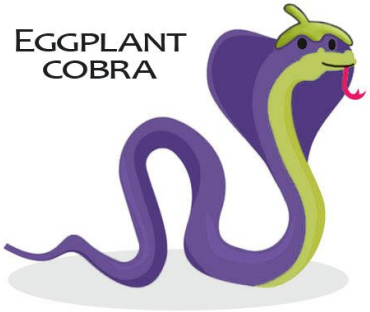


How do representations in the early visual hierarchy compare with the layers of a convolutional deep neural net?

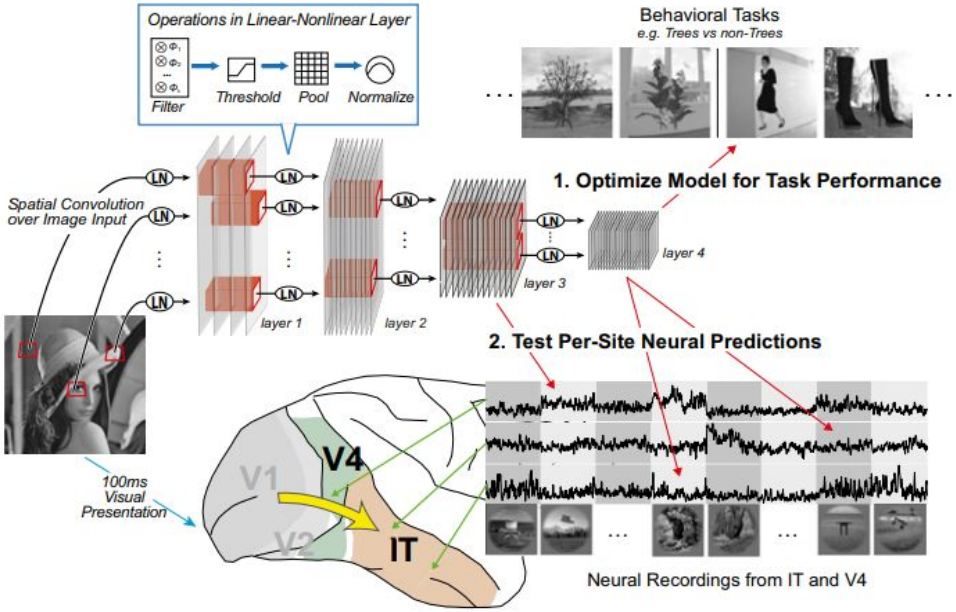
Pod: Eggplant Cobra

Mentor: Reza Abbasi-Asl (UCSF)

Theoretical Background

Visual system → Hierarchical organization

Convolutional Neural Networks (CNNs) → modeled after the structure of early visual cortex *(Lindsay et. al., 2020)*.



Yamins et al (2014)

How do representations in the ROIs of the early visual hierarchy compare with the layers of a convolutional deep neural net?

Hypothesis:
Higher layers in the visual hierarchy will be more representationally similar to higher layers of the deep neural net.

