Categories Versus Exemplars

Ensemble of Exemplar-SVMs for Object Detection and Beyond

What is this about?

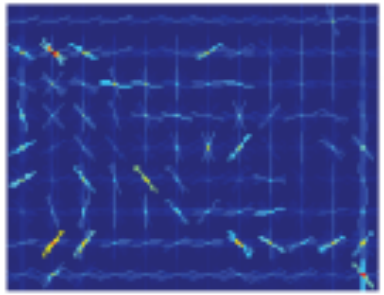
- Categories => Over generalization of objects

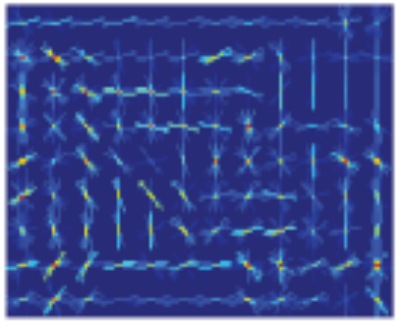
What is this about?

- Exploit distinction

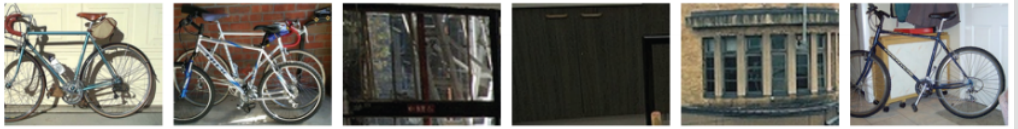
vs

IS versus IS NOT





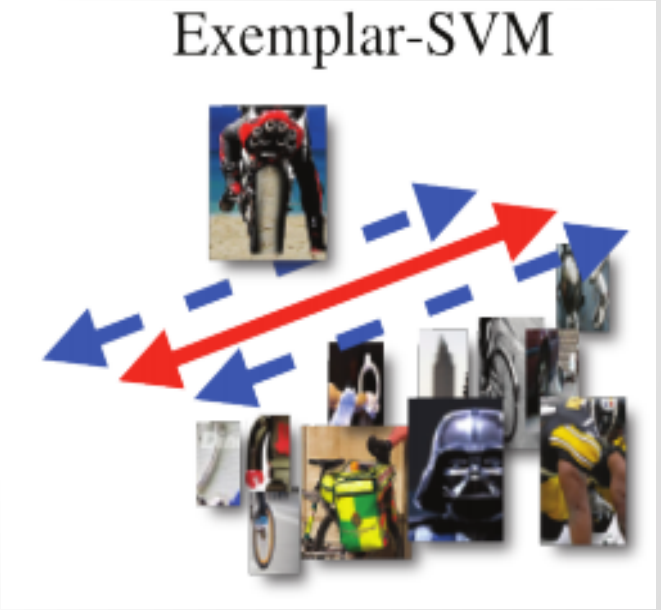
Exemplar-SVM based on per-exemplar distance



Exemplar-SVM

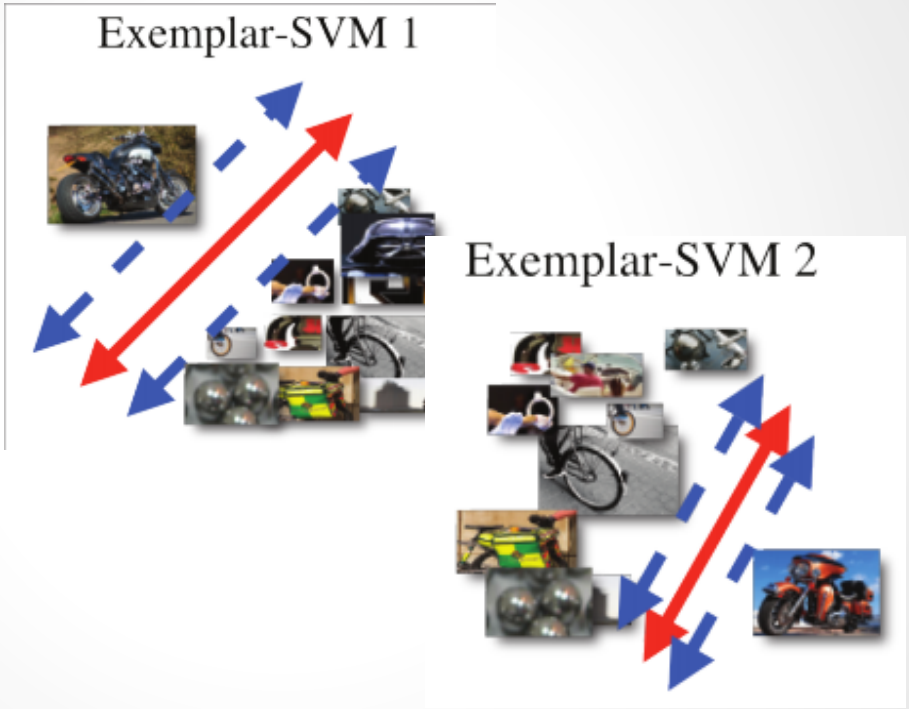
How is it done?

