Using Human Brain Activity to Guide Machine Learning

Ruth Fong
ruthfong@robots.ox.ac.uk

MBCC Workshop
October 29, 2017

"We hold these truths to be self-evident, that not *all cat gifs are created equal..." [Fong 15]*

"Cat"-ness

Which is the more "cat"-like photo?

Learnability

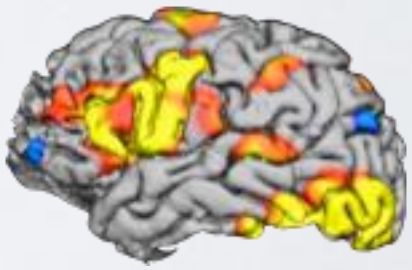
- We never ask a novice to learn all information at once.
- Yet, machine learning and computer vision algorithms typically use a "sink-or-swim", "all-or-nothing" approach to learning.

Motivating Questions

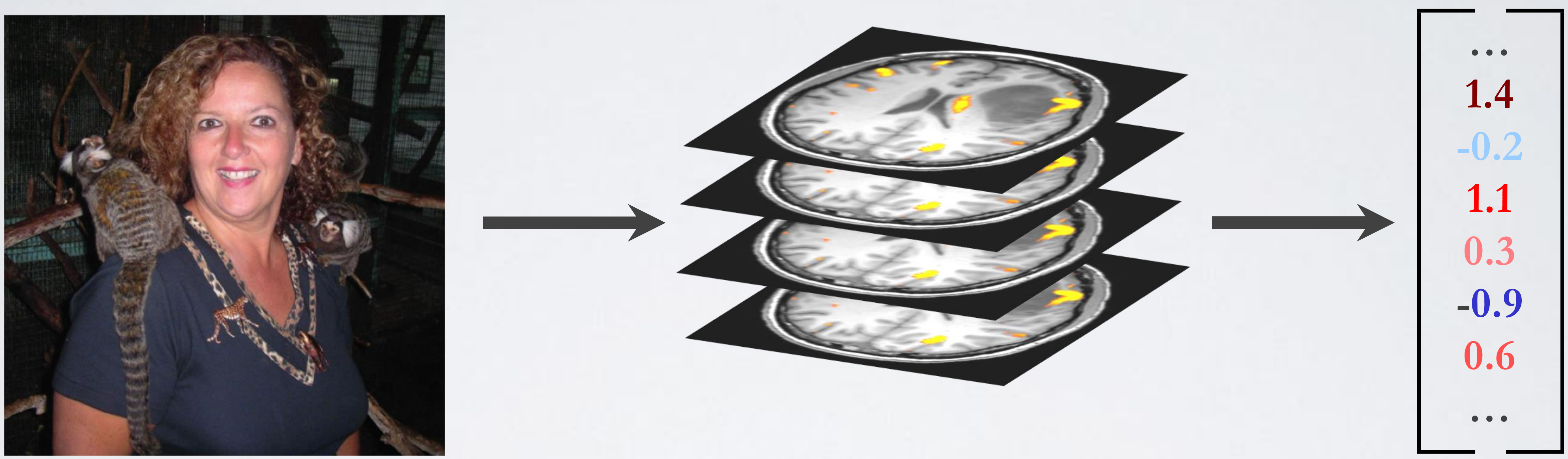
1. Can we learn how *canonical* an example of an object is from human brain activity?
2. Does using brain-derived annotations of *canonical-ness* improve object classification?
3. If so, which brain regions best improve object classification?

