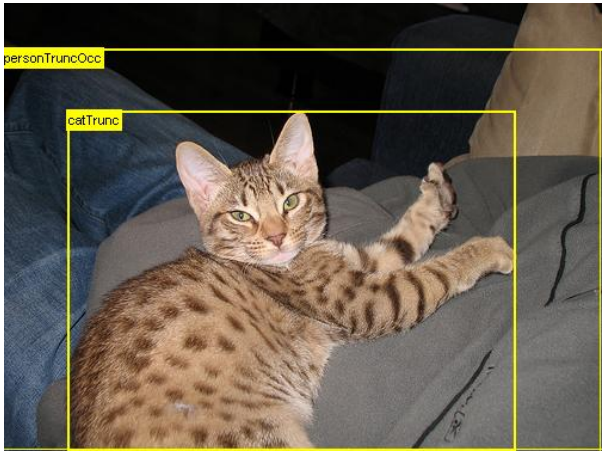
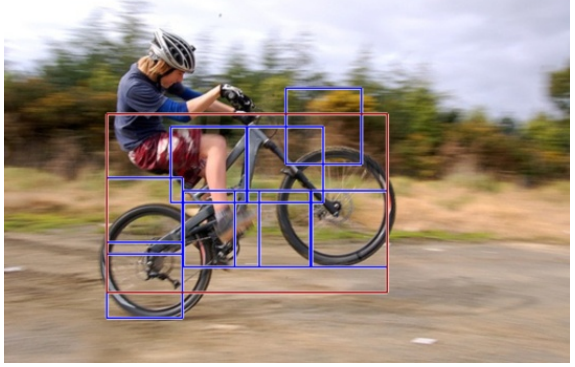
Semantic vs. Visual Subcategories in Computer Vision and Neuroscience (in IT)

Abhinav Shrivastava

Deformable Objects

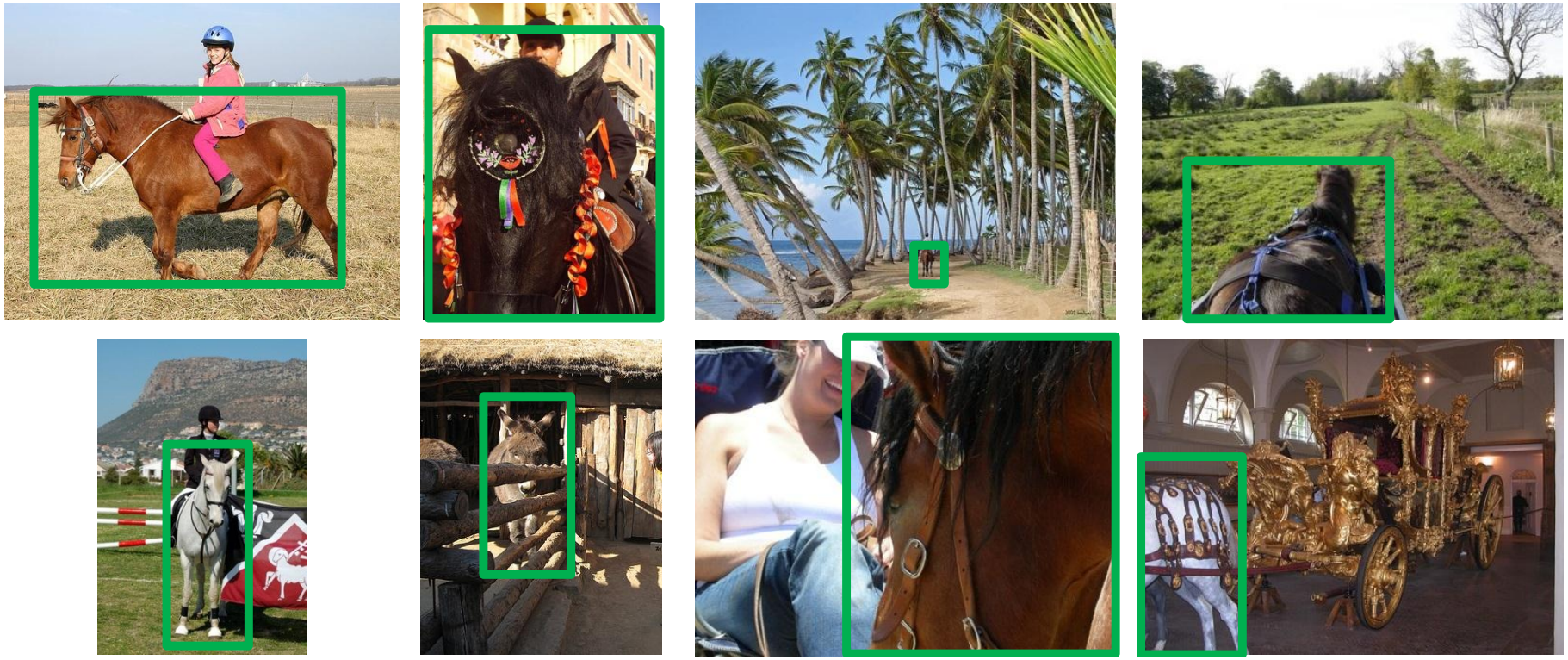





HONDA


MONGOOSE

Intra-class Diversity



Example images for “Horse” from PASCAL VOC

Variation due to camera viewpoints, object poses, occlusions, etc.

Slide from Divvala et al.

State-of-the-art Detector

Deformable Part-based Model (DPM)

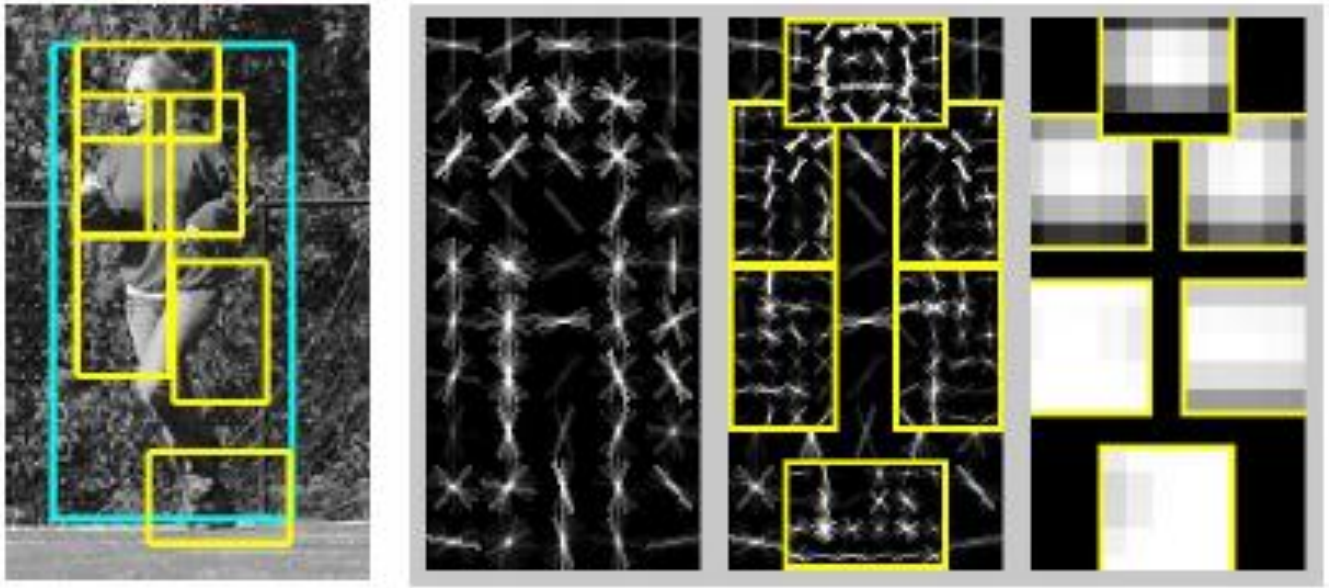


Image | Root Filter | Part Filters | Deformation Models

Parts in Broad-strokes

(a) Original (b) Foreshortening (c) Out-of-plane

Fig. 2: We show that 4 small, translating parts can approximate non-affine (e.g., perspective) warps.

Fig. from Yang & Ramanan

State-of-the-art Detector

Deformable Part-based Model (DPM)

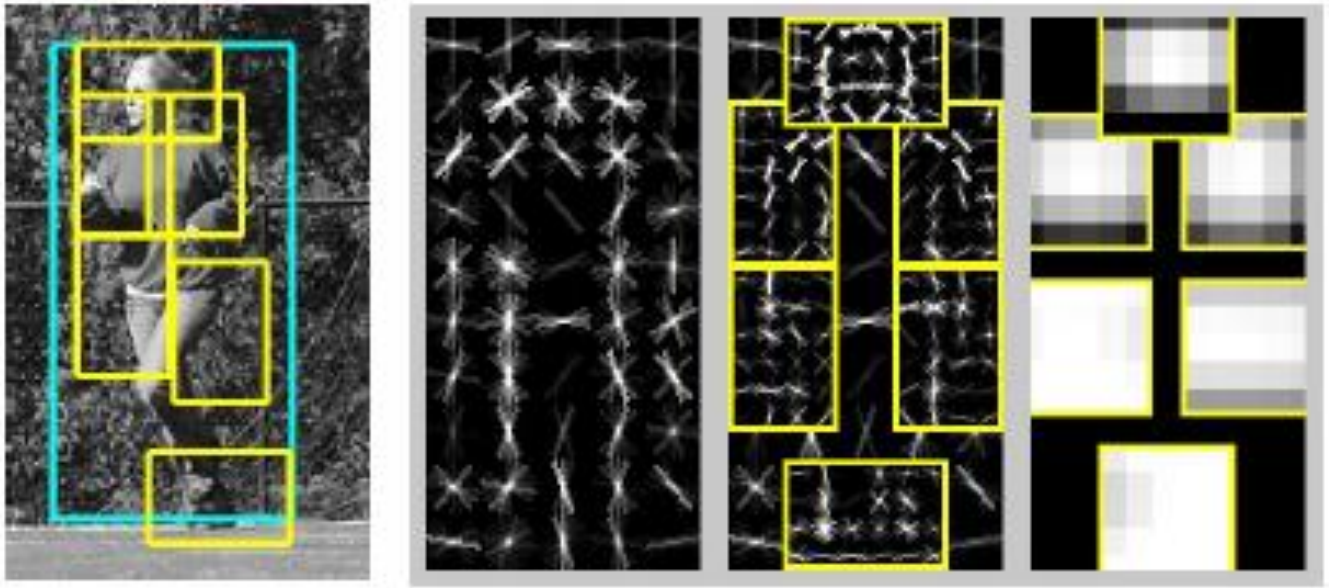
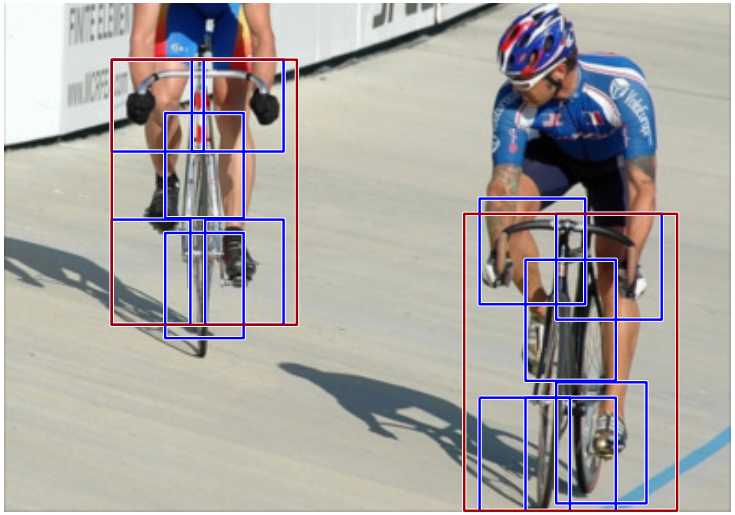
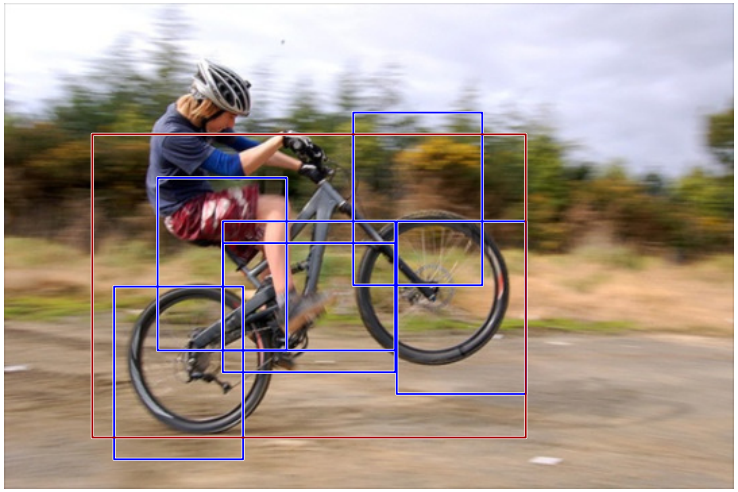


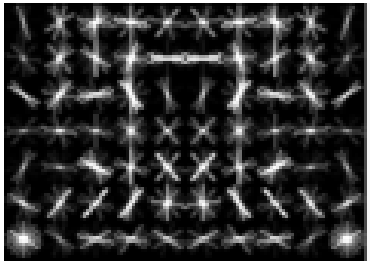
Image | Root Filter | Part Filters | Deformation Models

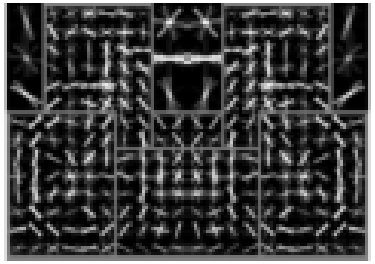
Is One Model Enough?

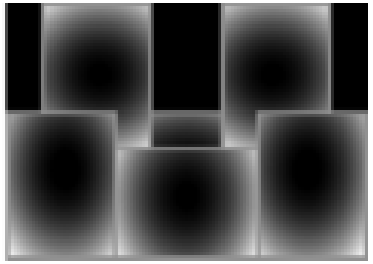
Mixture Models

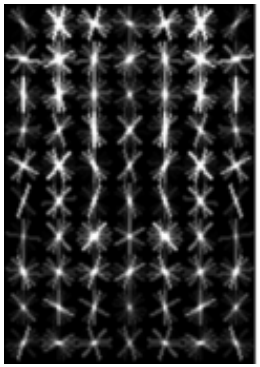


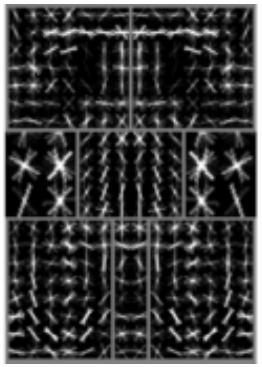


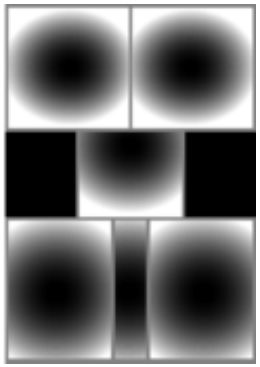






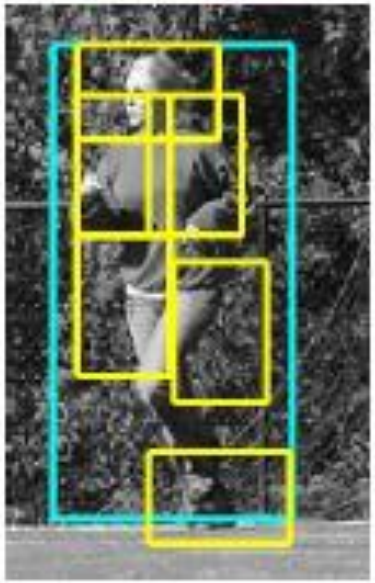




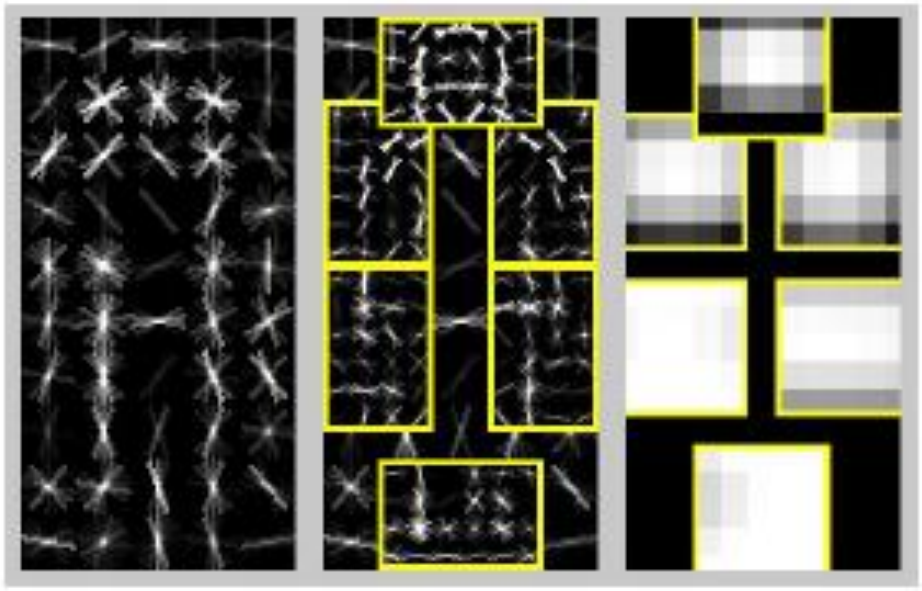


State-of-the-art Detector

Deformable Part-based Model (DPM)



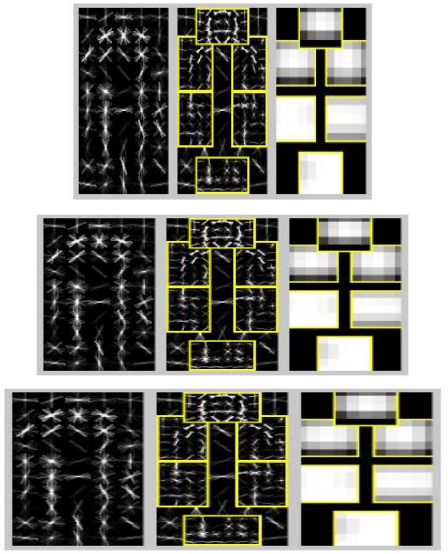
Image



Root Filter

Part Filters

Deformation Models



Mixture Models

State-of-the-art Detector

Deformable Part-based Model (DPM)

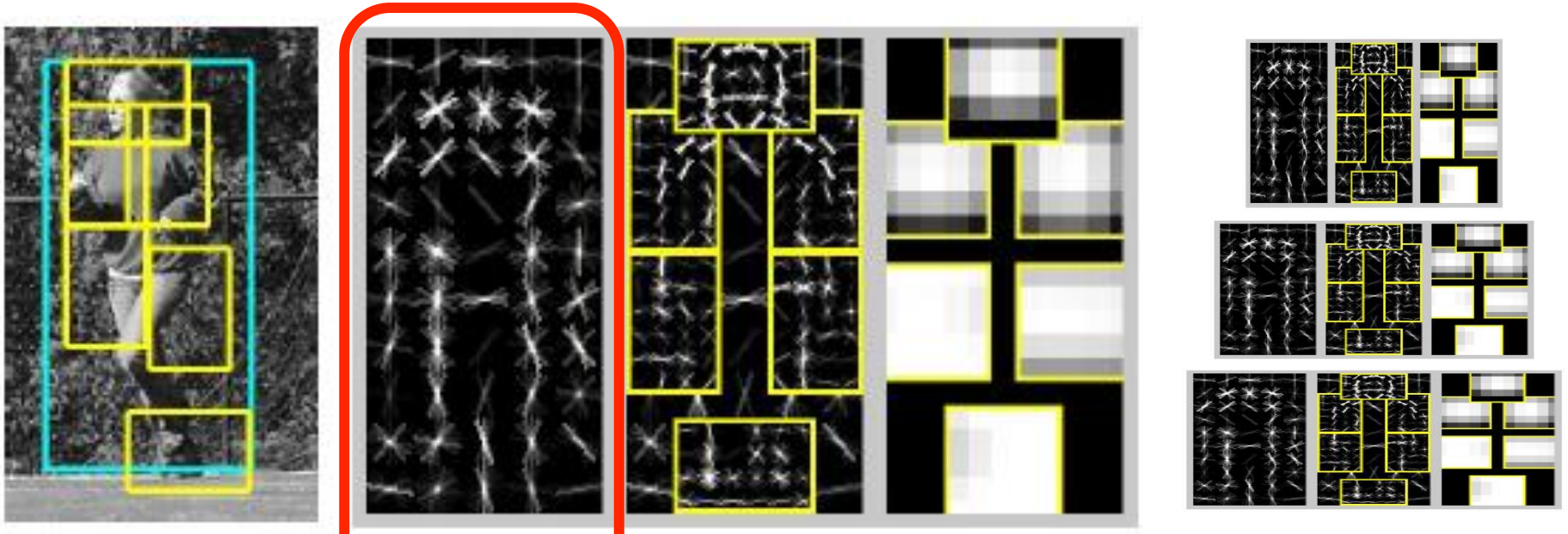


Image | Root Filter | Part Filters | Deformation Models | Mixture Models

Features, Data Mining & Classifier?

Papageorgiou & Poggio (2000)

Viola & Jones (2001)

Schneiderman et al. (2002, 03, 04 etc.)

Dalal & Triggs (2005), Freeman et al (1996), Lowe (1999)

....

State-of-the-art Detector

Deformable Part-based Model (DPM)

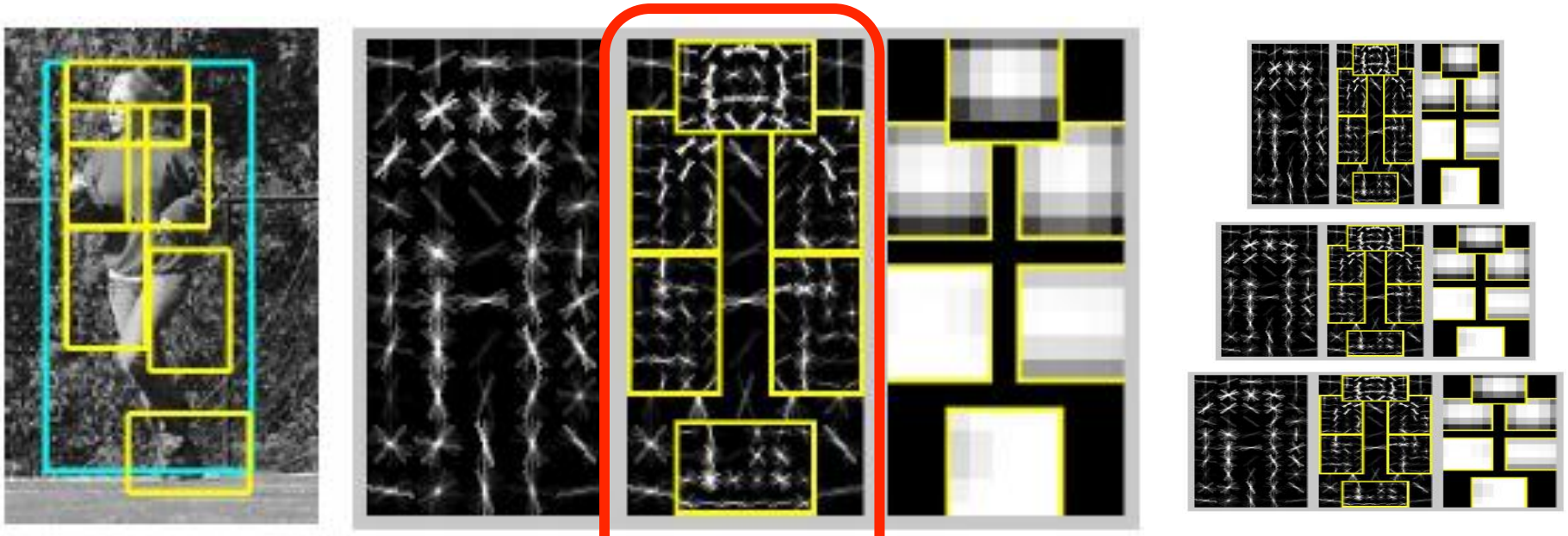


Image | Root Filter | Part Filters | Deformation Models | Mixture Models

Parts?

Mohal et al. (2001)

….

State-of-the-art Detector

Deformable Part-based Model (DPM)

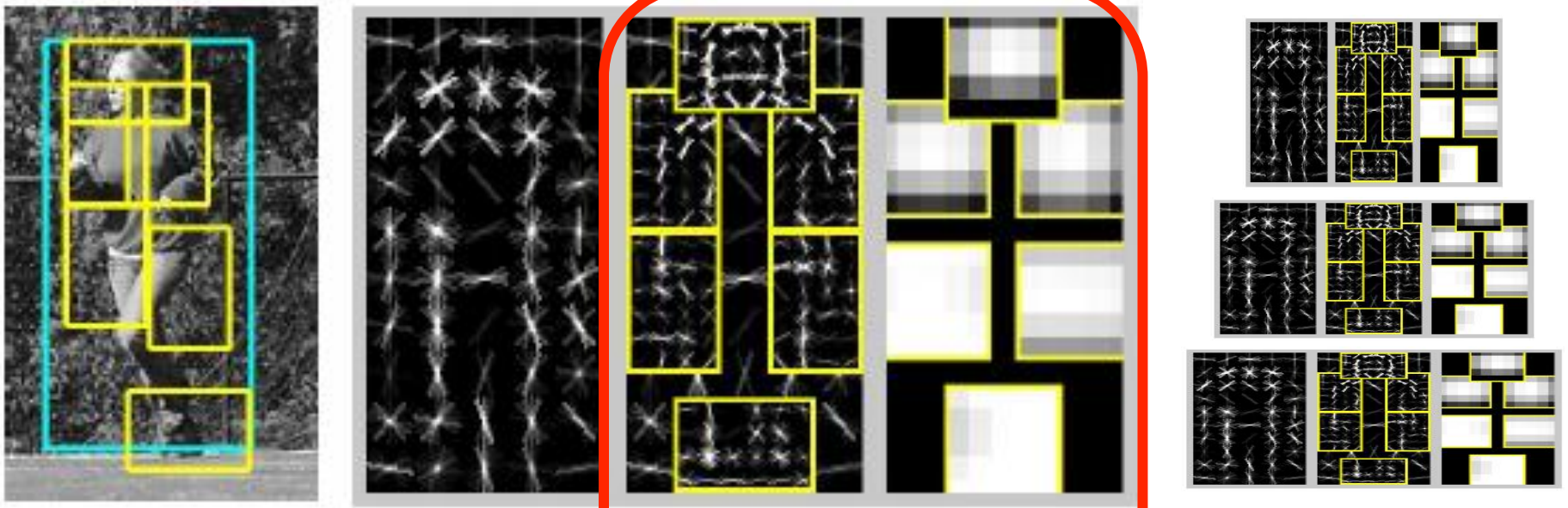


Image | Root Filter | Part Filters | Deformation Models | Mixture Models

Learning + Parts?

