DeepCor: Denoising fMRI data using Contrastive Variational Autoencoders

Aidas Aglinskas*
Department of Psychology and Neuroscience
Boston College
Boston, MA 02467
aidas.aglinskas@bc.edu

Yu Zhu
Department of Psychology and Neuroscience
Boston College
Boston, MA 02467
polly.yuzhu@gmail.com

Stefano Anzellotti
Department of Psychology and Neuroscience
Boston College
Boston, MA 02467
stefano.anzellotti@bc.edu

Abstract

Functional magnetic resonance imaging (fMRI) is widely used in neuroscience research to measure neural activity non-invasively with high spatial resolution. However, fMRI data is affected by noise that hinders researchers from making novel discoveries about the brain. In consideration of the complexity of noise sources and their interactions, we introduce and evaluate a denoising method which utilizes adversarial or deep generative models to disentangle and remove noise (DeepCor). The method is applicable to data from single participants, without requiring datasets with large numbers of individuals. DeepCor outperforms other denoising approaches on a variety of real datasets (StudyForrest, Adolescent Brain Cognitive Development, and THINGS-fMRI), more effectively enhancing BOLD signal responses to face selectivity in face selective regions, and place selectivity in place selective regions.

1 Introduction

Functional magnetic resonance imaging (fMRI) offers a unique window into the human brain, enabling researchers to probe the neural basis of cognition and behavior, and to identify biomarkers that may guide diagnosis and treatment of psychiatric and neurological disorders. However, the utility of fMRI is constrained by substantial noise from physiological processes, head motion, and scanner artifacts, which can obscure true neural signals—particularly in analyses at the single-subject level, where data are limited and individual differences must be preserved. Widely used denoising approaches such as CompCor [Behzadi et al., 2007, Muschelli et al., 2014] address this challenge by extracting principal components from noise-dominated regions-of-no-interest (e.g., white matter, cerebrospinal fluid) and linearly regressing them from gray matter time series. While effective to a degree, these methods assume a linear relationship between noise in regions-of-interest and regions-of-no-interest, an assumption often violated by complex nonlinear interactions between neural signals and noise, potentially limiting its denoising efficacy.

Building on recent advances in representation learning, we introduce DeepCor, a deep learning approach utilizing similar principles as CompCor, but capable of learning nonlinear relationships

*Corresponding author.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Data on the Brain & Mind.

between signal and noise. DeepCor comes in two variants: DeepCor-Adv uses an adversarial network setup Ganin and Lempitsky [2015] and DeepCor-Gen, which is built on generative AI principles, specifically Contrastive Variational Autoencoders Abid and Zou [2019]. Both variants offer a flexible, participant-level solution applicable to both resting-state and task-based fMRI.

Because establishing ground truth during resting-state fMRI is unclear - we evaluated denoising effectiveness using task-based fMRI data, specifically during a face and place perception task, which is a robust and widely studied paradigm. This task reliably produces distinct BOLD responses: face stimuli selectively activate the fusiform face area (FFA; [Kanwisher et al., 1997]), while place stimuli selectively activate the parahippocampal place area (PPA; [Epstein and Kanwisher, 1998]). These effects represent some of the most robust and reproducible contrasts in fMRI research. Prior test-retest studies highlight strong reliability of both FFA and PPA responses Zanto et al. [2014], making them ideal for assessing denoising performance. Despite the reliability, physiological and scanner-related noise can still degrade the clarity of these task-evoked contrasts. We hypothesized that improved denoising would enhance the strength and detectability of the contrasts between conditions in these regions.

Related works. Classical fMRI denoising methods include ICA-based approaches such as FIX and ICA-AROMA, which classify independent components as signal or noise for removal [Salimi-Khorshidi et al., 2014, Pruim et al., 2015], as well as GLMdenoise, which estimates noise regressors from task-unrelated voxels and improves cross-validated model fits [Kay et al., 2013]. Acquisition-based innovations such as multi-echo fMRI with ME-ICA [Kundu et al., 2017, Spreng et al., 2019, DuPre et al., 2021] and thermal-noise suppression via NORDIC [Moeller et al., 2021, Dowdle et al., 2023, Vizioli et al., 2021] further improve signal quality at the source. However, these methods require either manual labeling of components or specialized acquisitions not available in most existing datasets, limiting their accessibility. In contrast, CompCor operates directly on standard single-echo data by extracting principal components from white matter and the cerebrospinal fluid (CSF), avoiding manual labeling and the need for new acquisitions. As a result, CompCor is the most widely adopted denoising method in large-scale fMRI studies (e.g., fMRIPrep incorporates it by default [Esteban et al., 2019]), and often provides the strongest performance among classical approaches. Recent years have seen the emergence of deep learning approaches, which aim to overcome the limitations of classical methods. For such models to be viable in practice, they should: (1) generalize to both resting-state and task-based data, (2) be trainable on data from a single participant, and (3) provide open-source implementations for reproducibility. Several previous deep learning attempts are aimed automating manual denoising methods by identifying noise components [Theodoropoulos et al., 2021, Heo et al., 2022]; others focus exclusively on task-fMRI [Yang et al., 2020b] and others do not have publicly available code implementations [Zhao et al., 2020, Theodoropoulos et al., 2021, Yang et al., 2020a]. We identified DeNN [Yang et al., 2020a] as the only other method meeting these criteria and include it in our comparisons alongside DeepCor models.

