Using Human Behavior and Brain Activity to Guide Machine Learning

Walter J. Scheirer

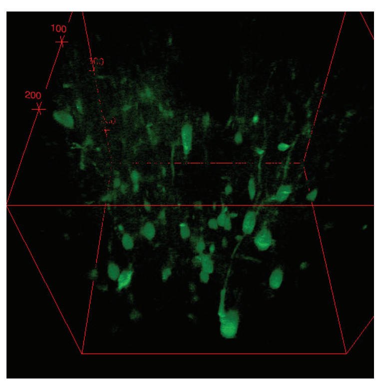
Computer Vision Research Laboratory,
Department of Computer Science and Engineering

UNIVERSITY OF NOTRE DAME

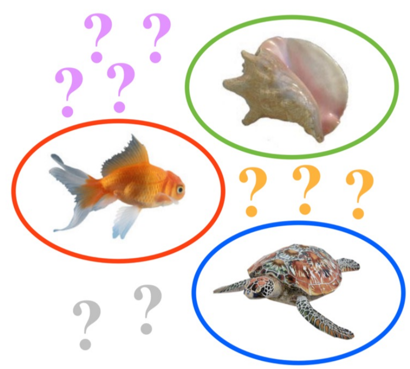
About me

- Joined Notre Dame Summer 2015
  - Ph.D. from the University of Colorado 2009
  - 2007 — 2012 Biometrics Startup Securics, Inc.
  - 2012 — 2015 Harvard University Center for Brain Science
- Research in Computer Vision and Machine Learning

Reverse engineering biological vision



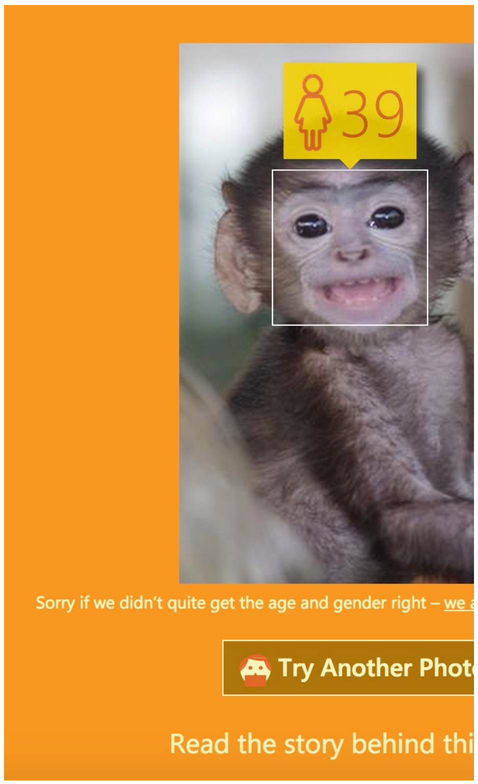
Tools for Neuroscience



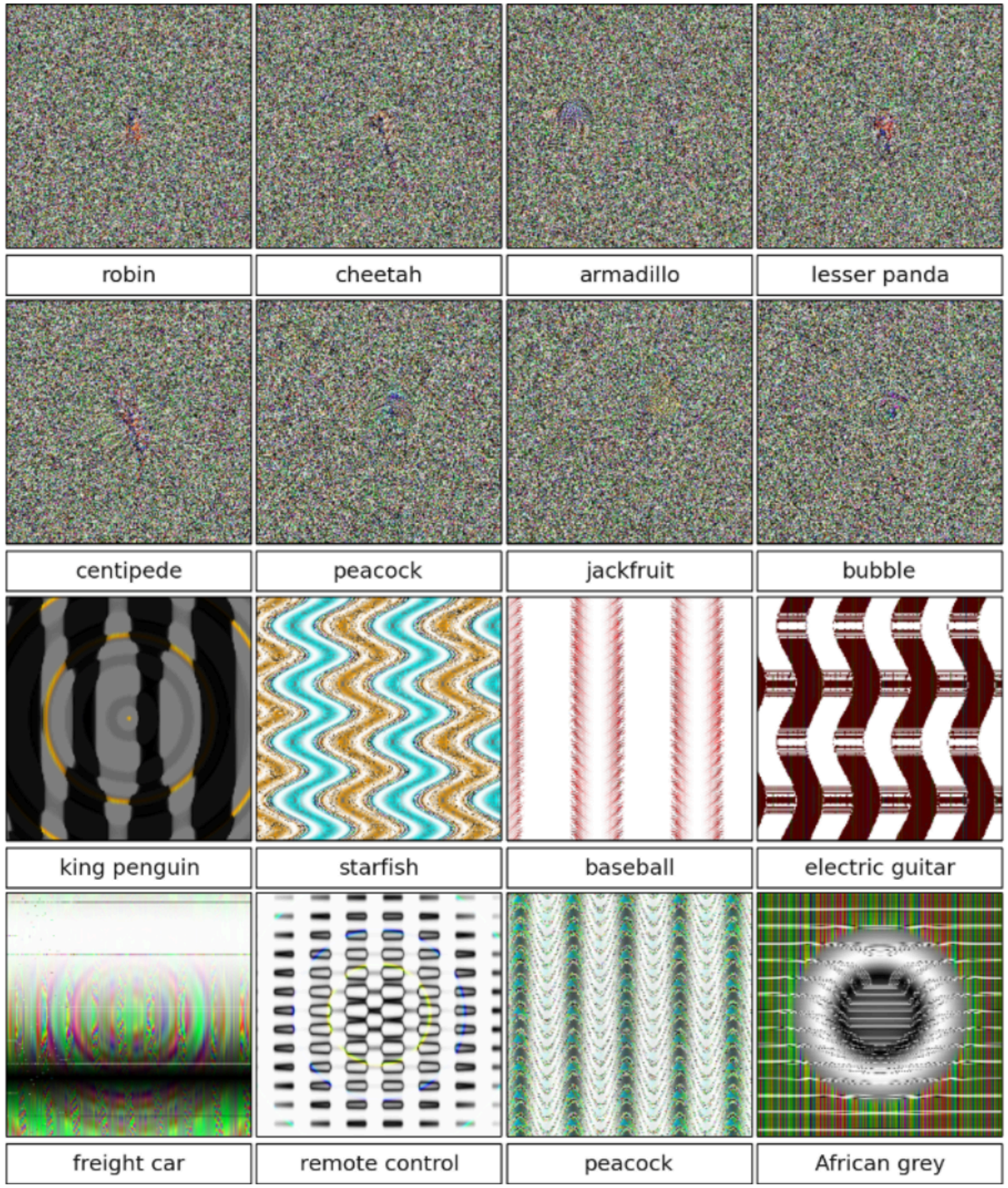
Statistical methods for visual recognition

Digital Humanities

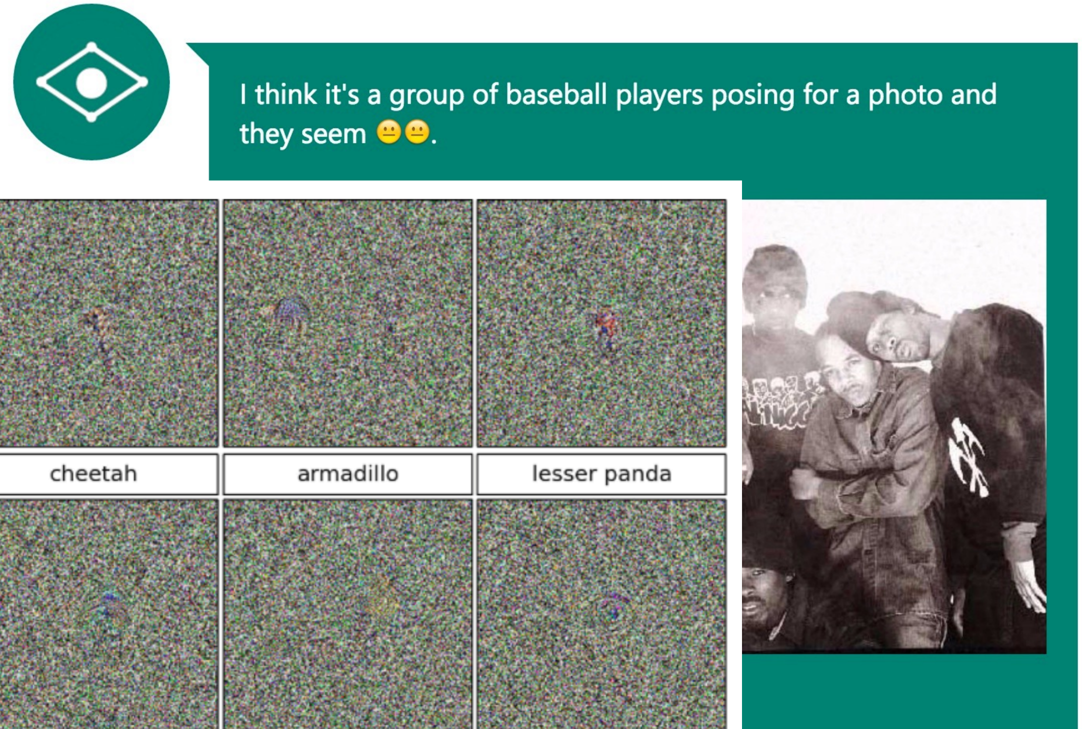
Trouble with Computer Vision…



https://how-old.net/



Nguyen et al. CVPR 2015

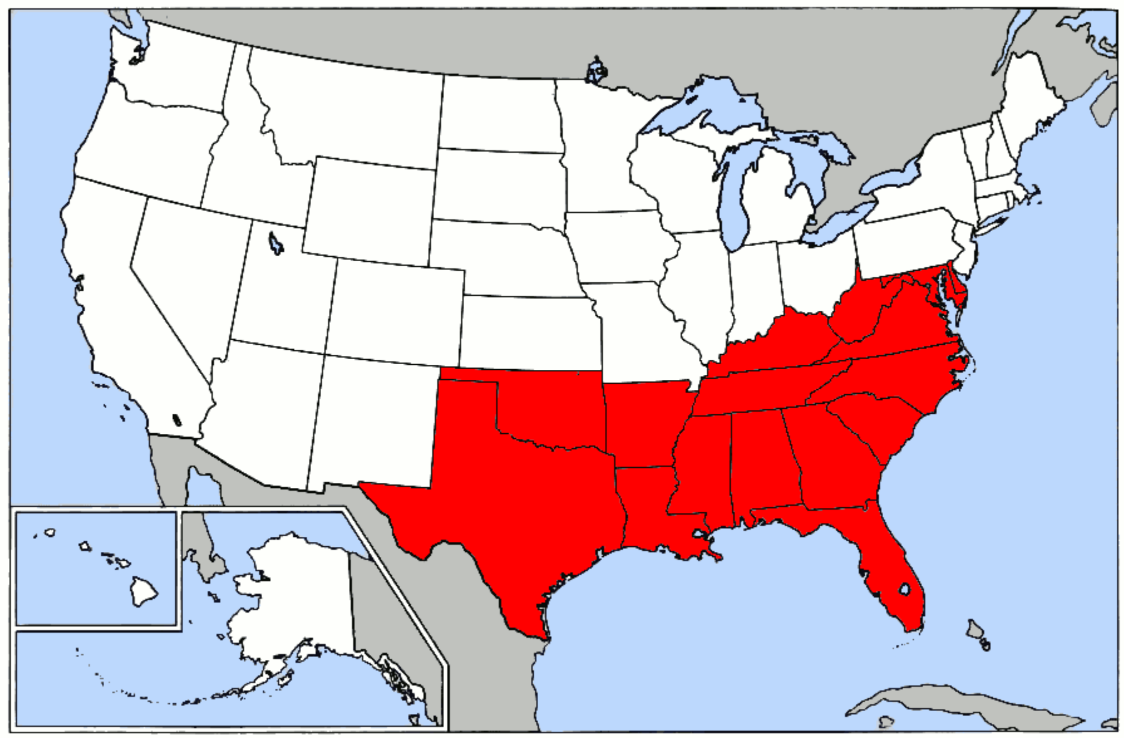


https://www.captionbot.ai/

Learnability

Imagine a newly arrived foreigner in the US…

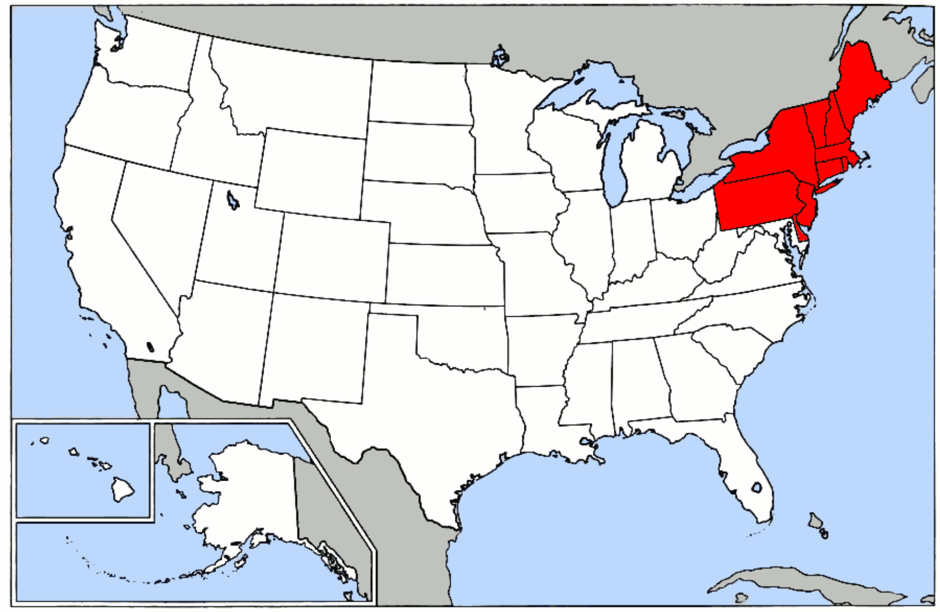
Could they recognize a person's origin based on their speech?



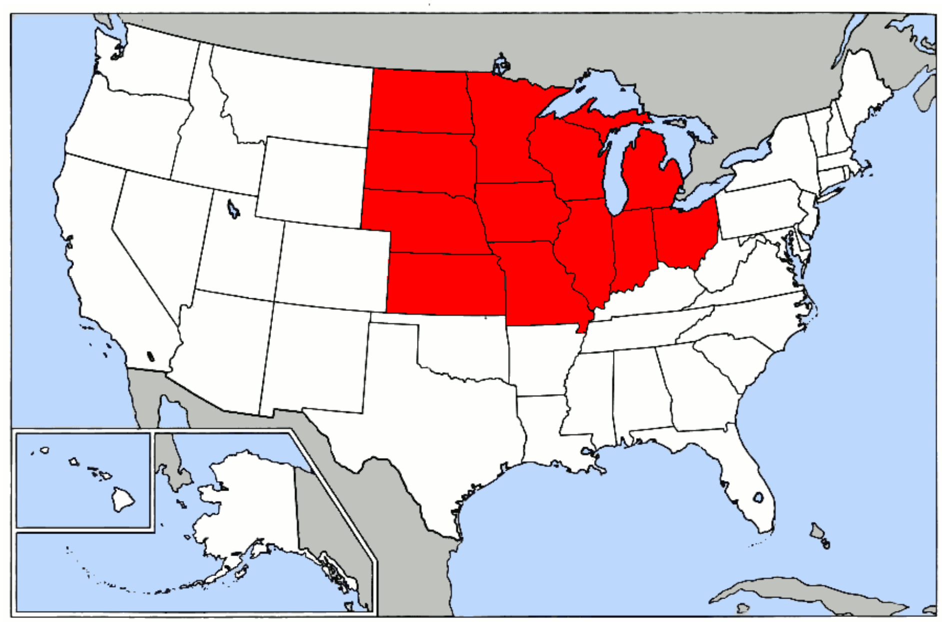
Map of USA Highlighting South

Learnability

What about the distinction between the Northeastern and the Mid-Western accents?



