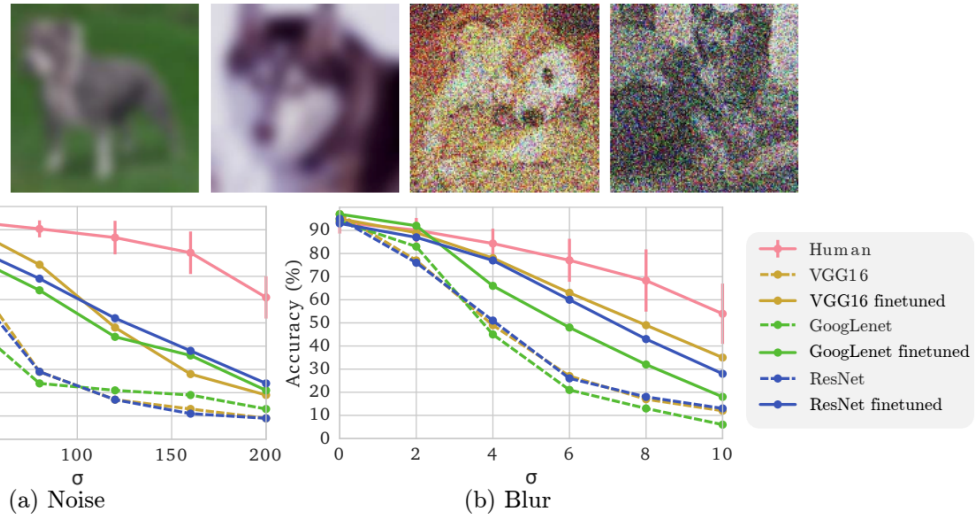
Improving Machine Vision Using Human Perceptual Representations

RT Pramod & SP Arun

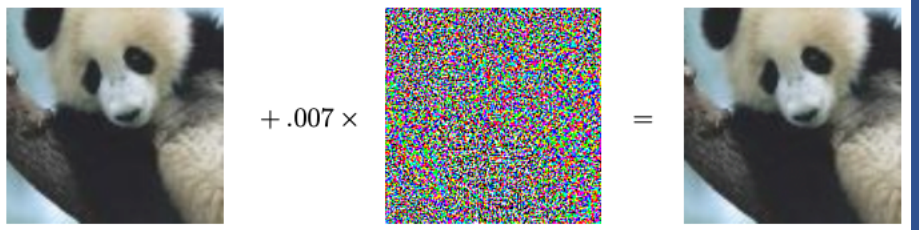
Performance gap between human and machine vision

Humans are better than CNNs at detecting cars and people in natural scenes *(Katti et al, Attn Percept Psychophys, 2019)*
Car detection: (Humans, 92% vs CNNs, 73%)
Person detection: (Humans, 94% vs CNNs, 84%)



Dodge & Karam, arXiv, 2017

Adversarial examples



Goodfellow et al, ICLR 2015

Geirhos et al, ICLR, 2019

Questions

- How closely do machine vision representations match human perception?
- Do machine vision models deviate systematically from human perception?
- Can we improve machine vision models using human perception?

Object recognition is similar in both machines and humans

Input Image

Feature extraction
in visual cortex

Binary decision
in prefrontal cortex

Output Label

Dog
Sheep
Motorcycle

Image Features
(Gabor, SIFT etc.)

Classifier
(LDA, SVM, etc.)

How do we compare Human and Machine Vision?

Input Image

Image Features
(Gabor, SIFT etc.)

Classifier
(LDA, SVM, etc.)

Output Label
Dog
Sheep
Motorcycle

How do we compare Human and Machine Vision?

Input Image

Image Features
(Gabor, SIFT etc.)

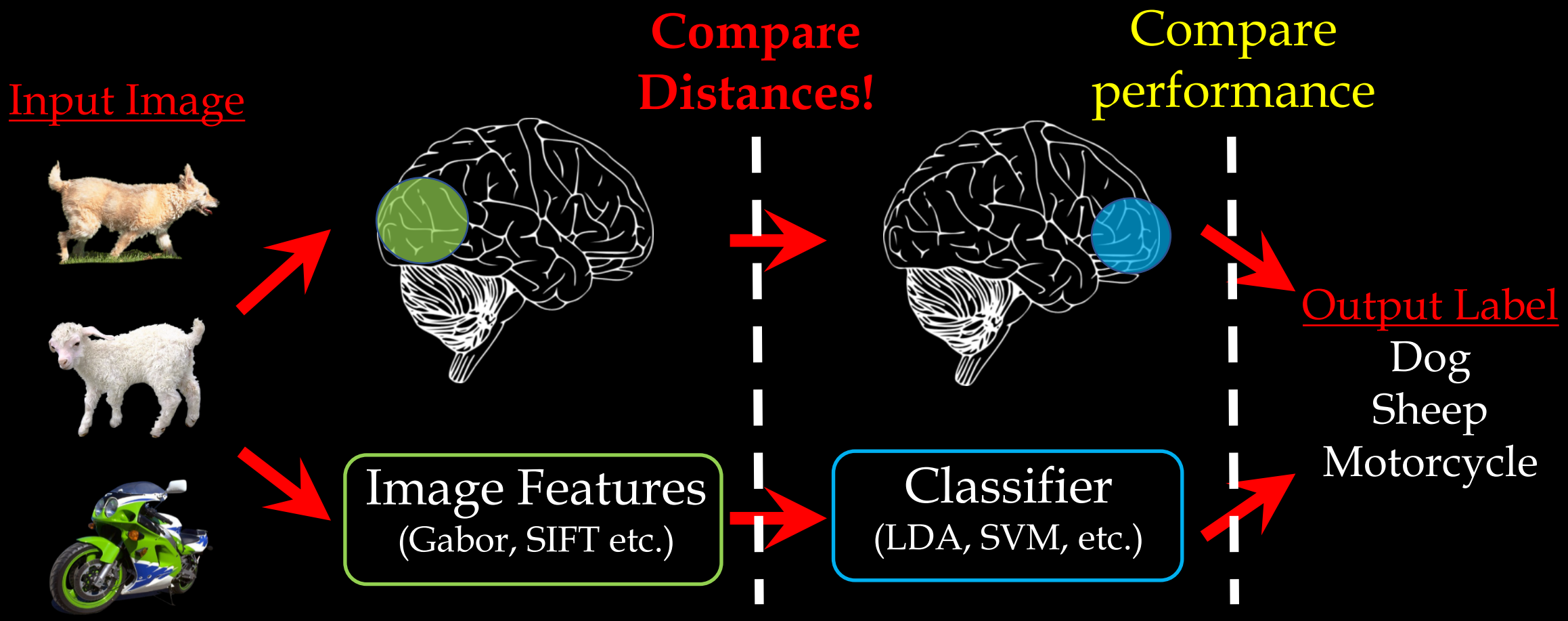
Compare features

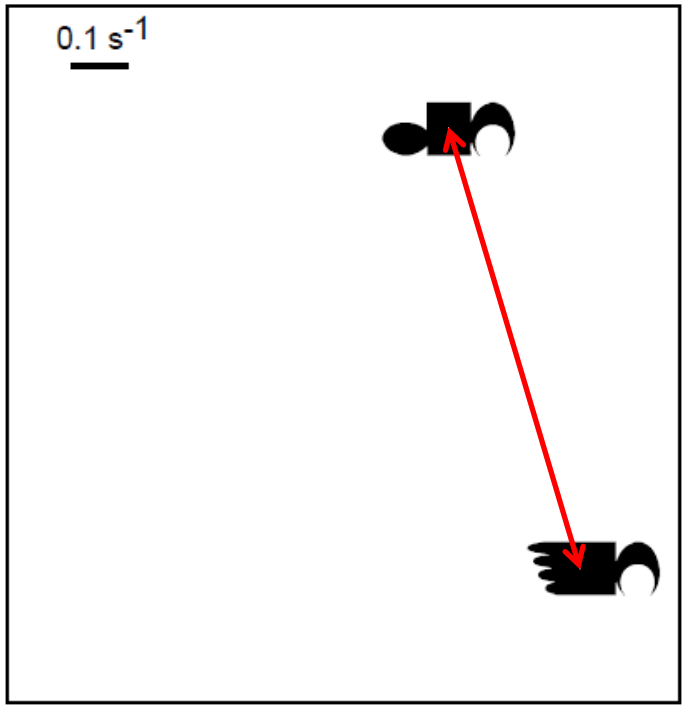
Classifier
(LDA, SVM, etc.)

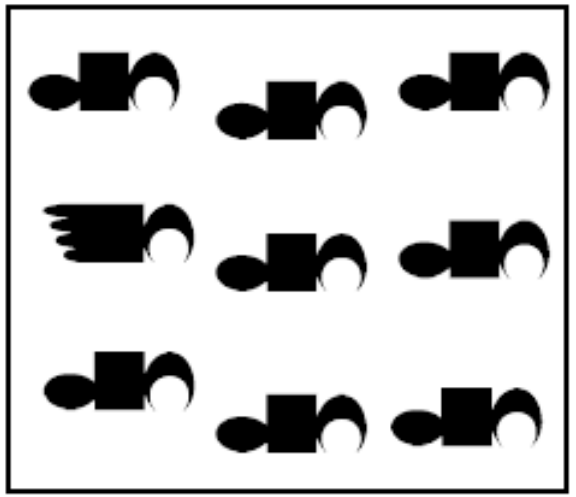
Compare performance

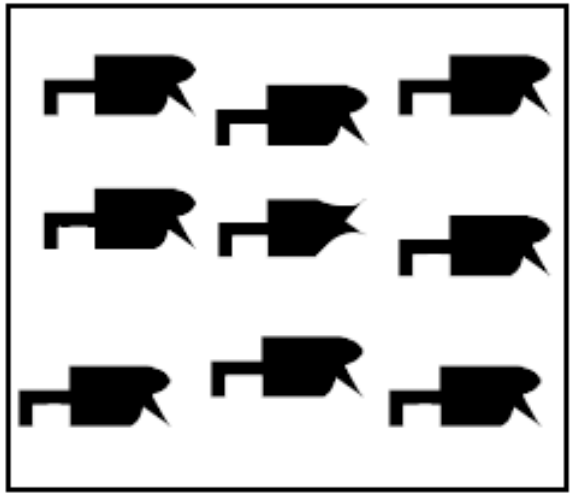
Output Label
Dog
Sheep
Motorcycle

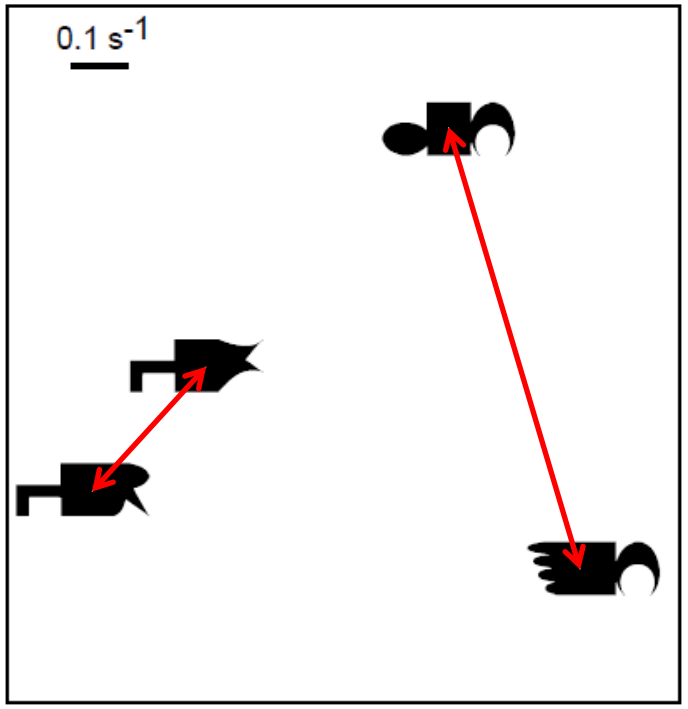
How do we compare Human and Machine Vision?