- SVMs are trained for one against rest
- Huge data is represented as decision boundary



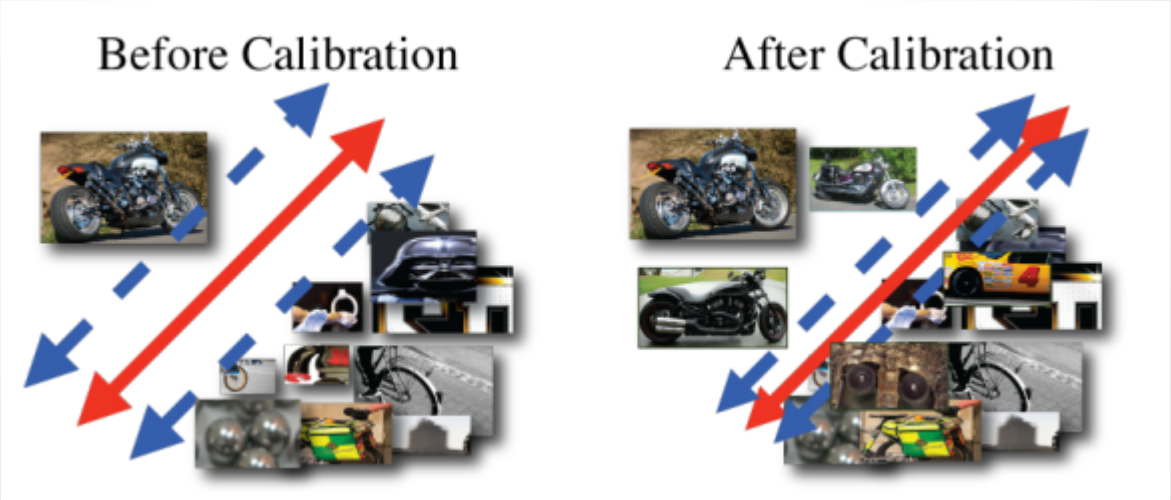
How is it done?

