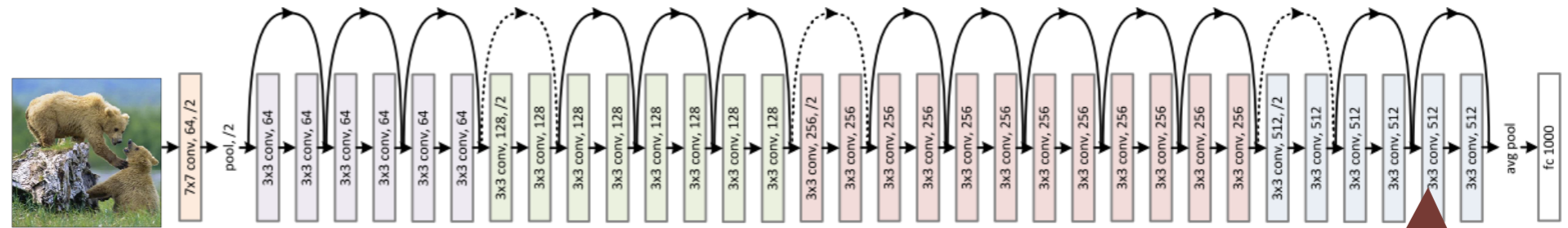# *Factorized visual representations in the primate visual system and deep neural networks*

Jack Lindsey
Columbia University
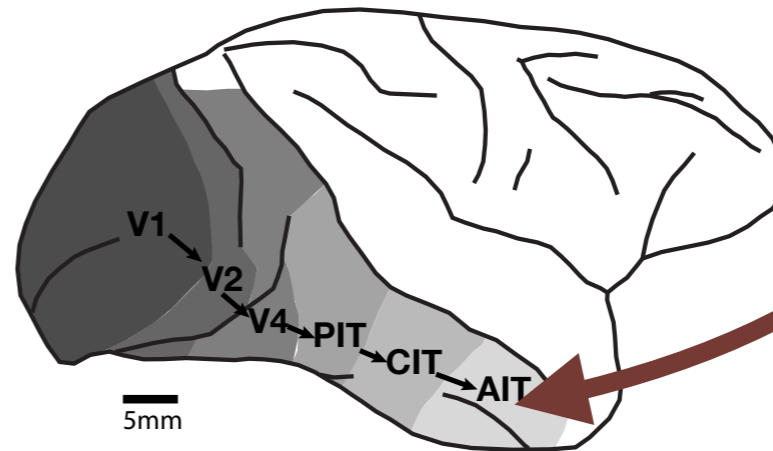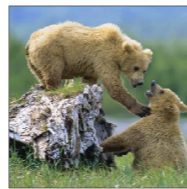
Work conducted with Elias Issa

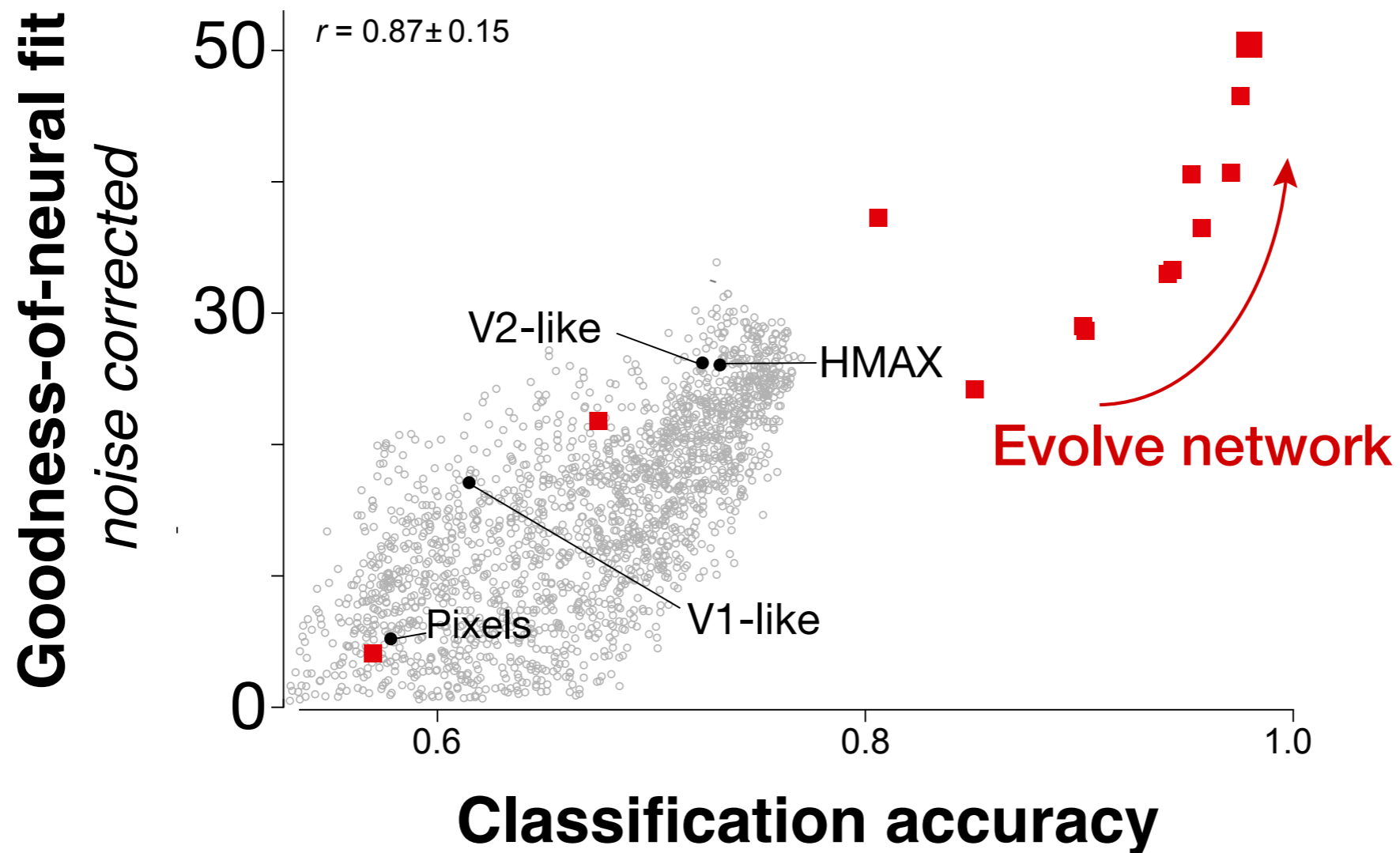# Deep neural network models predict neural activity in visual brain regions

How well does a model's image classification performance predict its fit to neural data?

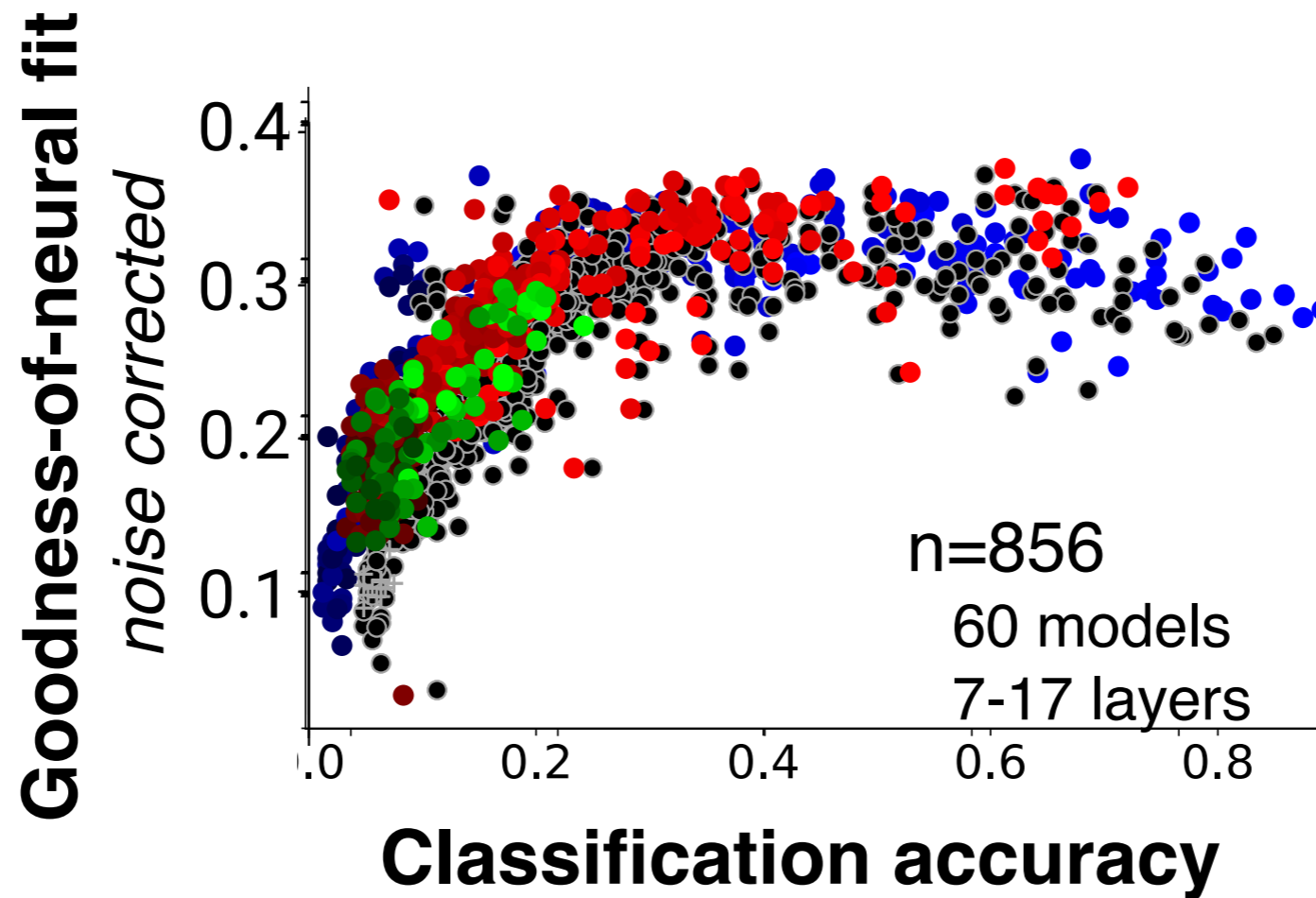Yamins & DiCarlo, *PNAS* 2014

How well does a model's image classification performance predict its fit to neural data?