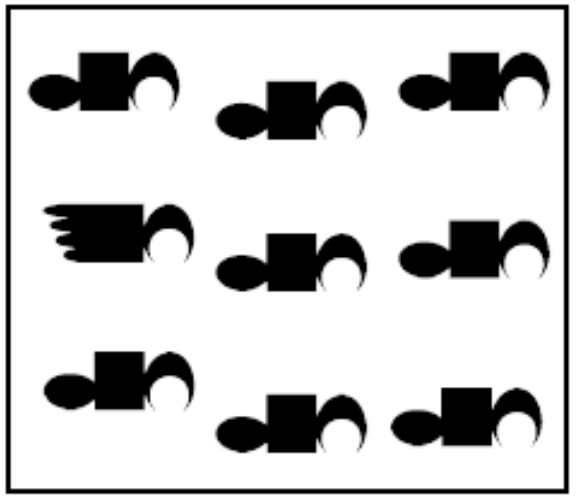


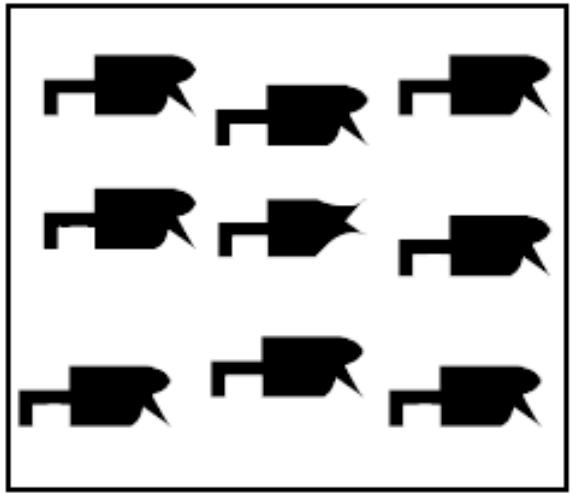

0.1 s-1

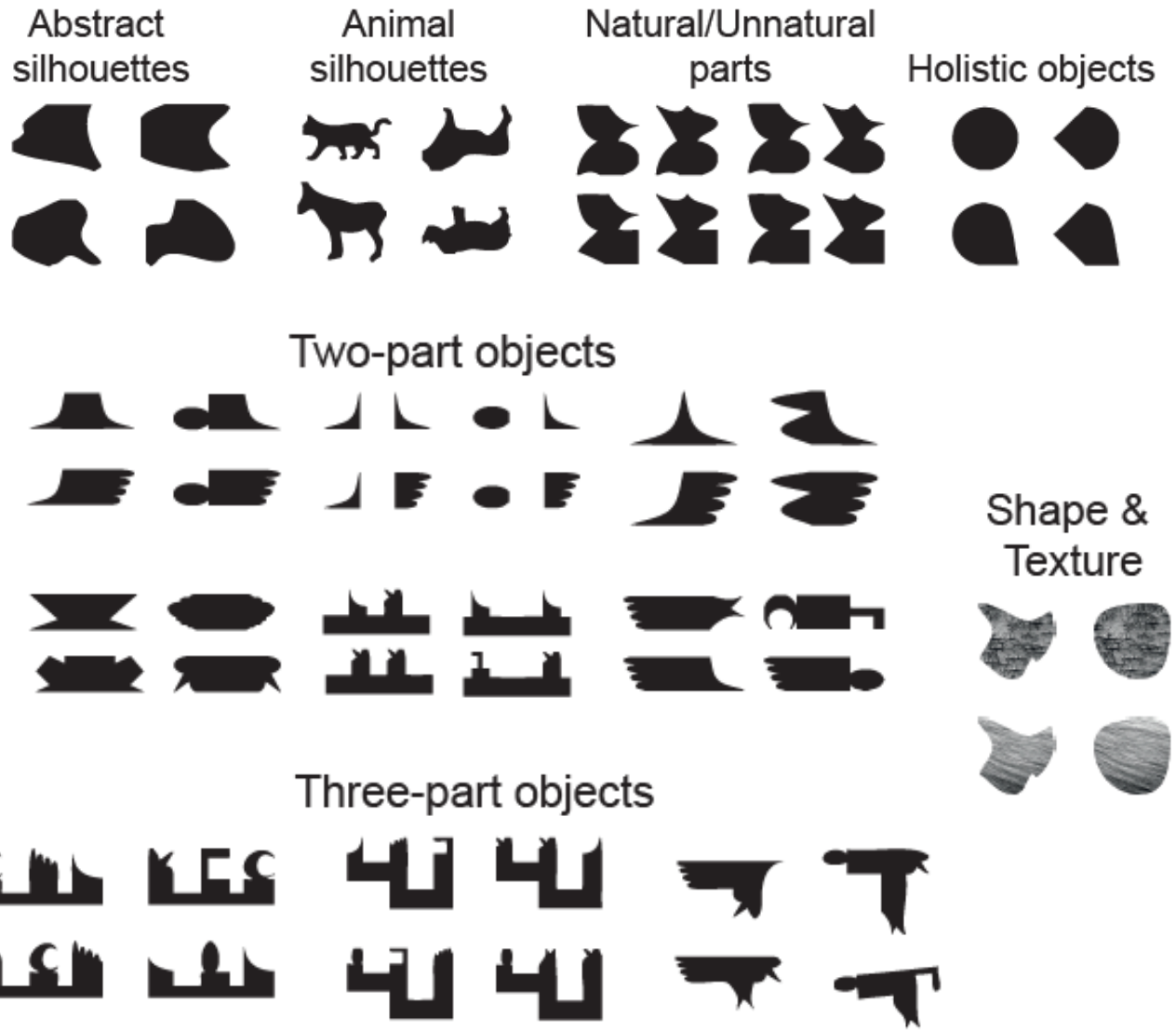


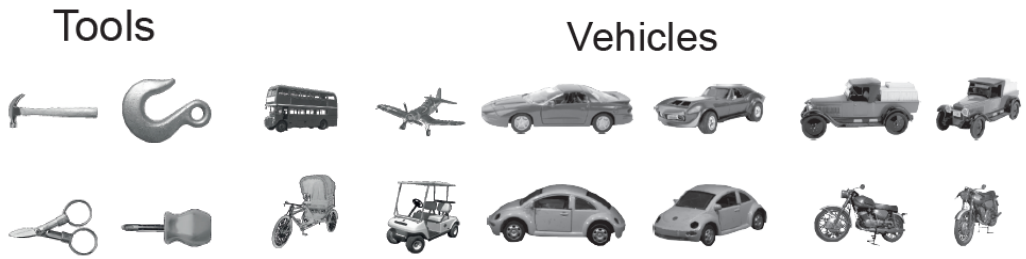



0.1 s-1

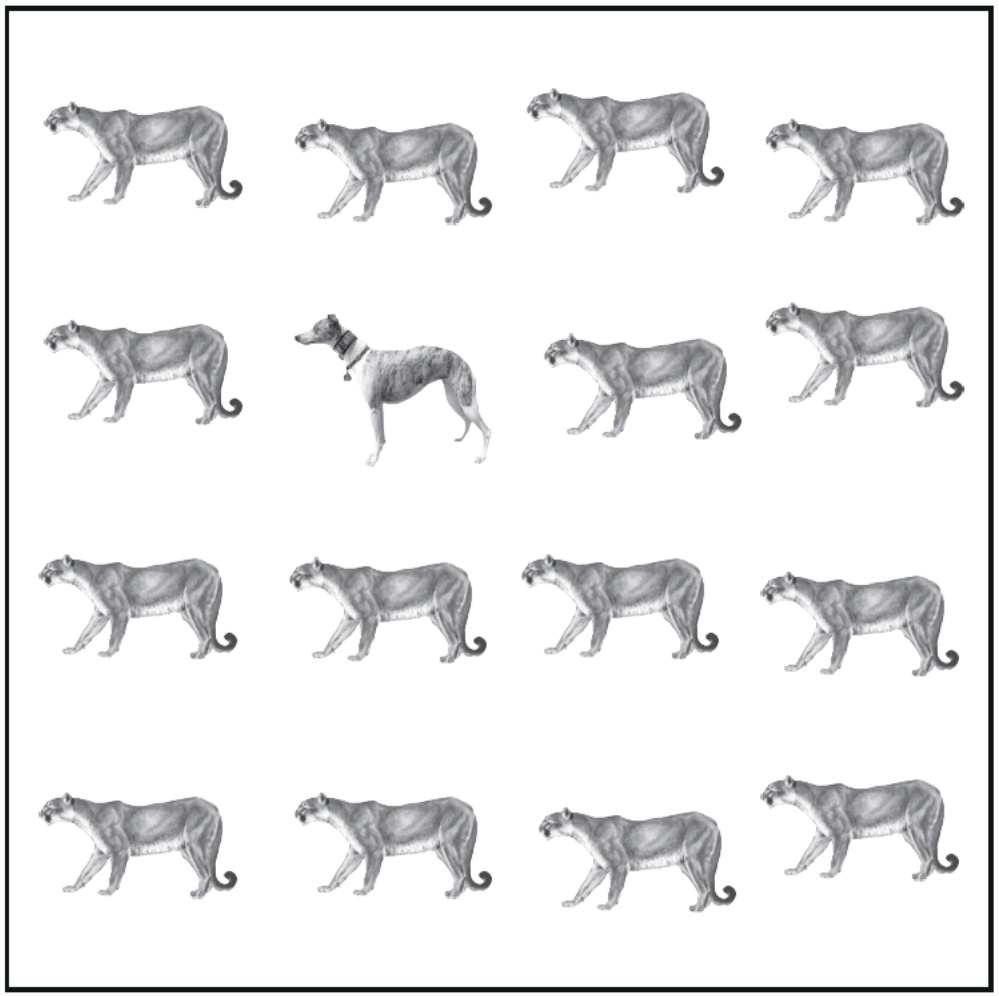


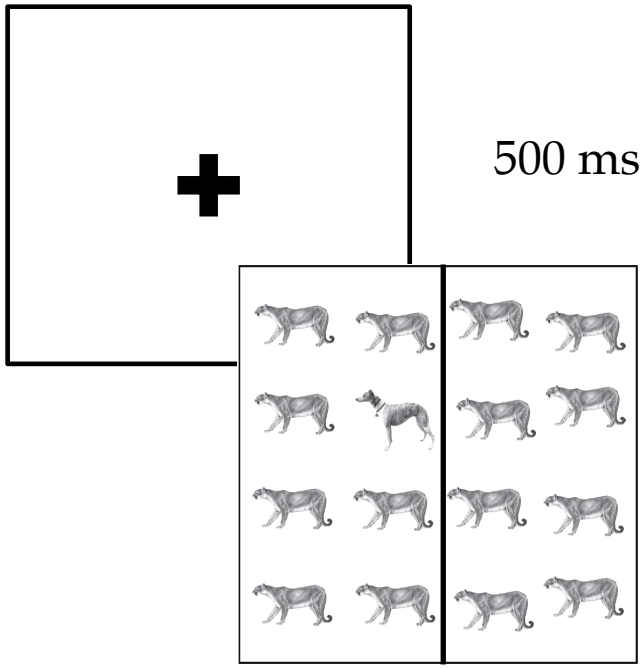






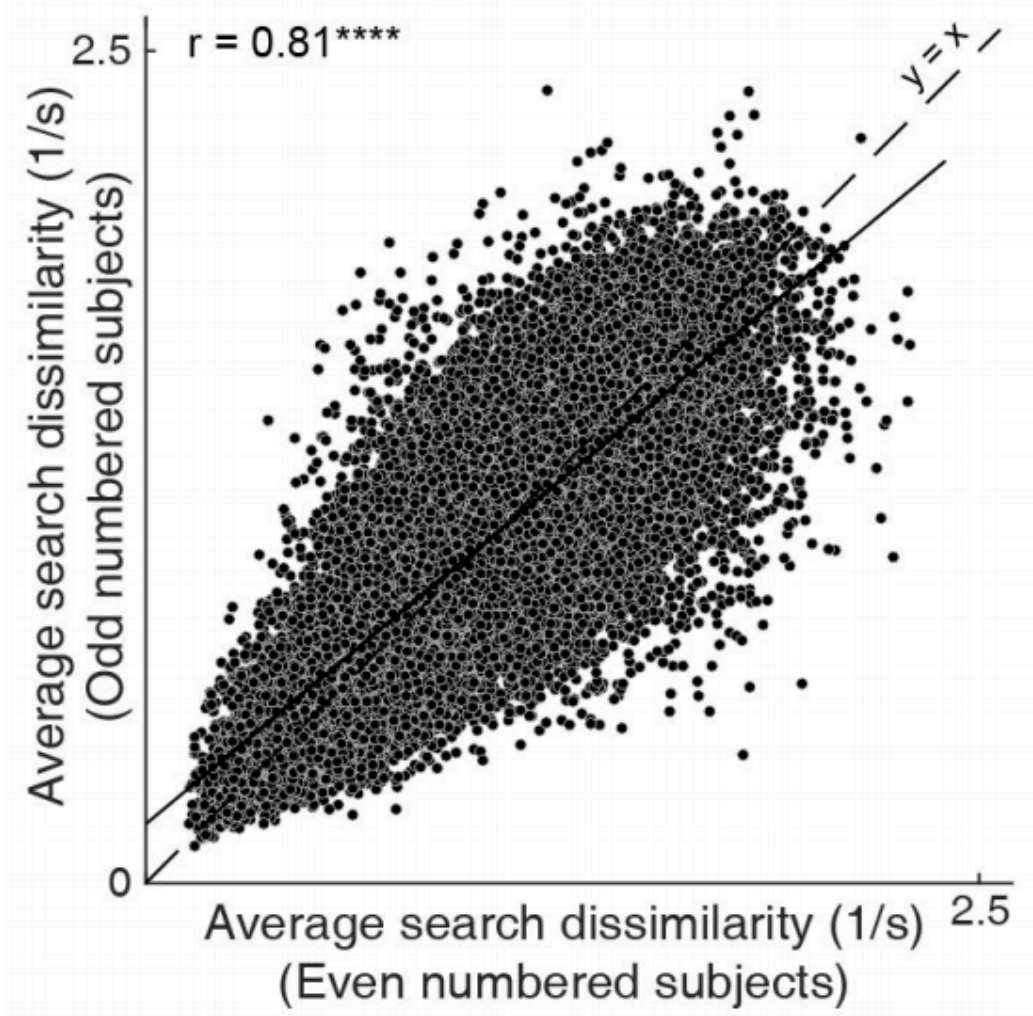
We collected a dataset of 2,801 objects (natural objects and silhouettes) and measured perceived dissimilarity for 26,675 pairs of objects across 269 human subjects!

