On fMRI preprocessing in natural image reconstruction

Mohammad Hadji, Yunlai Chen, Cornelia Wang, Hamed Yousefi

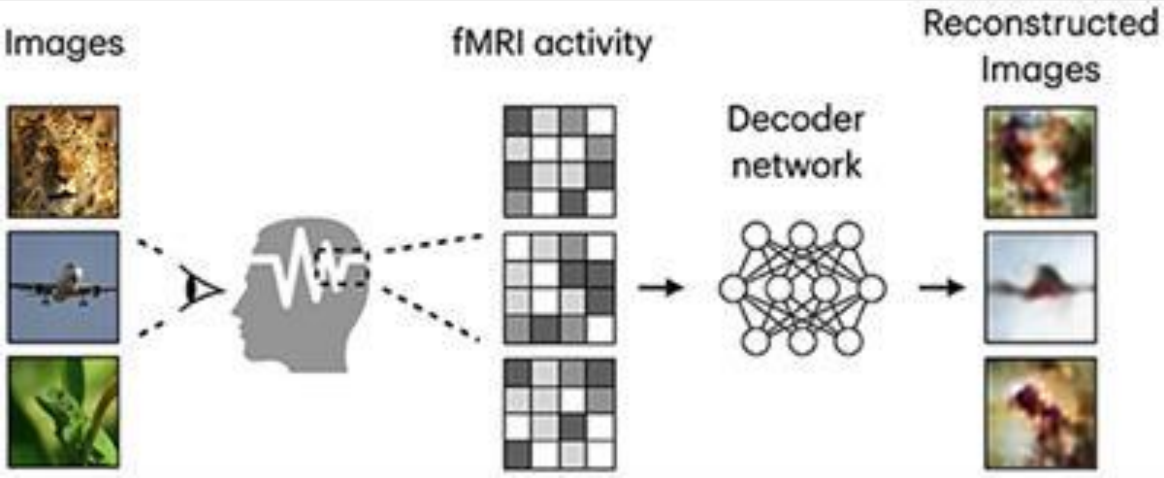
Washington University in St. Louis

Introduction

- Reconstruction of seen images drew attention due to its potential to decode the contents of dreams and mental imaginary, as well as to contribute on brain-computer-interfaces

- functional magnetic resonance imaging (fMRI) is the most popular neuroimaging modality for brain decoding

- **The deep learning approach:** Train a neural network to decode fMRI activity to image stimulus.



In supervised learning, acquiring sufficient image-fMRI pairs of data is challenging.

Self-supervised Autoencoder model:

This model was presented in:

"From voxels to pixels and back: Self-supervision in natural-image reconstruction from fMRI. "

Published on 33rd Conference on Neural Information Processing Systems (NeurIPS 2019)

Beliy et al., 2019

Problem: Decoding studies vary the model and the preprocessing pipeline

BigBiGan:

"applied a standard preprocessing pipeline: slice-time correction, realignment, and coregistation to the T1w anatomical image. "

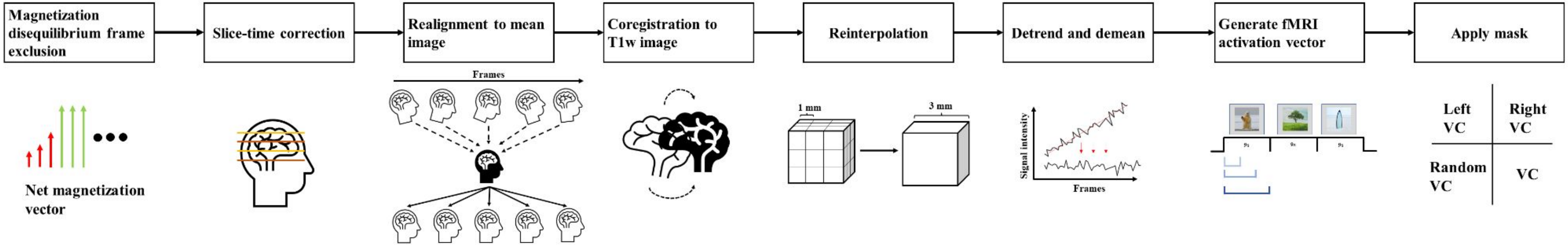
Autoencoder:

"To accommodate the center-fixation uncertainty, we introduced random shifts of the input images during Encoder training."

Our novelties:

Vary:

- Haemodynamic shift parameters
- Frames averaged in the stimulus window
- Brain mask



Methods

Dataset

T. Horikawa and Y. Kamitani, "Generic decoding of seen and imagined objects using hierarchical visual features," Nature communications, vol. 8, no. 1, pp. 1–15, 2017.

Two experiments (train, test) were presented to **five** subjects over two months

The training experiment:

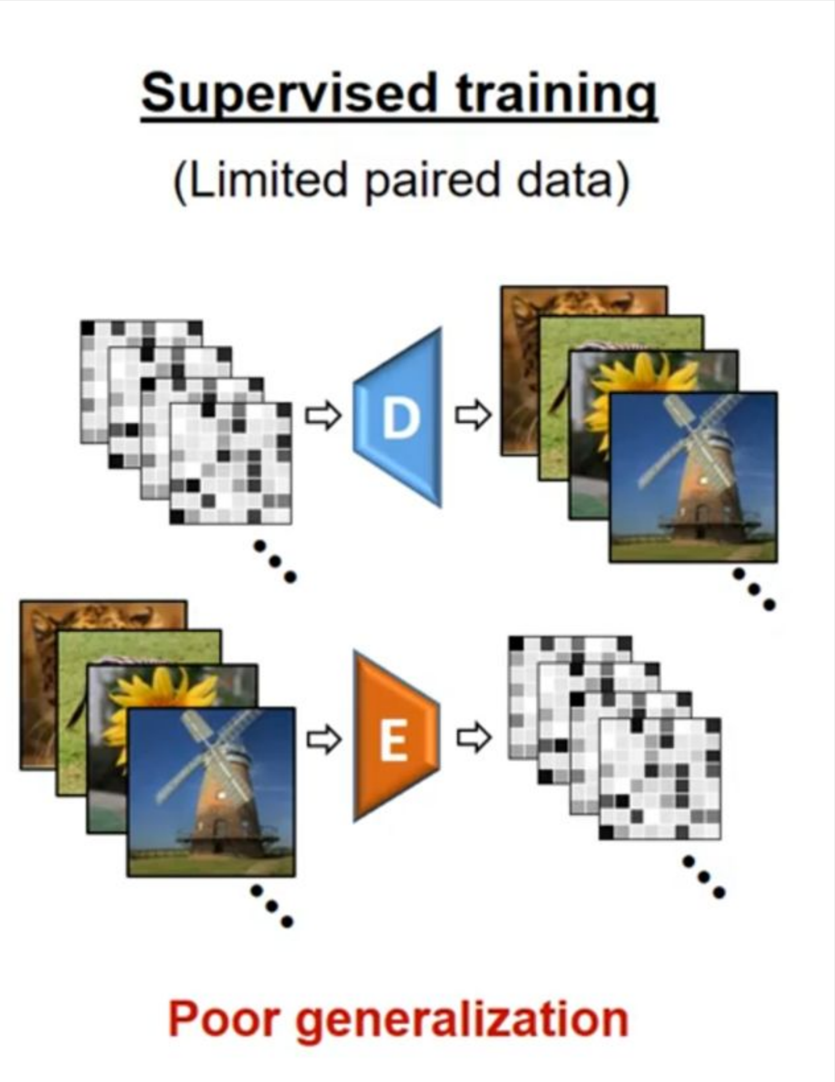
- ❑ Present **1200** images from ImageNet across 24 runs. Run duration: 9 minutes and 54 seconds.
- ❑ Each run, subjects viewed 50 unique images from the 1200 image set. Five 1-back tasks (55 images total presented per run).
- ❑ A 33 and 6-second rest period was added to the start and end of each run.