Map of USA Highlighting Northeast (cc) BY-SA 3.0 Wapcaplet



Map of USA Highlighting Midwest (cc) BY-SA 3.0 Wapcaplet

Learnability

Or the distinction between the people who originated from different parts of Brooklyn?

Brooklyn neighborhoods map 🅭 BY-SA 3.0 Peter Fitzgerald

The Practice of Teaching

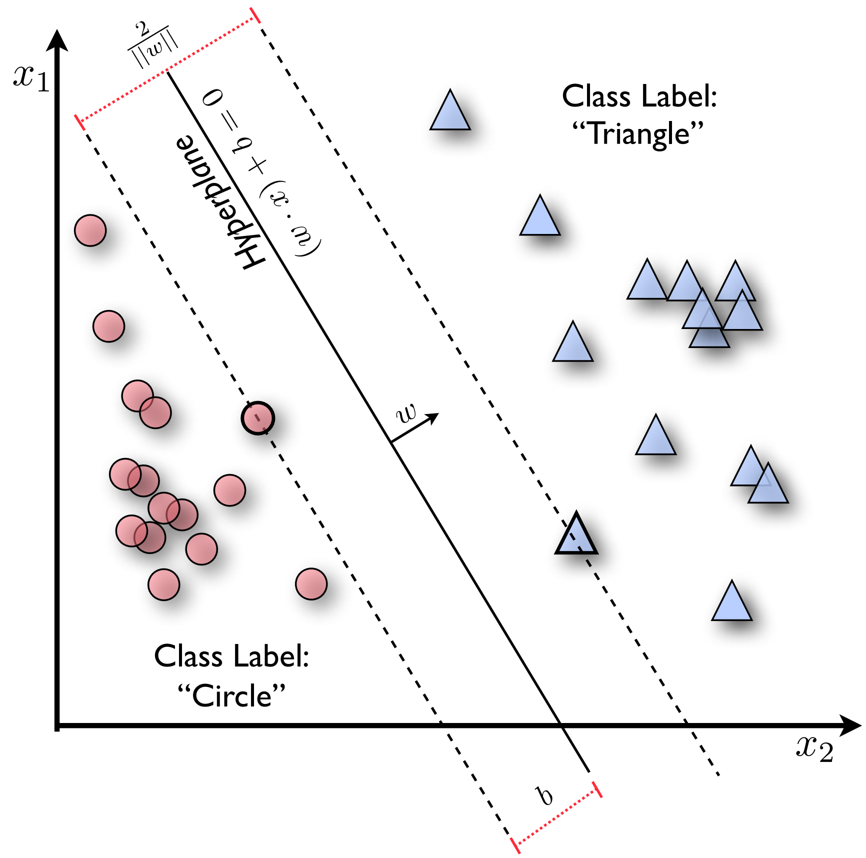
How would we teach a new arrival to identify accents?

1. Start with the easiest distinctions
2. Proceed with finer distinctions

We would never suggest that a novice learn all distinctions at the same time.

Supervised Learning

A “sink or swim” approach



No effort to tailor the learning to the human ability to learn from particular images.

Perceptual Annotation

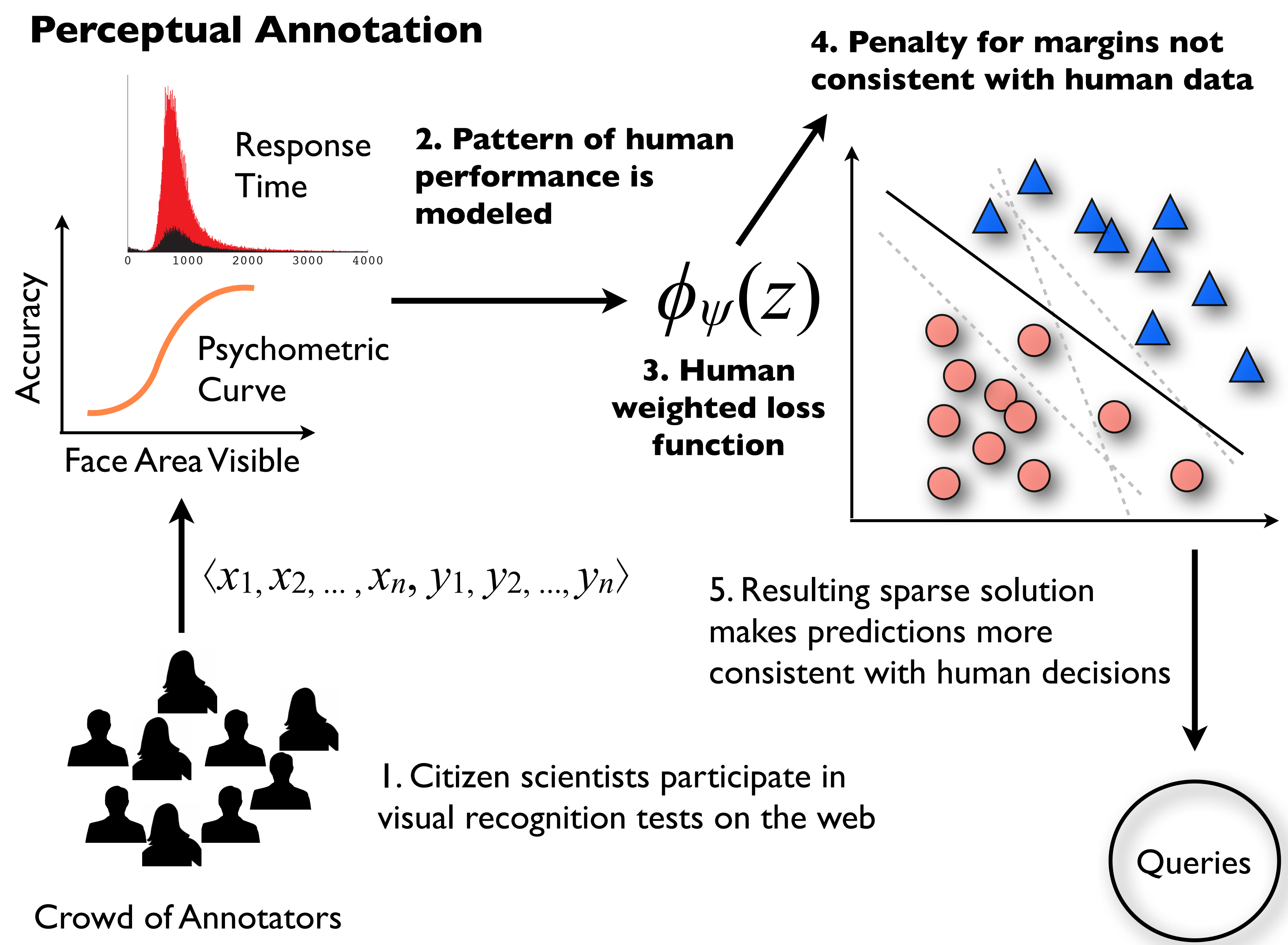
Much information about human capacities can be of direct value for machine learning:

Some images are learnable, and some are not.

Learnability varies with experience.

Some things are easily learned, other things take more time.

Such detailed information reflecting human capacity is what we call a perceptual annotation.



W. Scheirer, S. Anthony, K. Nakayama, and D.D. Cox, "Perceptual Annotation: Measuring Human Vision to Improve Computer Vision," IEEE T-PAMI, vol. 36, no. 8, August 2014.

Related Work:

Active Learning: B. Settles. Active Learning. Morgan & Claypool, 2012.

Useful Bias Estimation: C. Vondrick, H. Pirsiavash, A. Oliva, and A. Torralba, “Learning Visual Biases From Human Imagination,” NIPS, 2015.

Human Annotation Process Modeling: P. Welinder, S. Branson, S. Belongie, and P. Perona, “The Multidimensional Wisdom of Crowds,” NIPS, 2010.

Human vs. Computer Performance: R. Geirhos, D. H. J. Janssen, H. H. Schutt, J. Rauber, M. Bethge, and F. A. Wichmann, “Comparing Deep Neural Networks Against Humans: Object Recognition When the Signal Gets Weaker,”arXiv preprint1706.06969, 2017.

P. J. Phillips and A. J. O’Toole, “Comparison of human and computer performance across face recognition experiments,” Image and Vision Computing, vol. 32, no. 1, pp. 74–85, 2014.

Visual Psychophysics Using TestMyBrain.org

Visual Psychophysics

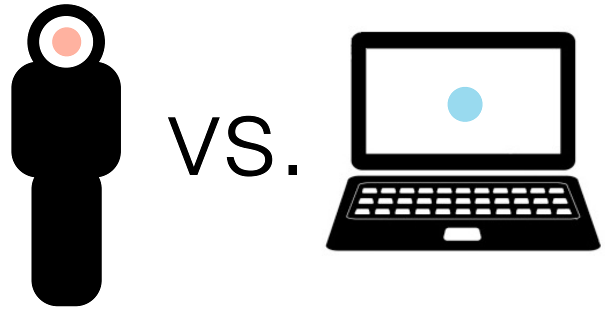
Probe psychological and perceptual thresholds through controlled manipulation of stimuli.

Careful management of stimulus construction, ordering and presentation allows for precise determination of perceptual thresholds.

Canonical Early Example*: minimum threshold for stimulation of an individual photoreceptor.

Sam Anthony
(Harvard Vision Lab)

S. Hecht, S. Shlaer, M. Pirenne, "Energy, quanta and vision," J. Gen. Physiol., 25 (1942), pp. 819–840.

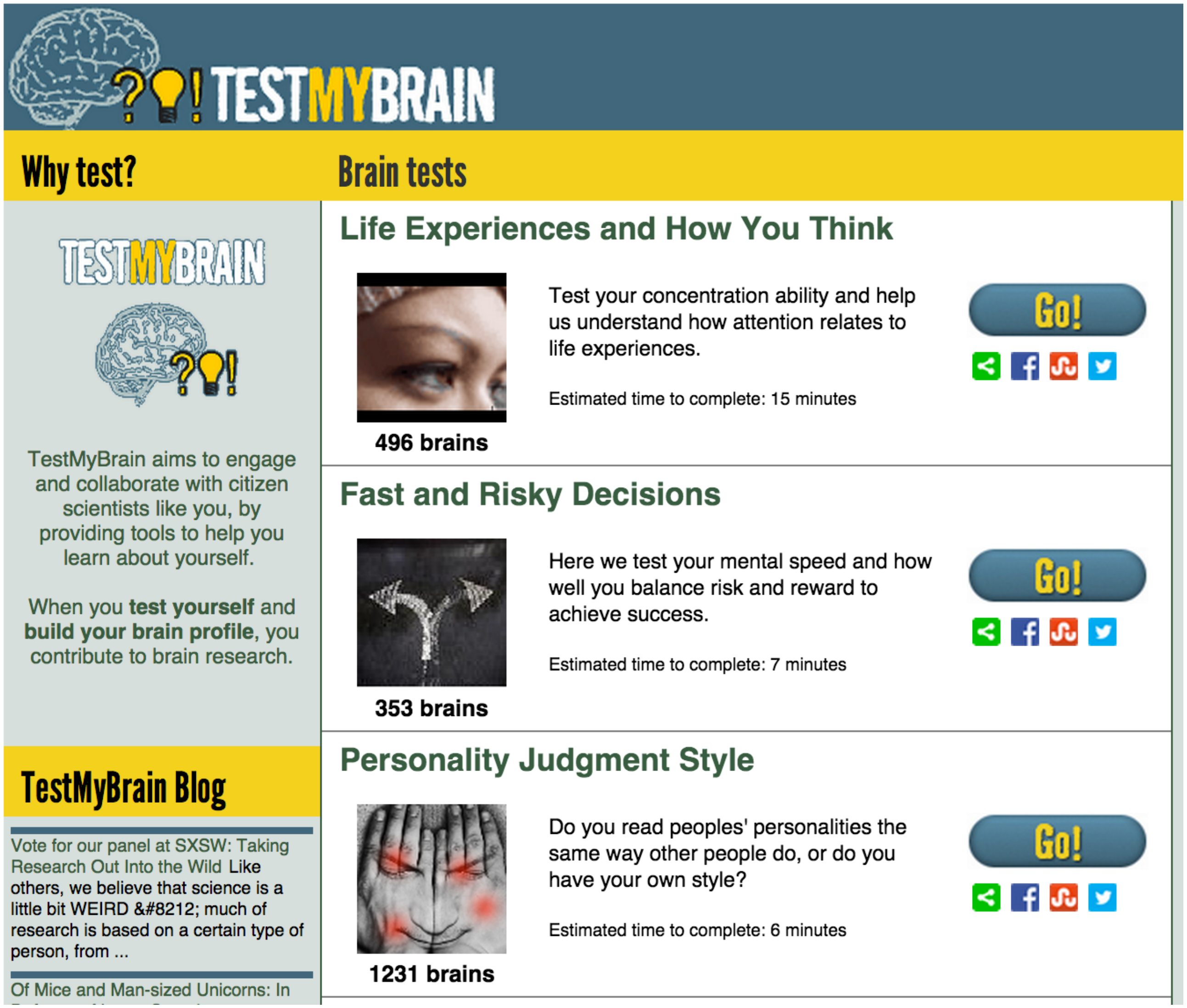


Face Detection: Identical face stimuli shown to humans and computer algorithms.

A selection of common algorithms, including commercial algorithms from Google and face.com (now part of Facebook).

Large-scale web samples captured on the TestMyBrain platform.