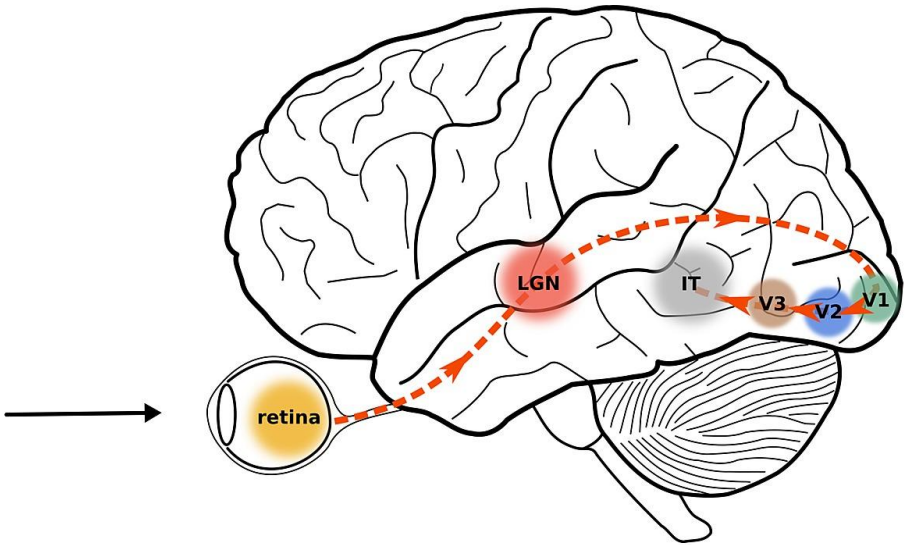
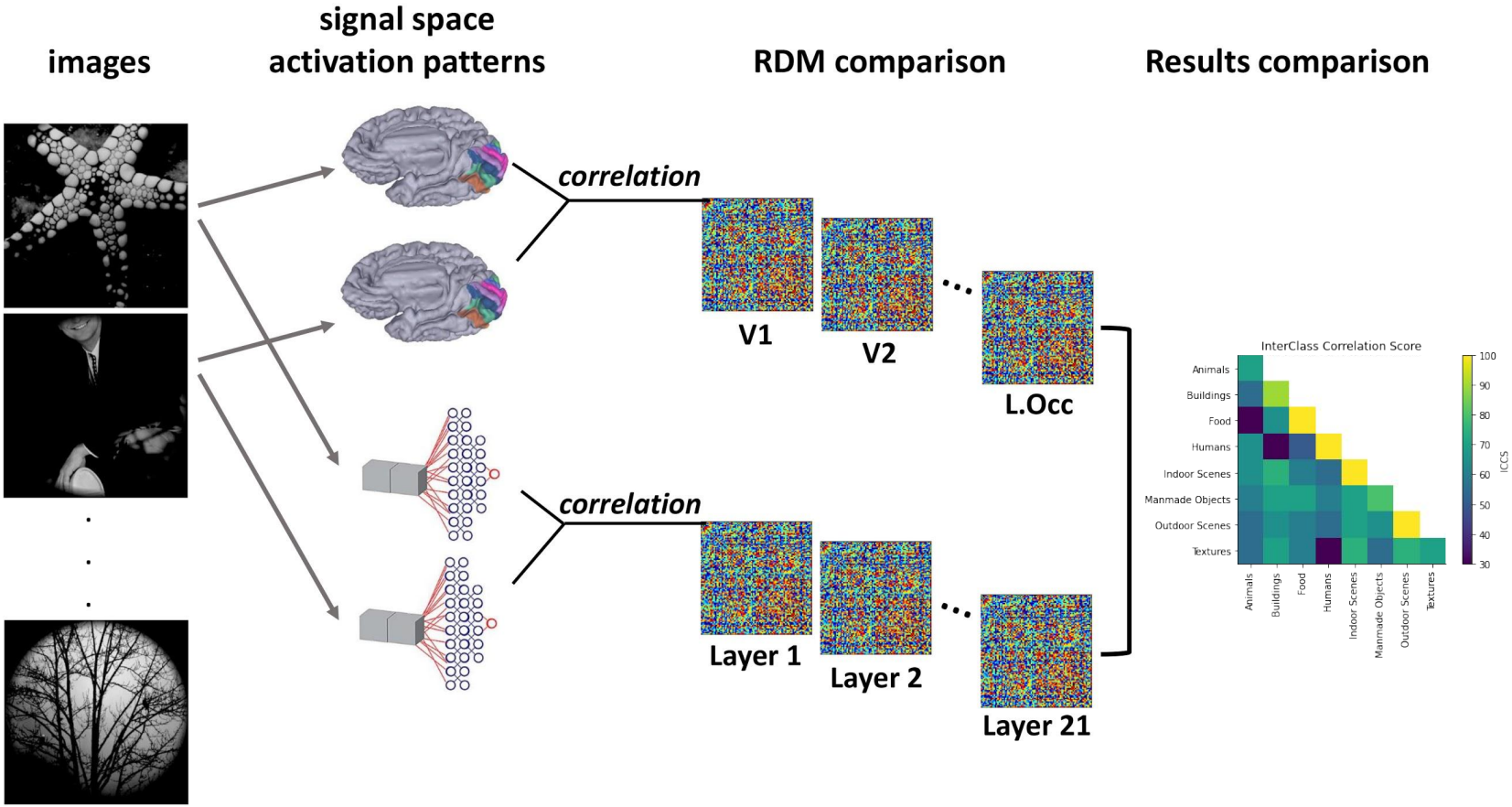


