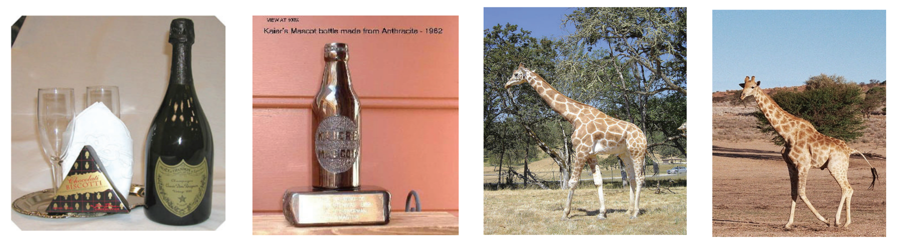
Shape Discovery from Unlabeled Image Collections

Yong Jae Lee and Kristen Grauman
University of Texas at Austin

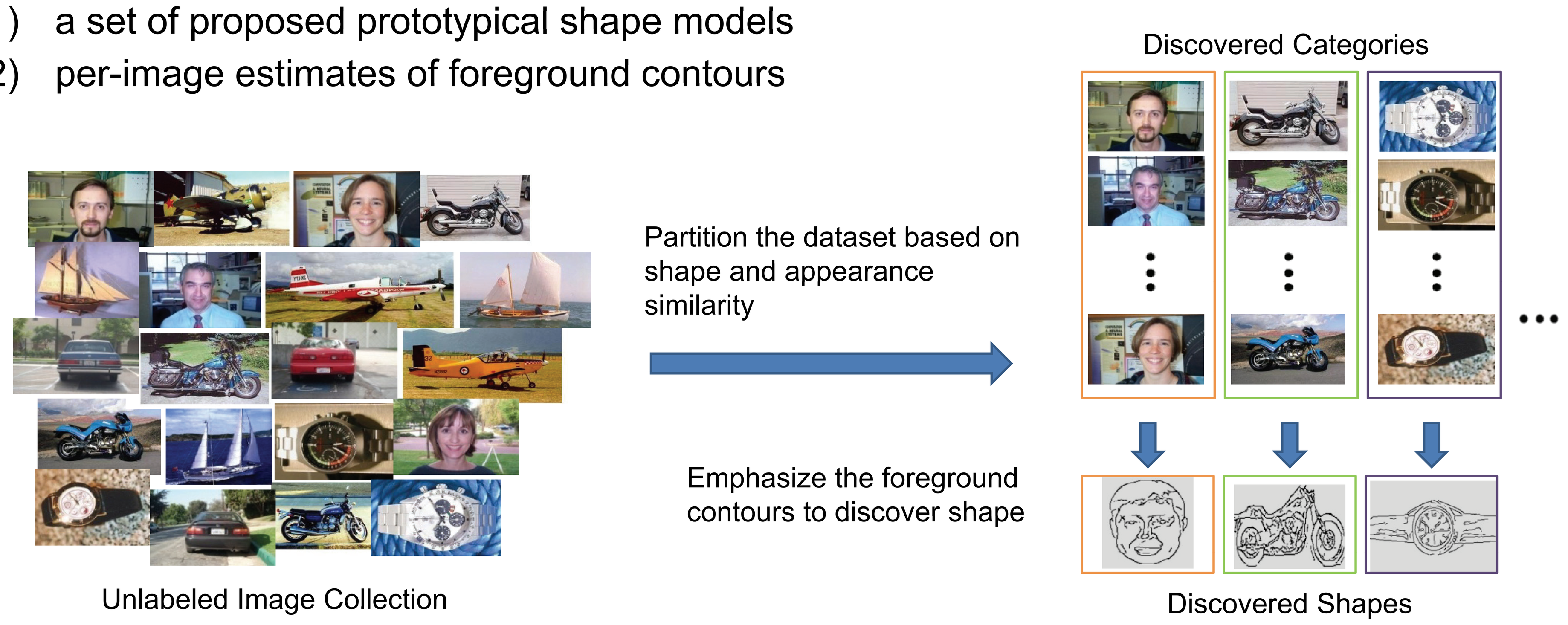
Problem

Can we discover common object shapes within unlabeled multi-category collections of images?