Felzenszwalb et al. (2000)

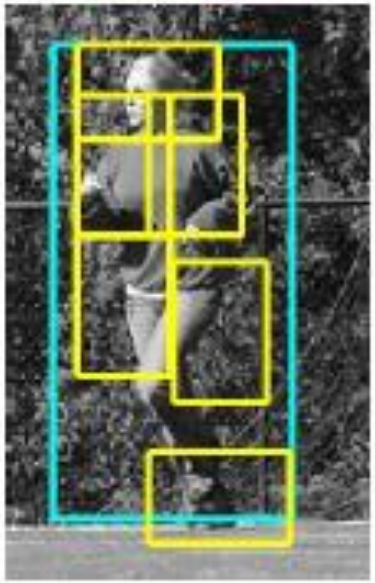
Mikolajczyk et al (2004)

Ioffe & Forsyth (1999)

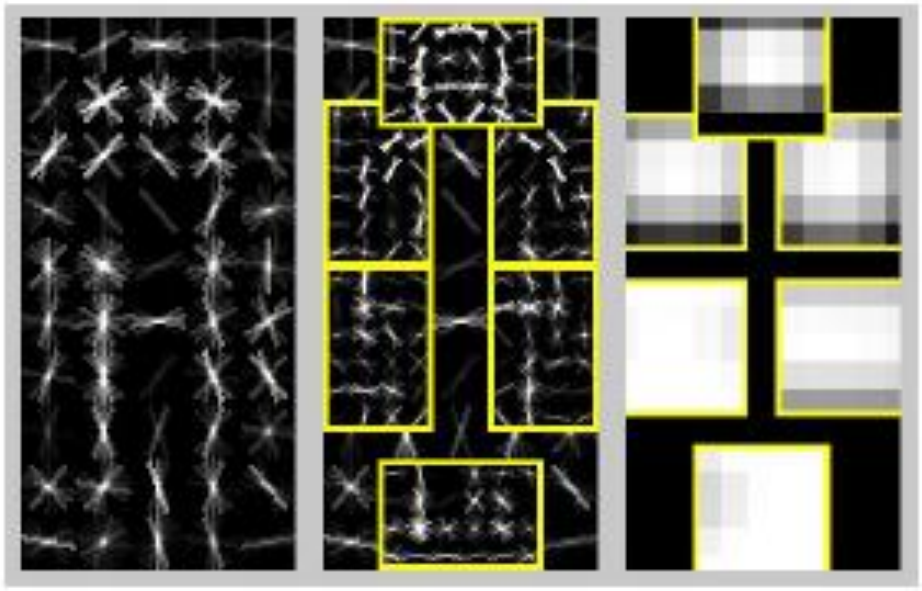
....

State-of-the-art Detector

Deformable Part-based Model (DPM)



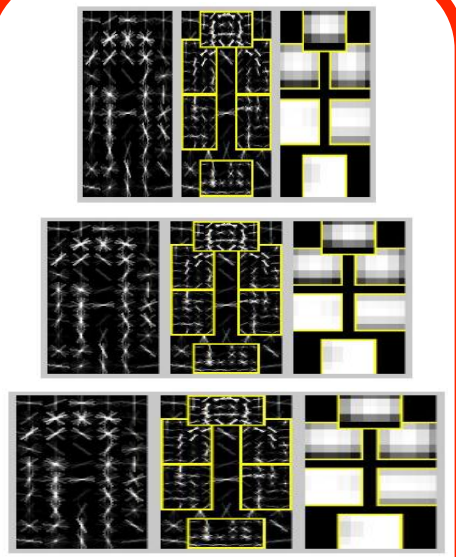
Image



Root Filter

Part Filters

Deformation Models



Mixture Models

Reduced Intra-class variation?

Schneiderman et al. (2002, 03, 04 etc.)

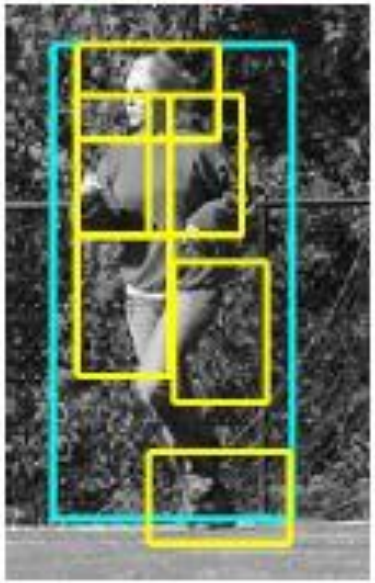
Felzenszwalb et al. (2009)

Chum & Zisserman (2007), Harzallah and Schmid (2008)

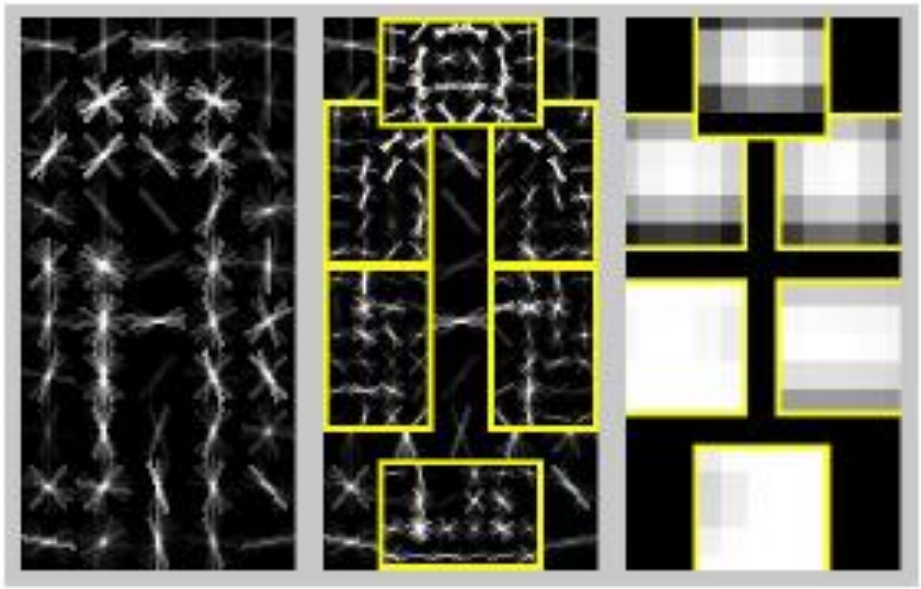
Deng et al. (2009), Divvala et al. (2012) ….

State-of-the-art Detector

Deformable Part-based Model (DPM)



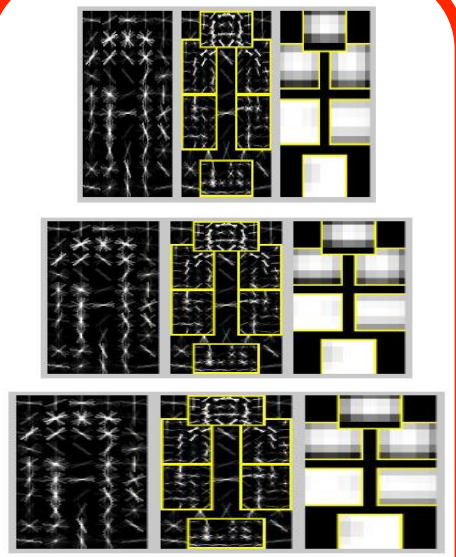
Image



Root Filter

Part Filters

Deformation Models



Mixture Models

Sub-categories

Reduced Intra-class variation?

Schneiderman et al. (2002, 03, 04 etc.)

Felzenszwalb et al. (2009)

Chum & Zisserman (2007), Harzallah and Schmid (2008)

Deng et al. (2009), Divvala et al. (2012) ….

Sub-categories in Computer Vision

Example and Images from Divvala et al. (2012)

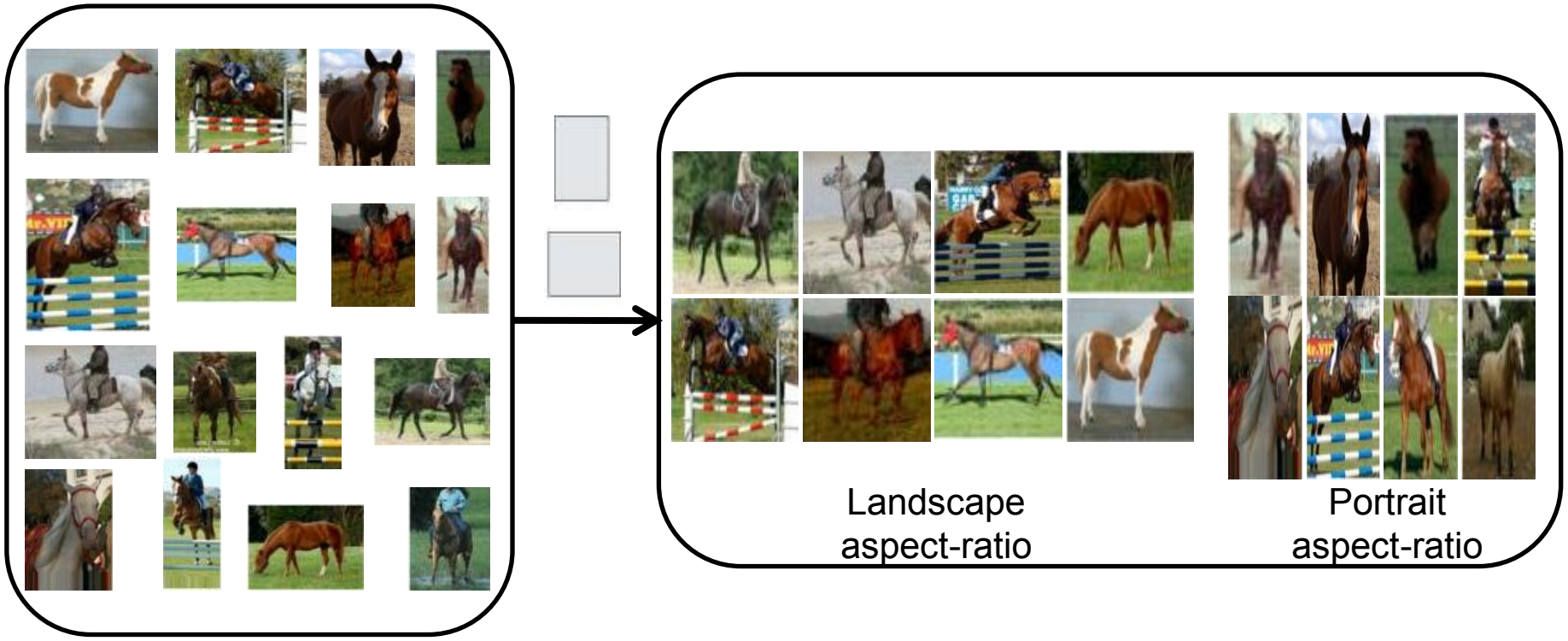
Aspect Ratio

Landscape

Portrait

Felzenszwalb et al. (2009)

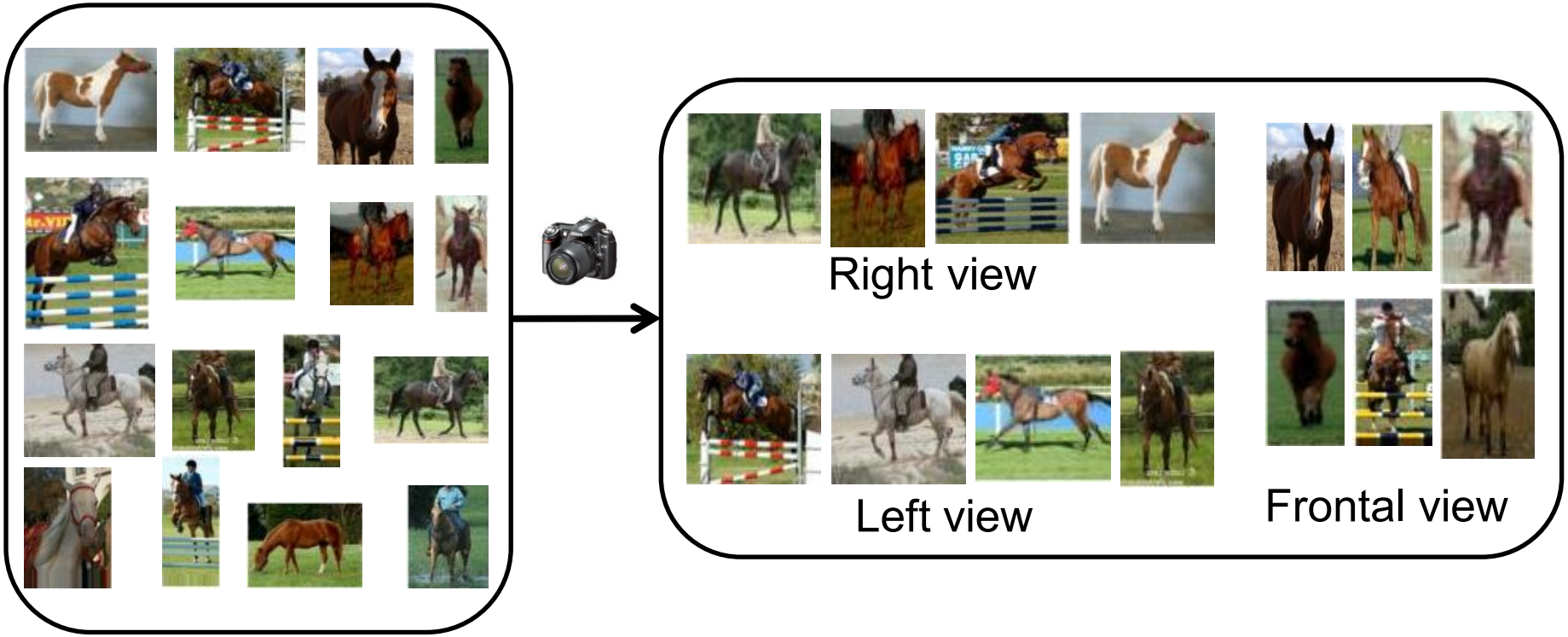
Aspect Ratio



Felzenszwalb et al. (2009)
etc.

Slide from Divvala et al. (2012)

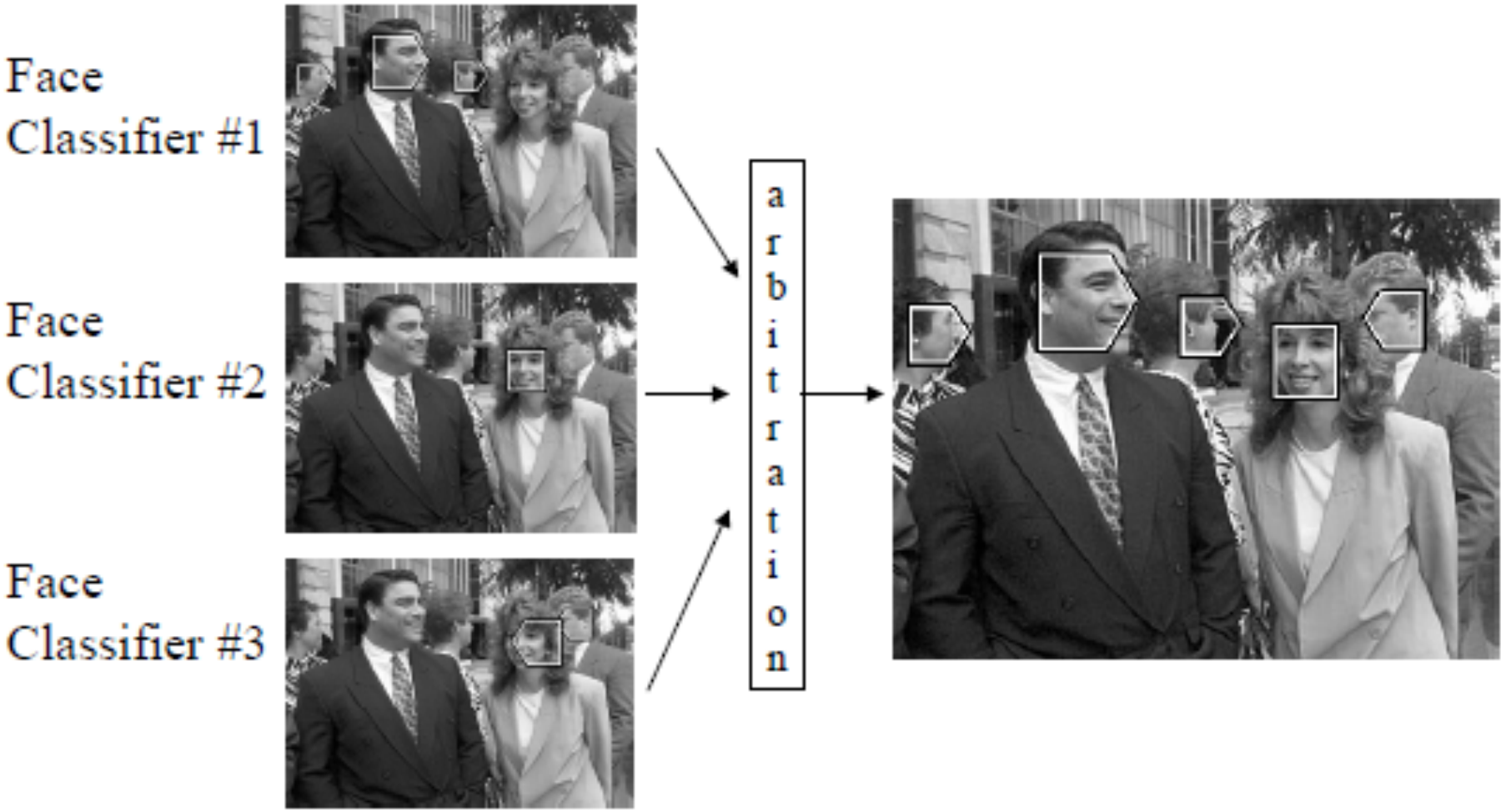
View-point



Schneiderman et al. (2002, 03, 04 etc.)
Chum & Zisserman 2007
Harzallah and Schmid 2008

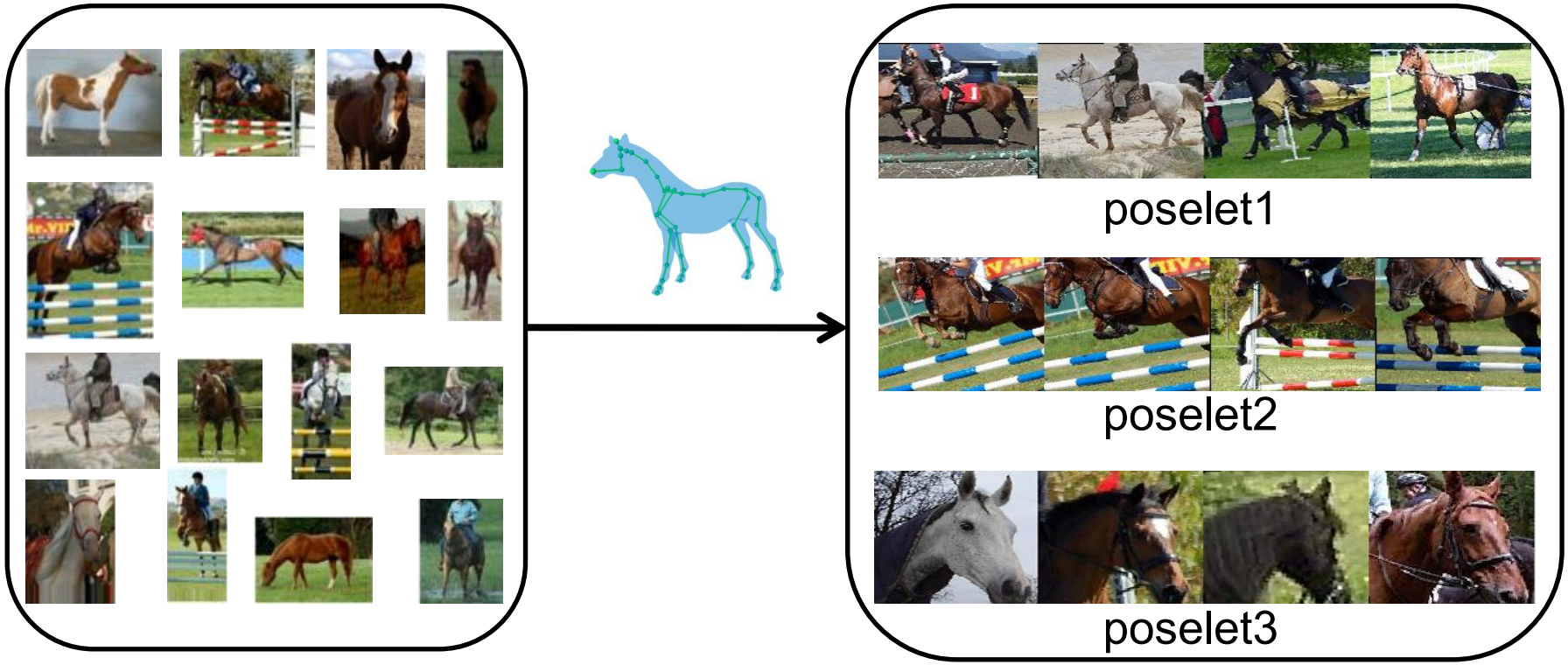
Image from Divvala et al. (2012)

View-point



Schneiderman et al. (2002, 03, 04 etc.)

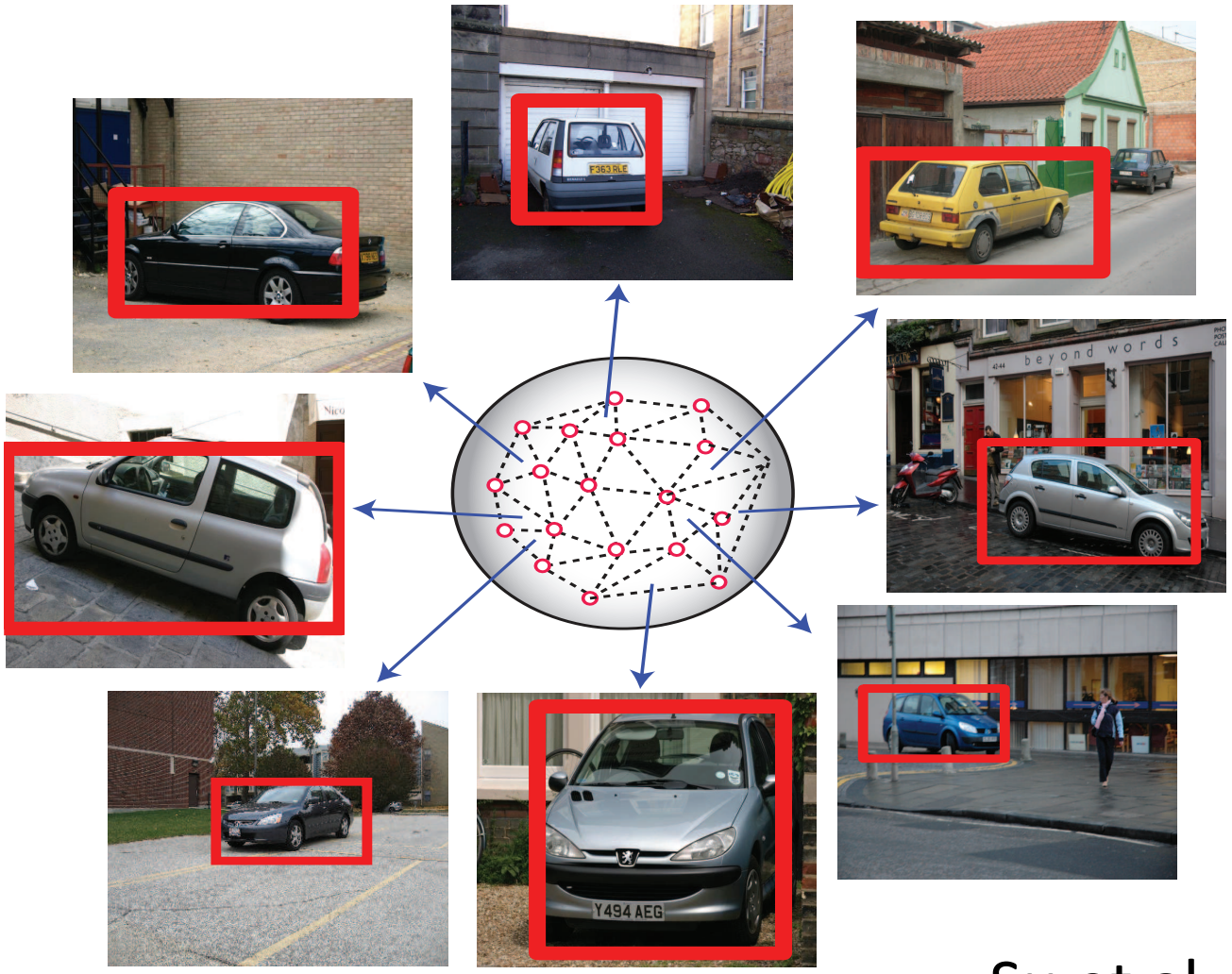
3D Configuration



Bourdev & Malik, 2009

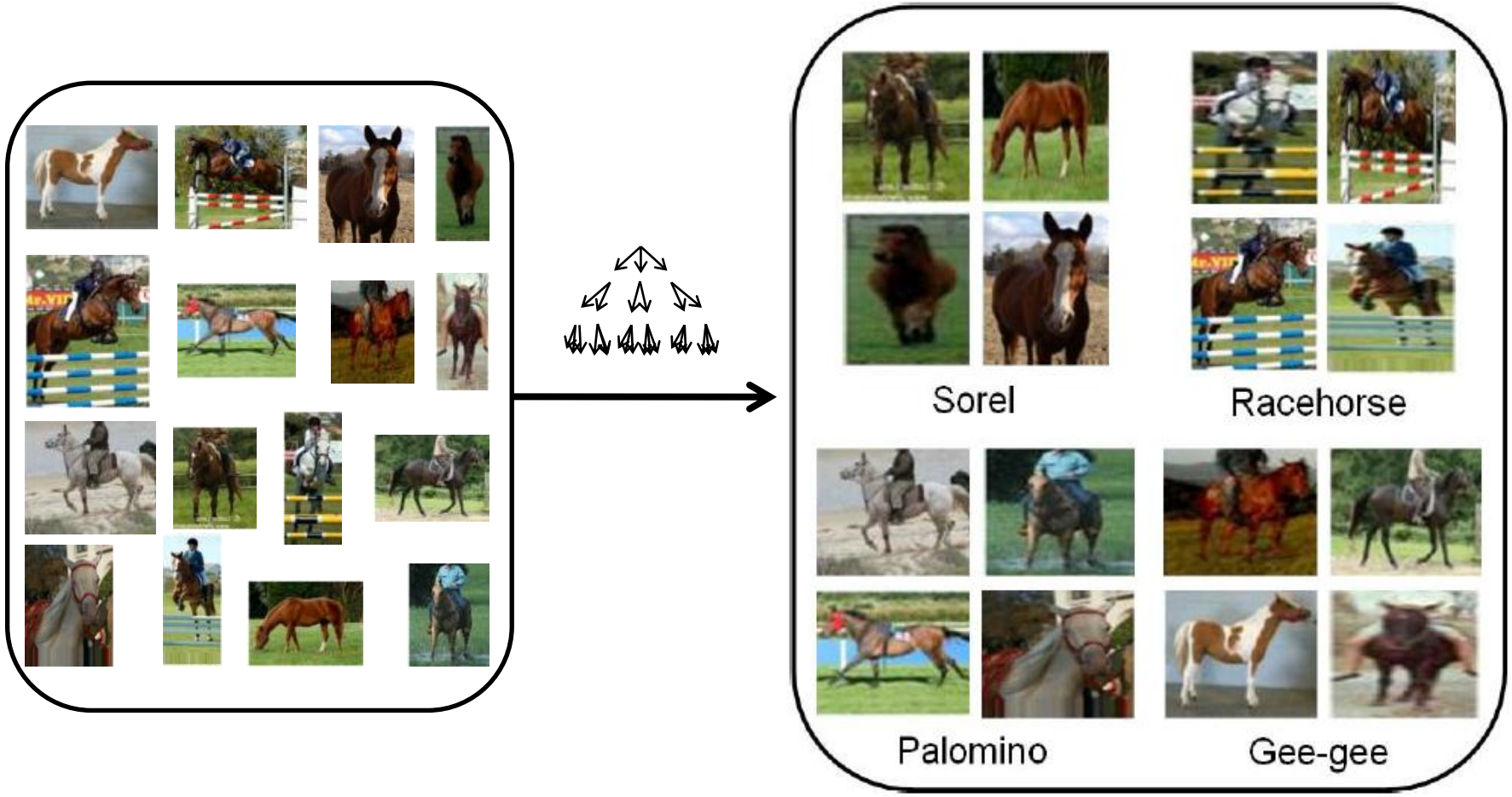
Slide from Divvala et al. (2012)