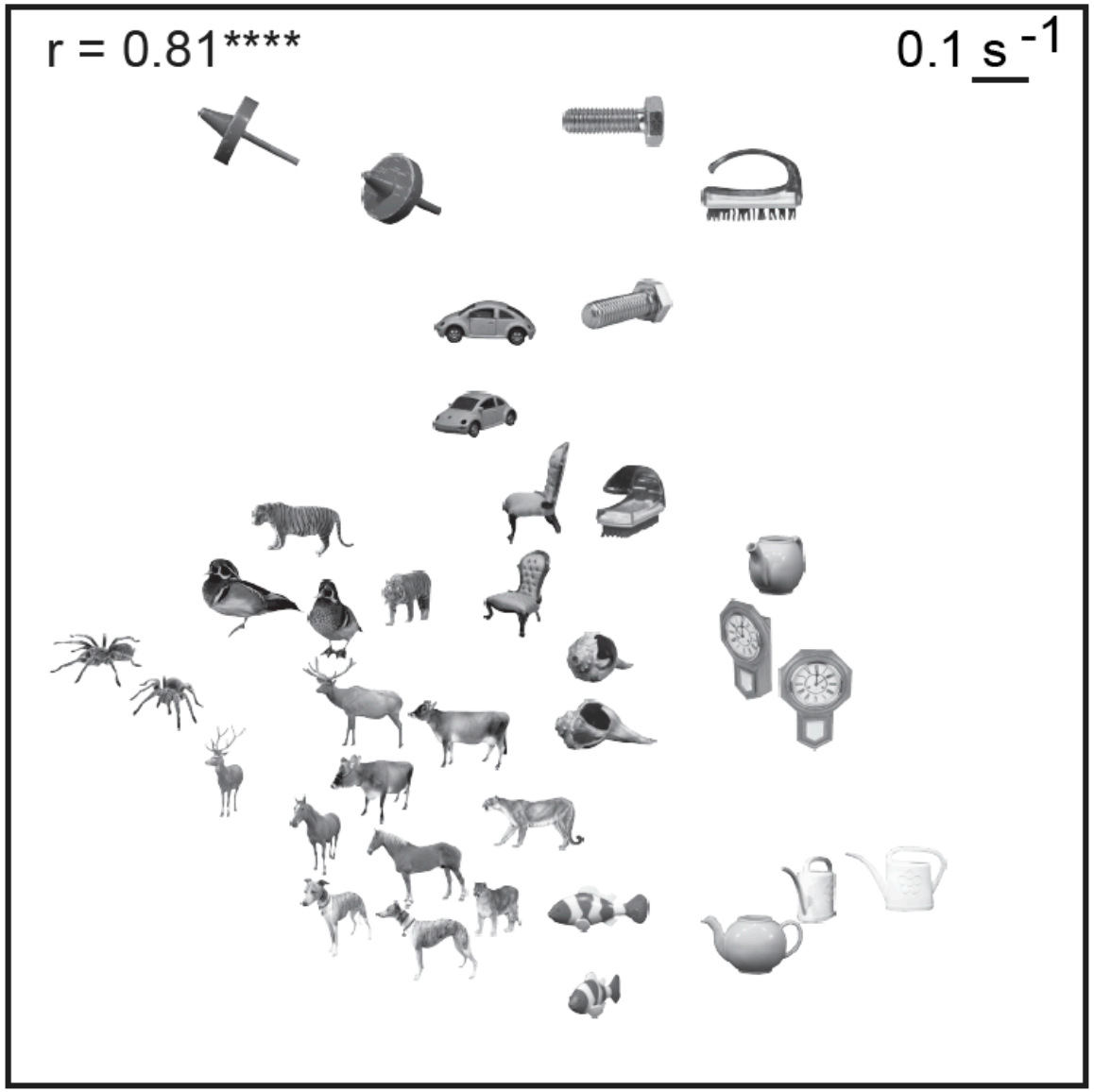




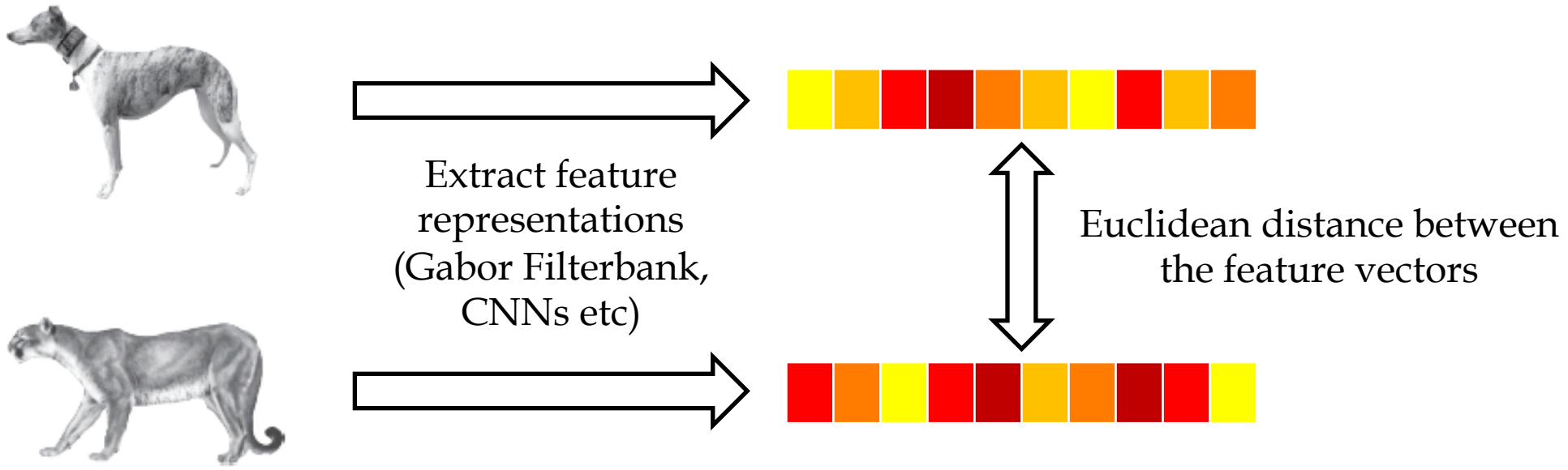
Reaction time, RT α similarity

1/RT α dissimilarity


r = 0.81****
y=x
2.5
0
Average search dissimilarity (1/s)
(Odd numbered subjects)
Average search dissimilarity (1/s)
(Even numbered subjects)
2.5


r = 0.81****
0.1 s -1

Estimating distances in machine vision models



We computed distances on 23 machine vision models! (including VGG-16, GoogleNet, ResNet-50, -100, -150)

Perceptual distances in humans

Measured as the reciprocal of Reaction Time in a visual search task.

Perceived distances were measured on ~26,000 pairs of objects from 269 human participants.

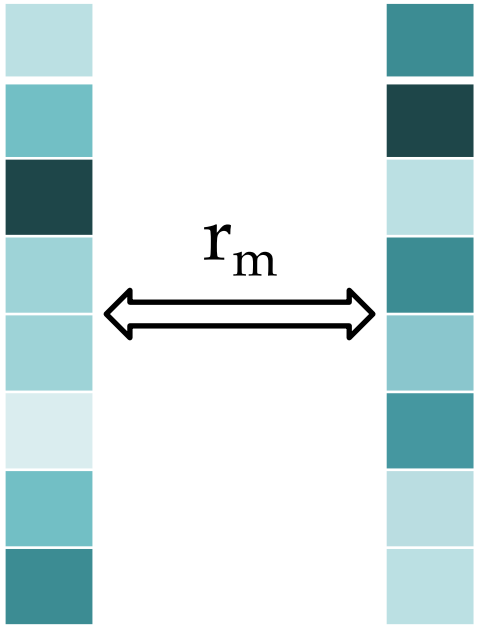
Distances in computer vision

Measured as the Euclidean distance between the feature vectors.

Feature distances were calculated on ~26,000 pairs of objects from 23 computer vision models (Fourier Descriptors, Gabor, SIFT, HoG, Gist, Convolutional Neural Networks)

These distances can be directly compared

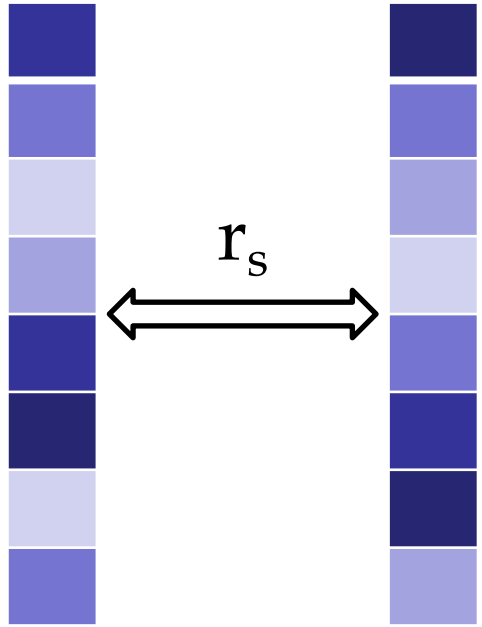
Comparing machine vision distances and human perceptual dissimilarities



r_m

Perceptual dissimilarity (humans) 26675 x 1

Feature distances (models) 26675 x 1



r_s

Perceptual dissimilarity (even subjects) 26675 x 1

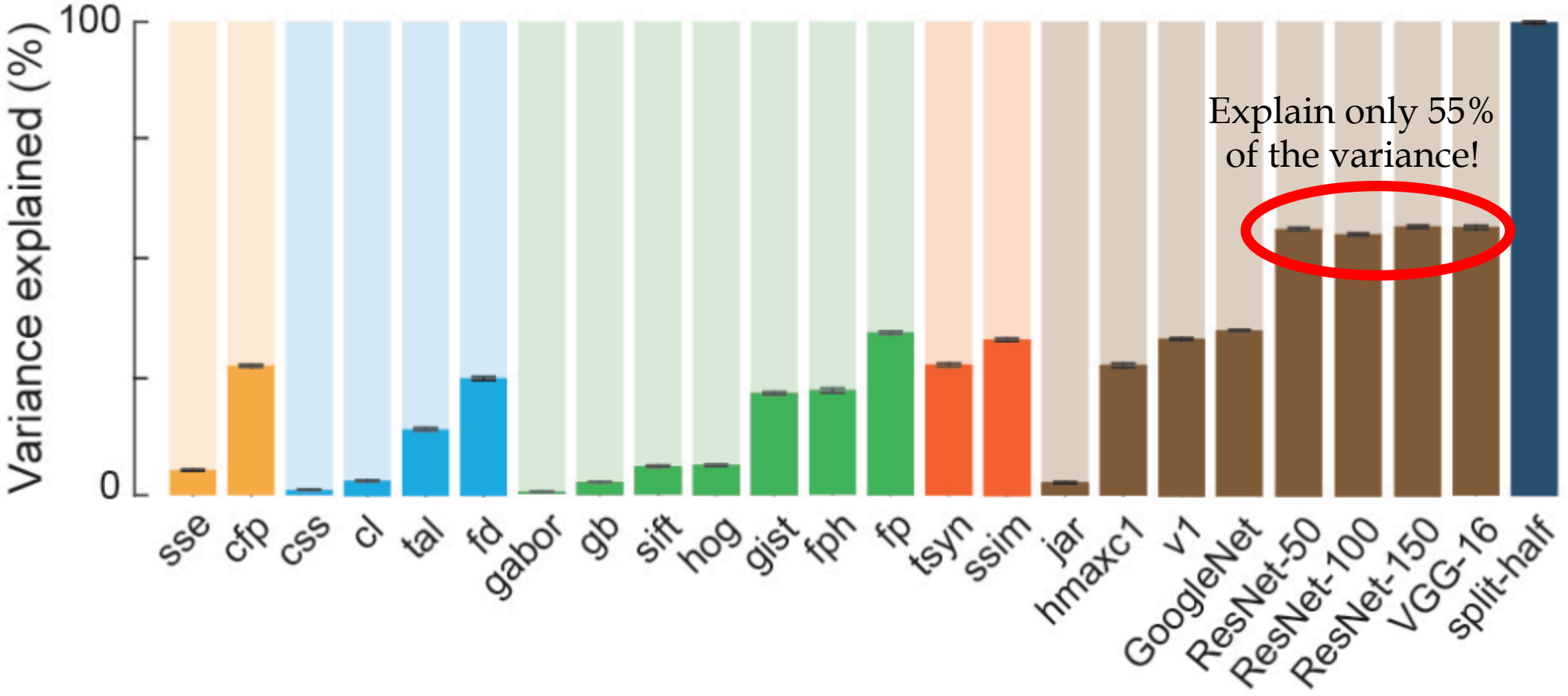
Perceptual dissimilarity (odd subjects) 26675 x 1

