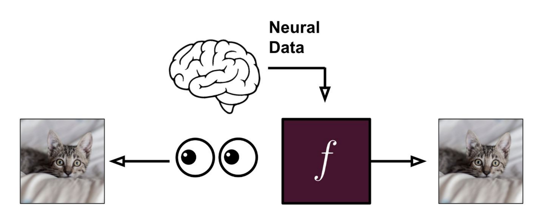
BRAID: Brain Representation to Artificial Image via Diffusion

Teo Imoto-Tar, Leela Noguchi

Research Question

- Image reconstruction from neural data
- Utilize pre-trained diffusion models but condition on fMRI

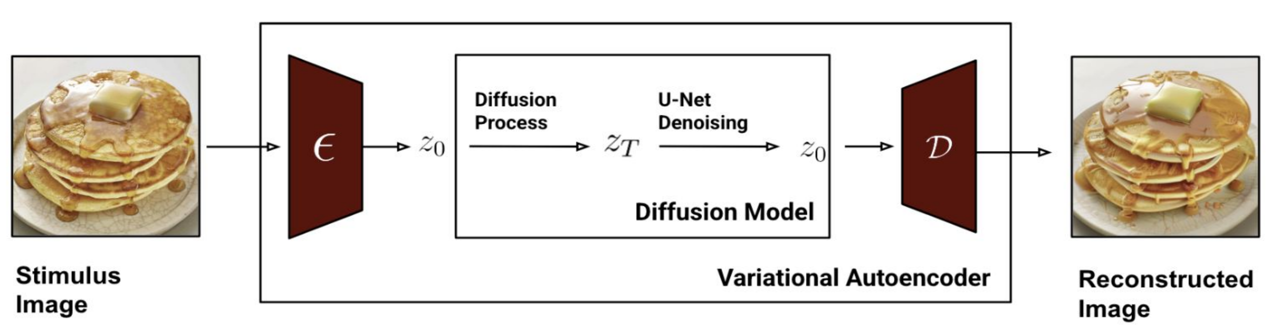


Objectives

- Can pretrained **diffusion models** be conditioned on fMRI data to reconstruct stimuli?
- What's the minimal parameter space and the maximum achievable reconstruction quality?

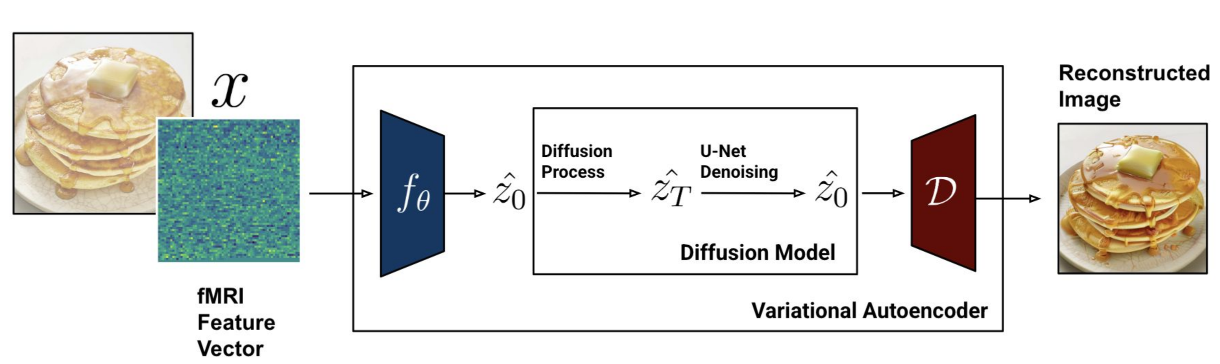
Baseline Model

SSD-1B Pretrained Latent Diffusion Model (LDM)
[Gupta et al., Arxiv, 2024] – **Img2Img**



Swap out LDM VAE encoder with our own **trained fMRI encoder** to produce similar initial latents to z_0

Model Architecture



Let $f_\theta : \mathbb{R}^{D_{\mathrm{fMRI}}} \to \mathbb{R}^k$ learned encoder that maps fMRI feature vector x to latent $\hat{z}_0 = f_\theta(x)$