How well does a model's image classification performance predict its fit to neural data?

Goodness-of-neural fit *noise corrected*

0.4
0.3
0.2
0.1

n=856
60 models
7-17 layers

Classification accuracy

.0   0.2   0.4   0.6   0.8

Architectures
ResNet-18,
ResNet-50,
ResNeXt,
VGG,
Inception…

Unsupervised auxiliary tasks
- BigBiGAN
- Relative patch location
- Rotation prediction
- Colorization
- etc

Self-Supervised contrastive /
cluster learning
- SimCLR
- MOCO
- BYOL
- etc

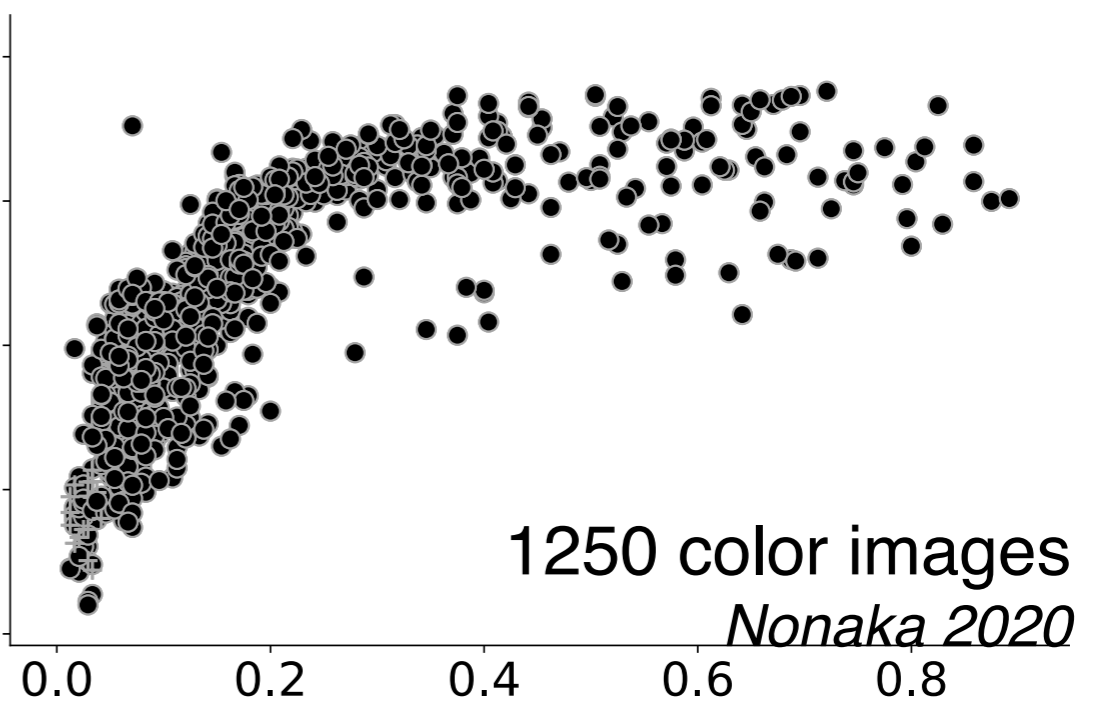Supervised Classification
- AlexNet
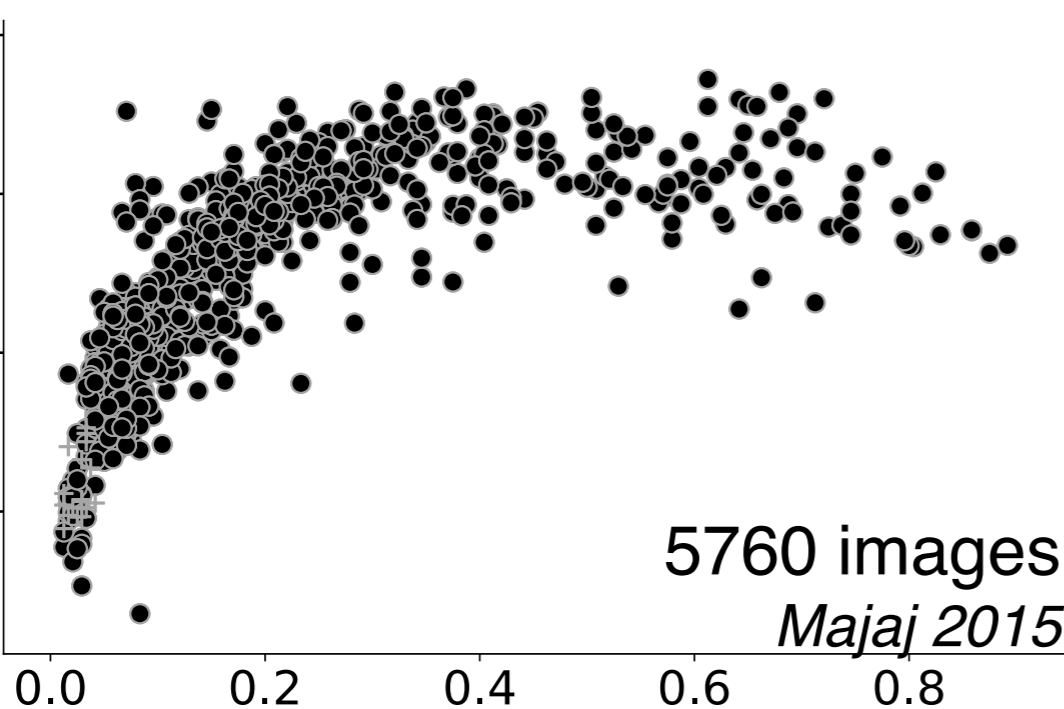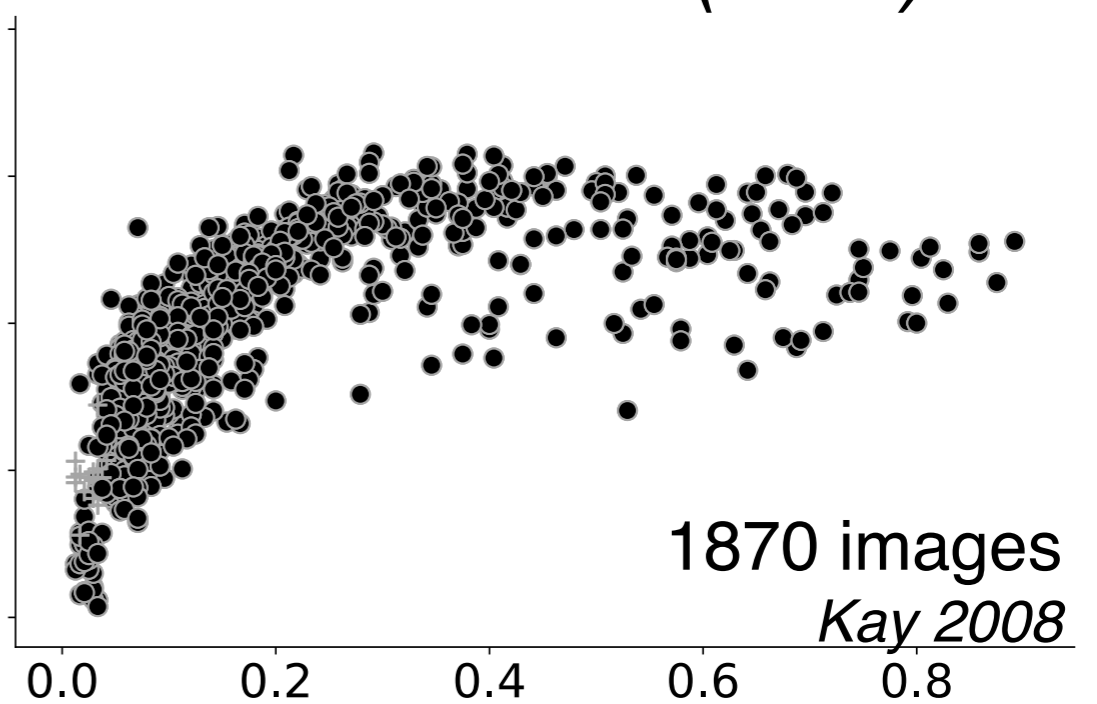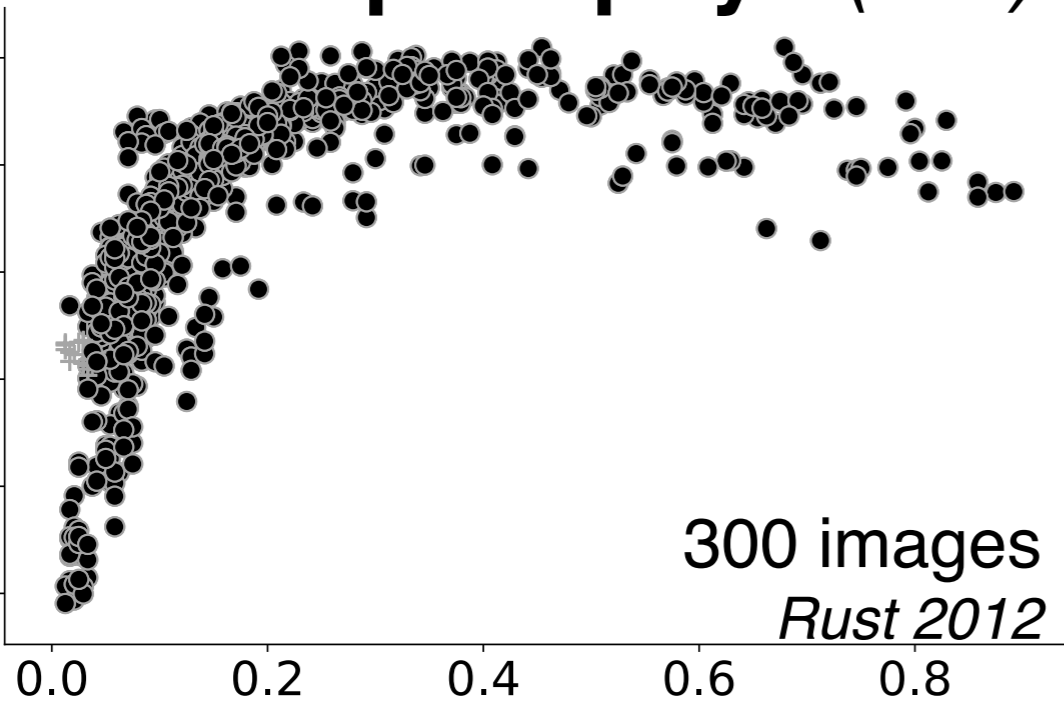- ResNet
- VGG
- Inception
- etc
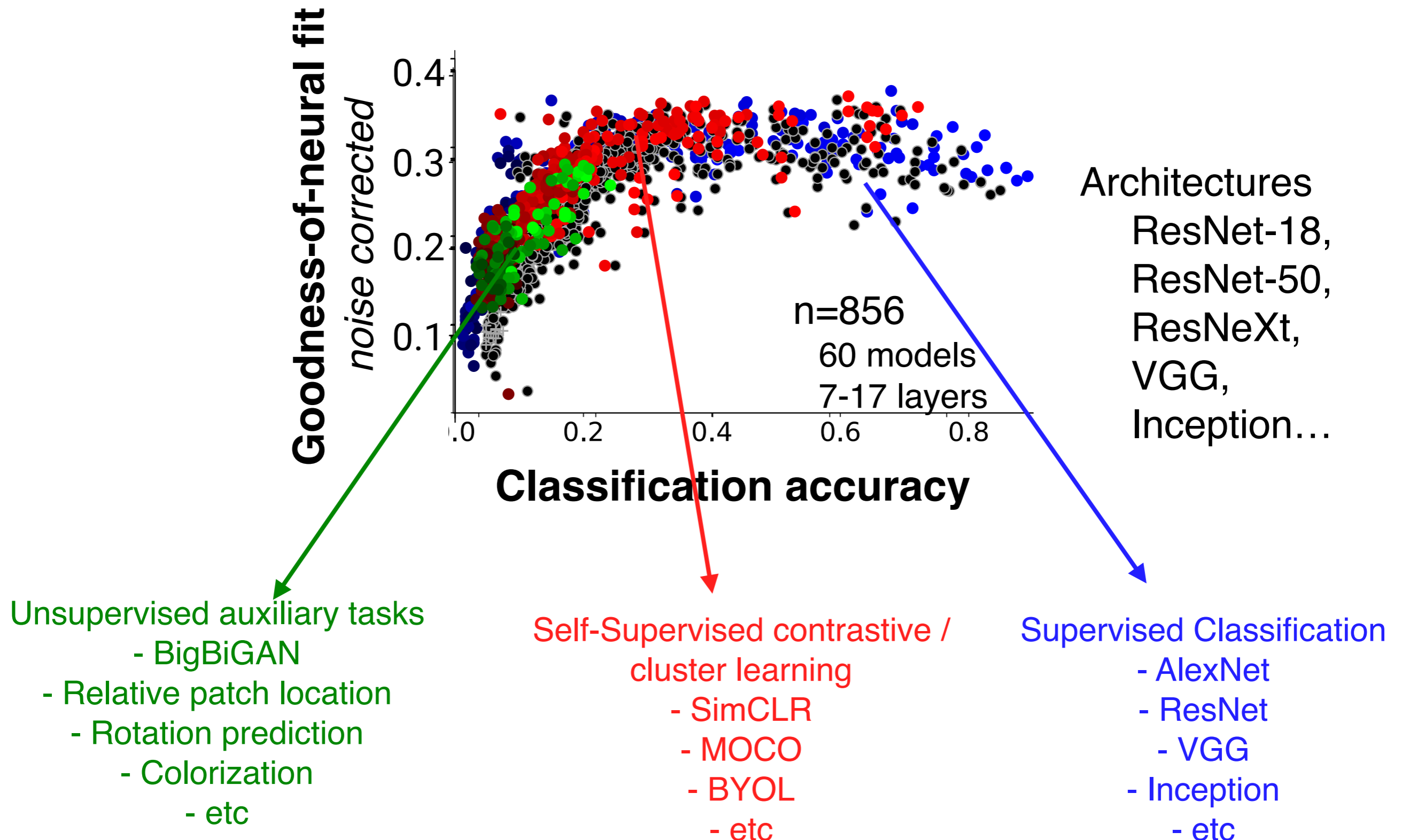
Trend holds across multiple datasets