Figure: Figshare.com- Jonas Kubilius

Methods



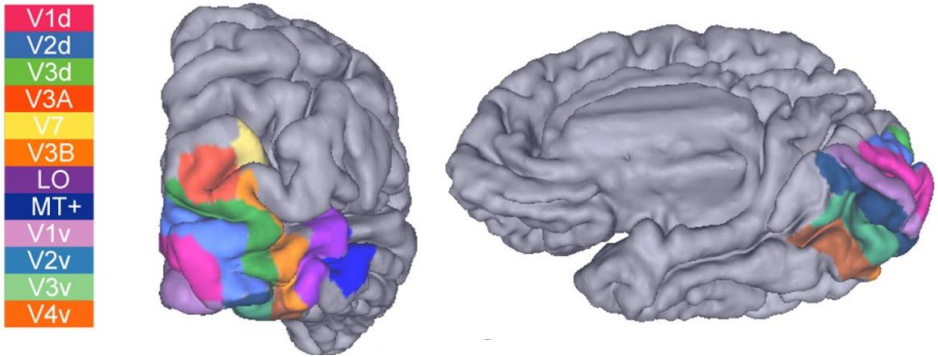
Kay et al (2008)

BOLD response patterns to naturalistic images

BOLD response amplitudes (z-scores)

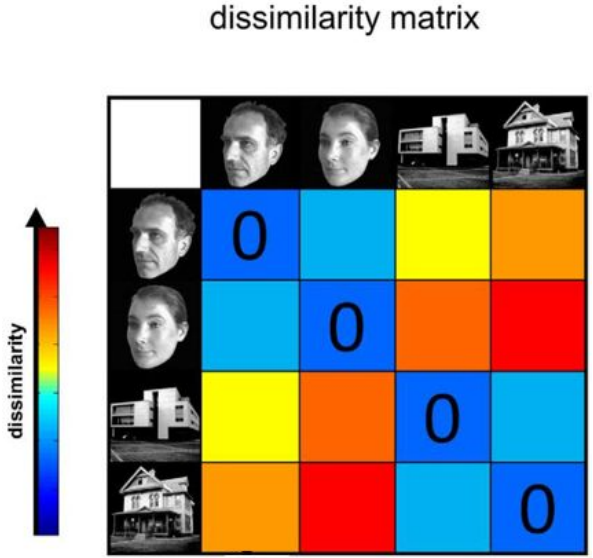
Regions of interest (ROIs):

V1, V2, V3, V3A, V3B, V4 & L.Occ.



Ban et al (2013)

Representational similarity analysis (RSA)