vs

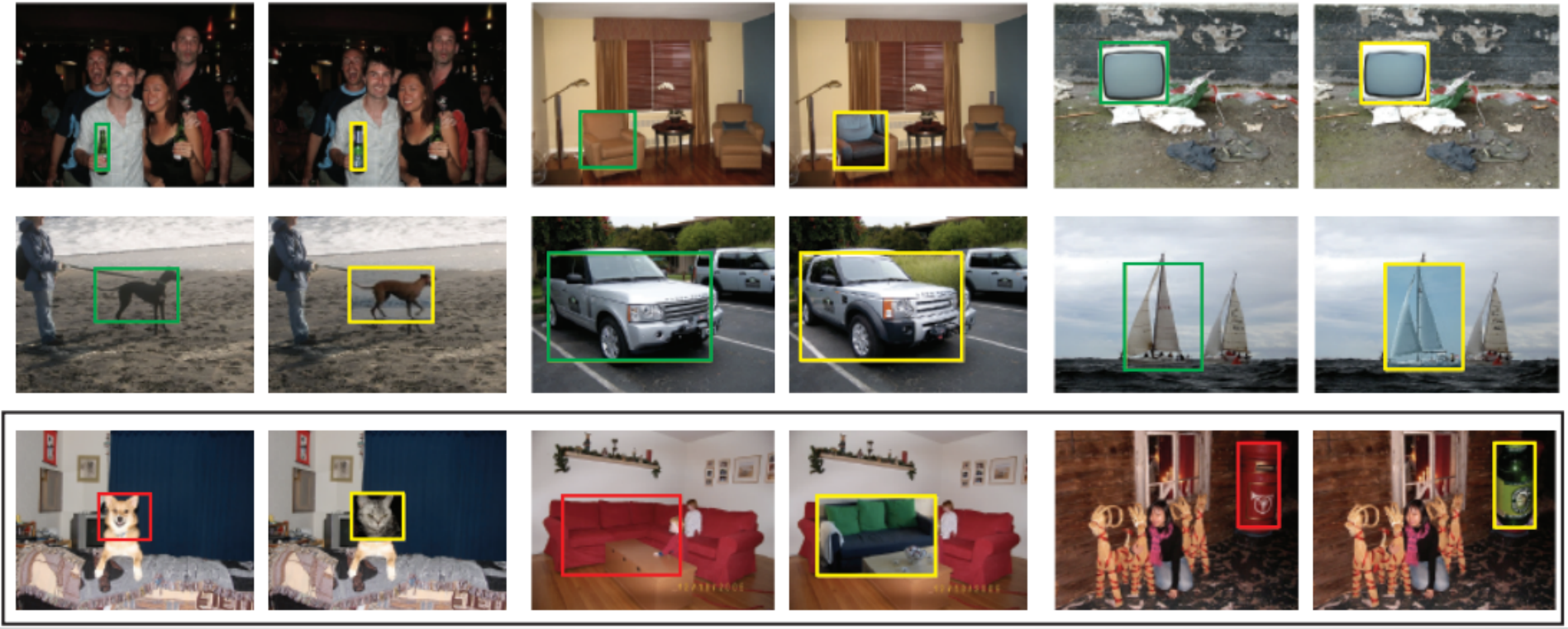


How is it done?

- To avoid overfitting, used regularized linear SVMs
- Calibration for comparable scores



Results



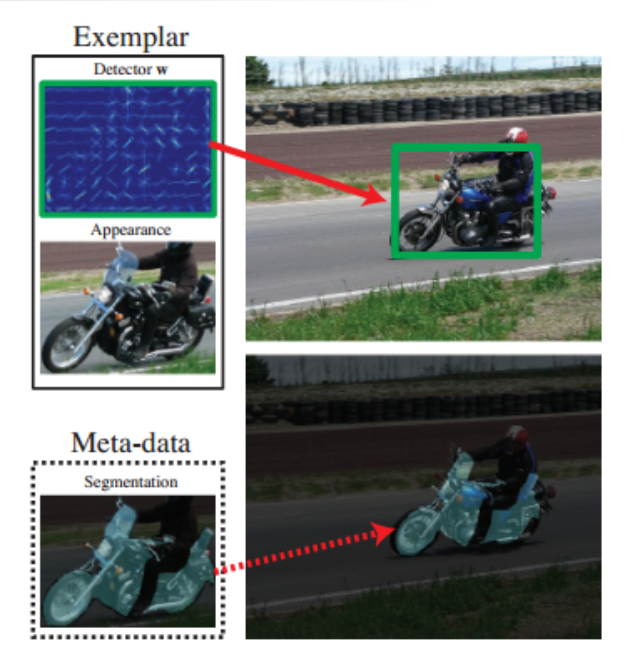
Results

Approach	aeroplane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	diningtable	dog	horse	motorbike	person	pottedplant	sheep	sofa	train	tvmonitor	mAP
NN	.006	.094	.000	.005	.000	.006	.010	.092	.001	.092	.001	.004	.096	.094	.005	.018	.009	.008	.096	.144	.039
NN+Cal	.056	.293	.012	.034	.009	.207	.261	.017	.094	.111	.004	.033	.243	.188	.114	.020	.129	.003	.183	.195	.110
DFUN+Cal	.162	.364	.008	.096	.097	.316	.366	.092	.098	.107	.002	.093	.234	.223	.109	.037	.117	.016	.271	.293	.155
E-SVM+Cal	.204	.407	.093	.100	.103	.310	.401	.096	.104	.147	.023	.097	.384	.320	.192	.096	.167	.110	.291	.315	.198
E-SVM+Co-occ	.208	.480	.077	.143	.131	.397	.411	.052	.116	.186	.111	.031	.447	.394	.169	.112	.226	.170	.369	.300	.227
CZ [6]	.262	.409	–	–	–	.393	.432	–	–	–	–	–	–	.375	–	–	–	–	.334	–	–
DT [7]	.127	.253	.005	.015	.107	.205	.230	.005	.021	.128	.014	.004	.122	.103	.101	.022	.056	.050	.120	.248	.097
LDPM [9]	.287	.510	.006	.145	.265	.397	.502	.163	.165	.166	.245	.050	.452	.383	.362	.090	.174	.228	.341	.384	.266