Macaque Ephys *(ITC)*

Human fMRI *(HVC)*

Goodness-of-neural fit
noise corrected

Classification accuracy

300 images
Rust 2012

1870 images
Kay 2008

5760 images
Majaj 2015

1250 color images
Nonaka 2020
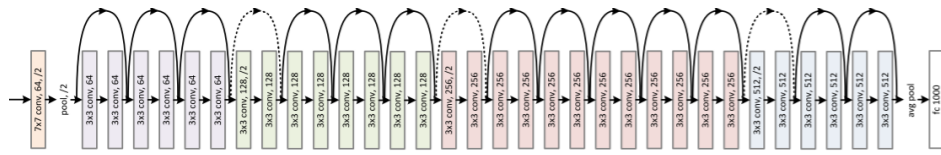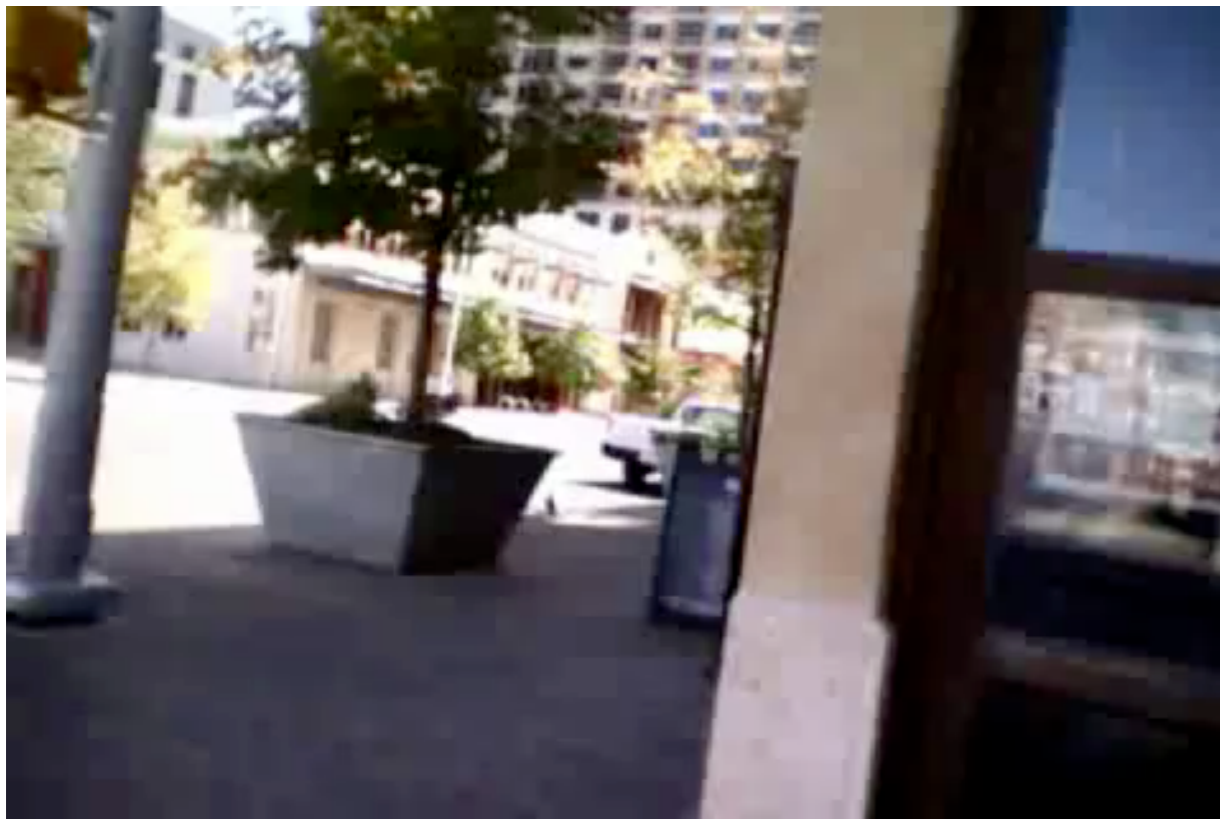
Moving beyond classification

What additional properties of a model ***make some better than others*** at predicting activity in the primate visual system?