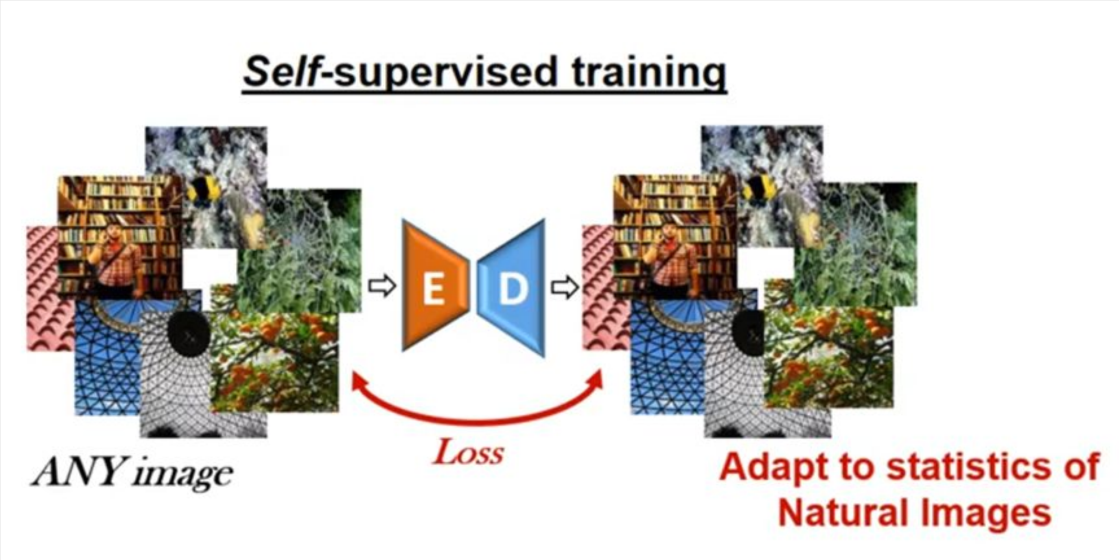
The test experiment:

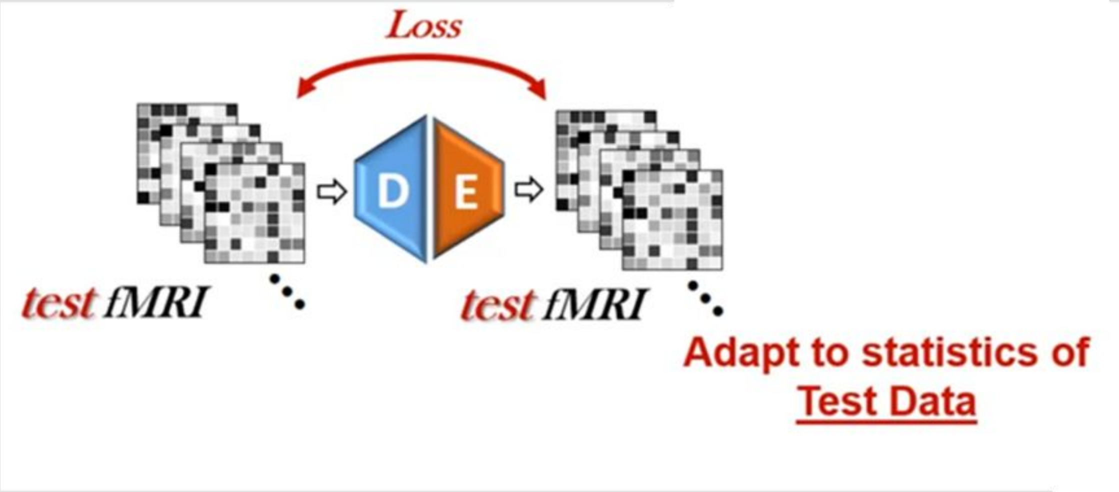
- ❑ A set of **50** images from ImageNet not shown during training were shown 35 times.
- ❑ All experimental parameters were identical to the training experiment.