2 Methods

Datasets. We evaluated the models on three fMRI datasets: StudyForrest category localizer task (N=14; 4 runs; studyforrest.org); subset of the Adolescent Brain Cognitive Development fMRI n-back task using faces and places (ABCD; N=33; 2 runs; abcdstudy.org) and THINGS-fMRI category localizer task (N=3; 6 runs; things-initiative.org).

Regions of Interest. For each dataset, we identified functional regions for Fusiform Face Area (FFA) and Parahippocampal Place Area (PPA). To localize the FFA in individual subjects, we first performed a group level GLM and identified the peak coordinates for the contrast faces > non-faces. We then drew a sphere around the coordinates and selected the n voxels in the sphere with the highest contrast values as individual-specific FFA. An identical procedure was used to localize the PPA, using places > non-places contrast. StudyForrest data were 3x3x3mm so we used a 9-mm-radius sphere and selected the top 80 voxels. ABCD and THINGS datasets were 2x2x2mm so we used 10-mm-radius sphere and selected the top 100 voxels per subject. To to establish a conservative baseline - ROIs were selected using data denoised with CompCor.

Metrics. As our primary performance metric, we quantified face and place specificity using contrast values for faces > non-faces (FFA) and places > non-places (PPA), averaged across runs and voxels to yield one value per subject. These contrast values, widely used in the literature, reflect the strength of category-selective responses. Because contrast estimates can be influenced by noise affecting either the target category (e.g., faces in FFA, places in PPA) or the baseline categories (non-faces, non-places), we also report "category responsivity" metric, defined as the correlation between the target (face or place) onset regressor convolved with the hemodynamic response function—and the BOLD signal in the corresponding ROI (FFA or PPA) voxels. This complementary measure assesses denoising performance specifically with respect to single category responses. **Statistical tests.** To establish whether deep learning models can be used to replace industry standard denoising approaches, we compared the latest model DeepCor-Gen-v2, which we hypothesize to be the most effective, to the industry-standard approach, which is CompCor denoising. We averaged across voxels and runs, resulting in one value per subject. Correlation values are first Fisher z-transformed. We used one-tailed, paired-samples t-test, and report uncorrected p-values.

Models tested. We compared metrics calculated on data that has been preprocessed with fMRIPrep with no additional denoising ("No Denoising"), denoising with CompCor, DeNN, DeepCor-Adv and DeepCor-Gen.

DeepCor-Adv is an adversarial 1D convolutional autoencoder. It consists of three components: encoder (three conv layers followed by Relu nonlinearities) used to encode both ROI and RONI time series to a latent representation; signal decoder (thee transposed conv layers followed by a sigmoid layer) used to produce denoised ROI time series from latent representation, and a structurally identical noise decoder used to produce noise features from latent representation. The encoder is used to map both ROI and RONI voxels to a latent representation. Crucially, when decoding RONI voxels, latent features are first passed through a gradient reversal layer. This way the encoder is forced to preserve information useful to reconstruct gray matter voxels containing signal while discouraging preserving information useful to reconstruct RONI voxels containing noise. The loss function is a sum of mean-squared-error (MSE) and normalized cross-correlation (NCC) between target pairs, specifically: between ROI inputs and signal outputs and between RONI voxels and noise outputs (see Appendix for details).

DeepCor-Gen-v1 is a Contrastive Variational Autoencoder (CVAE). It consists of two encoders (noise and signal encoders) projecting the data onto two 4-dimensional latent spaces (noise and signal features). A unified decoder takes in a concatenated representation (noise+signal) to produce a reconstruction. RONI data is passed through noise encoder only, while ROI data, consisting of noise and signal is passed through both noise and signal encoders. To denoise an ROI voxel, the latent noise features are set to 0 before decoding. Loss function is composed of an MSE reconstruction loss and Kullback-Leibler (KL) divergence term. DeepCor-Gen-v2 additionally incorporates voxel coordinate information, uses adversarial decoding to remove motion information from the signal features, and incorporates additional regularization parameters (see Appendix for details).

DeNN is an LSTM autoencoder. It consists of a 1D convolutional layer, followed by an LSTM layer and a time-distributed fully connected layer that produces K candidate outputs at each time point, and a selection layer chooses the best time series from candidate ones. The model's loss function minimizes the absolute value of the correlation between outputs generated for ROI and RONI voxels Yang et al. [2020c].

Training procedure. All models were trained for 100 epochs using Adam optimizer. DeepCor models used learning rate of 0.001. DeNN used a learning rate of 0.05 as suggested by the authors Yang et al. [2020c]. For DeepCor models, we employed an ensembling procedure, training 20 models and averaging the outputs - which was shown to improve reliability of CVAE models Aglinskas et al. [2025]. Models were trained on a compute cluster using NVIDIA V100 and A100 GPUs. Runtime of DeepCor models is about 2h per subject (StudyForrest).

3 Results

Category Selectivity. As expected, even without any denoising, both FFA and PPA showed robust category-selective responses, with contrast estimates ranging from 1.79 to 6.83. Averaging across datasets and domains (face selectivity in the FFA and place selectivity in the PPA), the mean pre-denoising contrast value was 3.48 (Tables 1 and 2). Applying CompCor increased the average to 3.87, representing an 11% improvement in category selectivity, relative to no denoising. The largest gains were achieved with the DeepCor-Gen models: DeepCor-Gen-v1 improved category selectivity to 4.51 (+30%), and DeepCor-Gen-v2 to 5.49 (+58%). We compared the best performing model, DeepCor-Gen-v2, to the widely used CompCor approach using one-tailed, paired-samples t-test. DeepCor-Gen-v2 improved over CompCor, across all datasets: StudyForrest ΔM=0.81, t(13) = 5.86, $p < .001$; ABCD ΔM=1.87, t(32) = 4.38 , $p < .001$ and THINGS ΔM=2.18, t(2) = 3.17 , p = 0.0435 (see Table 1 and Fig. S01).