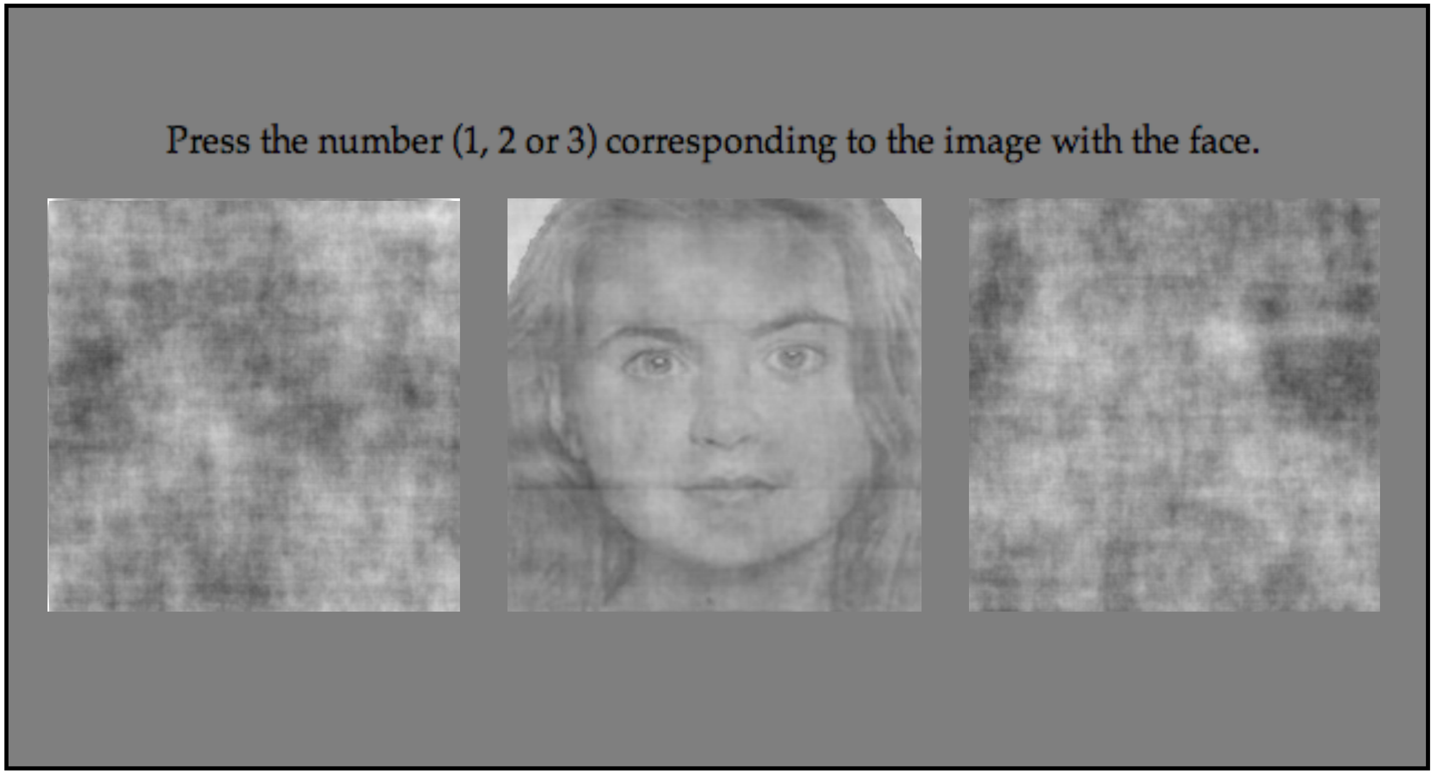
TestMyBrain.org



L. Germine, K. Nakayama, B. Duchaine, C. Chabris, G. Chatterjee, and J. Wilmer, “Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments,” Psychonomic Bulletin & Review, vol. 19, pp. 847–857, 2012.