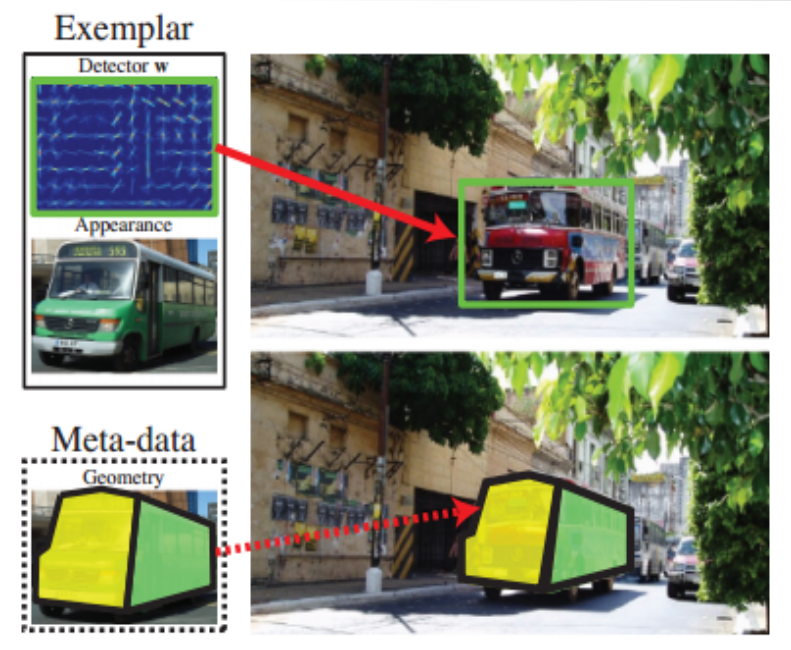
Table 1. **PASCAL VOC 2007 object detection results.** We compare our full system (ESVM+Co-occ) to four different exemplar based baselines including NN (Nearest Neighbor), NN+Cal (Nearest Neighbor with calibration), DFUN+Cal (learned distance function with calibration) and ESVM+Cal (Exemplar-SVM with calibration). We also compare our approach against global methods including our implementation of Dalal-Triggs (learning a single global template), LDPM [9] (Latent deformable part model), and Chum et al. [6]'s exemplar-based method. [The NN, NN+Cal and DFUN+Cal results for person category are obtained using 1250 exemplars]

Deliverables

- Segmentation Transfer



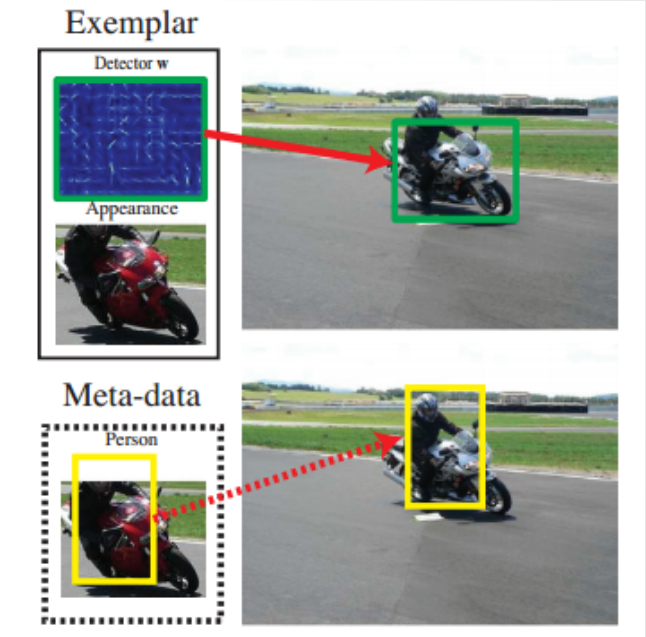
- Geometry Transfer



Deliverables

- 3D Model Transfer

- Priming



Relationship Between Category and Exemplars: A Neuroscientific Analysis

What?

- Behavioral Experiment
  - Forced-choice categorization
- fMRI Experiment
  - Passive viewing

Where?

- 5 ROIs:
  a. V1
  b. Parahippocampal Place Area
  c. Retrosplenial Cortex
  d. Lateral Occipital Complex
  e. Fusiform Face Area