View-point + 3D Configuration



Su et al., 2009

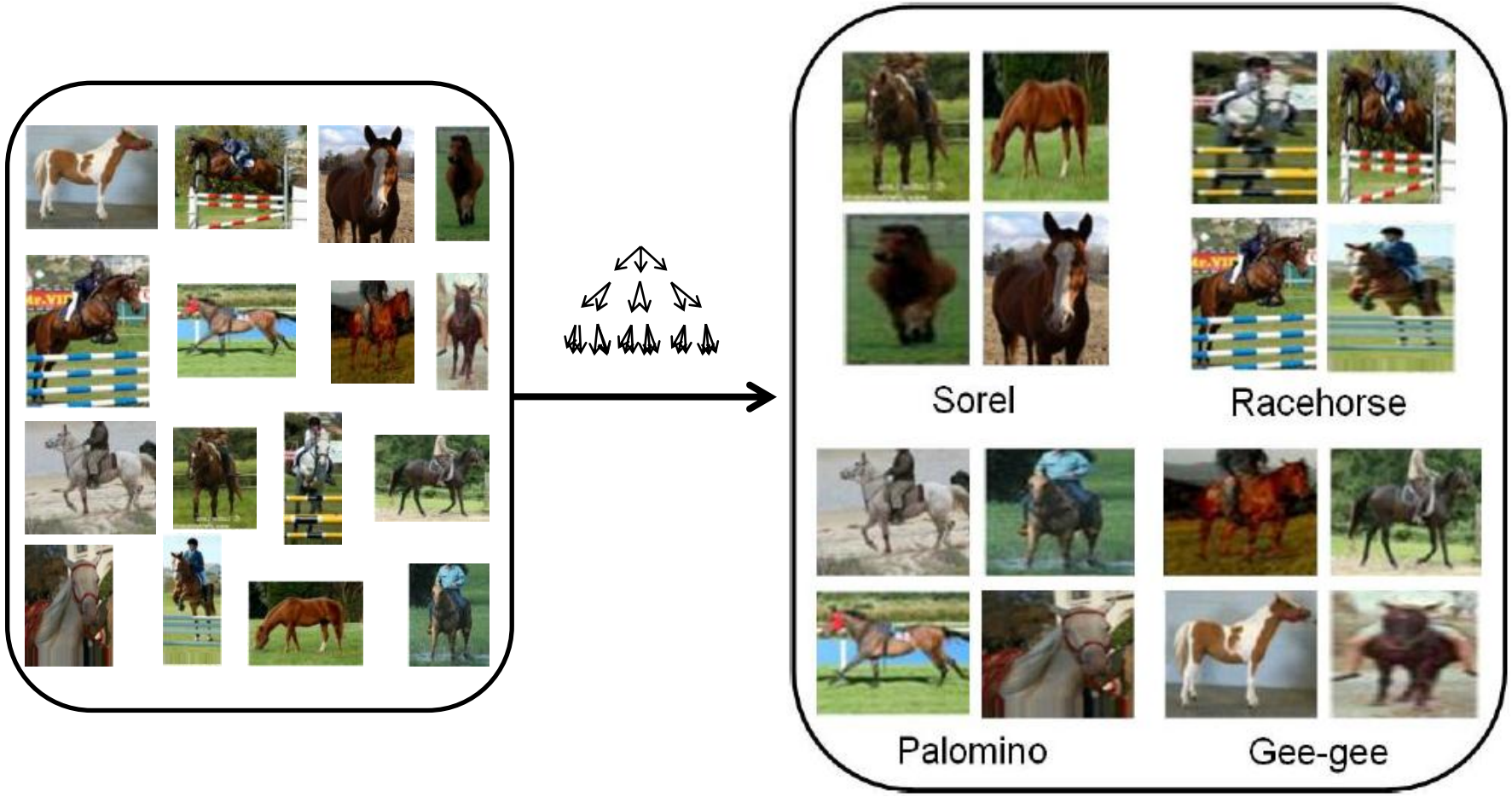
Semantic or Taxonomy



"ImageNet", Deng et al., 2009

Slide from Divvala et al. (2012)

Semantic or Taxonomy



Evidence in Neuro-science

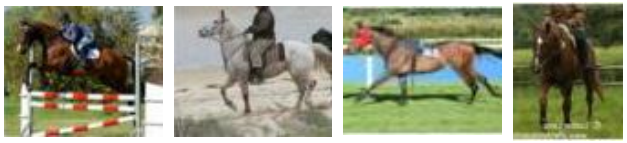
Kriegeskorte et al., 2008

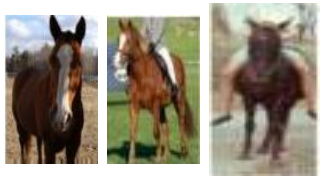
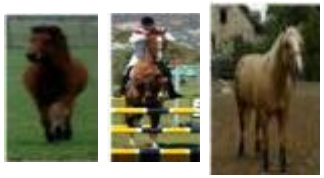
Kiani et al., 2007

"ImageNet", Deng et al., 2009

Slide from Divvala et al. (2012)


Right


Left



Front


poselet1


poselet2


poselet3



Racehorse





Gee-gee





Palomino





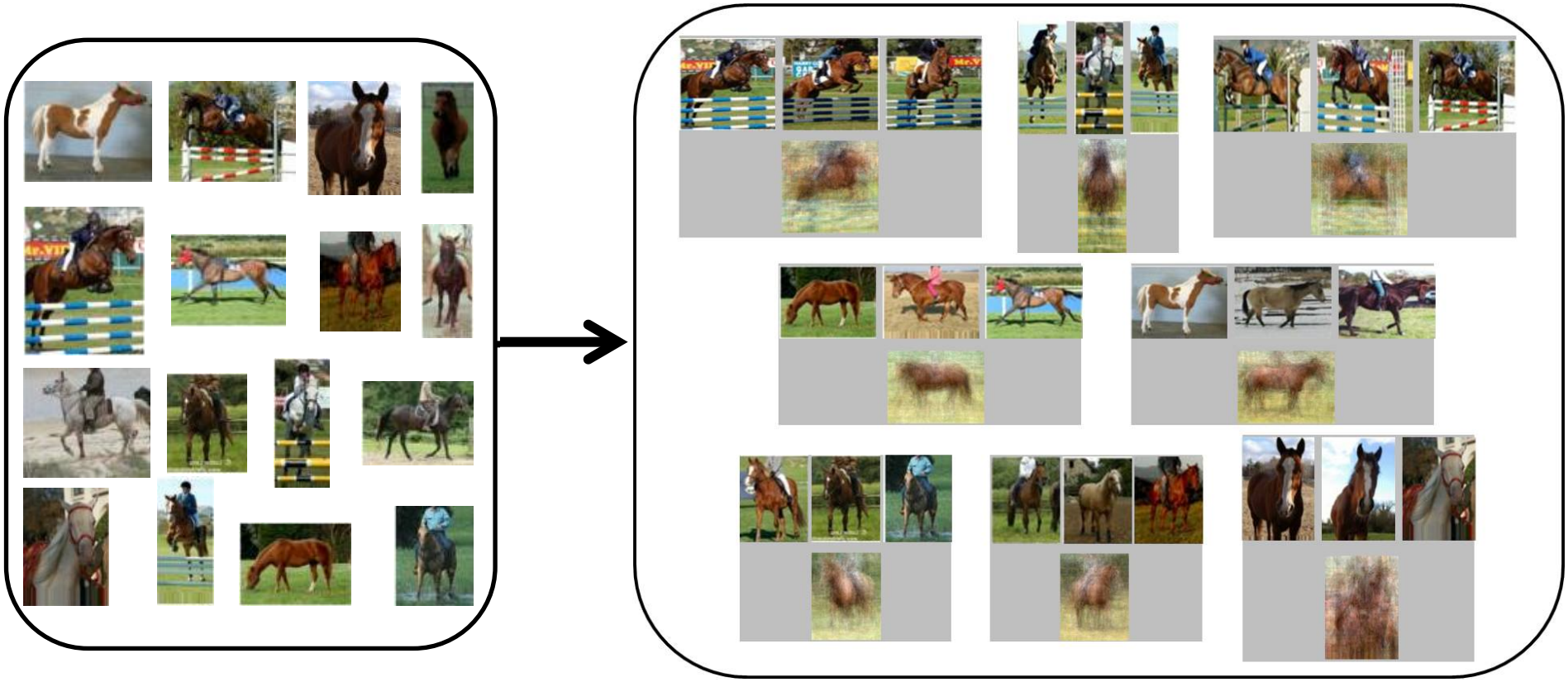
Sorel


Landscape


Portrait

Slide from Divvala et al. (2012)

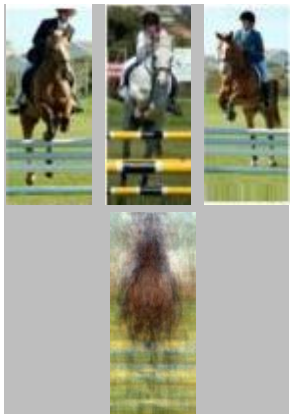
Visual Sub-categories

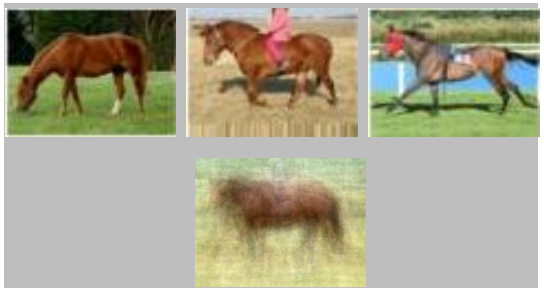


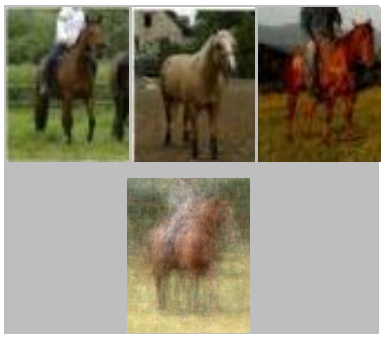
Evidence in Neuro-science
DiCarlo et al., 2013
…

Divvala et al. (2012)
Chen et al. (2014)

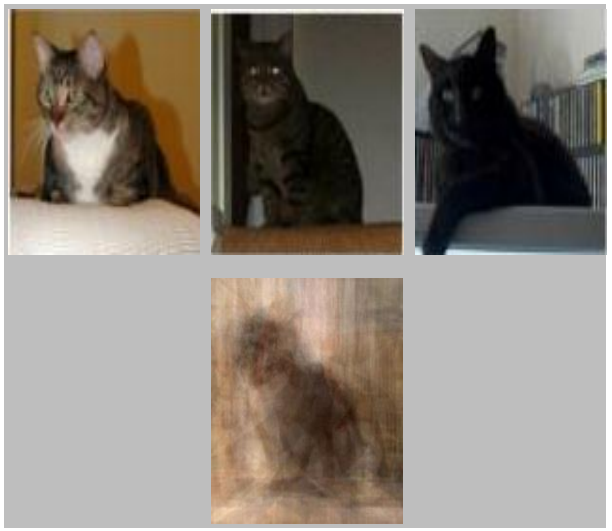
Visual Sub-category

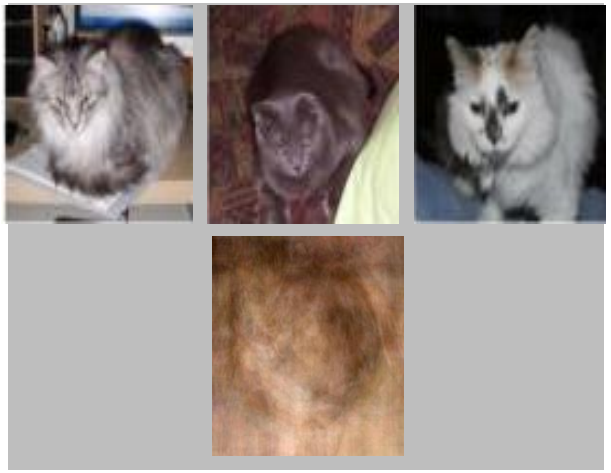


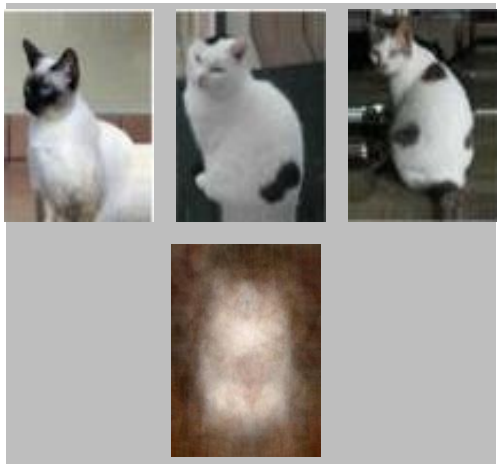


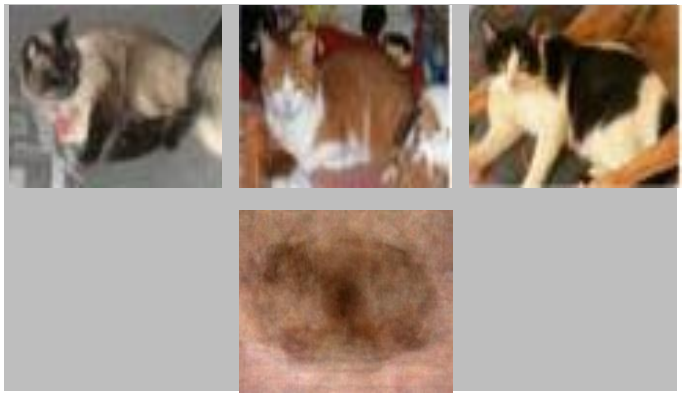


Visual Sub-categories









Visual Sub-categories

in very broad stokes

Visual Sub-categories

- Initialize sub-categories

All ground-truth ‘horse’ instances

K-means
feature-space
clustering

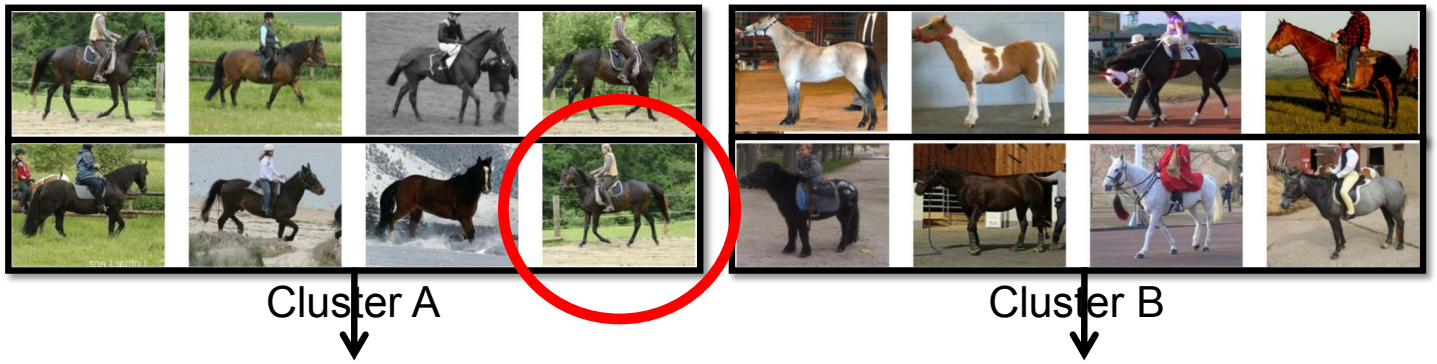
Visual Sub-categories

- Initialize sub-categories
  - Clustering (kmeans) on feature space (AR, HOG etc.)

Visual Sub-categories

- Initialize sub-categories
  - Clustering (kmeans) on feature space (AR, HOG etc.)
  - Train Models

- Latent Update or Re-clustering
  - Find cluster assignments **again** using learned models



Train Models w_a w_b

Apply on training images

Reassign members

argmax_{ab} (score(w_a), score(w_b))

Visual Sub-categories

1. Initialize sub-categories
   - Clustering (kmeans) on feature space (AR, HOG etc.)
   - Train models

2. Latent Update or Re-clustering
   - Find cluster assignments **again** using learned models

3. Re-train models.
   - Back to 2

Some Caveats

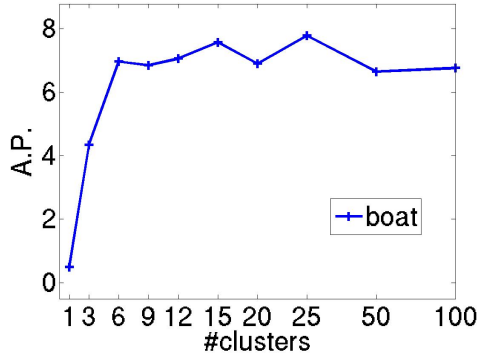
- Number of Sub-categories (K)
  - Large enough to capture all variation

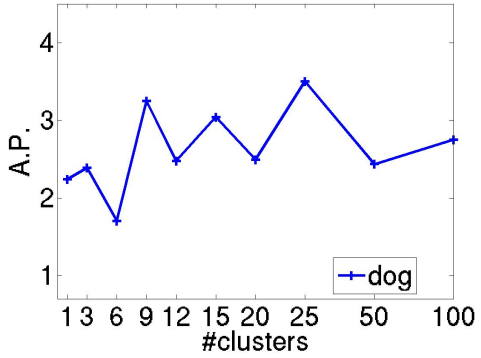
Some Caveats

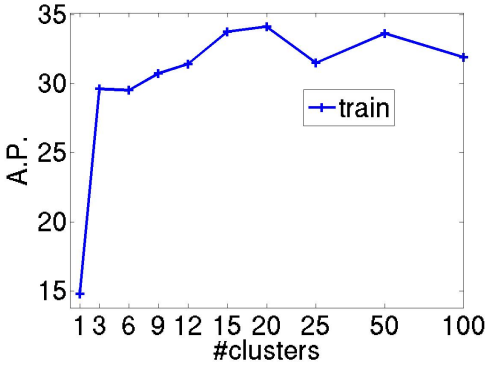
- Number of Sub-categories (K)
  - Large enough to capture all variation
  - But not too large
    - If too large, sub-categories fight for instances
    - Less training data for each sub-category

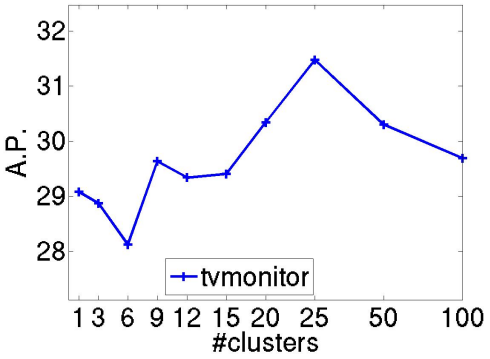
Some Caveats

- Number of Sub-categories (K)
  - Large enough to capture all variation
  - But not too large
  - Different sweet-spot for every category









Some Caveats

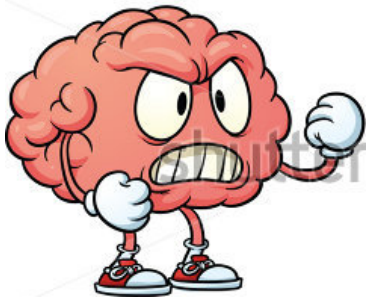
- Number of Sub-categories (K)
  - Large enough to capture all variation
  - But not too large
  - Different sweet-spot for every category
- Tricky calibration
  - Combining multiple clusters
  - Removing noisy clusters

Evolution of HoG based Object Detectors

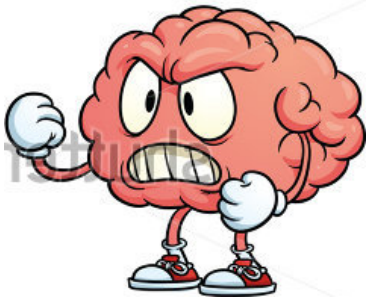
[?]	# of Mixtures	Type of Clustering	# of Parts	mAP on PASCAL
HOG' 05	1	--NA--	--NA--	0.17
DPM'08	1	--NA--	6	0.21
DPM'10	2	Aspect Ratio	6	0.26
DPM'11	6	Aspect Ratio	8	**0.32**
SUB'12	15	Appearance++	8	**0.35**
SUB'12	15	Appearance++	0	0.24
SUB'12	15	Appearance++	"1"	0.31
ESVM'11	N	--NA--	--NA--	0.23

What does Neuroscience literature say about Object Representation in IT

Object Representation in IT



Semantic vs. Visual



Kiani et al., 2007

Kriegeskorte et al., 2008

Bell et al., 2009

…

..

.

Logothetis et al., 1996

DiCarlo et al., 2012

Brincat et al., 2006

Kourtzi et al., 2011

Brincat et al., 2004

Yamane et al., 2008

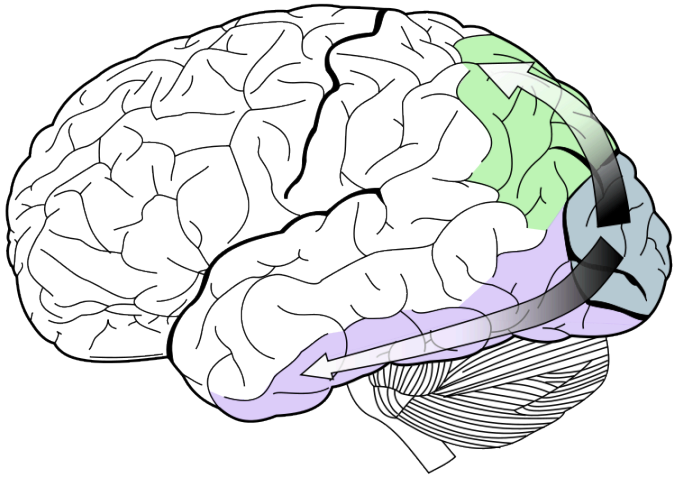
Kayaert et al., 2003/05

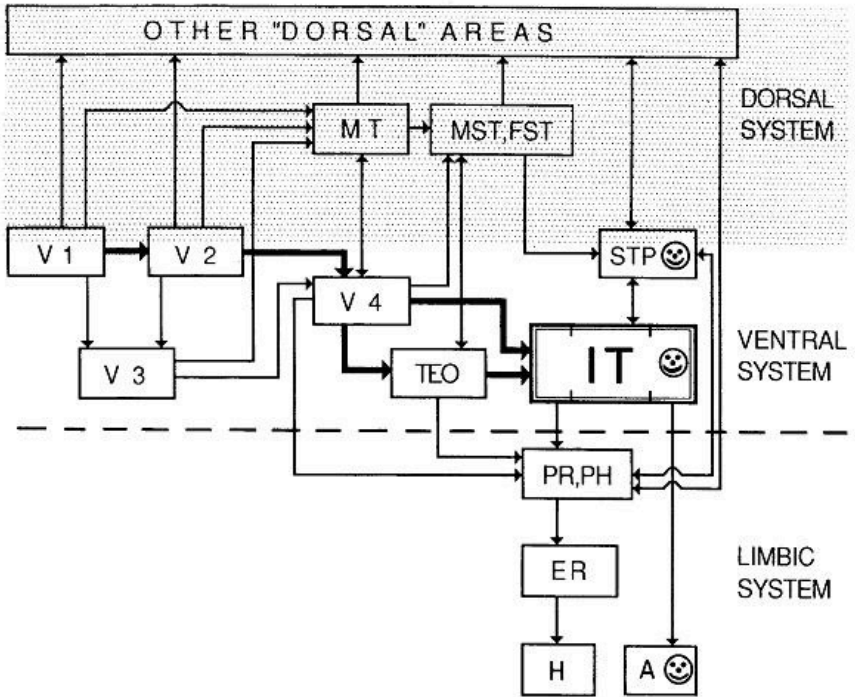
Op de Beeck et al., 2001

…

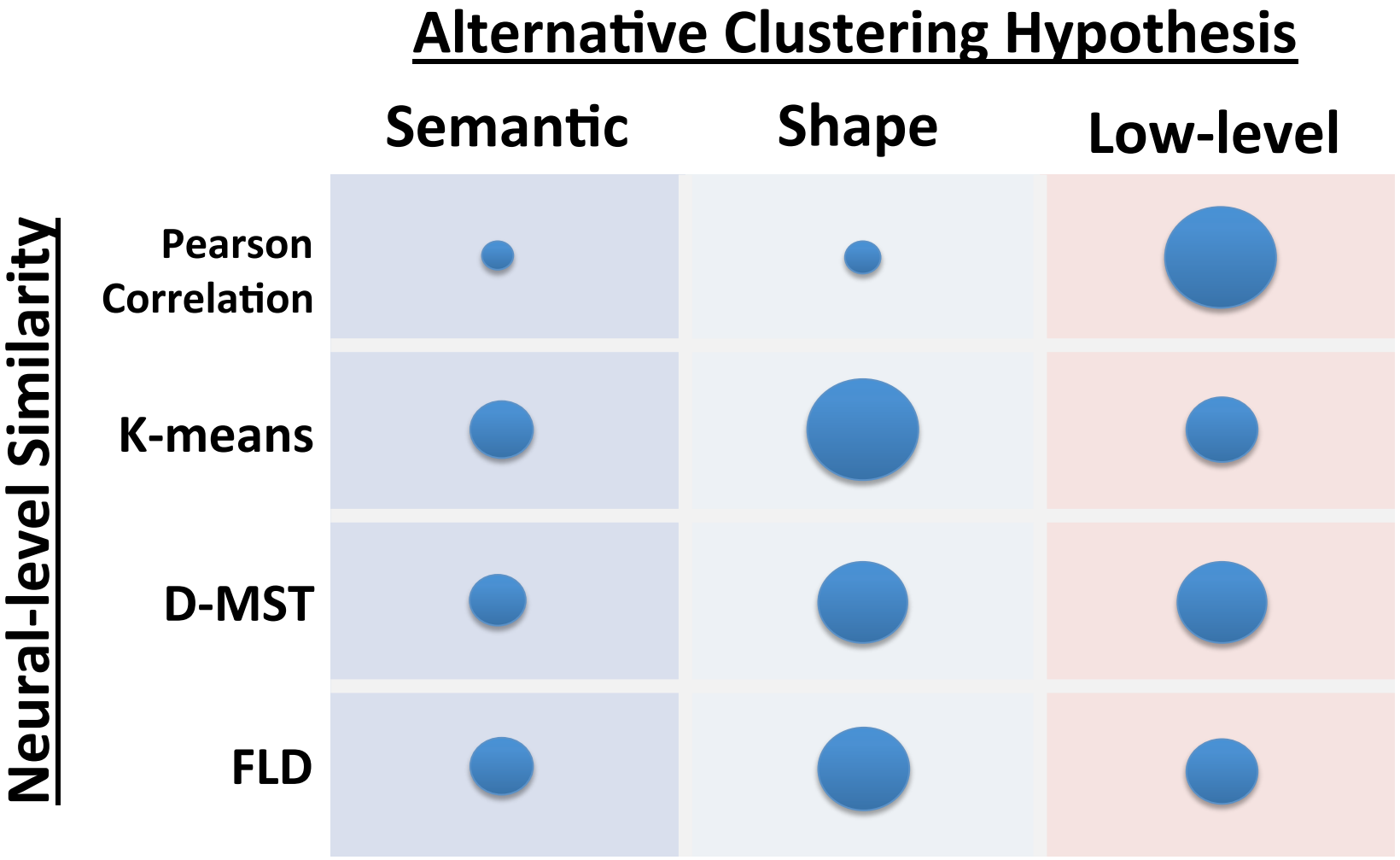
..

Infero-temporal Cortext (IT)





Structure of Visual Object Representation in IT



Neural-activity Recording Setup

- 2 Monkeys
- 94 well-isolated single units from anterior IT
- ~5x4 mm area (SCS and AMTS)
- No attempt done to target preferential cells
- Smaller AP and ML extent regions sampled as opposed to [14]

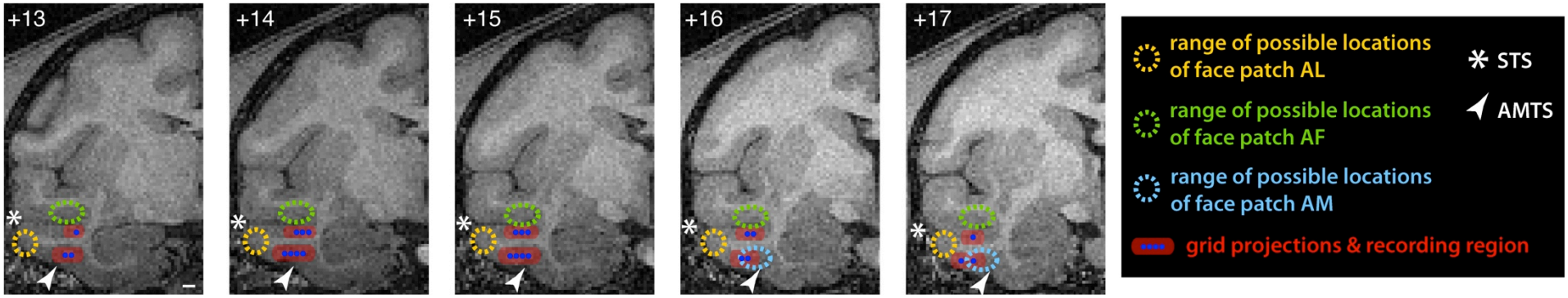


Figure 1. **Recording locations**.