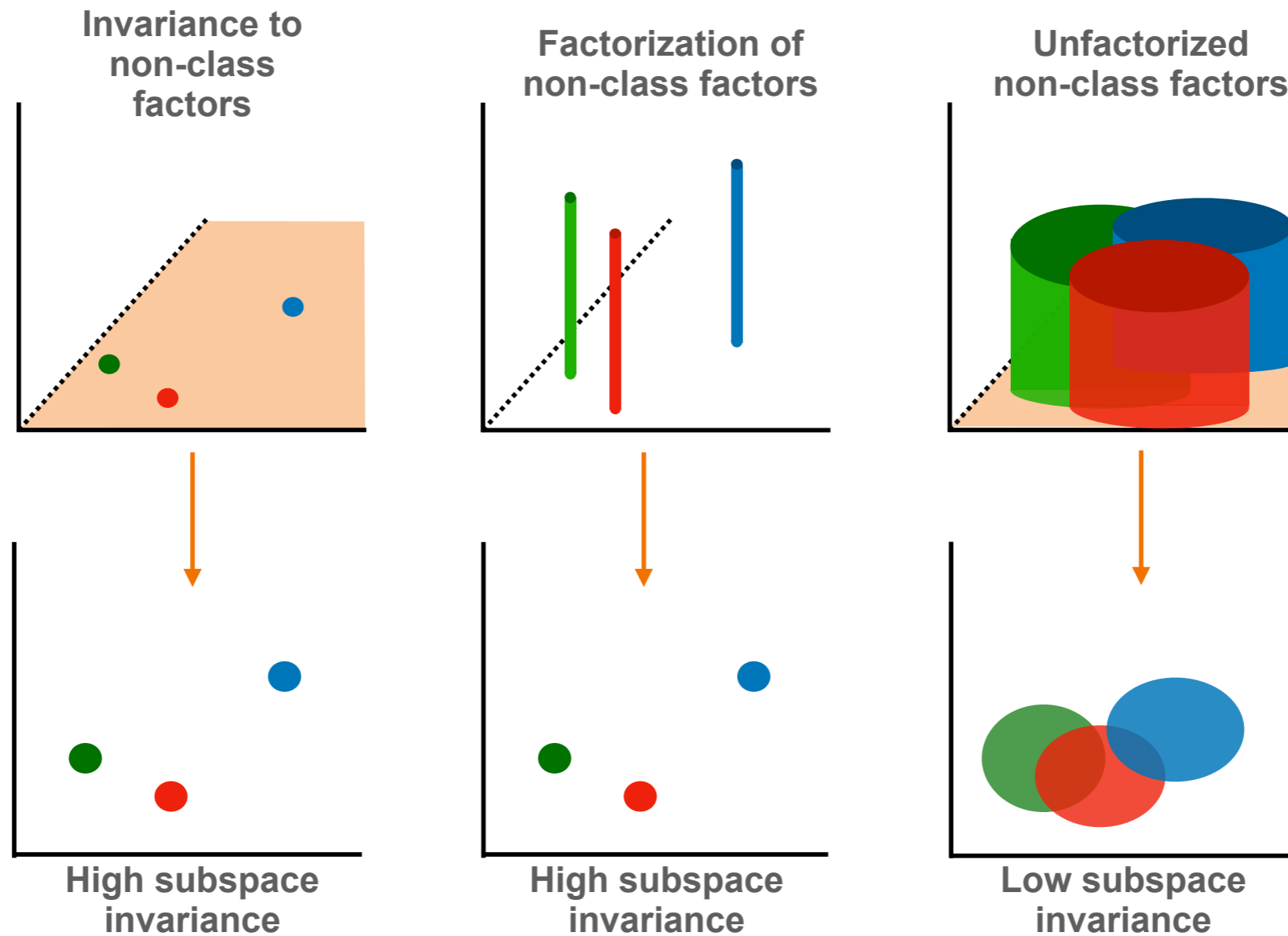# Conceptual gap between models & visual perception

Most **models optimize for classification of object identity,** *irrespective of* pose & scene composition.

The brain supports **a rich understanding of many scene properties**.

Candidate property: factorization

Factorization measures the extent to which different parameters drive overlapping vs. orthogonal variance in neural activity space
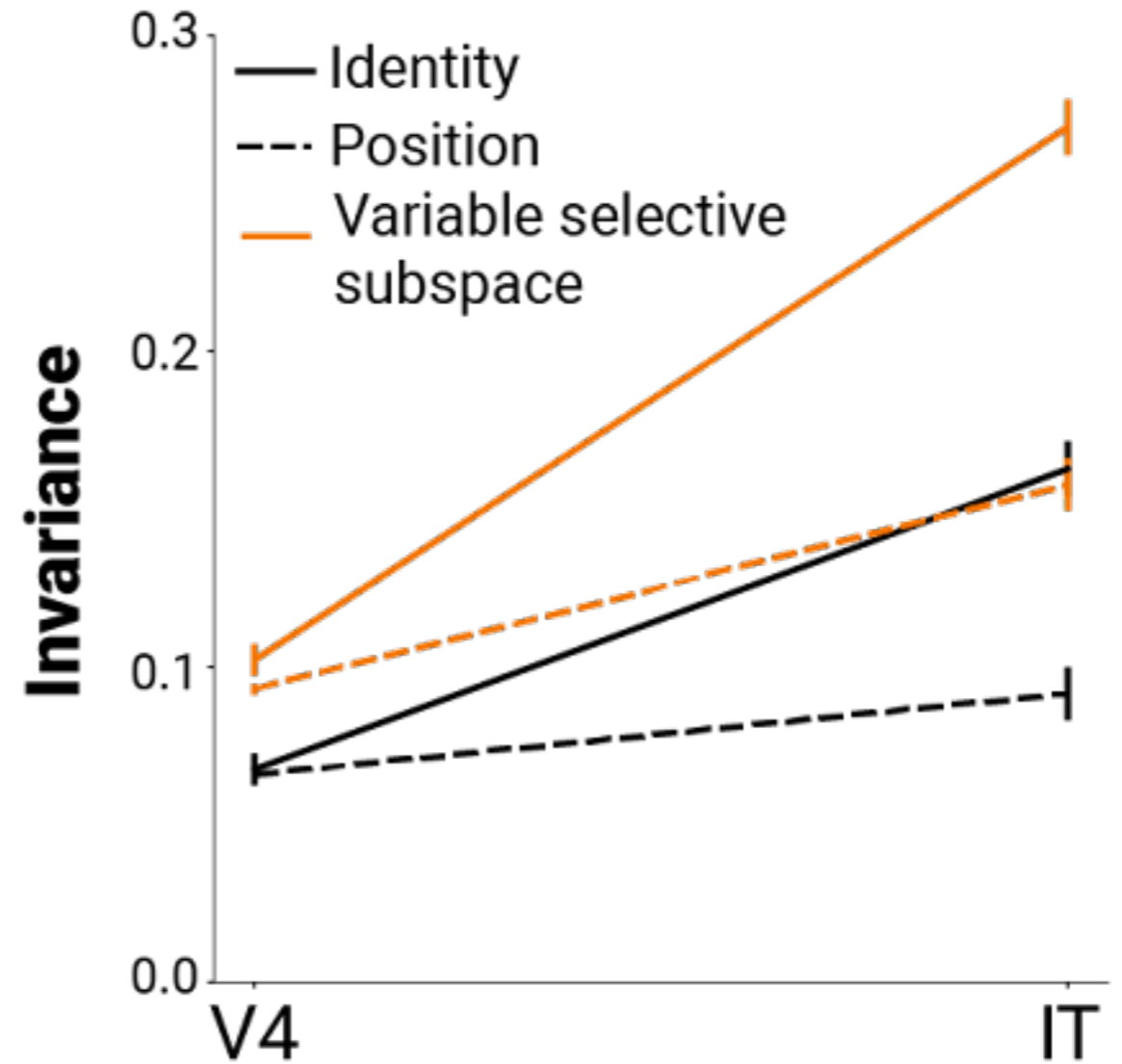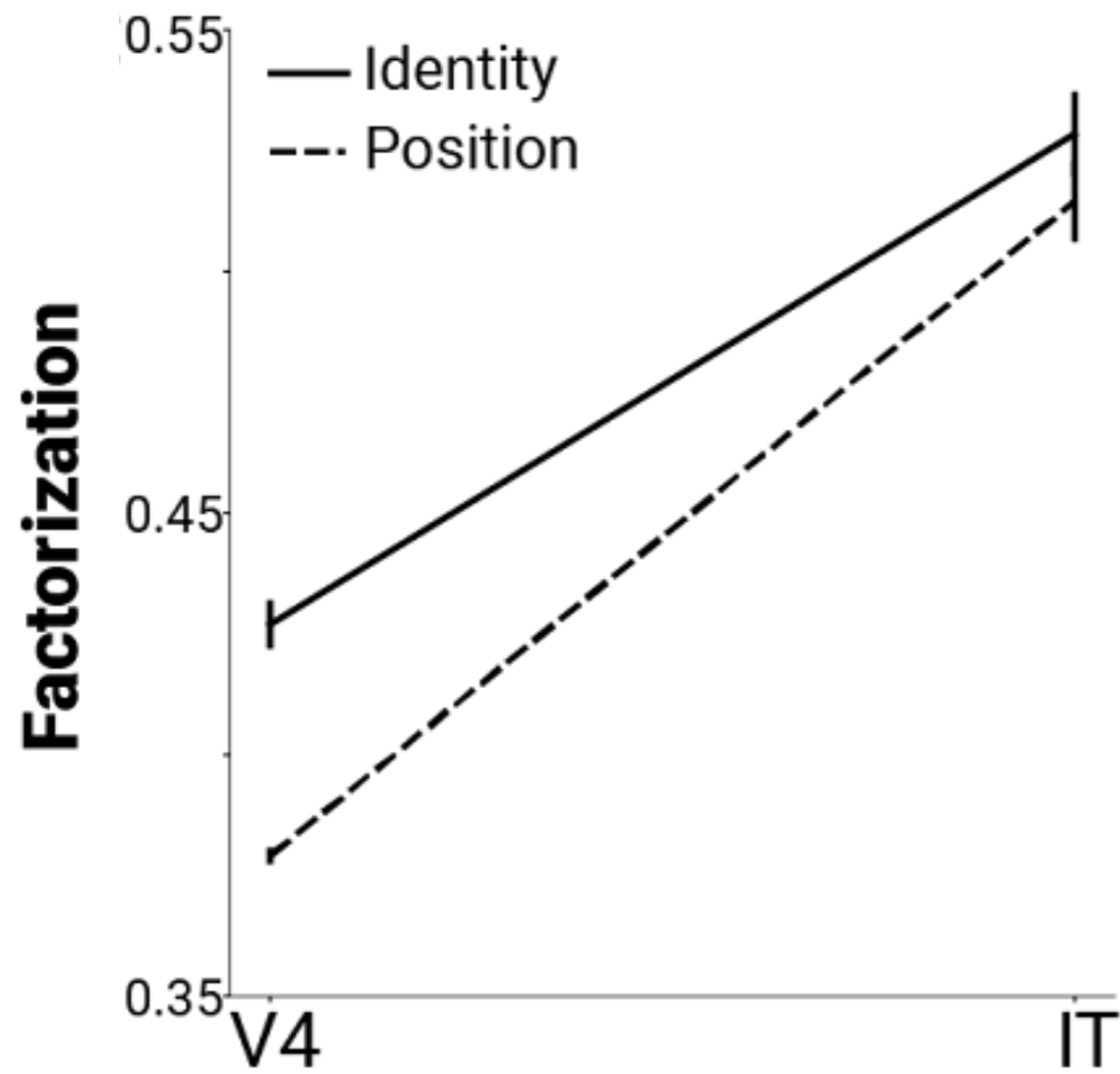
Metrics