Spearman-Brown corrected split-half correlation

$$r_c = \frac{2r_s}{1 + r_s}$$

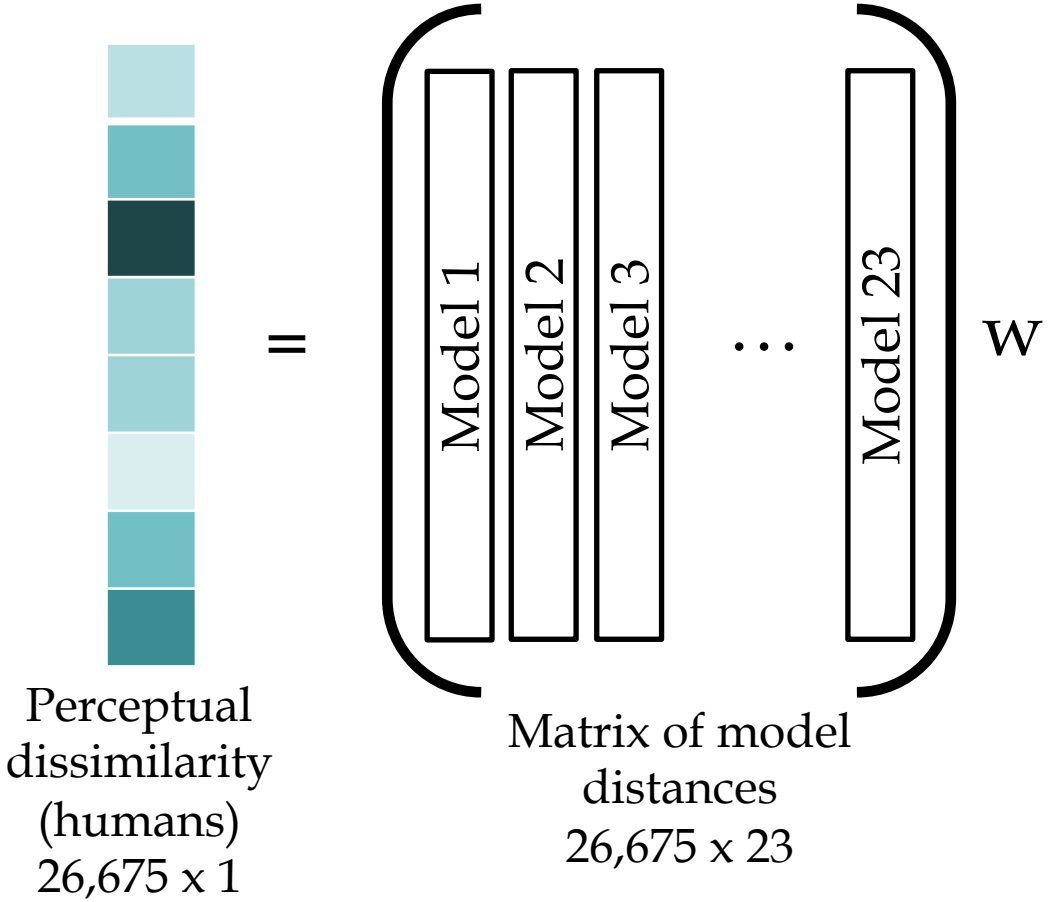
$$\%\ variance\ explained = \left(\frac{r_m}{r_c}\right)^2$$

Evaluating machine vision models on human perceptual dissimilarities

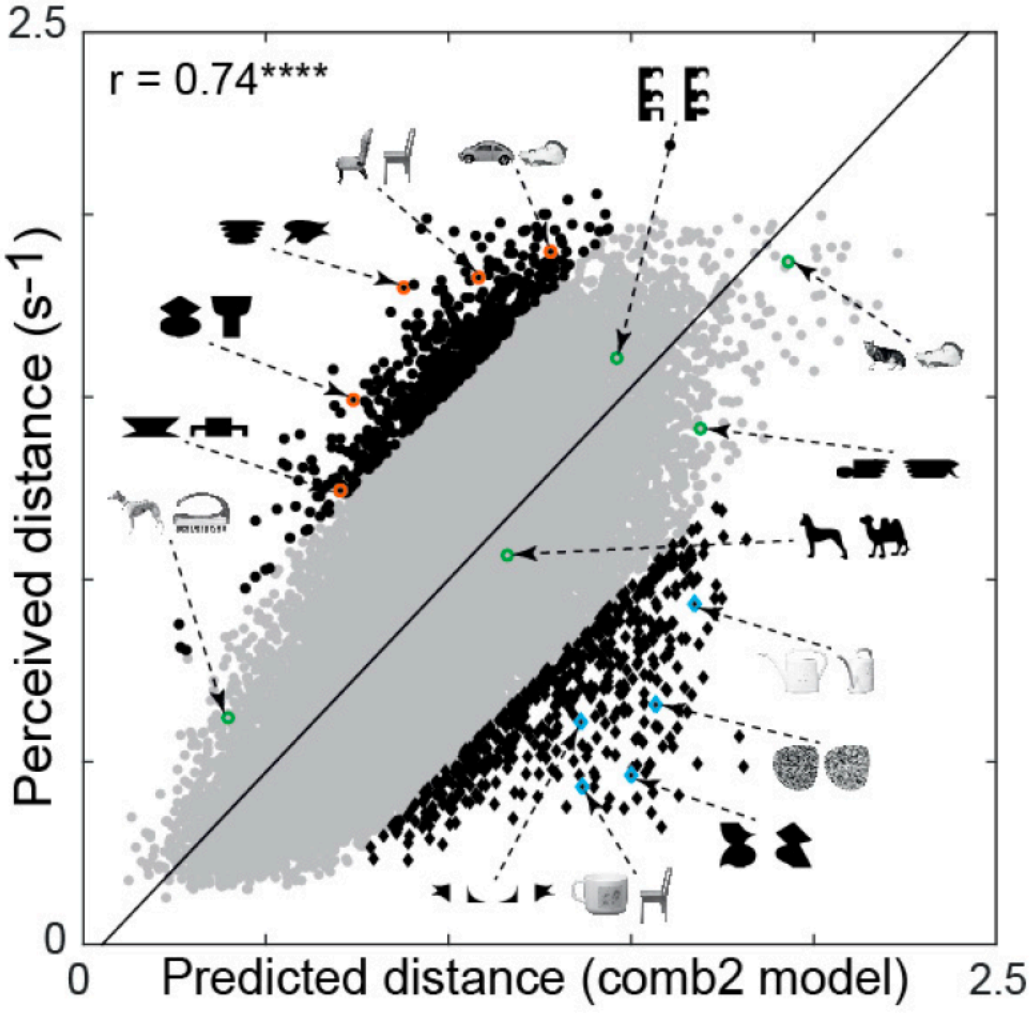


None of the machine vision models explain all the variance in human perceptual dissimilarities!