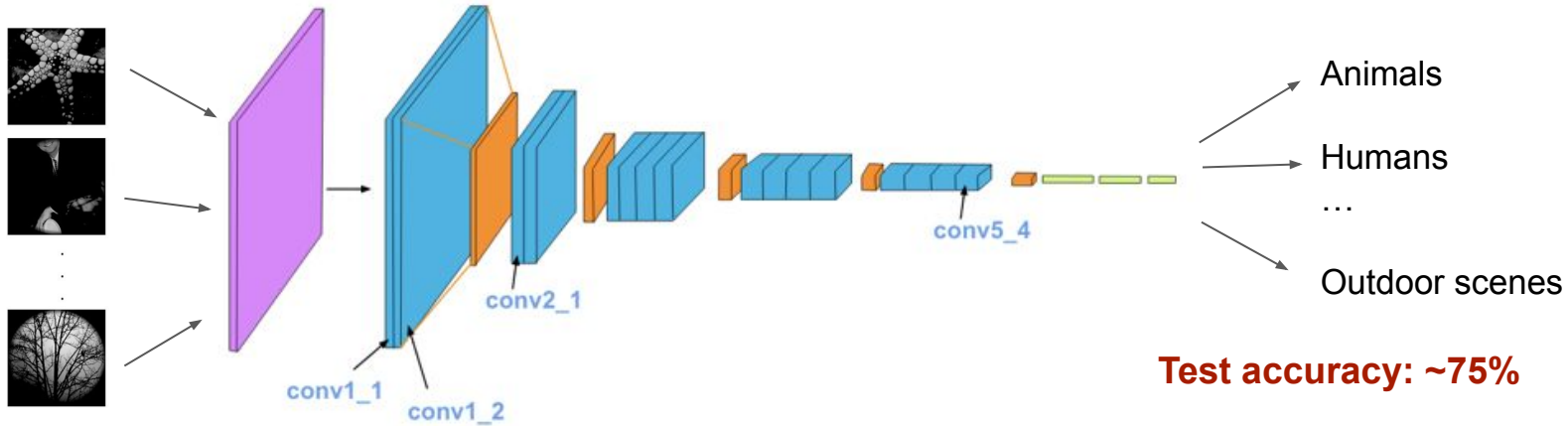
Kriegeskorte et al. (2008)

Neural networks

1) Label images with pretrained network

2) Build our own convolutional neural network to predict labels from images
   - Improve accuracy: add layers, change parameters (kernel