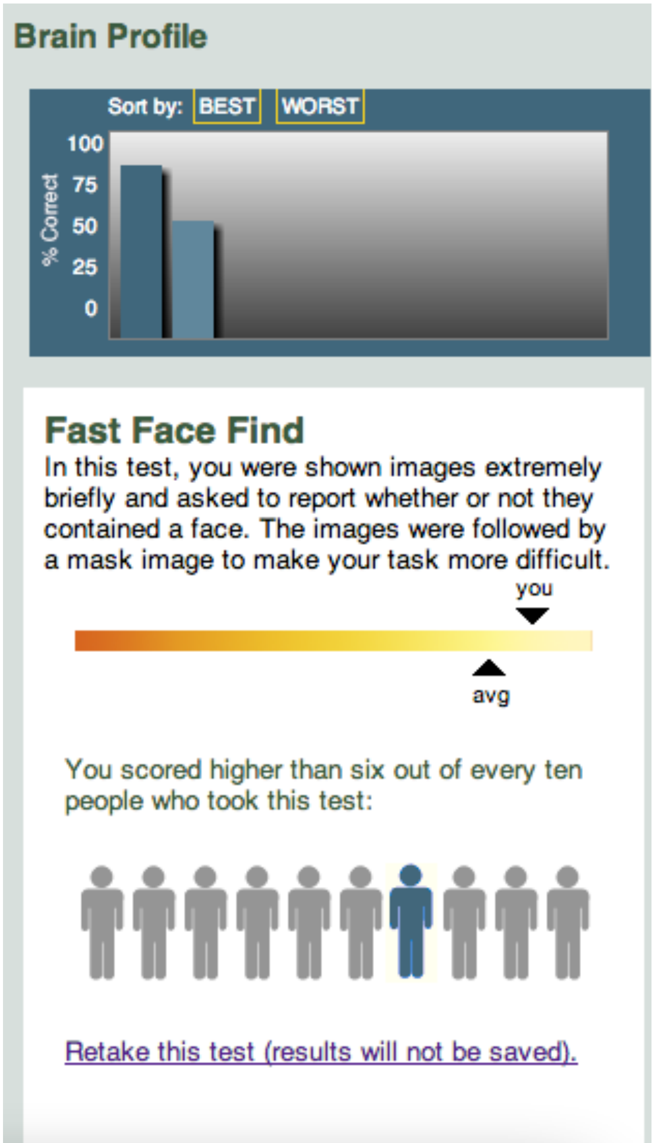
Behavioral Task

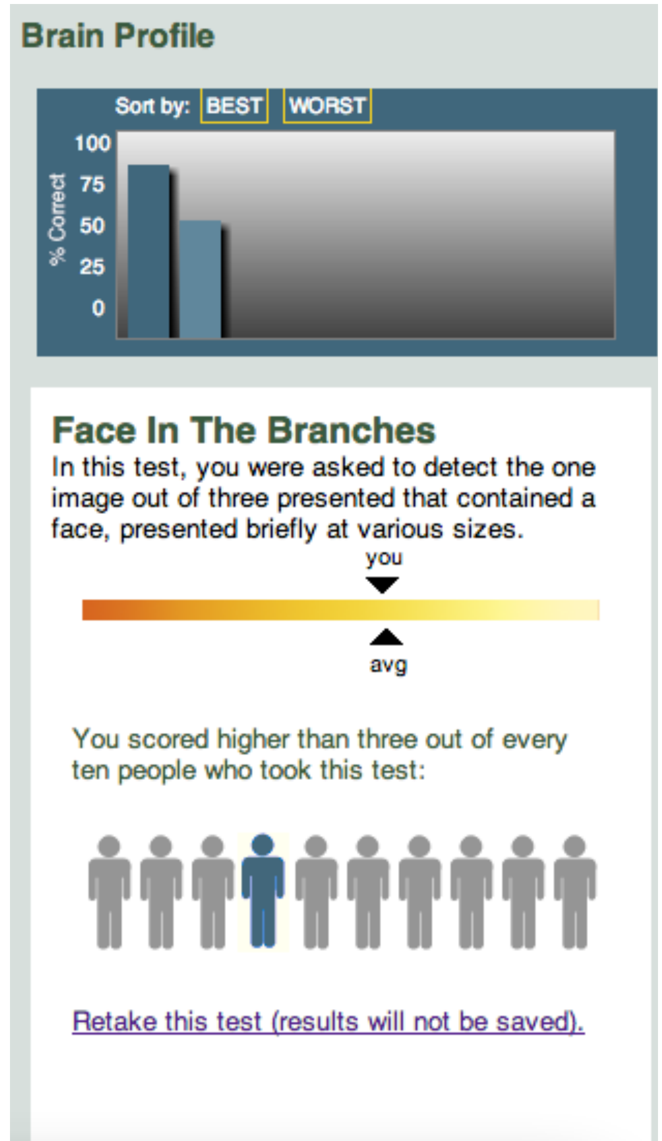
3 Alternative Forced Choice

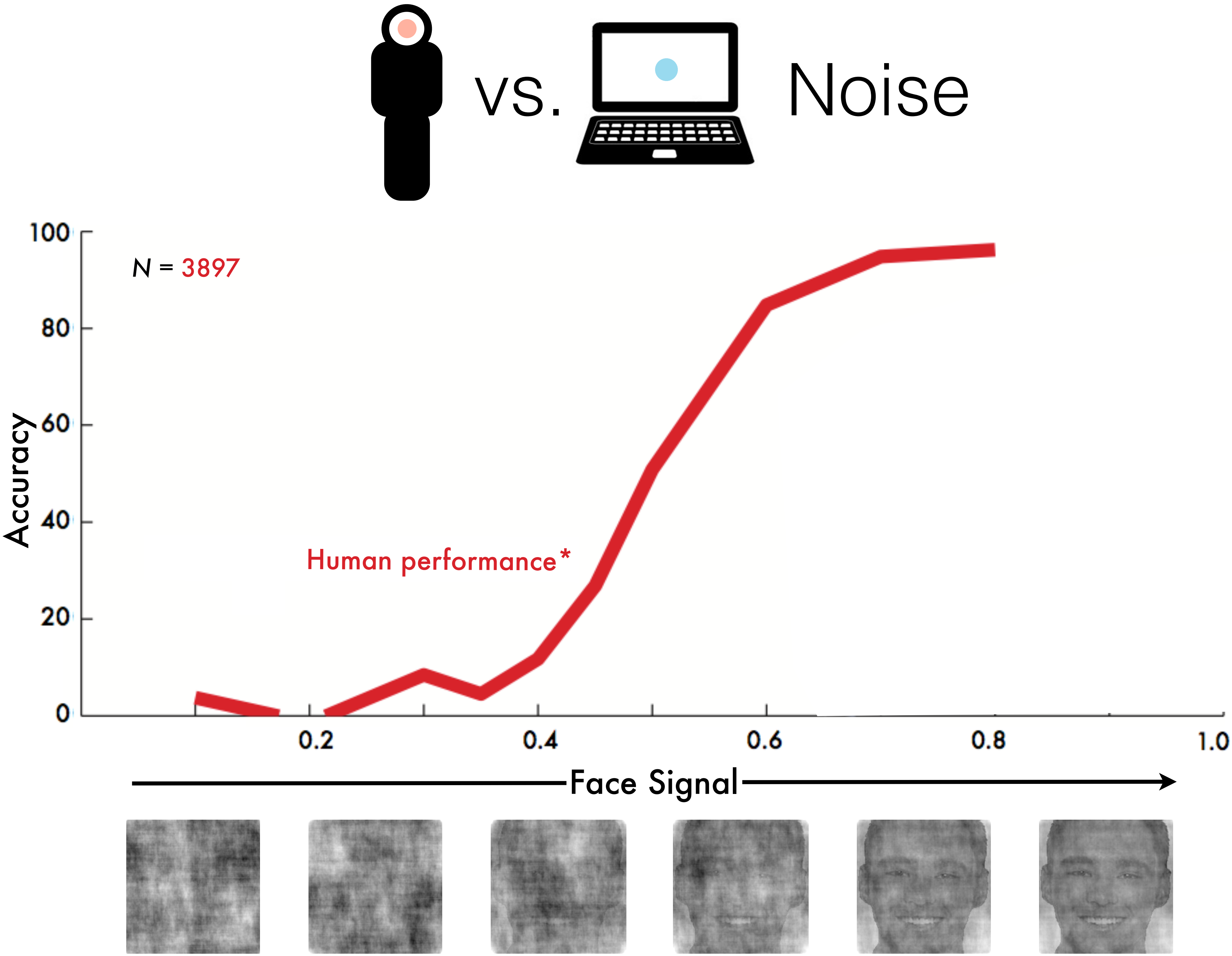


TESTMYBRAIN

Behavioral Task

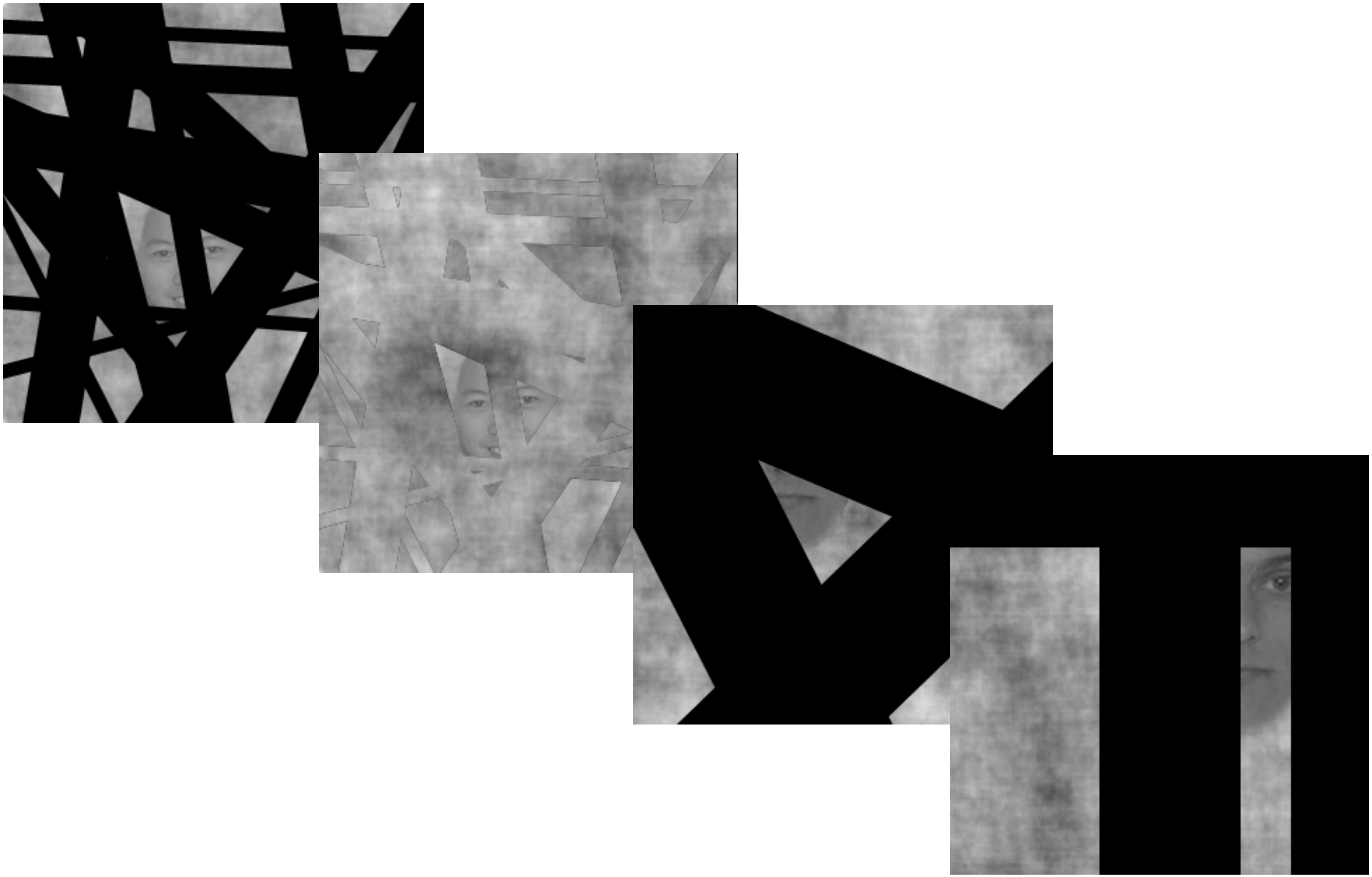


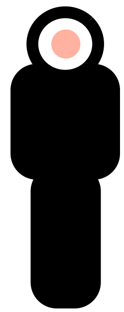


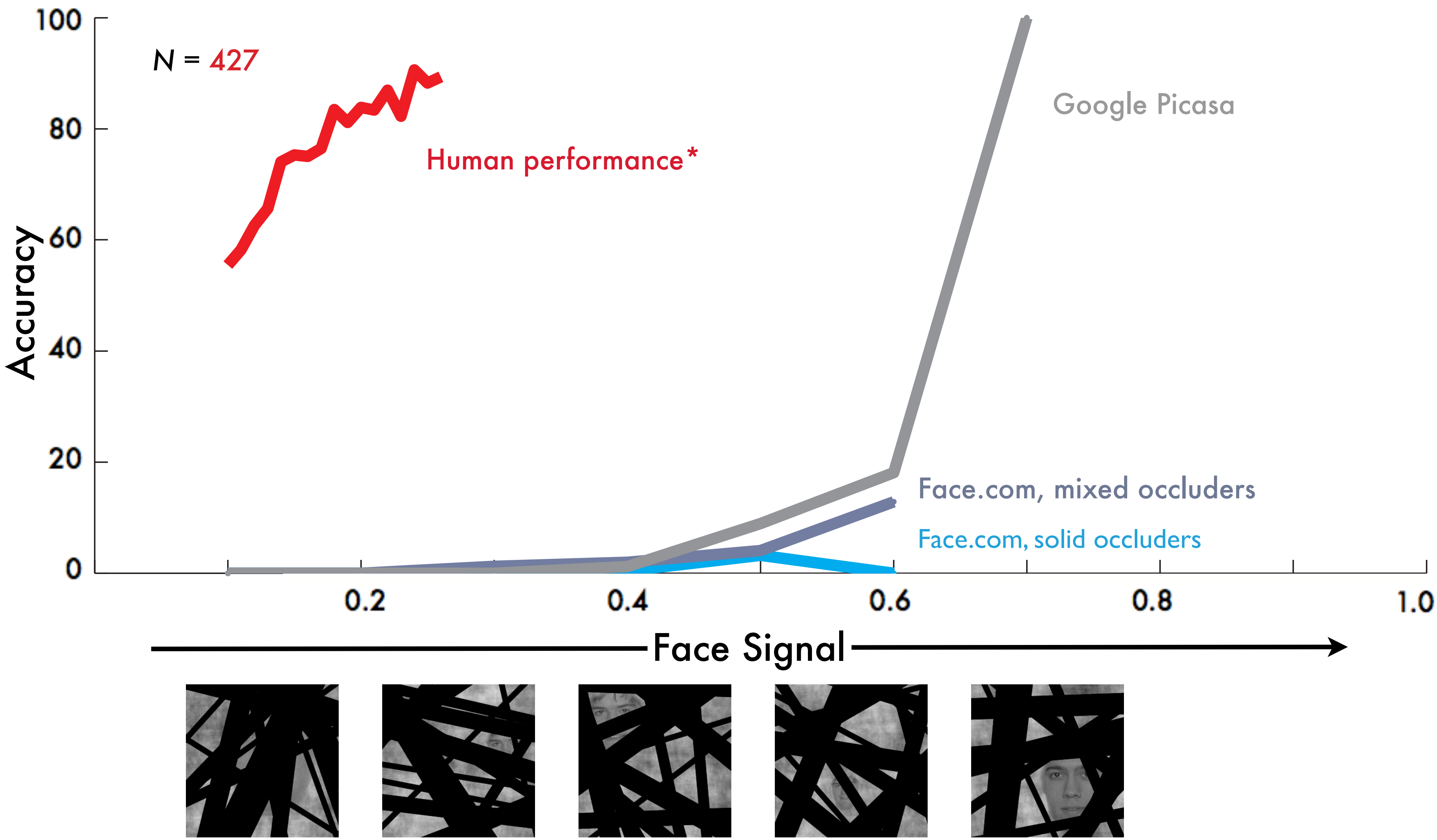


* normalized so chance is zero

Occlusion



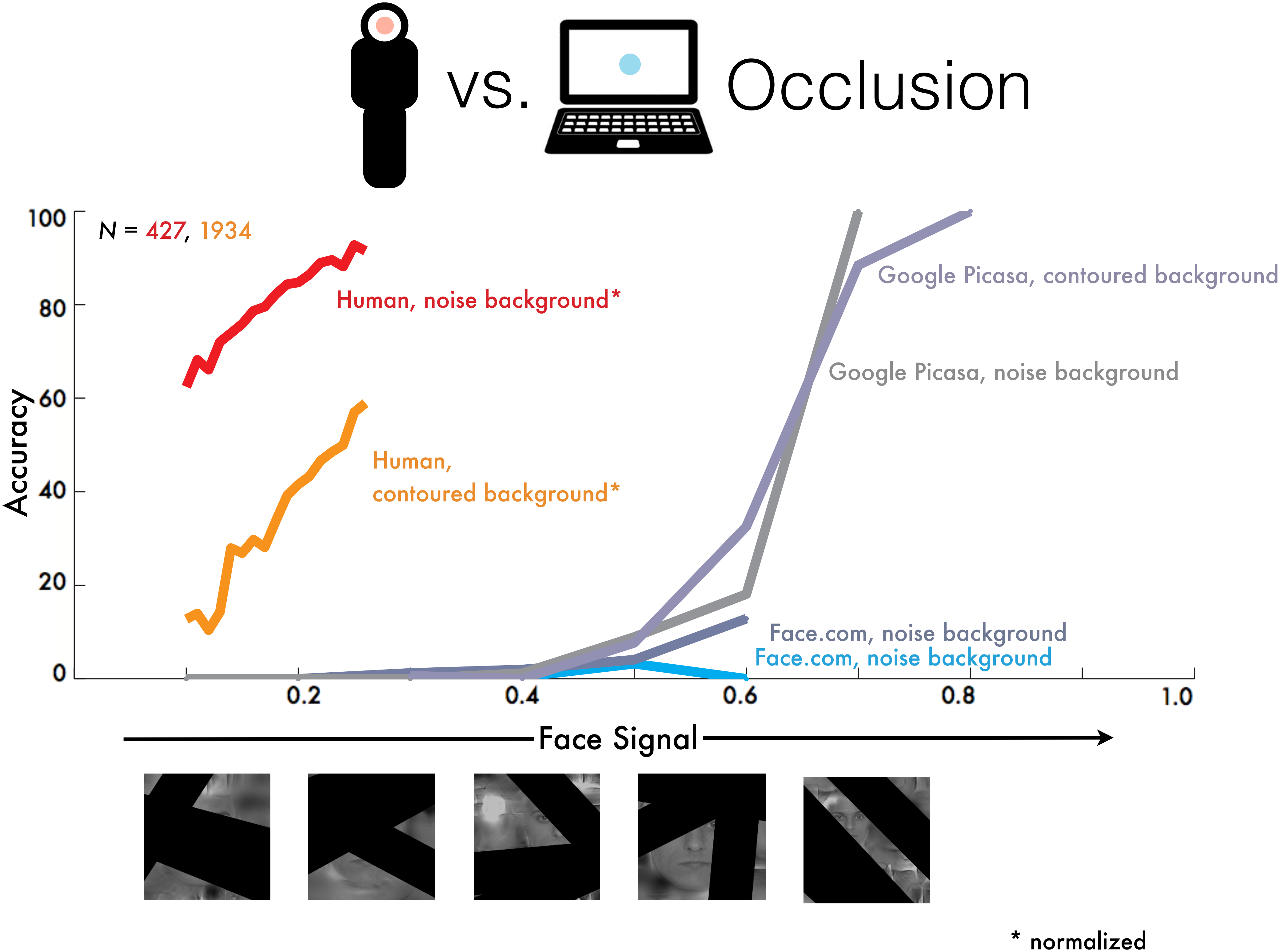
 vs.  Occlusion



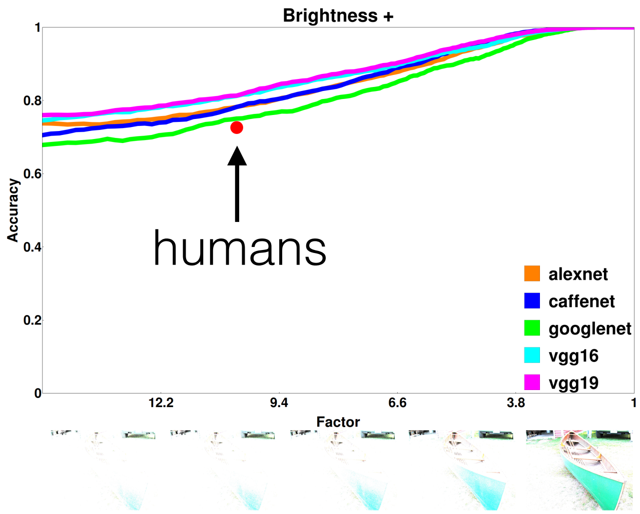
* normalized

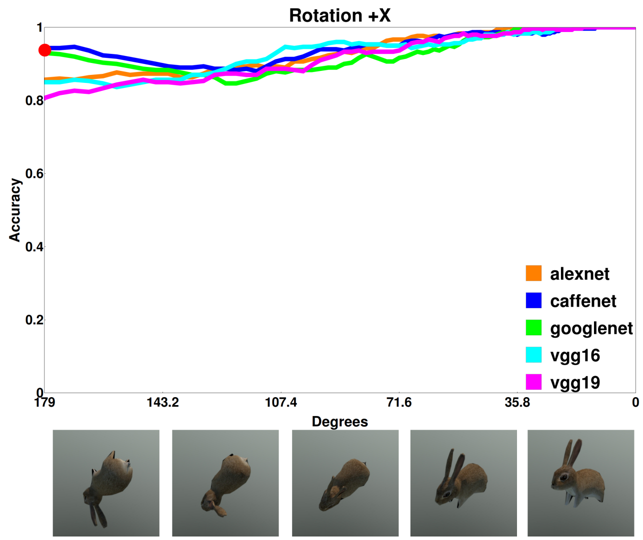
Black occluders with Portilla-Simoncelli Backgrounds

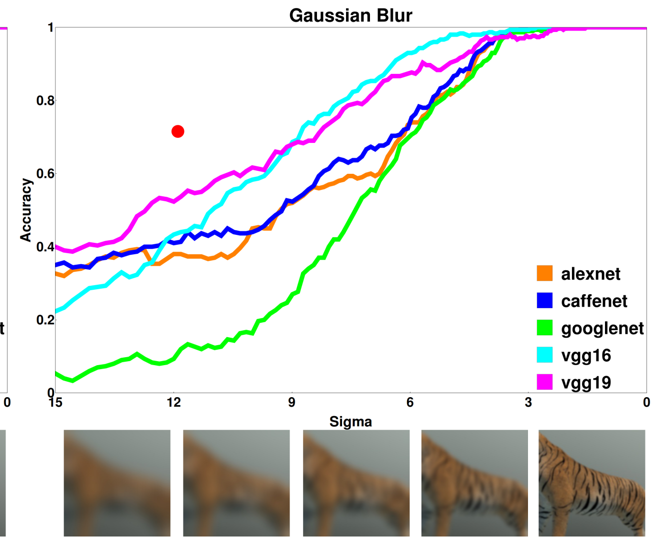
J. Portilla and E. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” IJCV, vol. 40, no. 1, 2000.


vs.
Occlusion
N = 427, 1934
Human, noise background*
Human,
contoured background*
Google Picasa, contoured background
Google Picasa, noise background
Face.com, noise background
Face.com, noise background
Accuracy
100
80
60
40
20
0
0.2
0.4
0.6
0.8
1.0
Face Signal
* normalized

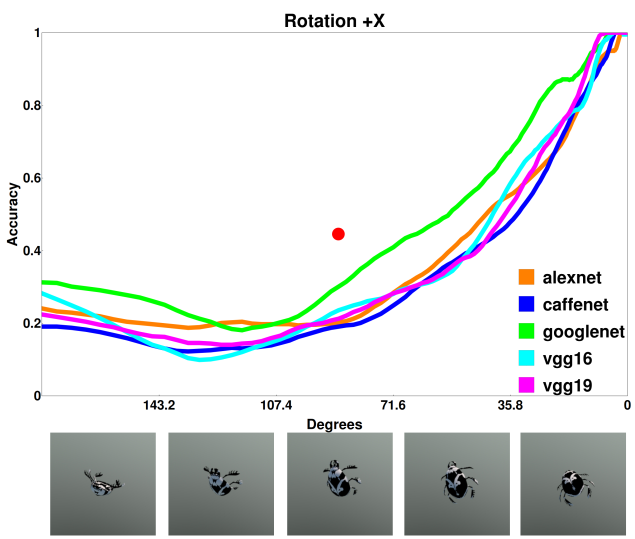
Two-alternative Forced Choice Task

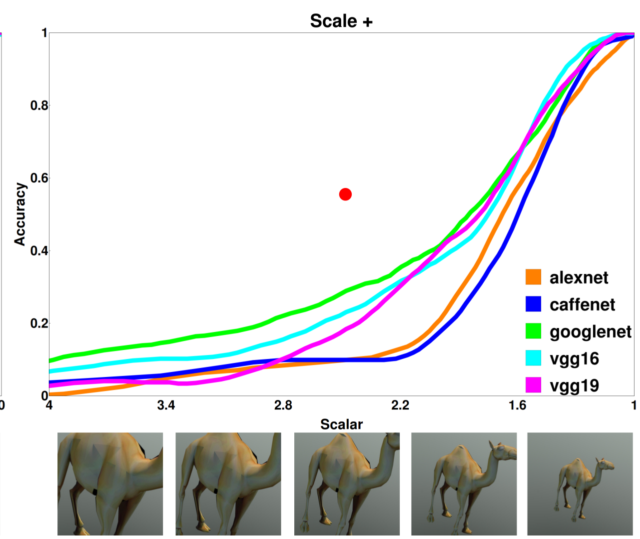






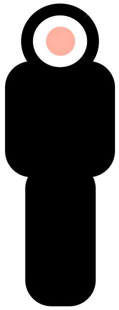
M-alternative Forced Choice Task





Brandon RichardWebster
(Notre Dame CVRL)

B. RichardWebster, S. Anthony, and W. J. Scheirer, “PsyPhy: A Psychophysics Driven Evaluation Framework for Visual Recognition,” arXiv:1611.06448, 2017



vs.

Summary

Humans beat even the best algorithms.

Algorithms have enormous problems with degradations like occlusion that people find trivial.

Contoured image backgrounds reduce human performance; people are still much better.