Invariance: $1 - \dfrac{\text{var}(P)}{\text{var}(\text{all})}$

Factorization: $1 - \dfrac{\text{var}(P \,|\, \text{subspace}[Q])}{\text{var}(P)}$

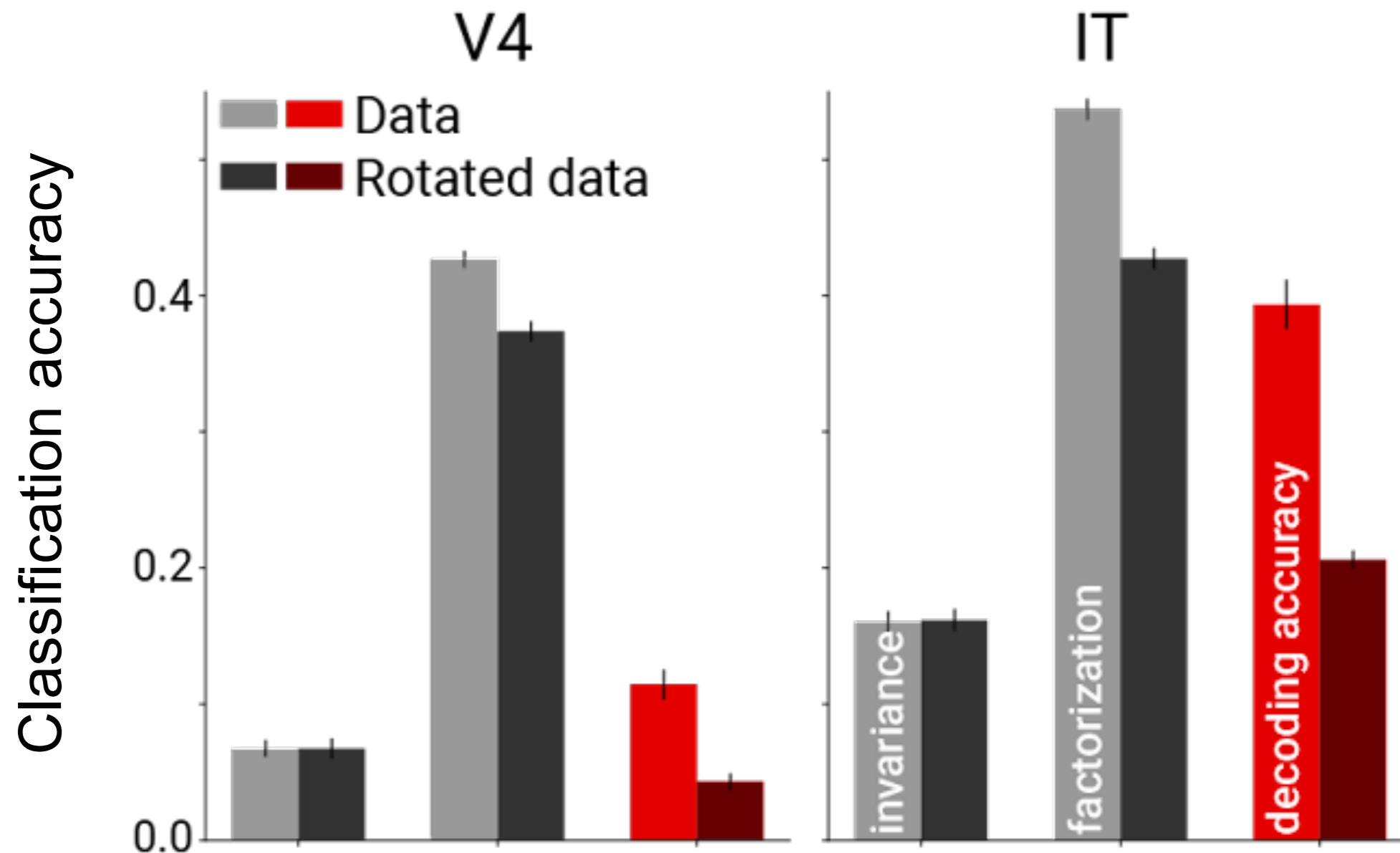## Notation

- var(P) = "variance induced by parameter P (controlling all other parameters)"
- subspace(P) = "top principal components of variance induced by parameter P"
- var(P | S) = "var(P) when projected onto S"

# Factorization increases across the visual stream

Factorization is important for object decoding

- Synthetic data with reduced factorization but otherwise identical statistics is much worse at supporting object decoding
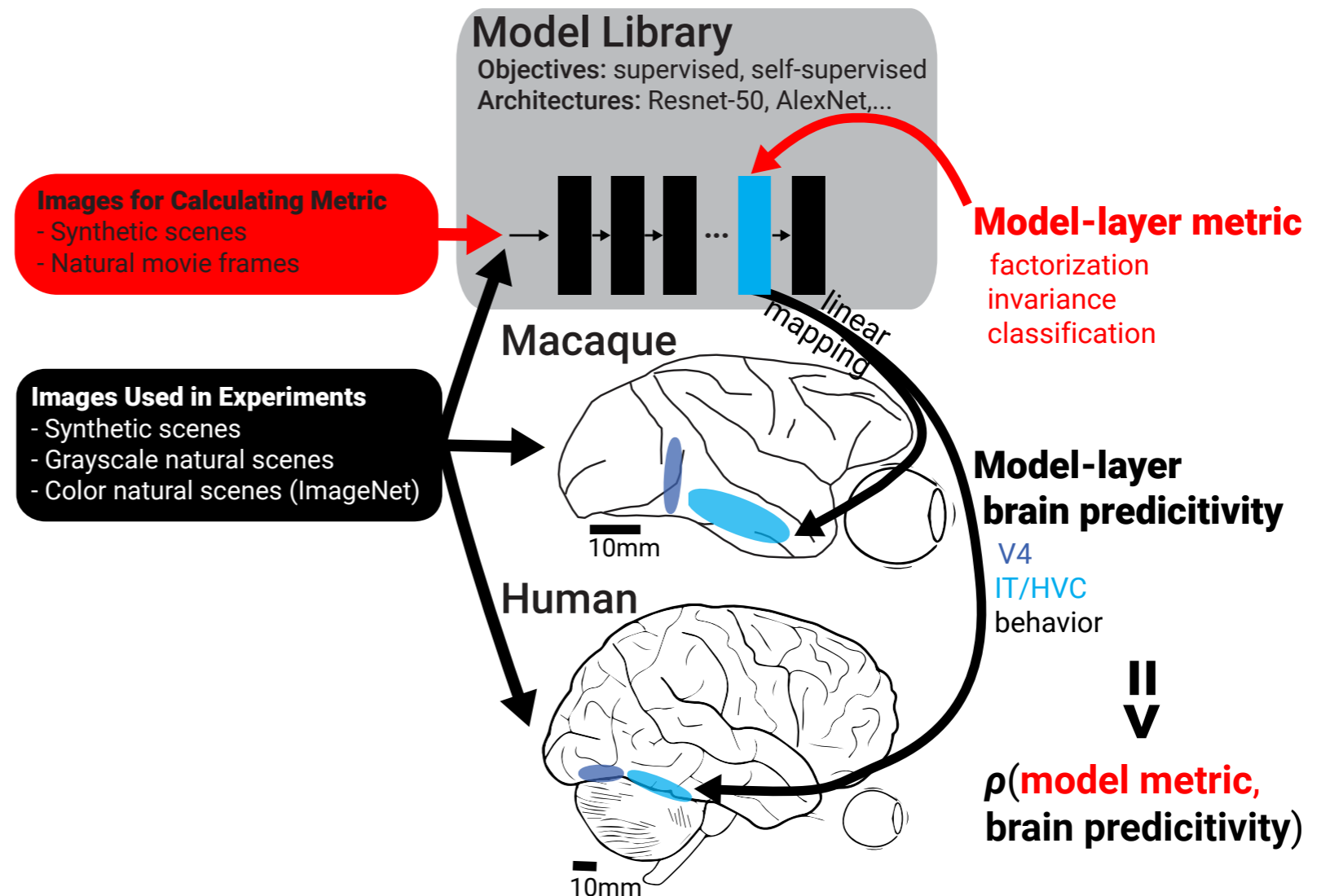
- Can't do much more than this with existing neural data