Results: Behavioral Experiment

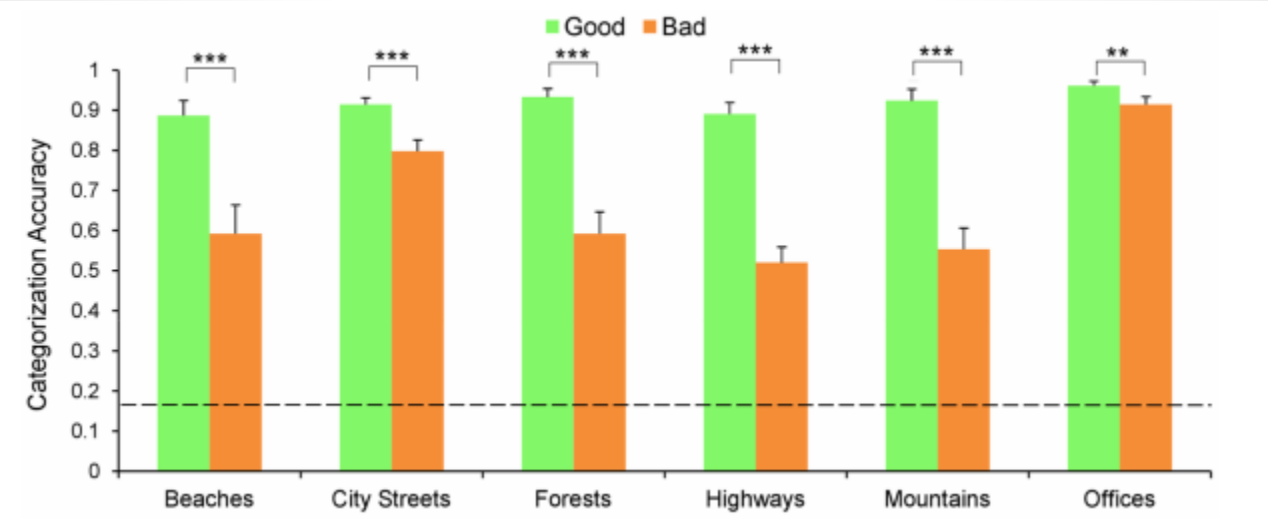


Image Analysis

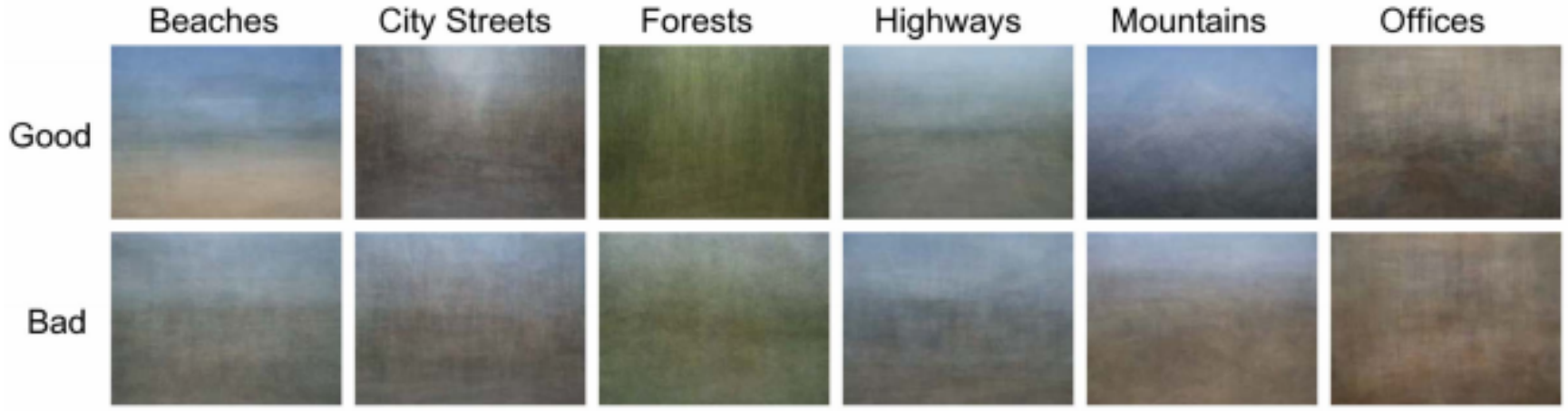
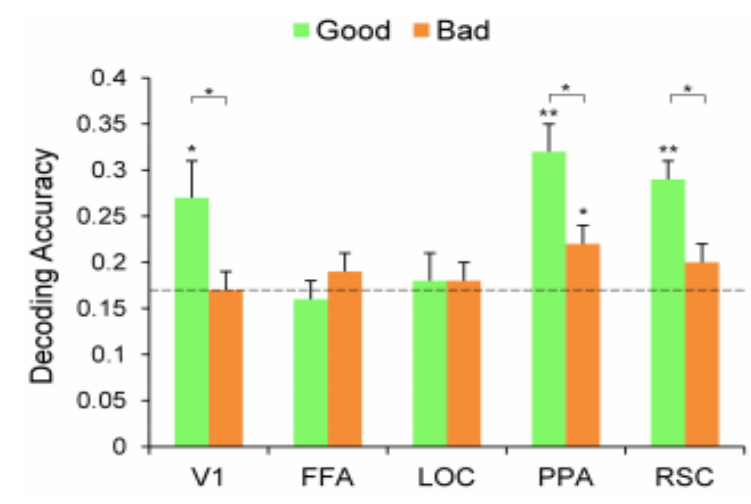


Figure 6. Average images. Pixel-wise RGB average images of good (first row) and bad (second row) exemplars across the categories (columns).

