A VQ-VAE framework for modeling physiological information in fMRI

Shiyu Wang[1] **Yamin Li**[1] **Ziyuan Xu**[1] **Haatef Pourmotabbed**[1] **Chang Li**[1]
Roza G. Bayrak[1] **Catie Chang**[1]
[1]Vanderbilt University
shiyu.wang.1@vanderbilt.edu

Abstract

The coupling between functional magnetic resonance imaging (fMRI) and autonomic physiological processes is not merely "noise", but captures important brain-body interactions. This relationship is likely bidirectional: autonomic fluctuations can influence brain activity and hemodynamics, and neuronal activity can modulate systemic physiological processes. This complexity makes it difficult for traditional linear approaches to fully characterize their relationship. Here, we introduce a novel application of vector quantized variational autoencoder (VQ-VAE) to characterize whole-brain patterns associated with respiration using the discretized fMRI latent space. Further, we demonstrate that this framework can be leveraged to assess the quality of fMRI-based reconstructions of low-frequency respiratory fluctuations when physiological recordings are missing or corrupted. The success of our model indicates that we can extract the non-linear relationship between the fMRI brain signals and bodily autonomic processes, revealing their intrinsic connections in the cross-modal latent space.

1 Introduction

Physiological processes in the body have been known to influence functional magnetic resonance imaging (fMRI) signals. However, while certain physiological processes induce fMRI artifacts (e.g., head motion), slow variations in autonomic processes, including natural changes in breathing rate and depth, are associated with low-frequency fluctuations in the blood-oxygen-level-dependent (BOLD) signal [1]. The latter effects are thought to arise from factors such as fluctuations in arterial carbon dioxide levels [2], sympathetic modulation of vascular tone [3], and neural control of autonomic activity [4–6]. These converging effects contribute to the complexity of the relationship between fMRI and physiological processes, alongside inter-subject variability as well as intra-subject, state-dependent changes [7].

Most existing approaches use linear models (via physiological impulse response functions [8–10]) to capture the brain-body relationship embedded in the interaction between fMRI signals and autonomic processes. However, these approaches might fail to capture any nonlinear or state-dependent whole-brain hemodynamics correlated with physiological processes. Here, we introduce a data-driven approach using vector quantized variational autoencoder (VQ-VAE) to capture a wider nonlinear repertoire of structured whole-brain patterns related to respiration.

The proposed framework could be extended to the practical challenge of handling missing or corrupted physiological recordings in fMRI datasets. This prevalent scenario has motivated the development of machine learning techniques for directly reconstructing physiological fluctuations from fMRI data alone [11–16]. Yet, as the performance of such methods can vary across individuals, it is

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Foundation Models for the Brain and Body.

critical to have a measure that quantifies the predictive uncertainty, and such a confidence score is not currently provided by existing methods. Here, we present a novel approach that leverages our proposed VQ-VAE framework to estimate a confidence score for fMRI-derived RV reconstructions. The goals of this work were twofold: (1) to develop an interpretable framework to explore brain patterns associated with the body, focusing on respiration volume (RV), and (2) to provide a quality assessment metric for fMRI-derived RV reconstructions.

2 Datasets and preprocessing

We included 1380 resting-state fMRI scans with concurrent respiratory belt recordings from the minimally preprocessed Human Connectome Project - Young Adult dataset (HCP-YA) (TR = 0.72 s, voxel size = 2mm isotropic, duration 14.4 min with 1200 volumes per scan) [17, 18]. Of these, we used 900 scans for training, 180 scans for validation, and 300 scans for testing, without including any test-set subject in the training or validation process. In addition, we drew upon 365 resting-state fMRI scans from the HCP-Aging dataset (HCP-A [19]) to test the generalizability of the model to an unseen dataset with a different age distribution. The HCP-A dataset (TR = 0.8 s, voxel size = 2mm isotropic, 6.5 min duration with 487 volumes per scan) underwent the same minimal preprocessing pipeline as the HCP-YA dataset [18, 19] . We extracted 512 ROIs from the DiFuMo atlas [20], bandpass-filtered the time courses to 0.01-0.15 Hz, z-normalized each ROI's time course, and resampled both HCP-YA and HCP-A scans to 1.44 s intervals. Low-frequency changes in breathing (respiratory volume; RV) during the scan were calculated by taking the standard deviation of the respiratory belt waveform in a 6-s window centered at each fMRI time point. This RV signal was then similarly bandpass-filtered to 0.01-0.15 Hz, z-normalized, and resampled to align with the resampled fMRI data points.