Main idea

We propose a method that analyzes a collection of unlabeled images and returns:
1) a set of proposed prototypical shape models
2) per-image estimates of foreground contours

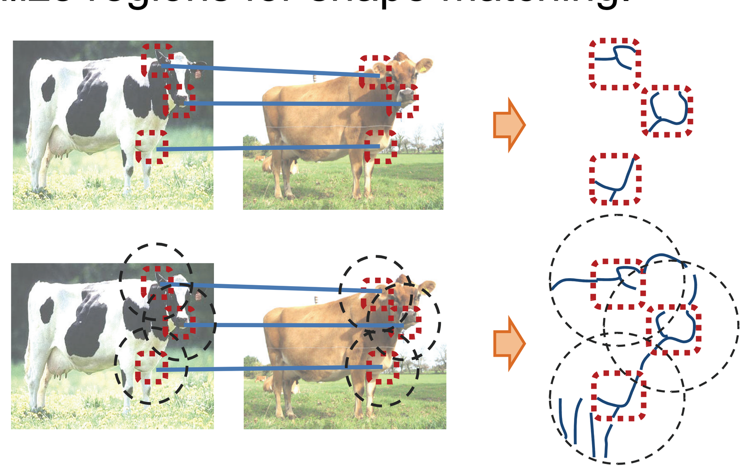


The goal is to identify which foreground contours in each image can form high quality shape clusters, and use any intra-cluster agreement to discover the underlying prototypical shapes.

Algorithm Overview

Anchoring Edge Fragments to Local Patches

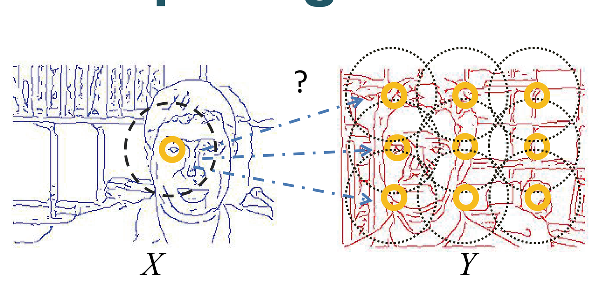
Since edge features often lack distinctiveness, we use patch appearance matches (e.g., SIFT) to initialize regions for shape matching.



Even with accurate patch matches, there's a limit to how much shape information can be captured.

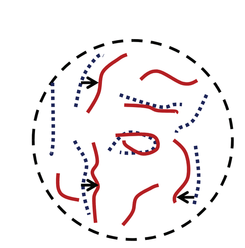
By anchoring edge fragments to patch features, we can produce more reliable matches *and* discover the object's shape.

Computing Feature Matches with Anchored Edge Features



Compare each feature in X to all features in Y

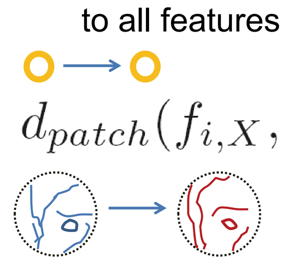
Align X's edgemap onto Y's edgemap based on the best matching feature