Results: fMRI Multivariate



Decoder prediction

Category presented

V1

	b	c	f	h	m	o
beaches	.32	.09	.10	.17	.19	.13
city streets	.08	.38	.15	.10	.13	.15
forests	.10	.32	.23	.10	.02	.21
highways	.23	.13	.10	.21	.17	.15
mountains	.15	.10	.13	.15	.28	.19
offices	.17	.21	.20	.06	.15	.21

V1

	b	c	f	h	m	o
beaches	.06	.23	.18	.12	.28	.12
city streets	.10	.23	.21	.17	.13	.15
forests	.15	.23	.10	.04	.30	.17
highways	.12	.25	.13	.08	.21	.21
mountains	.20	.15	.08	.23	.27	.07
offices	.17	.17	.08	.17	.17	.23

PPA

	b	c	f	h	m	o
beaches	.34	.13	.08	.10	.24	.10
city streets	.17	.38	.07	.11	.00	.27
forests	.19	.05	.32	.08	.21	.15
highways	.15	.27	.13	.17	.10	.18
mountains	.15	.00	.17	.13	.49	.06
offices	.10	.28	.11	.30	.00	.21

PPA

	b	c	f	h	m	o
beaches	.17	.10	.15	.10	.31	.16
city streets	.08	.21	.21	.16	.17	.17
forests	.19	.2	.22	.18	.23	.06
highways	.08	.24	.11	.17	.23	.17
mountains	.27	.13	.17	.11	.25	.06
offices	.10	.21	.06	.17	.13	.32

RSC

	b	c	f	h	m	o
beaches	.19	.17	.04	.21	.22	.17
city streets	.08	.51	.02	.15	.07	.17
forests	.17	.10	.23	.07	.34	.08
highways	.15	.30	.28	.17	.04	.07
mountains	.15	.05	.23	.11	.34	.12
offices	.08	.30	.17	.11	.04	.29

RSC

	b	c	f	h	m	o
beaches	.19	.21	.19	.07	.20	.15
city streets	.15	.30	.08	.15	.04	.27
forests	.21	.15	.23	.08	.15	.18
highways	.15	.19	.17	.15	.19	.15
mountains	.12	.19	.22	.15	.15	.17
offices	.12	.20	.15	.13	.21	.19

Good

Bad

Classical VS Prototypical Approach

While training models in CV, why are all training examples considered equal?

Does a Trade-off Exist?

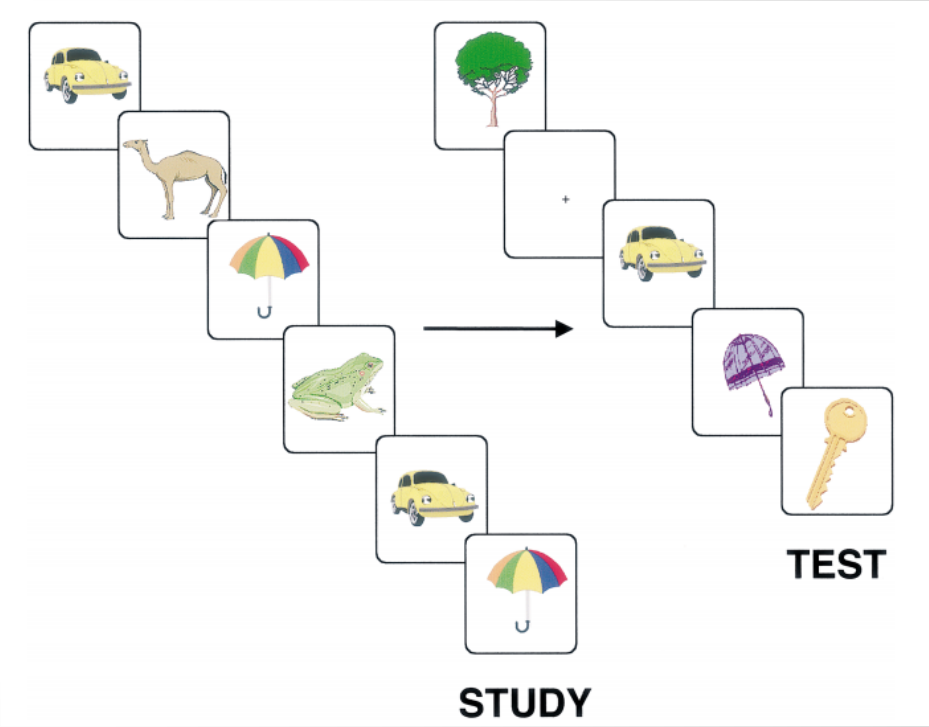
What?