Category Responsivity. When using the measure of correlation with the specific onset regressor (faces in the FFA and places in the PPA), results were consistent with previous analyses. Before any denoising, averaging across datasets and contrasts, the correlation between BOLD signal in the ROIs (FFA and PPA) and the regressor (face or place regressor) was on average 0.22. CompCor increased the average to 0.23 (+4%). Biggest improvements, again, came from DeepCor-Gen models, with improvements up to 0.29 (+29% improvement) after DeepCor-Gen-v1 and to 0.33 (+49% improvement) after DeepCor-Gen-v2. DeepCor-Gen-v2 outperformed CompCor across all datasets: StudyForrest ΔM=0.09, t(13) = 7.51 , $p < .001$; ABCD ΔM=0.08, t(32) = 5.23 , $p < .001$, and THINGS ΔM=0.12, t(2) = 3.09 , p = 0.0455.

DeNN denoising decreased category selectivity to negligible amounts, according to both category-selectivity (M = -0.03) and correlation measures (M = 0.005). Across all comparisons DeNN results were not statistically different from 0, one-sample t-test, all $p > .05$. We include an expanded discussion on these findings in the appendix.

Table 1: Category Selectivity. Face and Place selectivity contrast values after applying different denoising methods. "Avg." denotes the average contrast value, averaged across face and place contrasts and across datasets. "Gain" denotes improvement in contrast estimate after applying denoising relative to no denoising expressed as a percentage (i.e., doubling of the contrast estimate would be a 100% improvement).

	Face Selectivity (FFA)			Place Selectivity (PPA)			Average	
	Forrest	ABCD	things	Forrest	ABCD	things	Avg.	Uplift
No Denoising	1.79	3.73	2.73	2.12	6.83	3.69	3.48	0%
CompCor	1.93	4.67	3.03	2.30	7.09	4.17	3.86	11%
DeNN	-0.06	-0.03	0.00	0.01	0.00	-0.12	-0.03	<0%
DeepCor-Adv	1.81	4.16	3.14	2.30	8.73	4.26	3.74	7%
DeepCor-Adv-large	1.72	4.04	3.00	2.19	8.46	4.02	3.91	12%
DeepCor-Gen-v1	1.92	3.80	3.45	2.87	8.64	6.38	4.51	30%
DeepCor-Gen-v2	**2.36**	**5.30**	**4.35**	**3.50**	**10.19**	**7.24**	**5.48**	**58%**

Table 2: Category Responsivity results. Correlations with single category regressor. Boxcar function regressor corresponding to the onset of the face- or place- stimuli block was first convolved with the hemodynamic response function. Correlations were performed voxel-wise and then averaged.

	Face Regressor (FFA)			Place Regressor (PPA)			Average	
	Forrest	ABCD	things	Forrest	ABCD	things	Avg.	Uplift
No Denoising	0.14	0.16	0.22	0.23	0.30	0.26	0.22	0%
CompCor	0.16	0.18	0.24	0.25	0.28	0.27	0.23	4%
DeNN	0.00	0.03	0.00	0.01	0.00	-0.01	0.01	<0%
DeepCor-Adv	0.15	0.18	0.25	0.25	0.38	0.29	0.24	11%
DeepCor-Adv-large	0.14	0.17	0.24	0.24	0.37	0.28	0.24	9%
DeepCor-Gen-v1	0.20	0.18	0.27	0.33	0.33	0.41	0.29	29%
DeepCor-Gen-v2	**0.22**	**0.23**	**0.32**	**0.37**	**0.39**	**0.44**	**0.33**	**49%**

4 Discussion

Improving signal-to-noise in fMRI data is a long standing challenge that can unlock better diagnosis for psychiatric and neurodevelopmental disorders, and improve our understanding of brain function, informing diverse fields like neuroscience, cognitive science and artificial intelligence. Our results demonstrate that DeepCor substantially improves the recovery of task-evoked signals compared to widely used the CompCor method. We tested DeepCor models across diverse datasets, spanning sampling density (between 2 and 6 per subject) and tasks (block localizer and n-back matching task). These findings suggest that deep learning models, especially deep generative models like DeepCor-Gen-v2 can capture complex nonlinear noise–signal interactions that linear regression and component-based methods cannot, while remaining applicable to single-participant data. **Limitations.** Our evaluation focused on face/place contrasts, which are robust and widely used, but may not capture the full spectrum of fMRI applications. Performance in resting-state fMRI, naturalistic paradigms and event-related designs remain to be fully established. While we demonstrated improvements in signal quality, we did not yet examine downstream impacts on decoding or clinical prediction tasks.

References

Abubakar Abid and James Zou. Contrastive variational autoencoder enhances salient features, 2019. URL `https://arxiv.org/abs/1902.04601`.

Aidas Aglinskas, Alicia Bergeron, and Stefano Anzellotti. Understanding heterogeneity in psychiatric disorders: A method for identifying subtypes and parsing comorbidity. *Psychiatry and Clinical Neurosciences*, 79(7): 406–414, Jul 2025. doi: 10.1111/pcn.13829. Epub 2025 Apr 30.

