2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871803

Mind the gap: functional network connectivity interpolation between schizophrenia patients and controls using a variational autoencoder

Xinhui Li[1,2], Eloy Geenjaar[1,2], Zening Fu[2], Sergey Plis[2,3], and Vince Calhoun[1,2]

***Abstract*— Mental disorders such as schizophrenia have been challenging to characterize due in part to their heterogeneous presentation in individuals. Most studies have focused on identifying groups differences and have typically ignored the heterogeneous patterns within groups. Here we propose a novel approach based on a variational autoencoder (VAE) to interpolate static functional network connectivity (sFNC) across individuals, with group-specific patterns between schizophrenia patients and controls captured simultaneously. We then visualize the original sFNC in a 2D grid according to the samples in the VAE latent space. We observe a high correspondence between the generated and the original sFNC. The proposed framework facilitates data visualization and can potentially be applied to predict the stage that a subject falls within a disorder continuum as well as characterize individual heterogeneity within and between groups.**

I. INTRODUCTION

Schizophrenia is a lifelong mental disorder which affects perception and behavior with a prevalence of just under 1% in the population and it is among the top causes of disability worldwide [1], [2]. Relying on expert assessment or patient self-report data, the current diagnostic system is insufficient to accurately capture biological deficits and symptom progression [3]. Recent neuroimaging studies have shown that schizophrenia can be characterized by functional dysconnectivity [4], [5] and supervised learning approaches have been developed to predict disease conditions from functional connectivity data [6], [7]. However, this disorder is known to be highly heterogeneous, and there is debate over whether this reflects multiple diseases within a clinical syndrome or heterogeneity within the disorder itself [8]. Group averages and supervised classification approaches are not sufficient for its characterization. There is thus a need to develop unsupervised learning approaches that can characterize individual variability within and between groups.

Unsupervised learning algorithms such as variational autoencoders (VAEs) [9] are capable of learning low-dimensional latent representations from data without target labels. A VAE consists of an encoder and a decoder, where the encoder reduces high-dimensional inputs into low-dimensional distributions and the decoder then reconstructs the inputs according to the samples from those latent distributions. There are two main advantages to using a VAE. Firstly, the encoder performs dimension reduction and can learn latent distributions from the data in an unsupervised manner. Secondly, the decoder can generate continuous artificial data that resembles real data by sampling from the latent distributions. Recent work has demonstrated that latent representations of neuroimaging data learned by a VAE can be used to predict schizophrenia diagnosis and achieve competitive performance to supervised approaches [10].

Here, we propose to use a VAE to interpolate NeuroMark [11] static functional network connectivity (sFNC) collected from controls and individuals with schizophrenia. To better visualize sFNC patterns across subjects, we make use of a 2D grid layout to organize similar or distinct individual sFNC matrices. A VAE is then used to interpolate the sFNC matrices along this grid. We display the generated sFNC matrices sampled from the VAE latent distribution and the original sFNC matrices ordered by VAE two-dimensional features in the 2D grid respectively. We observe a high correspondence between the generated and the original sFNC matrices. The generated result supports both group-specific patterns as well as a continuous alteration between patients and controls. Our study highlights the benefits of leveraging unsupervised generative models to capture individual variability within a group and continuous sFNC progression patterns between groups, and to interpolate over the pattern alterations across subjects.

II. METHODS

A. Dataset

We utilize a subset of the functional biomedical informatics network (fBIRN) dataset [12], which contains 144 subjects with schizophrenia and 145 controls. The fMRI data was processed with the NeuroMark pipeline [11]. Subject-specific functional components and corresponding time courses were then estimated using an adaptive-ICA method. The sFNC was subsequently computed as the correlation between the time courses of 53 intrinsic connectivity networks (ICNs). These 53 ICNs are from 7 functional domains including the subcortical (SC), auditory (AU), sensorimotor (SM), visual (VI), cognitive control (CC), default mode (DM) and cerebellar (CB) domains (Fig. 1) [11].

B. Model Architecture

The fBIRN dataset $X = \{x_i\}_{i=1}^{N}$ consists of N samples x_i, where N is equal to 289 subjects and x_i is a vector with size 1378×1, the flattened upper triangle of a 53×53

*This work was supported by NIH grants R01MH118695 and R01EB006841 and NSF grant #2112455.

[1]The Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

[2]Tri-institutional Center for Translational Research in Neuroimaging and Data Science: Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA.

[3]The Department of Computer Science, Georgia State University, Atlanta, GA, USA.