size, batch size etc.), image augmentation
   - Accuracy ~ 40% (compared to 12.5% chance level)

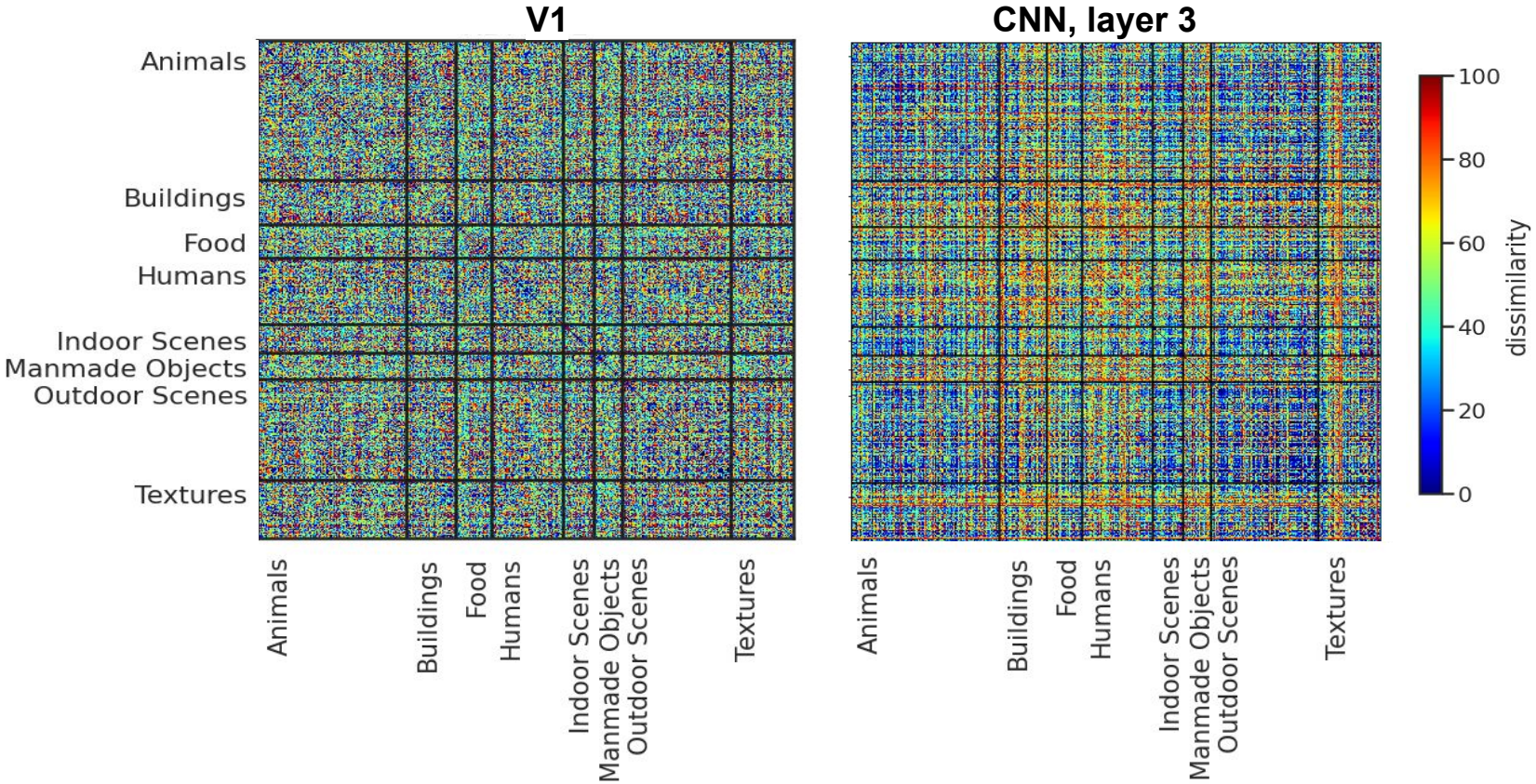
3) Use pretrained network (VGG-19) to predict labels from images with high accuracy:
   Pooling layers: 3, 6, 11, 16 and 21