Neural Priming!

?

What?

- Categorical pair of tokens
- Visually different
- Shoeboxing Experiment



Results

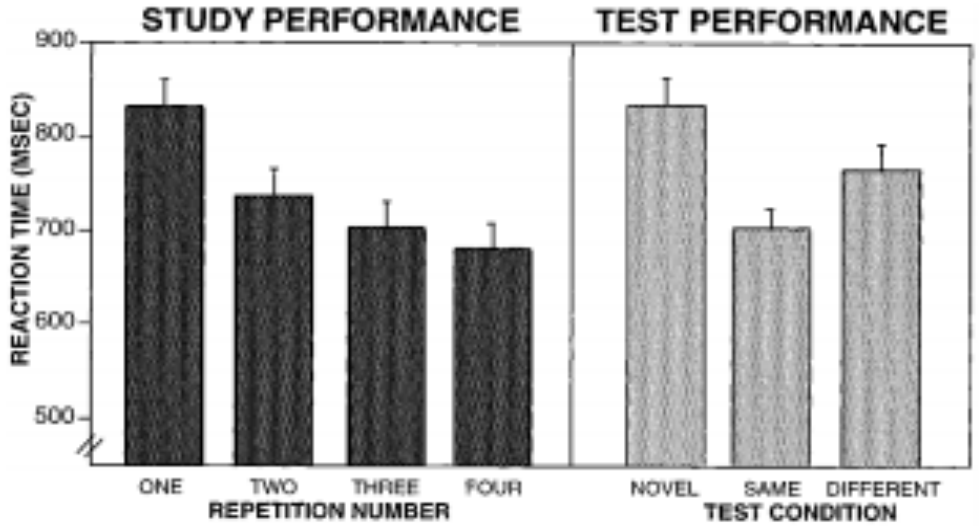
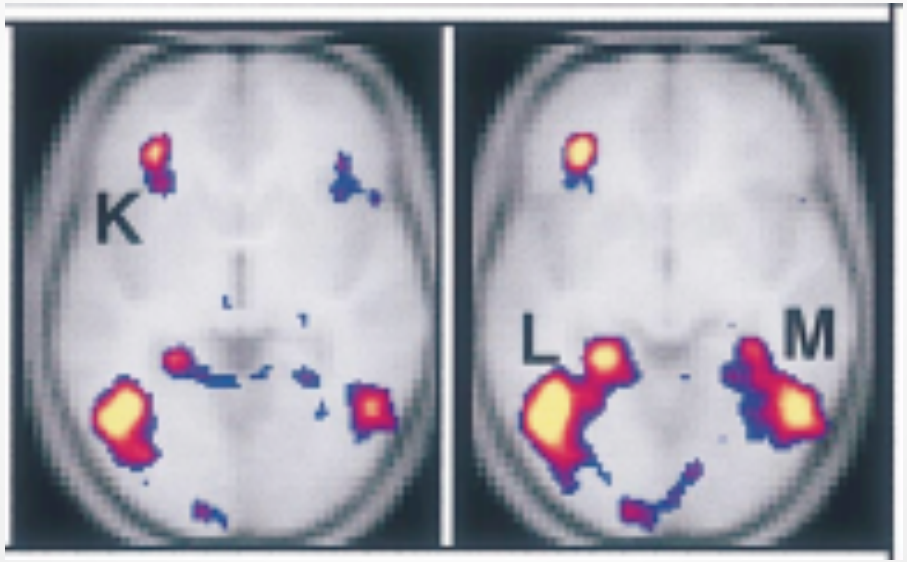


Fig. 2. Mean response latencies for the object classification task across the four repeated presentations of objects during the study phase (left panel) and for novel, repeated same, and repeated different objects in the test phase (right panel). Error bars indicate standard errors of the mean.