Corresponding author: Xinhui Li, xinhuili@gatech.edu

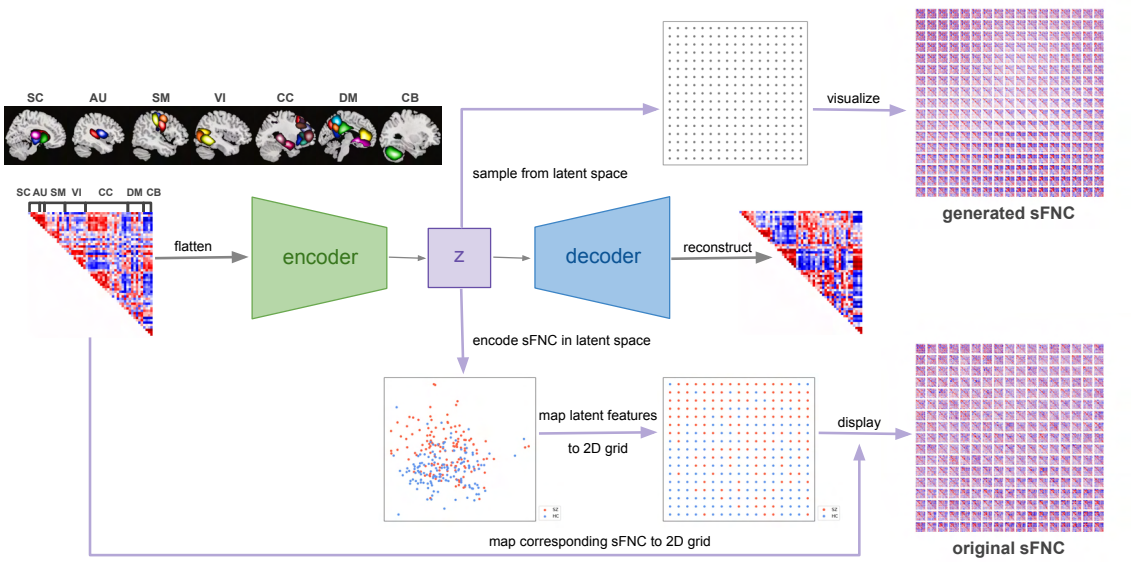

Fig. 1. **Schematic of interpolation framework.** The upper triangle of the sFNC matrix is flattened and used as the input to the VAE. During the training stage (gray arrow), the VAE is trained to learn a latent representation z by minimizing the reconstruction loss between the inputs and the reconstructed outputs, while the KL divergence regularizes the difference between the prior distribution $p(z)$ and the approximate posterior distribution $q_\phi(z|x)$. During the inference stage (purple arrow), sFNC matrices are generated by uniformly sampling the latent representation z. Also the original sFNC matrices are fed to the trained encoder, the two-dimensional features in the latent space are mapped to a 2D grid using the Jonker-Volgenant algorithm, and then the corresponding original sFNC matrices are mapped and displayed in the 2D grid.

sFNC matrix. Assume that the latent variable z denoting the underlying factors in the dataset is sampled from a prior distribution $p(z)$ and the data x is sampled from the conditional likelihood distribution $p_\theta(x|z)$ parameterized by θ. According to Bayes' rule, the latent posterior distribution $p_\theta(z|x)$ can be written as

$$p_\theta(z|x) = \frac{p(z)p_\theta(x|z)}{p_\theta(x)} = \frac{p(z)p_\theta(x|z)}{\int p(z)p_\theta(x|z)dz}.$$

The computation of the posterior $p_\theta(z|x)$ is intractable because of the integral $\int p(z)p_\theta(x|z)dz$. Thus, we need to assume a simpler proposal distribution to perform variational inference and approximate the true distribution. The KL divergence $D_{KL}(q_\phi(z|x) \parallel p_\theta(z|x))$ is used to quantify the distance between the true distribution $p_\theta(z|x)$ and a multivariate Gaussian approximation $q_\phi(z|x)$ with a diagonal covariance structure $\mathcal{N}(z; \mu, \sigma^2 I)$ parameterized by ϕ. The evidence lower bound (ELBO) can be derived as

$$\begin{aligned}\mathcal{L}(\theta,\phi) &= -D_{KL}(q_\phi(z|x) \parallel p_\theta(z|x)) + \log p_\theta(x) \\ &= -D_{KL}(q_\phi(z|x) \parallel p(z)) + \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)],\end{aligned}$$

where $\mathcal{L}(\theta,\phi) \leq \log p_\theta(x)$. An encoder and a decoder parameterized by variational parameters ϕ and generative parameters θ are trained to maximize the ELBO $\mathcal{L}(\theta,\phi)$ and thus maximize the log-likelihood of generating real data samples $\log p_\theta(x)$. To train the model using backpropagation, we use the reparameterization trick [9]: $z = \mu + \sigma \odot \epsilon$ where $\epsilon \sim \mathcal{N}(0, I)$. A detailed derivation of the variational inference and learning algorithm can be found in [9].

