The Hierarchy Architecture of the Visual Cortex

By: Ying Zhou, Naeun Oh, Rathi K
Pod : Complicated-Chihuahuas | Team : Brownies A La Mode
Thanks to : Sthithpragya Gupta (Pod TA), Hamed Karimi (Project TA), Mingbo Cai (Mentor)

neuromatch
academy

Introduction

Visual cortex

: Anatomical and functional hierarchy structure

: Different regions process different information.

Deep neural network

: Recognize images with high accuracy.

: Different layers extract different information, from basic feature to object recognition.
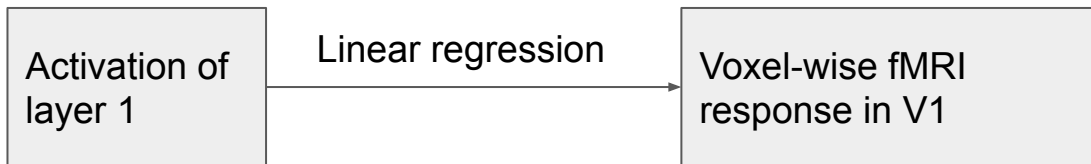
Does visual cortex encode images in a way that DNN model does?

If so, what does this imply about the hierarchy of the visual cortex?

Hypotheses

Does visual cortex encode images in a way that DNN model does?

Hypothesis 1.1: fMRI responses in visual cortex can be predicted by the output of AlexNet model

| Activation of layer 1 | Linear regression → | Voxel-wise fMRI response in V1 |

Hypothesis 1.2 The dissimilarity matrices of fMRI responses and model layers between different images have similar structures.

Hypotheses

Does visual cortex encode images in a way that DNN model does? ✔

What does this imply about the hierarchy of the visual cortex?

The AlexNet model extract features from low to high levels

Hypothesis 2.1: Low hierarchical regions represent spatial information as the initial layers of AlexNet.

Hypothesis 2.2: Higher hierarchical regions represent categorical information as the deeper layers of AlexNet.

Dataset : Kay et al.(2008) fMRI Dataset

1. Stimuli

1750 images (train) + 120 images (test)

Normalized grayscale, Background cropped

Labeled via DNN trained on imagenet

Passively viewed while fixating

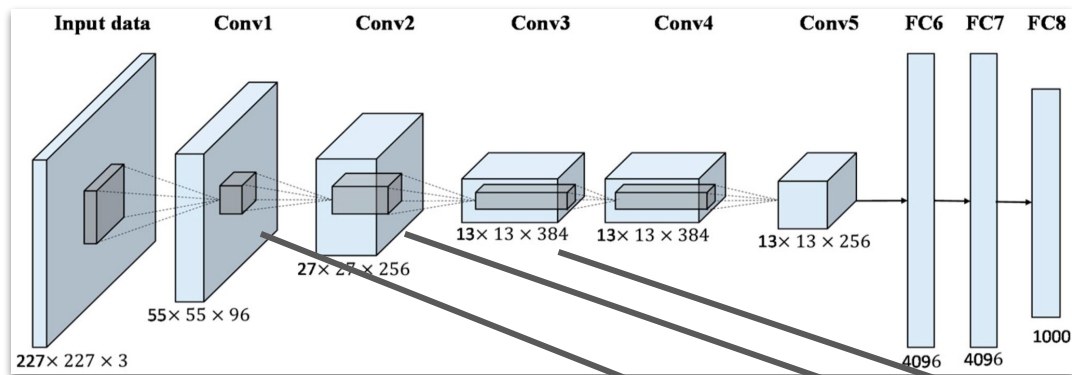**2. fMRI response**

8480 voxels (2x2x2.5mm)

Average of 13 trials, preprocessed, normalized
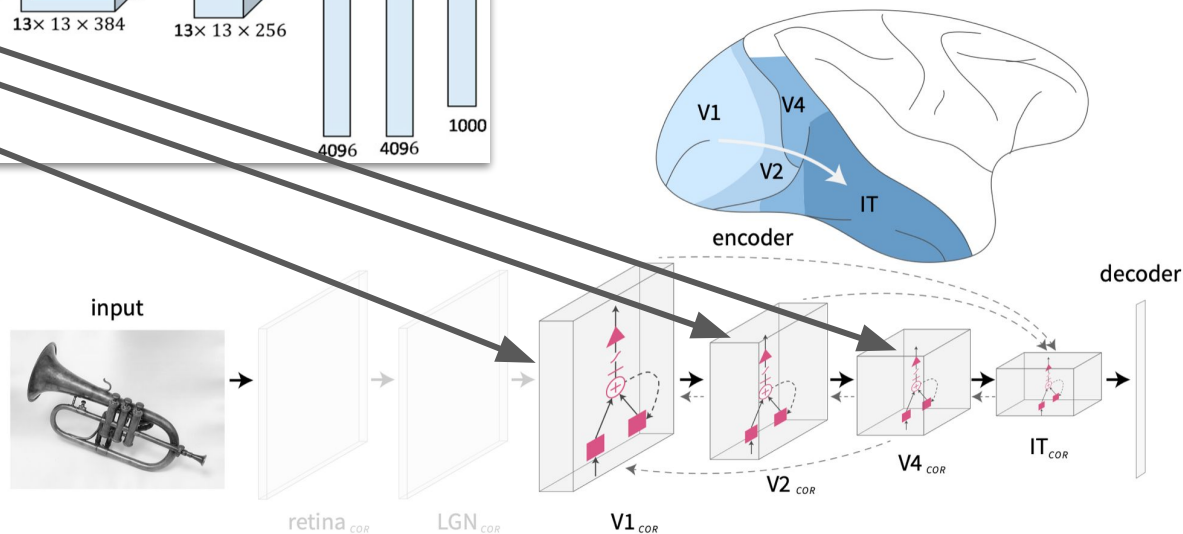
7 visual cortex ROIs - V1, V2, V3, V3A, V3B, V4, LatOcc