*We limited the analysis to subject 2 and 3

Autoencoder overview

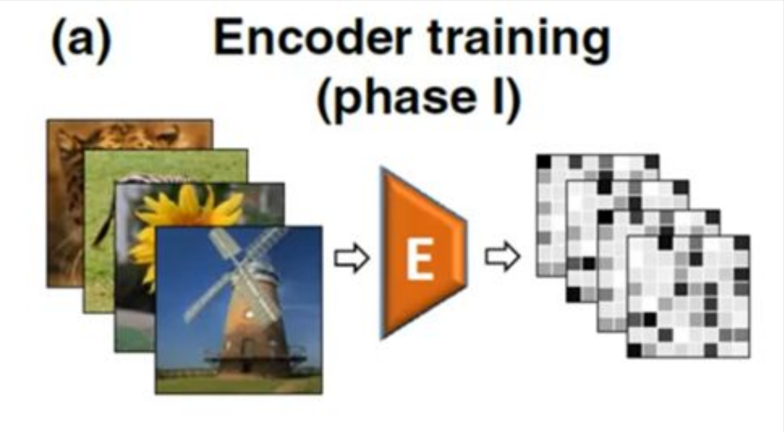




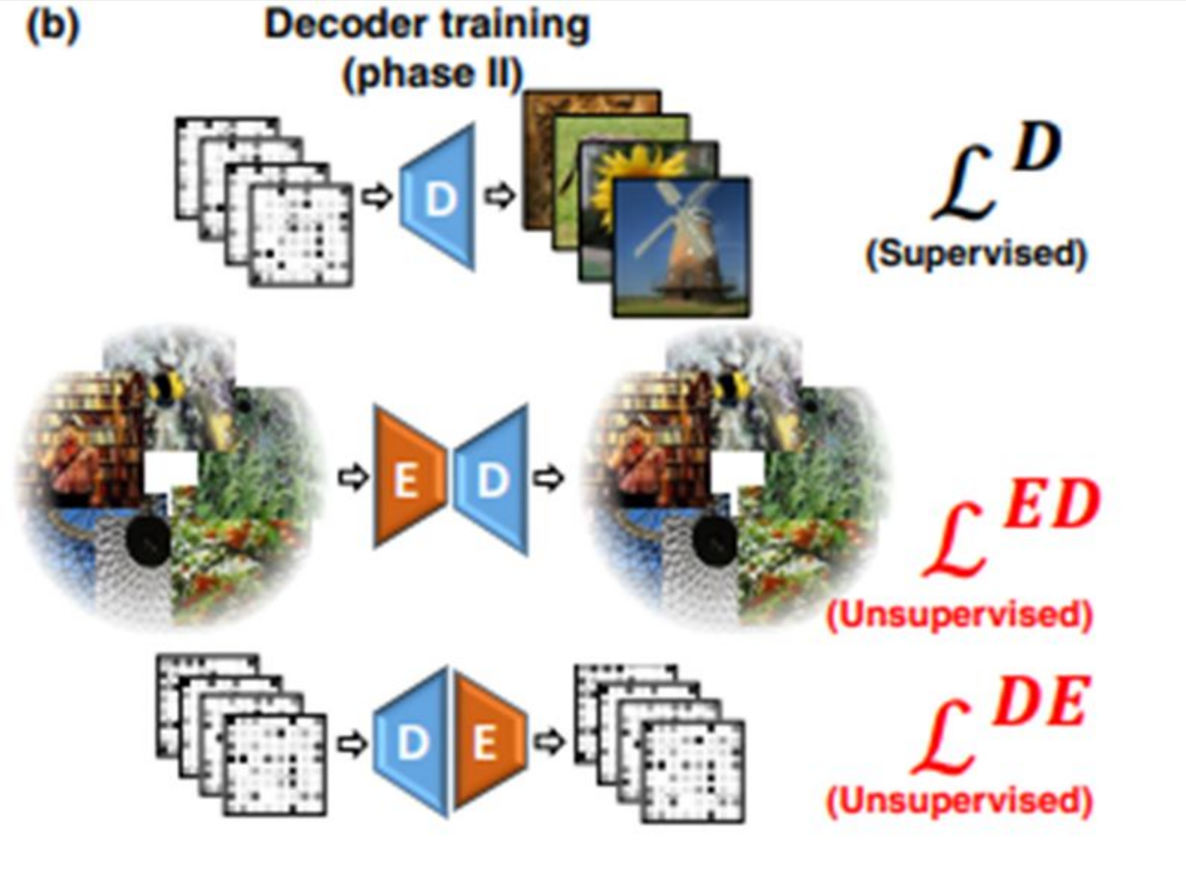


Training

Train the encoder on {fMRI, Image} pairs

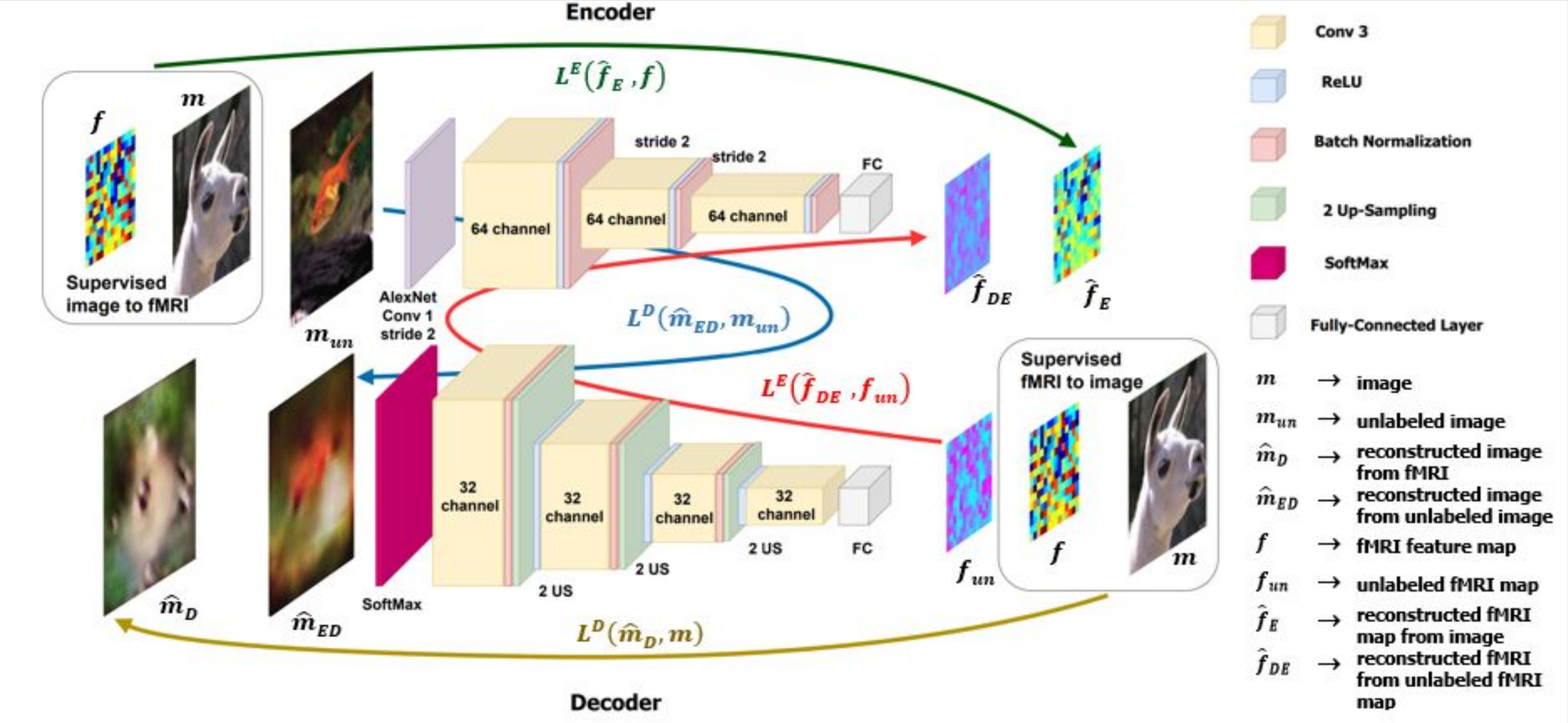


Train the decoder on {fMRI, Image} pairs and unlabeled image- and fMRI samples



- 2-stage training: first train encode in a supervised way; then fix the encoder and train decoder
- **50k** unlabeled natural images on image net used for unsupervised training