vs

+



Biologically-Informed Learning Paradigm



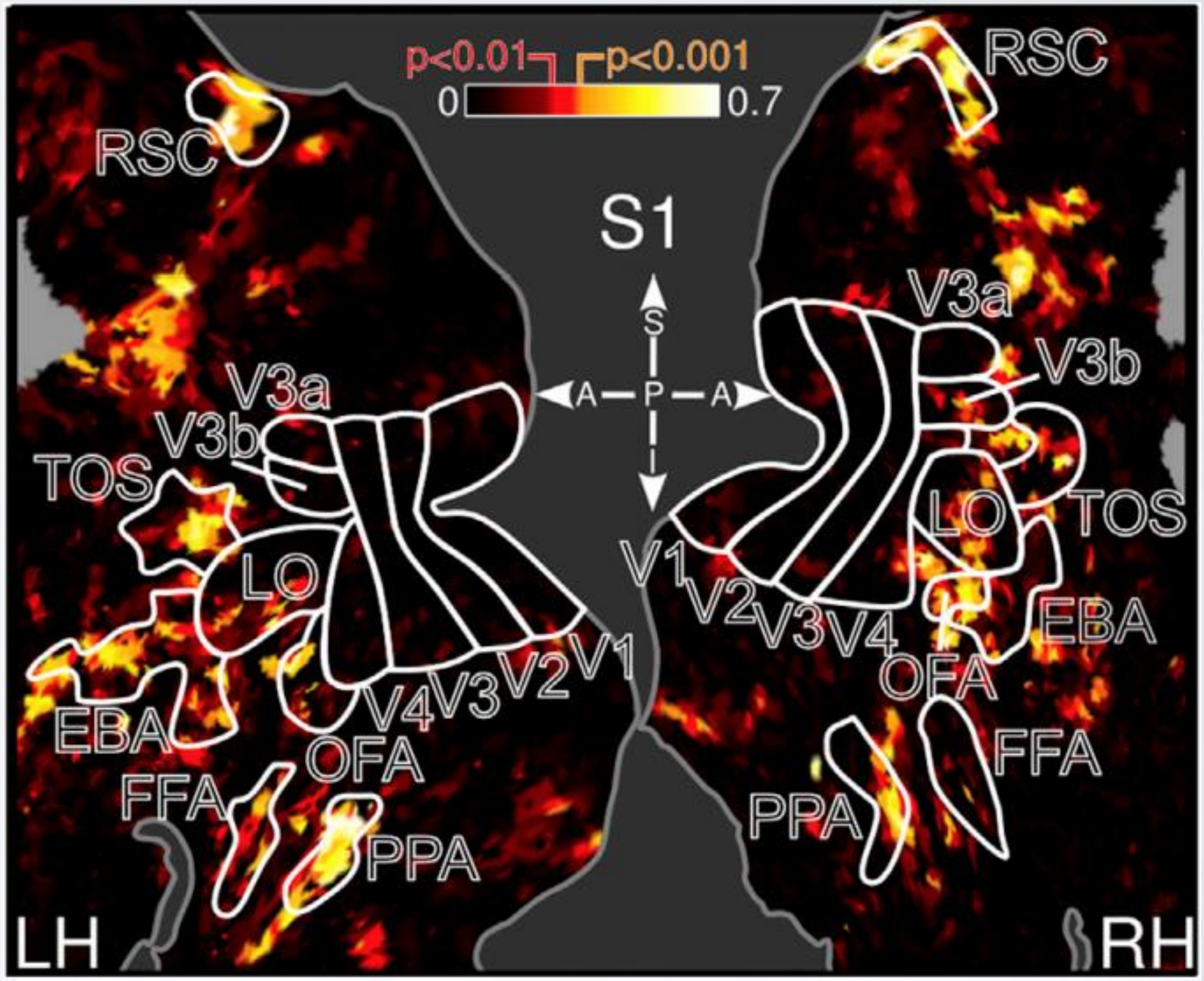
Stimulus *fMRI Scans* *Activity Vectors*

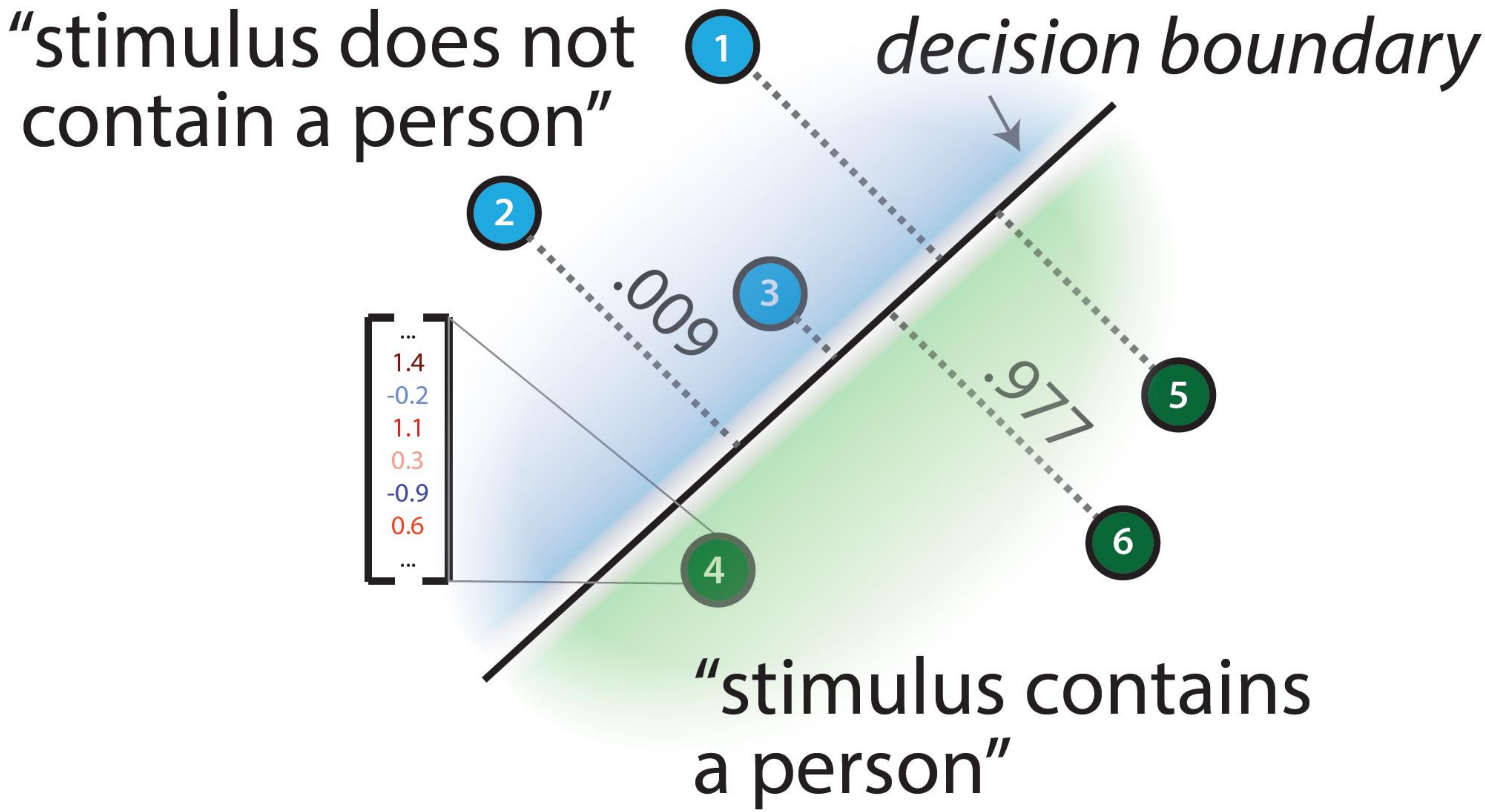
1. Collect Activity Vectors

Data Collection: Stansbury et al., 2013

Brain Areas

- 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)