3 Methods

3.1 Reconstruction confidence score

We formulated the (ground-truth) Respiratory Volume Reconstruction Confidence Score (RVConf) as the Pearson correlation between the ground-truth RV signal and the averaged RV reconstructions from three models: a bidirectional long short-term memory model (Bi-LSTM)[15], a transformer sequence-to-sequence model (Tf-seq2seq)[16] and a transformer sequence-to-one model (Tf-seq2one)[16]. This correlation value was then linearly scaled to a [0, 1] range.

3.2 Model architecture

The Vector Quantized Variational Autoencoder (VQ-VAE) has been widely used in the computer vision field [21], and previous works have applied the VQ-VAE structure to fMRI data to decode brain state [22] and reconstruct image stimuli presented during the fMRI scan session [23]. Here, rather than solely using the quantized latent space, we introduce a codebook usage histogram derived from VQ-VAE to capture the latent patterns in fMRI. We hypothesized that the VQ-VAE codebook index during the vector quantization process could serve as a low-dimensional representation of fMRI data. Similar to linear methods like fMRI co-activations patterns (CAPS)[24], here we obtained the discretized and compressed representation of each fMRI time point using their codebook assignment.

Each fMRI frame was processed by a transformer-based encoder (embedding dimension = 256, depth = 8 layers, 4 attention heads) and quantized using a multi-stage VQ-VAE (Figure 1). The quantized representations were decoded back into fMRI space via a transformer decoder (embedding dimension = 128, depth = 4, nhead = 4). To summarize the latent distribution, we computed a codebook usage histogram for each successive 60 s fMRI window (41 frames after downsampling to a temporal resolution of 1.44 s). Frame contributions were weighted by a Gaussian kernel (bandwidth = 6 frames) to emphasize central time points. The multi-stage hierarchical quantizer discretizes latent representations progressively: early stages capture global patterns, whereas later stages encode fine-grained details. We applied a 7-stage vector quantization scheme but focused our codebook usage analysis on the first three stages. This choice is based on the assumption that respiratory effects on fMRI are spatially widespread and primarily global, therefore better represented at the early quantization levels than at later, finer-grained stages. Nevertheless, future work could examine this assumption and assess whether the interaction between respiration and BOLD signal is also reflected in the later quantization stages. The codebook size was 1024 with an embedding dimension of 128.