- Trained on CHPC. Each dataset takes **~1h** on Tesla T4 and **~15min** on V100.

Autoencoder architecture



Quantitative analysis:

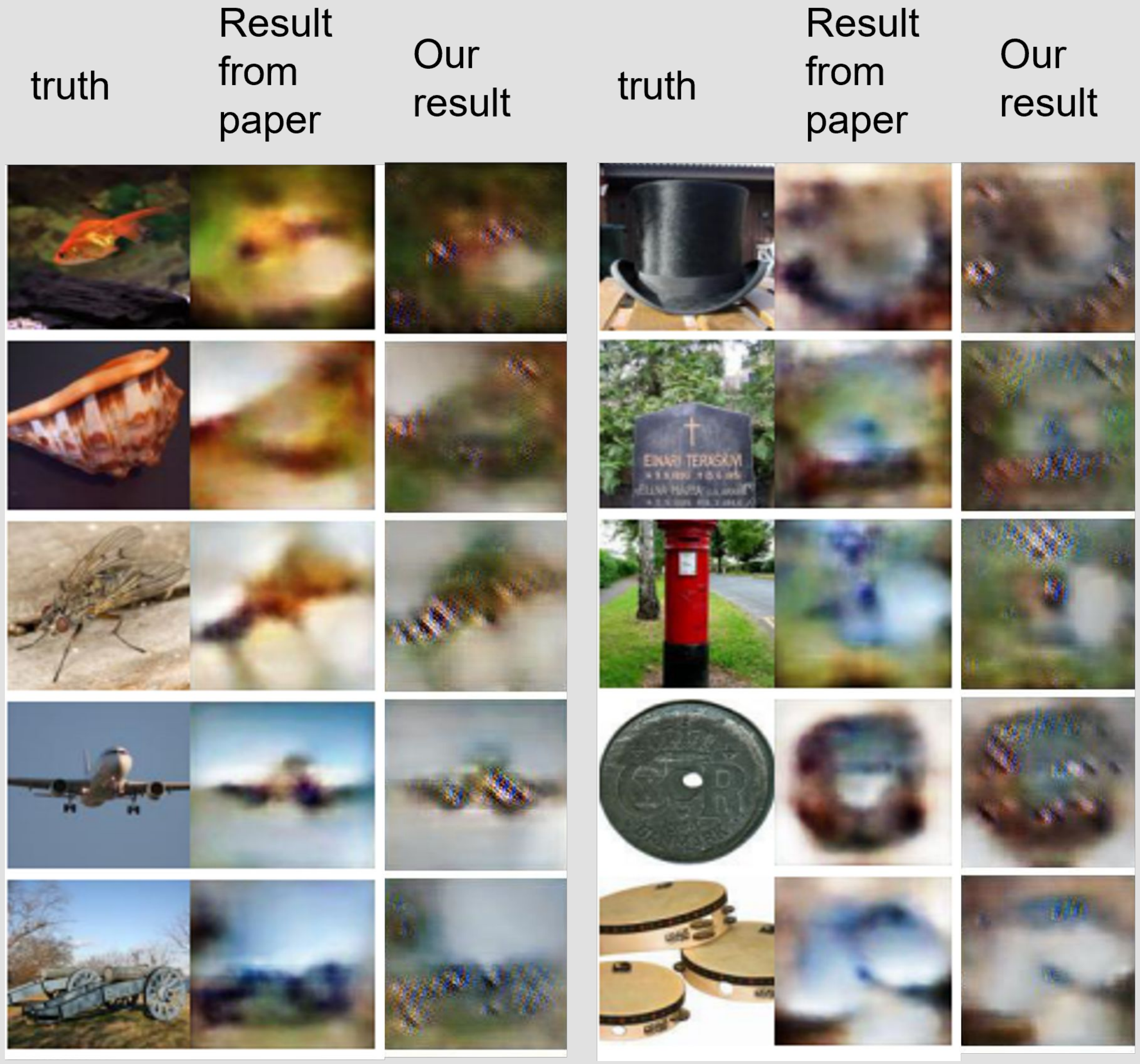
$$CosineProx(x,y) = \frac{\boldsymbol{x}.\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{({\mu_x}^2 + {\mu_y}^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

$$Correlation = \frac{\sum_m \sum_n (x_{mn} - \mu_x)(y_{mn} - \mu_y)}{\sqrt{(\sum_m \sum_n (x_{mn} - \mu_x)^2)((\sum_m \sum_n (y_{mn} - \mu_y)^2))}}$$