- The blue dots show the projections of the recording chamber grid-point locations from the top of the skull to the ventral bank of the superior temporal sulcus (STS) and the ventral surface lateral to the anterior middle temporal sulcus (AMTS).
- The projections are shown over a sequence of MRI images (spanning a 13–17 anteroposterior range; Horsley-Clarke coordinates) that were collected, for one of the monkeys, before the chamber implant surgery. Only the grid locations in which the electrode was inserted at least once are shown.
- The red-shaded areas highlight the estimated cortical span that was likely sampled during recording, given that:
  - each electrode penetration usually spanned the whole depth of the targeted cortical bank (either STS or AMTS); and
  - the upper bound of the variability of each recording location along the mediolateral axis (due to bending of the electrode during insertion) can be estimated as +/- 2 mm [80].
- The figure also shows the range of possible locations of the three anterior face patches (AL, AF and AM) according to [33], so as to highlight their potential overlap with the recording locations.

Stimuli Setup

- 213 Gray-scale
- 5 images/sec
- Simple object detection task (?)

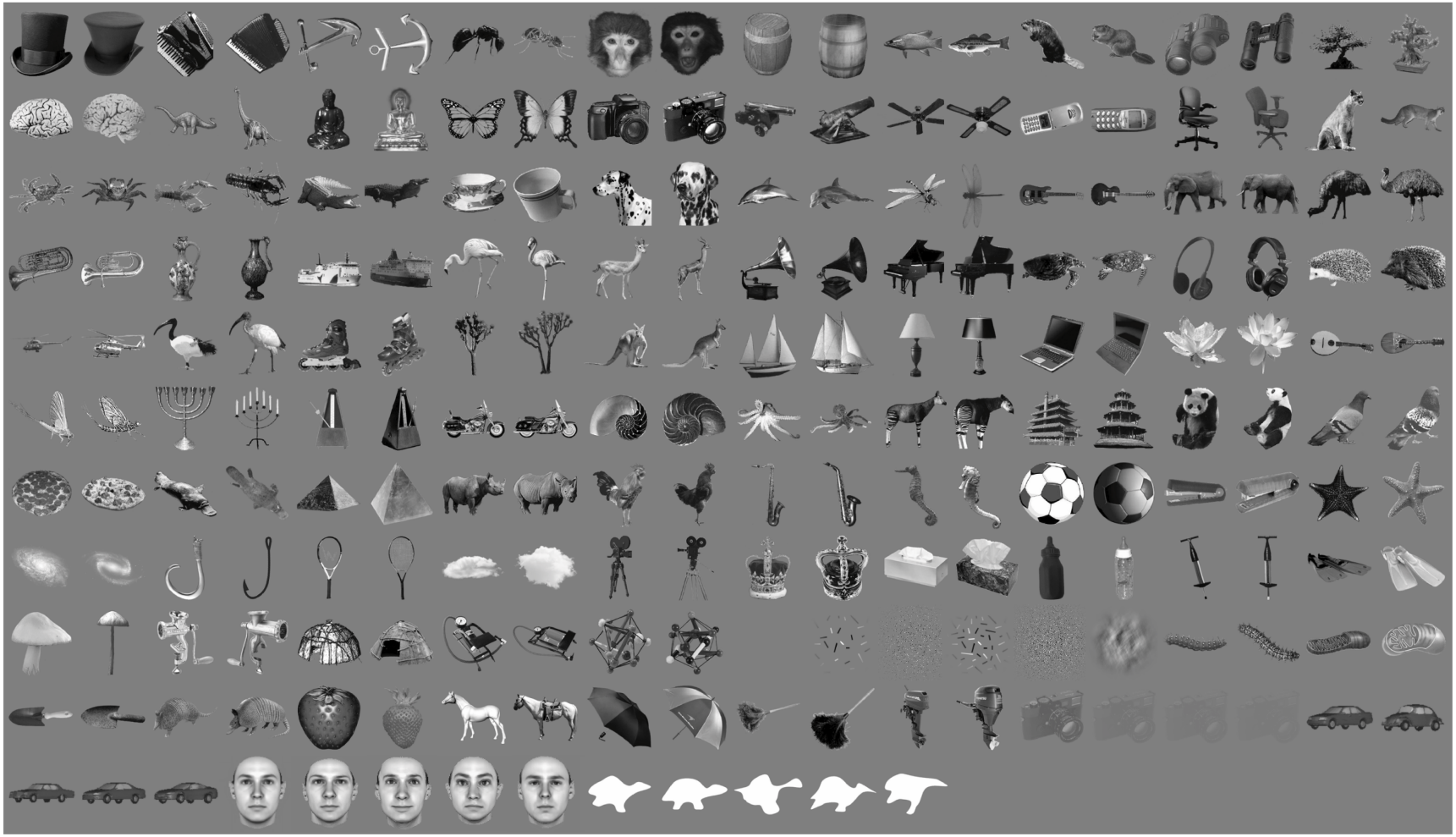


Figure 2. The stimulus set. The full set of 213 objects used in our study.

i) 188 images of real-world objects belonging to 94 different categories (e.g., two hats, two accordions, two monkey faces, etc.);

ii) 5 cars, 5 human faces, and 5 abstract silhouettes;

iii) 5 patches oftexture (e.g., random dots and oriented bars);

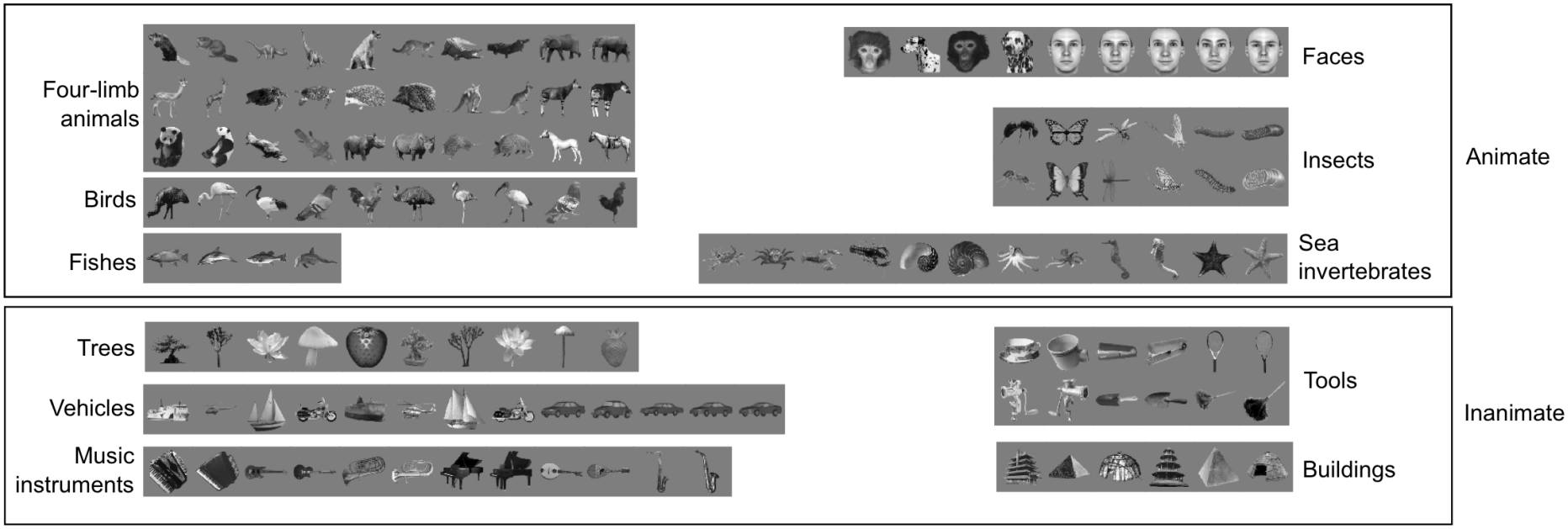
iv) a blank frame;

v) 4 low contrast (10%, 3%, 2% and 1.5%) images of one of the objects (a camera).

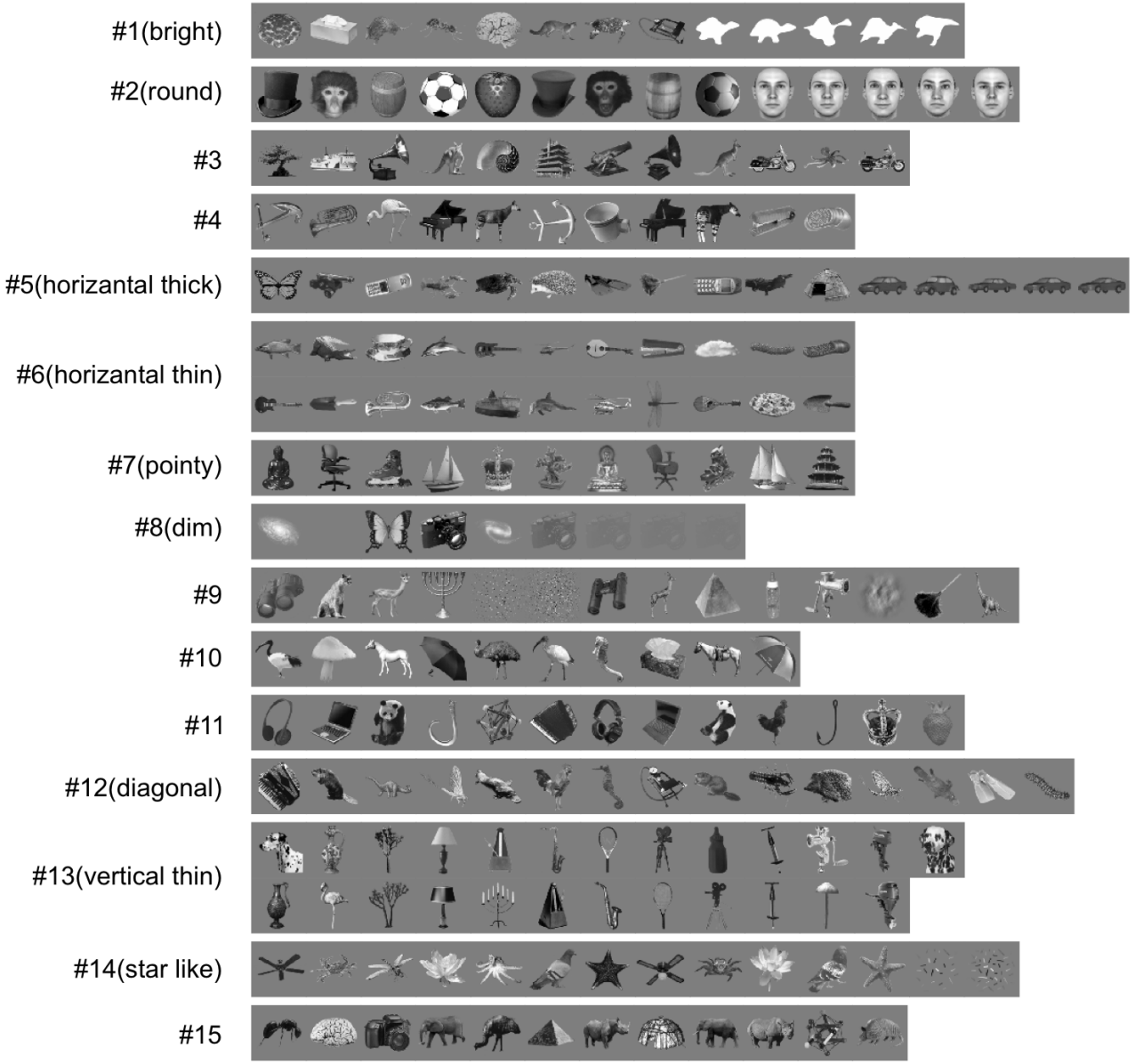
Image-level Clustering

- Semantic
- Shape-based
- Low-level

Semantic Categories



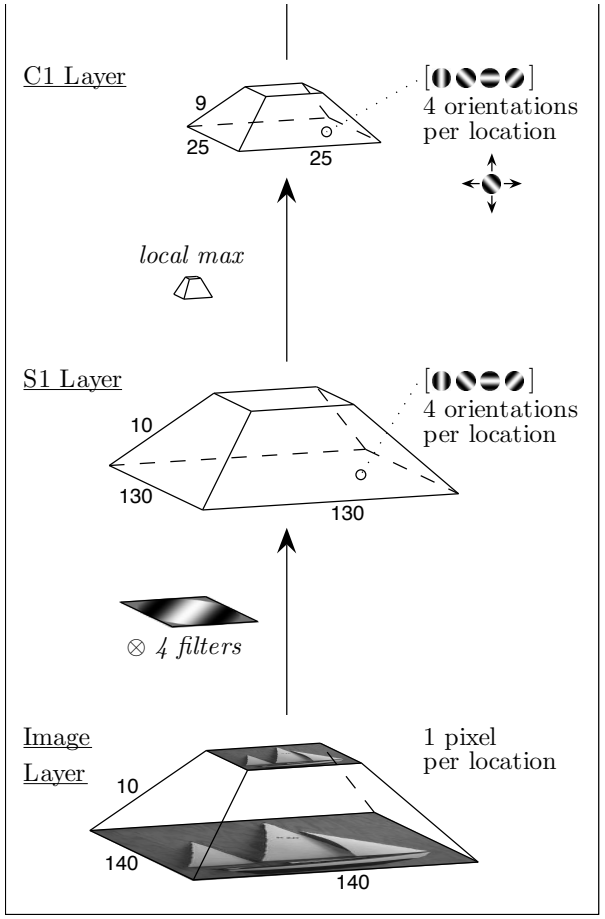
Shape-based Categories

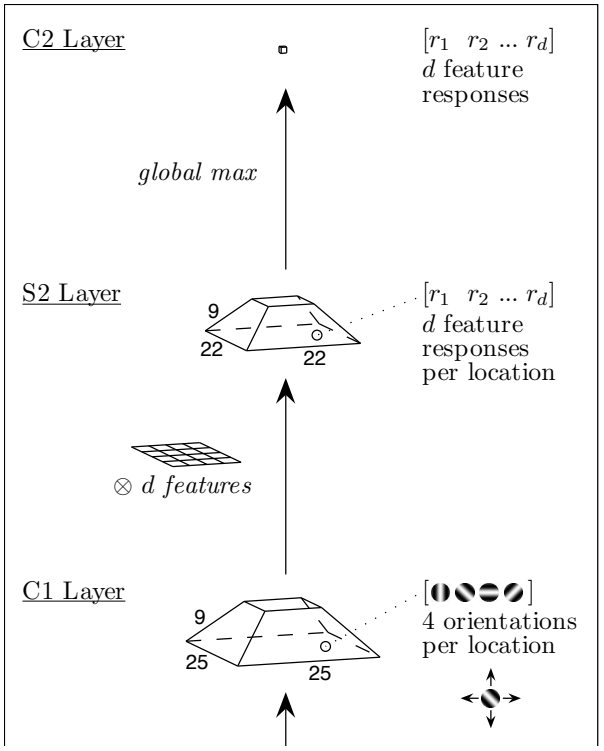


Shape-based features

Local Invariance Layer

Gabor filter Layer





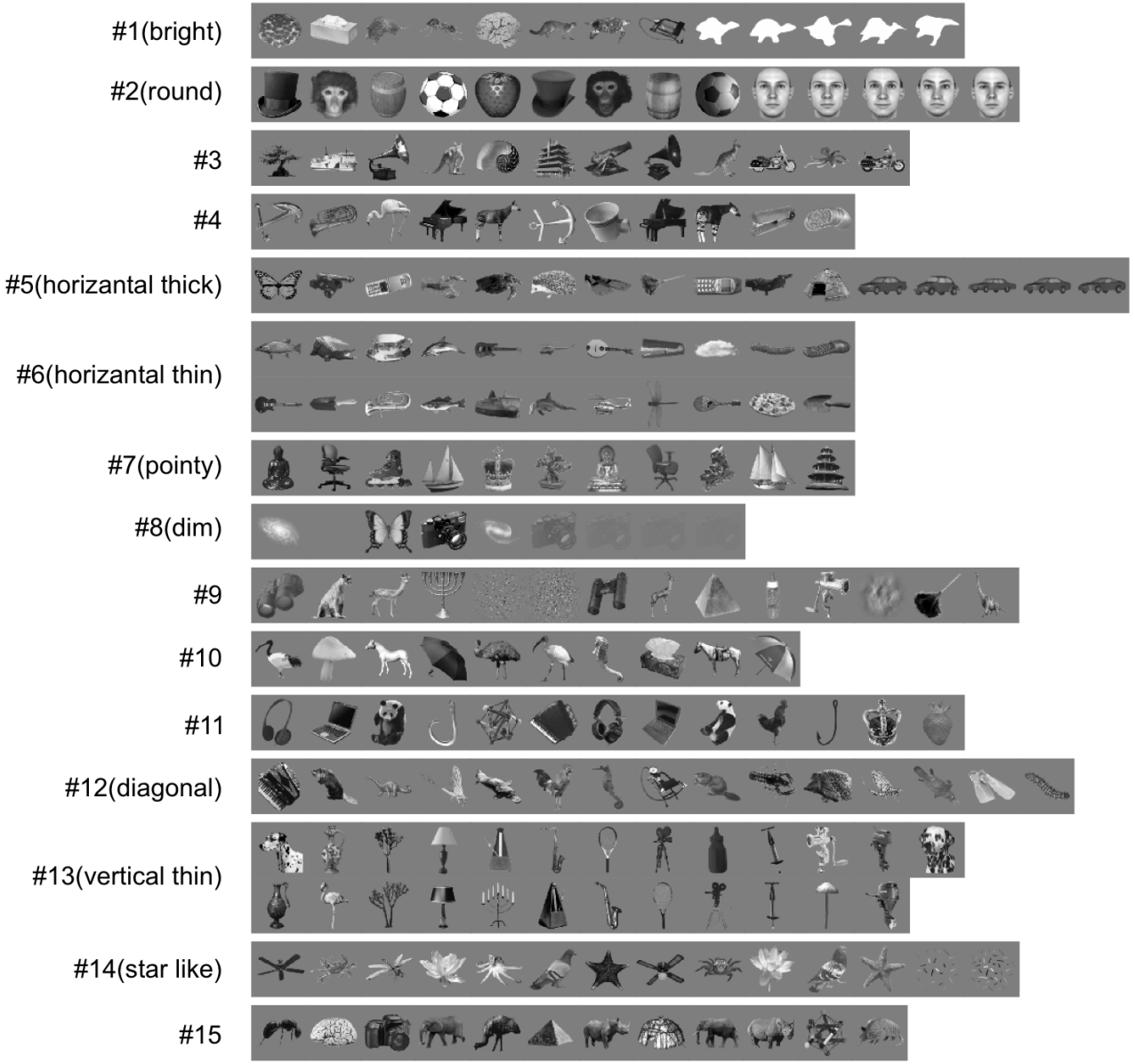
Global Invariance Layer

Intermediate Feature Layer

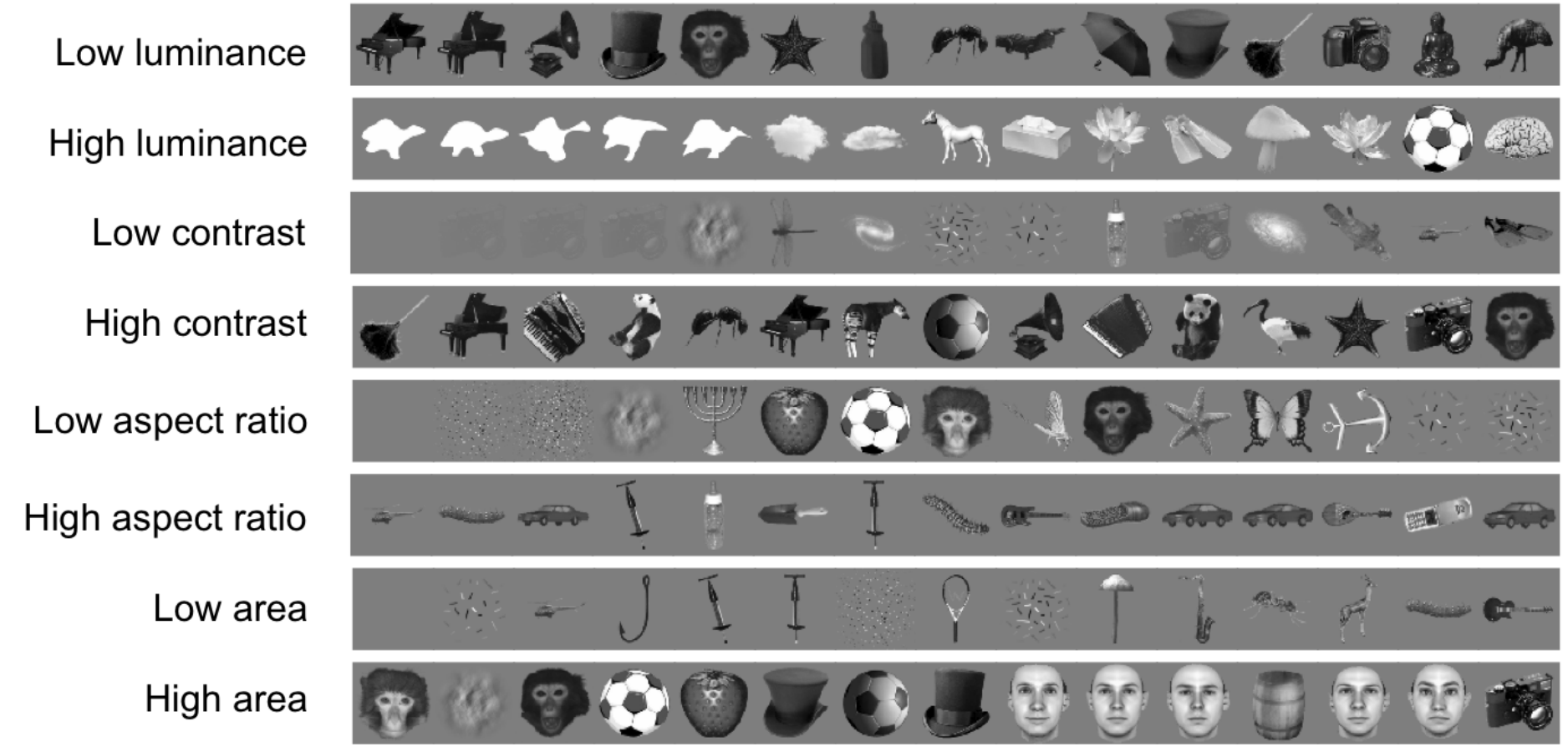
Local Invariance Layer

Mutch J, Lowe DG
Object Class Recognition and Localization Using Sparse Features with Limited Receptive Fields.
IJCV 2008