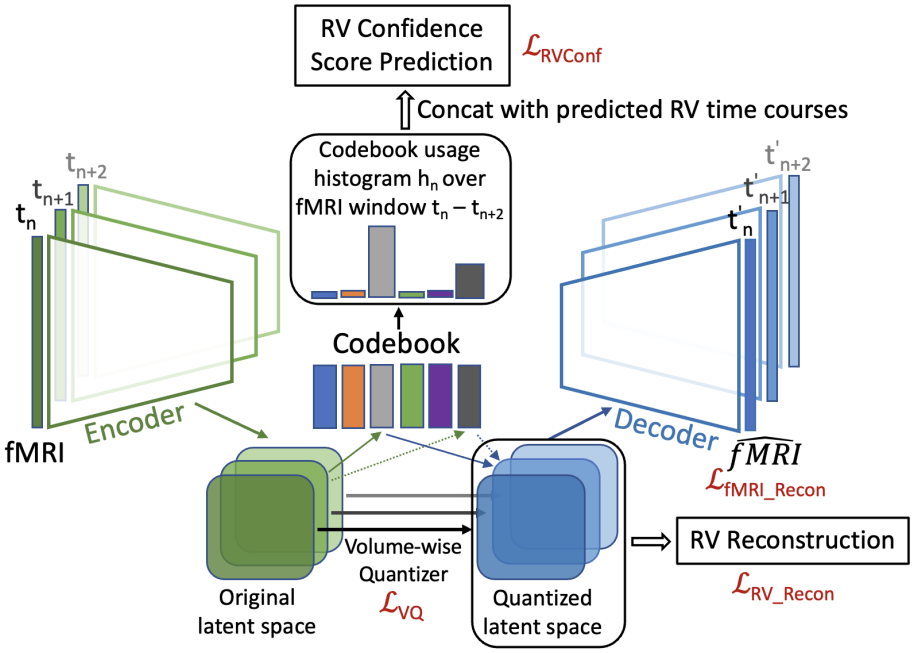

Figure 1: **Overall framework.** Vector quantized variational autoencoder (VQ-VAE) framework for modeling fMRI-physiology relationship. Each fMRI frame is encoded via a transformer-based encoder, and discretized using a learned vector quantization codebook. The quantized latent space is then passed through a transformer-based decoder to reconstruct the original fMRI frame. For the set of fMRI frames within a given 1-min window of data, this quantized latent space also reconstructs the corresponding (1-min) segment of the RV time course (for illustation, only 3 frames are shown here). A codebook usage histogram is aggregated and is used to predict the RVConf score of this window, together with 3 vectorized RV reconstructions. The loss function consists of 1) RV prediction confidence loss ($\mathcal{L}_{\text{RVConf}}$), a binary cross entropy loss with target and predicted confidence scores both rescaled within the [0, 1] range, 2) fMRI reconstruction loss ($\mathcal{L}_{\text{fMRI\_Recon}}$), defined as the MSE loss, 3) vector quantization loss ($\mathcal{L}_{\text{VQ}}$), and 4) RV reconstruction loss ($\mathcal{L}_{\text{RV\_Recon}}$).

The discretized codebook usage histogram was concatenated with the three vectorized reconstructed respiration signals and passed through a fully connected multilayer perceptron (MLP) to predict the RV reconstruction confidence score (RVConf). The quantized latent representation of N=41 fMRI frames was supervised to reconstruct the RV time course over each 1-minute window using an LSTM head, encouraging the quantized fMRI latent space to retain sufficient respiration-related information. The overall loss function consists of 4 components: 1) an fMRI reconstruction mean squared error (MSE) loss $\mathcal{L}_{\text{fMRI\_Recon}} = MSE(f\hat{M}RI, fMRI)$, 2) a vector quantization (VQ) loss $\mathcal{L}_{\text{VQ}}$ to regularize the discretization process, 3) a binary cross-entropy (BCE) loss for RVConf reconstruction $\mathcal{L}_{\text{RVConf}} = BCE(RV\hat{C}onf, RVConf)$ with both target and predicted RVConf scores rescaled to continuous [0, 1] range, and 4) an RV time course reconstruction loss $\mathcal{L}_{\text{RV\_Recon}} = 1 - \rho(\hat{RV}, RV)$ for each 1-minute window, where $\rho(\cdot, \cdot)$ denotes the Pearson Correlation Coefficient between the target and reconstructed RV signals.

3.3 Baseline comparisons and ablation tests

To assess the contribution of individual components to the RVConf prediction, we conducted the following ablation tests: 1) predict RVConf with the codebook usage histogram and the 3 reconstructed RV signals from Bi-LSTM, Tf-seq2seq and TF-seq2one, but remove the RV reconstruction task from the quantized latent space; and 2) predict RVConf solely with the 3 reconstructed RV signals, excluding fMRI codebook histogram. As a static reference baseline, we correlated RVConf with the average pairwise correlation among reconstructed RV signals from the three established methods, assuming that reconstruction convergence (i.e., higher inter-method similarity) reflects higher confidence.

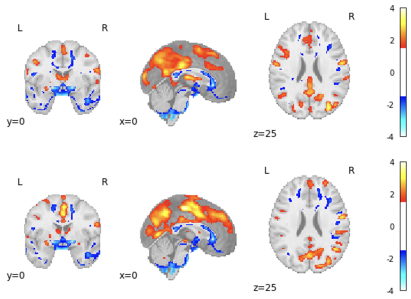

Figure 2: Top two codebook embeddings most strongly associated with respiration reconstruction accuracy in the HCP-A dataset, projected into anatomical space.

3.4 Experimental details

The model structures and hyperparameters for the Bi-LSTM, Tf-seq2seq and TF-seq2one models were directly adopted from previous papers [15, 16]. We chose a batch size of 96 and a learning rate of 1e-5 with 4 warmup epochs for all experiments. We used Adam optimizer with default parameters $\beta 1 = 0.9$, $\beta 2 = 0.95$. The optimal training epoch was selected as the one with the lowest validation loss. Pearson correlation coefficient and mean absolute error between the ground truth RVConf and predicted RVConf were used as model evaluation metrics. Experiments 1 and 2 were performed on an NVIDIA A6000 GPU, and experiment 3 was performed on an NVIDIA GeForce RTX 2080 Ti.

4 Results and discussion

Table 1: RV reconstruction confidence score prediction performance from the full model, ablation studies and linear baseline on both HCP-YA and HCP-A dataset. CB, codebook; Pred. RV, predicted RV time course; Avg. Pred. RV Corr, average pairwise correlation among the 3 predicted RV signals; Corr, Pearson correlation coefficient; MAE, mean absolute error (mean $\pm$ std). **Bold**: the best.