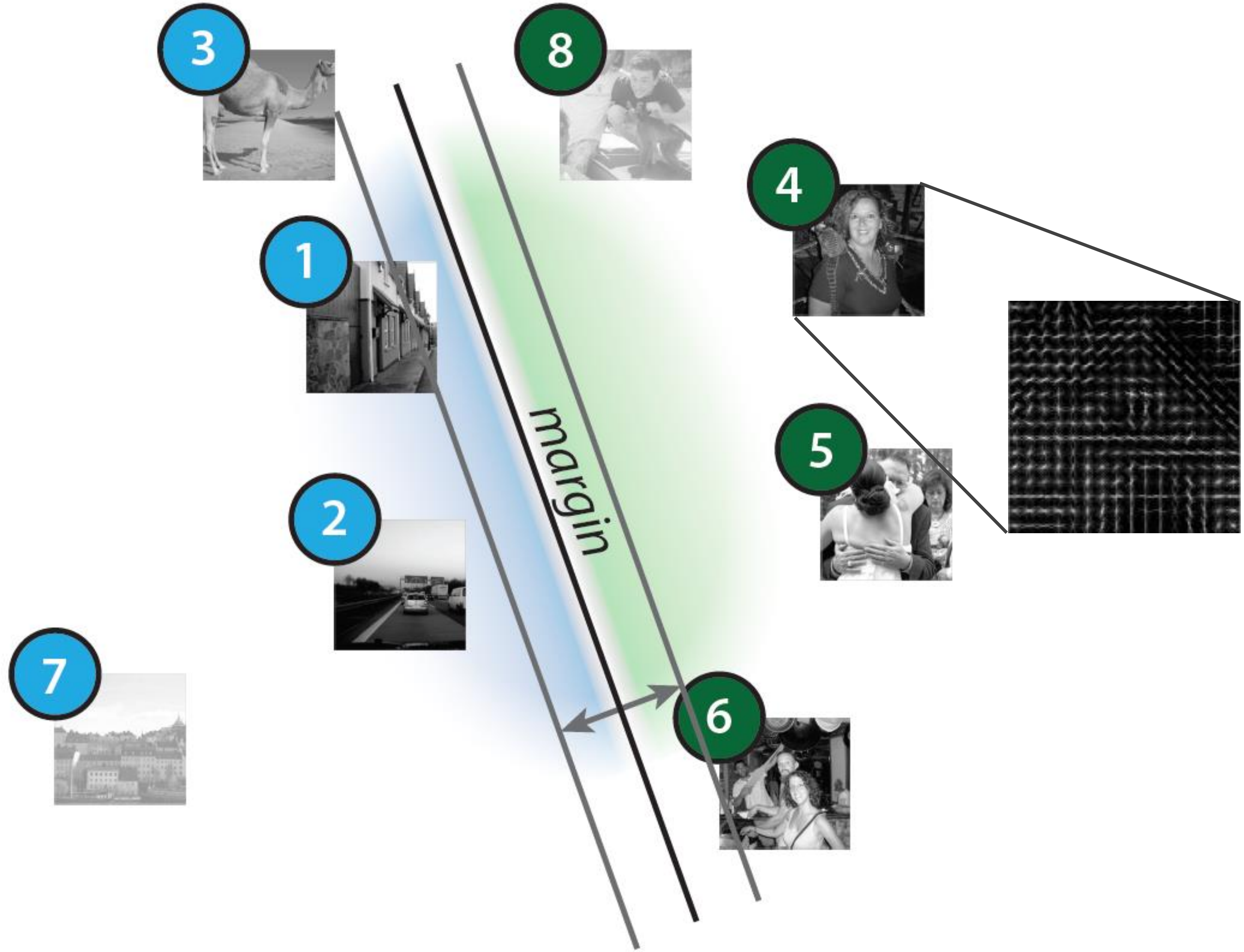




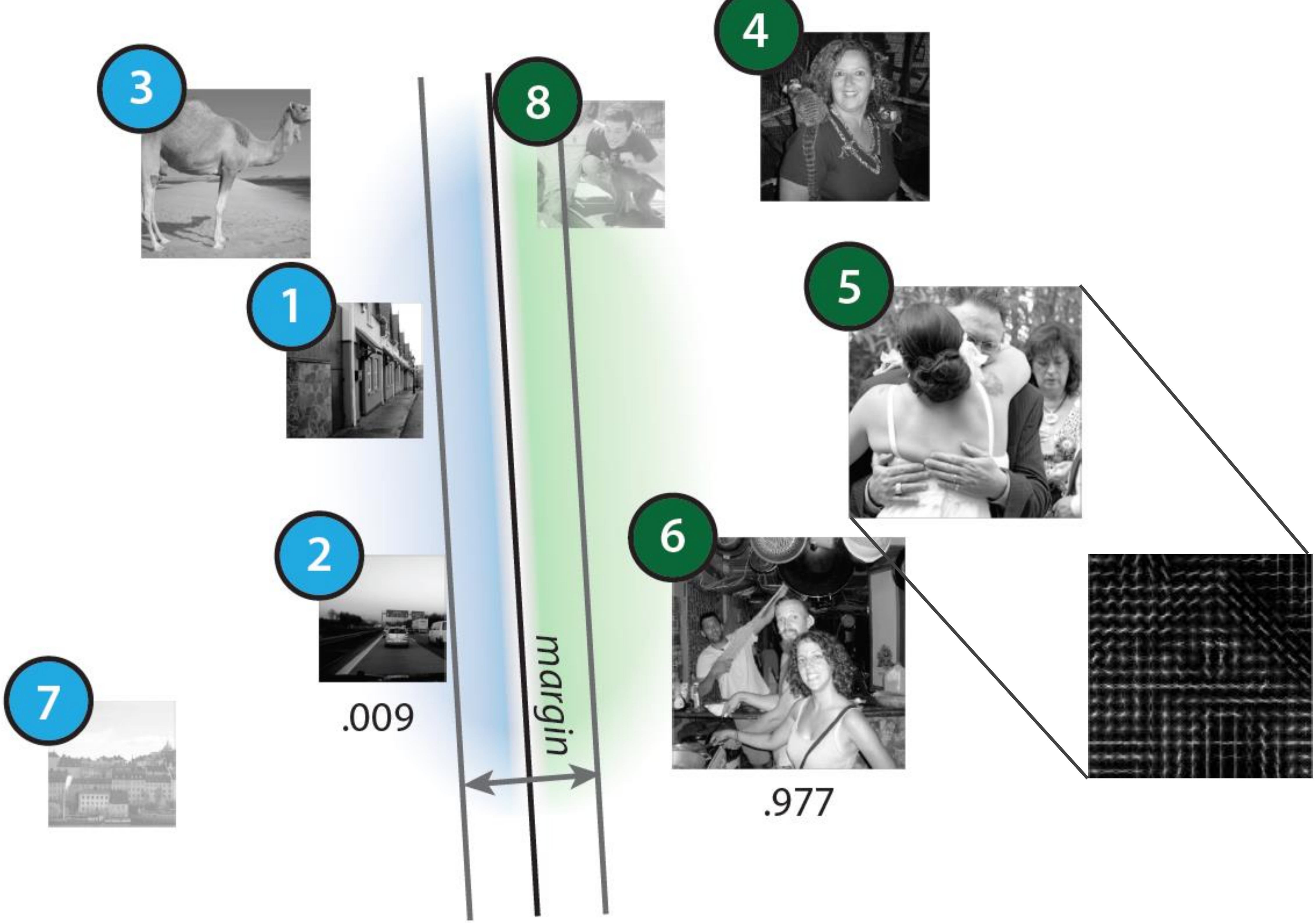
2. Train a classifier on fMRI Activity Vectors

P(human | EBA)

Based on the brain activity in the EBA region, what's the probability that an image contains a human?



3. Train "Vanilla-Flavored" Image Classifier as our Baseline



4. Train "Activity-Weighted" Classifier