We implement a multilayer perceptron (MLP) as the encoder ($q_\phi(z|x)$) and decoder ($p_\theta(x|z)$) in the VAE. The MLP consists of five dense layers with 256, 128, 64, 32, and 16 output units in the MLP encoder and a symmetric MLP decoder. Each layer uses a hyperbolic tangent ($tanh$) activation function and L_1 regularization. The model is trained for 300 epochs using the Adam optimizer [13] with a learning rate of 0.001 and a batch size of 16. By optimizing the ELBO $\mathcal{L}(\theta,\phi)$, the VAE is able to learn the underlying factors in the dataset and generate new data by interpolating between points using the learned factors.

The model is implemented in Python using the Keras Application Programming Interface [14], which runs on top of the TensorFlow open source machine learning framework [15]. The Google Colaboratory platform with an NVIDIA Tesla K80 GPU was used as the development environment.

C. Interpolation Framework

An overview of the interpolation framework is presented in Fig. 1. The VAE is trained to learn a latent representation z by minimizing the reconstruction loss between the inputs and the reconstructed outputs, while the KL divergence regularizes the difference between the prior distribution $p(z)$ and the approximate posterior distribution $q_\phi(z|x)$. After the training stage, we utilize the decoder to generate sFNC matrices by uniformly sampling points in the latent space

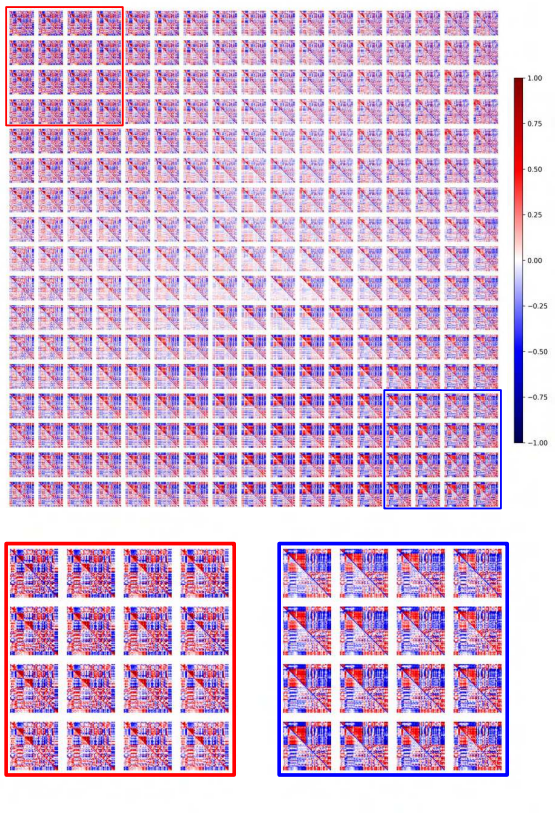

Fig. 2. **Generated sFNC matrices.** Top: The generated sFNC matrices. Bottom: Two 4×4 zoom-in grids showing patient (left) and control (right) patterns, respectively. The VAE interpolation result can capture individual variability within a group and continuous patterns between groups.