Perceptual Annotations

What information are we recording from a psychophysics experiment for machine learning training?

1. Per Image Avg. Accuracy

2. Per Image Avg. Reaction Time

Perceptual Annotation for SVM

Classification Risk

$$\underset{f}{\operatorname{argmin}} \left\{ R_{\mathcal{I}}(f) := \int_{\mathbb{R}^d \times \mathbb{N}} L(x, y, f(x)) P(x, y) \right\}$$

Ideal Risk Loss Function Joint Distribution

A. Smola, "Learning with Kernels," Ph.D. dissertation, Technische Universitat Berlin, Berlin, Germany, 1998.

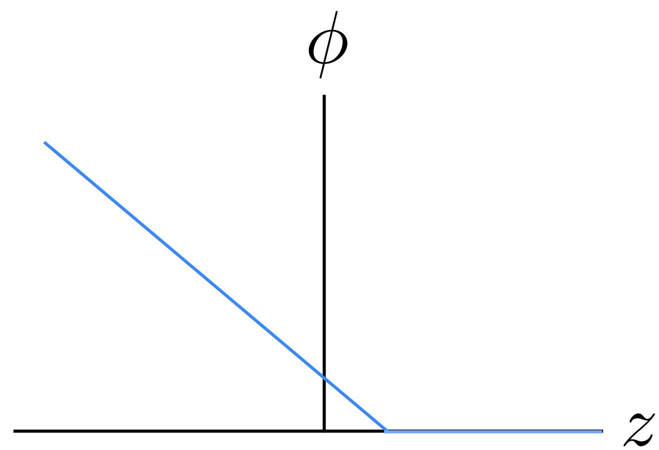
Loss Functions

A prediction during training is calculated as the output of a classifier multiplied by its label:

$$z = yf(x)$$

Typical Loss Function: Hinge Loss

$$\phi(z) = max(0, 1 - z)$$



Non-linear nature of psychometric curves for visual recognition tasks suggests a much different model.

Human Weighted Loss

Besides data x and labels y, assume we also have a cost c for each training sample:

$$\phi_\psi(x, z) = \max(0, (1 - z) + \mathrm{M}(x, z))$$

where

$$\mathrm{M}(x, z) = \begin{cases} c_x & \text{if } z < 1 \\ 0, & \text{otherwise} \end{cases}$$

Human Weighted Loss

c can take on one of two types of values:

A static penalty (*e.g.*, 0 if a sample doesn't have a perceptual annotation)

A point on the psychometric curve (*e.g.*, accuracy or reaction time)

*All training samples do not require an associated perceptual annotation.

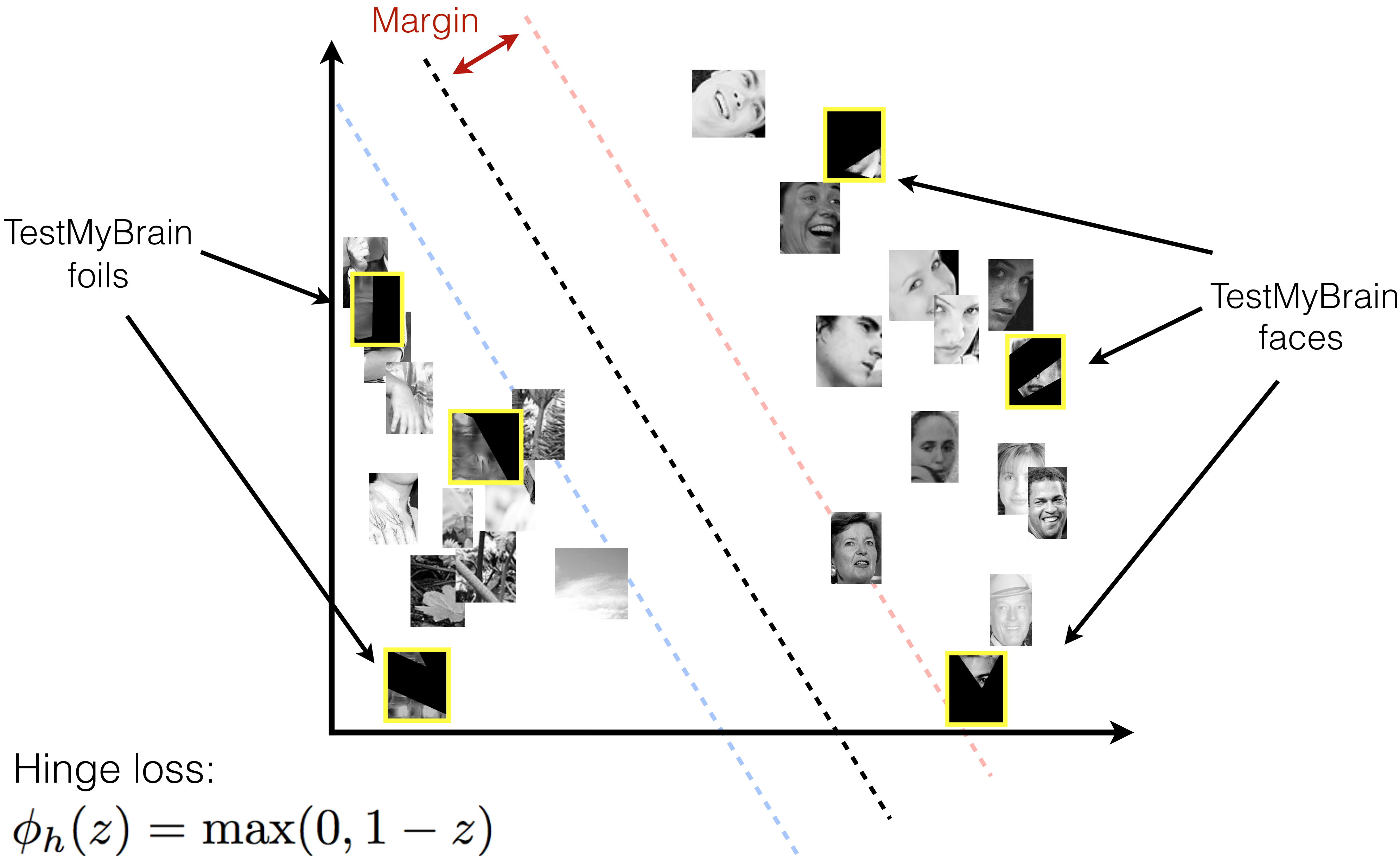
Optimization Problem

For the linear binary case, solve the following optimization problem:

$$min\frac{1}{2}||w||^2 + C\sum_{l=1}^{L}\phi_{\psi}(x_l, y_l f(x_l))$$

Perceptual Annotations

Train a Face Classifier

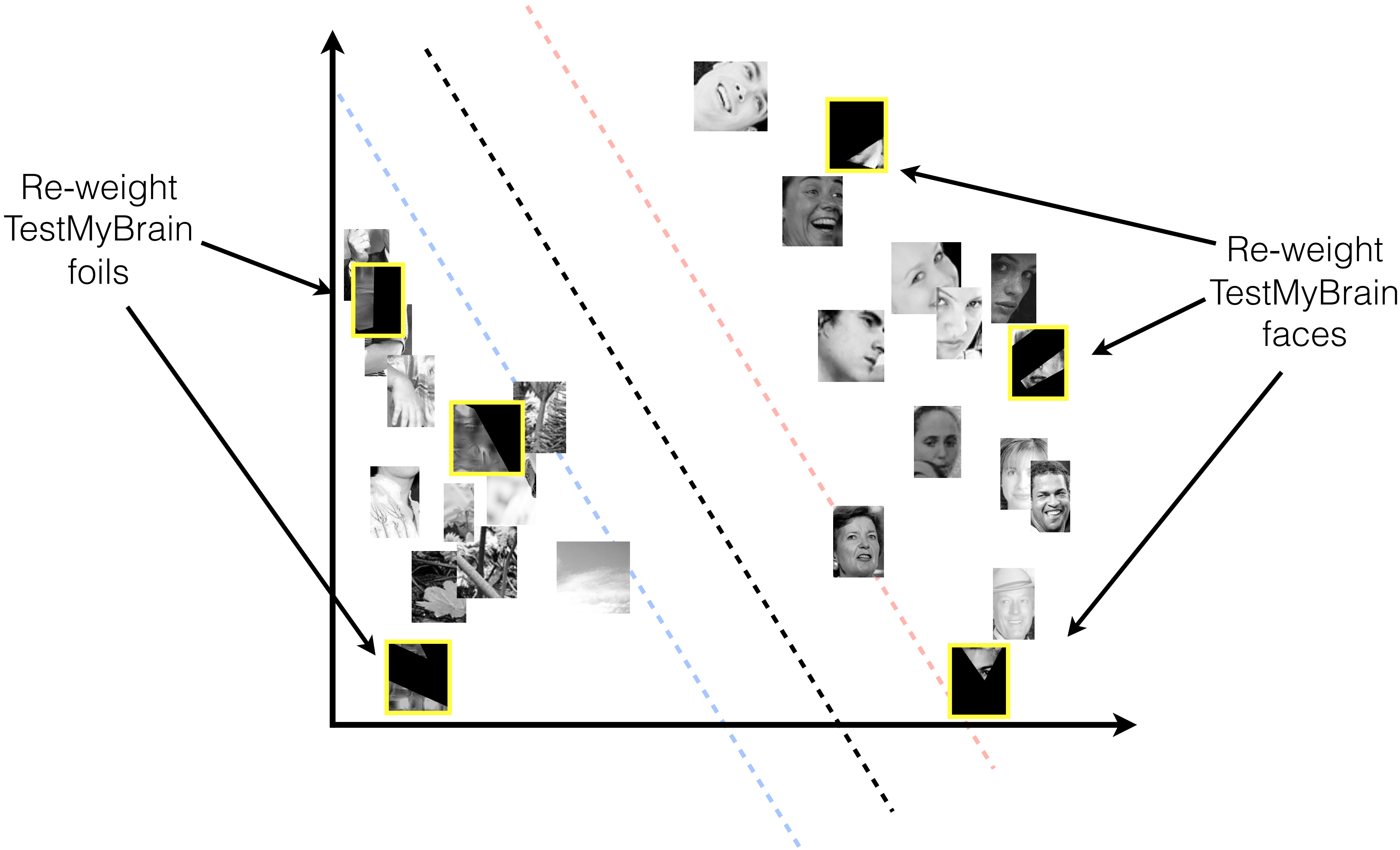


Hinge loss:

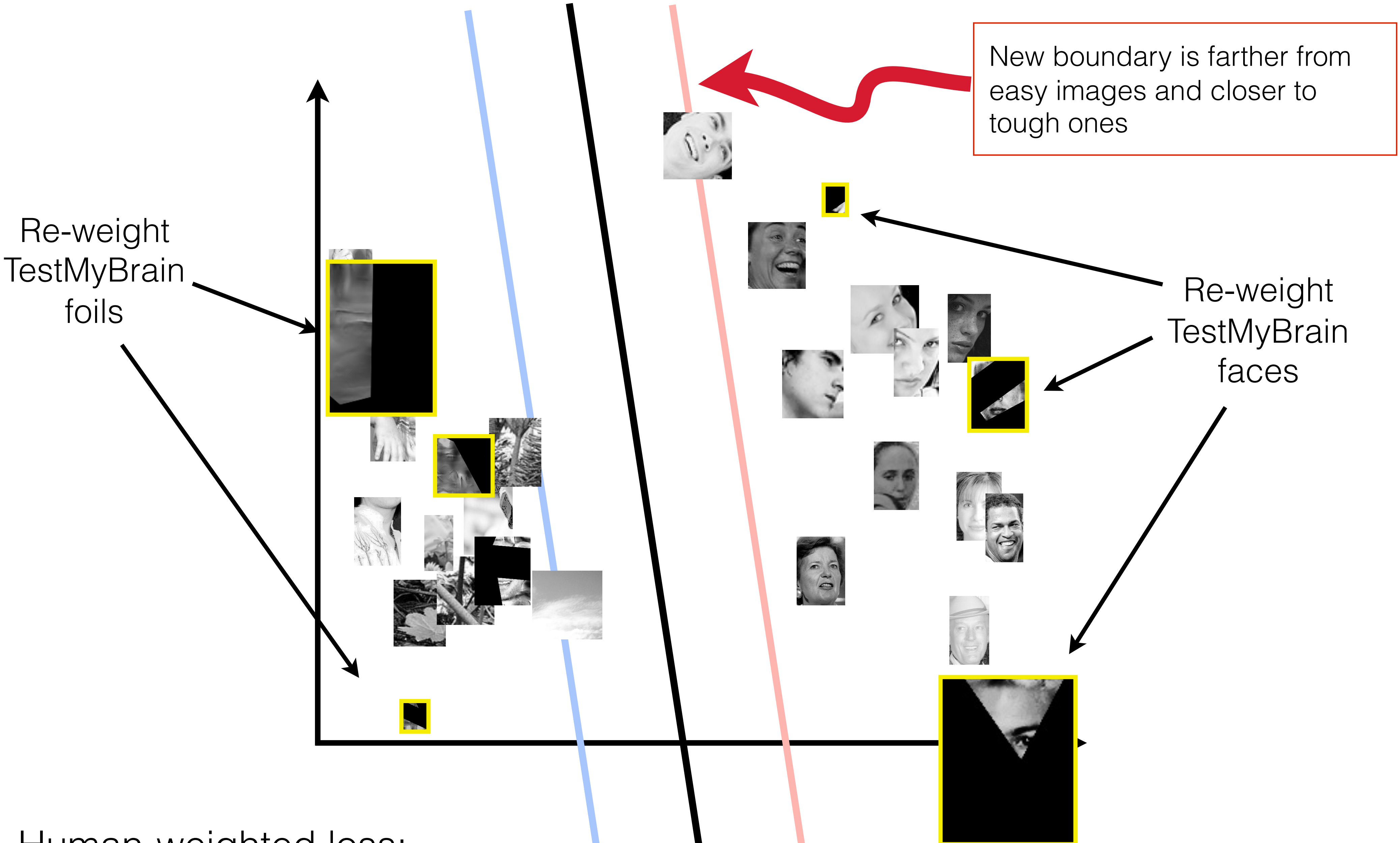
$$\phi_h(z) = \max(0, 1 - z)$$

$$\textit{where} \quad z = yf(x)$$

Re-weight TestMyBrain images



Re-weight TestMyBrain images



Human-weighted loss:

$$\phi_\psi(x, z) = \max(0, (1 - z) + \mathrm{M}(x, z))$$

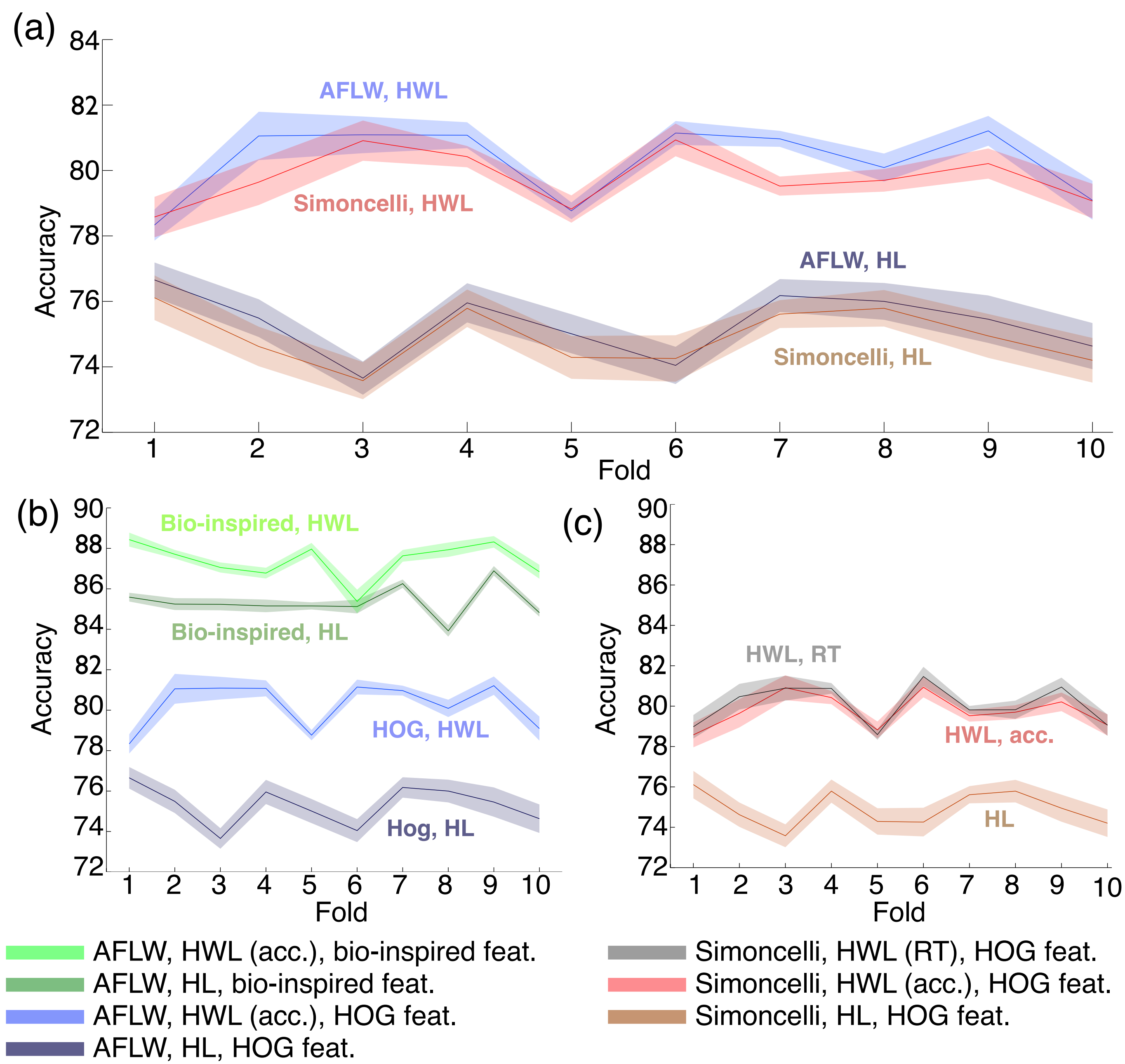
Case Study: Face Detection

FDDB: Face Detection Dataset and Benchmark

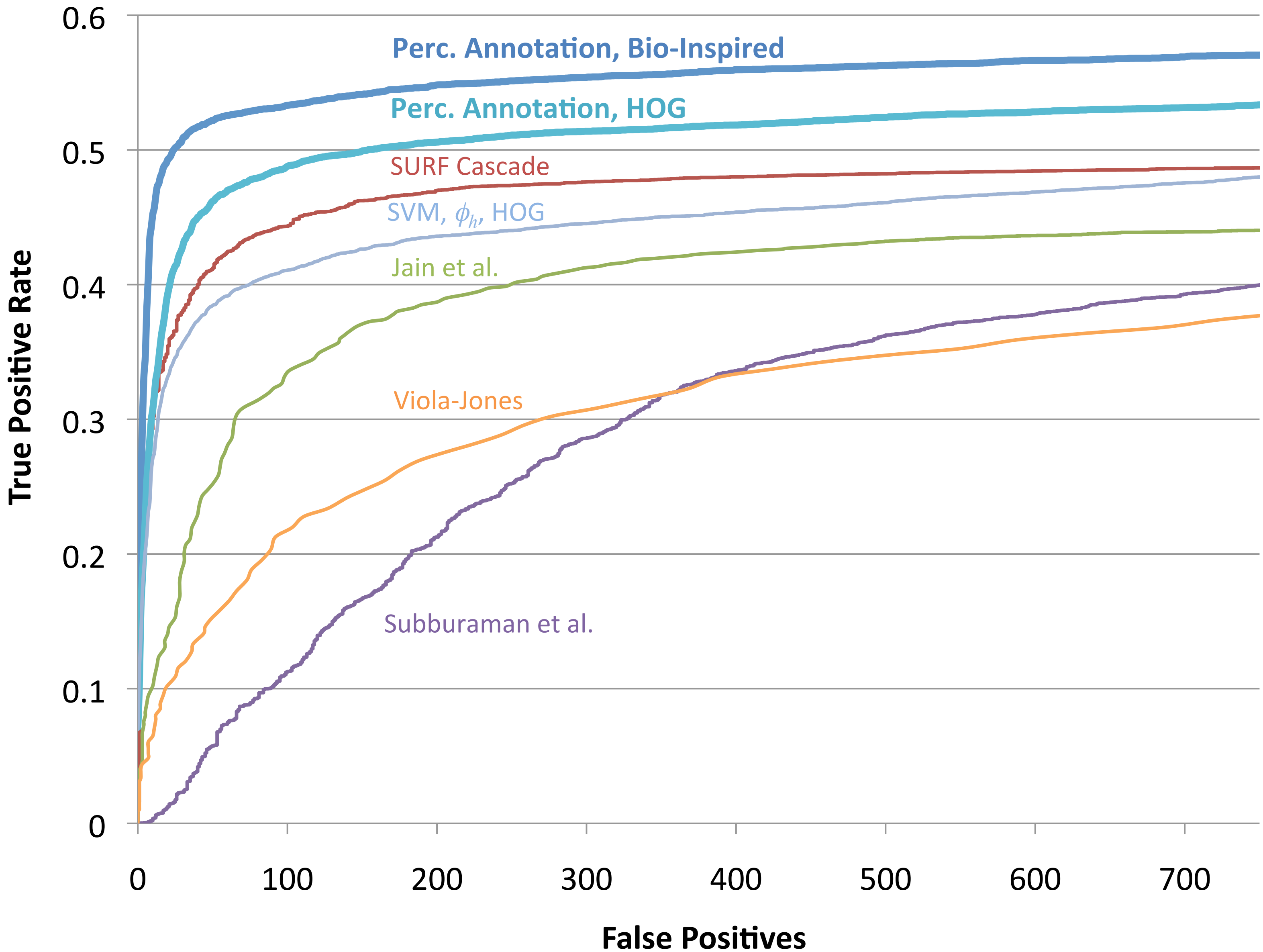
- 2,845 images with a total of 5,171 faces
- A wide range of challenges including occlusions, difficult poses, and low resolution and out-of-focus faces
- The specification of face regions as elliptical regions
- Both grayscale and color images
- 10-fold cross-validation style testing

V. Jain and E. Learned-Miller, “FDDB: A Benchmark for Face Detection in Unconstrained Settings,” UMass Amherst Tech Report UM-CS-2010-009, 2010

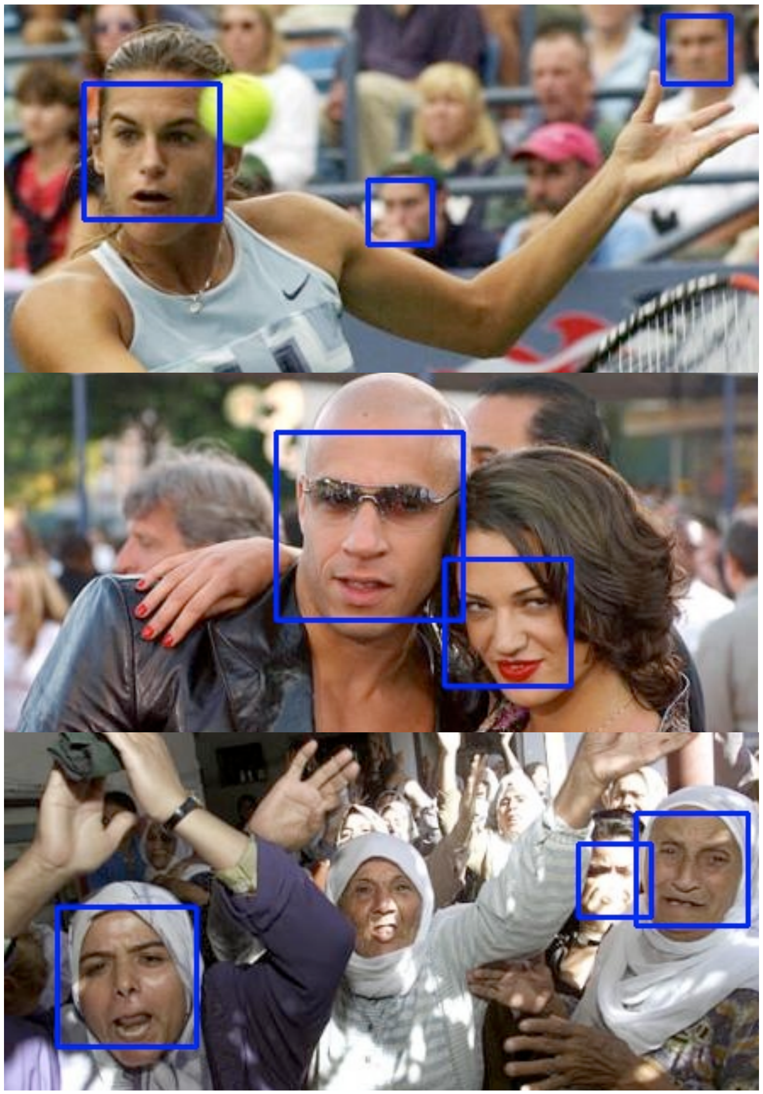
Effect: HL Replaced by HWL



FDDB Continuous Score Metric



Example Detections



Viola-Jones

Perceptual Annotation

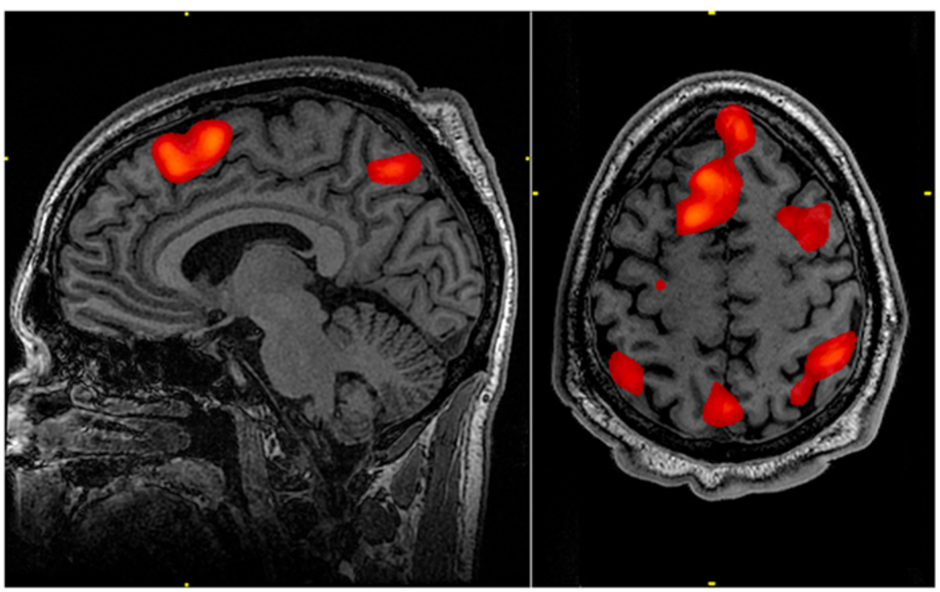
Using Human Brain Activity to Guide Machine Learning

fMRI

A more direct way to measure brain activity

Non-invasive experimentation with humans

Uses blood flow as a proxy for neuronal activations



Spatial resolution good enough to identify Brodmann areas

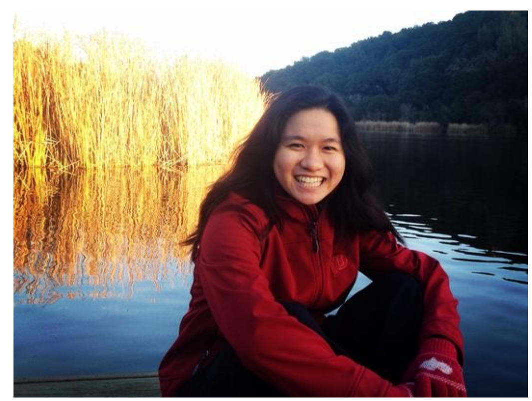
“Neurally-Weighted” Machine Learning

Collect fMRI measurements of human brain activity from subjects viewing images

Infuse these data into the training process of an object recognition learning algorithm

Goal: a solution that is more consistent with the human brain (like perceptual annotation)