Each fragment in X is fitted to the nearest best matching fragment in Y

$d_{patch}(f_{i,X}, f_{j,Y})$

$d_{shape}(f_{i,X}, f_{j,Y})$



symmetric chamfer distance

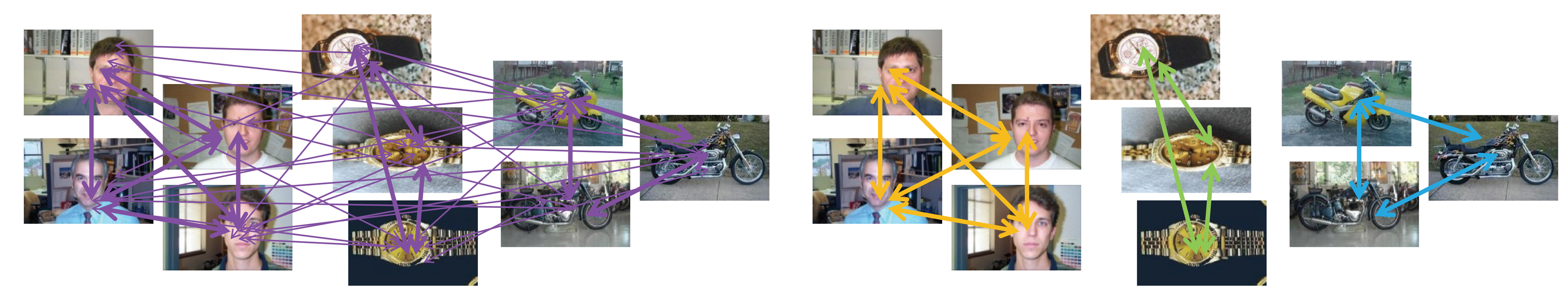
* Dotted circles represent the Gaussian weighting on the edge fragments with the highest weight at feature center.

Grouping Cluttered Images with Similar Shapes

The overall directed patch and shape distance from image X to image Y is the average over the component feature distances between each $f_{i,X}$ in X and its best matching feature $f_{j(i)^*,Y}$ in Y:

$$D_{patch}(X,Y) = \frac{1}{|X|}\sum_{i=1}^{|X|} d_{patch}(f_{i,X}, f_{j(i)^*,Y}), \qquad D_{shape}(X,Y) = \frac{1}{|X|}\sum_{i=1}^{|X|} d_{shape}(f_{i,X}, f_{j(i)^*,Y}).$$

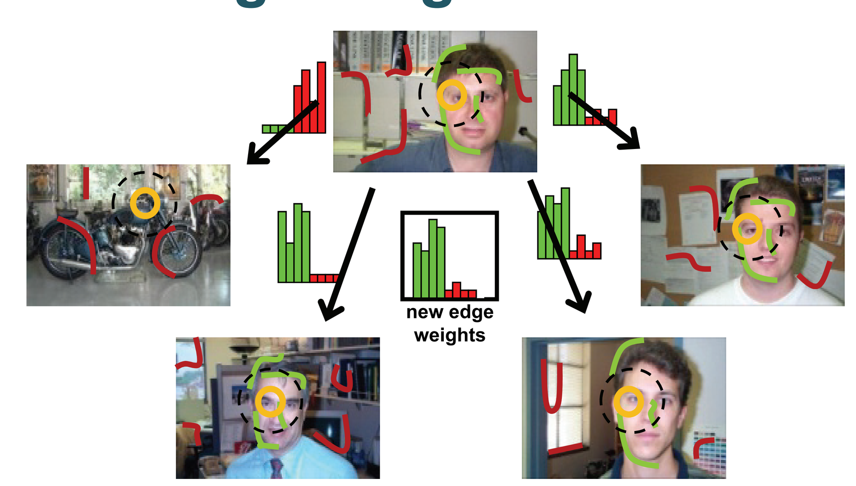
We obtain a symmetric cost via the sum: $D'(X,Y) = D(X,Y) + D(Y,X) = D'(Y,X)$.