- No large-scale datasets of visual responses to parametrically varied scenes

- Solution: large-scale analysis of deep neural network models

A synthetic stimulus set to independently control scene variables

Foreground object identity

Background identity

Object pose

Lighting

Viewpoint
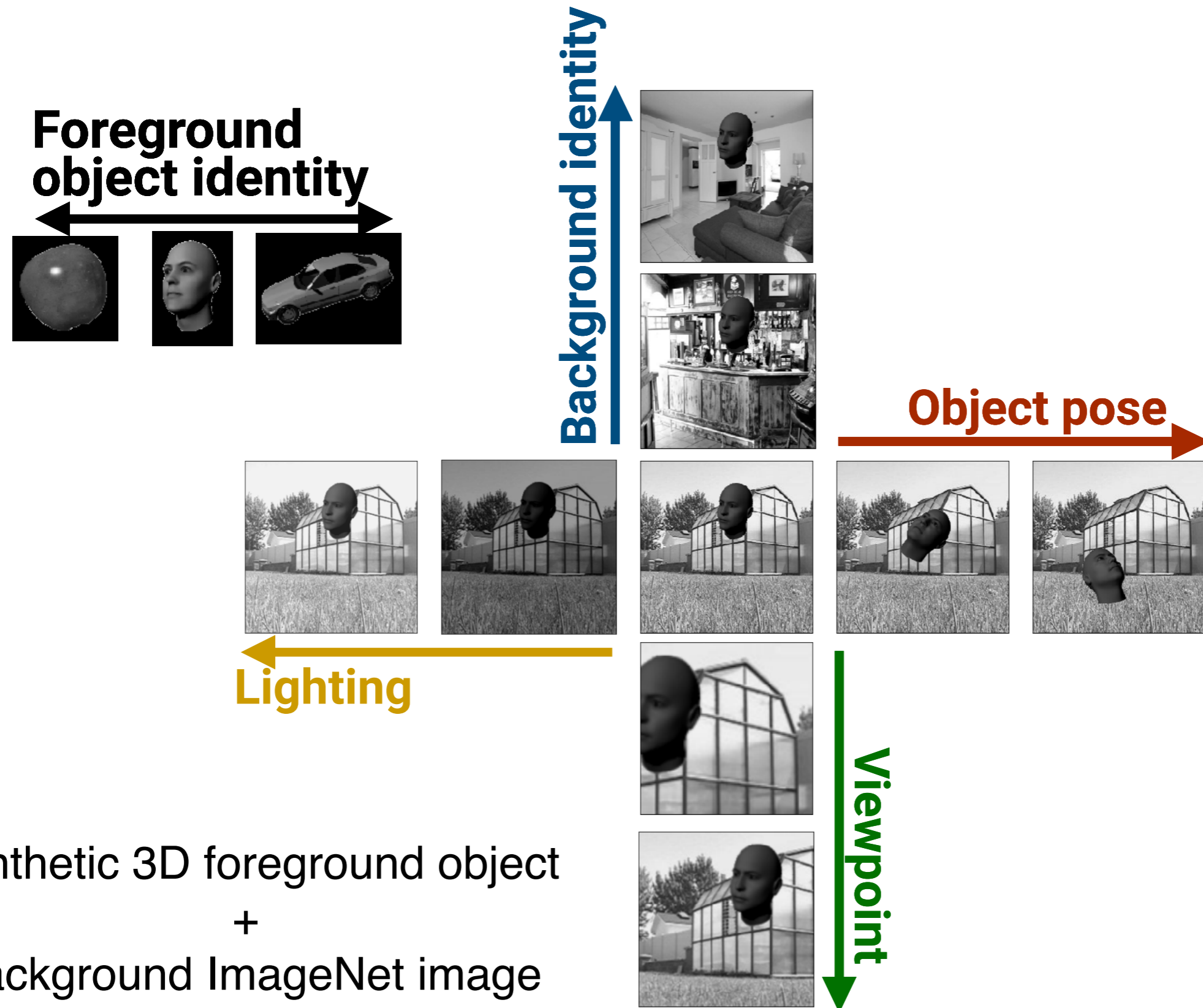
Synthetic 3D foreground object
+
Background ImageNet image

Strategy: what model properties are associated with match to neural data?

Model Library
Objectives: supervised, self-supervised
Architectures: Resnet-50, AlexNet,...

Images for Calculating Metric
- Synthetic scenes
- Natural movie frames

Model-layer metric
factorization
invariance
classification

Images Used in Experiments
- Synthetic scenes
- Grayscale natural scenes
- Color natural scenes (ImageNet)

Macaque

10mm

linear mapping

Model-layer brain predicitivity
V4
IT/HVC
behavior

Human

10mm

ρ(**model metric**, **brain predicitivity**)

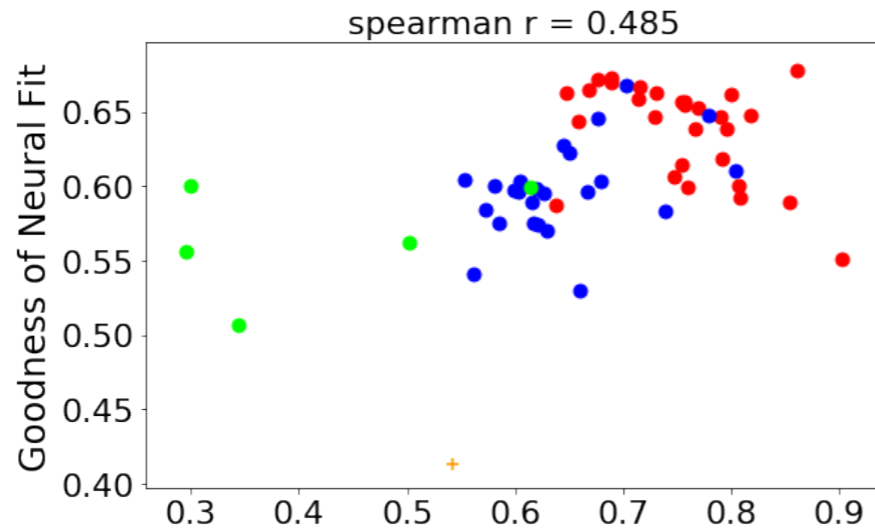**1) Compute scene encoding properties of model**

- *Factorization + invariance metrics*

- *Involves >1k PCA runs over >100k dimensions*
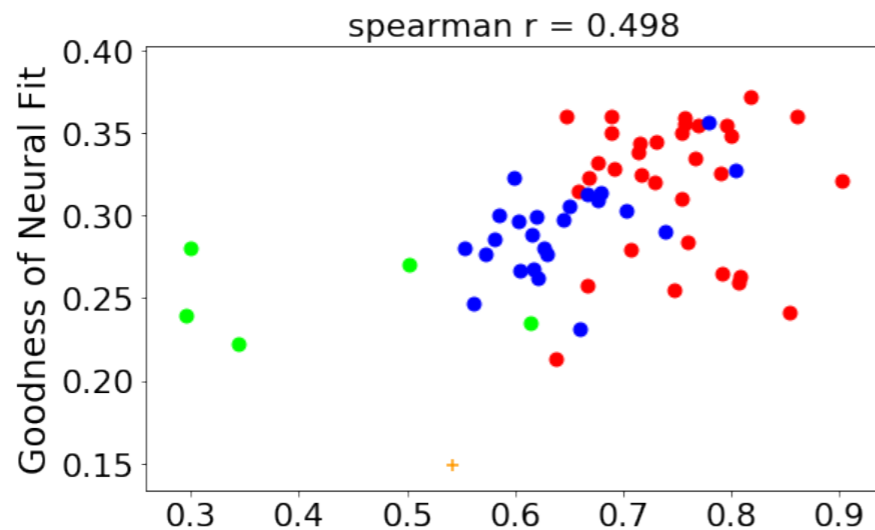
2) Assess how well models match neural data