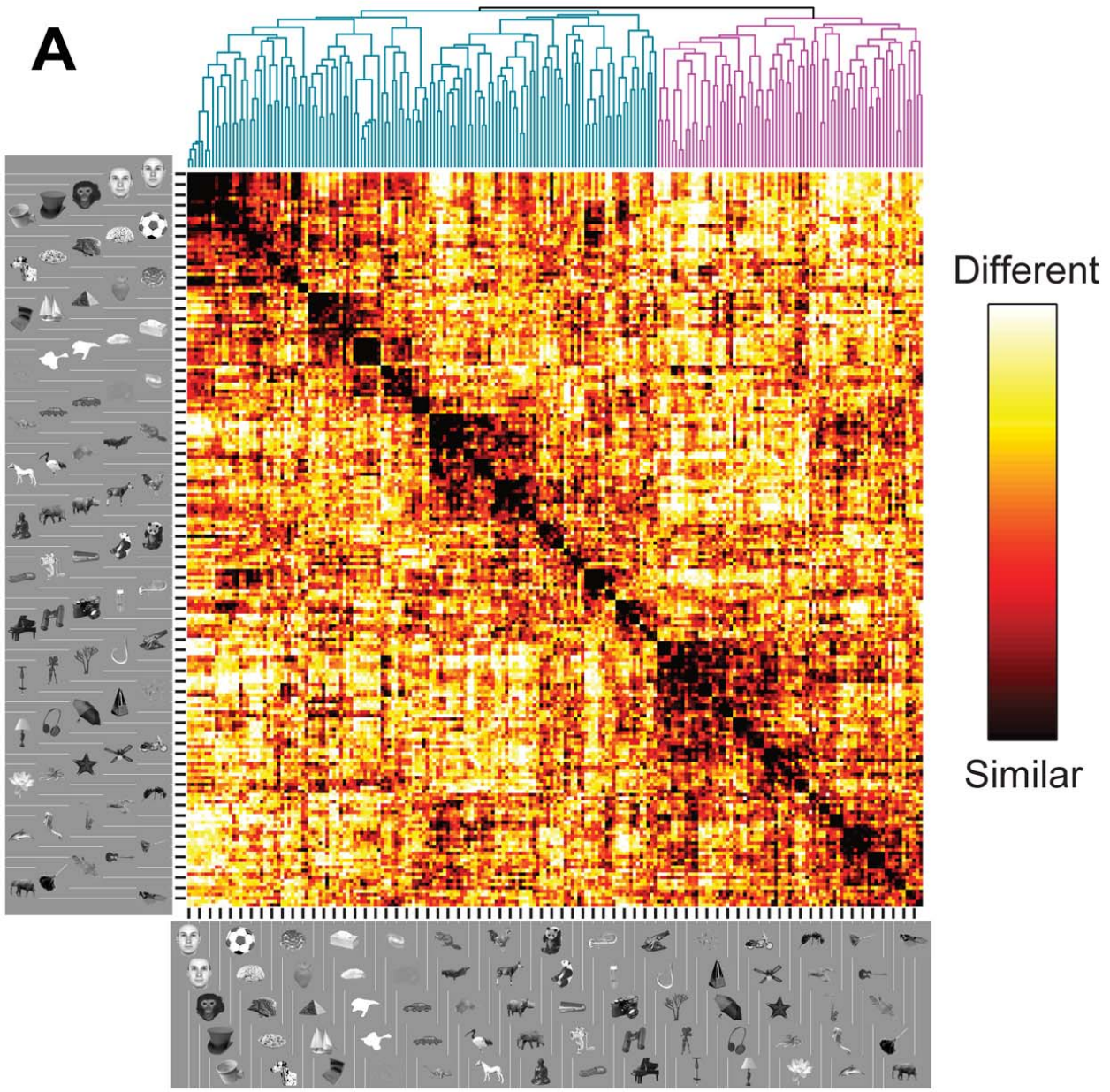
Shape-based Categories



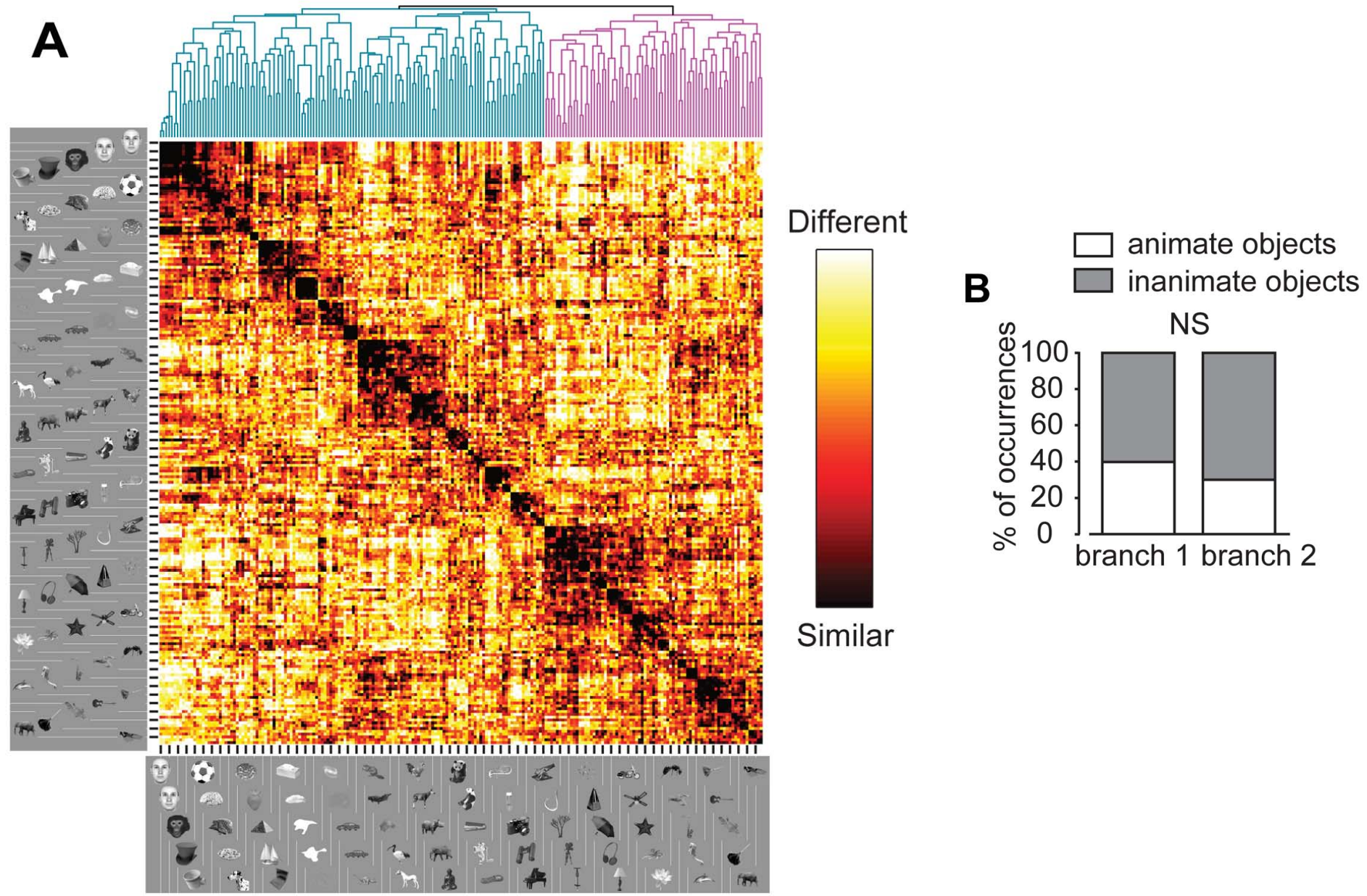
Low-level Categories



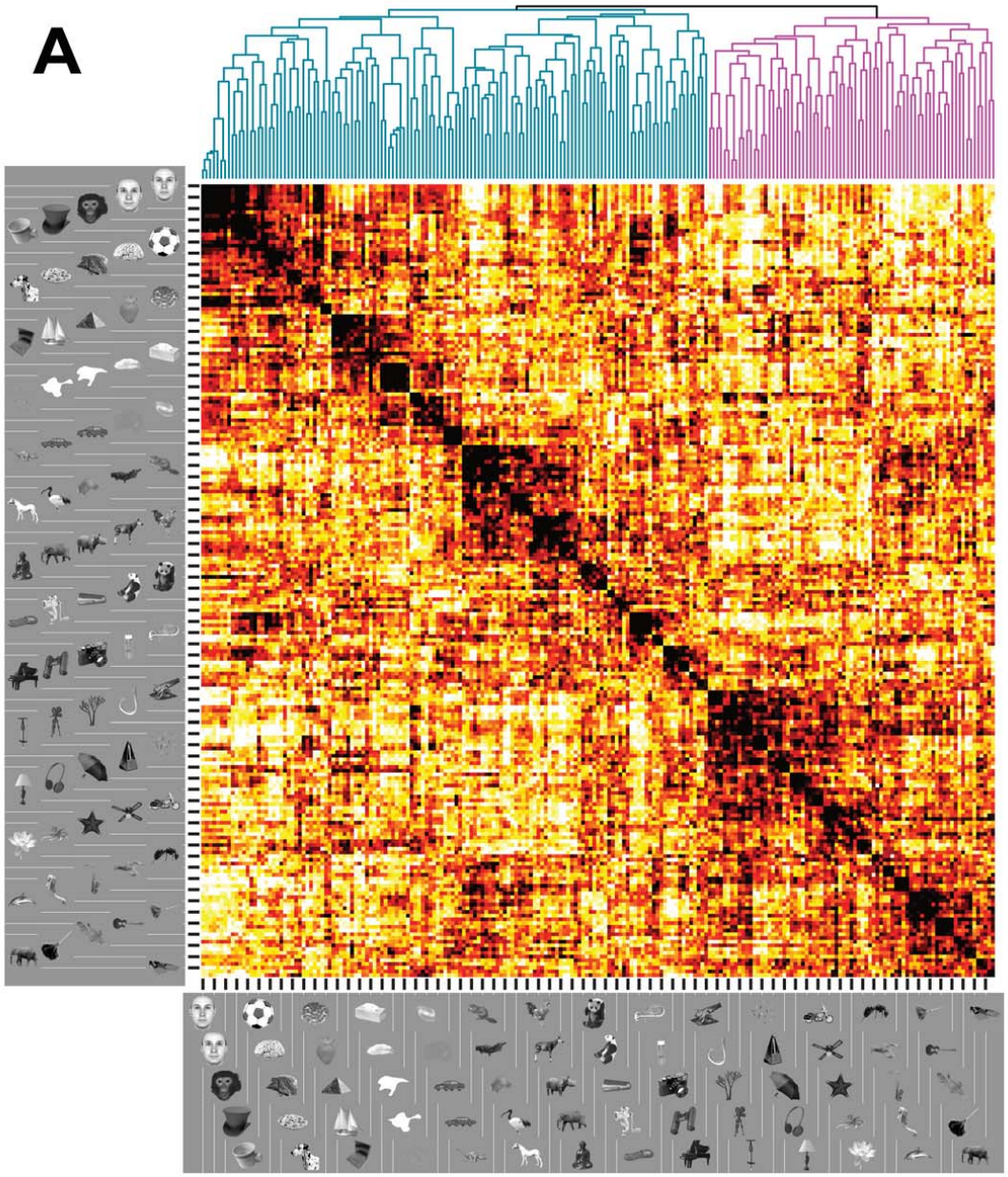
Pearson correlation coefficient

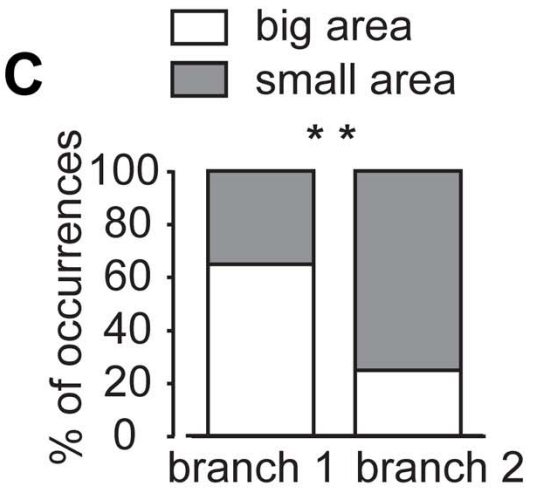


Inanimate vs. Animate



Thin (elongated) vs. Thick (roundish)





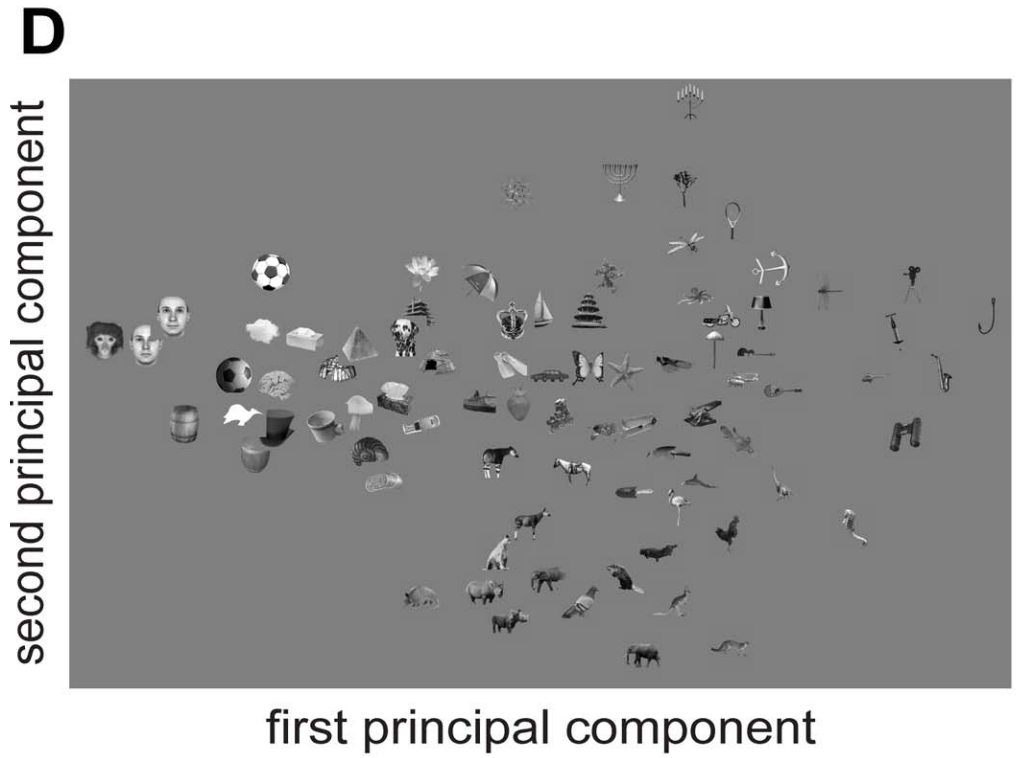
PCA on Neuronal population vectors

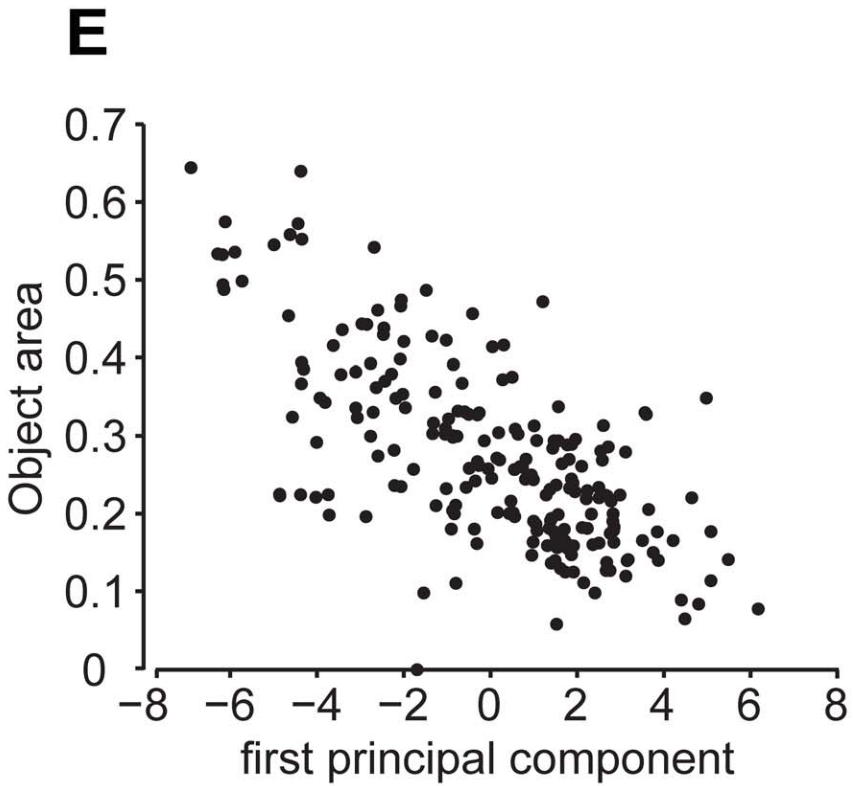
- Variance by 2 components ~15% (low).

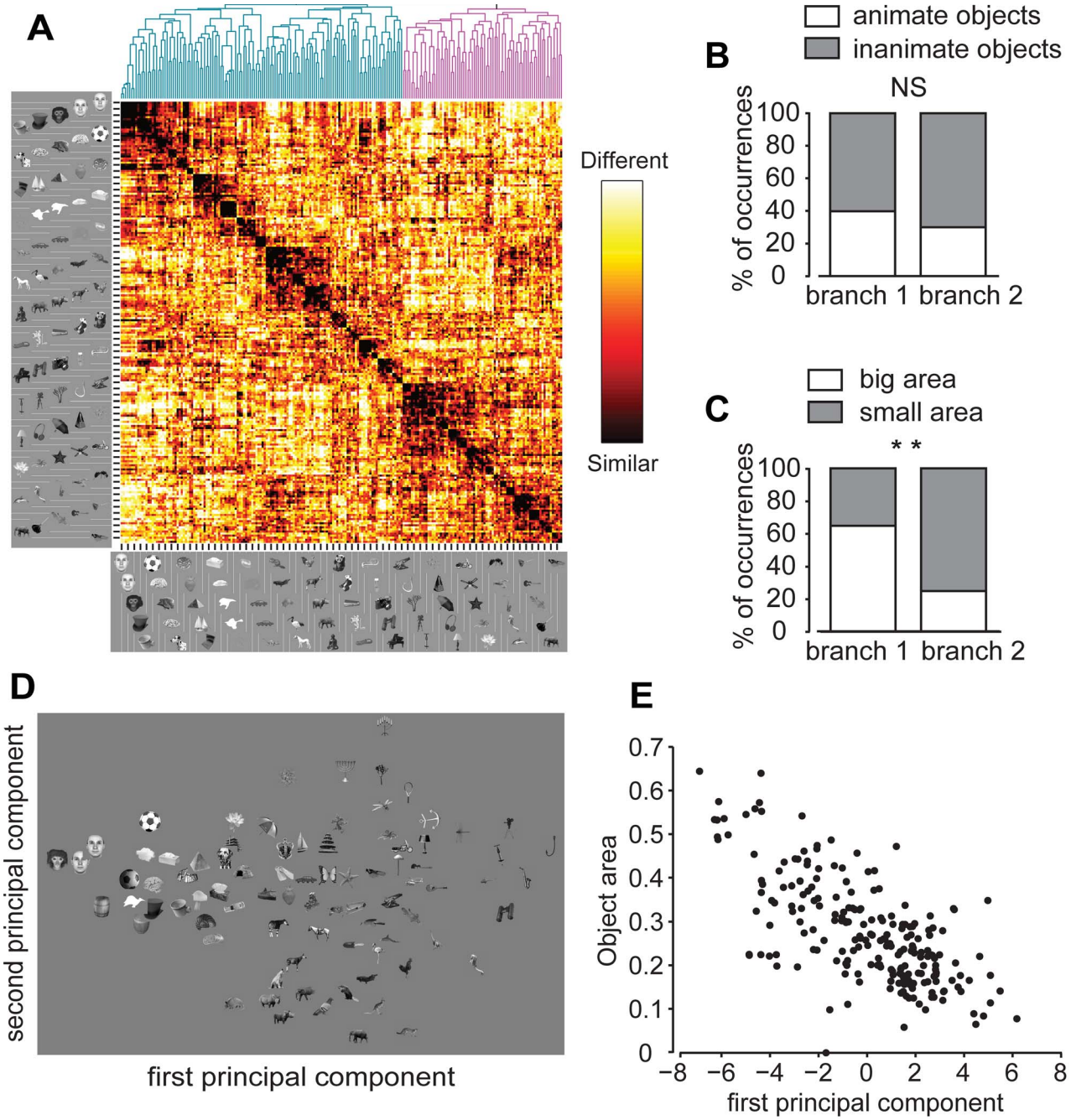
- High-level IT neurons won't capture all (highly varied) visual properties

- Goal is **not** to find dimensions that account for most variations
- Just check if any component could be associated with some global property.

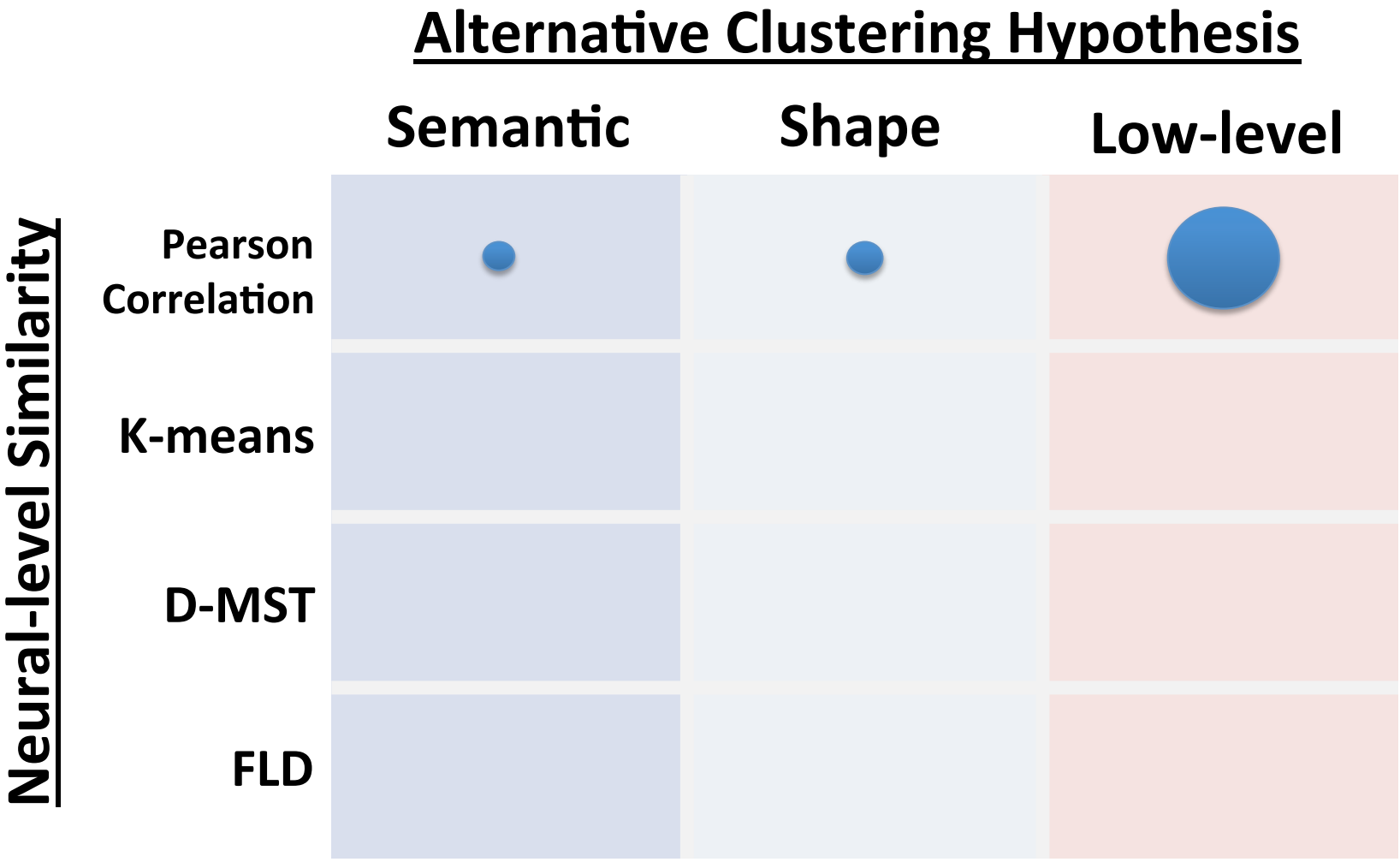
PCA on Neuronal population vectors



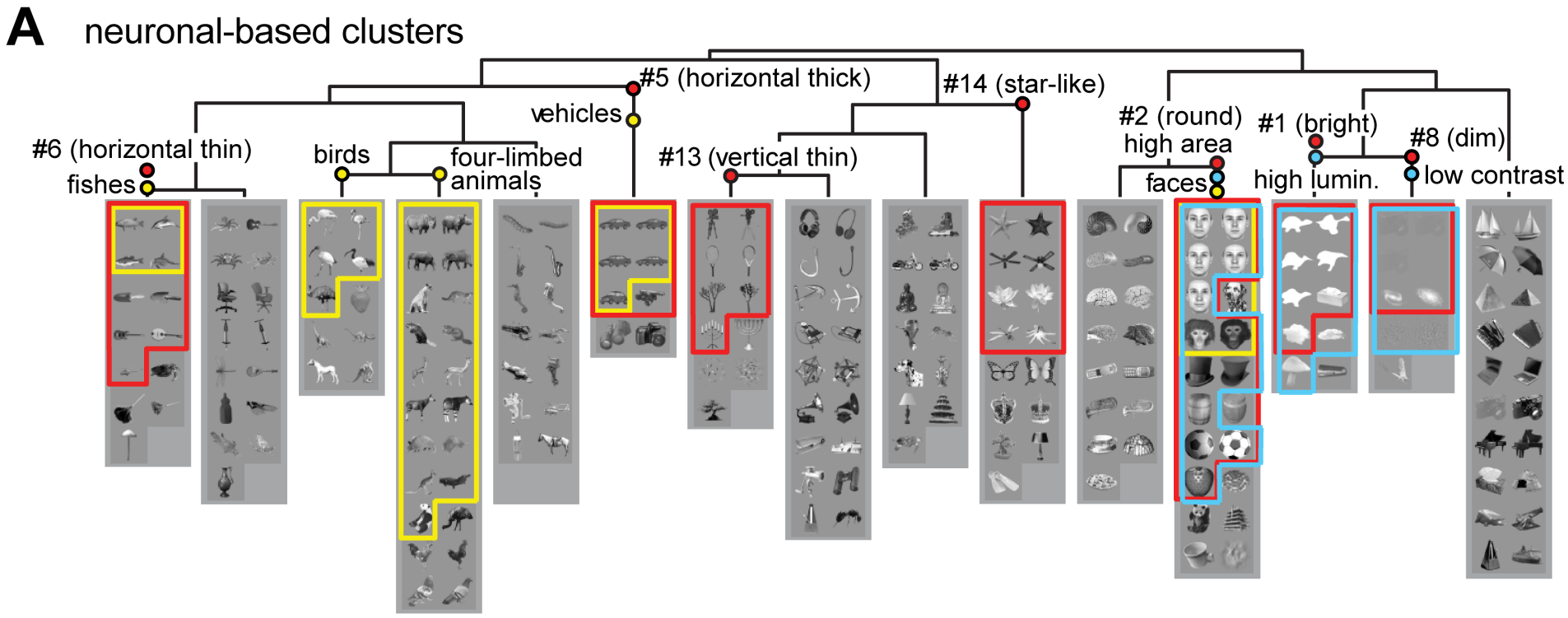


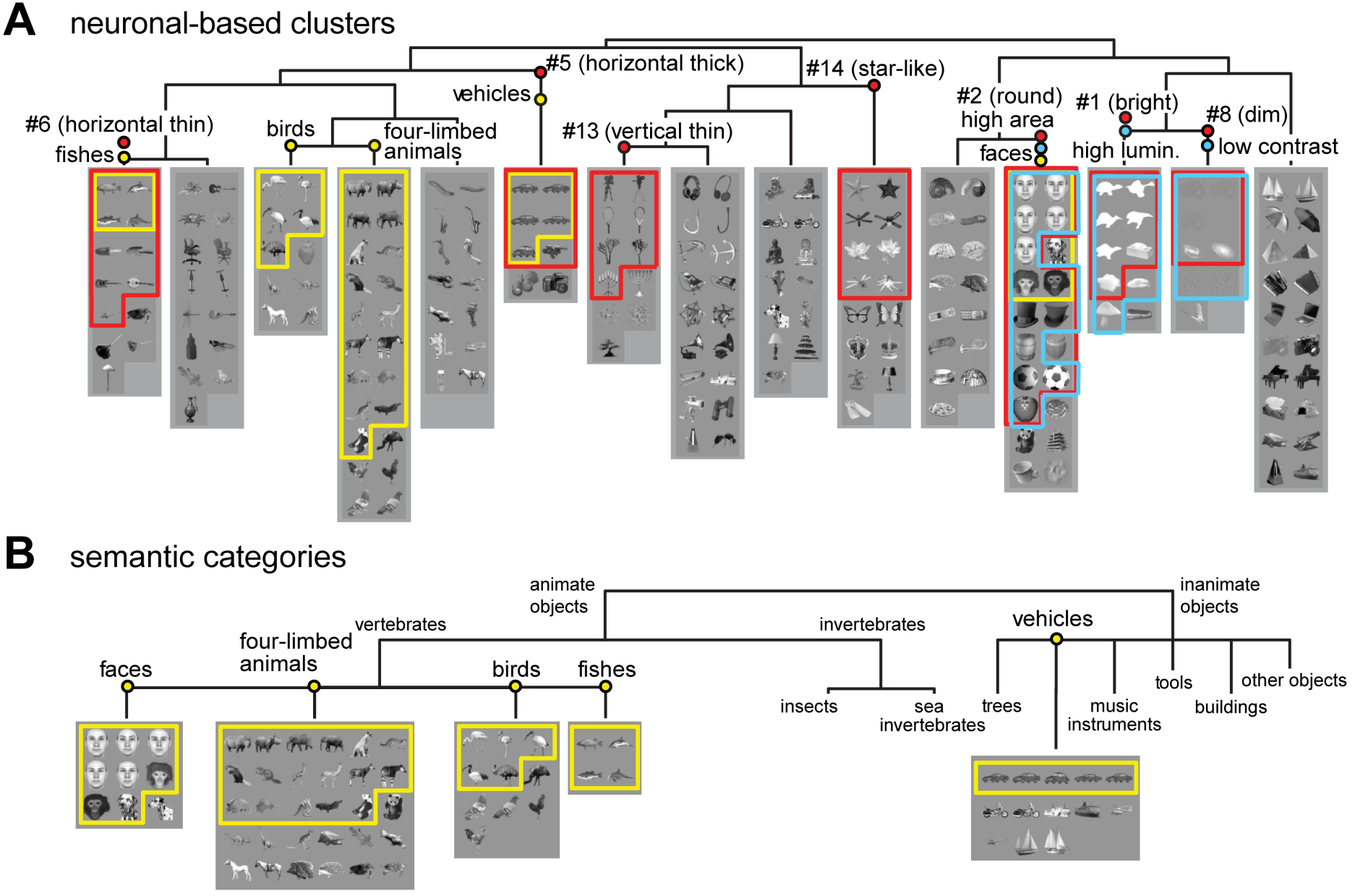

A
Different
Similar
B
animate objects
inanimate objects
NS
% of occurrences
100
80
60
40
20
0
branch 1
branch 2
C
big area
small area
* *
% of occurrences
100
80
60
40
20
0
branch 1
branch 2
D
second principal component
first principal component
E
Object area
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−8
−6
−4
−2
0
2
4
6
8
first principal component