croquet ball

bubble

komondor

tailed frog

Pomeranian

starfish

impala

African elephant
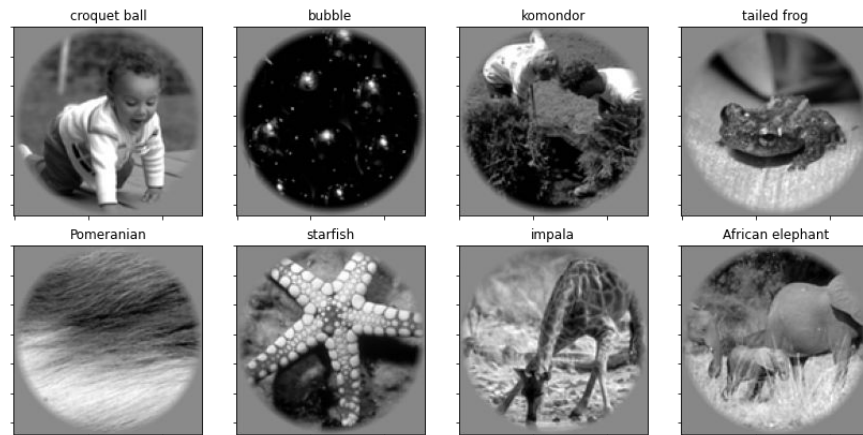
Hypothesis 1.1

fMRI responses in visual cortex can be hierarchically predicted by the layer output of AlexNet
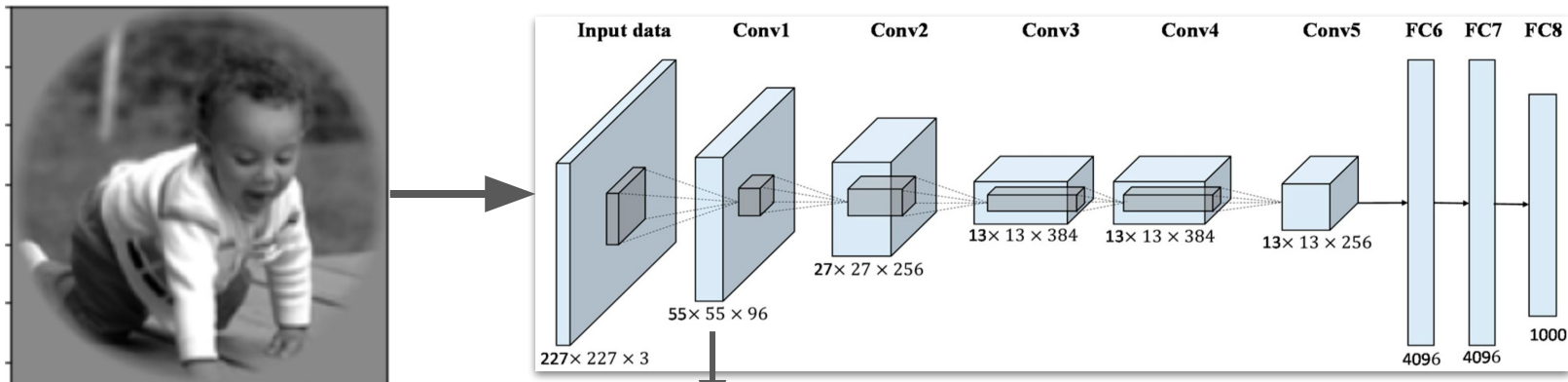