(weight examples based on activity weight from neural data)

Activity Weighted Loss

Binary Classification

Datum:	$\vec{x} \in \mathbb{R}^D$
Label:	$y \in \{+1, -1\}$
Function:	$f : \mathbb{R}^D \to \{+1, -1\}$
Prediction:	$z = y \cdot f(\vec{x})$

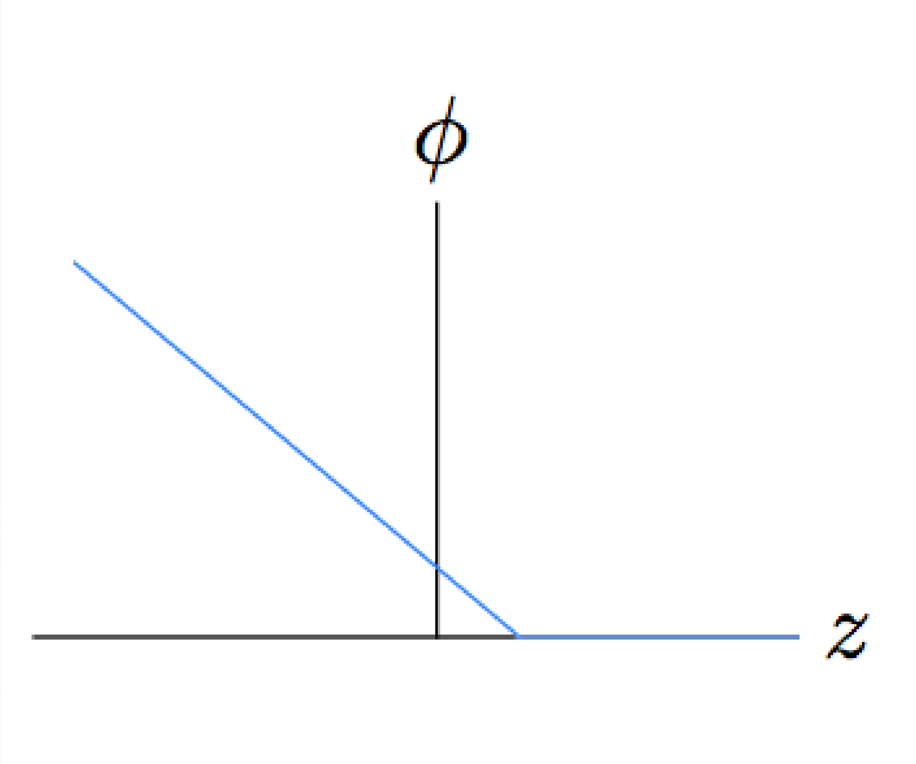
Loss Functions

- Prediction given by *z*:

$$z = y \cdot f(\vec{x})$$

- Typical Loss Function: Hinge Loss

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



Activity Weighted Loss

- Suppose we had a cost, *c*, for training sample, in addition to the datum vector, *x*, and its label, *y*.

$$\phi(\vec{x}, z) = \max(0, (1 - z)\, M(\vec{x}, z))$$

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

Optimization Problem

$$\min \frac{1}{2}||\vec{w}||^2 + C \sum_{l=1}^{L} \phi(\vec{x}_l, y_l \cdot f(\vec{x}_l))$$

Experimental Setup

Experimental Setup

- 4 mutually exclusive object categories (1386 total images)
  - Humans (219 images, 34%)
  - Animals (180 images; 28%)
  - Buildings (151 images; 23%)
  - Foods (59 images; 9%)

Experimental Setup

1. Generate image features (HOG or CaffeNet) for each image

2. Divide images into training and test sets (80% training; 20% test)

3. Learn activity weights from fMRI data of training images using an RBF-kernel SVM and 5-fold cross validation

4. Use activity weights to train another SVM on image features from training set (baseline: no weights)

5. Test SVM models with image features from test set

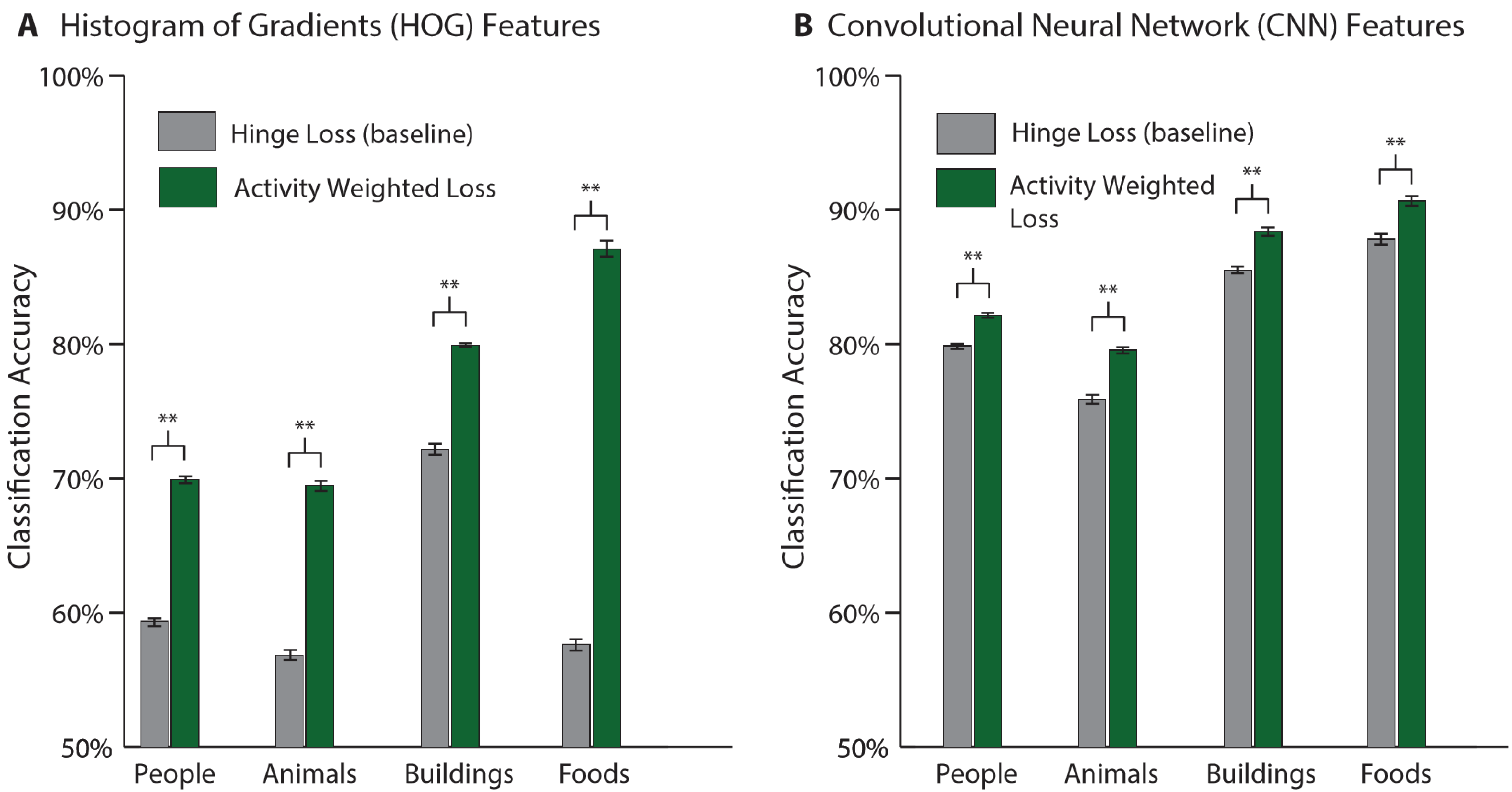
Experimental Setup

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

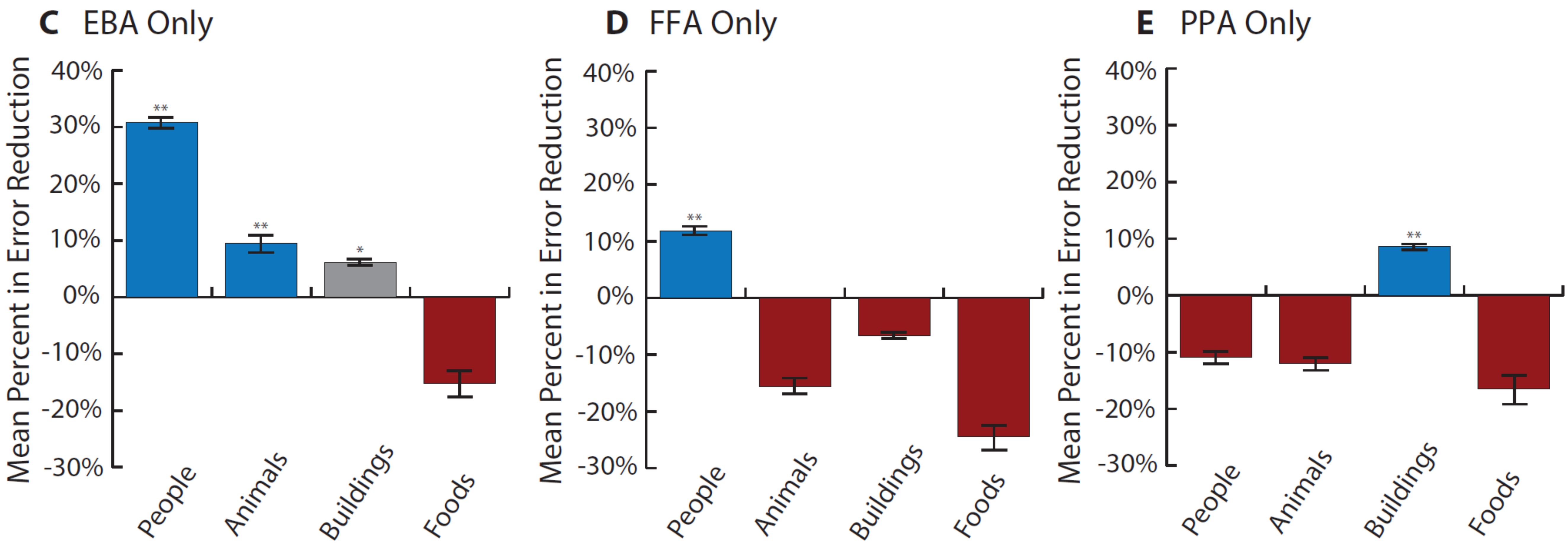
2. Set up 127 parallel experiments for the 127 combinations of 7 ROIs.

2. Set up 5 balanced classification problems.

Results

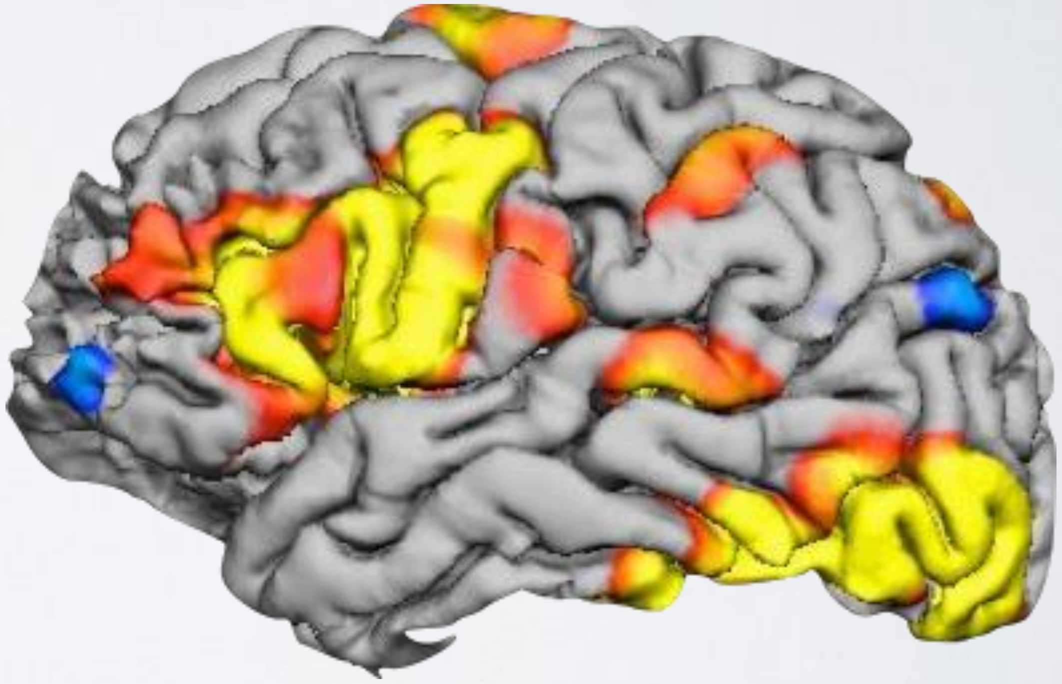


A Histogram of Gradients (HOG) Features
B Convolutional Neural Network (CNN) Features
100%
90%
80%
70%
60%
50%
Classification Accuracy
Hinge Loss (baseline)
Activity Weighted Loss
**
People
Animals
Buildings
Foods