- *Fit linear* ~~sion~~ *from model neurons* ~~a ne~~ */ voxels / behavior in* ~~nse~~ *same images*

~~res~~ ~~ge~~ *regressions over >100k dimensions*

Example: viewpoint factorization

Ephys 1

spearman r = 0.485

fMRI 1

spearman r = 0.262

Ephys 2

spearman r = 0.498

fMRI 2

spearman r = 0.492
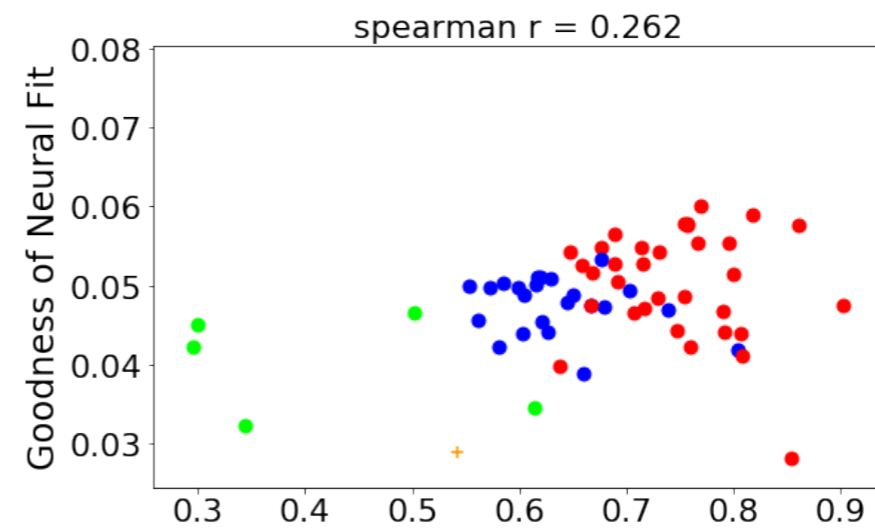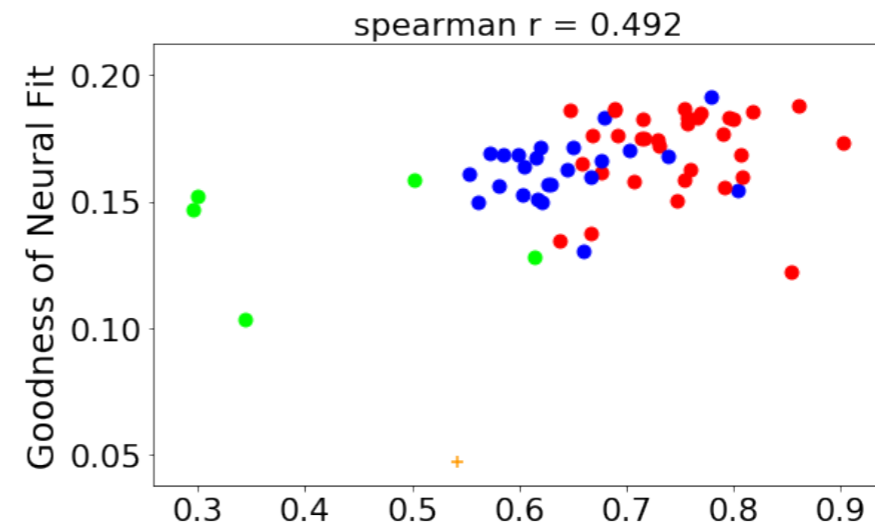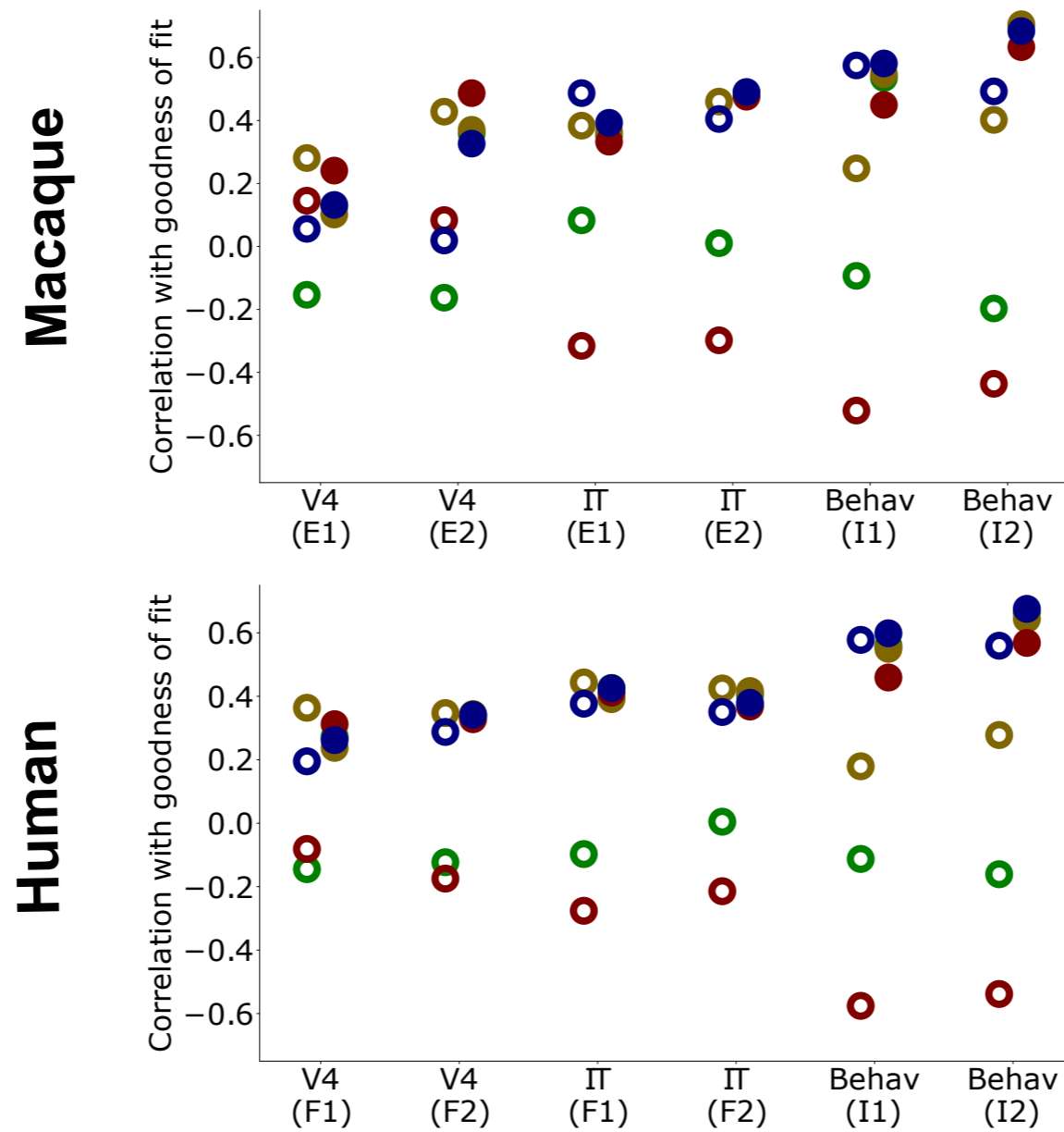
- Shown here: last layer of each network
- Take correlation values, average across datasets + layers
- Repeat for other metrics

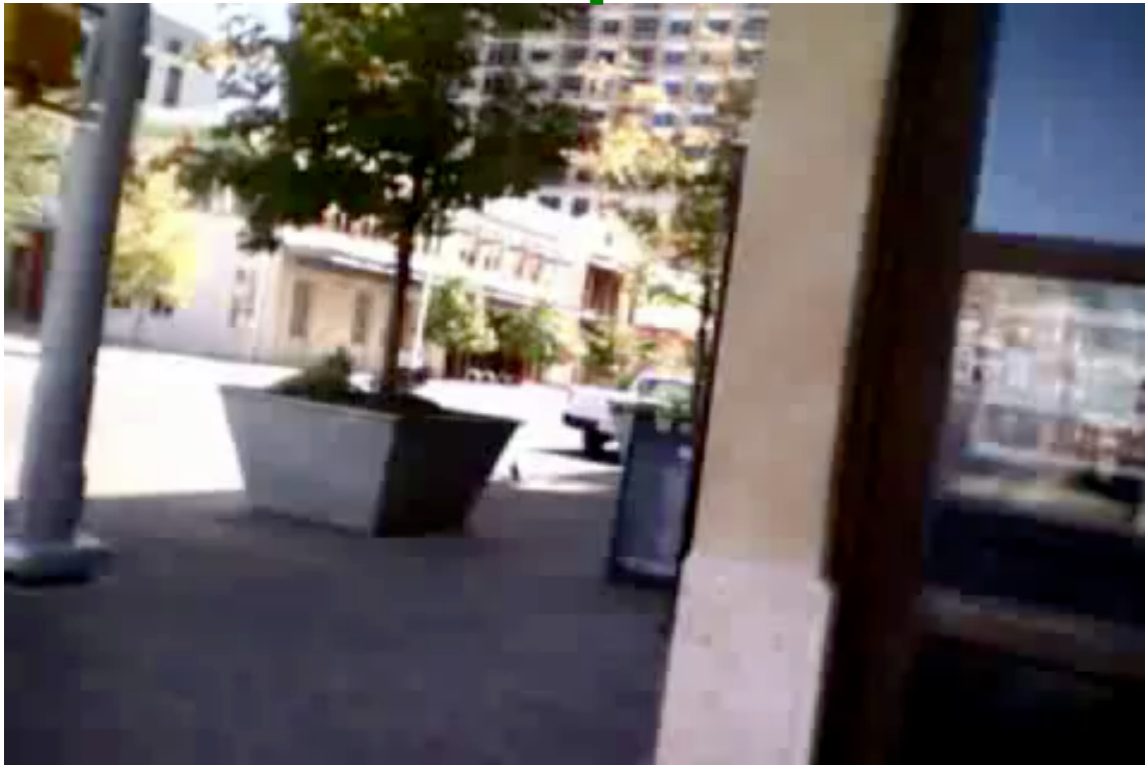# More brain-like models tend to factorize scene variables