AlexNet Model

Hierarchical organization of visual system

Hypothesis 1.1

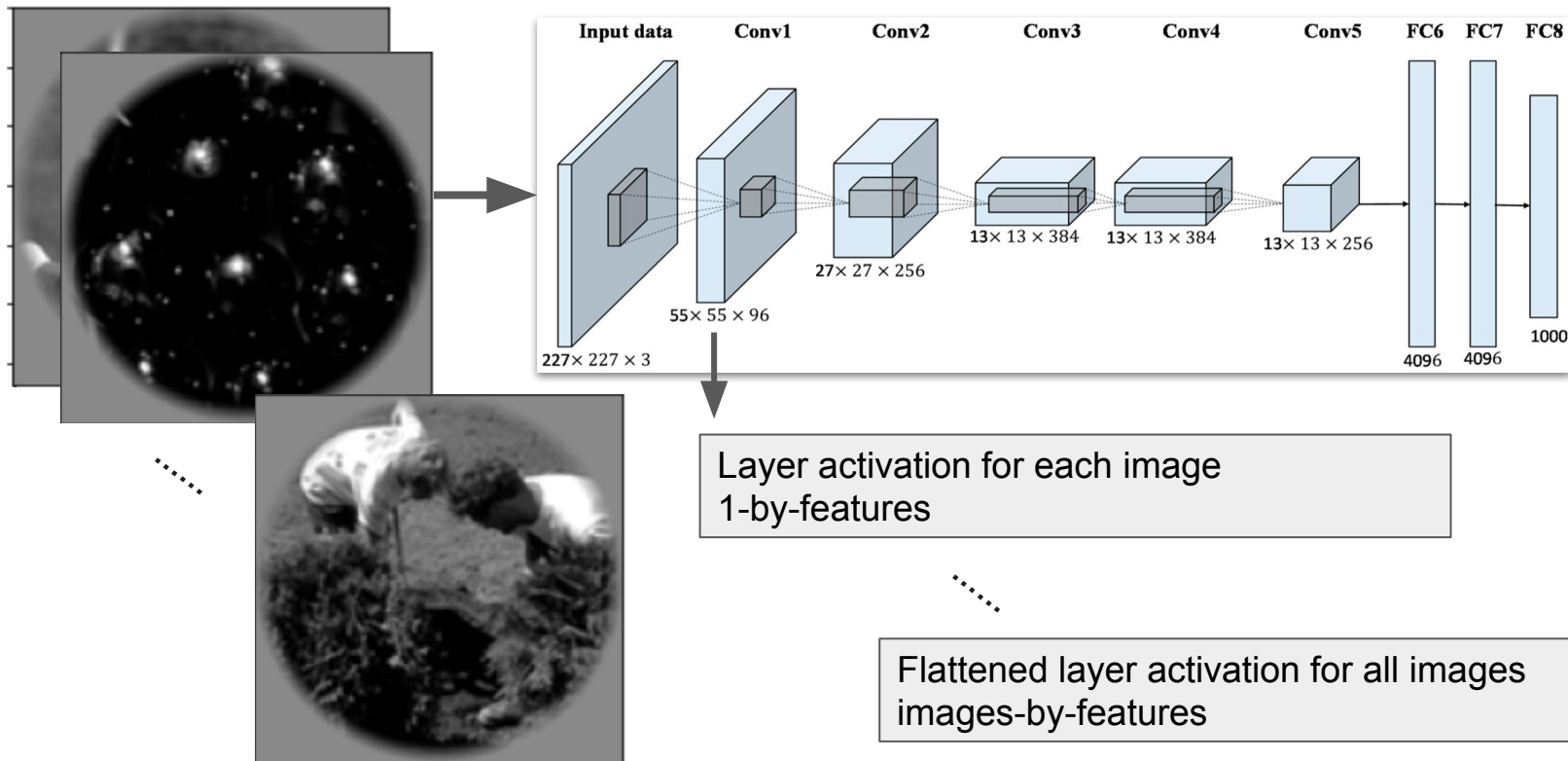**Extract each layer's output of a pre-trained AlexNet for each image**

Layer activation for each image
1-by-features
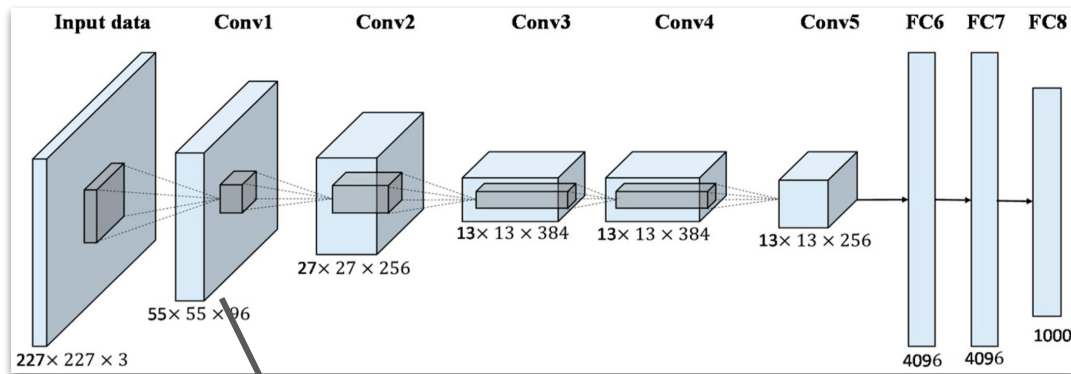
5 convolutional layers + 3 fully connected layers