Combining distances from all machine vision models



comb2 model



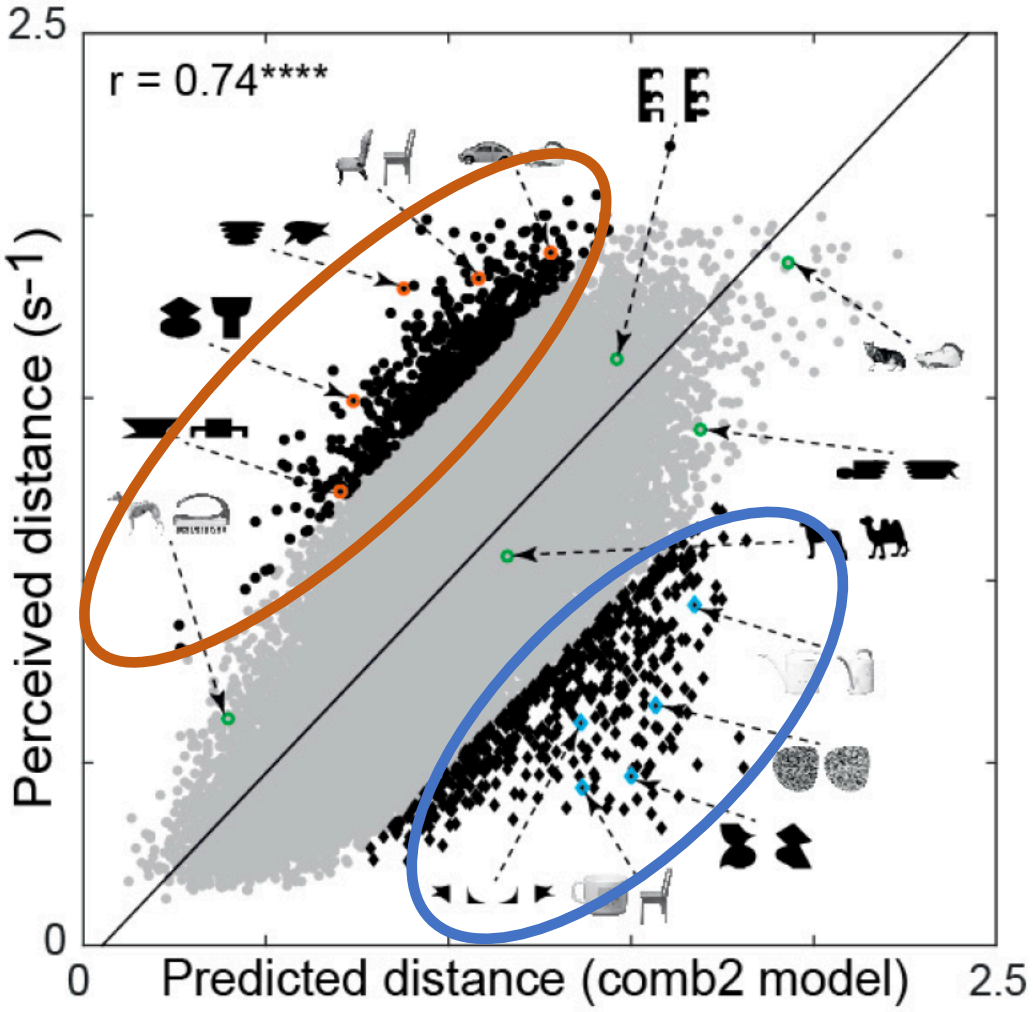
68.1% variance explained

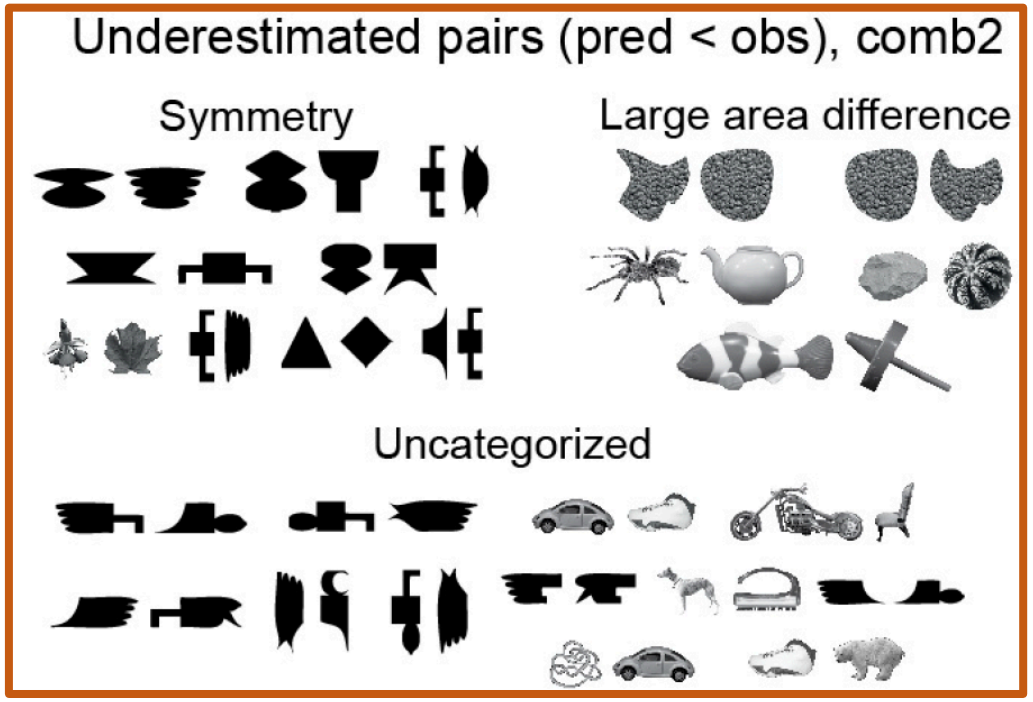
Questions

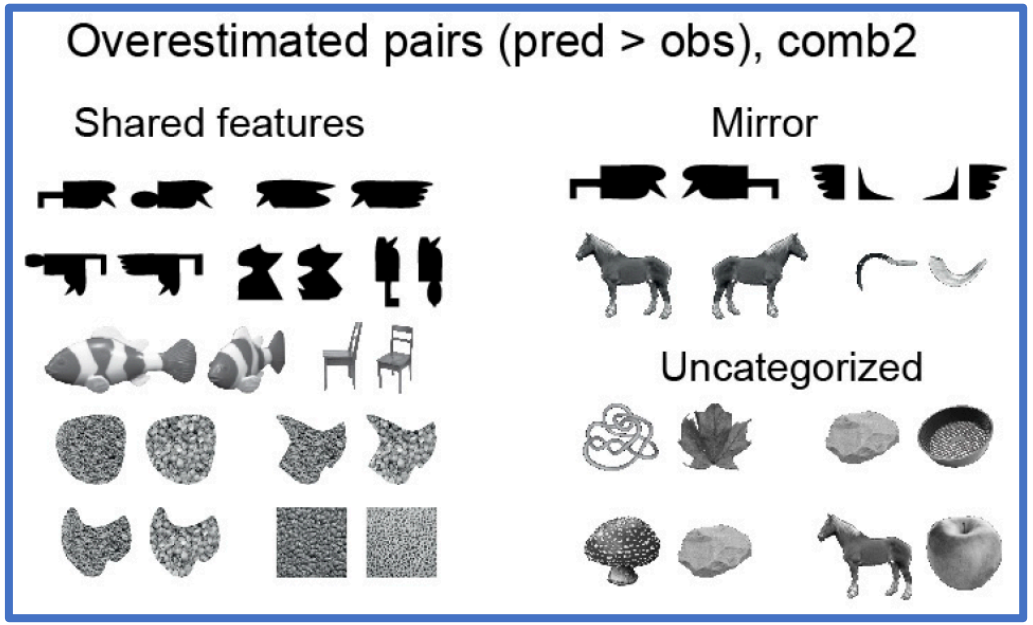
- How closely do machine vision representations match human perception?
- Deep neural networks explain more than half the variance in the human perceptual data with a combined model explaining ~68% of the variance.
- Do machine vision models deviate systematically from human perception?
- Can we improve machine vision models using human perception?

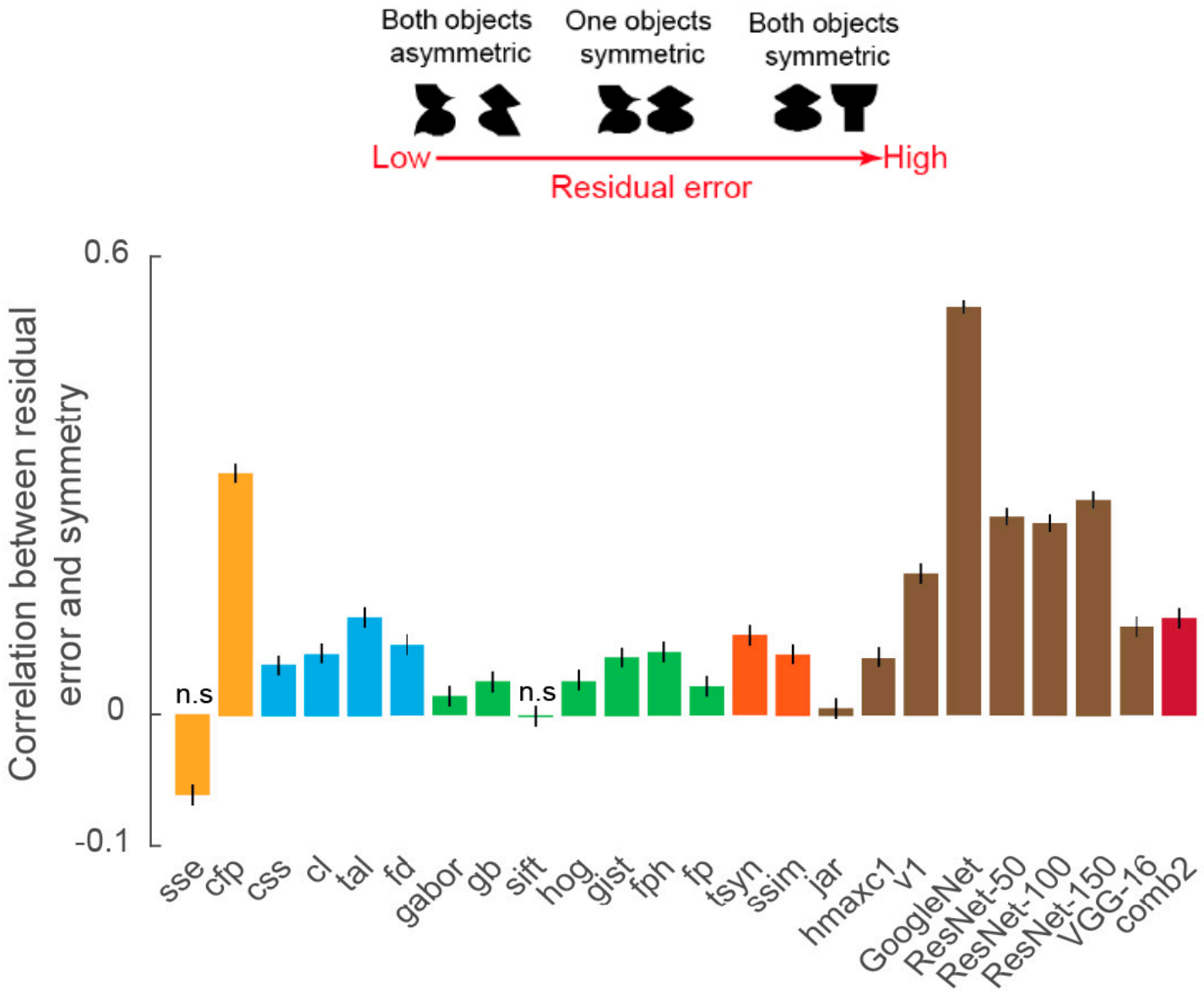
Questions

- How closely do machine vision representations match human perception?
- Deep neural networks explain more than half the variance in the human perceptual data with a combined model explaining ~68% of the variance.
- Do machine vision models deviate systematically from human perception?
- Can we improve machine vision models using human perception?


r = 0.74****
2.5
0
Perceived distance (s-1)
0
Predicted distance (comb2 model)
2.5


Underestimated pairs (pred < obs), comb2
Symmetry
Large area difference
Uncategorized

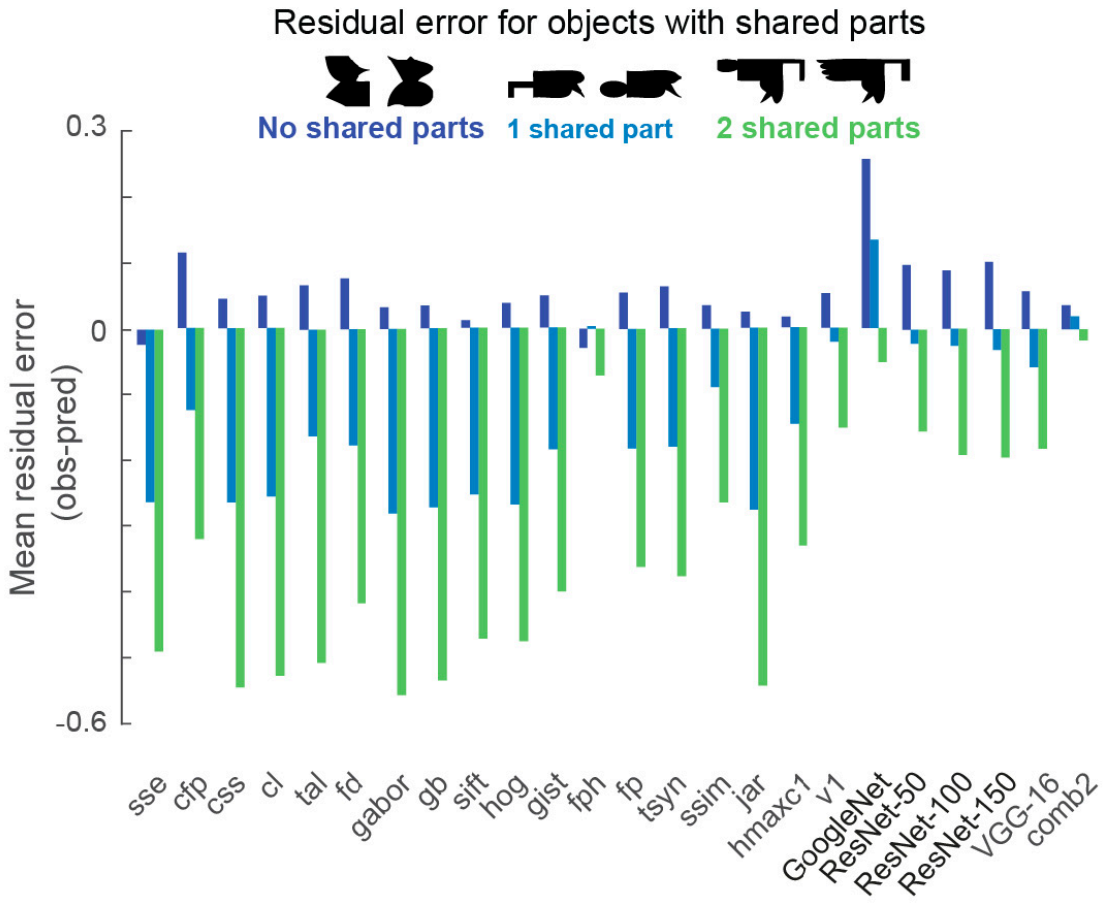

Overestimated pairs (pred > obs), comb2
Shared features
Mirror
Uncategorized

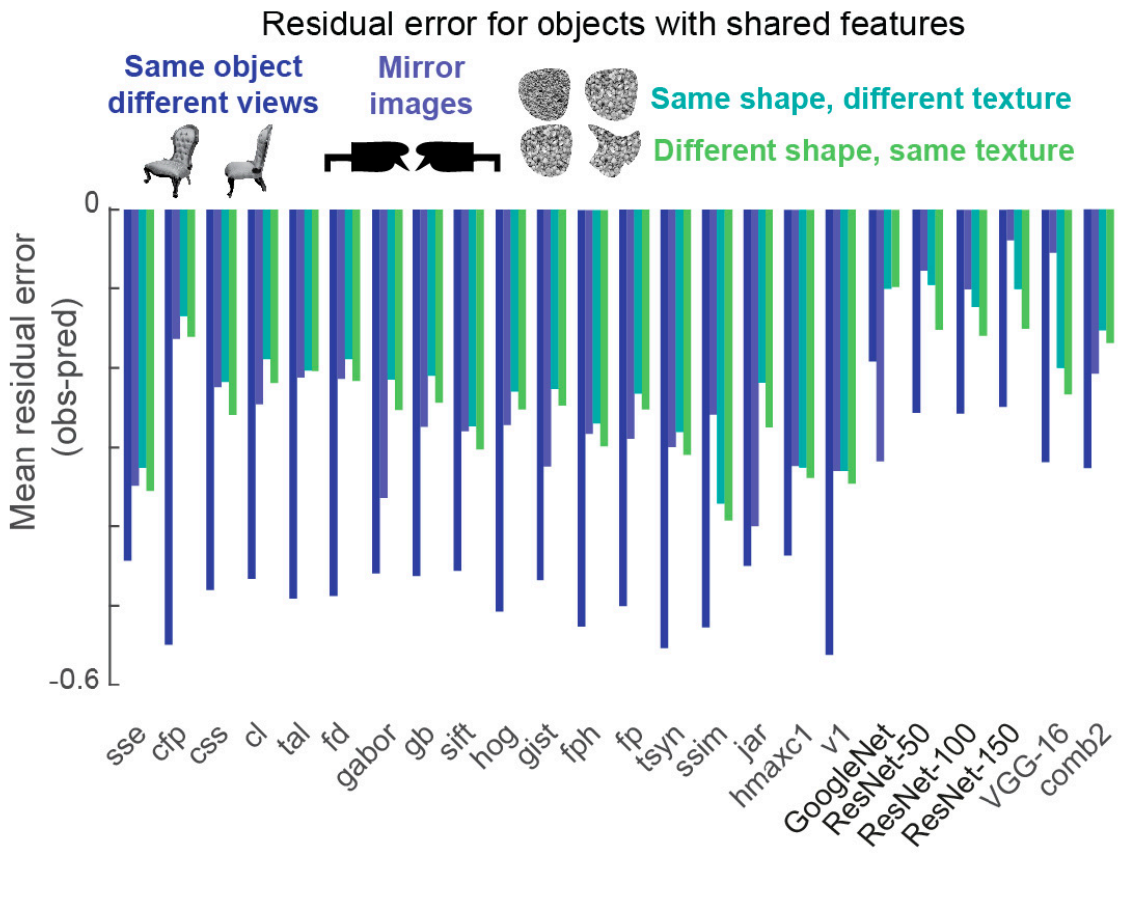


$$S_v = 1 - \frac{\Sigma \text{abs}(A - \text{flipv}(A))}{\Sigma \text{abs}(A + \text{flipv}(A))}$$

$S_{sym} = \max(S_v, S_h)$

Symmetric objects are more distinctive in perception than in machine vision models!