Y. Behzadi, K. Restom, J. Liau, and T. T. Liu. A component based noise correction method (compcor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1):90–101, 2007. doi: 10.1016/j.neuroimage.2007.04.042.

L. T. Dowdle et al. Evaluating increases in sensitivity from NORDIC for diverse fMRI acquisition strategies. *NeuroImage*, 270:119949, 2023. doi: 10.1016/j.neuroimage.2023.119949.

E. DuPre, T. Salo, R. D. Markello, et al. Te-dependent analysis of multi-echo fMRI with `tedana`. *Journal of Open Source Software*, 6(66):3669, 2021. doi: 10.21105/joss.03669.

Russell Epstein and Nancy Kanwisher. A cortical representation of the local visual environment. *Nature*, 392 (6676):598–601, 1998. doi: 10.1038/33402.

O. Esteban et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, 16(1):111–116, 2019. doi: 10.1038/s41592-018-0235-4.

Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation, 2015. URL `https://arxiv.org/abs/1409.7495`.

Heo, Shin, Hung, Lin, Zhang, Shen, and Kam. Deep attentive spatio-temporal feature learning for automatic resting-state fmri denoising. *NeuroImage*, 254:119127, 2022. doi: 10.1016/j.neuroimage.2022.119127. Code available at `https://github.com/username/Automatic-rsfMRI-noise-detection`.

Nancy Kanwisher, Josh McDermott, and Marvin M. Chun. The fusiform face area: a cortical region specialized for the perception of faces. *Journal of Neuroscience*, 17(11):4302–4311, 1997. doi: 10.1523/JNEUROSCI.17-11-04302.1997.

K. N. Kay, A. Rokem, J. Winawer, R. F. Dougherty, and B. A. Wandell. GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Frontiers in Neuroscience*, 7:247, 2013. doi: 10.3389/fnins.2013.00247.

P. Kundu, V. Voon, P. Balchandani, M. V. Lombardo, B. A. Poser, and P. A. Bandettini. Multi-echo fMRI: A review of applications in fMRI denoising and analysis of BOLD signals. *NeuroImage*, 154:59–80, 2017. doi: 10.1016/j.neuroimage.2017.03.033.

S. Moeller et al. NORDIC: Noise reduction with distribution corrected PCA. *NeuroImage*, 226:117539, 2021. doi: 10.1016/j.neuroimage.2020.117539.

J. Muschelli, M. B. Nebel, B. S. Caffo, A. D. Barber, J. J. Pekar, and S. H. Mostofsky. Reduction of motion-related artifacts in resting state fMRI using acompcor. *NeuroImage*, 96:22–35, 2014. doi: 10.1016/j.neuroimage.2014.03.028.

R. H. R. Pruim, M. Mennes, J. K. Buitelaar, and C. F. Beckmann. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, 112:267–277, 2015. doi: 10.1016/j.neuroimage.2015.02.064.

G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90:449–468, 2014. doi: 10.1016/j.neuroimage.2013.11.046.

R. N. Spreng et al. Take a deep breath: Multiecho fMRI denoising effectively removes head motion artifacts. *Proceedings of the National Academy of Sciences*, 116(44):21970–21972, 2019. doi: 10.1073/pnas.1909848116.

Theodoropoulos, Chatzichristos, and Van Huffel. Automatic artifact removal of resting-state fmri with deep neural networks. arXiv preprint, 2021. No public code available.

L. Vizioli et al. Lowering the thermal noise barrier in functional brain mapping. *Nature Communications*, 12: 5181, 2021. doi: 10.1038/s41467-021-25431-8.

Yang, Zhuang, Sreenivasan, Mishra, and Cordes. Disentangling time series between brain tissues improves fmri data quality using a time-dependent deep neural network. *NeuroImage*, 223:117340, 2020a. doi: 10.1016/j.neuroimage.2020.117340. Partial code release; same code base as above.

Yang, Zhuang, Sreenivasan, Mishra, Curran, and Cordes. A robust deep neural network for denoising task-based fmri data: An application to working memory and episodic memory. PubMed Central, PMCID: PMC6980789, 2020b. Code available at `https://github.com/username/DeNN-task-fMRI-denoising`.

Zhengshi Yang, Xiaowei Zhuang, Karthik R. Sreenivasan, Virendra Mishra, Tim Curran, and Dietmar Cordes. A robust deep neural network for denoising task-based fmri data: An application to working memory and episodic memory. *Medical Image Analysis*, 60:101622, Feb 2020c. doi: 10.1016/j.media.2019.101622.

Theodore P. Zanto, Judy Pa, and Adam Gazzaley. Reliability measures of functional magnetic resonance imaging in a longitudinal evaluation of mild cognitive impairment. *NeuroImage*, 84:443–452, 2014. doi: 10.1016/j.neuroimage.2013.08.063.

Zhao, Li, Jiao, Du, and Fan. A 3d convolutional encapsulated long short-term memory (3dconv-lstm) model for denoising fmri data. PubMed Central, PMCID: PMC7687287, 2020. No public code available.

Supplementary Material

A Code availability

Code used to train the models used in this manuscript is made publically availabe at the following URL: anonymous.4open.science/r/pub-neurips-workshop-DBM-3EEF/.