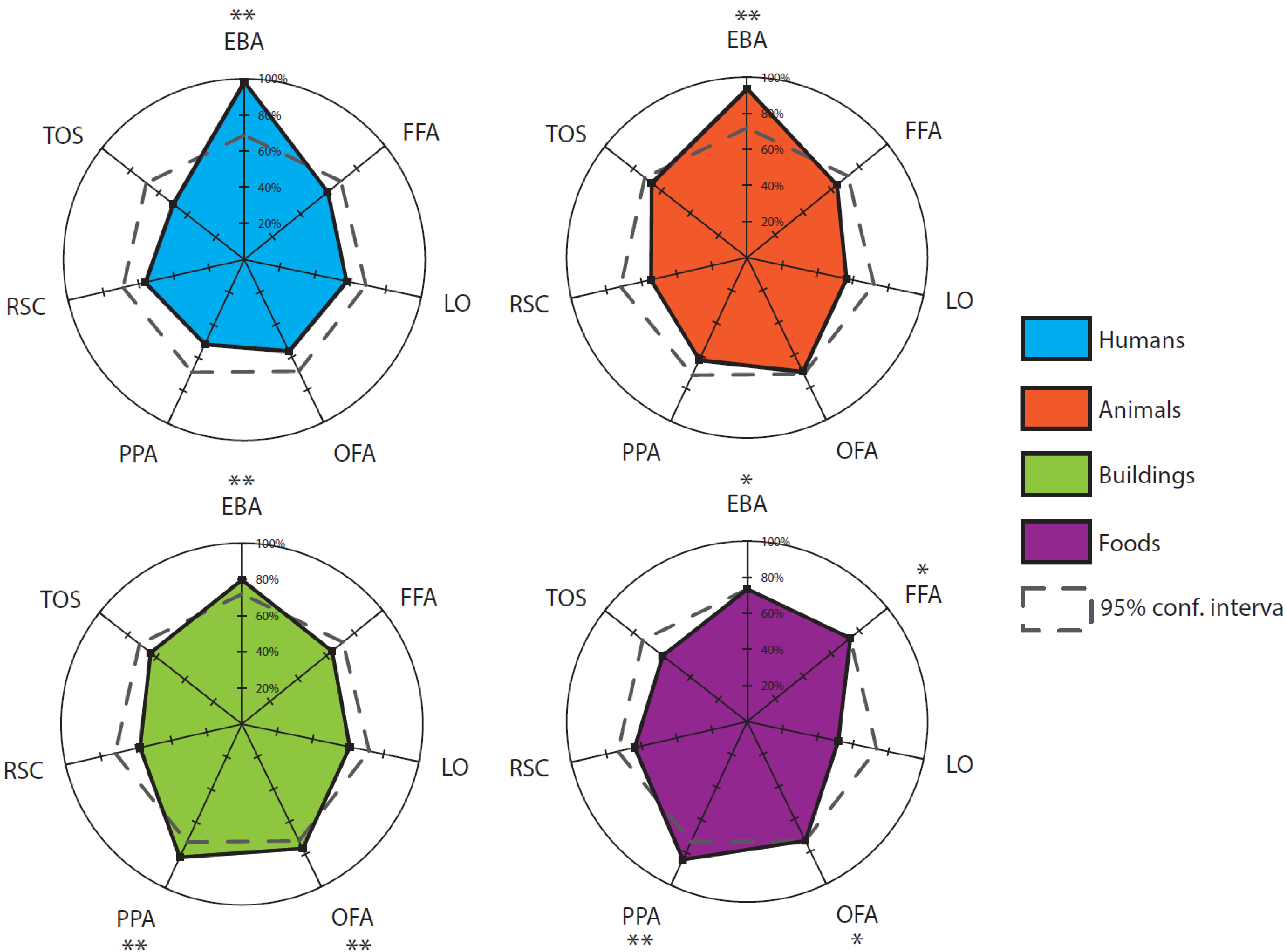

C EBA Only
D FFA Only
E PPA Only
Mean Percent in Error Reduction
40%
30%
20%
10%
0%
-10%
-20%
-30%
**
*
People
Animals
Buildings
Foods

Which Regions Are Most Helpful?

- 7 Regions:
  - eba, ffa, ppa, lo, rsc, tos, ofa
- 127 Combinations:
  - eba+ffa, eba+ffa+ppa, etc.