Objects with shared features or parts are more similar in perception than in most machine vision models!

Questions

- How closely do machine vision representations match human perception?
- Deep neural networks explain more than half the variance in the human perceptual data with a combined model explaining ~68% of the variance.

- Do machine vision models deviate systematically from human perception?
- Yes! Specifically, symmetric objects are more distinctive in human perception than in most machine vision models.

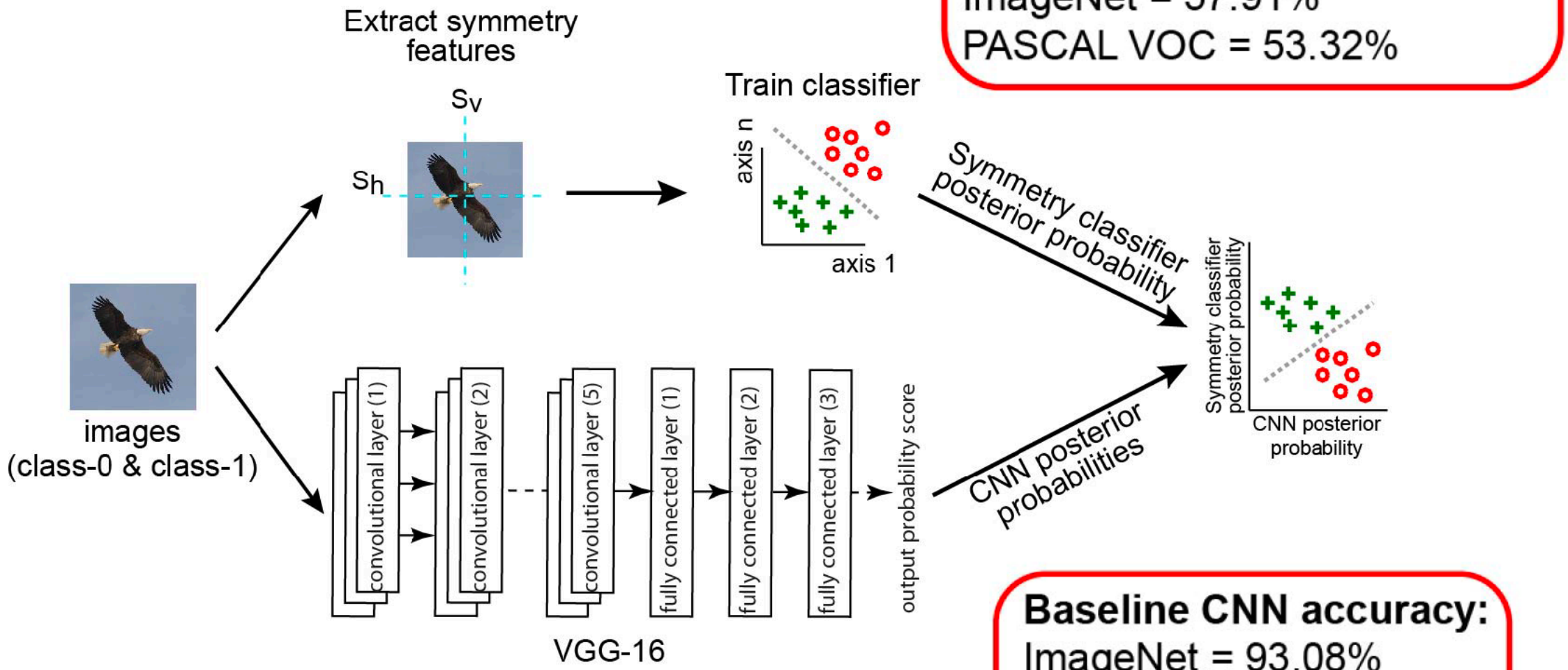
- Can we improve machine vision models using human perception?

Questions

- How closely do machine vision representations match human perception?
- Deep neural networks explain more than half the variance in the human perceptual data with a combined model explaining ~68% of the variance.

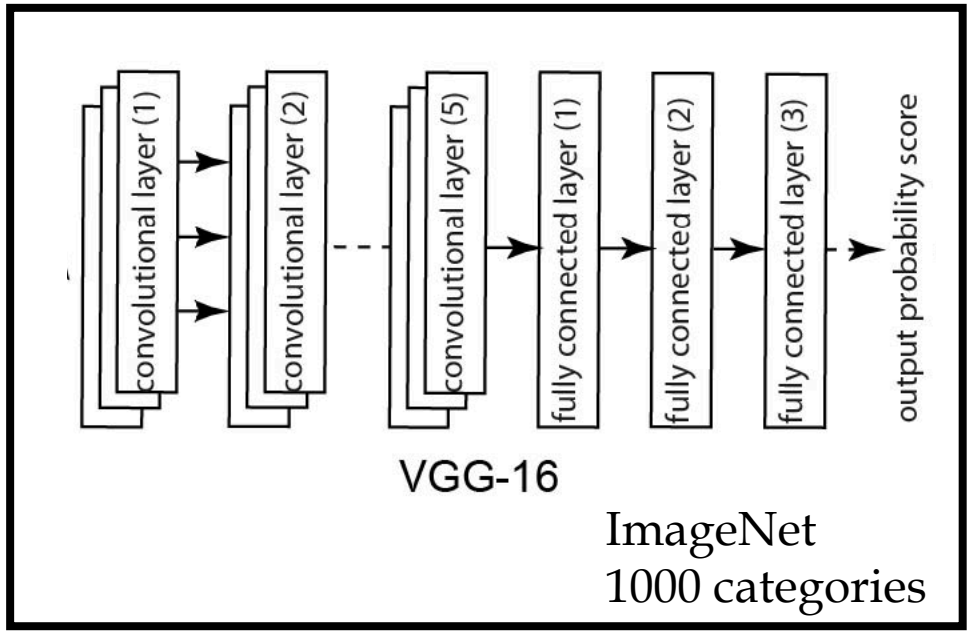
- Do machine vision models deviate systematically from human perception?
- Yes! Specifically, symmetric objects are more distinctive in human perception than in most machine vision models.

- Can we improve machine vision models using human perception?

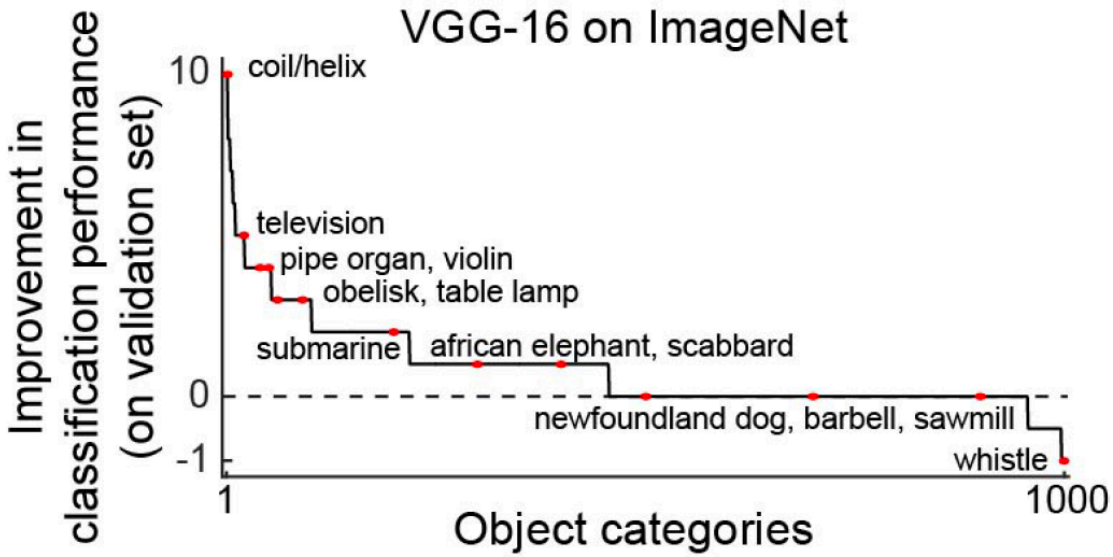


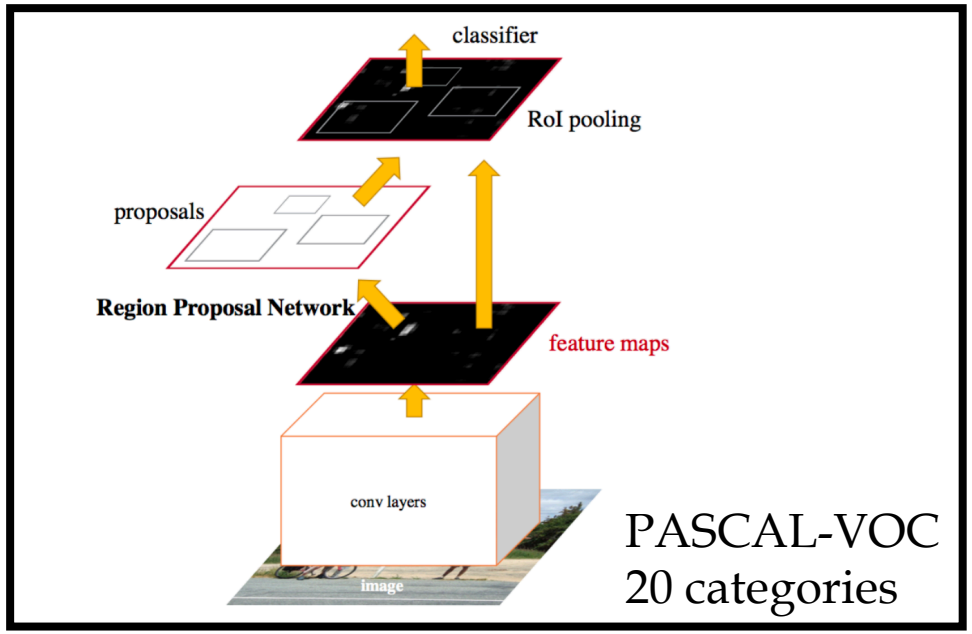
Symmetry classifier accuracy:
ImageNet = 57.91%
PASCAL VOC = 53.32%