Hypothesis 1

extract the each layer's output of a pre-trained AlexNet for all the images

Layer activation for each image
1-by-features

Flattened layer activation for all images
images-by-features

Hypothesis 1.1

Train voxel-wise encoding models to predict fMRI responses

True response in V1

Input data Conv1 Conv2 Conv3 Conv4 Conv5 FC6 FC7 FC8

13× 13 × 384   13× 13 × 384

13× 13 × 256

27× 27 × 256

55× 55 × 96

227× 227 × 3

1000

4096   4096

MSE

Prediction in V1

Layer activation → Linear Regression (+LASSO) → ① ② ③ ④ ••• Ⓝ Voxel number → Predict

Response

Voxel number
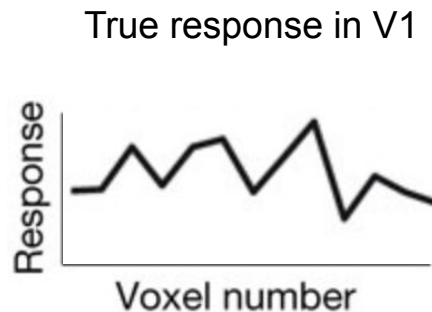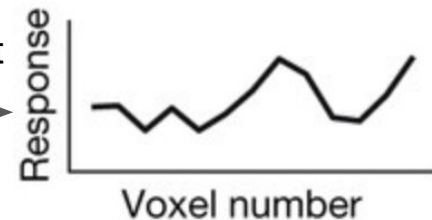
Hypothesis 1.1: Prediction Performance from layers to regions
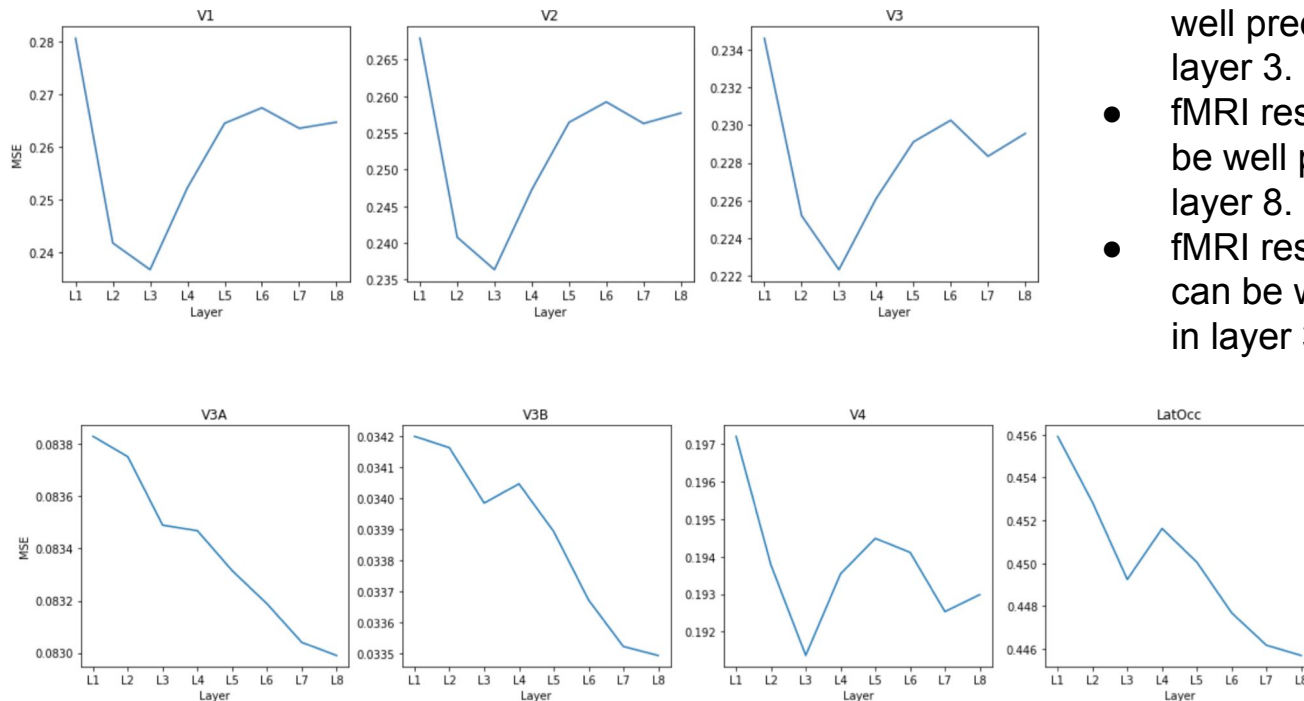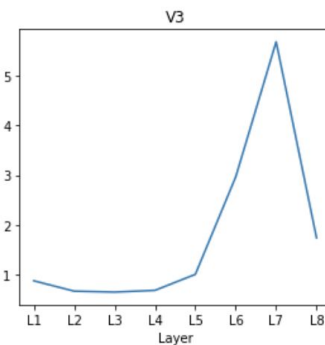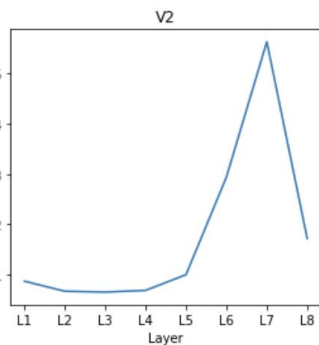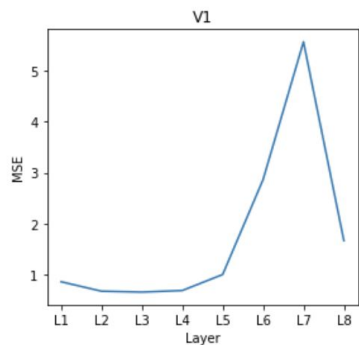

