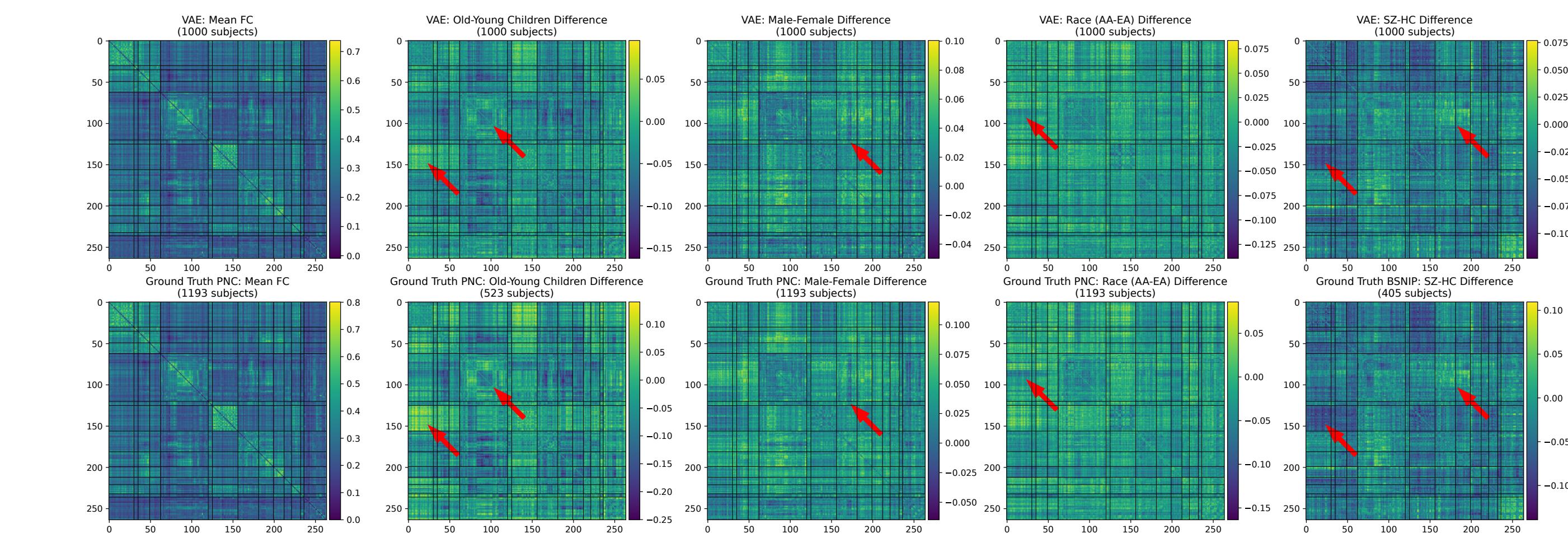


Figure 4. Group FC differences using real data and using synthetic data generated by DemoVAE conditioned on appropriate demographic input. DemoVAE qualitatively recapitulates group differences in the PNC (mean, age, sex, race) and BSNIP (SZ diagnosis) datasets.

DemoVAE Removes Demographic Bias

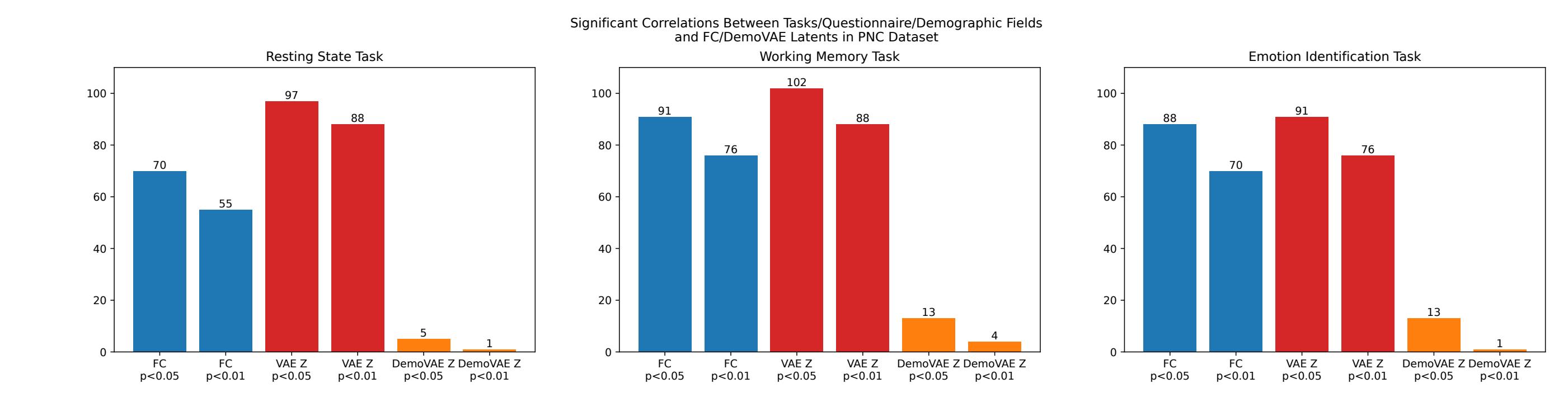


Figure 5. Correlation of computerized battery and clinical fields with fMRI FC data versus traditional VAE or decorrelated DemoVAE latents in the PNC dataset. There were a total of 169 fields in the PNC dataset. Once de-biased with DemoVAE, FCs are significantly correlated with only a few fields.