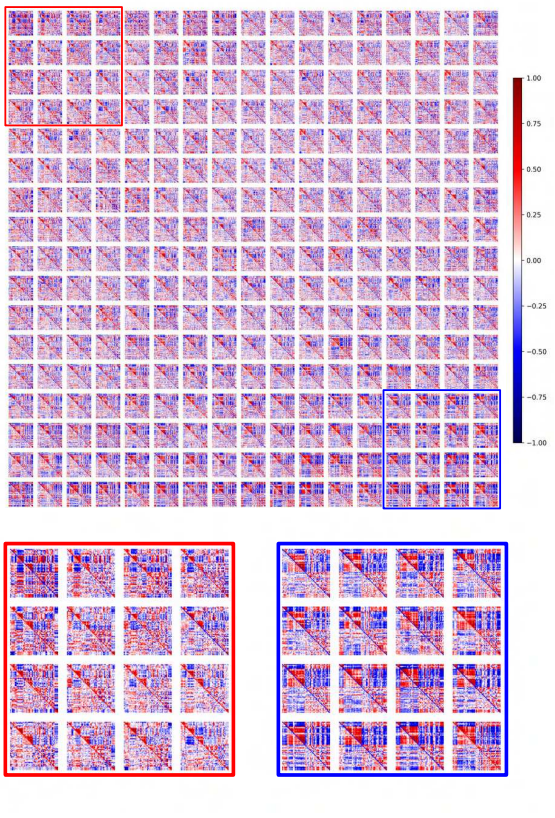

Fig. 3. **Original sFNC matrices.** Top: The original sFNC matrices. Bottom: Two 4×4 zoom-in grids showing patient (left) and control (right) patterns, respectively. Individual differences and group-specific patterns can be visualized according to the VAE latent features.

and visualize the generated sFNC matrices in a 2D grid with 17×17 nodes. We subsequently feed the original training data to the trained encoder and extract the representative two-dimensional features in the latent space. The Jonker-Volgenant algorithm [16] is then used to map these latent features into nodes in a 2D grid by minimizing the pairwise Euclidean distance between the latent samples and the 2D grid nodes. Code for the Jonker-Volgenant algorithm can be found at `https://github.com/gatagat/lap`. This allows us to visualize the generated sFNC matrices and the original sFNC matrices in the 2D grid, respectively.

III. RESULTS

Our primary results show a high correspondence between the generated sFNC matrices (Fig. 2) and the original sFNC matrices (Fig. 3). To better visualize the results, we retain the upper triangle of the sFNC matrix, remove the element-wise mean across all subjects in the lower triangle, and color code the diagonal based on the diagnosis (red: patients; blue: controls) for each sFNC matrix. By examining the group-specific patterns in both 2D grids, we observe focal modularity in the top left corner dominated by the patient group, and highly modular and polar patterns in the bottom right corner dominated by the control group. For example, the patient group shows low functional connectivity between the AU and SM domains and between the SM and VI domains. In contrast, the control group shows strong function connection among the AU, SM and VI domains. The 2D grid layout also facilitates understanding the sFNC pattern alterations between schizophrenia patients and controls. Specifically, the changes can be visualized in the connectivity between the SC and CB domains, between the SC and AU domains, as well as between the SC and SM domains. Thus, the generated results can capture representative patterns within a group and continuous pattern alterations between groups. The original sFNC matrices in the 2D grid can help to examine individual variability in the dataset. For example, the sFNC in the Fig. 3 bottom right corner shows a patient with a typical control pattern, suggesting that the patient may have mild symptoms.

Pearson correlations further support the correspondence between the generated and the original results (Fig. 4). The highest correlation is 0.91. The median correlation across all subjects is 0.73, with a median of 0.72 in the patient group and a median of 0.74 in the control group. Matrices in the upper 7 rows show relatively low correlations with a median of 0.67 and 71.4% of these matrices represent patients, indicating more variability in the patient group.

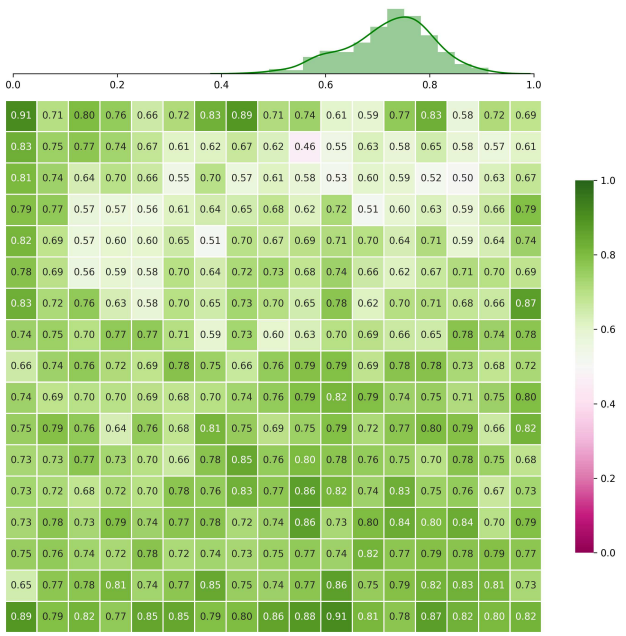

Fig. 4. **Pearson correlation between the generated and the original sFNC matrices.** The VAE can achieve fairly good reconstructions with a median correlation of 0.73 across all subjects ($p < 0.05$ for all pairs).