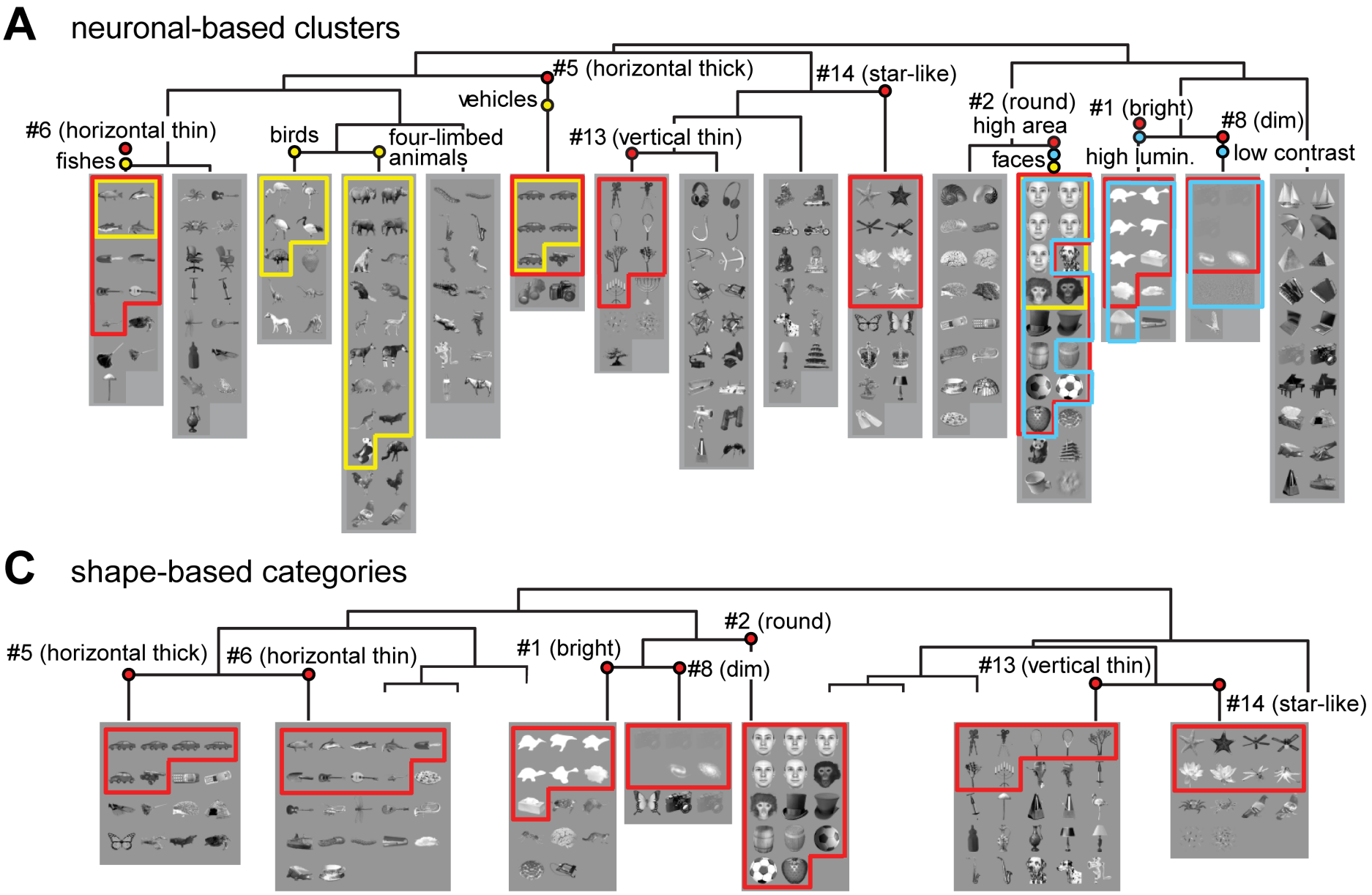
Structure of Visual Object Representation in IT

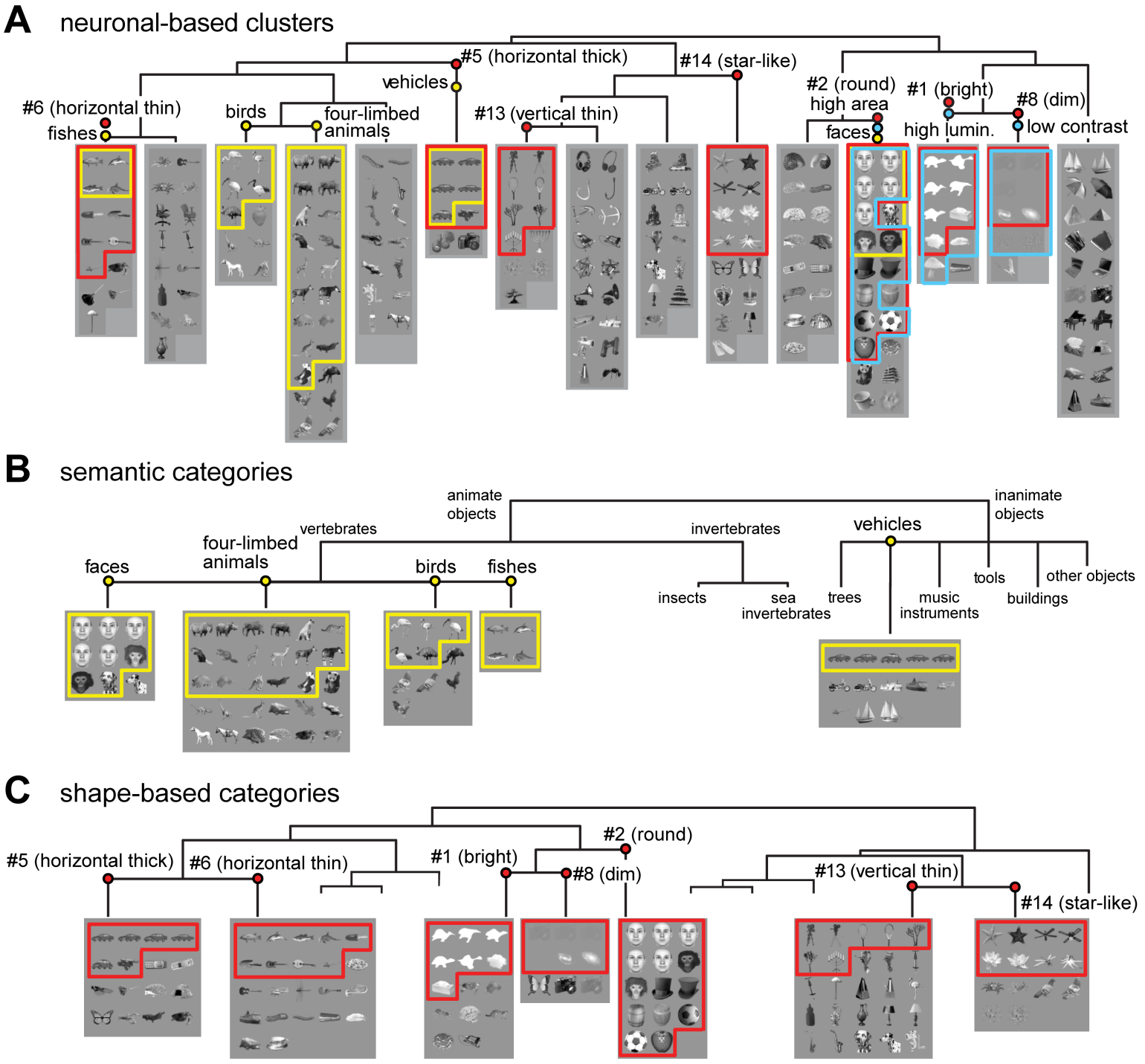


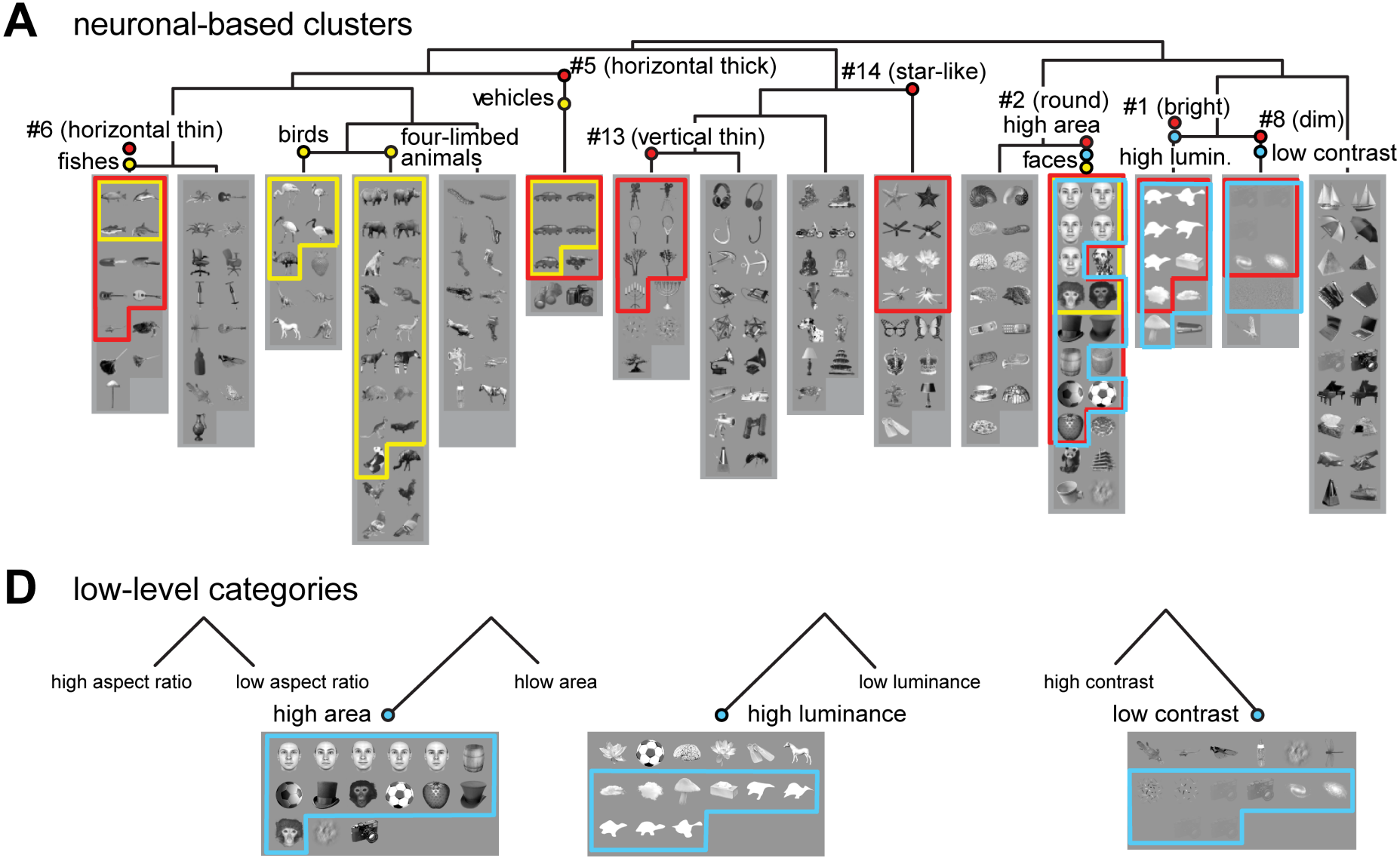
K-means on Neuronal Population Vectors

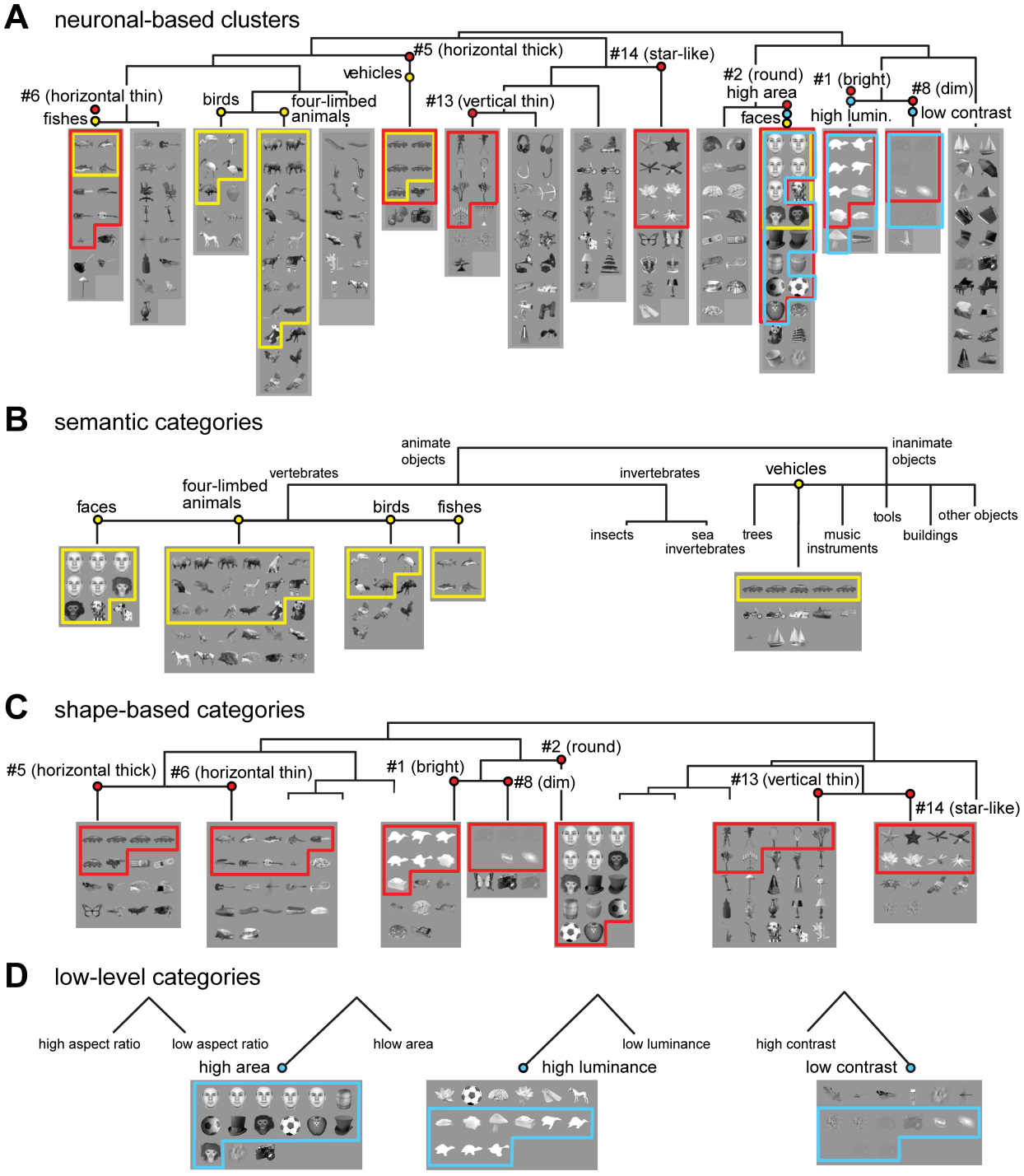


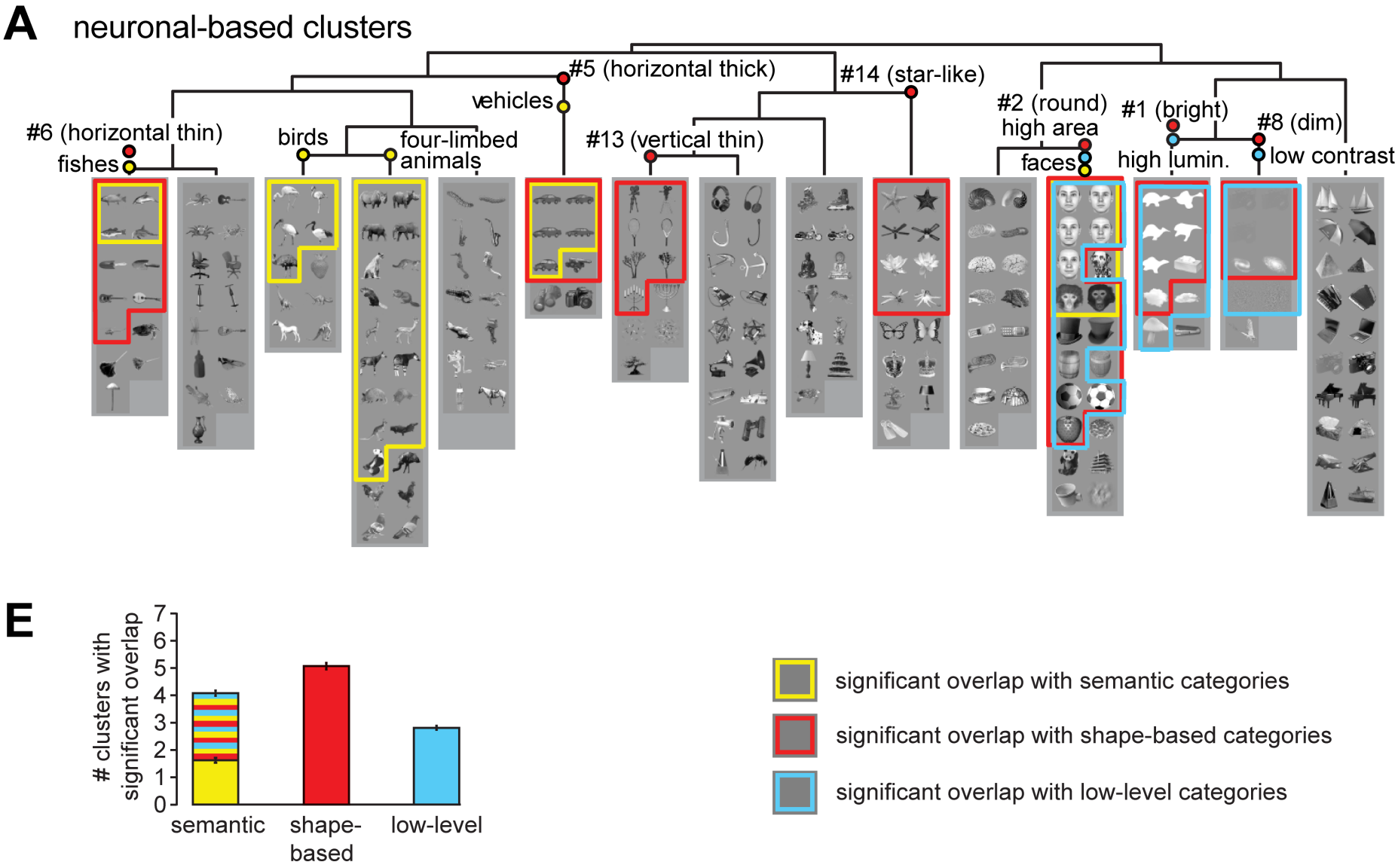

A
neuronal-based clusters
#6 (horizontal thin)
fishes
birds
four-limbed animals
vehicles
#5 (horizontal thick)
#13 (vertical thin)
#14 (star-like)
#2 (round) high area
faces
#1 (bright)
high lumin.
#8 (dim)
low contrast
B
semantic categories
animate objects
inanimate objects
vertebrates
invertebrates
vehicles
faces
four-limbed animals
birds
fishes
insects
sea invertebrates
trees
music instruments
tools
buildings
other objects


A
neuronal-based clusters
#5 (horizontal thick)
#14 (star-like)
vehicles
#6 (horizontal thin)
birds
four-limbed animals
#13 (vertical thin)
#2 (round) high area
#1 (bright)
#8 (dim)
fishes
faces
high lumin.
low contrast
C
shape-based categories
#2 (round)
#5 (horizontal thick)
#6 (horizontal thin)
#1 (bright)
#8 (dim)
#13 (vertical thin)
#14 (star-like)


A
neuronal-based clusters
#5 (horizontal thick)
#14 (star-like)
vehicles
#2 (round)
high area
#1 (bright)
#8 (dim)
#6 (horizontal thin)
birds
four-limbed
animals
#13 (vertical thin)
faces
high lumin.
low contrast
fishes
B
semantic categories
animate
objects
inanimate
objects
vertebrates
invertebrates
vehicles
four-limbed
animals
faces
birds
fishes
insects
sea
invertebrates
trees
music
instruments
tools
buildings
other objects
C
shape-based categories
#2 (round)
#5 (horizontal thick)
#6 (horizontal thin)
#1 (bright)
#8 (dim)
#13 (vertical thin)
#14 (star-like)


A
neuronal-based clusters
#6 (horizontal thin)
fishes
birds
four-limbed animals
vehicles
#5 (horizontal thick)
#13 (vertical thin)
#14 (star-like)
#2 (round) high area
faces
#1 (bright)
high lumin.
#8 (dim)
low contrast
D
low-level categories
high aspect ratio
low aspect ratio
high area
hlow area
low luminance
high luminance
high contrast
low contrast


A
neuronal-based clusters
#5 (horizontal thick)
#14 (star-like)
vehicles
#2 (round)
high area
#1 (bright)
#8 (dim)
#6 (horizontal thin)
birds
four-limbed animals
#13 (vertical thin)
faces
high lumin.
low contrast
fishes
B
semantic categories
animate objects
inanimate objects
vertebrates
invertebrates
vehicles
four-limbed animals
faces
birds
fishes
insects
sea invertebrates
trees
music instruments
tools
buildings
other objects
C
shape-based categories
#2 (round)
#5 (horizontal thick)
#6 (horizontal thin)
#1 (bright)
#8 (dim)
#13 (vertical thin)
#14 (star-like)
D
low-level categories
high aspect ratio
low aspect ratio
high area
hlow area
low luminance
high luminance
high contrast
low contrast


A
neuronal-based clusters
#6 (horizontal thin)
fishes
birds
four-limbed
animals
vehicles
#5 (horizontal thick)
#13 (vertical thin)
#14 (star-like)
#2 (round)
high area
faces
#1 (bright)
high lumin.
#8 (dim)
low contrast
E
clusters with
significant overlap
0
1
2
3
4
5
6
7
semantic
shape-
based
low-level
significant overlap with semantic categories
significant overlap with shape-based categories
significant overlap with low-level categories

Animate vs. Inanimate

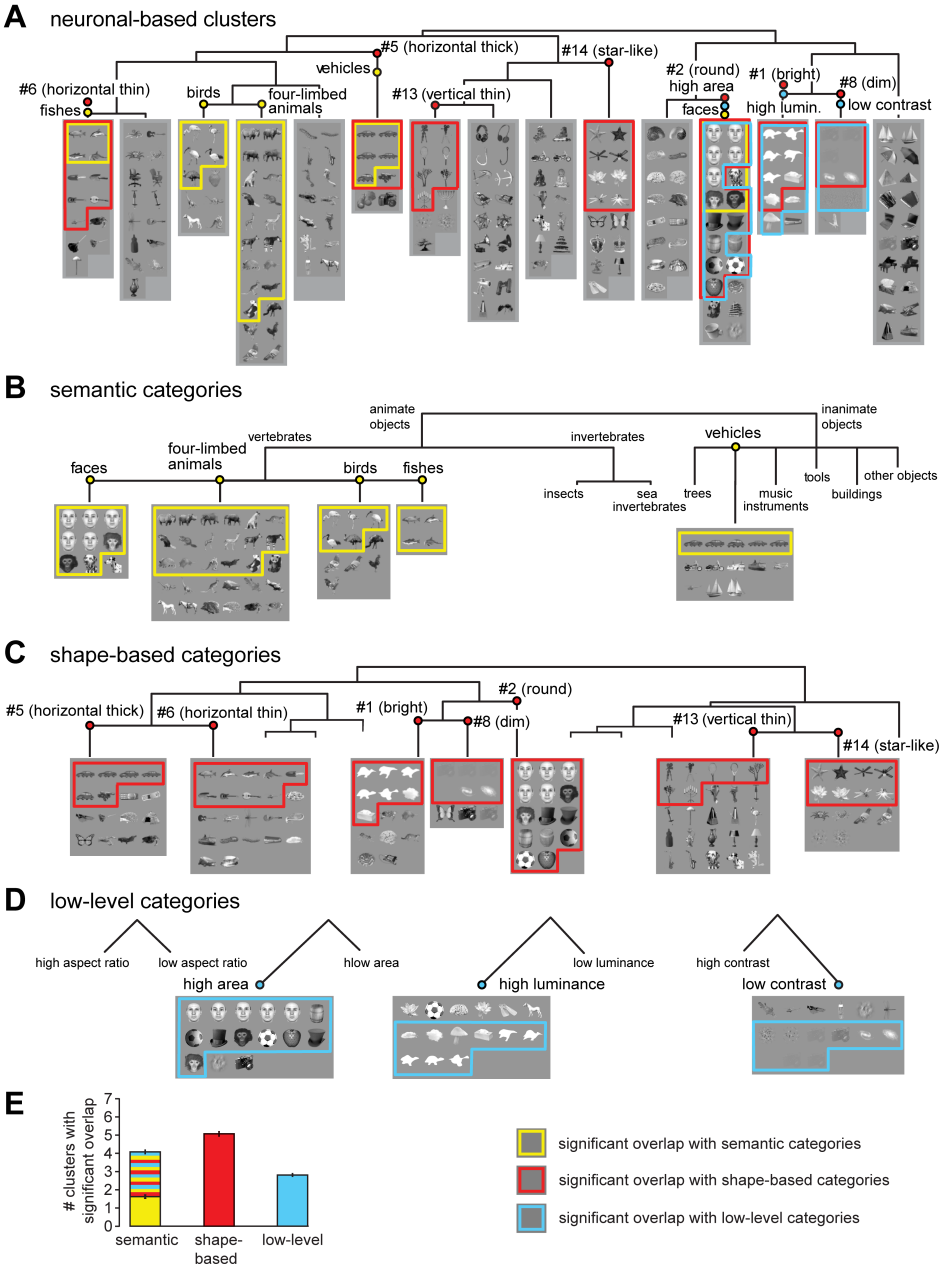
- K=2 (100 runs)
- |animate_C1 –animate_C2| = ~7%
- Not significantly larger than chance (p=0.39)

- Again similar to pearson coefficient.