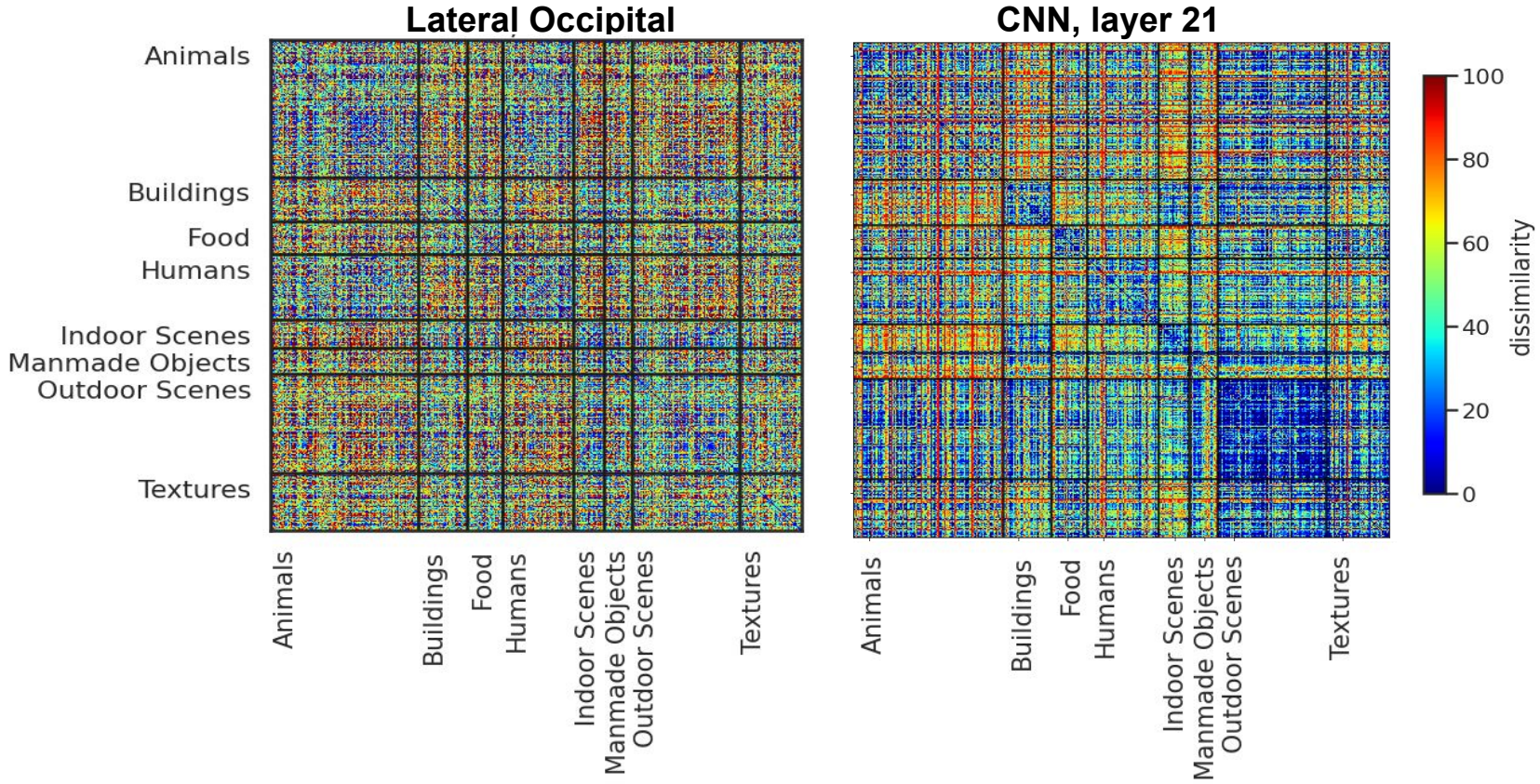
Test accuracy: ~75%

VGG-19 Figure: Medium.com

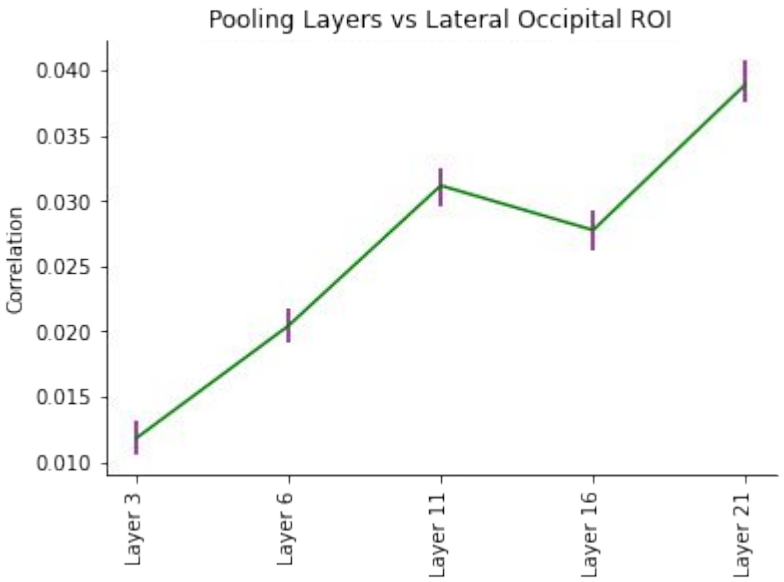
Results

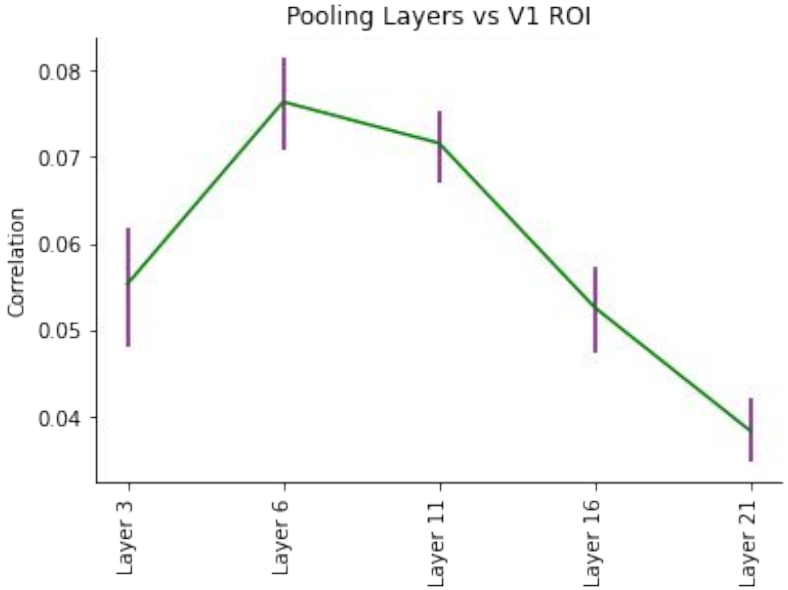

V1
CNN, layer 3
Animals
Buildings
Food
Humans
Indoor Scenes
Manmade Objects
Outdoor Scenes
Textures
dissimilarity
100
80
60
40
20
0

Results


Lateral Occipital
CNN, layer 21
Animals
Buildings
Food
Humans
Indoor Scenes