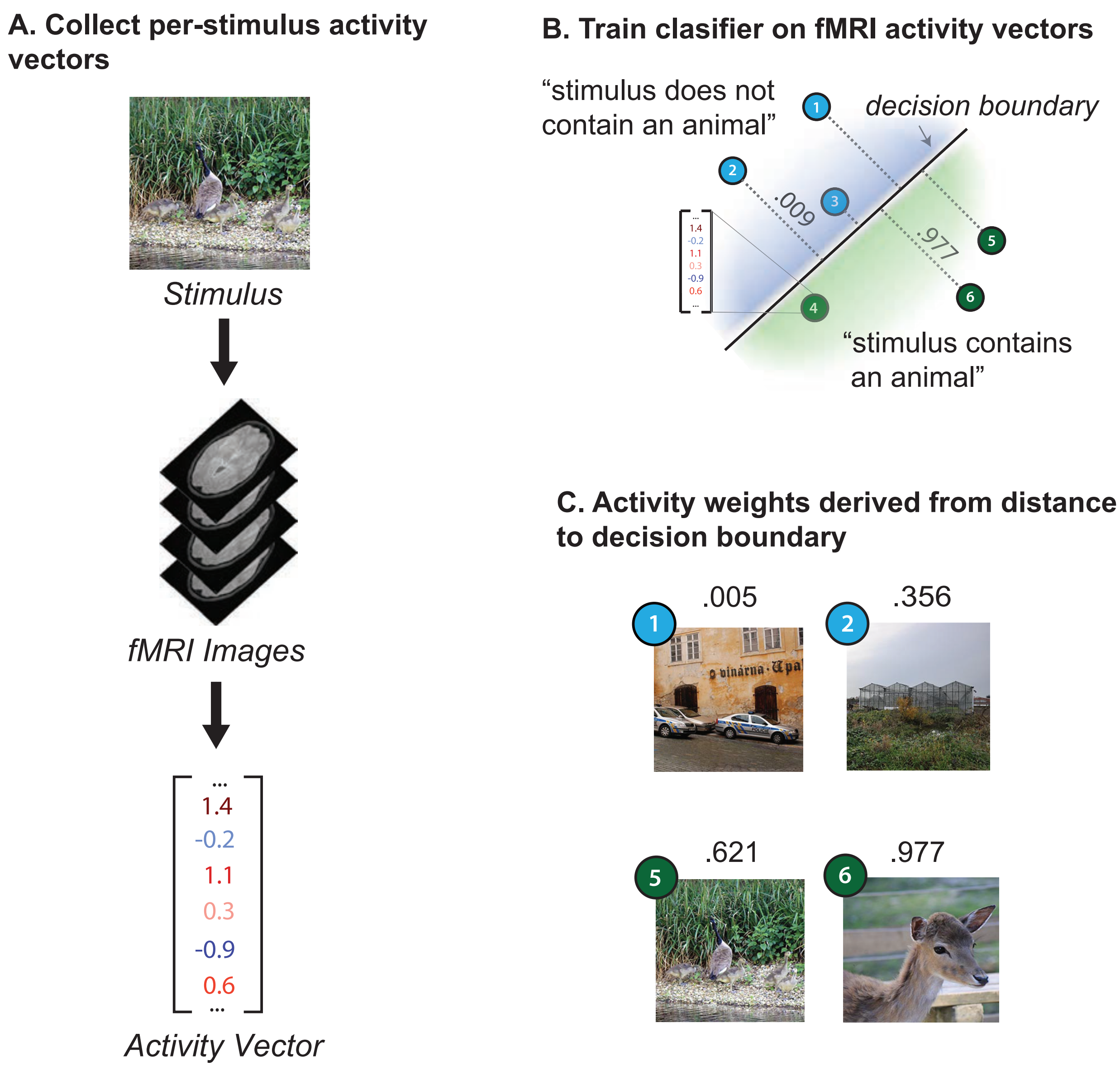
After training, neurally-weighted classifiers are able to classify images without requiring additional neural data



Ruth Fong (Oxford Visual Geometry Group)

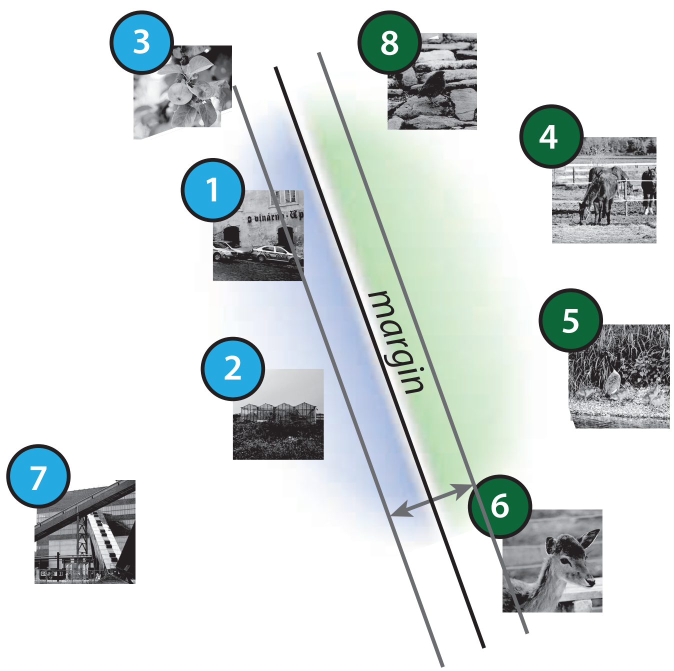
R. Fong, W. J. Scheirer, and D. D. Cox, “Using Human Brain Activity to Guide Machine Learning,” to appear in Scientific Reports, 2017.

Phase 1: Derive per-stimulus activity weights from fMRI data

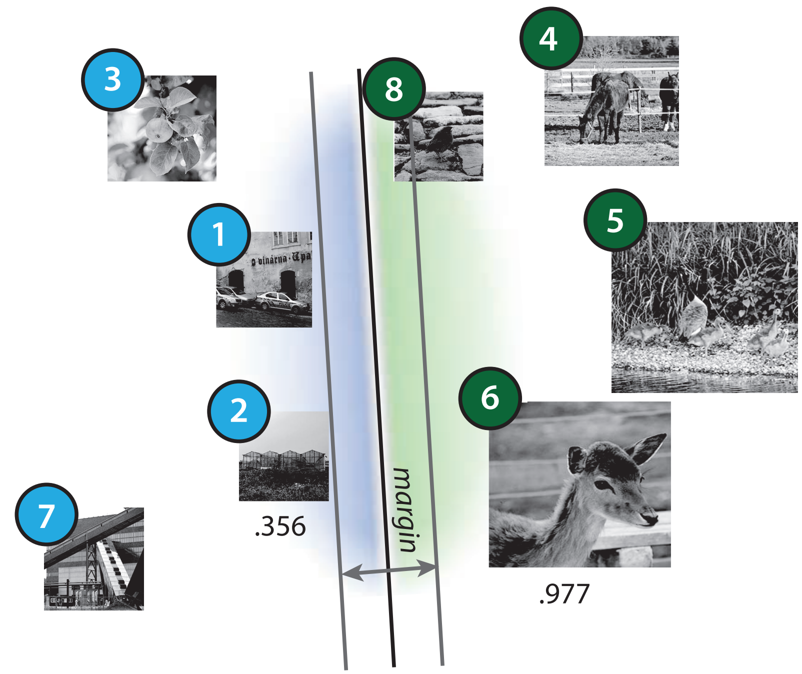


Phase 2: Train Image Classifier

D. Conventional image classifier training



E. Margins reweighted by activity data



fMRI Experimental Setup

Data collected by the Gallant lab at UC Berkeley*

One adult subject viewed 1,386 color 500 × 500 pixel images of natural scenes, while being scanned in a 3.0 Tesla MRI machine

Response amplitude values for 67,600 voxels were available for each image

3,569 were labeled as being part of 1 of 13 visual ROIs, including those in the early visual cortex

7 ROIs for higher level visual processing

*D. E. Stansbury, T. Naselaris, and J. L. Gallant, “Natural Scene Statistics Account for the Representation of Scene Categories in Human Visual Cortex”, Neuron 79, 2013

Brain Areas

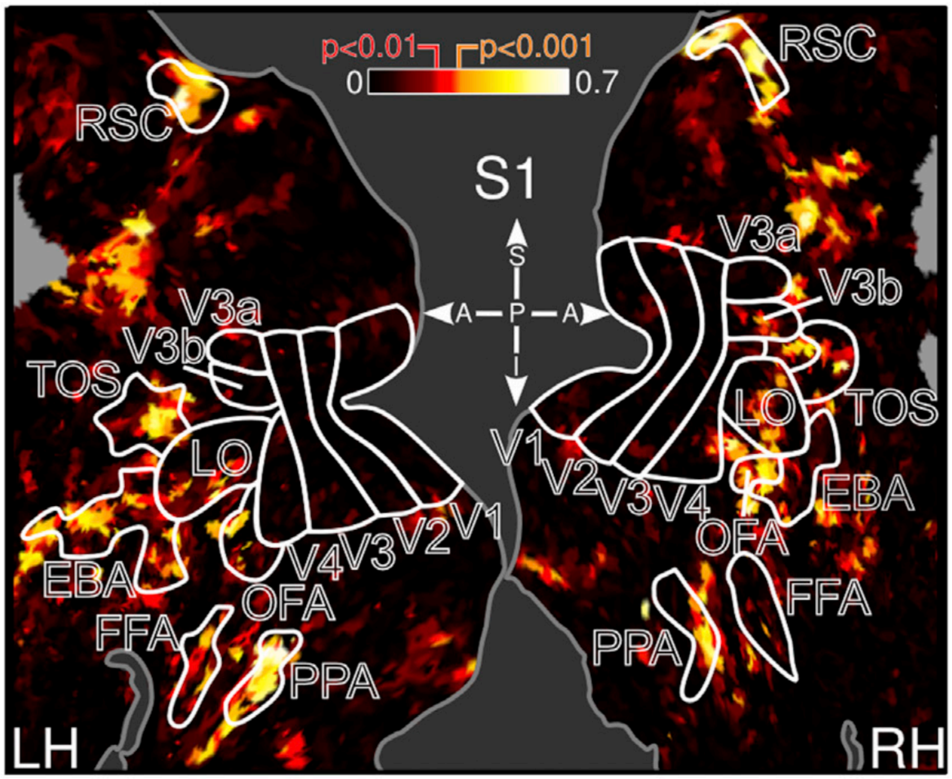


Figure Credit: Stansbury et al. Cell 2013

extrastriate body area (EBA)

fusiform face area (FFA)

lateral occipital cortex (LO)

occipital face area (OFA)

parahippocampal place area (PPA)

retrosplenial cortex (RSC)

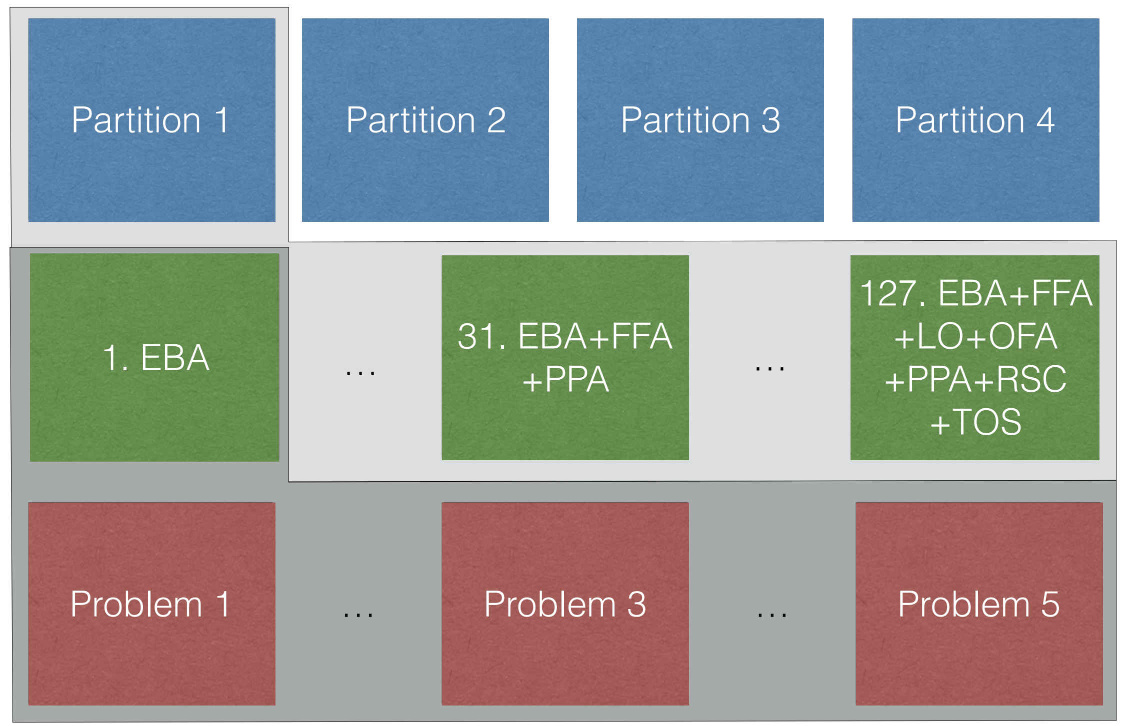
transverse occipital sulcus (TOS)

Machine Learning Experimental Setup

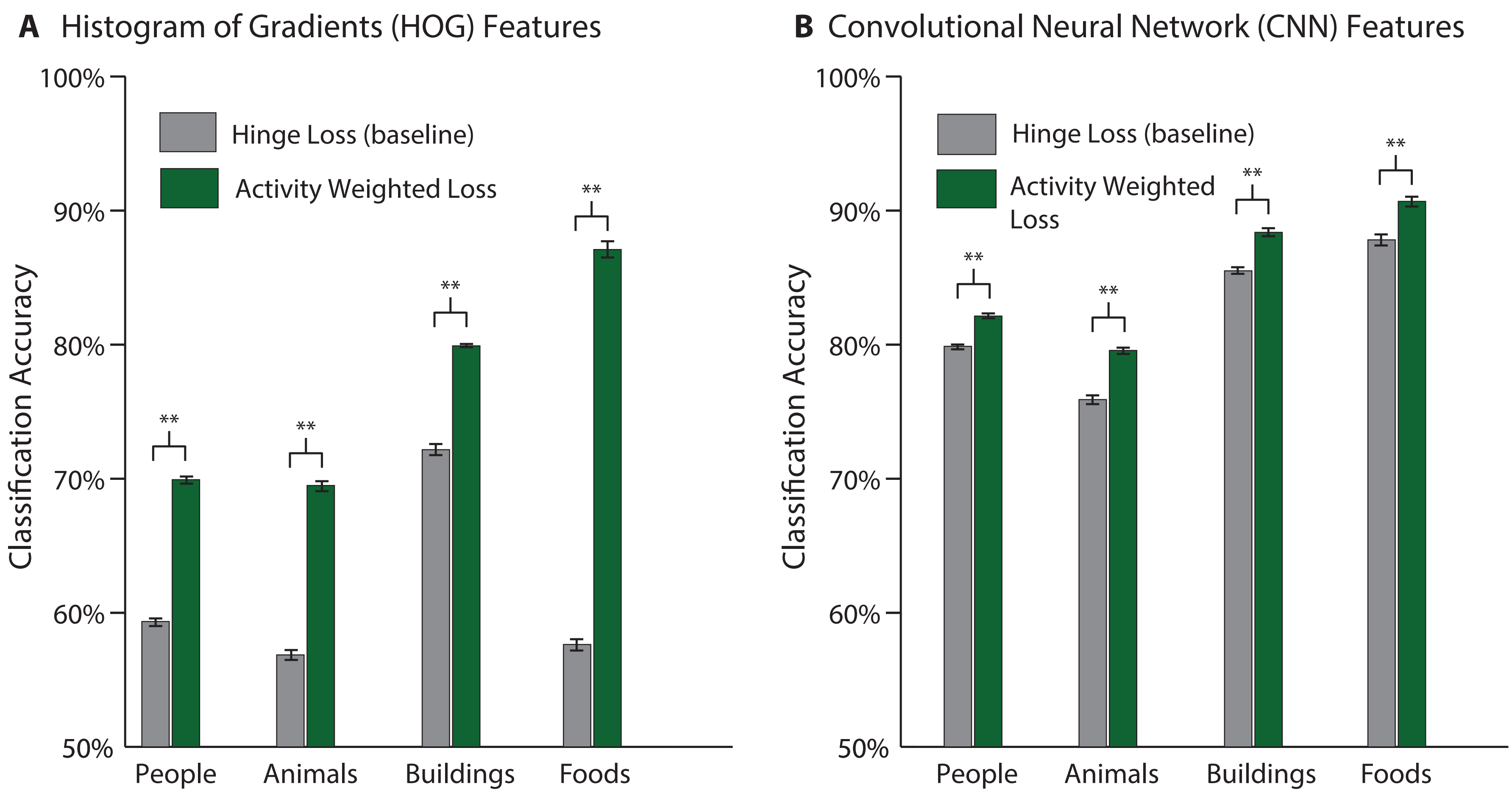
1. Set up 4 partitions that randomly split training (80%) and test (20%) data.