Given the distances between all pairs of the N unlabeled images, we perform spectral clustering [Ng et al., NIPS 2001] with an N x N affinity matrix A that reflects the strongest shape and appearance correspondences found between images:

$$A_{r,s} = \exp\left(-\frac{1}{\sigma^2} D'_{patch}(X_r, X_s) * D'_{shape}(X_r, X_s)\right), \quad \text{for all } r, s = 1, \ldots, N.$$

Inferring Foreground Contours

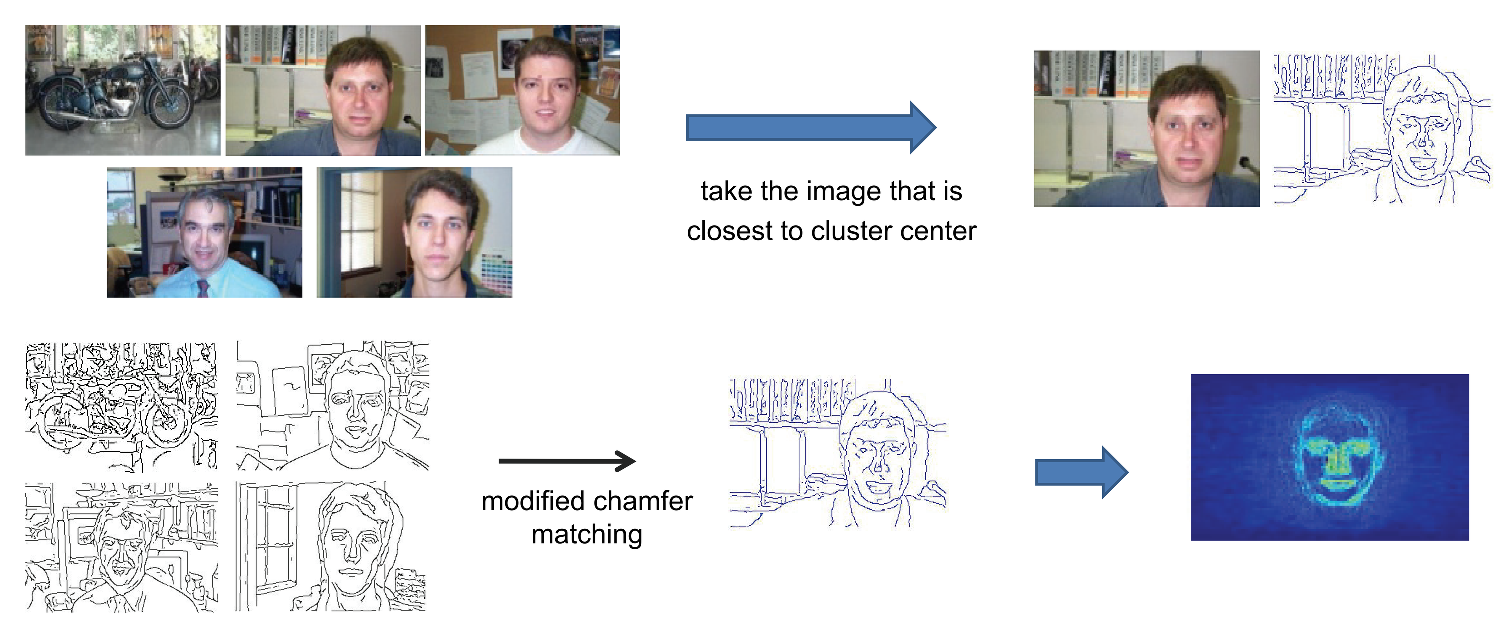


Next, we analyze the pattern of the intra-cluster edge matches. We identify the fragments with the most consistent correspondences and increase their weights.

Prototypical Shape Formation

Not all images in a cluster will contain the same object or agree in terms of shape.

→ Create a simple vote space with the computed edge weights such that the common shape can be reinforced in the output while parts that agree less can be discarded.