Latent Diffusion

Forward (Noising) Process: Iteratively add Gaussian noise to clean latent

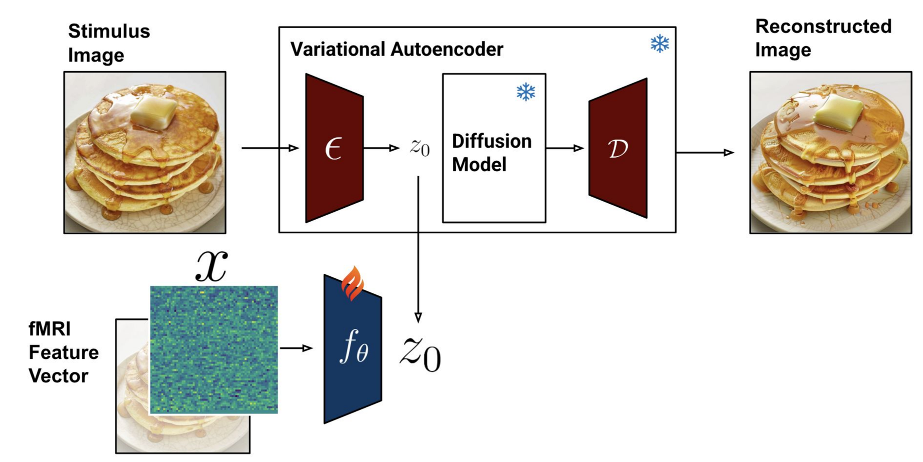
$$q(z_t \mid z_0) = \mathcal{N}\big(z_t;\, \sqrt{\bar{\alpha}_t}\, z_0,\, (1-\bar{\alpha}_t)\mathbf{I}\big) \text{ where } \bar{\alpha}_t = \textstyle\prod_{s=1}^{t}(1-\beta_s)$$

Backward (Denoising) Process: Iteratively remove noise from latent

$$z_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(z_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\epsilon_\phi(z_t,t)\right) + \sigma_t\,\eta, \qquad \eta \sim \mathcal{N}(0,\mathbf{I}), \quad t = T, T-1, \ldots, 1.$$

We follow the formulation in [Rombach *et al.*, CVPR 2022]

Cross Modal Framework



a) Stimulus image
b) Initial latent
c) fMRI feature vector

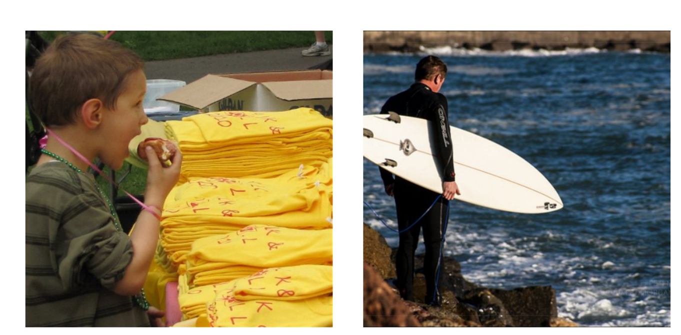
Dataset

Natural Scenes Dataset (2021) Overview [Allen et al., Nature, 2022]
- fMRI recorded with cooperative subject viewing controlled image stimuli

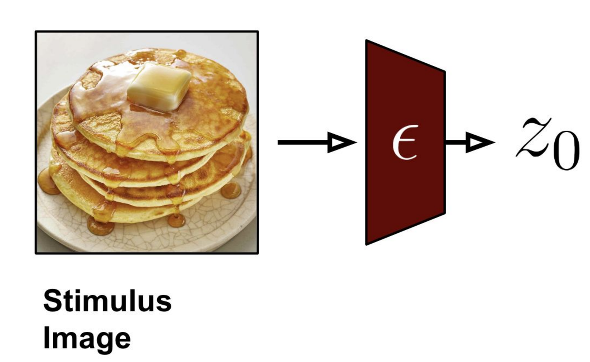
1. **COCO Stimulus Images** [Tsung-Yi Lin et al., Arxiv, 2014]
2. **"β"** = single-trial GLM (Generalized Linear Model)
   a. Estimated activity per voxel in response to an image stimulus
   b. ROI = Region of Interest