B ROI and RONI inputs

Consider a functional MRI (fMRI) dataset with dimensions $L \times W \times H \times T$, where $L \times W \times H$ represents the total number of voxels in the brain volume and T denotes the number of timepoints in the fMRI time series. To extract the temporal signals, We applied a mask to the RONI to obtain the timecourses of the m RONI voxels, and a separate mask to the ROI to retrieve the timecourses of the n ROI voxels. The ROI mask was constructed by thresholding the gray matter probability map generated during segmentation, including only voxels with a probability $p > 0.50$. Conversely, the RONI mask was derived by combining the probability maps for white matter and cerebrospinal fluid and applying the same probability threshold of $p > 0.50$. To ensure consistency and comparability, each voxel's timecourse was standardized by subtracting its mean and scaling by its standard deviation, resulting in normalized time series with zero mean and unit variance.

C Architecture of the DeepCor-Adv model

DeepCor-Adv. DeepCor-Adv is an adversarial 1D convolutional autoencoder that learns, within a single subject, to retain information that is predictive of gray-matter (GM) BOLD time courses while actively suppressing information that is predictive of nuisance sources measured in regions-of-no-interest (RONI; white matter and CSF). The model is composed of a *shared encoder* and two *mirrored decoders*. The encoder takes in a batch of GM or RONI voxel time series and compresses them into a latent representation; the signal decoder maps this latent code back into a denoised GM-like time series, whereas the noise decoder reconstructs RONI-like time series. A gradient-reversal layer (GRL) is placed in front of the noise decoder: it leaves the forward pass unchanged but multiplies the backward gradient by $-\lambda$ (here $\lambda = 1$), so that any latent feature that makes RONI reconstruction easier will, by design, push the encoder to *discard* it. This mechanism implements an adversarial trade-off: the encoder is rewarded for preserving structure needed to reconstruct GM, and simultaneously penalized for representing structure that would explain RONI.

Concretely, the encoder consists of three 1D convolutional blocks with ReLU nonlinearities (kernel size 3, stride 2, padding 1), which progressively downsample the sequence length while increasing channel capacity. This produces a compact latent sequence z that is shared across branches. The signal decoder mirrors this hierarchy with three ConvTranspose1d layers (kernel 3, stride 2, padding 1, output_padding 1), followed by a Sigmoid function. The noise decoder has the same transposed-convolutional structure but operates on $\mathrm{GRL}(z)$, ensuring that its training signal exerts an *opposing* pressure on the encoder.

Training optimizes reconstruction on both branches with losses defined between inputs and their respective reconstructions. We use a mean-squared error (MSE) objective, and a normalized cross-correlation (NCC) objective that also emphasizes temporal alignment independent of amplitude.

$$\mathcal{L}_{\text{MSE}} = \text{MSE}(\text{GM},\ \hat{\text{GM}}) + \text{MSE}(\text{RONI},\ \hat{\text{RONI}})$$
$$\mathcal{L}_{\text{NCC}} = [1 - \text{NCC}(\text{GM},\ \hat{\text{GM}})] + [1 - \text{NCC}(\text{RONI},\ \hat{\text{RONI}})]$$

The GM term compels the encoder and signal decoder to preserve task-relevant BOLD structure; the RONI term trains the noise decoder *and*, via the GRL, yields a *negative* gradient to the encoder whenever z contains features that explain RONI. As a result, the easiest solution for the encoder should be to learn a representation that is maximally informative for GM and minimally informative for RONI. Overall, the combination of a shared encoder, mirrored decoders, and gradient-reversal delivers a compact latent representation that disentangles neural signal from nuisance without relying on explicit labels of noise sources or multi-subject supervision.

Table S1: DeepCor-Adv architecture. Shapes shown as (Channels, Length). DeepCor-Adv-large has an additional Conv1d layer with 64 channels.

Block	**Layer**	**Hyperparameters**	**Output shape**
Encoder (shared for GM and RONI inputs)			
	Conv1d	in=1, out=16, k=3, s=2, p=1	(16, $L/2$)
	ReLU	—	(16, $L/2$)
	Conv1d	in=16, out=32, k=3, s=2, p=1	(32, $L/4$)
	ReLU	—	(32, $L/4$)
	Conv1d	in=32, out=d_z, k=3, s=2, p=1	(d_z, $L/8$)
	ReLU	—	(d_z, $L/8$)
Signal decoder (GM reconstruction from z)			
	ConvTranspose1d	in=d_z, out=32, k=3, s=2, p=1, out_pad=1	(32, $L/4$)
	ReLU	—	(32, $L/4$)
	ConvTranspose1d	in=32, out=16, k=3, s=2, p=1, out_pad=1	(16, $L/2$)
	ReLU	—	(16, $L/2$)
	ConvTranspose1d	in=16, out=1, k=3, s=2, p=1, out_pad=1	(1, L)
	Sigmoid	—	(1, L)
Noise decoder (RONI reconstruction from GRL(z))			
	GRL	$\lambda = 1.0$	(d_z, $L/8$)
	ConvTranspose1d	in=d_z, out=32, k=3, s=2, p=1, out_pad=1, **bias=False**	(32, $L/4$)
	ReLU	—	(32, $L/4$)
	ConvTranspose1d	in=32, out=16, k=3, s=2, p=1, out_pad=1, **bias=False**	(16, $L/2$)
	ReLU	—	(16, $L/2$)
	ConvTranspose1d	in=16, out=1, k=3, s=2, p=1, out_pad=1, **bias=False**	(1, L)
	Sigmoid	—	(1, L)