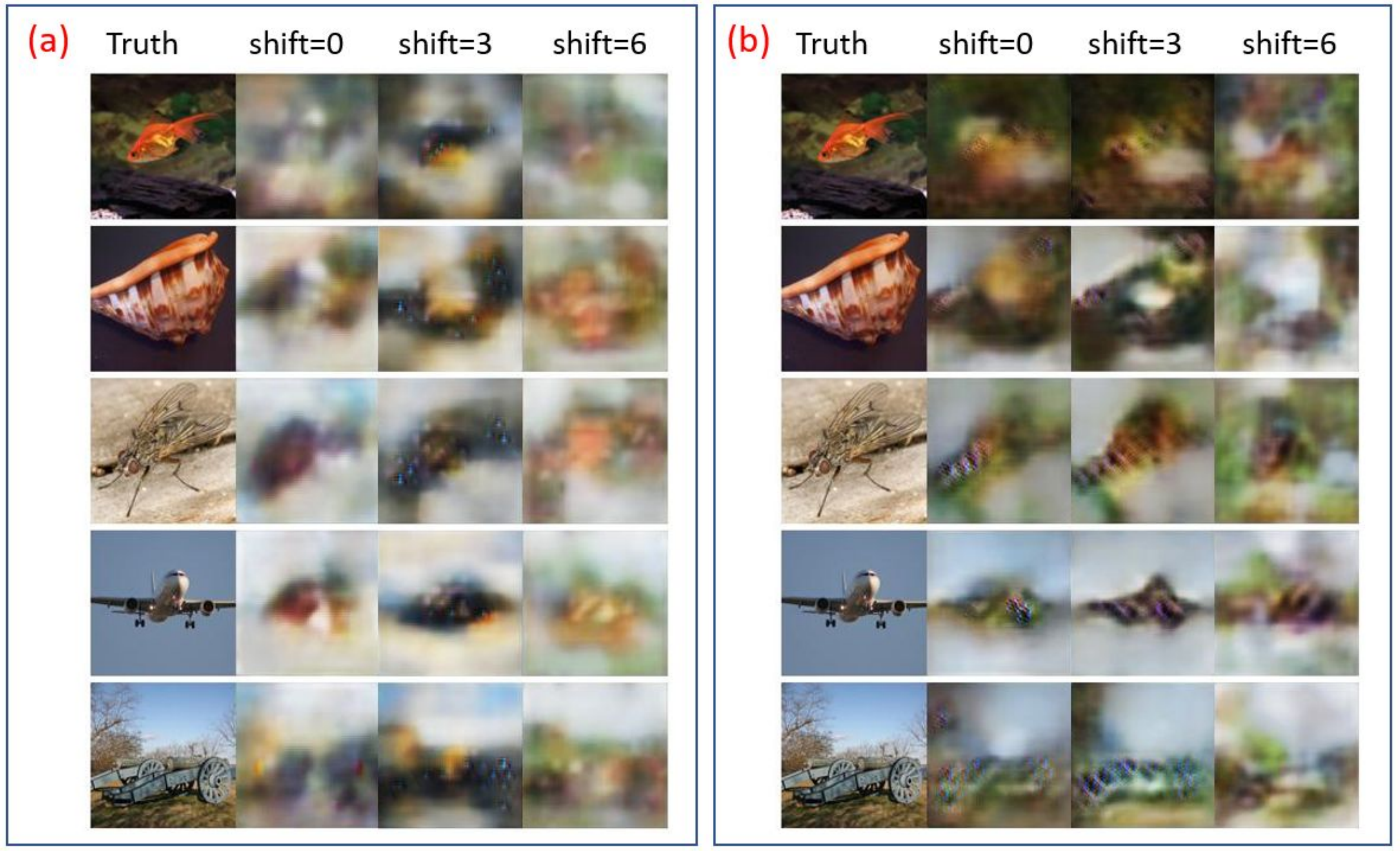
Results/Discussion

Replicating original preprocessing pipeline



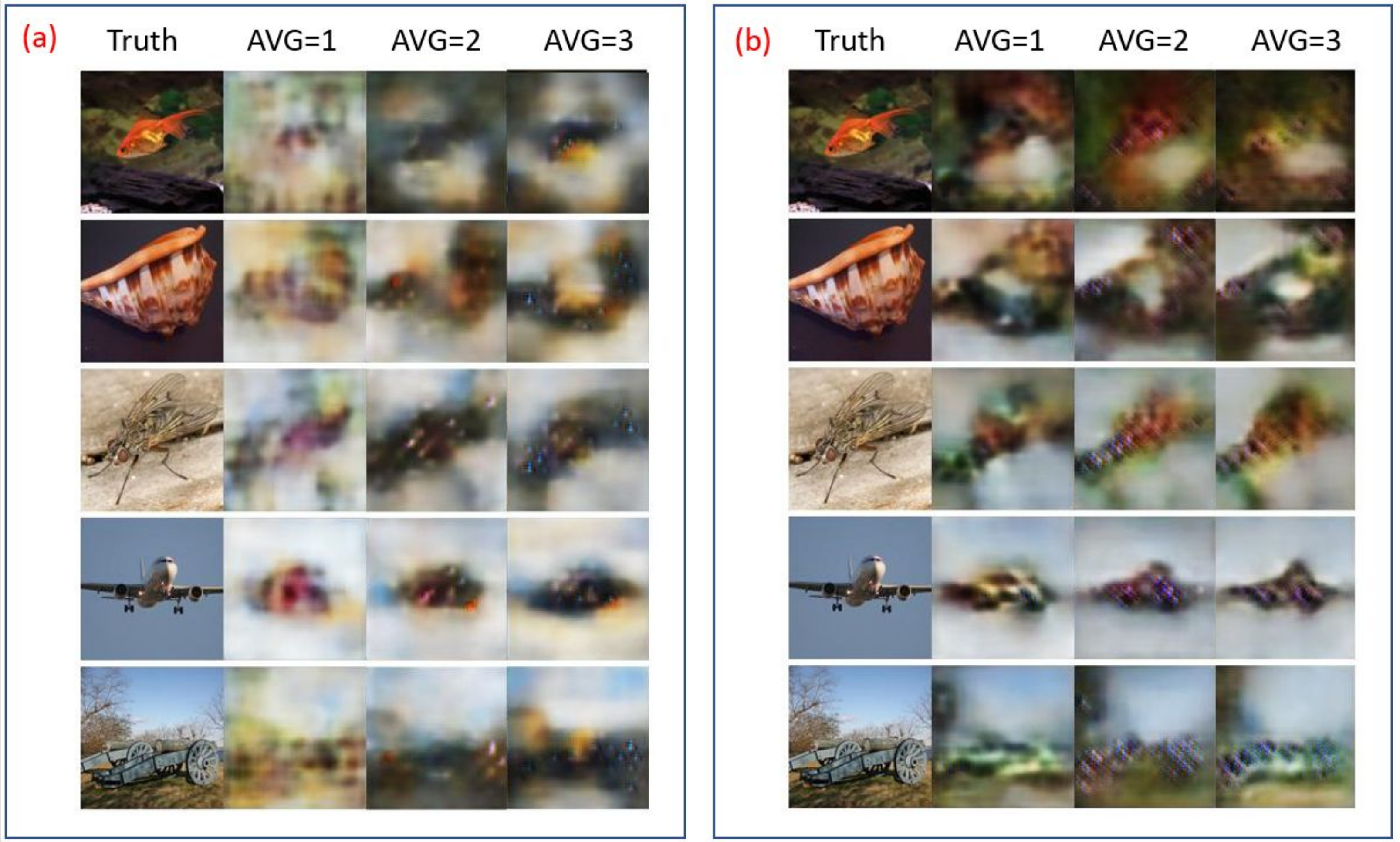
Test-set reconstructions from Subject 3.