Data Preprocessing

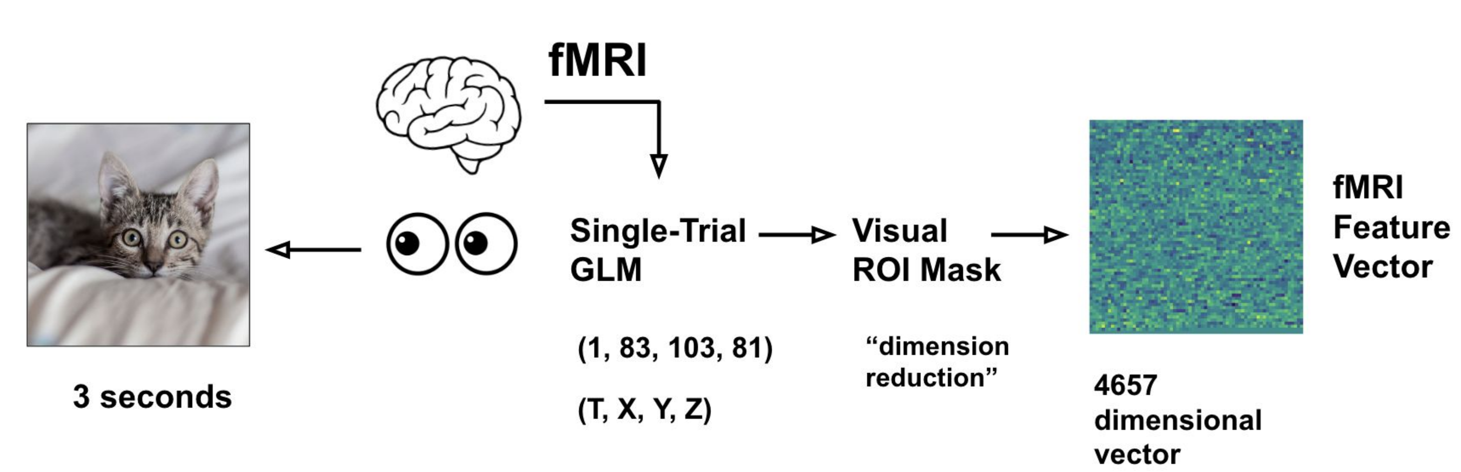
COCO Image stimulus



Retrieve Initial Latent

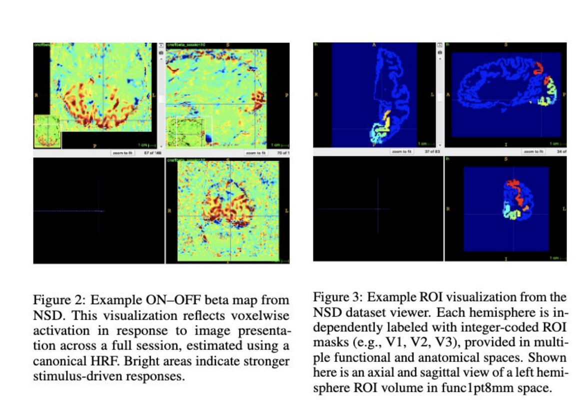


fMRI Preprocessing Pipeline



β

Single-Trial GLM



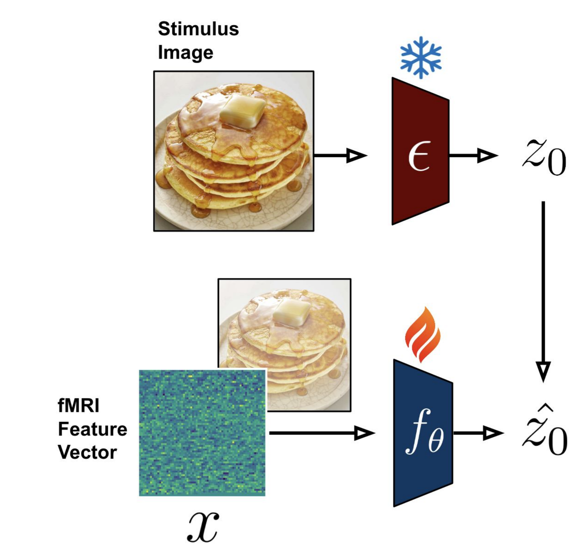
Visual ROI (Region of Interest) Mask

Training

Training f_θ to produce $\hat{z}_0$ that matches target z_0

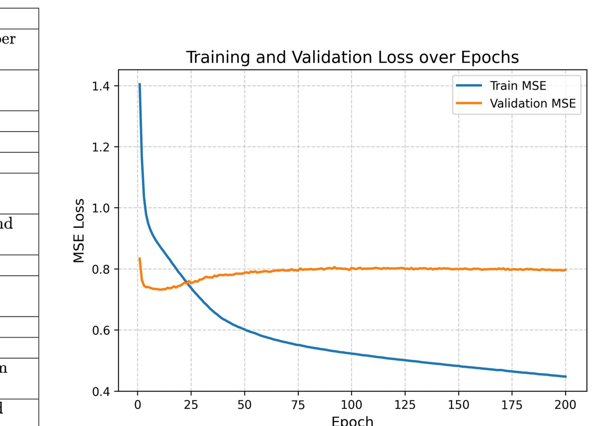
Objective Function: L2 loss

$$\mathcal{L}_{\mathrm{enc}} = \|f_\theta(x) - z_0\|_2^2$$

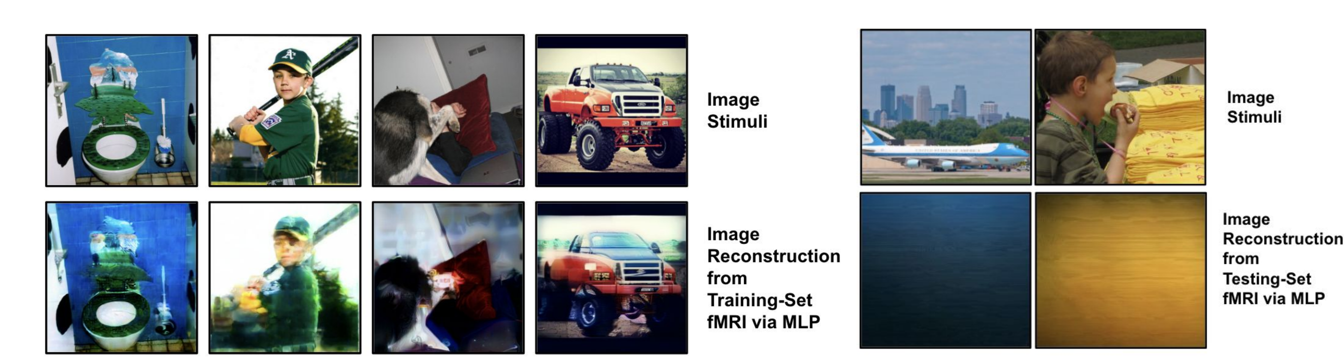


MLP Hyperparameters

Hyperparameter	Value	Description
Input dimension	4,657	Dimensionality of fMRI input vector per sample
Output dimension	16,384	Dimensionality of target latent from SDXL VAE (64×64×4)
Hidden layers	[2048, 4096]	Two-layer MLP with increasing width
Normalization	LayerNorm	Applied after each linear layer
Activation function	ReLU	Used after each LayerNorm
Dropout rate	0.5	Applied after ReLU in each layer for regularization
Loss function	Mean Squared Error (MSE)	Measures L2 loss between predicted and target latent
Optimizer	Adam	Adaptive learning rate optimizer
Learning rate	1×10^{-4}	Fixed learning rate used throughout training
Batch size	512	Number of samples per mini-batch
Epochs	200	Full passes through the training data
Normalization of inputs	z-score	Mean/std normalization per voxel from training set
Target normalization	Latent z-score	Latents normalized using precomputed mean/std
Train/Val/Test split	90/5/5	Split ratio over the 7,500 total fMRI-latent pairs



Inference



Conclusion

- Demonstrates lightweight models can achieve meaningful cross modal alignment
- Highlights strong potential for student-teacher distillation to transfer knowledge from large foundational models

References

- [1] E. J. Allen et al., "A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence," *Nature Neuroscience*, vol. 25, pp. 116–126, 2022.
- [2] Y. Gupta, V. V. Jaddipal, H. Prabhala, S. Paul, and P. Von Platen, "Progressive Knowledge Distillation of Stable Diffusion XL Using Layer Level Loss," *arXiv preprint arXiv:2401.02677*, 2024.
- [3] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context," *arXiv preprint arXiv:1405.0312*, 2014.
- [4] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, 2022.