D Architecture of the DeepCor-Gen models

The **DeepCor-Gen-v1** is an autoencoder comprising two probabilistic encoders and a decoder. The first encoder, $q_{\phi_s}(s|x)$, extracts signal-related features, while the second encoder, $q_{\phi_z}(z|x)$ captures noise-specific features. The decoder $f_\theta(s, z)$, reconstructs the original measurement by processing the concatenated outputs of the two encoders. Each encoder is built with four one-dimensional (1D) convolutional layers, followed by two parallel fully-connected layers that parameterize the mean and standard deviation of the latent distribution. The decoder architecture includes a fully connected layer bridging to the latent space, four 1D transposed convolutional layer, and a final convolutional layer for output reconstruction. **DeepCor-Gen-v2** builds on this by making the signal–noise factorization more explicit, adding regularization to reduce leakage of nuisance structure into the signal latent and incorporating coordinate and motion information. Architecturally, v2 retains the two-encoder/one-decoder design but enforces an additive decomposition of each measurement: instead of reconstructing a target sample with a single decode from $[z; s]$ as in v1, the model reconstructs it as the *sum of two complementary decodes*, one that is forced to rely only on the putative signal reconstruction and the other only on the nuisance reconstruction. This compositional reconstruction pressure encourages z and s to account for non-overlapping variance, yielding a cleaner separation at decode time when we set the nuisance code to zero for denoising. Input is expanded from 1D to 4D, where each timepoint has associated xyz voxel coordinates. MSE loss reconstruction is only calculated between 1D time series inputs and reconstructions: this way coordinate information can be used by convolutional filters if it enabled more accurate reconstruction. To further discourage contamination of the signal latent with motion and other confounds, v2 introduces *auxiliary confound decoders* that read out known confound information (subject motion: 3 translation and 3 rotation parameters) directly from each latent (one head from z, one from s). These heads enable explicit supervision: z is trained to *predict* confounds, while s is trained *adversarially* not to (via a gradient-reversal layer available in the module), thereby pushing confound-related information into z. The loss is correspondingly expanded: in addition to the β-weighted KL regularization used in v1, v2 includes an NCC loss, like the one used in DeepCor-Adv; cross-decoding loss that penalizes reconstruction of GM voxels from RONI voxels and a smoothness loss that penalizes high variance. Taken together, these changes make v2 more expressive more identifiable (through additive factorization), and more robust to nuisance leakage (through confound supervision and adversarial discouragement), while preserving the simple inference procedure of decoding with the nuisance code clamped to zero for denoised outputs.

Table S2: DeepCor-Gen-v1 (contrastive VAE) architecture. Shapes shown as (Channels, Length). L denotes the input sequence length. Latent dim. was set to 8.

Block	Layer	Hyperparameters	Output shape
Signal encoder E_z			
	Conv1d + BN + LeakyReLU	in=1, out=64, k=3, s=2, p=`pad[4]`	(64, L_1)
	Conv1d + BN + LeakyReLU	in=64, out=128, k=3, s=2, p=`pad[3]`	(128, L_2)
	Conv1d + BN + LeakyReLU	in=128, out=256, k=3, s=2, p=`pad[2]`	(256, L_3)
	Conv1d + BN + LeakyReLU	in=256, out=256, k=3, s=2, p=`pad[1]`	(256, L_4)
	Flatten	—	$(256 \times L_4)$
	Linear (mean)	$256 \times L_4 \to d$	(d)
	Linear (log-var)	$256 \times L_4 \to d$	(d)
Noise encoder E_s (same backbone as E_z)			
	Conv1d + BN + LeakyReLU	in=1, out=64, k=3, s=2, p=`pad[4]`	(64, L_1)
	Conv1d + BN + LeakyReLU	in=64, out=128, k=3, s=2, p=`pad[3]`	(128, L_2)
	Conv1d + BN + LeakyReLU	in=128, out=256, k=3, s=2, p=`pad[2]`	(256, L_3)
	Conv1d + BN + LeakyReLU	in=256, out=256, k=3, s=2, p=`pad[1]`	(256, L_4)
	Flatten	—	$(256 \times L_4)$
	Linear (mean)	$256 \times L_4 \to d$	(d)
	Linear (log-var)	$256 \times L_4 \to d$	(d)
Decoder D (shared)			
	Linear (decoder_input)	$[z; s] \in \mathbb{R}^{2d} \to 256 \times L_4$	(256, L_4)
	ConvTranspose1d + BN + LeakyReLU	in=256, out=256, k=3, s=2L_3)	
	ConvTranspose1d + BN + LeakyReLU	in=256, out=128, k=3, s=2L_2)	
	ConvTranspose1d + BN + LeakyReLU	in=128, out=64, k=3, s=2L_1)	
	ConvTranspose1d + BN + LeakyReLU	in=64, out=64, k=3, s=2	
	Conv1d (final)	in=64, out=1, k=3, p=1	(1, L)

Table S3: DeepCor-Gen-v2 (contrastive VAE with disentanglement and confound supervision). Shapes shown as (Channels, Length). L denotes the input sequence length; L_i are intermediate temporal sizes computed by `compute_padding`. Default `hidden_dims` $= [64, 128, 256, 256]$, asymmetric latent widths $(d_z, d_s) = (8, 8)$.