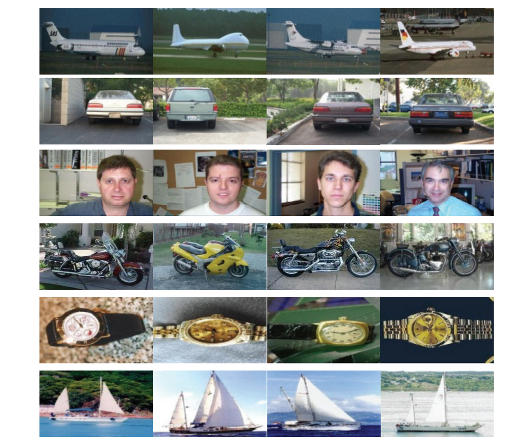


Algorithm Flow Chart: Recap

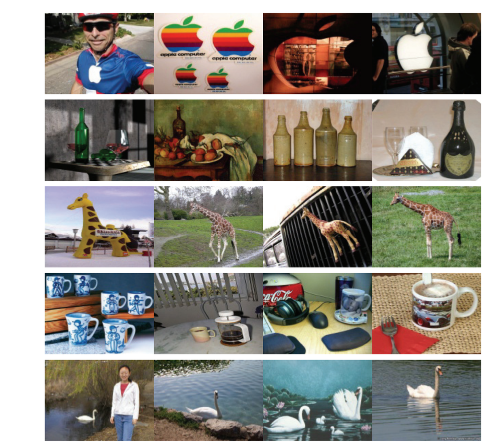
Compute Feature Matches with Anchored Edge Features → Cluster Images with Similar Shapes given Feature Distances → Infer Foreground Contours by Analyzing Intra-Cluster Match Patterns → Vote to Form Prototypical Shapes with Foreground Contours

Results

We present results to analyze our method's unsupervised category and shape discovery using the Caltech, ETHZ shape, and LabelMe datasets.



Caltech Dataset
[Fei-Fei et al. 2004]
- 6 categories
- 100 images per category



ETHZ Shape Dataset
[Ferrari et al. 2007]
- 5 categories
- 32-87 images per category (total 255)
- Bounding Box and Expanded regions

Unsupervised Category Discovery

CT-Categories	Our Method	Patch-only	Kim et al.	G & D	L & G
A,C,F,M	98.03 ± 0.66	87.37	98.55	86.00	88.82
A,C,F,M,W	96.92 ± 0.63	83.78	97.30	N/A	N/A
A,C,F,M,W,K	96.15 ± 0.52	83.53	95.42	N/A	N/A

ETHZ-Categories	Our Method	Patch-only
A,B,G,M,S (bbox)	95.85	78.89
A,B,G,M,S (expanded)	76.47	61.25

Table 1. Category discovery accuracies measured by mean purity for the categories of the Caltech [A: Airplanes, C: Cars, M: Motorbikes, W: Watches, K: Ketches] (left) and ETHZ [A: Applelogos, B: Bottles, G: Giraffes, M: Mugs, S: Swans] (right) datasets.