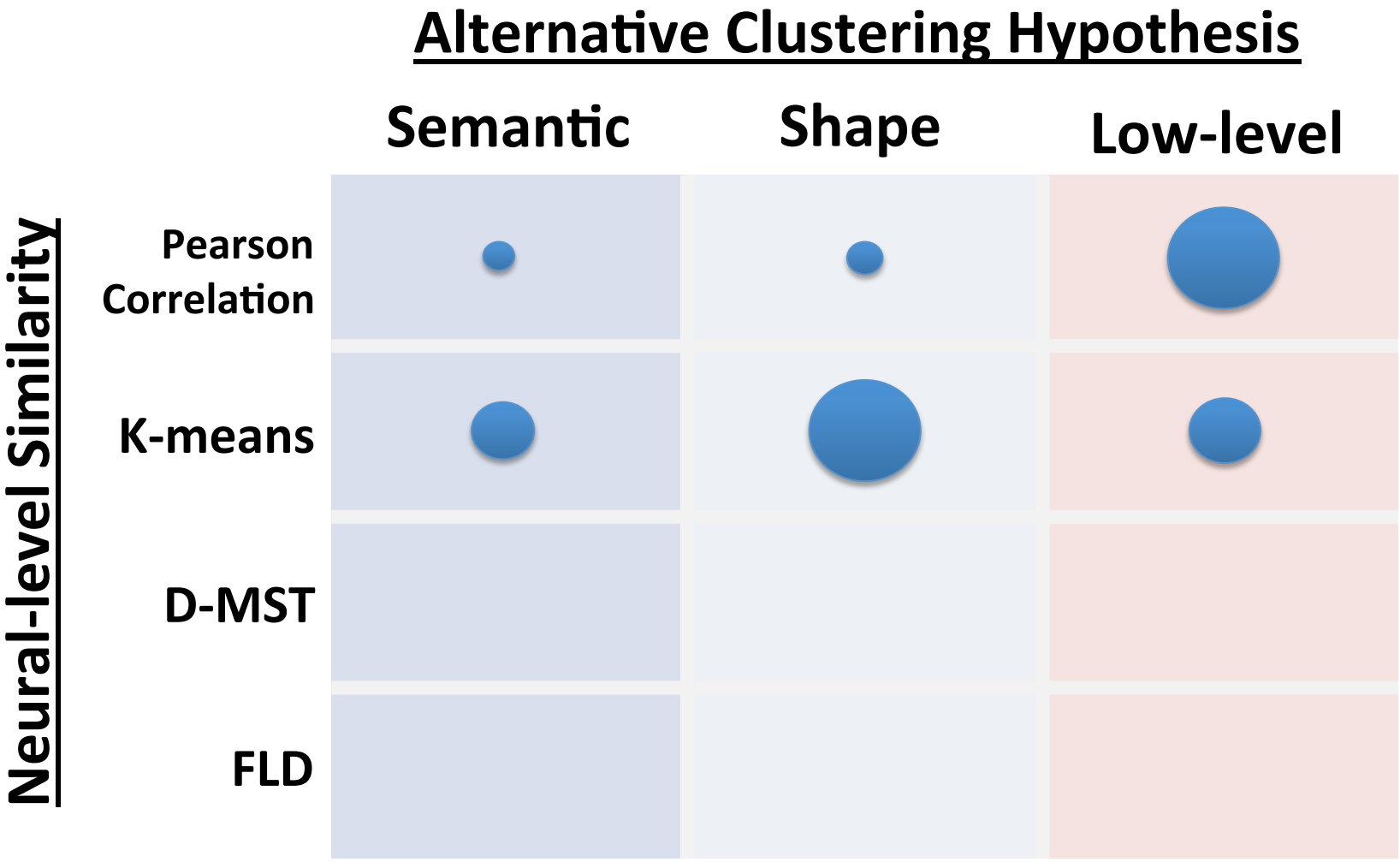
K-Means Analysis

- Most clusters explainable by visual similarity
- Both shape & **low-level**
- Few semantic categories do exist:

  Birds, four-limbed animals



Structure of Visual Object Representation in IT



D-MST Clustering

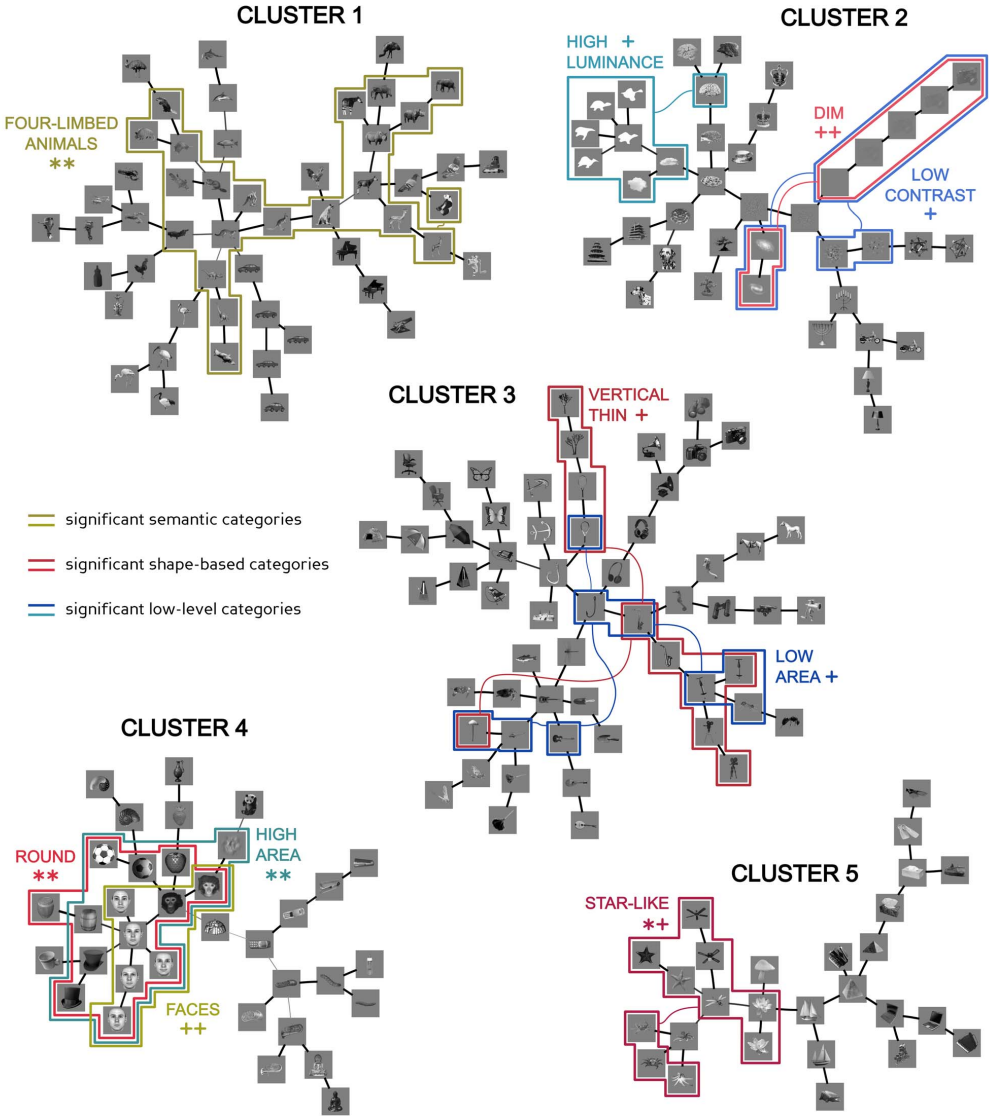
- Un-supervised Clustering

- Combines advantages of both:
  - K-means like partitions -- allow studying overlaps
  - Hierarchical approaches – fine-grained relationships b/w objects

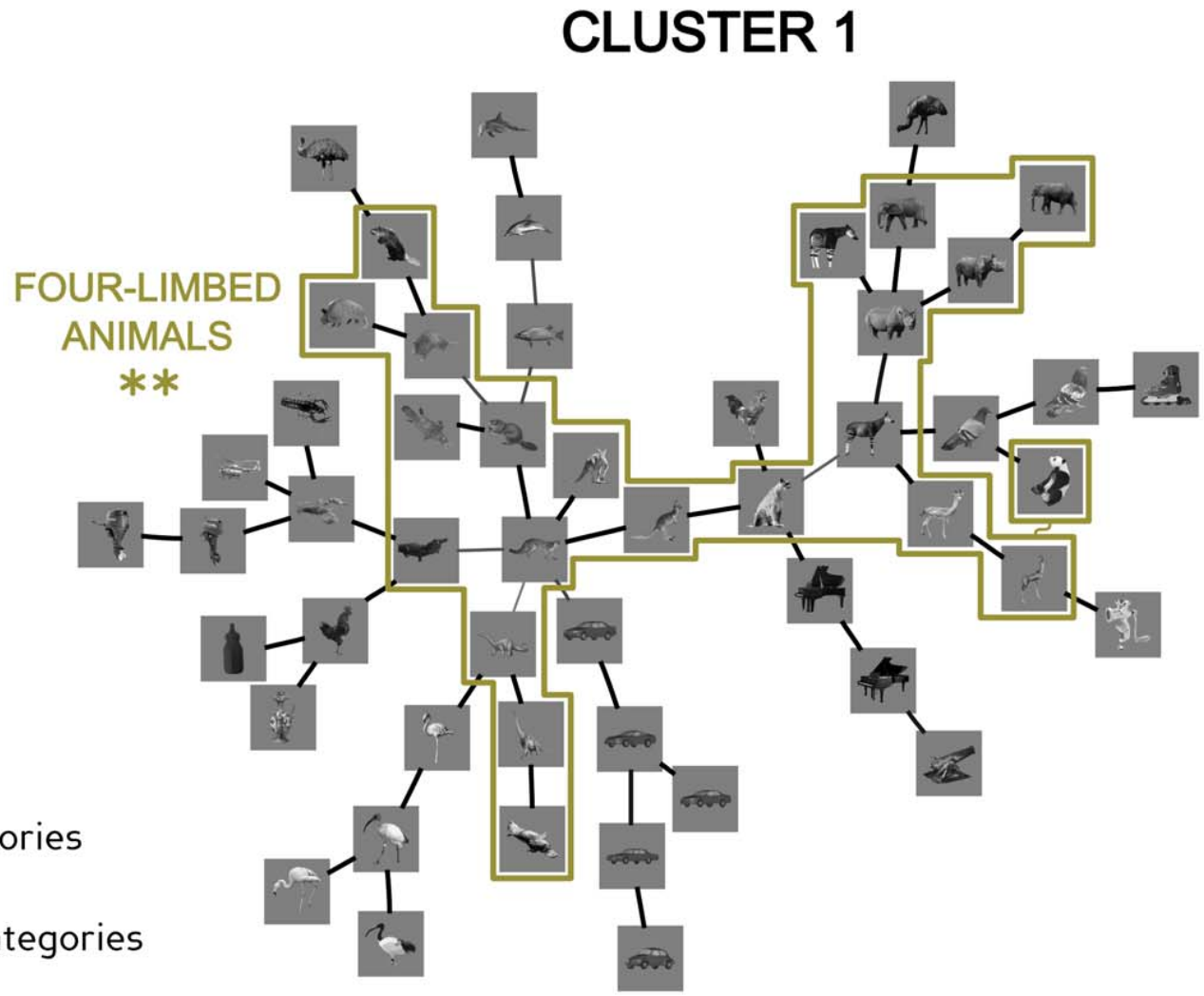
- Allows non-spherical clusters
  - As opposed to kmeans

- Outputs a forest – richer information about topology/structure of data

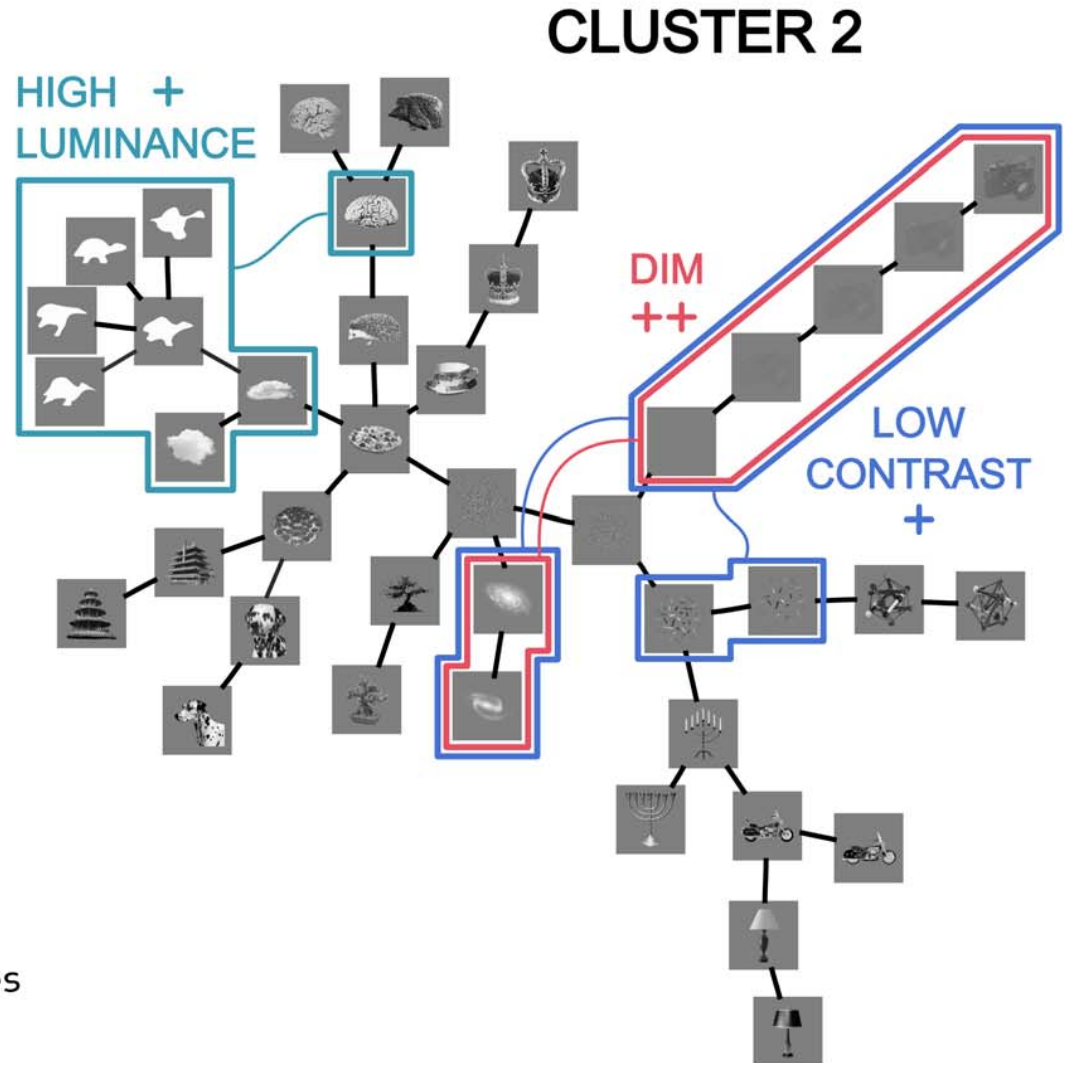
D-MST Clustering



D-MST Clustering



D-MST Clustering

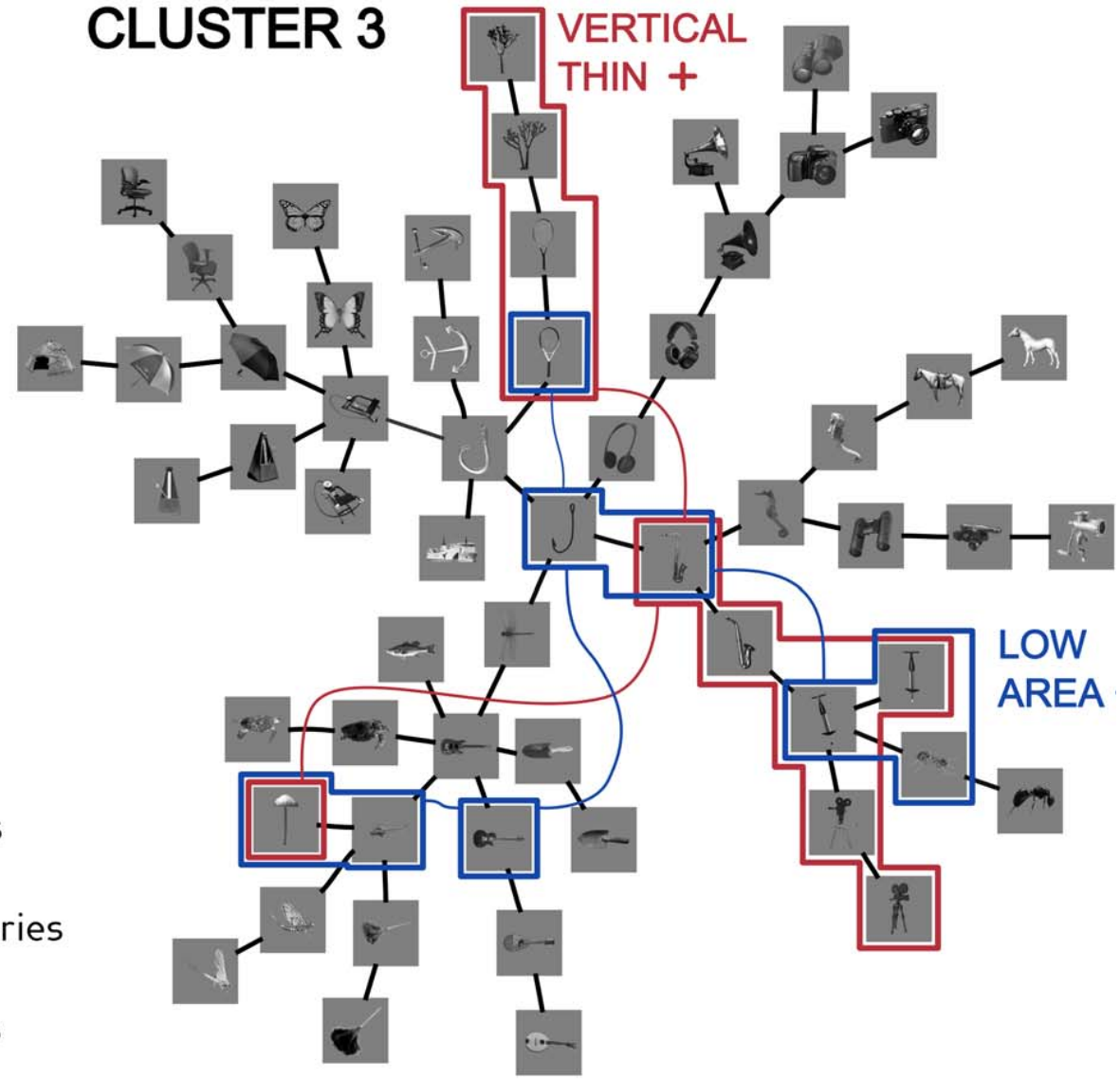


significant semantic categories

significant shape-based categories

significant low-level categories

D-MST Clustering

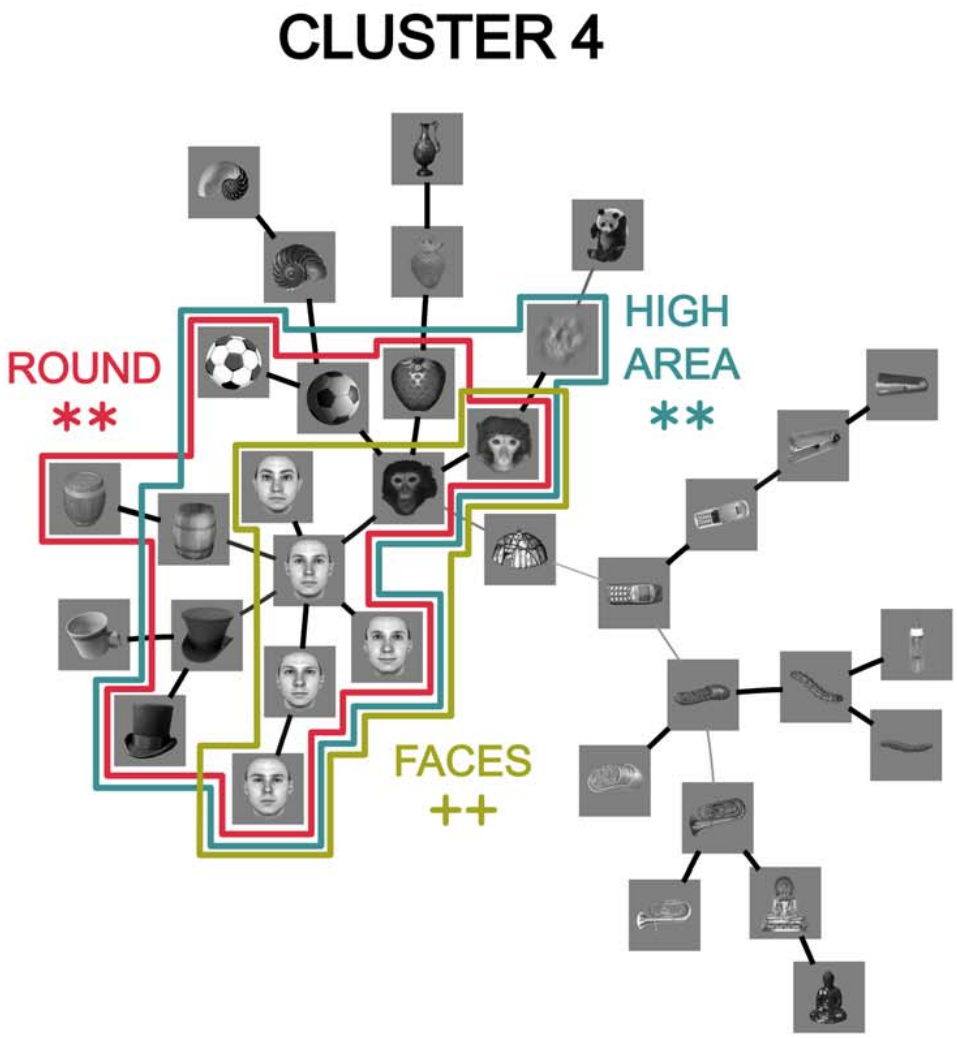


significant semantic categories

significant shape-based categories

significant low-level categories

D-MST Clustering

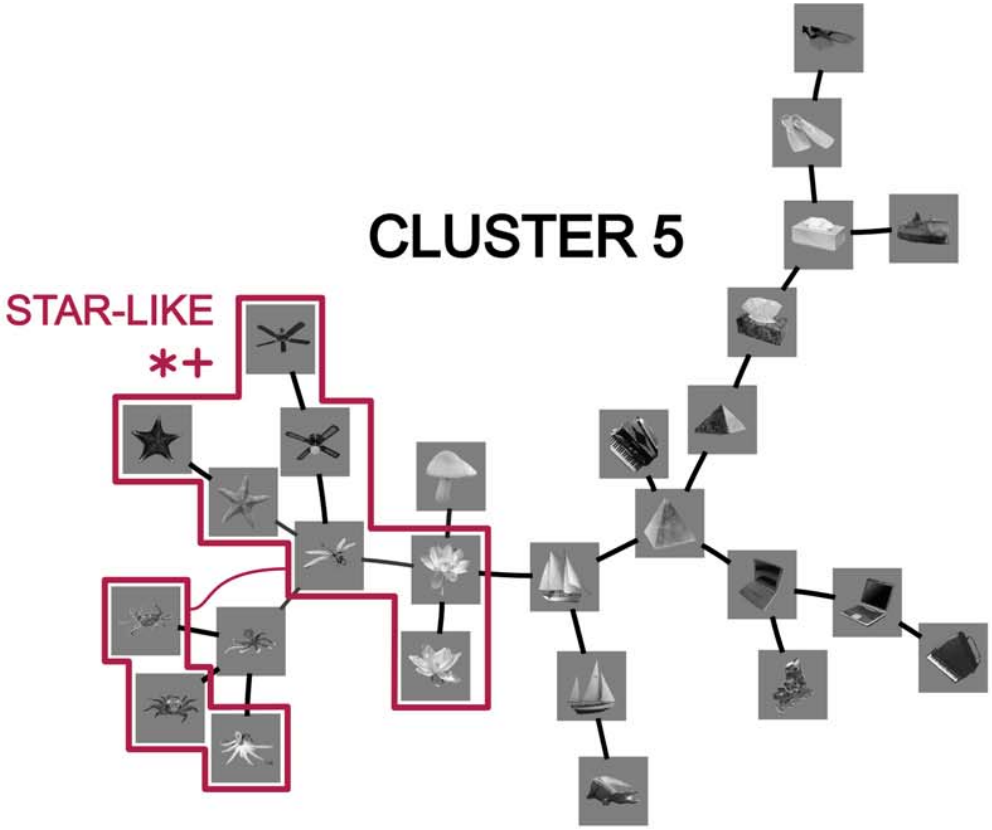


significant semantic categories

significant shape-based categories

significant low-level categories

D-MST Clustering



significant semantic categories

significant shape-based categories

significant low-level categories

Overlap (this paper vs. Kiani et a.)

- Compensating for existence of multiple (very similar) exemplars of same objects (i.e., twins)

- When using 'buggy' Overlap
  - Some more semantic overlaps
  - Most of those overlaps explained by shape or low-level overlaps

Table 1. Overlapping between semantic categories and D-MST neuronal-based clusters.

Category	D-MST Cluster	Ratio 1	Ratio 2	Overlap	*p* (twins)	Signif.	p (obj.)	Signif.
Four-limb. anim.	1	0.73	0.96	0.71	0.0000	**	0.0000	**
Faces	4	0.78	1.00	0.78	0.0023	++	0.0000	**
Fishes	1	0.75	1.00	0.75	0.0742		0.0007	*+
Sea invertebr.	5	0.50	0.86	0.46	0.0840		0.0004	**
Birds	1	1.00	0.48	0.48	0.1048		0.0003	**
Music instr.	3	0.50	0.75	0.43	0.1140		0.0012	*+
Vehicles	1	0.46	0.67	0.37	0.2617		0.0065	++
Insects	3	0.58	0.47	0.35	0.3635		0.0192	+
Tools	3	0.58	0.44	0.33	0.4587		0.0365	+
Trees	5	0.30	1.00	0.30	0.6240		0.0979	
Buildings	5	0.33	1.00	0.33	0.8883		0.1471	

The table reports the overlap (fifth column) between each semantic category (first column) and the D-MST neuronal-based cluster (second column) containing the best matching sub-tree of contiguous objects, according to a score defined as the ratio between the intersection of the sub-tree with the category and their union (fifth column). Significance of the overlap was computed by permuting (1,000,000 times) either sets of twin objects (forth- and third-to-last columns) or individual objects (second-to-last and last columns) across the categories of a given clustering hypotheses: Holm-Bonferroni corrected $p<0.01$ (**) and $p<0.05$ (* and *+); and uncorrected $p<0.01$ (++ and *+) and $p<0.05$ (+). For comparison with [14], two other overlap metrics (Ratio 1 = the fraction of objects in the category overlapping with the cluster; and Ratio 2 = the fraction of objects in the cluster overlapping with the category) are also reported.

Table 2. Overlapping between shape-based categories and D-MST neuronal-based clusters.

Category	D-MST Cluster	Ratio 1	Ratio 2	Overlap	*p* (twins)	Signif.	*p* (obj.)	Signif.
#2 (round)	4	1.00	1.00	1.00	0.0000	**	0.0000	**
#14 (star-like)	5	0.71	0.91	0.67	0.0007	*+	0.0000	**
#8 (dim)	2	0.78	0.78	0.64	0.0097	++	0.0000	**
#13 (vertical thin)	3	0.52	0.68	0.42	0.0347	+	0.0002	**
#6 (horiz. thin)	3	0.41	1.00	0.41	0.0520		0.0003	**
#1 (bright)	2	0.57	0.66	0.44	0.0748		0.0004	**
#5 (horiz. thick)	1	0.44	0.87	0.41	0.0927		0.0008	*+
#12 (diagonal)	1	0.47	0.50	0.32	0.4299		0.0392	+
#15	1	0.50	0.50	0.33	0.4878		0.0368	+
#10	3	0.45	0.50	0.31	0.5313		0.0667	
#11	3	0.31	1.00	0.30	0.5347		0.0582	
#4	1	0.45	0.41	0.28	0.7109		0.1694	
#7 (pointy)	5	0.27	0.60	0.23	0.9279		0.4949	
#9	1	0.29	0.50	0.22	0.9451		0.5630	
#3	2	0.33	0.40	0.22	0.9530		0.5768	

The table reports the overlap (fifth column) between each shape-based category (first column) and the D-MST neuronal-based cluster (second column) containing the best matching sub-tree of contiguous objects. Same table structure and symbols as in Table 1.

doi:10.1371/journal.pcbi.1003167.t002

Table 3. Overlapping between low-level categories and D-MST neuronal-based clusters.

Category	D-MST Cluster	Ratio 1	Ratio 2	Overlap	*p* (twins)	Signif.	*p* (obj.)	Signif.
High area	4	0.93	1.00	0.93	0.0000	**	0.0000	**
Low contrast	2	0.60	0.82	0.53	0.0103	+	0.0000	**
Low area	3	0.60	0.69	0.47	0.0333	+	0.0001	**
High luminance	2	0.53	0.80	0.47	0.0352	+	0.0001	**
Low aspect ratio	2	0.40	0.86	0.37	0.1910		0.0049	++
High aspect ratio	4	0.33	0.83	0.31	0.4760		0.0454	+
Low luminance	1	0.33	0.42	0.28	0.9240		0.5116	
High contrast	1	0.33	0.36	0.21	0.9761		0.7167	

The table reports the overlap (fifth column) between each low-level category (first column) and the D-MST neuronal-based cluster (second column) containing the best matching sub-tree of contiguous objects. Same table structure and symbols as in Table 1.

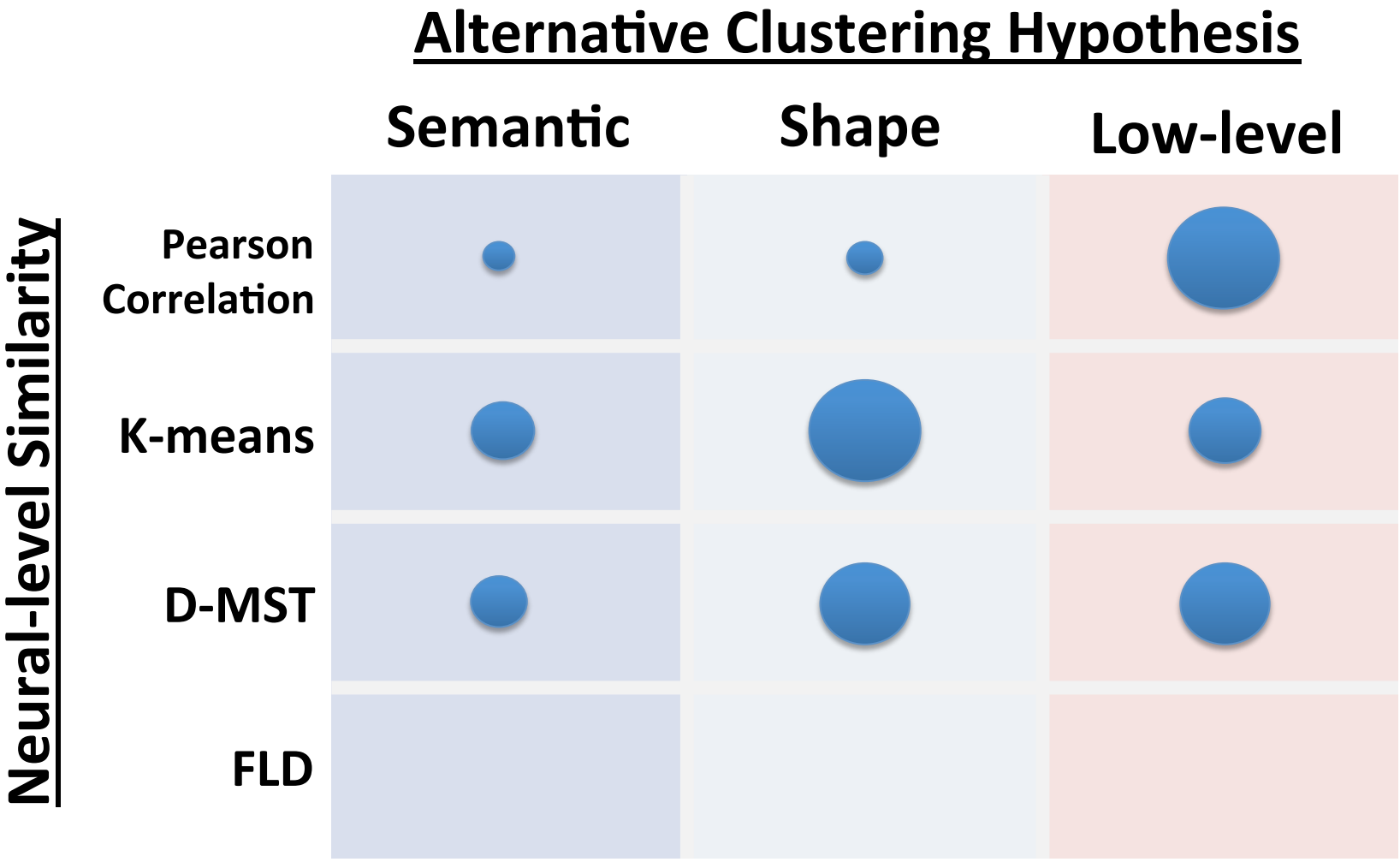
D-MST Clustering Analysis

The object clustering produced by the D-MST algorithm suggests the existence of a rich multi-level object representation in IT,

which is largely driven by the similarity of visual objects across a spectrum of visual properties,

ranging from low-level image attributes to complex combinations of shape features that are often hard to model and quantify.