Temporal modification: Varying the haemodynamic shift parameters



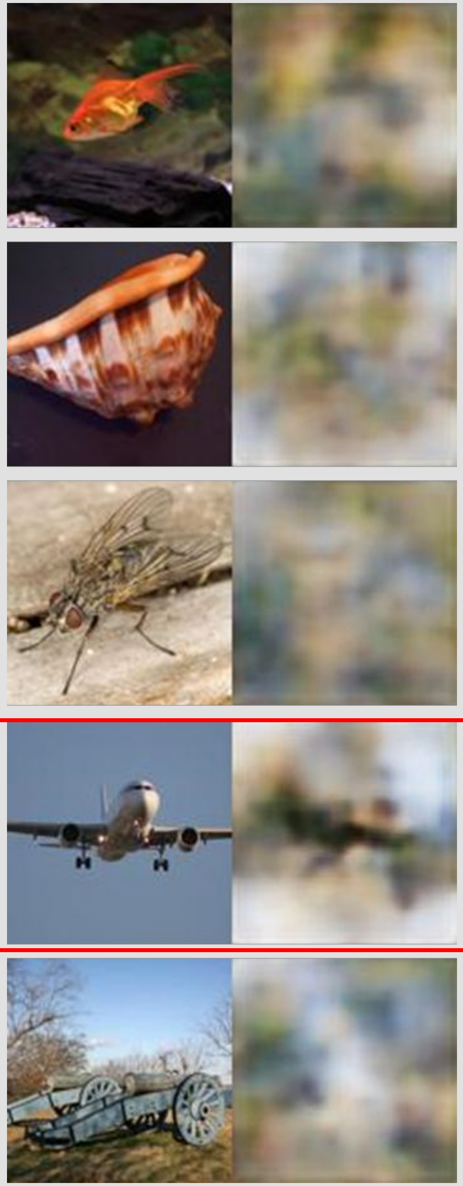
(a) Reconstruction from Subject 2's visual stimuli. (b) Reconstruction from Subject 3's visual stimuli.

Temporal modification: Vary frames averaged in stimuli window

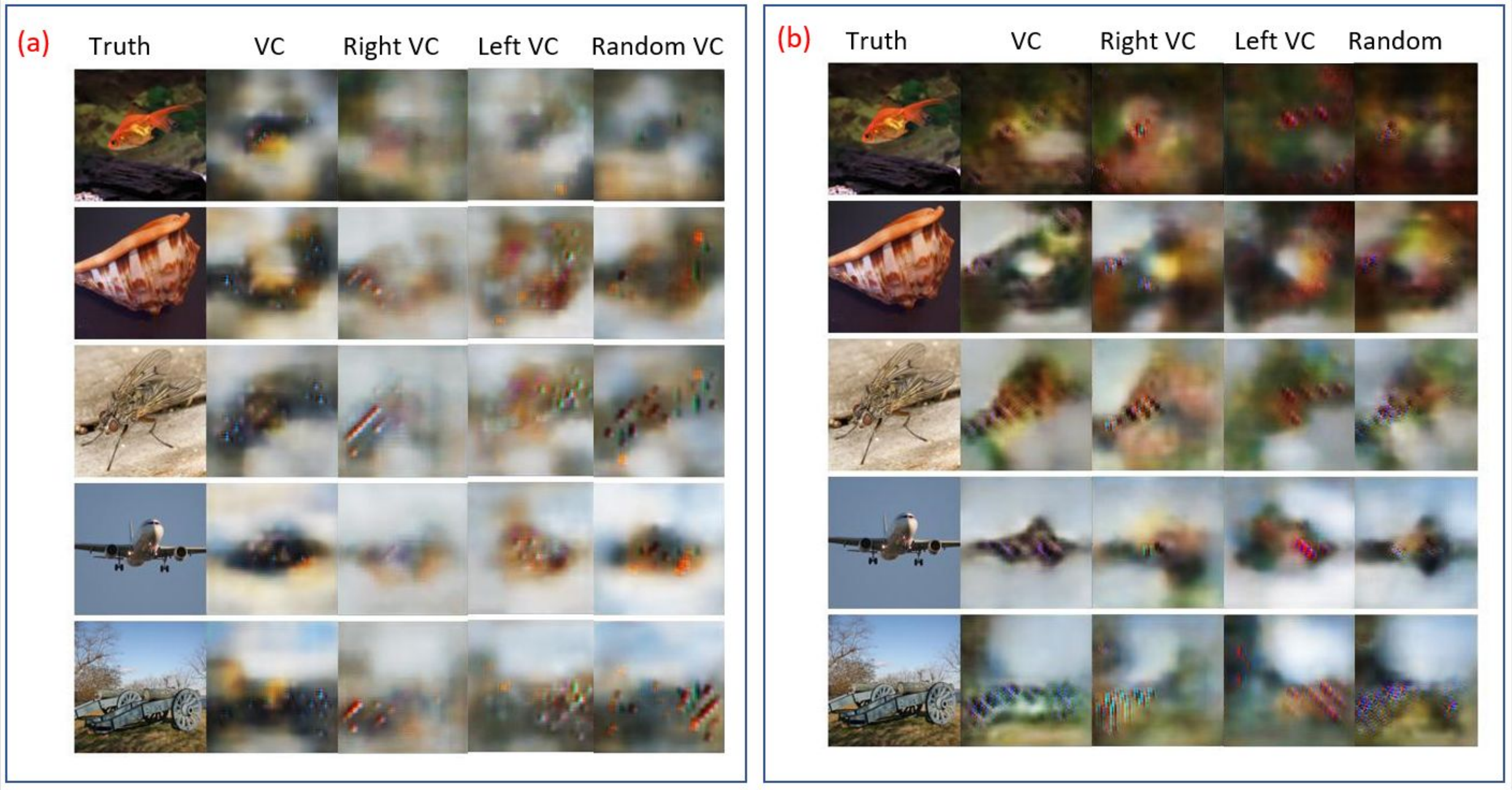


(a) Reconstruction from Subject 2's visual stimuli. (b) Reconstruction from Subject 3's visual stimuli.