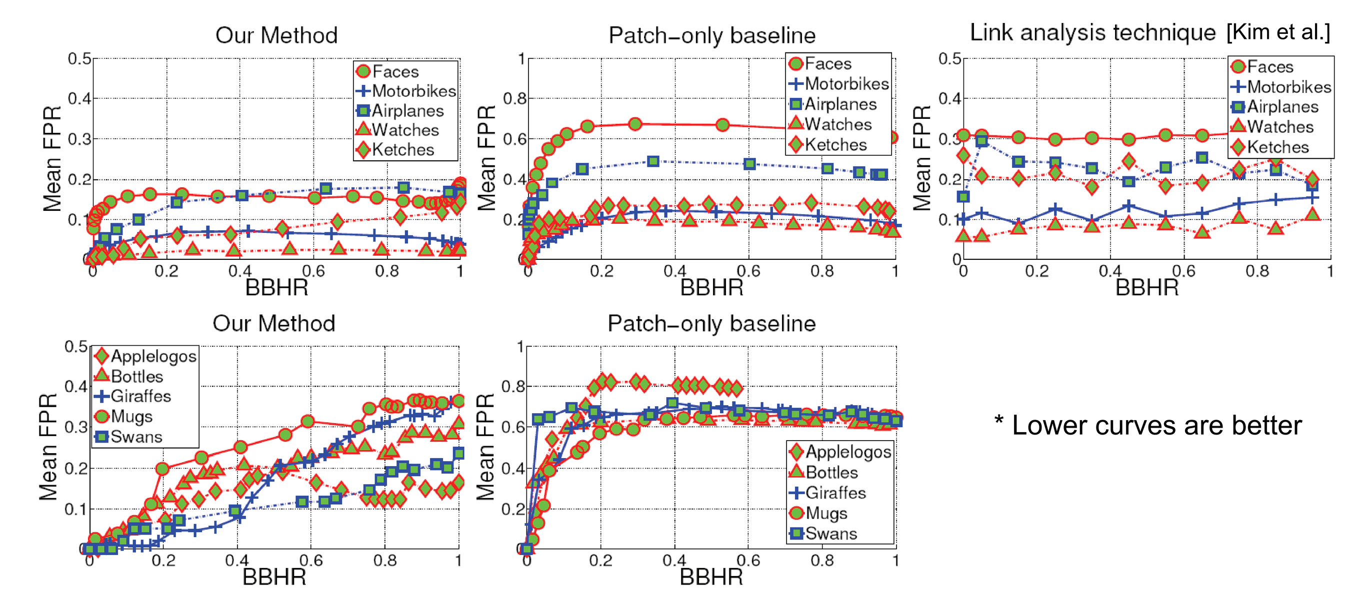
- Purity: measures the extent to which a cluster contains images of a single dominant class.

* Baseline methods (appearance-based):
Patch-only: we use the same steps as our method, but use only patch features (without shape information).
* Kim et al., CVPR 2008., G & D: Grauman and Darrell, CVPR 2006., L & G: Lee and Grauman, BMVC 2008.

Foreground Shape Discovery: Localization

Caltech Images

ETHZ Images (expanded)