IV. DISCUSSION

This study demonstrates the ability of the unsupervised generative model to capture representative group-specific sFNC patterns and interpolate the continuous pattern alterations between schizophrenia patients and controls (Fig. 2). In addition, our work highlights the benefits of a 2D grid layout to visualize individual sFNC differences. As presented in Fig. 3, a 2D grid can be used to display the original individual sFNC matrices according to the latent samples and visualize both individual and group-specific patterns in the dataset. We observe relatively homogeneous sFNC patterns in the control group and highly heterogeneous patterns in the patient group. This finding supports the fact that schizophrenia is a highly heterogeneous disorder and that there may exist subgroups among patients, which could potentially be identified with our framework.

Moreover, the correlation result (Fig. 4) between the generated and the original sFNC shows that the VAE can learn the representative latent distributions from the data and achieve fairly good reconstructions, especially in the control group. Since schizophrenia is known to be highly heterogeneous, the vanilla VAE cannot fully capture the heterogeneity in the patient group, resulting in slightly lower correlations in patients compared with controls.

It is worth noting that we made an assumption that the sFNC changes between schizophrenia patients and controls can be interpolated. However, it is still unknown whether there is a categorical or a continuous boundary between groups. Our study suggests that there exist both group-specific sFNC patterns and intermediate continuous patterns which can be interpolated over two groups.

Future work will consider further characterizing the heterogeneity of disease syndromes using hierarchical clustering or mixture model priors in VAEs, identifying clinical transition states and disease subtypes, and incorporating large datasets with autism spectrum disorder and depression to capture disease progression across multiple mental disorders.

V. CONCLUSION

We utilize an unsupervised generative model to interpolate sFNC matrices across subjects while also capturing individual variability within a group and sFNC progression patterns between groups. We demonstrate a high correspondence between the generated and the original sFNC matrices. By displaying individual sFNC in the 2D grid, we can observe both individual and group-specific patterns. Our approach is a promising step toward individualized disorder progression prediction and provides insights into the characterization of individual heterogeneity within and between groups.

ACKNOWLEDGMENT

This work was supported by NIH grants R01MH118695 and R01EB006841 and NSF grant #2112455. Additionally, Xinhui Li and Eloy Geenjaar were supported by the Georgia Tech/Emory NIH/NIBIB Training Program in Computational Neural-engineering (T32EB025816).

REFERENCES

[1] B. Moreno-Küstner, C. Martín, and L. Pastor, "Prevalence of psychotic disorders and its association with methodological issues. a systematic review and meta-analyses," *PloS one*, vol. 13, no. 4, 2018.
[2] S. Saha *et al.*, "A systematic review of the prevalence of schizophrenia," *PLOS Medicine*, vol. 2, no. 5, p. e141, 2005.
[3] C. Su *et al.*, "Deep learning in mental health outcome research: a scoping review," *Translational Psychiatry*, vol. 10, no. 1, 2020.
[4] Q. Yu *et al.*, "Brain connectivity networks in schizophrenia underlying resting state functional magnetic resonance imaging," *Current topics in medicinal chemistry*, vol. 12, no. 21, p. 2415–2425, 2012.
[5] E. Damaraju *et al.*, "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *NeuroImage: Clinical*, vol. 5, pp. 298–308, 2014.
[6] B. Rashid *et al.*, "Classification of schizophrenia and bipolar patients using static and dynamic resting-state fmri brain connectivity," *NeuroImage*, vol. 134, p. 645–657, 2016.
[7] L.-L. Zeng *et al.*, "Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity mri," *EBioMedicine*, vol. 30, pp. 74–85, 2018.
[8] M. T. Tsuang, M. J. Lyons, and S. V. Faraone, "Heterogeneity of schizophrenia: Conceptual models and analytic strategies," *British Journal of Psychiatry*, vol. 156, no. 1, p. 17–26, 1990.
[9] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2014.
[10] E. Geenjaar *et al.*, "Fusing multimodal neuroimaging data with a variational autoencoder," 2021.
[11] Y. Du, Z. Fu, *et al.*, "Neuromark: An automated and adaptive ica based pipeline to identify reproducible fmri markers of brain disorders," *NeuroImage: Clinical*, vol. 28, p. 102375, 2020.
[12] D. B. Keator *et al.*, "The function biomedical informatics research network data repository," *NeuroImage*, vol. 124, pp. 1074–1079, 2016.
[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.
[14] F. Chollet *et al.*, "Keras," https://github.com/fchollet/keras, 2015.
[15] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, pp. 265–283.
[16] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, pp. 325–340, 2005.