- fMRI response in V1-V3 can be well predicted by activations in layer 3.
- fMRI response in V3A and V3B can be well predicted by activations in layer 8.
- fMRI response in V4 and LatOcc can be well predicted by activations in layer 3 and 8.
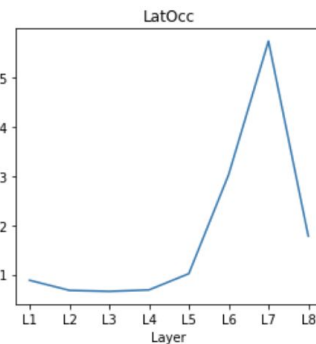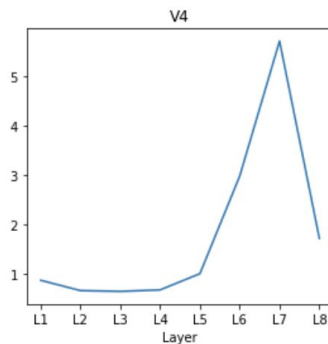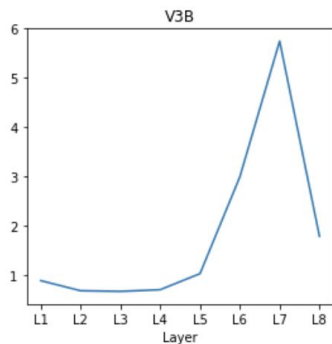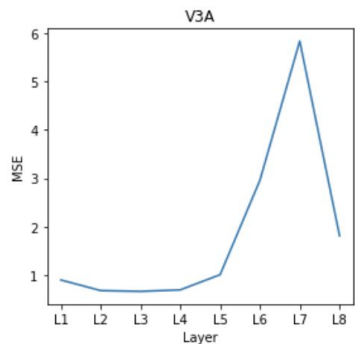
Hypothesis 1.1

- fMRI response in V3 is more predictable by AlexNet
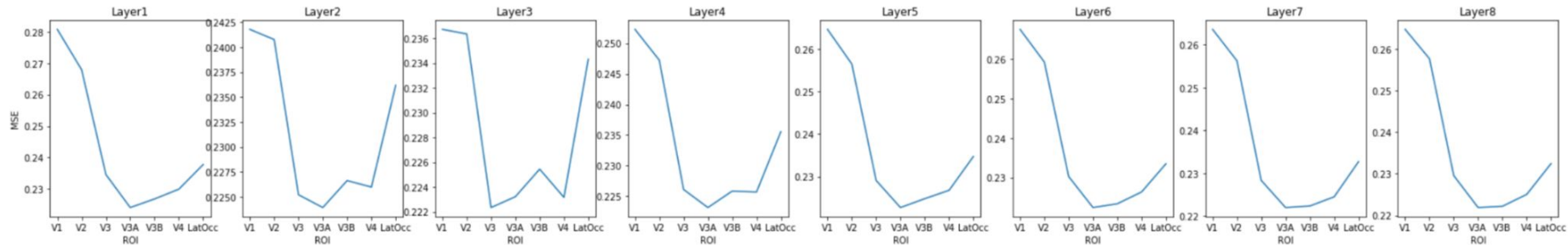
# Hypothesis 1.1: Prediction performance of un-trained AlexNet model