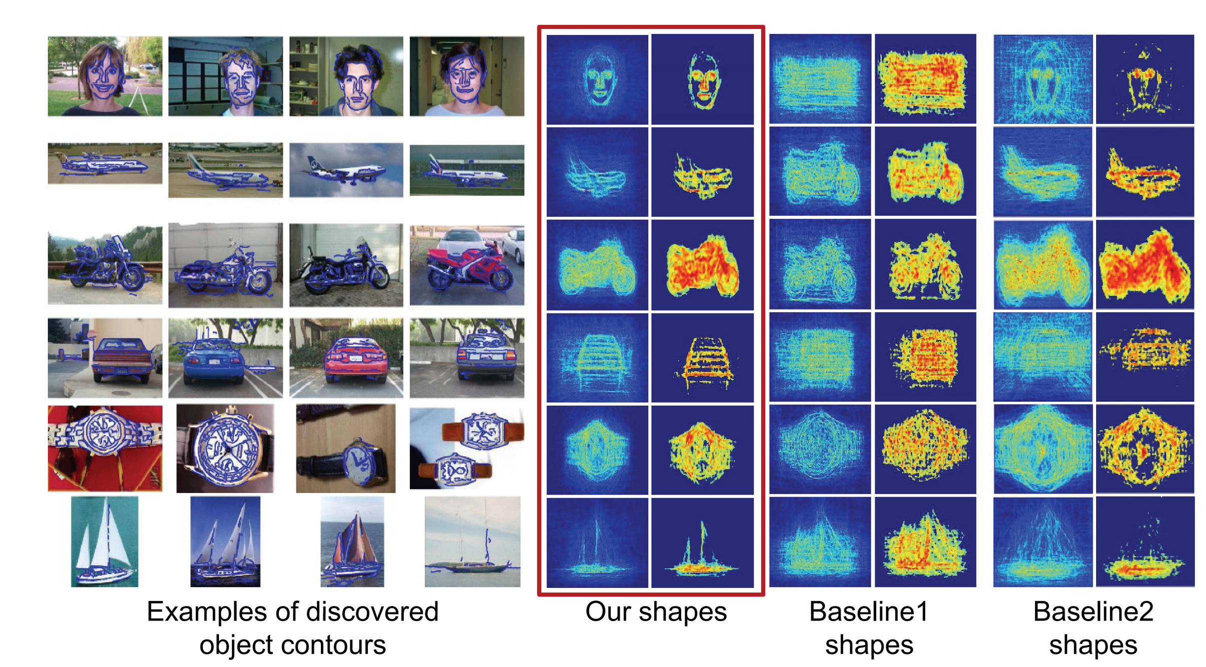
* Lower curves are better

- Bounding Box Hit Rate (BBHR): measures the % of images in the dataset that have at least $h = 5$ foreground features selected, as a function of the selection threshold applied to the feature weights.
- False Positive Rate (FPR): measures the average % of selected features falling outside of the bounding box.

Our method outperforms a state-of-the-art method and achieves very good localization rates.

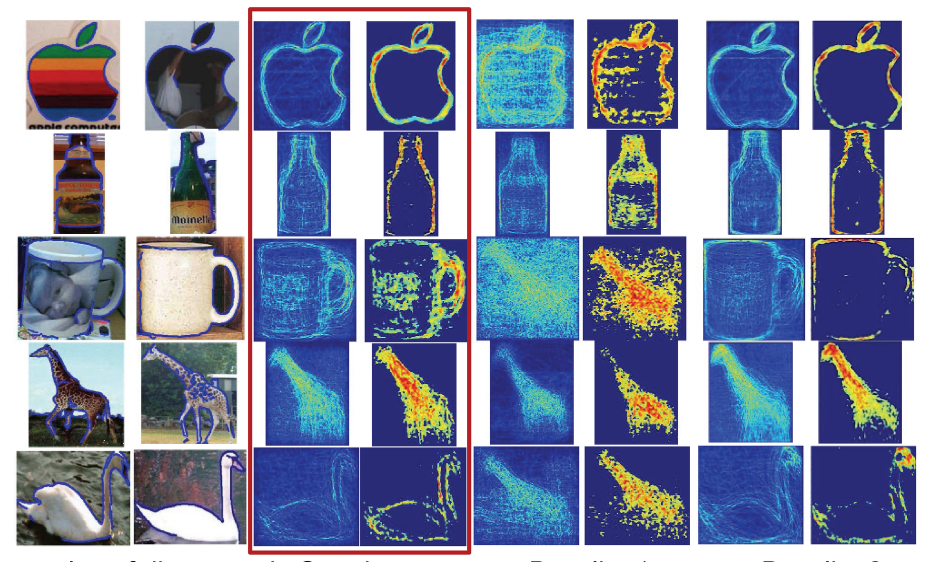
Foreground Shape Discovery: Prototypical Shape