Baseline CNN accuracy:
ImageNet = 93.08%
PASCAL VOC = 86.41%

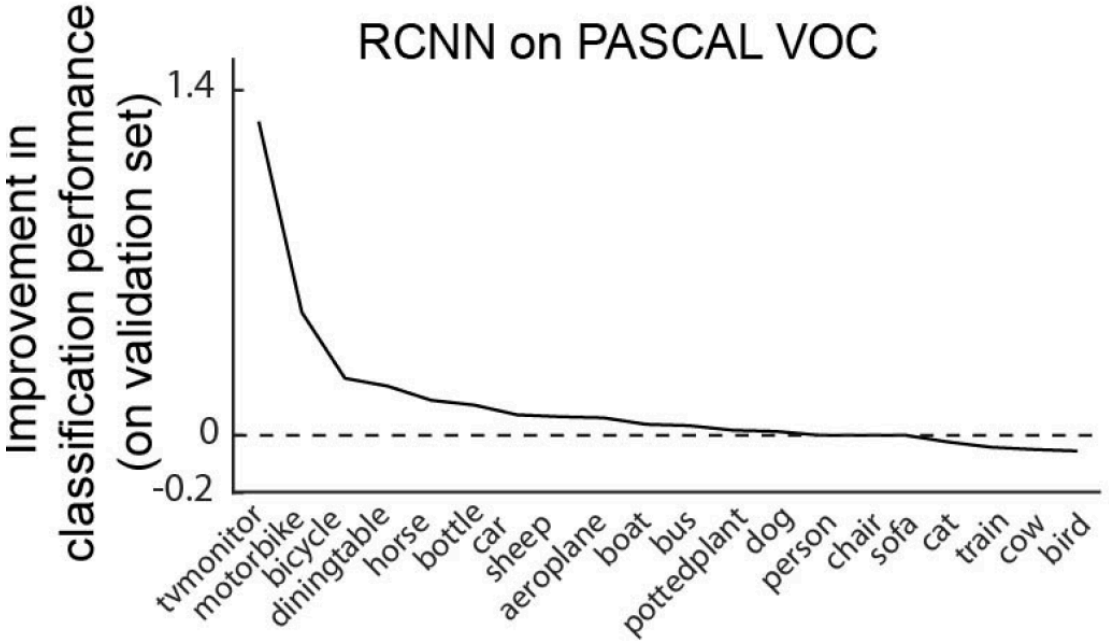


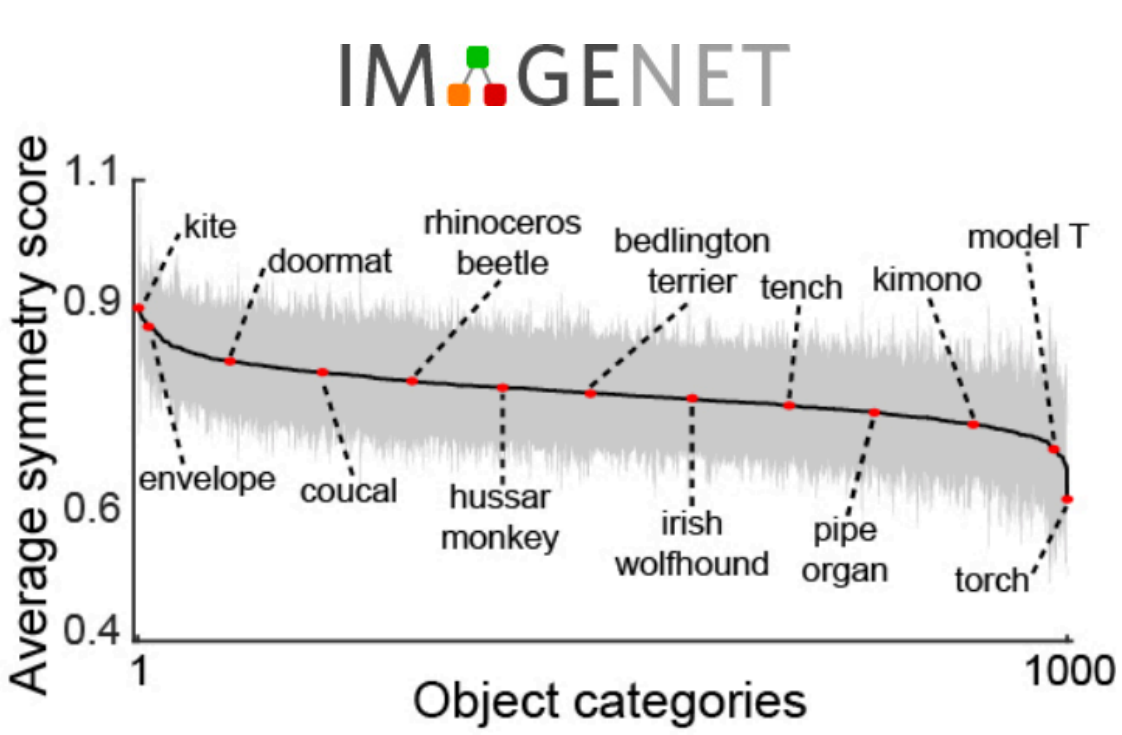
ImageNet
1000 categories

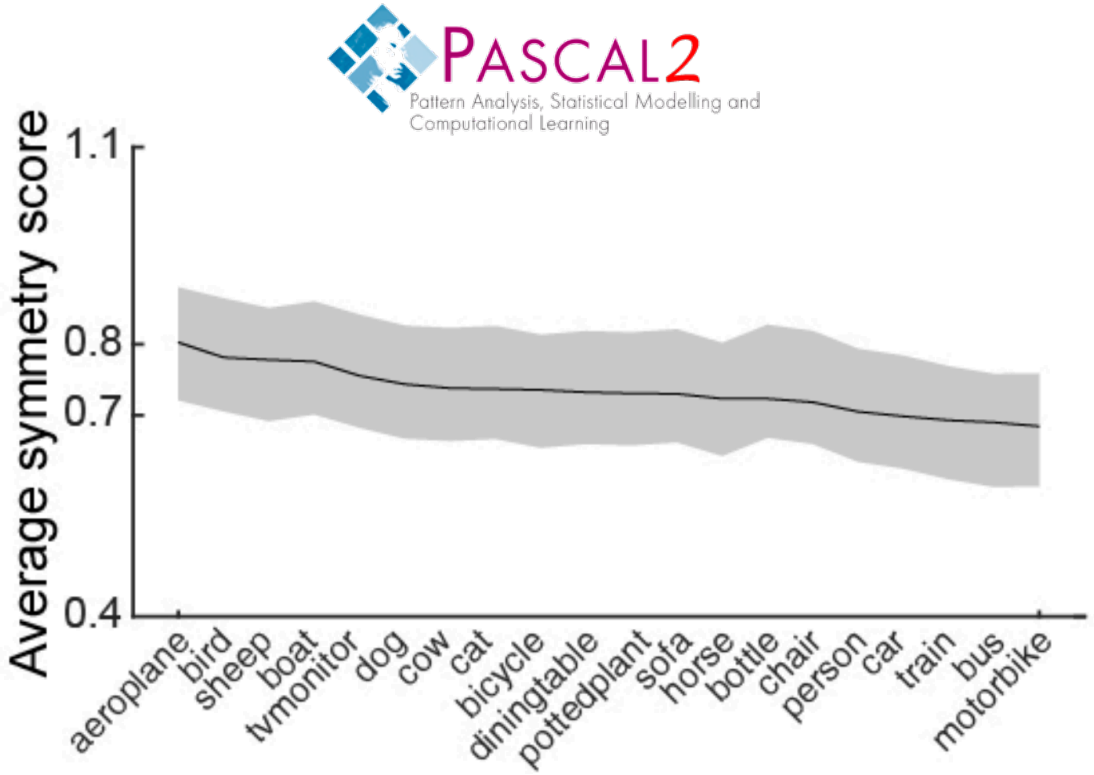




PASCAL-VOC
20 categories





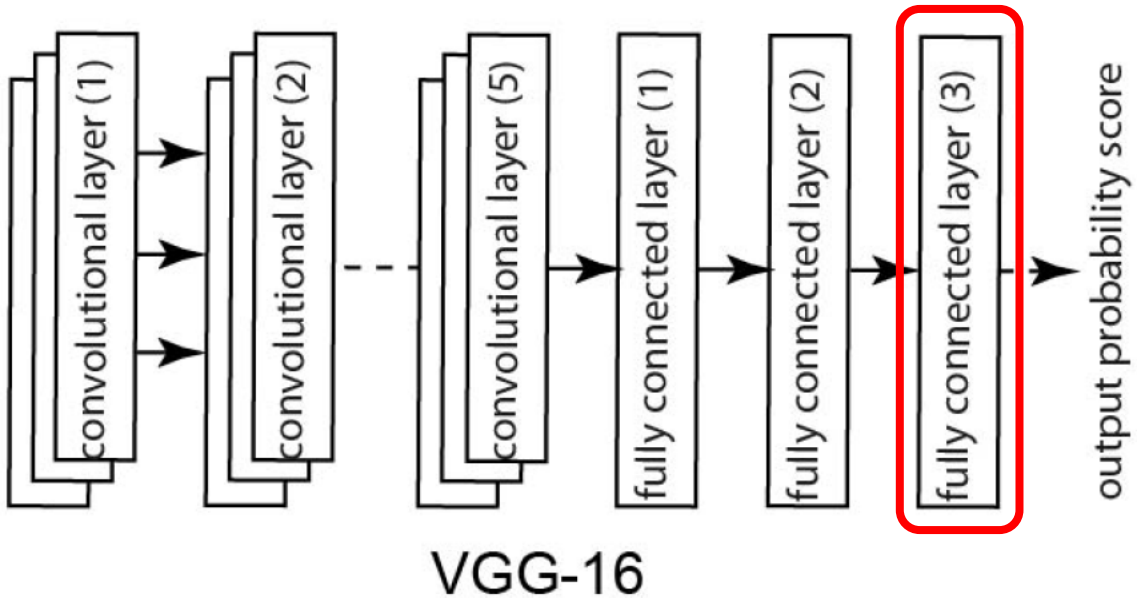


Lesser gains in performance on PASCAL-VOC dataset due to the images being less symmetric.

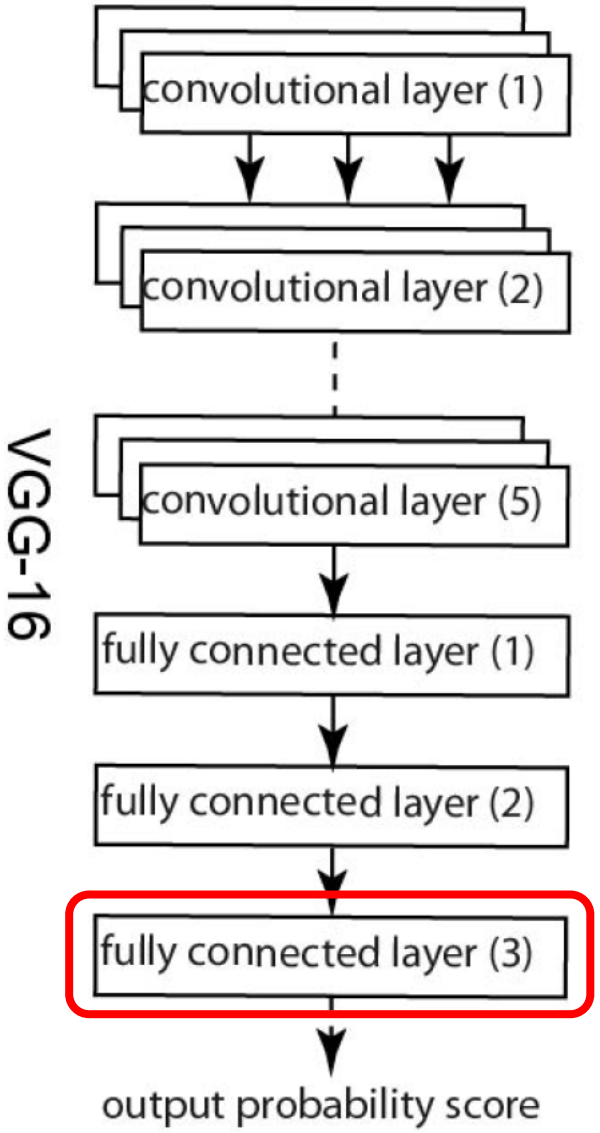
In general, the augmentation procedure can lead to significant gains in performance depending on the biases present in the dataset.

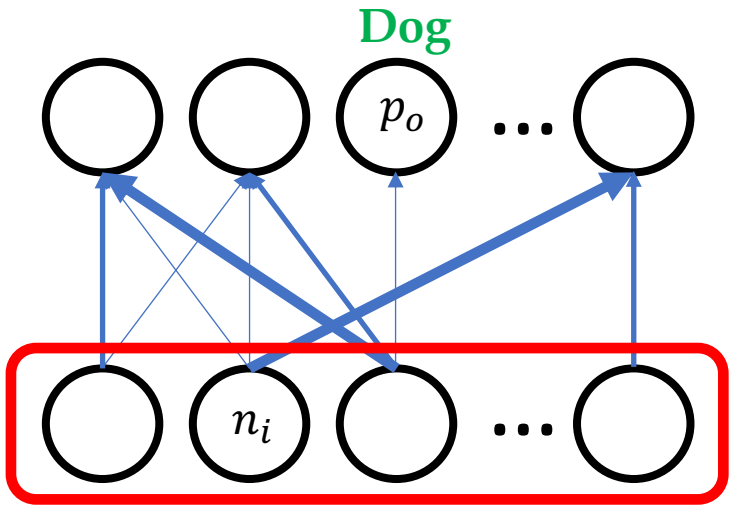
Why did this augmentation procedure work?