Factorization computed from natural videos

Viewpoint

Object motion

Macaque   **Human**

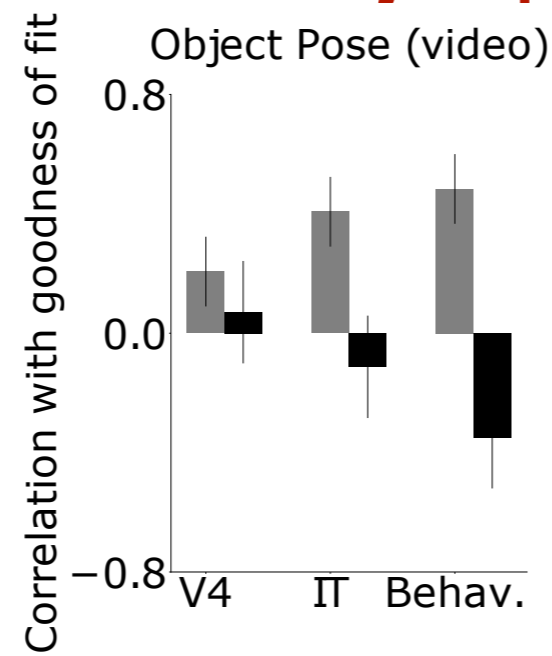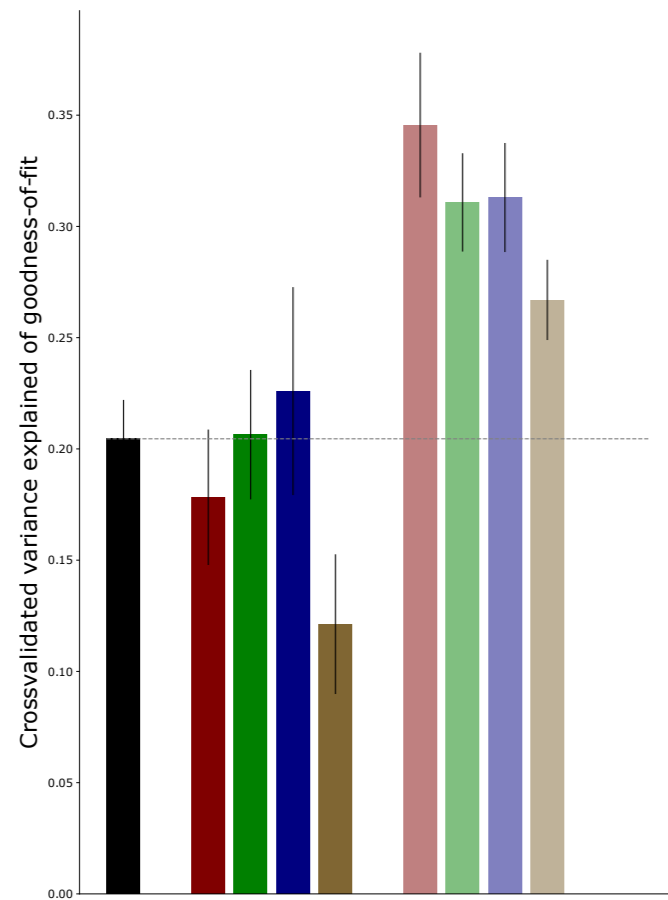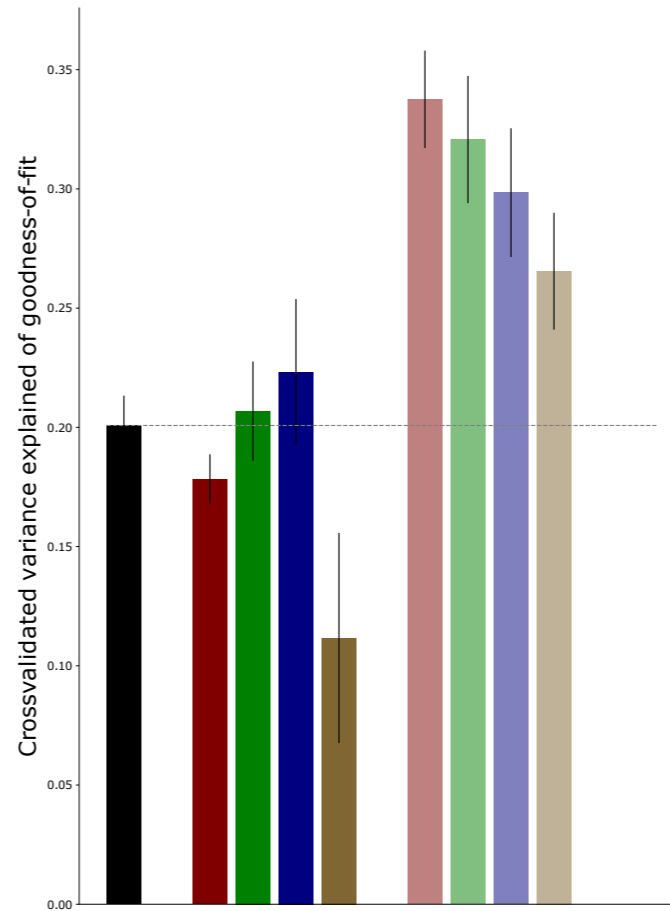**Viewpoint (video)**

Object pose (video)
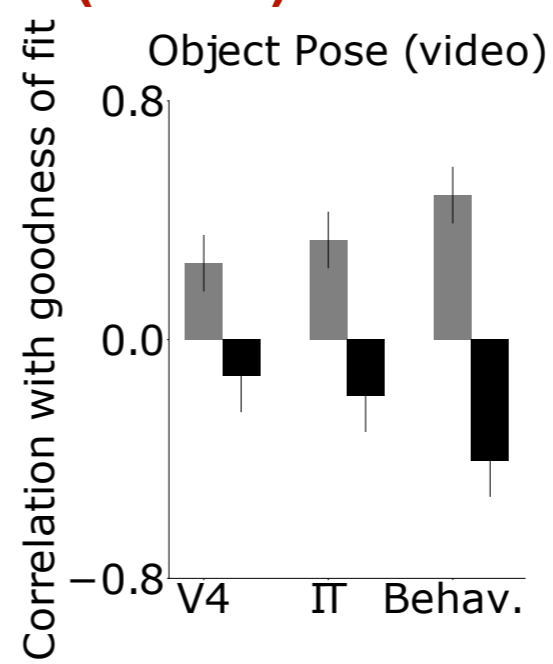
Factorization complementary to classification performance as an indicator of brain-like models

V4

IT
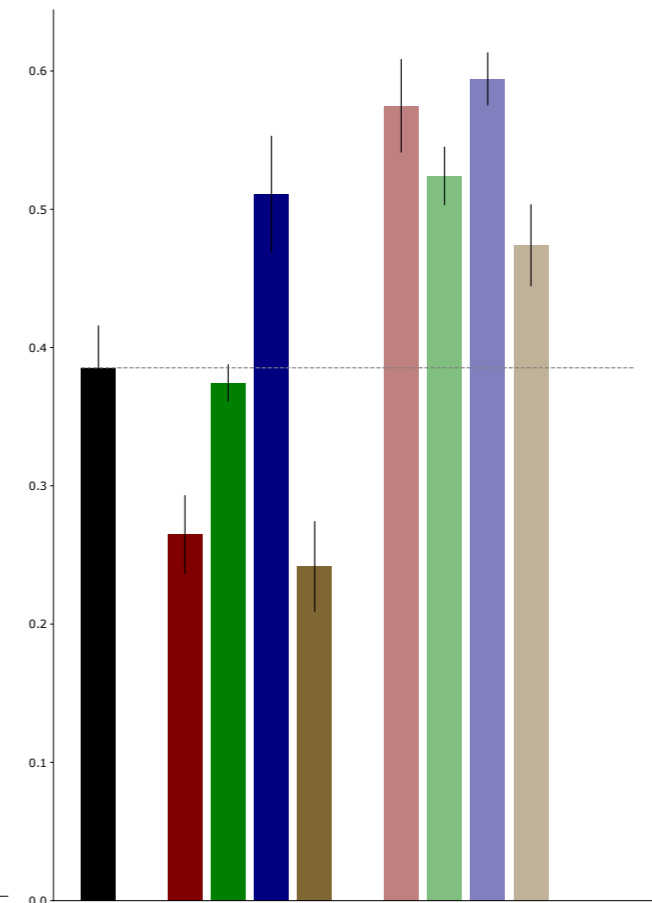
Behavior

Classification performance

Viewpoint fact. (+ class)

Object pose fact. (+ class)

Background fact. (+ class)

Lighting fact. (+ class)

Follow-up directions

- Auxiliary objective functions to directly incentivize factorization
  - Shown to improve out-of-distribution model performance (Ying et al., 2023)

- Unsupervised discovery of parameters being factorized by a given representation

- What is the biological substrate of parameter-driven subspaces?

Thank you!