Model	RV Recon Loss?	VQ-VAE CB Usage?	Nonlinear?	HCP-YA Corr ↑	HCP-YA MAE ↓	HCP-A Corr ↑	HCP-A MAE ↓
VQ-VAE (Full)	✓	✓	✓	**0.4920**	**0.0954±0.0052**	**0.5369**	**0.0943±0.0077**
VQ-VAE (No RV Recon)	✗	✓	✓	0.4780	0.0964±0.0053	0.5258	0.0952±0.0077
Pred. RV	✗	✗	✓	0.4684	0.0910±0.0058	0.4972	0.1004±0.0090
Avg. Pred. RV Corr	✗	✗	✗	0.3421	-	0.4245	-

As shown in Table 1, the best RV reconstruction confidence score prediction was achieved with the full model when tested on either the HCP-YA dataset or the unseen HCP-A dataset, and for both Pearson correlation coefficient (Corr) and mean absolute error metric (MAE ± std). Figure 2 depicts fMRI patterns associated with the two codebook embeddings that were most strongly associated with respiration reconstruction accuracy in the HCP-A datasets. While exhibiting some qualitative similarities, the upper panel has stronger weights in precuneus, while the lower panel has stronger weights in posterior cingulate cortex and medial frontal cortex. Negative weights in both maps are consistent with prior work indicating the contribution of sympathetic nervous system activity to fMRI signals in white matter [25]. The precuneus and posterior cingulate cortex have also been implicated in autonomic processing, supporting their involvement in brain-body coupling [26, 27].

The proposed VQ-VAE architecture provides an interpretable framework for modeling the relationship between fMRI BOLD signal and respiration by learning latent codebook representations aligned with respiration dynamics. Furthermore, in cases where physiological recordings are missing or corrupted, the framework produces an RV reconstruction confidence score that helps users to assess the reliability of low-frequency respiration signals predicted by previously established methods. Future work could extend the current framework to model the relationship between fMRI and cardiac activity and to incorporate a larger number of simultaneous fMRI–physiology recordings into the training dataset.

References

[1] Kevin Murphy, Rasmus M Birn, and Peter A Bandettini. Resting-state fmri confounds and cleanup. *Neuroimage*, 80:349–359, 2013.

[2] Richard G Wise, Kojiro Ide, Marc J Poulin, and Irene Tracey. Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. *Neuroimage*, 21(4):1652–1664, April 2004.

[3] Pinar Senay Özbay, Catie Chang, Dante Picchioni, Hendrik Mandelkow, Miranda Grace Chappel-Farley, Peter van Gelderen, Jacco Adrianus de Zwart, and Jeff Duyn. Sympathetic activity contributes to the fMRI signal. *Commun. Biol.*, 2(1):421, November 2019.

[4] Christina Zelano, Heidi Jiang, Guangyu Zhou, Nikita Arora, Stephan Schuele, Joshua Rosenow, and Jay A Gottfried. Nasal respiration entrains human limbic oscillations and modulates cognitive function. *J. Neurosci.*, 36(49):12448–12467, December 2016.

[5] Wenyu Tu and Nanyin Zhang. Neural underpinning of a respiration-associated resting-state fMRI network. *Elife*, 11, October 2022.

[6] Yameng Gu, Feng Han, Lucas E Sainburg, Margeaux M Schade, Orfeu M Buxton, Jeff H Duyn, and Xiao Liu. An orderly sequence of autonomic and neural events at transient arousal changes. *Neuroimage*, 264: 119720, December 2022.

[7] Feliberto de la Cruz, Andy Schumann, Stefanie Köhler, Karl-Jürgen Bär, and Gerd Wagner. Impact of the heart rate on the shape of the cardiac response function. *Neuroimage*, 162:214–225, November 2017.

[8] Catie Chang, John P Cunningham, and Gary H Glover. Influence of heart rate on the BOLD signal: The cardiac response function. *Neuroimage*, 44(3):857–869, February 2009.

[9] Rasmus M Birn, Monica A Smith, Tyler B Jones, and Peter A Bandettini. The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes in respiration. *Neuroimage*, 40(2): 644–654, April 2008.

[10] Michalis Kassinopoulos and Georgios D Mitsis. Identification of physiological response functions to correct for fluctuations in resting-state fMRI related to heart rate and respiration. *Neuroimage*, 202:116150, November 2019.

[11] Serdar Aslan, Lia Hocke, Nicolette Schwarz, and Blaise Frederick. Extraction of the cardiac waveform from simultaneous multislice fMRI data using slice sorted averaging and a deep learning reconstruction filter. *Neuroimage*, 198:303–316, September 2019.