Manmade Objects
Outdoor Scenes
Textures
100
80
60
40
20
0
dissimilarity

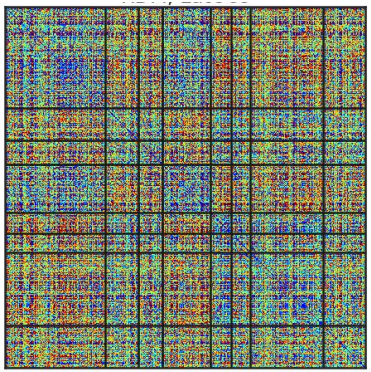
Results


Pooling Layers vs Lateral Occipital ROI
Correlation
0.040
0.035
0.030
0.025
0.020
0.015
0.010
Layer 3
Layer 6
Layer 11
Layer 16
Layer 21

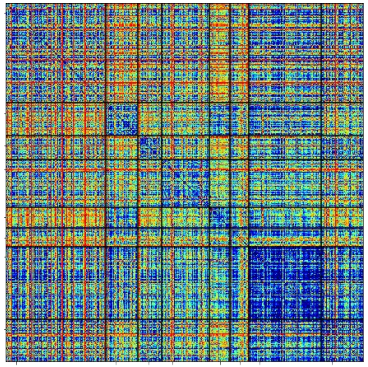

Pooling Layers vs V1 ROI
Correlation
0.08
0.07
0.06
0.05
0.04
Layer 3
Layer 6
Layer 11
Layer 16
Layer 21