We hypothesize that the more important units in the penultimate fully-connected layer in the deep neural network for object recognition have weaker representation of symmetry.

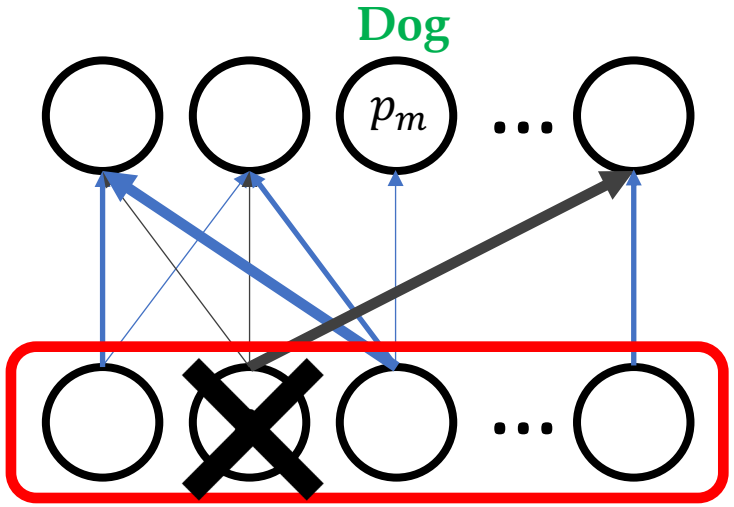


VGG-16





Normal network



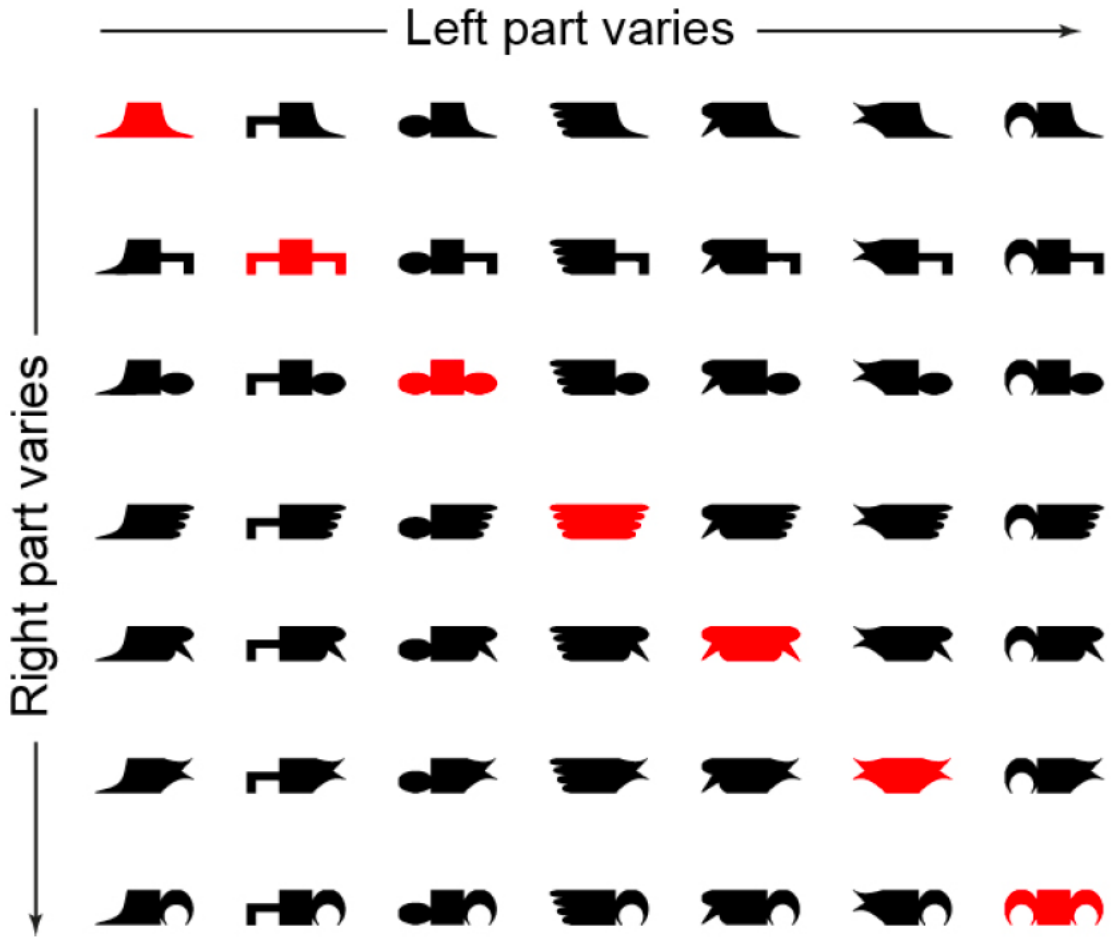
Modified (ablated) network

Node importance

$$\delta(n_i) = \frac{1}{20}\sum_{j=1}^{20}\left(p_o(c_j) - p_m(c_j)\right)$$

Nodes with large values of δ are considered to be more important for classification

Quantifying symmetry representation



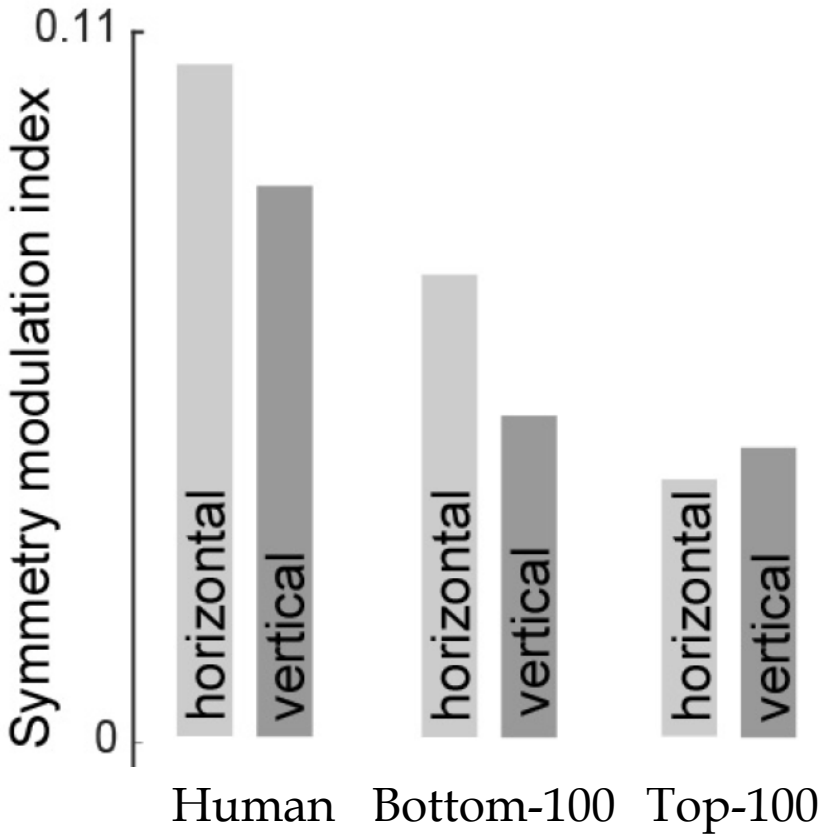
Perception

Symmetry Modulation Index

$$SMI = \frac{d_{sym} - d_{asym}}{d_{sym} + d_{asym}}$$

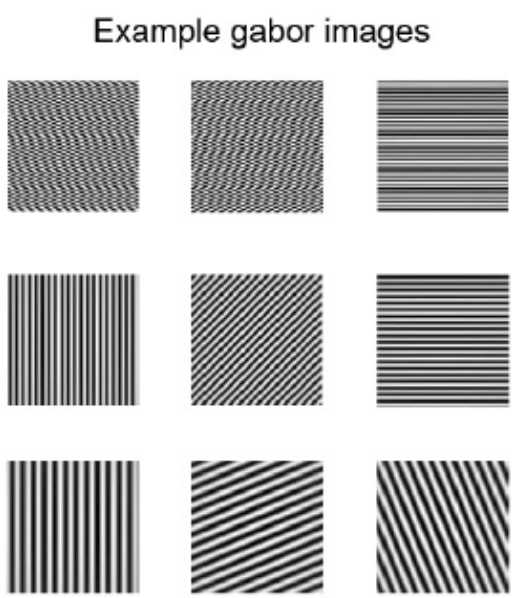
Used both horizontal and vertical versions of these objects

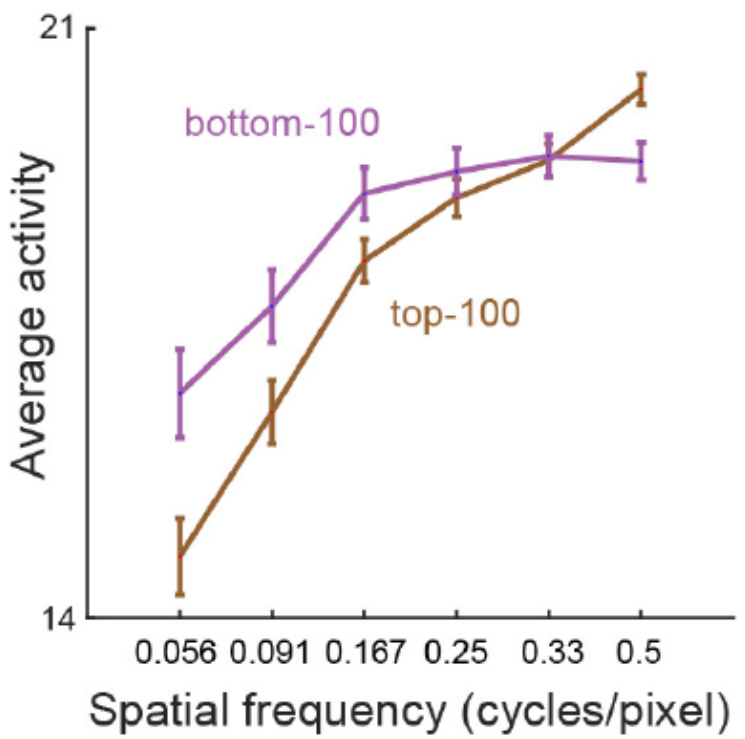
Symmetry Modulation Index

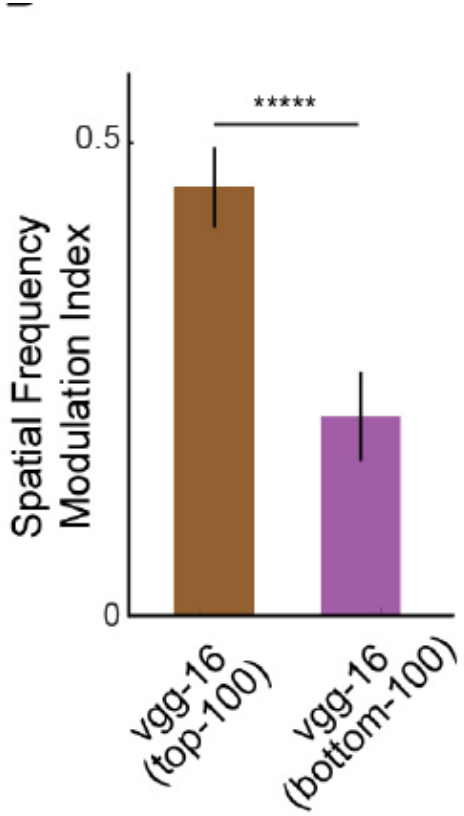


As hypothesized, the units more important for object classification have weaker representation of symmetry!

Different feature representation in the top-100 and bottom-100 units







Top-100 units respond more to higher spatial frequency content

Questions

- How closely do machine vision representations match human perception?
- Deep neural networks explain more than half the variance in the human perceptual data with a combined model explaining ~68% of the variance.

- Do machine vision models deviate systematically from human perception?
- Yes! Specifically, symmetric objects are more distinctive in human perception than in most machine vision models.

- Can we improve machine vision models using human perception?
- Yes! Object recognition performance of a CNN can be improved by augmenting with symmetry related information.

Summary

- All machine vision models (including deep neural networks) show systematic biases from human perception.
- Fixing one such bias (symmetry) led to improvements in object recognition performance of a widely used deep neural network.
- We further showed that the improvement was probably due to weak advantage for symmetry among units important for recognition.

SP Arun

Cracking the code

Vision Lab @ IISc

IISc-Dissimilarities between Individual Objects dataset

https://osf.io/q72cs/