Block	**Layer**	**Hyperparameters**	**Output shape**
Signal encoder E_z			
	Conv1d + LeakyReLU	in=1, out=64, k=3, s=2, p=`pad[4]`	(64, L_1)
	Conv1d + LeakyReLU	in=64, out=128, k=3, s=2	(128, L_2)
	Conv1d + LeakyReLU	in=128, out=256, k=3, s=2	(256, L_3)
	Conv1d + LeakyReLU	in=256, out=256, k=3, s=2	(256, L_4)
	Flatten	—	$(256 \times L_4)$
	Linear (mean)	$256 \times L_4 \to d_z$	(d_z)
	Linear (log-var)	$256 \times L_4 \to d_z$	(d_z)
Noise encoder E_s (same backbone as E_z)			
	Conv1d + LeakyReLU	in=1, out=64, k=3, s=2	(64, L_1)
	Conv1d + LeakyReLU	in=64, out=128, k=3, s=2	(128, L_2)
	Conv1d + LeakyReLU	in=128, out=256, k=3, s=2	(256, L_3)
	Conv1d + LeakyReLU	in=256, out=256, k=3, s=2	(256, L_4)
	Flatten	—	$(256 \times L_4)$
	Linear (mean)	$256 \times L_4 \to d_s$	(d_s)
	Linear (log-var)	$256 \times L_4 \to d_s$	(d_s)
Main decoder D (additive reconstruction from z and s)			
	Linear (decoder_input)	$[z; s] \in \mathbb{R}^{d_z + d_s} \to 256 \times L_4$	(256, L_4)
	ConvTranspose1d + LeakyReLU	in=256, out=256, k=3, s=2	(256, L_3)
	ConvTranspose1d + LeakyReLU	in=256, out=128, k=3, s=2	(128, L_2)
	ConvTranspose1d + LeakyReLU	in=128, out=64, k=3, s=2	(64, L_1)
	ConvTranspose1d + LeakyReLU	in=64, out=64, k=3, s=2	(64, L)
	Conv1d (final)	in=64, out=1, k=3, p=1	(1, L)
Auxiliary confound decoders (supervised disentanglement)			
Decoder from z	ConvTranspose1d	in=d_z, out=128, k=L, s=1, bias=False	(128, L_3)
	Conv1d + ReLU	in=128, out=32, k=3, s=1, p=1	(32, L_2)
	Conv1d + ReLU	in=32, out=16, k=3, s=1, p=1	(16, L_1)
	Conv1d + Sigmoid	in=16, out=6, k=3, s=1, p=1	(6, L)
Decoder from s	ConvTranspose1d	in=d_s, out=128, k=L, s=1, bias=False	(128, L_3)
	Conv1d + ReLU	in=128, out=32, k=3, s=1, p=1	(32, L_2)
	Conv1d + ReLU	in=32, out=16, k=3, s=1, p=1	(16, L_1)
	Conv1d + Sigmoid	in=16, out=6, k=3, s=1, p=1	(6, L)

E Discussion of findings using DeNN model.

Despite using the code provided by the authors as instructed in their GitHub repository (github.com/CCLRCBH-BIC/DeNN), we were surprised by the underwhelming results it produced. We have identified two potential reasons for the DeNN models failure and discuss how DeepCor overcomes these challenges. First, DeNN uses the same network and weights regardless of whether the input time series is from a voxel of interest (e.g. in grey matter) or from a voxel of no interest (e.g. in white matter or cerebrospinal fluid). By contrast, DeepCor uses two separate branches for processing signal and noise information: while the "noise" branch is used for all voxels, the "signal" branch is only used for voxels from regions of interest, enabling DeepCor to learn specialized weights for the extraction of information about the neural signal. Second, DeNN relies on a loss function that minimizes the absolute value of the correlation between outputs generated for inputs from grey matter and for inputs from regions of no interest. However, in the absence of other constraints, the outputs can become uncorrelated while also losing information about the original responses. DeepCor, thanks to its encoders that compute distinct latent spaces for signal and noise, also separates between these two sets of features. However, the additional constraint of having to reconstruct the original measurements through a decoder forces the encoders to preserve information, encouraging a more comprehensive characterization of signal and noise.