Observations from avg 1 shift 0 parameters in subject 3

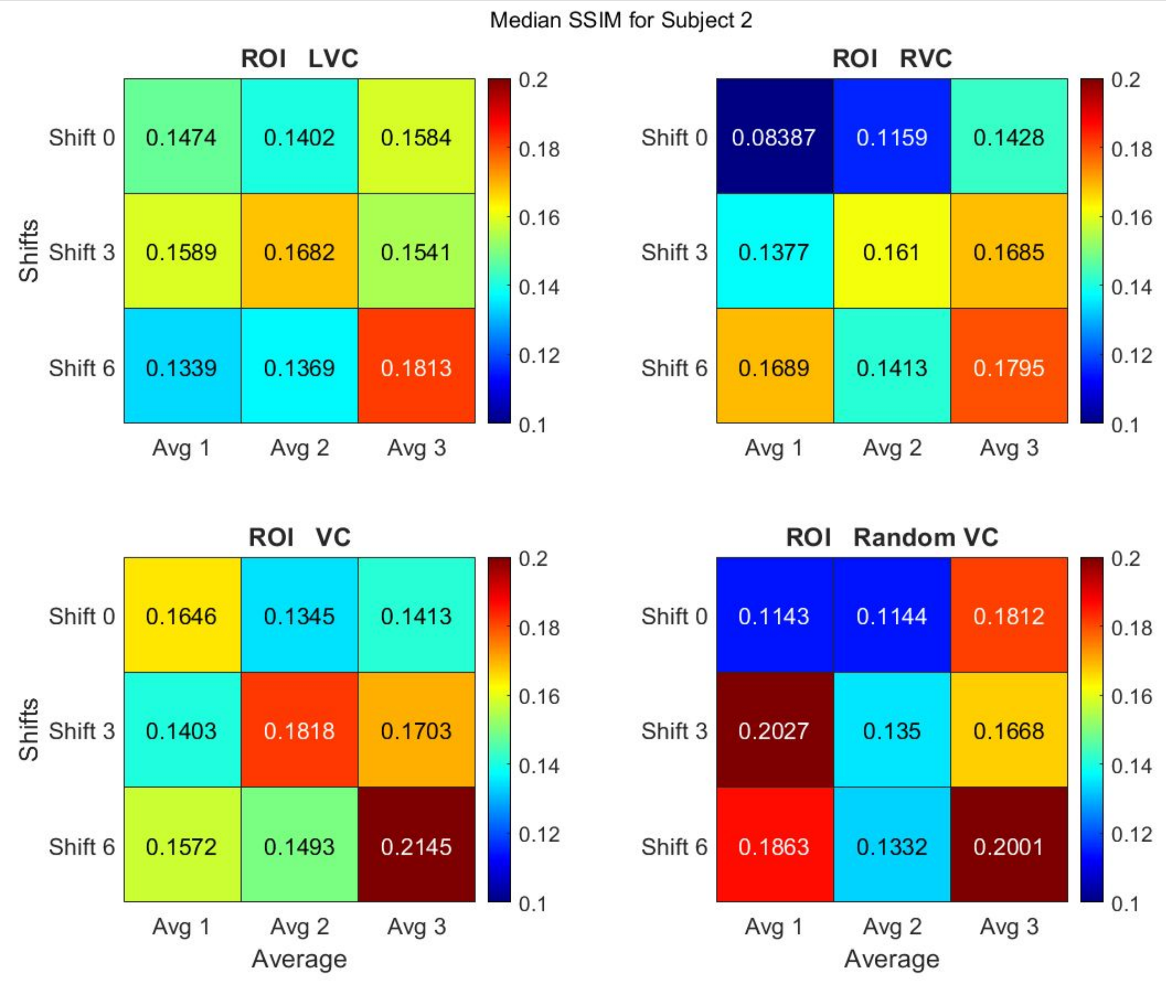


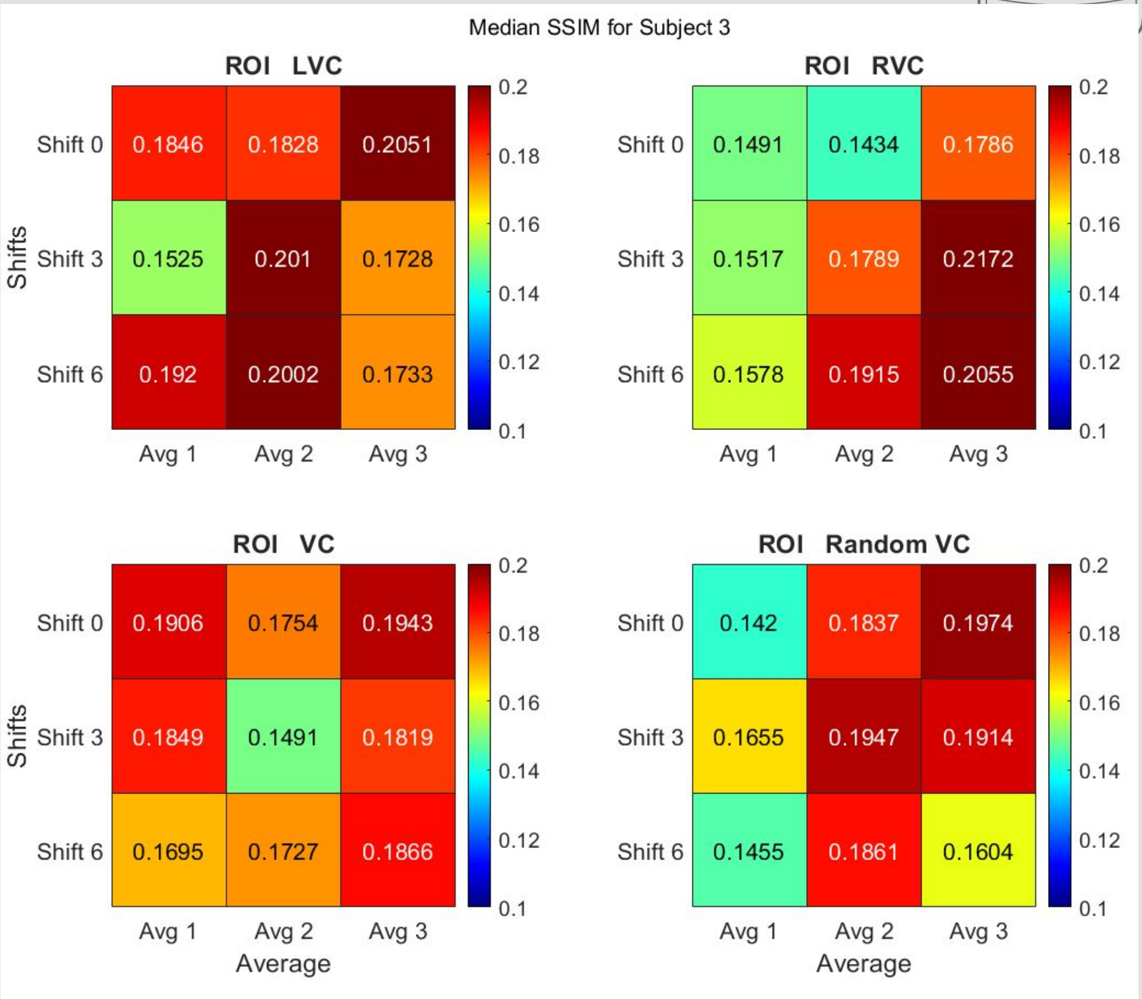
Spatial Modification: Four ROIs. Whole visual Cortex, left visual cortex, right visual cortex, random visual cortex.