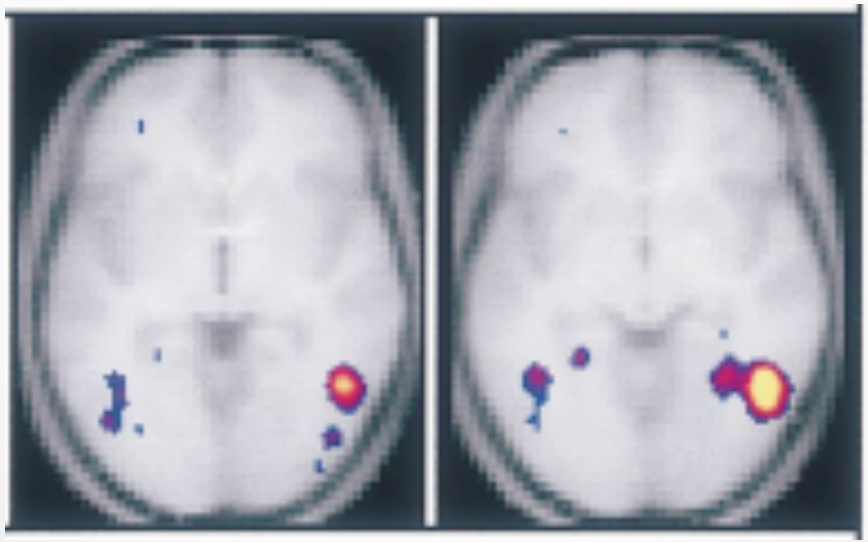
No Symmetry

Novel > All repeated

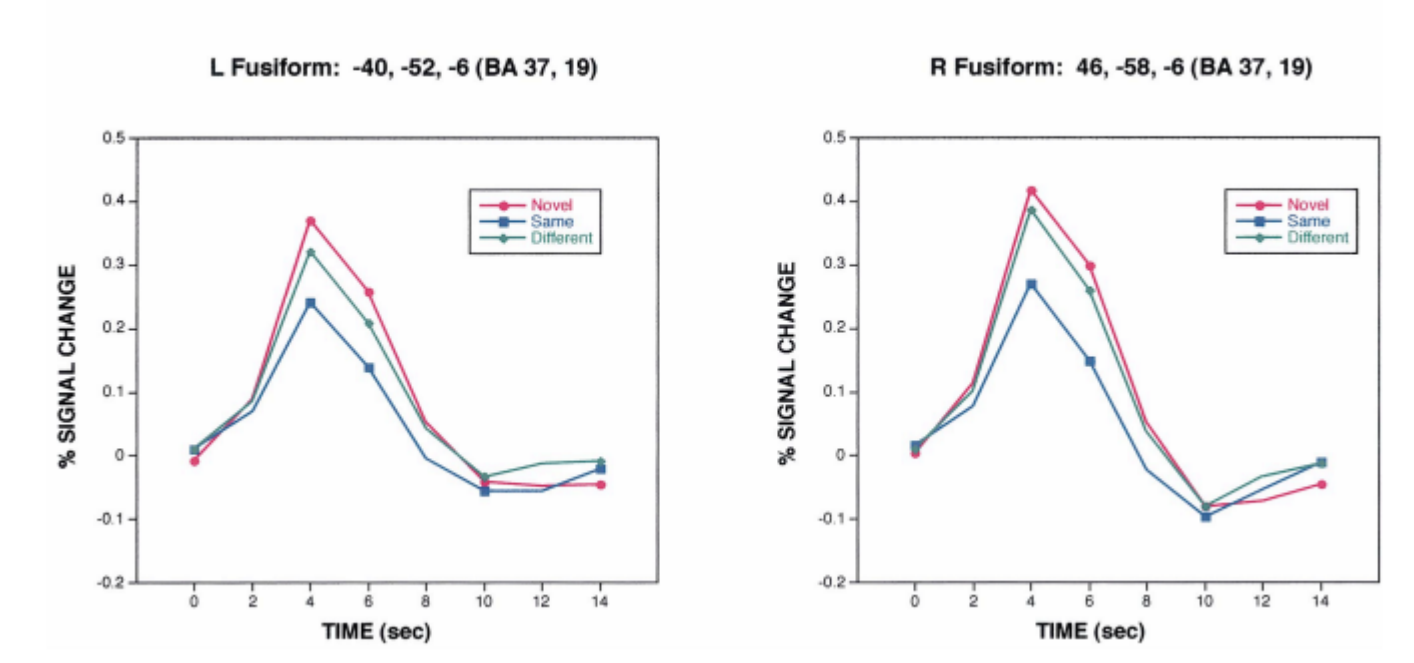


No Symmetry

Repeated Different > Repeated Same



Results


L Fusiform: -40, -52, -6 (BA 37, 19)
% SIGNAL CHANGE
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
0
2
4
6
8
10
12
14
TIME (sec)
Novel
Same
Different
R Fusiform: 46, -58, -6 (BA 37, 19)
% SIGNAL CHANGE
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
0
2
4
6
8
10
12
14
TIME (sec)
Novel
Same
Different

The overview

- Tokens from one category vary largely
- Not all represent the idea
- Brain uses the two systems in unison
- Can we?