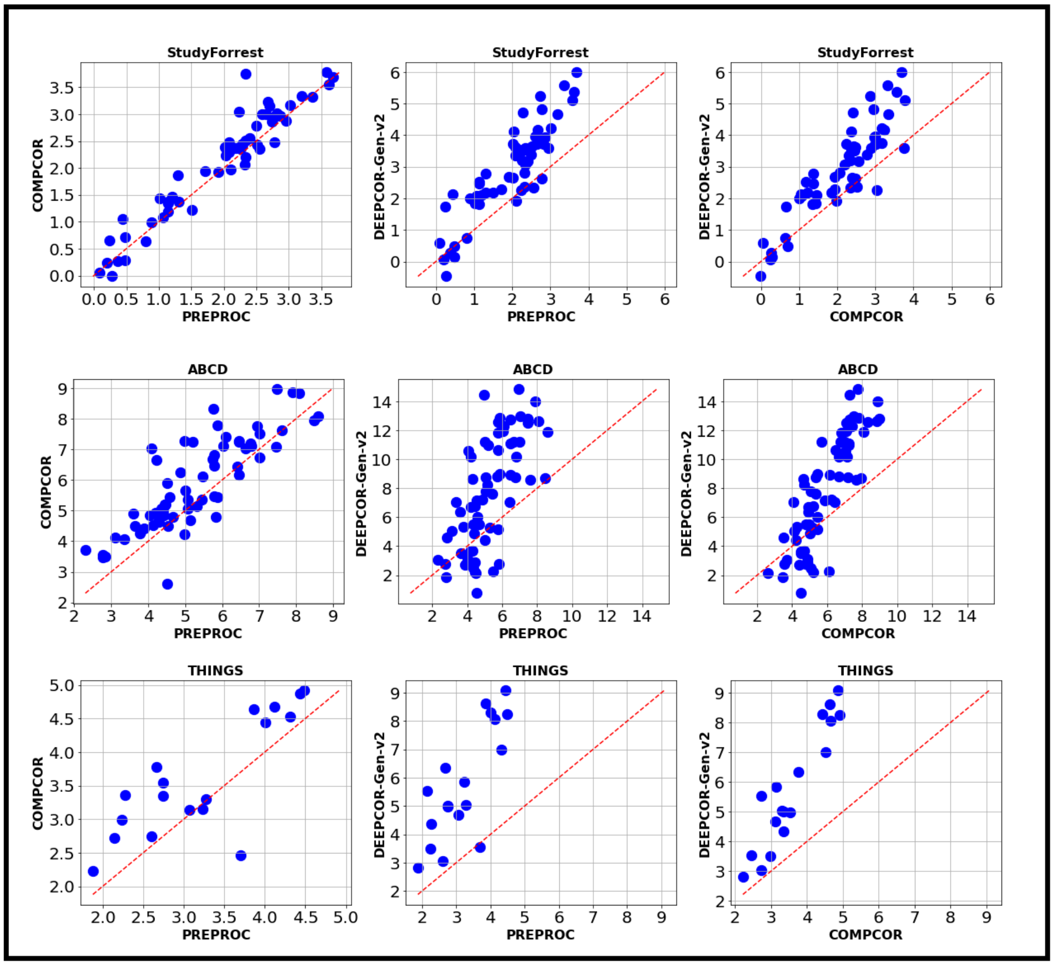

Figure S1: Scatterplots showing contrast estimates (average of face and place contrasts for each dataset) comparing: no denoising, CompCor and DeepCor-Gen-V2. Each run is plotted separately so the total number of dots is number of subjects times the number of runs. Red line denotes parity.

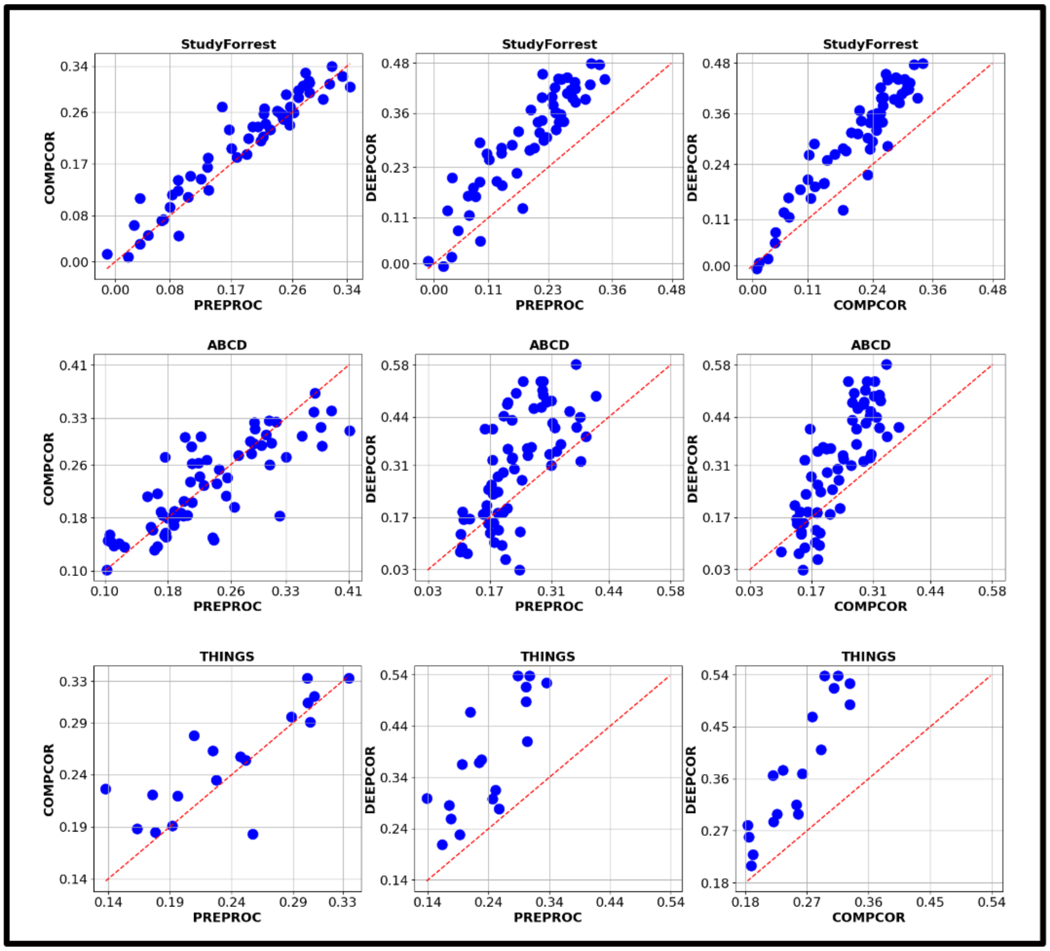

Figure S2: Scatterplots showing category responsivity estimates (average of face and place regressors with the BOLD signal in FFA and PPA, respectively) comparing: no denoising, CompCor and DeepCor-Gen-V2. Each run is plotted separately so the total number of dots is number of subjects times the number of runs. Red line denotes parity.