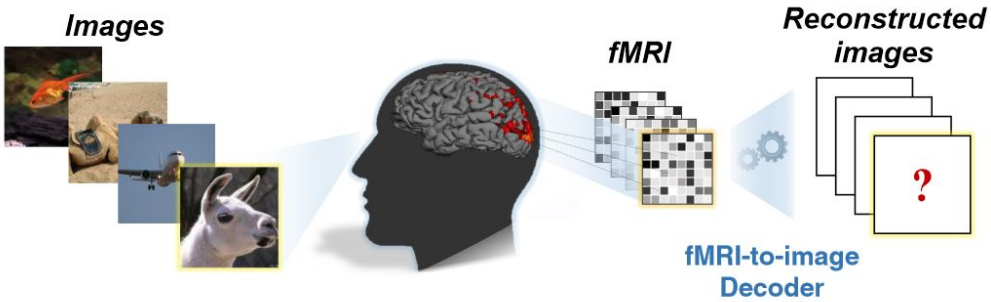
Self Supervised fMRI->Image Decoder

Marc Laugharn

Problem statement

Given fMRI scans of the visual cortex taken while looking at pictures, can we reconstruct what the subject was seeing?



Attempted re-implementation of 2019 paper "From voxels to pixels and back: Self-supervision in natural-image reconstruction from fMRI"

Dataset

(fMRI scan, image) pairs

Training set:

- 1200 images and fMRI scans (3D voxel values, represented as single 4643-d vector)

Test set:

- 50 images
- 35 * 50 fMRI scans, averaged

5 subjects, just limited to 1 subject (bc diff subjects had diff voxel arrangements/dimensions)

Approach

Problem: very small amount of data

Solution: learn multiple mappings and insert a lot more extraneous data:

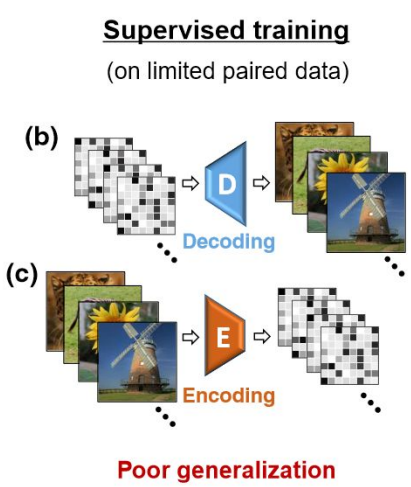
Decoder D: fMRI -> Image

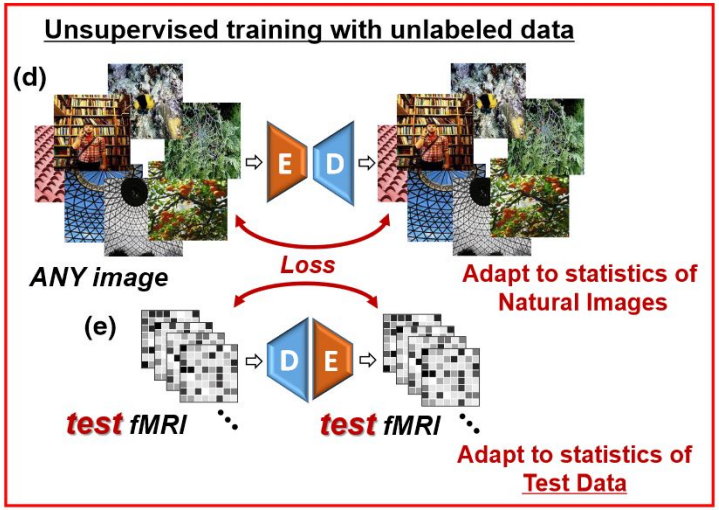
Encoder E: Image -> fMRI

DE: Image -> Image

ED: fMRI -> fMRI

Can use any dataset for these, just have the objective that you get out whatever you put in!





Training plan

1. Train the encoder E first
2. Freeze E and then train D by optimizing D, DE, and ED simultaneously

Encoder: image -> MRI

- Based on a convolutional neural network
- ResNet fine-tuned model + custom model head
  - ResNet originally has a layer at the end to pick a target class
  - Remove this last layer, and replace it with our own custom model ‘head’ that outputs a vector of voxel values
  - Freeze the image recognizing weights of ResNet and only train our model ‘head’

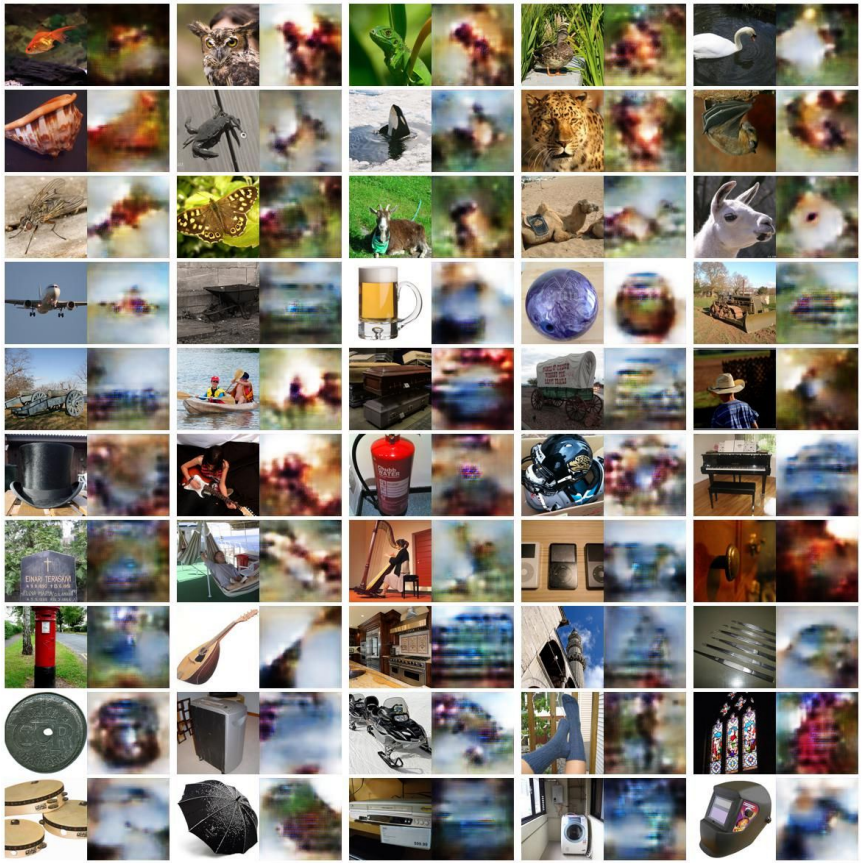
Decoder: MRI -> Image

Fully connected neural network layers + series of deconvolutions and upscaling layers

How to train D, DE, and ED at the same time

- 60% of the time, objective is D: fMRI -> Image
- 30% of the time, objective is ED: Image -> Image
  - 50% of the time, use training set images
  - 50% of the time, sample from 50000 imagenet images
    - Supposed to teach statistics of natural images
    - How to make do with limited dataset
- 10% of the time, objective is DE: fMRI -> fMRI
  - Authors suggest to allow model to use test fMRIs for this objective
  - Apparently OK since it’s not the main objective of learning D
  - Couldn’t get this to improve training

The paper’s results..



My results..

https://github.com/mlaugharn/braincoder

Blurrier, but some decoded images quite match the input; especially black objects on white background

Other things I tried

- Voxel adjacency matrix weighting
- Graph neural networks
- Perceptual loss criteria