- Sanity check
- The CNN architecture alone cannot induce the similarity between DNN and visual cortex
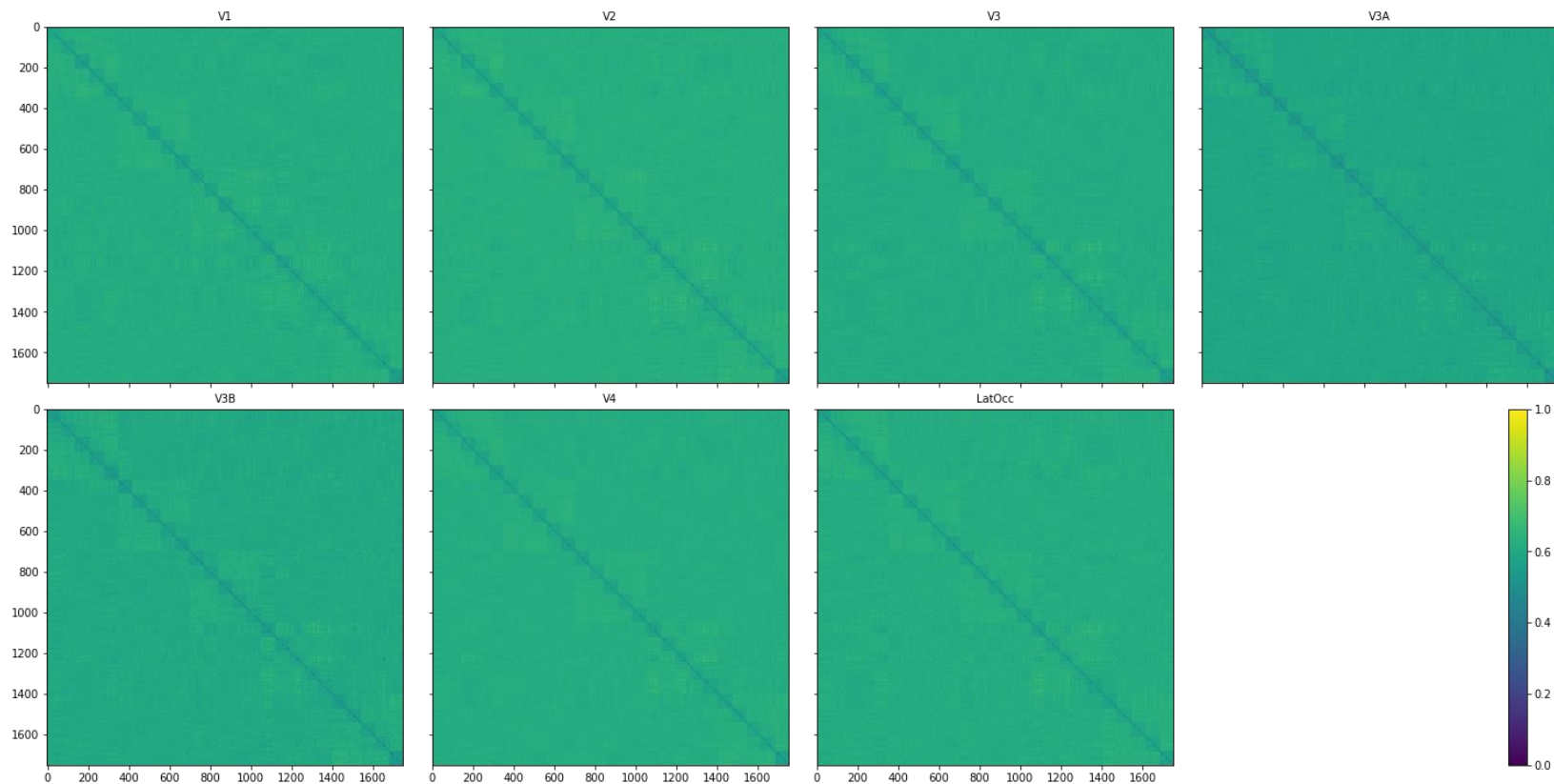
Hypothesis 1.2

Dissimilarity matrices of fMRI responses and model layers have similar representational structures.