Caltech Images



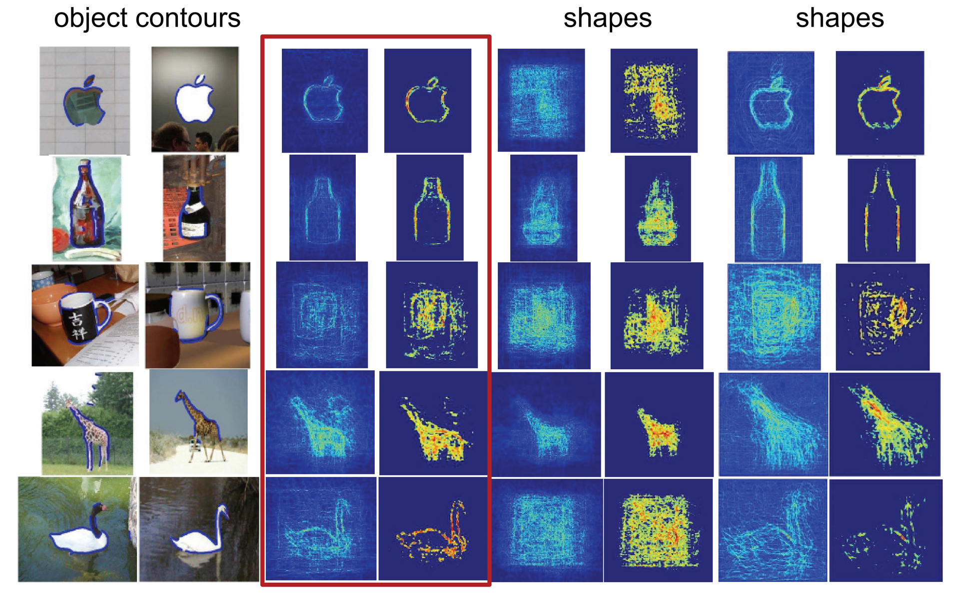
Foreground Shape Discovery: Prototypical Shape

ETHZ Images (bounding box)



Examples of discovered object contours | Our shapes | Baseline1 shapes | Baseline2 shapes

ETHZ Images (expanded)

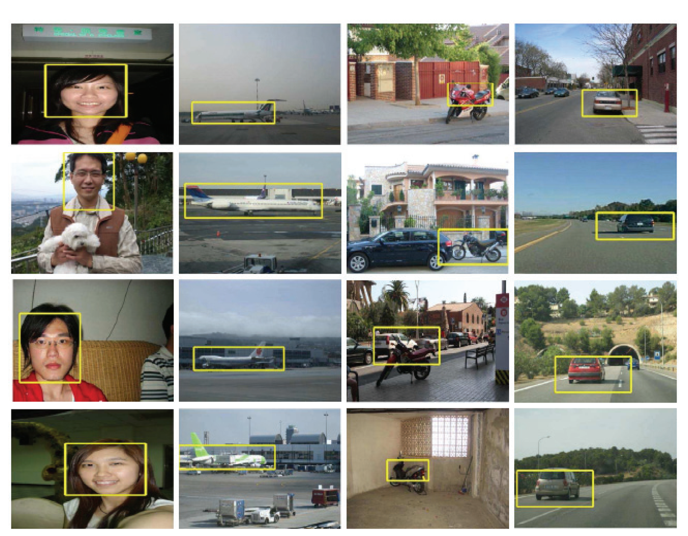


Examples of discovered object contours | Our shapes | Baseline1 shapes | Baseline2 shapes

- Baseline1: a shape-only method in which all edges are weighted equally when computing image similarities. We cluster the images with an affinity matrix computed from the symmetric chamfer distance between their edgemaps.
- Baseline2: we manually partition the images into "perfect" clusters, and simply average the aligned edge images with confidence weights given by Pb detector [Martin et al., TPAMI 2004].

Generalization to Detection in Novel Images

- Detection task on LabelMe dataset
- Match prototypical shapes to the test images using modified chamfer matching
- $a_0 = (BB_{gt} \cap BB_d)/(BB_{gt} \cup BB_d)$
- Average a_0 : [F: 0.47, A: 0.43, M: 0.38, C: 0.31]
- Chance Detection: [F: 0.03, A: 0.02, M: 0.03, C: 0.02]



Faces | Airplanes | Motorbikes | Cars Rear

Examples of correct detections

Faces | Airplanes | Motorbikes | Cars Rear

Examples of incorrect detections

Conclusions

- We have proposed the first method to discover common object shapes within unlabeled multi-category collection of images.
- We have shown the strength of our patch-anchored shape matching by comparing against baseline methods that use each feature in isolation, as well as against previous appearance-only unsupervised learners.