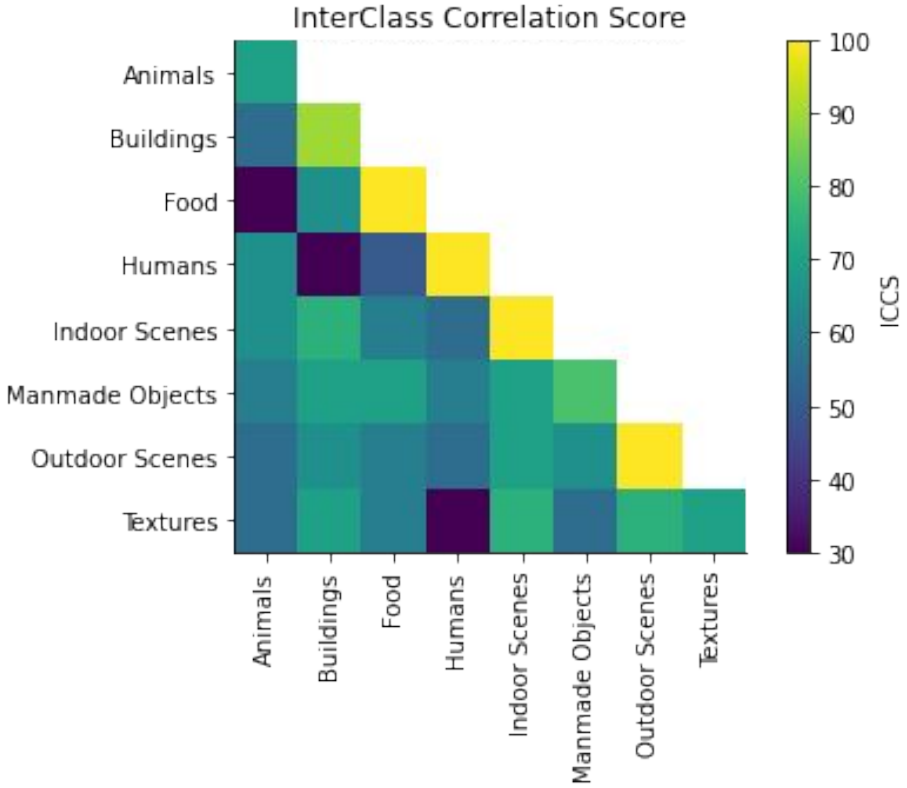
Results

Lateral Occipital



CNN layer 21

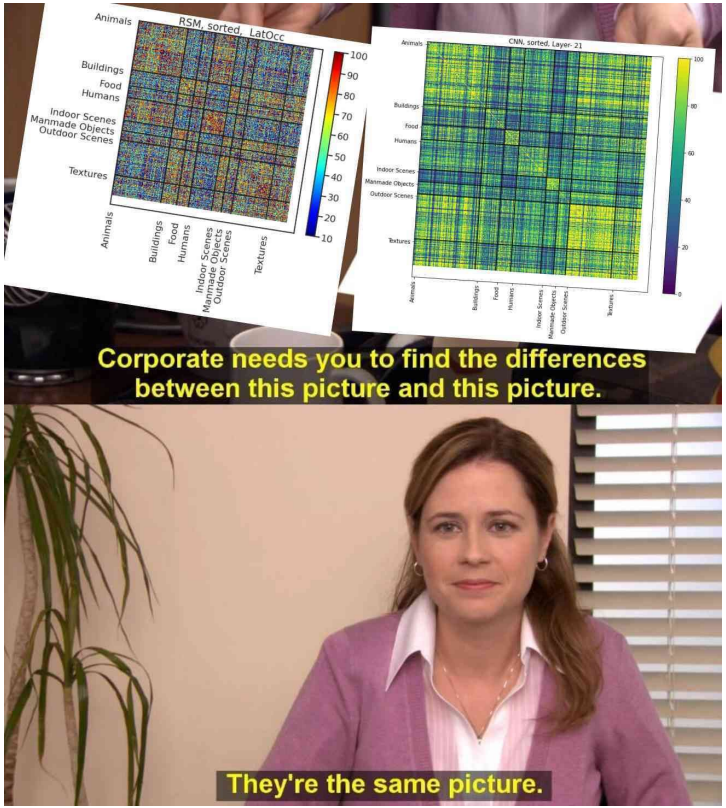




Discussion

- Early and Late CNN correlate best with Early and Late Visual Hierarchy
- Semantic structure observable in both RSAs
- Absolute correlation remains low!
- Further research?

Questions?


RSM, sorted_LatOcc
Animals
Buildings
Food
Humans
Indoor Scenes
Manmade Objects
Outdoor Scenes
Textures
CNN, sorted, Layer- 21
Corporate needs you to find the differences between this picture and this picture.
They're the same picture.

EGGPLANT COBRA