1. RDM of neural responses of each brain region to images
2. RDM of features of each model layer to images
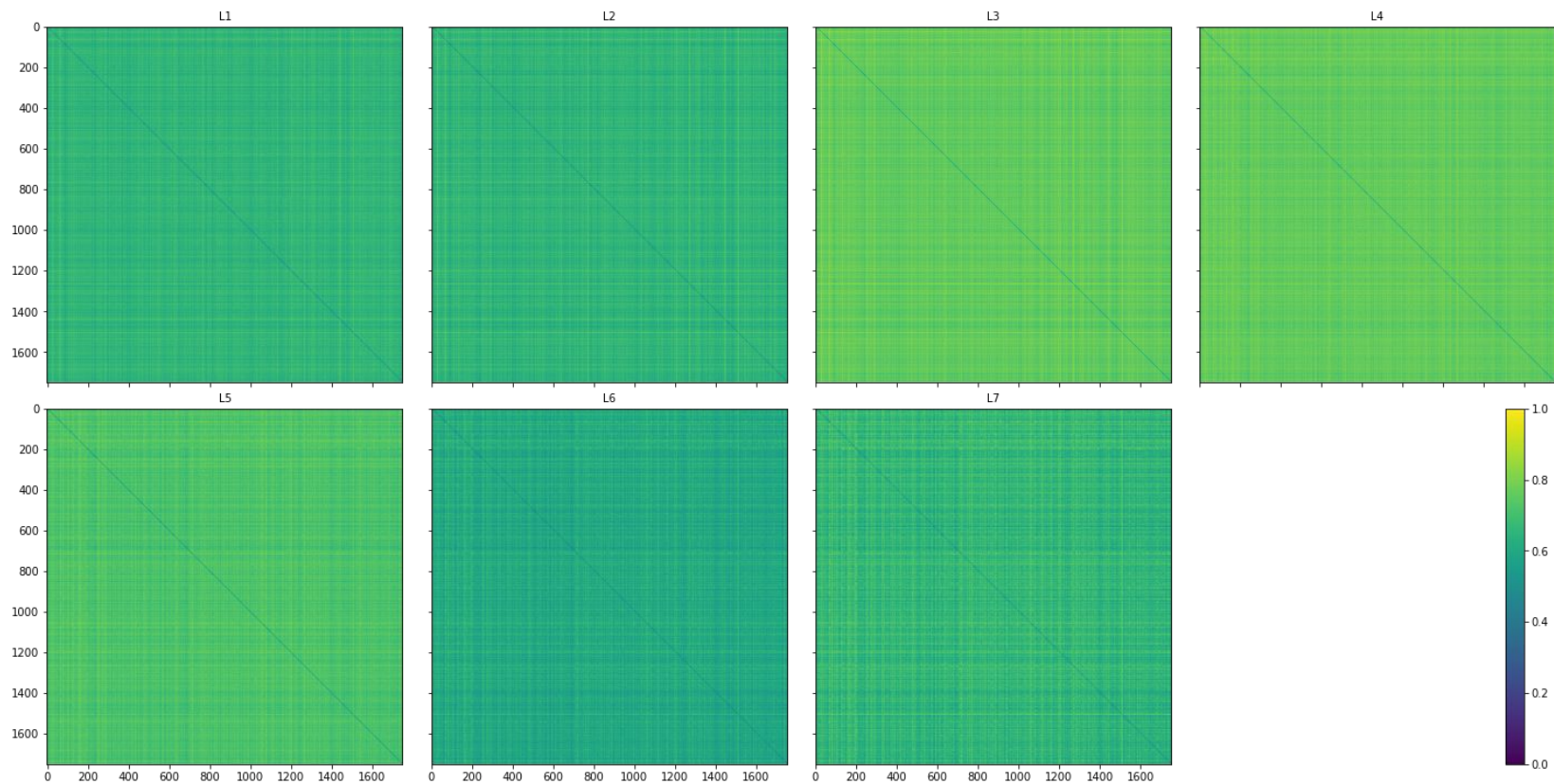3. Calculate similarities (Spearman's r) between RDM of a given layer and RDM of a region
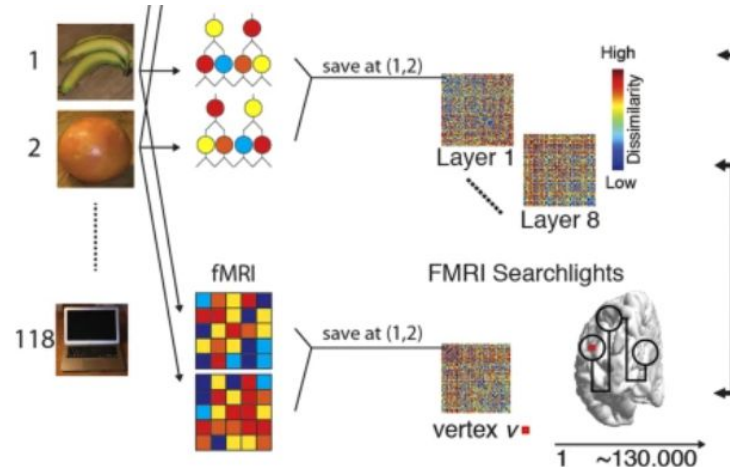
Cichy et al.(2016)