Structure of Visual Object Representation in IT



Unsupervised to Supervised Analysis

- Unsupervised approaches (K-means, D-MST)
  - Discover “natural” internal structure
  - No assessment of how much information does it convey about a given object set (?)
  - Based on average firing rates – do not take into account trial-by-trial variability of neuronal responses

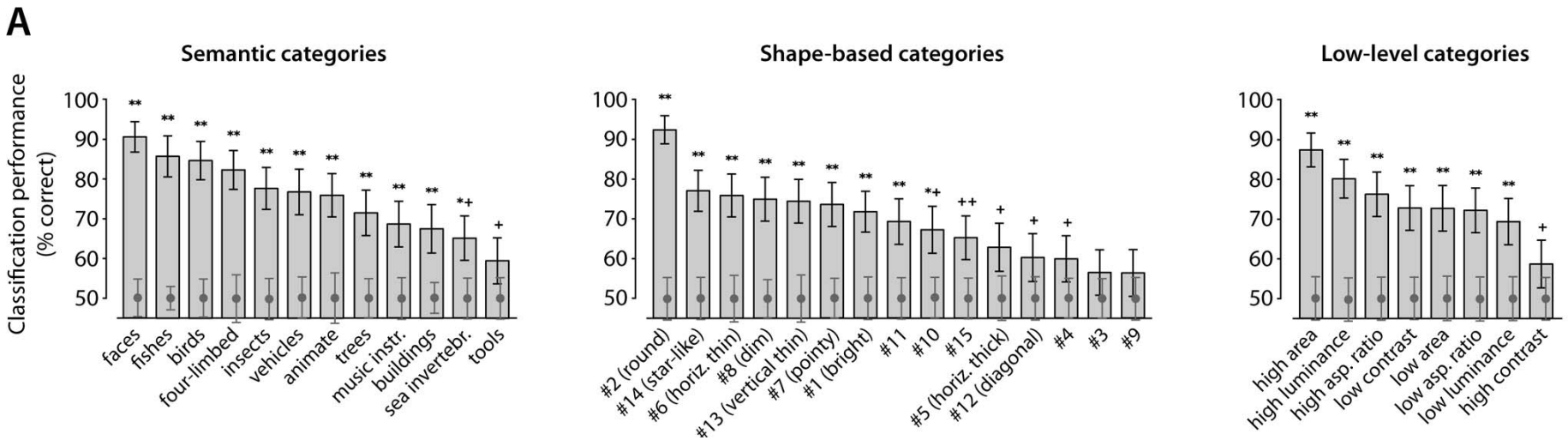
- Supervised decoding approaches
  - Discriminant-based linear classifiers
    - Quiroga et al. 2009, DiCarlo et al. 2005/09/10 etc.
  - Linear read-out schemes (?)

Fisher Linear Discriminants (FLDs)

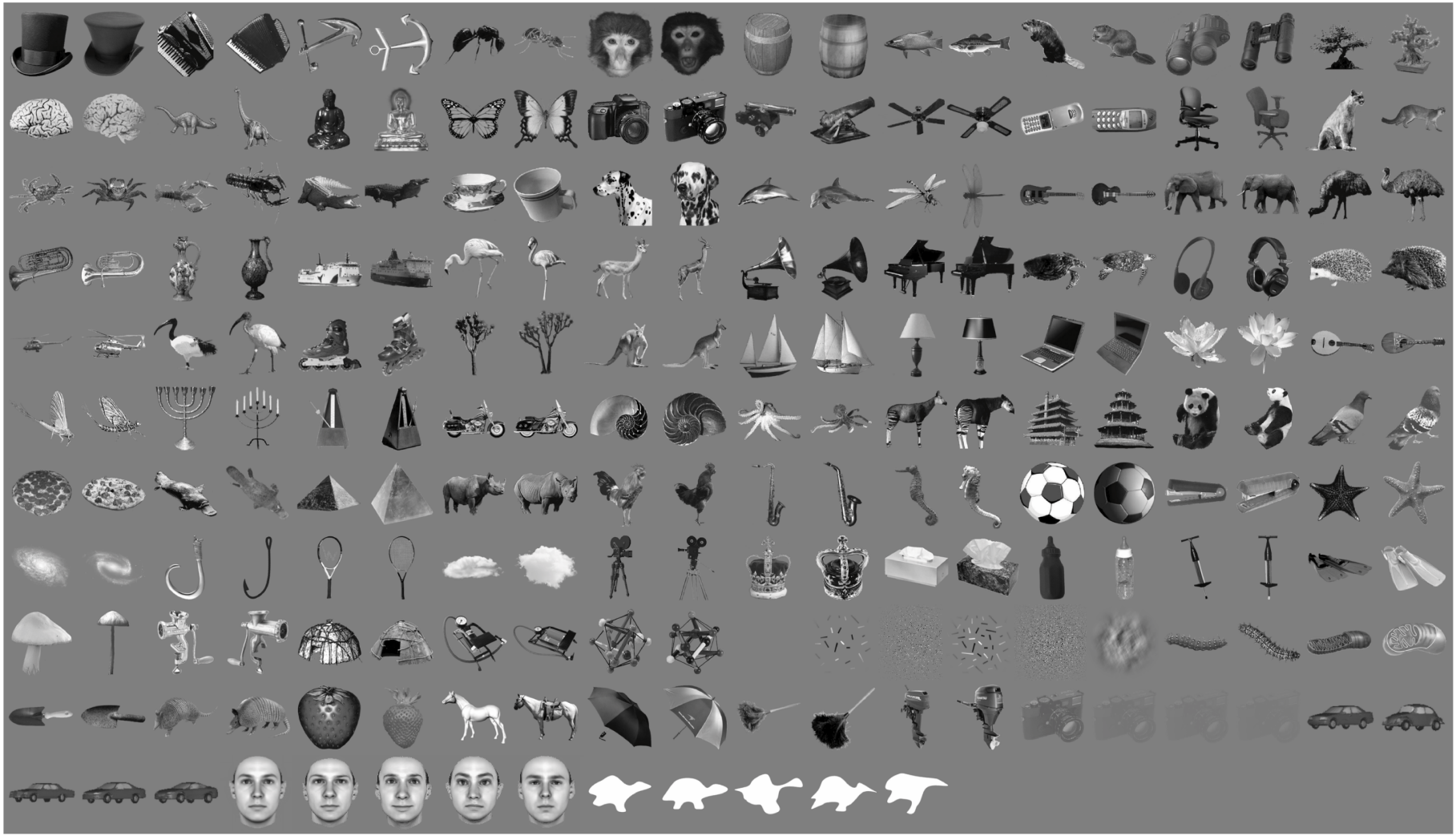
- Given neuronal response, perform a binary classification task for each object (e.g., faces vs. everything else, round vs. everything else etc.)

- Capability of FLDs to classify objects not used in training – population vectors for different presentation of a **given object in a given category** were excluded from training set.
  - A given face in faces category*****

FLDs Compared to Categories



Again, ‘twins’ problematic



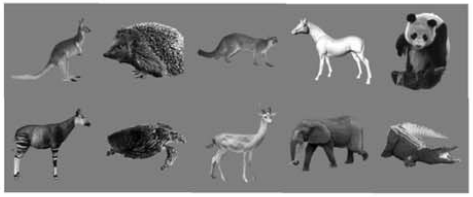
Pruned Sets

- Isolate semantics from shape and low-level etc.

Examples of *pruned* semantic categories

birds

four-limbed animals



Examples of *pruned* shape-based categories

round

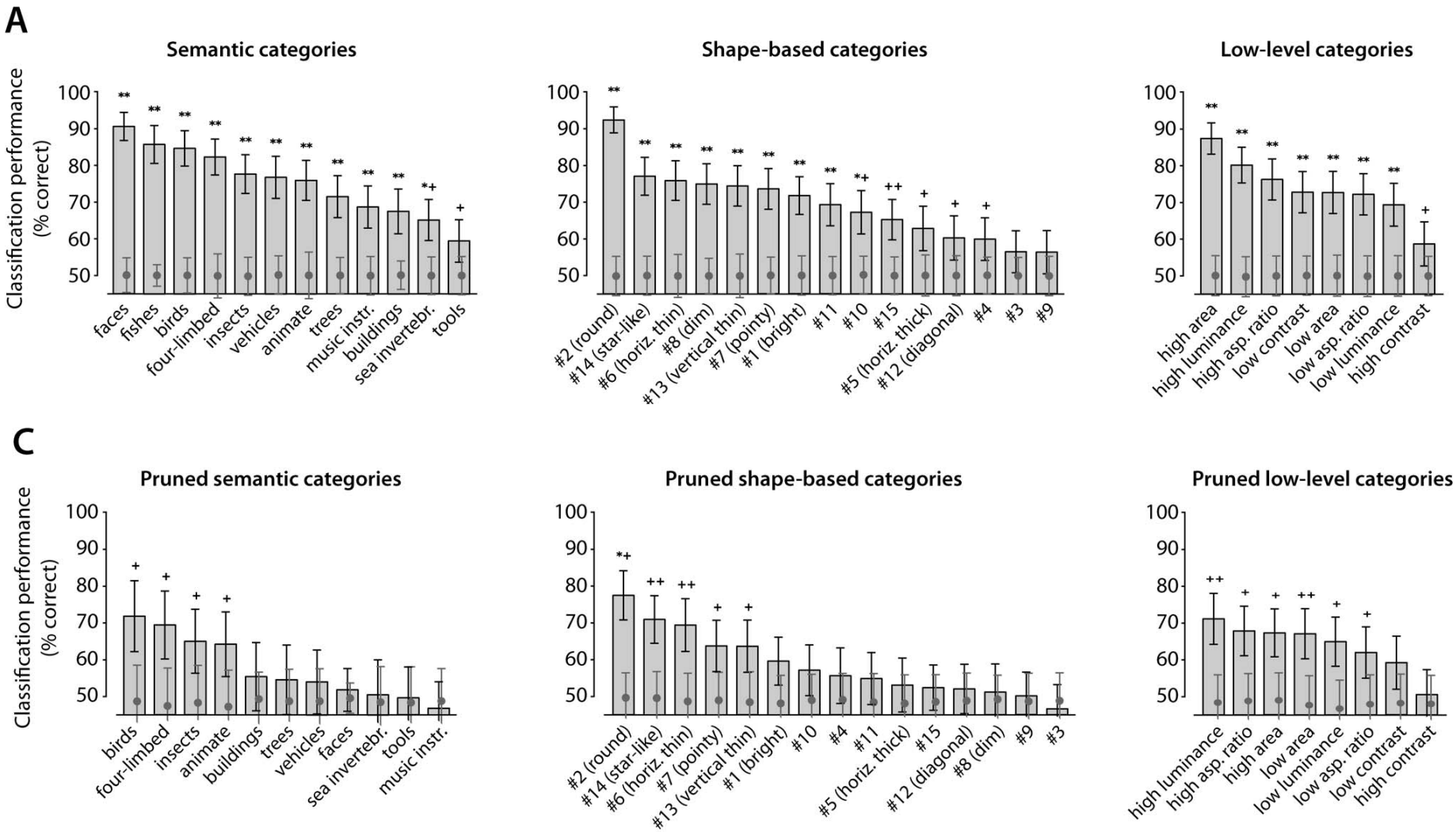
star-like

Examples of *pruned* low-level categories

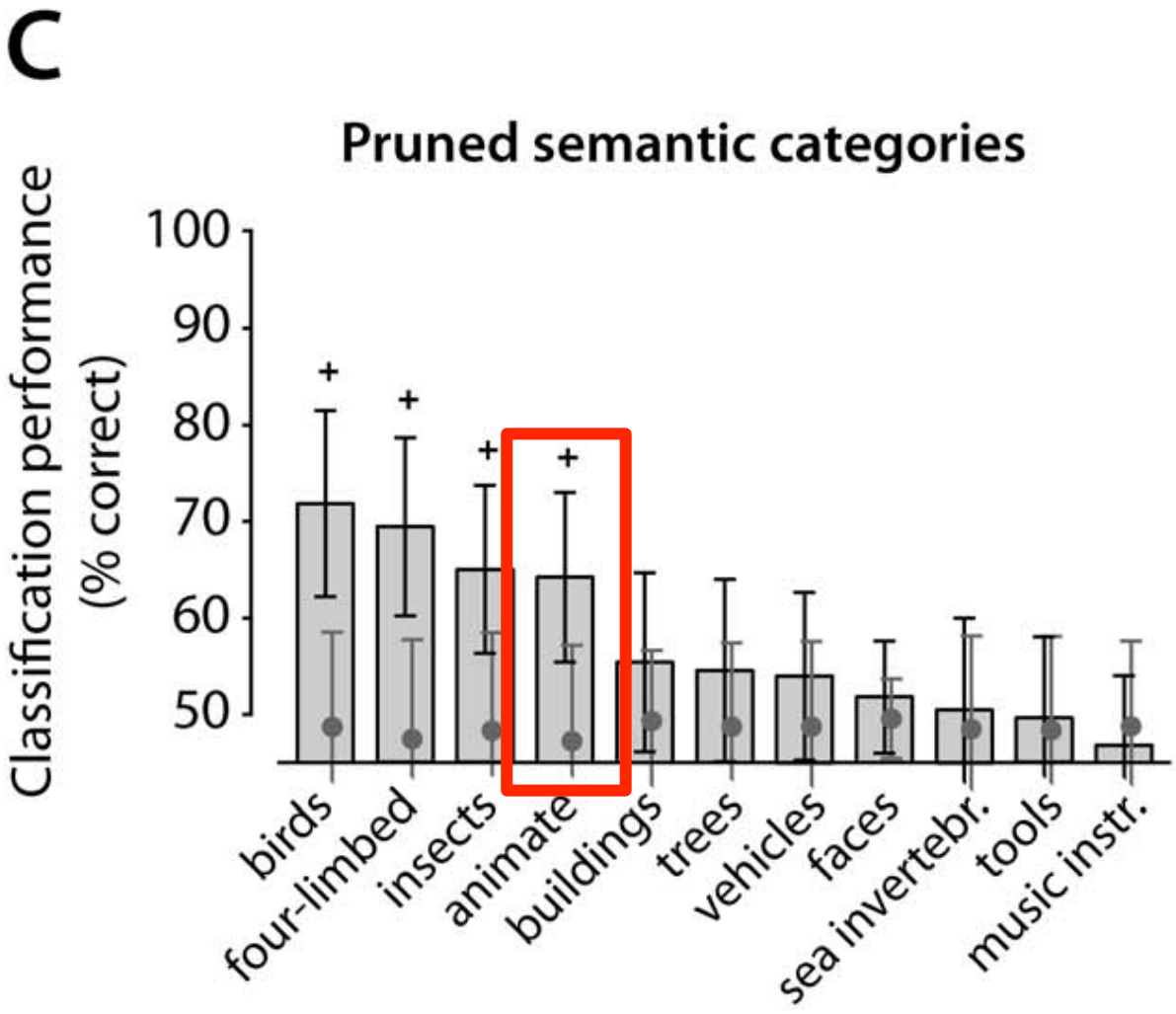
high luminance

high aspect ratio

Actual vs. Pruned



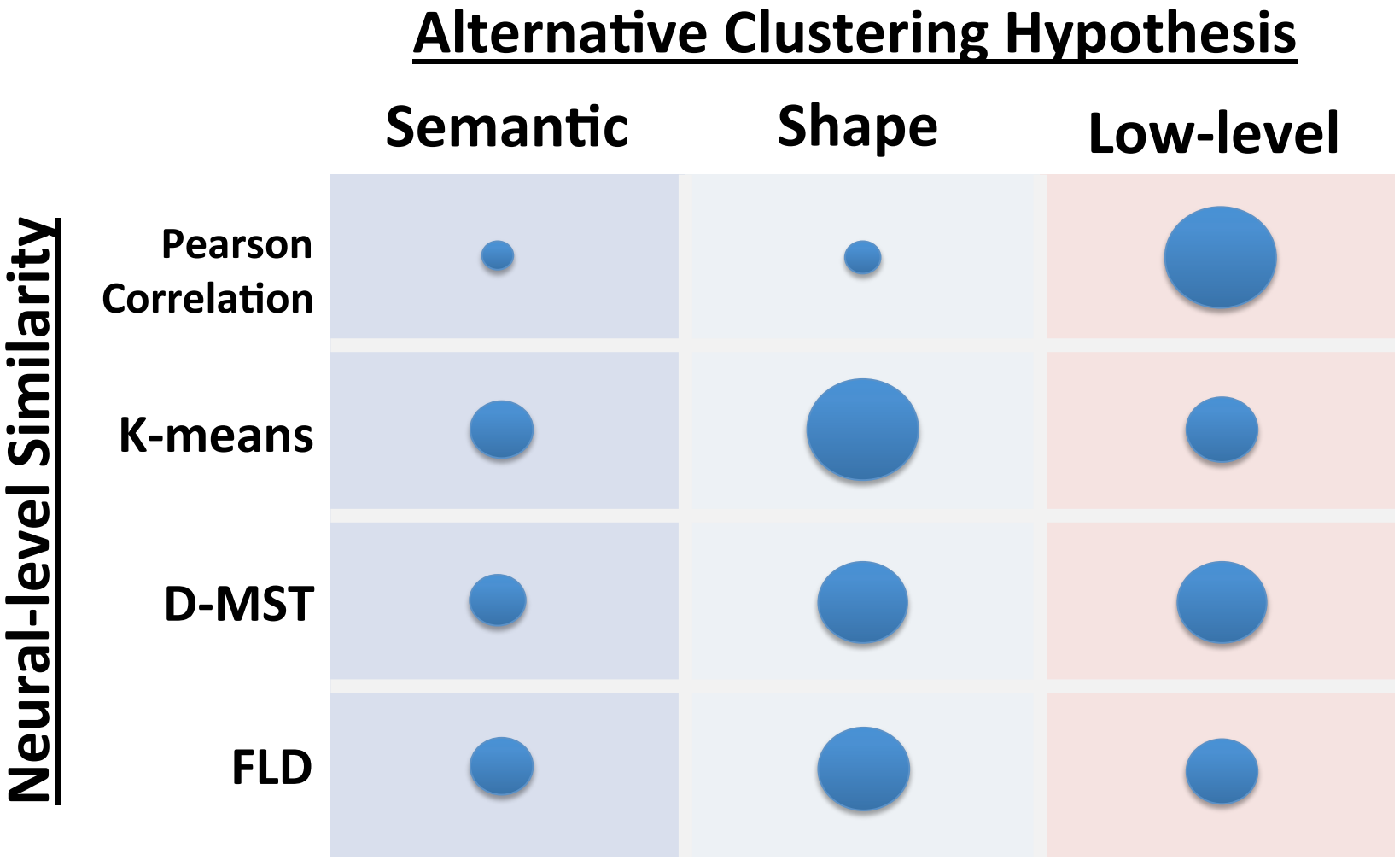
Animate vs. Inanimate



Animate vs. Inanimate

- FLDs, being supervised approaches, do not need to follow the ‘‘natural’’ object segregation in the IT representation
- Given the high dimensionality of the representation space, FLDs could find a hyperplane segregating the two main animate groups (i.e., four-limbed animals and the faces) from the inanimate objects, even if those groups belong to different ‘‘natural’’ clusters.

Structure of Visual Object Representation in IT



Multiple Clustering hypothesis together

- Supervised and Unsupervised have complementary information

Table 4. Semantic categories significantly represented in IT according to the D-MST and the FLD analyses.

Category	Signif. D-MST (twins' sets perm.)	Signif. FLD (pruned cat.)	Signif. D-MST & FLD
Four-limb. anim.	**	+	✓
Faces	++		
Birds		+	
Insects		+	

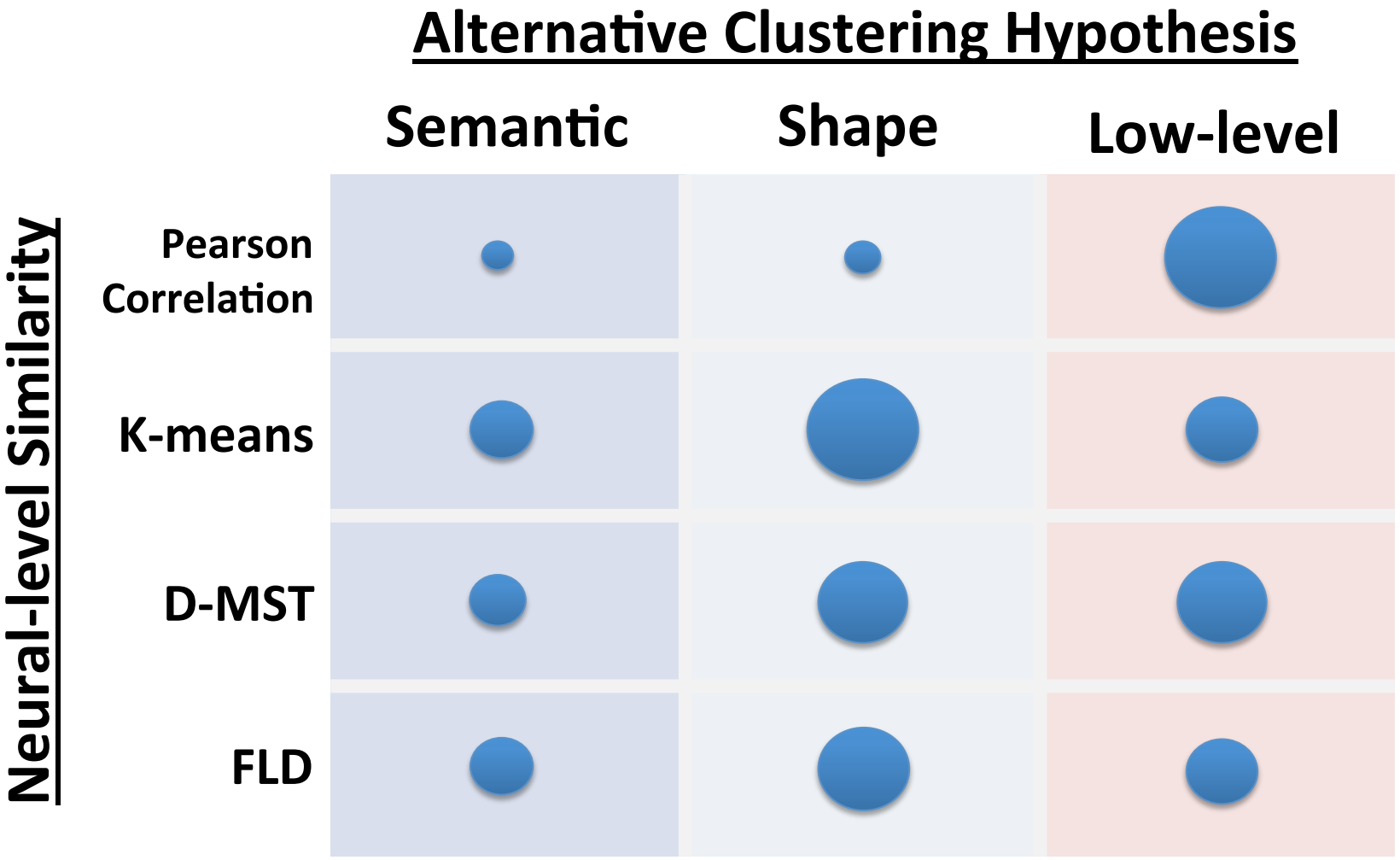
Table 5. Shape-based categories significantly represented in IT according to the D-MST and the FLD analyses.

Category	Signif. D-MST (twins' sets perm.)	Signif. FLD (pruned cat.)	Signif. D-MST & FLD
#2 (round)	**	*+	✓
#14 (star-like)	*+	++	✓
#8 (dim)	++		
#13 (vertical thin)	+	+	✓
#6 (horiz. thin)		++	
#7 (pointy)		+	

Table 6. Low-level categories significantly represented in IT according to the D-MST and the FLD analyses.

Category	Signif. D-MST (twins' sets perm.)	Signif. FLD (pruned cat.)	Signif. D-MST & FLD
High area	**	+	✓
Low contrast	+		
Low area	+	++	✓
High luminance	+	++	✓
Low aspect ratio		+	
High aspect ratio		+	
Low luminance		+	

Structure of Visual Object Representation in IT



Conclusions

- Used array of Supervised and Unsupervised approaches

- Visual objects in neuronal representation space
  - Coarse clustering – low-level visual properties
  - Finer-grain structure – higher-level shape features
  - Little role played by semantics
    - four-limbed animals robustly recorded everywhere
    - (may be evolution?)

Comparisons with [14, 15]

- [14, 15] couldn't find any visual-similarity metric that could reproduce object clusters…
- Apart from four-limbed animals (and birds), no other semantic segregation
  - Insects discriminable by FLDs, but not in kmeans or DMST
  - Faces discriminable, but explained by round (pruned sets)

Comparisons with [14, 15]

Animate vs. Inanimate

- Not in kmeans and DMST
- FLDs could segregate
  - high-dimensionality might be a reason

Strongly suggests no sharp segregation within IT (at least in the ones sampled here). But not randomly scattered..

[34] -- in the body-selective regions of monkey IT, objects do not primarily segregate according to whether they belong to the animate or the inanimate categories

Comparisons with [14, 15]

- Protocols and regions are comparable (not exactly same)
- Analysis with 'twins' compensated for!

- Lower # of objects (213 vs. 1084 in Kiani et al.)
- Smaller population recorded (94 vs. 674 IT neurons)
  - See paper for more detailed discussions on affect of these
- Different extent of IT sampled
  - Theirs is much smaller extent as opposed to Kiani et al.
  - Possibility of picking up on face-selective cells

Conclusion of Comparisons

To conclude, it is hard to infer what methodological differences may be at the root of the discrepancies between our study and [14,15]. Above, we have listed some of the differences that could be crucial. Ultimately, however, only a re-analysis of Kiani and colleagues' data with our analytical/statistical approaches, or, better, a full new set of recordings (e.g., with grayscale versions of the images used by Kiani and colleagues) could shed more light on the causes of these discrepancies. Both approaches are clearly beyond the scope of this study, but could be an interesting target of future investigations by ours or other groups.

Disclaimer

- Validity and implications of findings
  - Please read.. ☺

My (biased) Conclusion

- Excellent paper (both Vision and Neuroscience)
- Learning Visual Models
  - Visual sub-category will make its task easier
- But don't throw away semantics all together
- Enough evidence for both semantic and visual in lot of studies – but take everything with a grain of salt
- Find some combined hierarchy?
  - Animals, Vehicles (semantic)

Thank You!