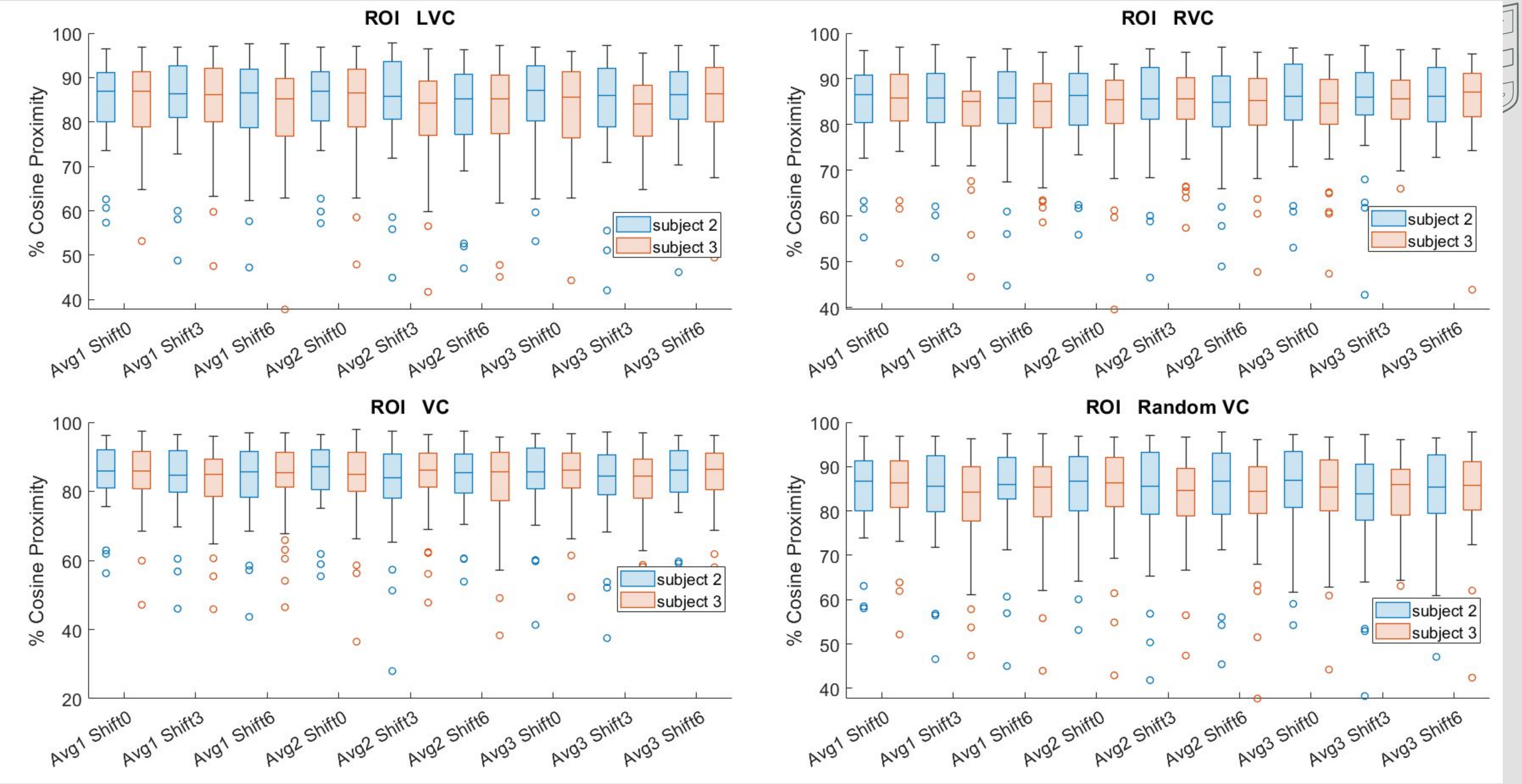


(a) Reconstruction from Subject 2's visual stimuli. (b) Reconstruction from Subject 3's visual stimuli.

Grid search over spatial and temporal modifications







ROI LVC
ROI RVC
ROI VC
ROI Random VC
% Cosine Proximity
subject 2
subject 3
Avg1 Shift0
Avg1 Shift3
Avg1 Shift6
Avg2 Shift0
Avg2 Shift3
Avg2 Shift6
Avg3 Shift0
Avg3 Shift3
Avg3 Shift6

Limitation and future work

- ❑ A key limiting factor for this task is small-size data. Although some work tried to address this challenge by using pretrained model on external images, a large-scale imaging data is still required.
- ❑ We only used two subjects and plan to use all 5 subjects.
- ❑ The variability between preprocessing methods isn’t quantified statistically with multiple comparisons
- ❑ Fail to generate significant increases in performance.

Next steps:

- ❑ We plan to use more ROIs.
- ❑ We repeat the process but with different models.

References

T. Horikawa and Y. Kamitani, “Generic decoding of seen and imagined objects using hierarchical visual features,” Nature communications, vol. 8, no. 1, pp. 1–15, 2017.

R. Beliy, G. Gaziv, A. Hoogi, F. Strappini, T. Golan,and M. Irani, “From voxels to pixels and back: Self-supervision in natural-image reconstruction from fmri,”Advances in Neural Information Processing Systems, vol. 32, 2019.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in 2009 IEEE conference on computer vision and pattern recognition, Ieee, 2009, pp. 248–255

S. Huang, W. Shao, M.-L. Wang, and D.-Q. Zhang,“Fmri-based decoding of visual information from human brain activity: A brief review,”
International Journal of Automation and Computing, vol. 18, no. 2, pp. 170–184,2021