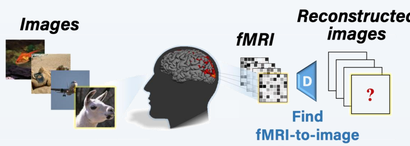
Self-Supervised Natural Scene Reconstruction and Rich Semantic Classification from Brain Activity

Guy Gaziv[1,*], Roman Beliy[1,*], Niv Granot[1,*], Assaf Hoogi[1], Francesca Strappini[2], Tal Golan[3], Michal Irani [1]

1. Dept. of Computer Science and Applied Math, The Weizmann Institute of Science
2. Dept. of Neurobiology, The Weizmann Institute of Science
3. Zuckerman Institute, Columbia U.

*Equal contributors
Corresponding author: michal.irani@weizmann.ac.il

Goal: Reconstruct images and decode semantic categories from fMRI recordings



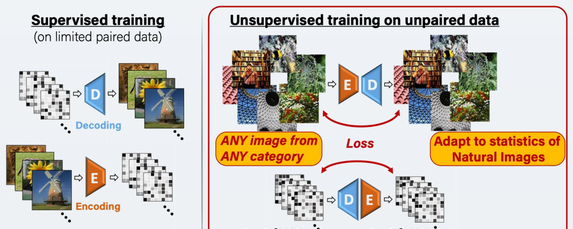
Problems:

- Insufficient training data (~1200 pairs from 150 semantic classes)
  - ➔ does not span natural images and their rich semantic space
- Differs between brains ➔ cannot be combined across subjects

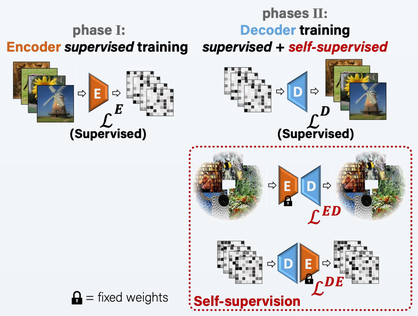
We propose:

Training on many additional *"unpaired" data:*

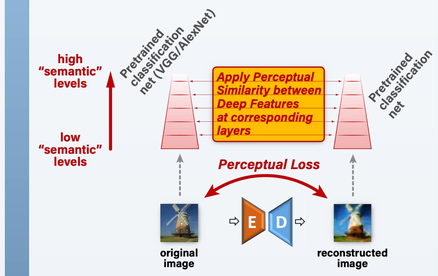
(i) Images without fMRI recordings ➔ Adapts to statistics of *Natural Images*
(ii) fMRI recordings without images ➔ Adapts to statistics of *input test data*



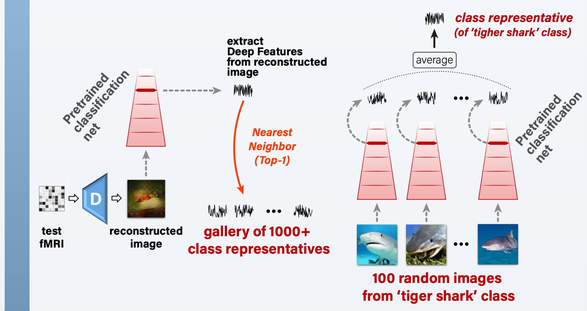
Two training phases:



Adding perceptual (semantic) constraints



Classifying against a gallery of class representatives

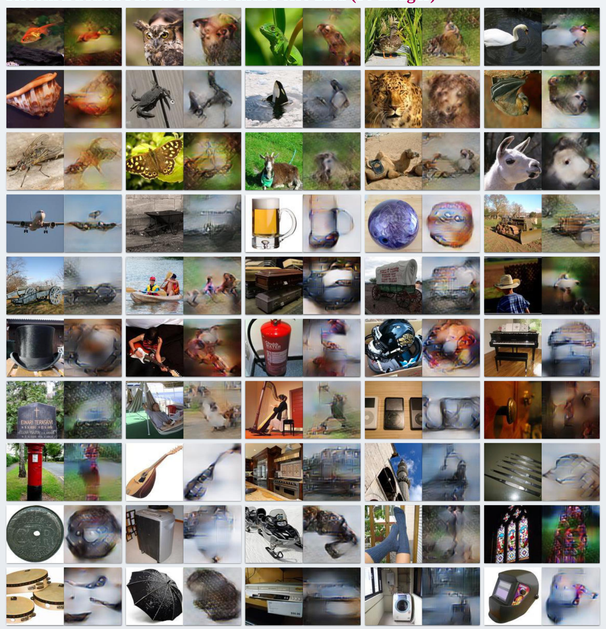


E-D: trained on unpaired images *(with no fMRI)*
(50,000 natural images from 1000 rich ImageNet classes *without their class labels*)
➔ *Adapts to the statistics of natural images and novel semantic classes*