2. Set up 127 parallel experiments for the 127 combinations of 7 ROIs.

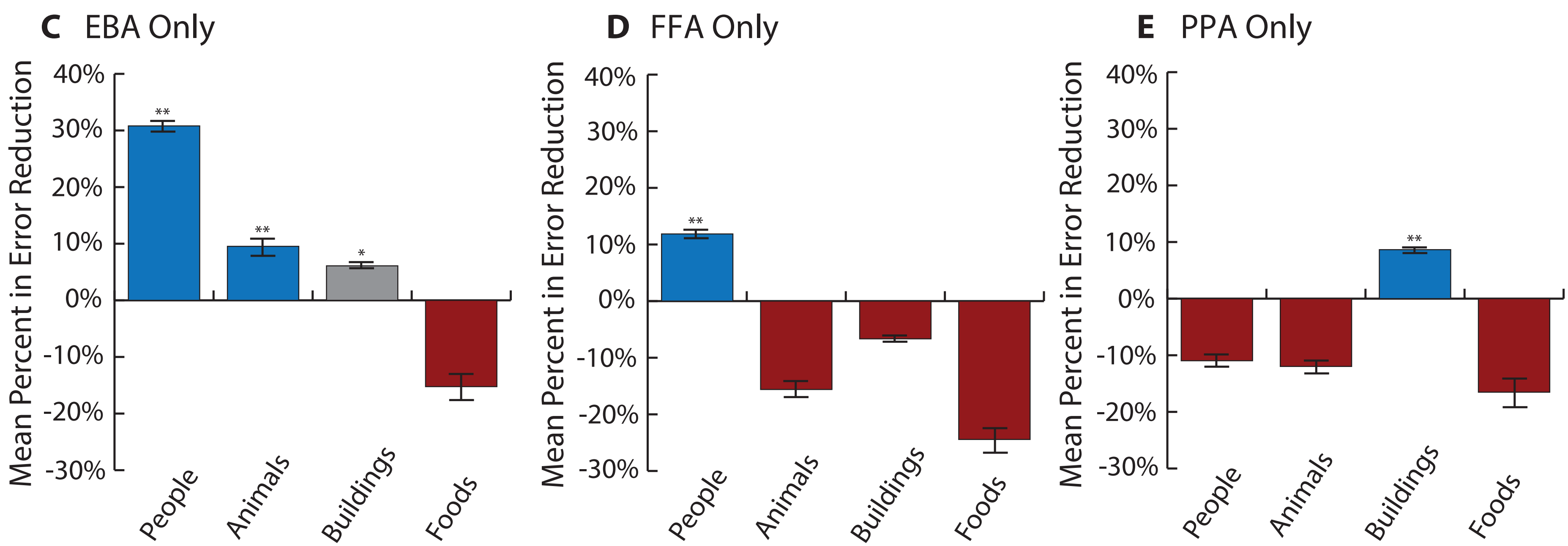
2. Set up 5 balanced classification problems.



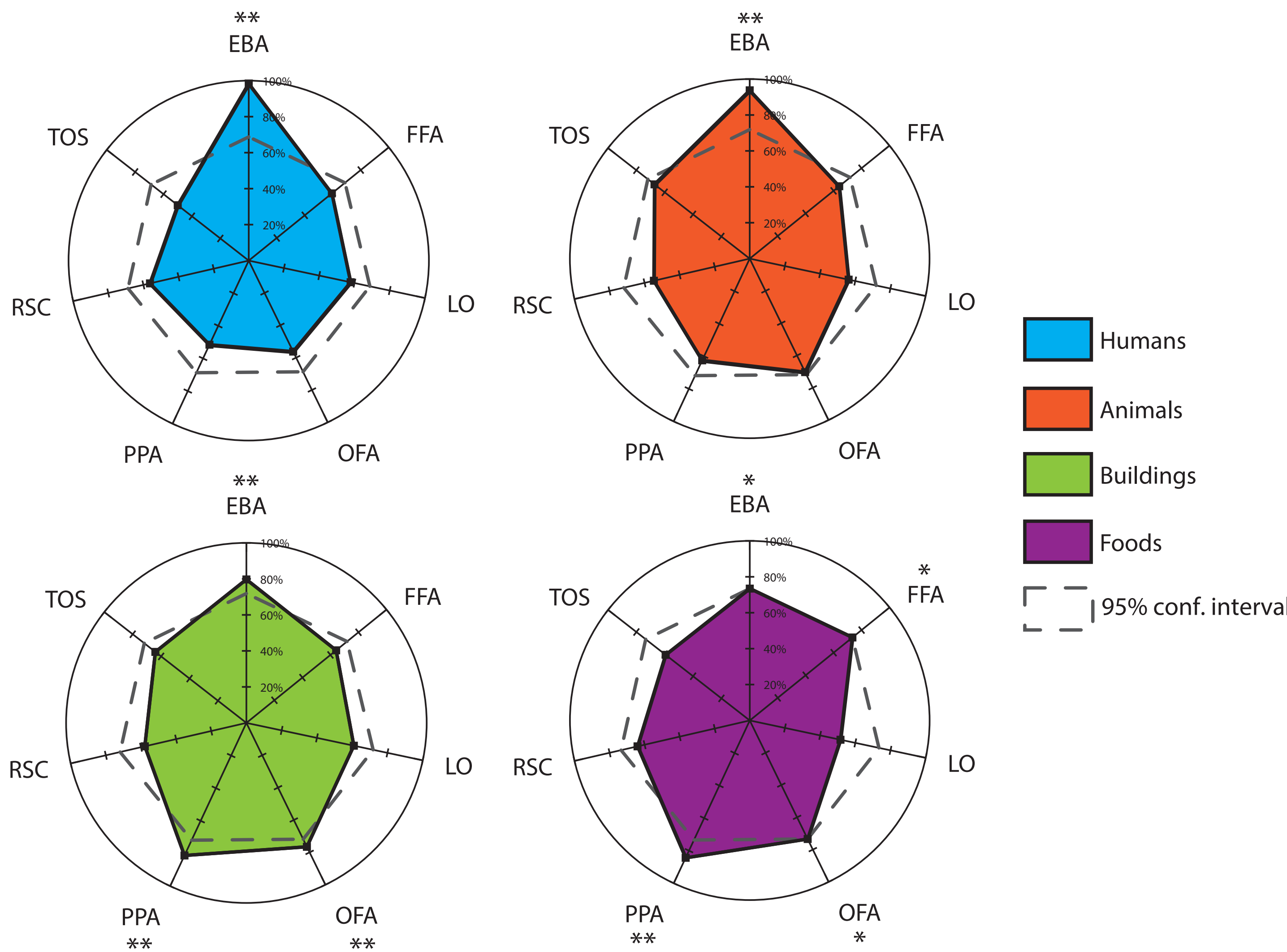
Side-by-side comparisons of the mean classification accuracy



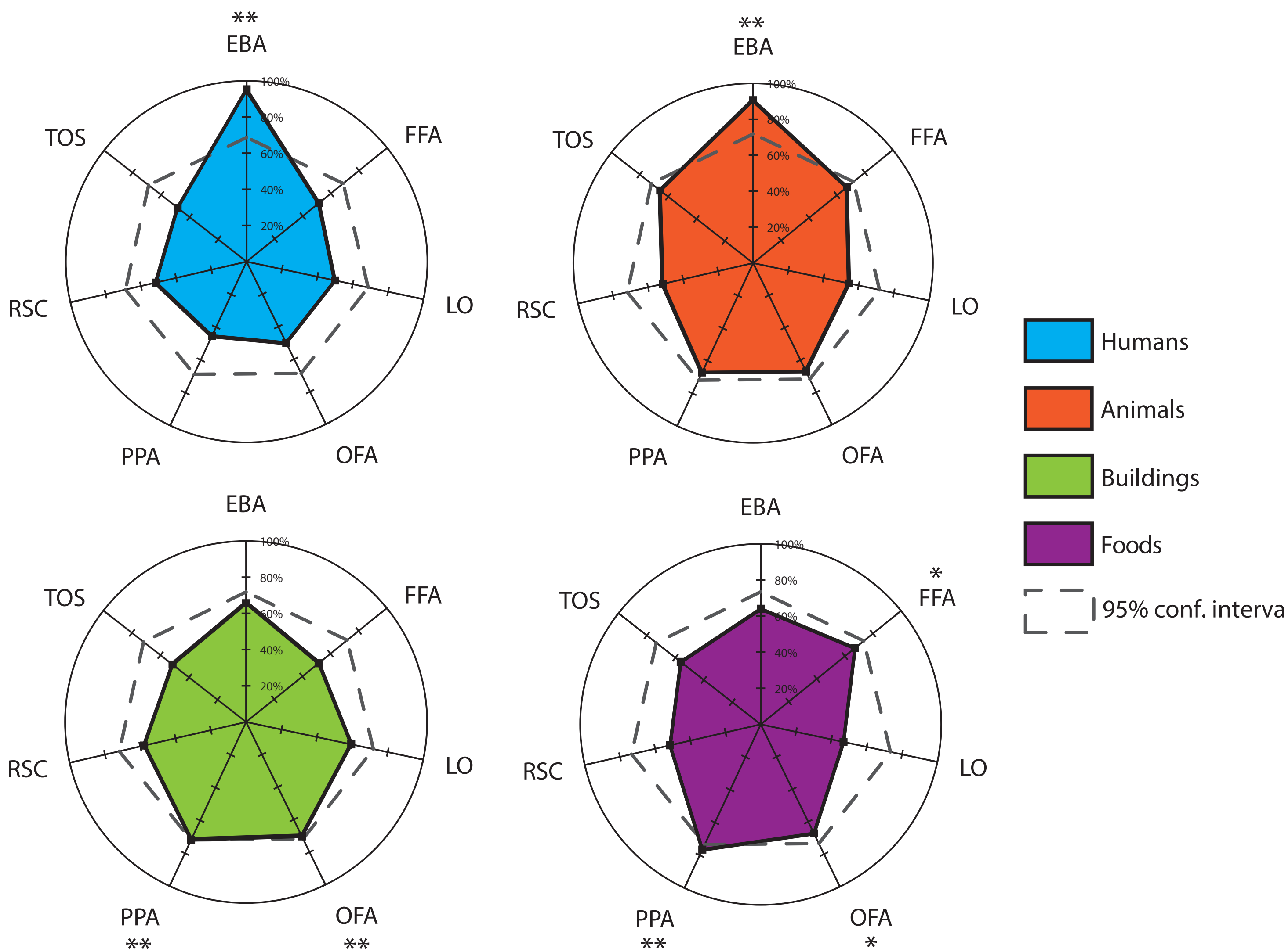
Mean error reductions gained when conditioning classifiers on brain activity from individual ROIs



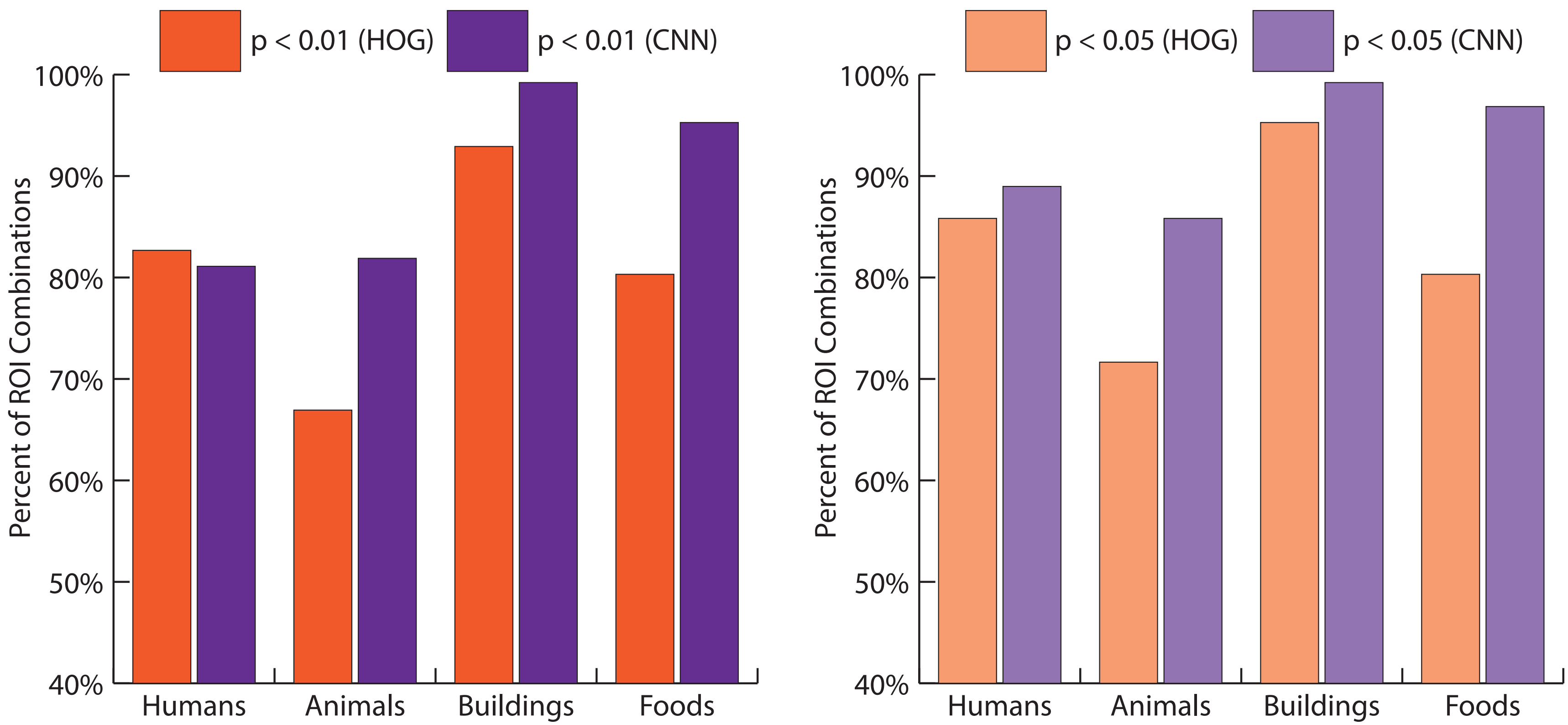
Analysis of ROIs (HOG)



Analysis of ROIs (CNN)



Significance of All Combinations



Wrapping Up…

Resources

Code:

https://github.com/coxlab/perceptual-annotation

Data:

http://www.wjscheirer.com/datasets/perceptual_annotation/

TestMyBrain:

http://TestMyBrain.org

Questions?