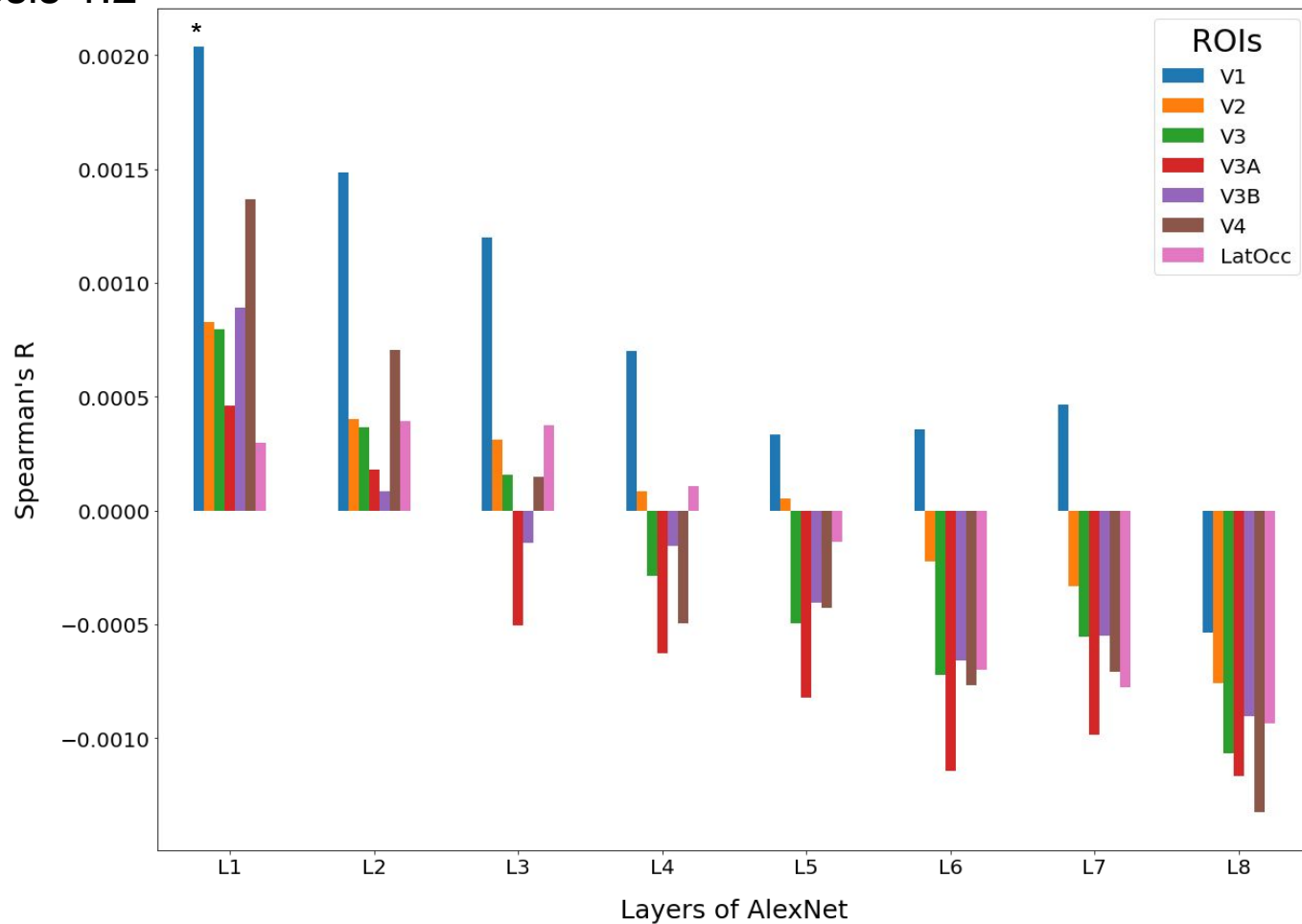# Hypothesis 1.2 : RDM of fMRI responses in ROIs

Hypothesis 1.2 : RDM of feature activations in layers

Hypothesis 1.2

Similarities between RDM of layers and RDM of regions

Discussion & Thoughts

Although our research questions have already been investigated in past research (Cichy et al., 2016; Mohsenzadeh et al. 2020), our results only partially support the hypothesis.

Most enjoyable yet difficult part of the process was trying to find the right model and methodology to test our research question.

Our preliminary work may be improved by testing with other DNN models, time-series fMRI responses, and categorical information of images (hypothesis 2).

Future Works with Hypothesis 2

Hypothesis 2.1: Low hierarchical regions represent low-level information as do the initial layers of AlexNet model (e.g., spatial information).

Method: To a model fine-tuned to our images, feed spatially transformed images (e.g. translation, scale, rotation, clutter) to the DNN model and observe how it changes the performance of layers predicting fMRI responses in each region.
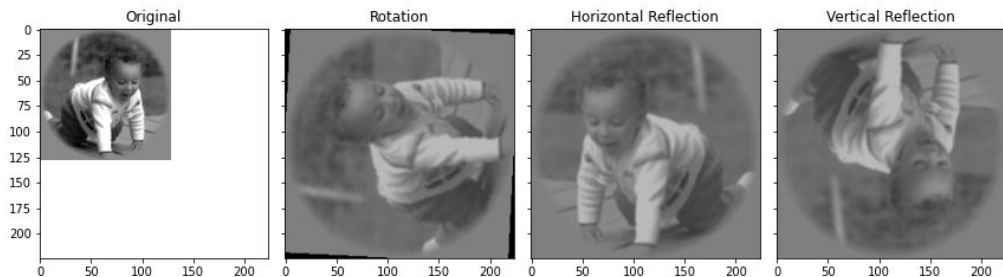
Hypothesis 2.2: Higher hierarchical regions represent high-level information as do the deeper layers of the DNN model (e.g., categorical information).

Method: Calculate representational dissimilarity matrices of fMRI responses to images sorted by categories in each brain region.