Analysis of ROIs (HOG)



Next Steps

1. Extend to CNNs

2. Compare with other tasks and sources of guidance

3. Guide feature learning more directly

Related works:

1. Focal Loss: Lin et al., ICCV 2017

2. Learn similar features to EEG prediction: Spampinato et al., CVPR 2017

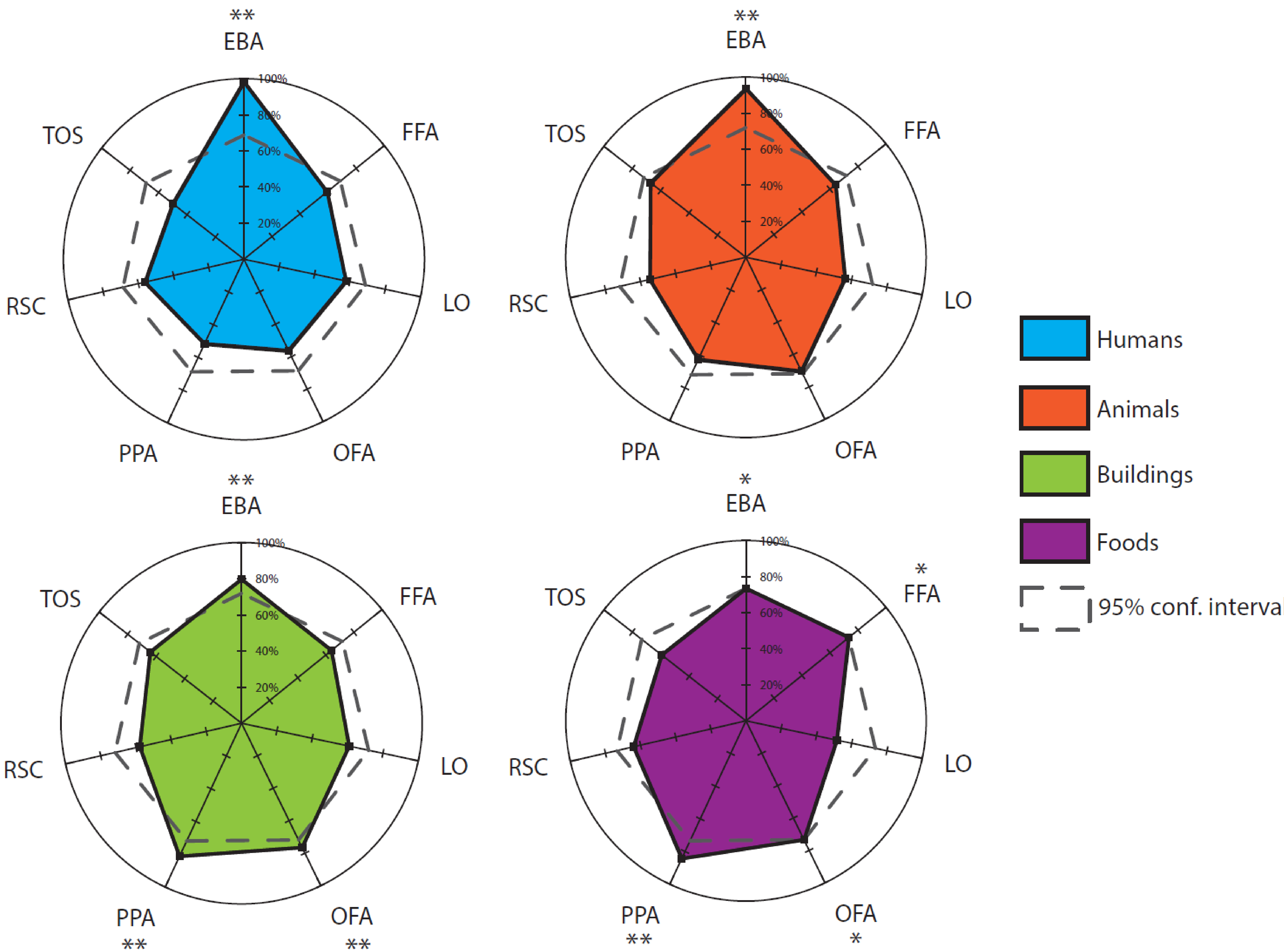
3. CNNs vs. fMRI vs. EEG: Cichy et al., Scientific Reports 2016

Thanks

David Cox, Walter Scheirer, Dustin Stansbury,
Cox Lab, Gallant Lab, “Bear”

Questions?

Analysis of ROIs (HOG)



		Mean Acc*	> Avg. Mean Acc (66.36%)
1	**EBA**	**72.02%**	**1**
2	**EBA, FFA**	**70.90%**	**1**
3	**EBA, LO**	**69.80%**	**1**
	...		
64	**EBA, FFA, LO, OFA, PPA, RSC, TOS**	**69.93%**	**1**
65	FFA	64.64%	0
66	FFA, LO	64.86%	0
67	FFA, OFA	65.67%	0
	...		
127	FFA, LO, OFA, PPA, RSC, TOS	67.15%	1
	Overall Average:	66.36%	*54.33%*
	Average for all bolded ROI combinations:	*69.91%*	98.44%

		Mean Acc*	> Avg. Mean Acc (66.36%)
1	**EBA**	**72.02%**	**1**
2	EBA, FFA	70.90%	1
3	EBA, LO	69.80%	1
	...		
64	**EBA, FFA, LO, OFA, PPA, RSC, TOS**	**69.93%**	**1**
65	FFA	64.64%	0
66	**FFA, LO**	**64.86%**	**0**
67	**FFA, OFA**	**65.67%**	**0**
	...		
127	FFA, LO, OFA, PPA, RSC, TOS	67.15%	1
	Overall Average:	66.36%	*54.33%*
	Average for all bolded ROI combinations:	*67.06%*	60.94%