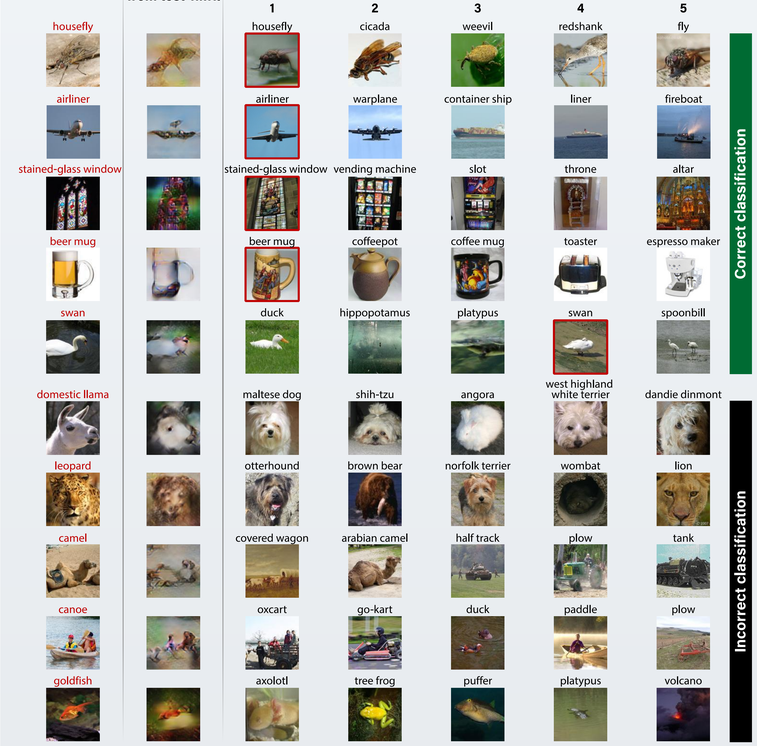
D-E: trained on unpaired fMRI *(unknown image)*
(Specifically, the unpaired *test-fMRI*, without any corresponding images)
➔ *Adapts to the statistics of the new (unpaired) test-data*

Our Results (for 'fMRI on ImageNet' [Horikawa et al. 2017]):

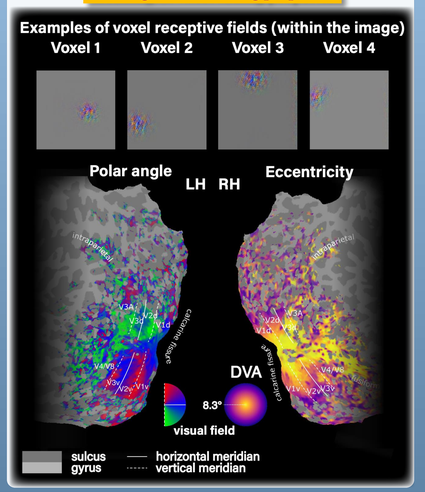
Reconstruction results for the entire test data (50 images)



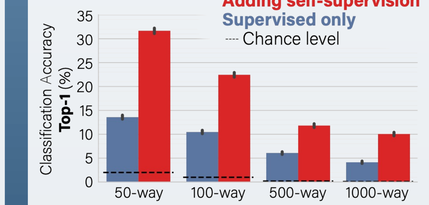
Semantic classification results – predicting class label out of 1000+ classes

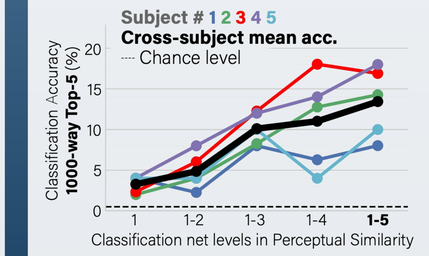


Our models capture biologically meaningful voxel tuning properties



Self-supervision with Perceptual Similarity allows classification to rich and novel semantic categories





Adding self-supervision on *unpaired data* improves reconstruction