[12] Tom Ash, John Suckling, Martin Walter, Cinly Ooi, Claus Tempelmann, Adrian Carpenter, and Guy Williams. Detection of physiological noise in resting state fMRI using machine learning. *Hum. Brain Mapp.*, 34(4):985–998, April 2013.

[13] Jorge A Salas, Roza G Bayrak, Yuankai Huo, and Catie Chang. Reconstruction of respiratory variation signals from fMRI data. *Neuroimage*, 225:117459, 2021.

[14] Roza G. Bayrak, Jorge A. Salas, Yuankai Huo, and Catie Chang. A deep pattern recognition approach for inferring respiratory volume fluctuations from fmri data. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 428–436, Cham, 2020. Springer International Publishing. ISBN 978-3-030-59728-3.

[15] Roza G. Bayrak, Colin B. Hansen, Jorge A. Salas, Nafis Ahmed, Ilwoo Lyu, Yuankai Huo, and Catie Chang. From brain to body: Learning low-frequency respiration and cardiac signals from fmri dynamics. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 553–563, Cham, 2021. Springer International Publishing. ISBN 978-3-030-87234-2.

[16] Shiyu Wang, Ziyuan Xu, Laurent M. Lochard, Yamin Li, Jiawen Fan, Jingyuan E. Chen, Yuankai Huo, Mara Mather, Roza G. Bayrak, and Catie Chang. Reconstructing physiological signals from fMRI across the adult lifespan. In Olivier Colliot and Jhimli Mitra, editors, *Medical Imaging 2025: Image Processing*, volume 13406, page 134060M. International Society for Optics and Photonics, SPIE, 2025. doi: 10.1117/12.3046980. URL `https://doi.org/10.1117/12.3046980`.

[17] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy E J Behrens, Essa Yacoub, and Kamil Ugurbil. The WU-minn human connectome project: An overview. *Neuroimage*, 80:62–79, October 2013.

[18] Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, David C Van Essen, and Mark Jenkinson. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, October 2013.

[19] Susan Y Bookheimer, David H Salat, Melissa Terpstra, Beau M Ances, Deanna M Barch, Randy L Buckner, Gregory C Burgess, Sandra W Curtiss, Mirella Diaz-Santos, Jennifer Stine Elam, Bruce Fischl, Douglas N Greve, Hannah A Hagy, Michael P Harms, Olivia M Hatch, Trey Hedden, Cynthia Hodge, Kevin C Japardi, Taylor P Kuhn, Timothy K Ly, Stephen M Smith, Leah H Somerville, Kâmil Uğurbil, Andre van der Kouwe, David Van Essen, Roger P Woods, and Essa Yacoub. The lifespan human connectome project in aging: An overview. *Neuroimage*, 185:335–348, January 2019.

[20] Kamalaker Dadi, Gaël Varoquaux, Antonia Machlouzarides-Shalit, Krzysztof J Gorgolewski, Demian Wassermann, Bertrand Thirion, and Arthur Mensch. Fine-grain atlases of functional modes for fMRI analysis. *Neuroimage*, 221(117126):117126, November 2020.

[21] Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. *arXiv [cs.LG]*, November 2017.

[22] Yanwu Yang and Thomas Wolfers. Hierarchical characterization of brain dynamics via state space-based vector quantization. *arXiv [eess.IV]*, June 2025.

[23] Sikun Lin, Thomas C Sprague, and Ambuj K Singh. Mind reader: Reconstructing complex images from brain activities. *Neural Inf Process Syst*, abs/2210.01769, September 2022.

[24] Xiao Liu, Nanyin Zhang, Catie Chang, and Jeff H Duyn. Co-activation patterns in resting-state fMRI signals. *Neuroimage*, 180(Pt B):485–494, October 2018.

[25] Pinar S Özbay, Catie Chang, Dante Picchioni, Hendrik Mandelkow, Thomas M Moehlman, Miranda G Chappel-Farley, Peter van Gelderen, Jacco A de Zwart, and Jeff H Duyn. Contribution of systemic vascular effects to fMRI activity in white matter. *Neuroimage*, 176:541–549, August 2018.

[26] Jin Fan, Pengfei Xu, Nicholas T Van Dam, Tehila Eilam-Stock, Xiaosi Gu, Yue-Jia Luo, and Patrick R Hof. Spontaneous brain activity relates to autonomic arousal. *J. Neurosci.*, 32(33):11176–11186, August 2012.

[27] G Valenza, R Sclocco, A Duggento, L Passamonti, V Napadow, R Barbieri, and N Toschi. The central autonomic network at rest: Uncovering functional MRI correlates of time-varying autonomic outflow. *Neuroimage*, 197:383–390, August 2019.