Learning the Easy Things First: Self-Paced Visual Category Discovery

Yong Jae Lee and Kristen Grauman
University of Texas at Austin

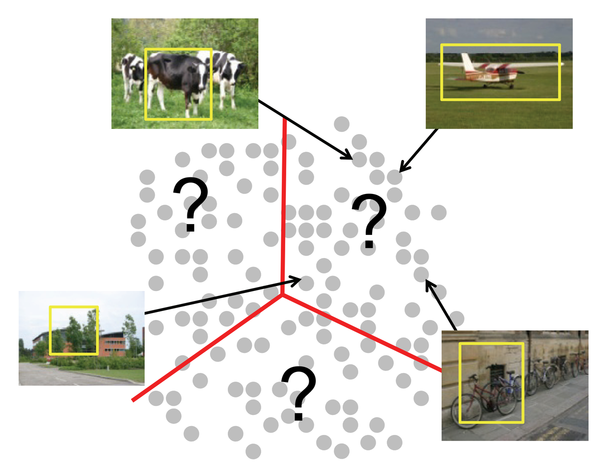
Problem

Existing methods treat discovery as a one-pass "batch" procedure: input is set of unlabeled images, output is *k* discovered categories. [Sivic et al., 2005, Russell et al., 2006, Kim et al., 2008, Lee and Grauman, 2010, …]

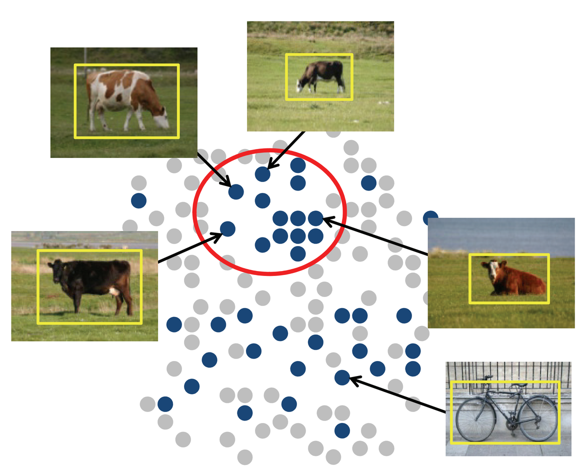
Paying equal attention to all instances makes the grouping sensitive to outliers and denies the possibility of exploiting inter-object context cues during discovery.

Our idea

Focus on the easier instances first, and gradually discover new models of increasing complexity.



Traditional Batch *k*-way



Single Easiest (Ours)

What makes some image regions easier than others?

- "Objectness": regions spanning a single object exhibit more regularity in appearance, making them more apparent to group.
- "Context-Awareness": regions surrounded by familiar objects have stronger context that can make a grouping more apparent.

Why should it matter in what order objects are discovered?

- As the system gradually accumulates category models, each new (more difficult) discovery benefits from the object-level context provided by earlier discoveries.

Algorithm overview

Initialize with stuff categories → Identify easy instances in undiscovered pool → Discover next prominent category → Expand discovered category knowledge

t++; Iterate until only hard instances remain

Exemplar-based category models

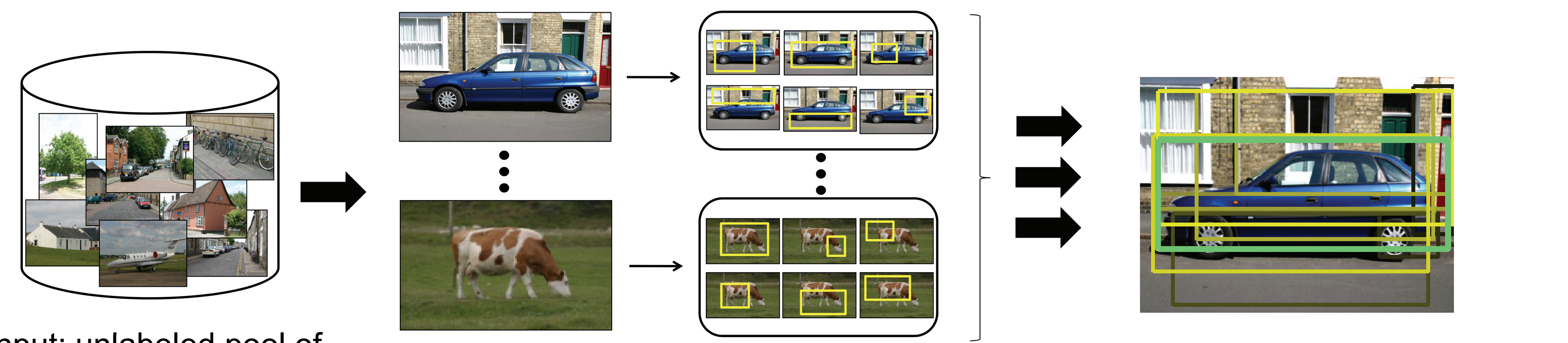
- We use an exemplar based model to represent familiar classes:

$$P(r|c_j) \propto \frac{1}{T}\sum_{m=1}^{T}\frac{1}{|c_j|}\sum_{l\in c_j} K_m(r,l)$$

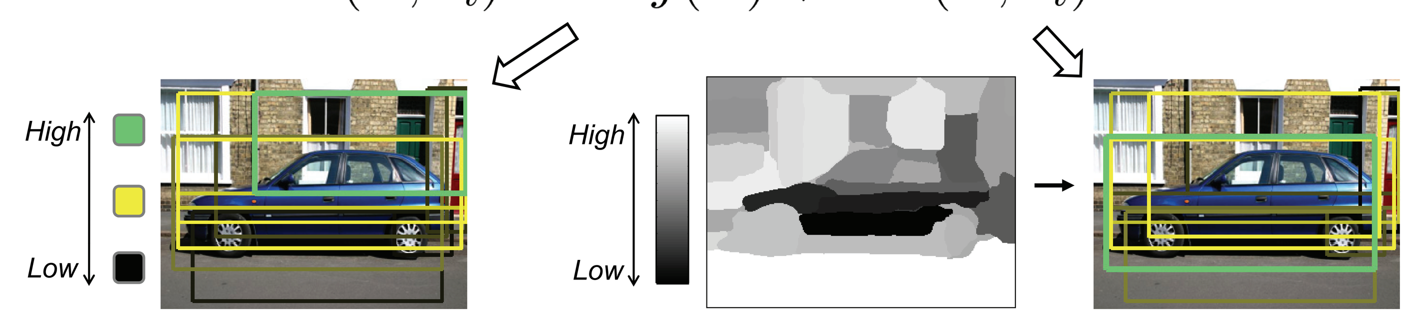
T: feature types K: chi-square kernel
r: window/region c_j: familiar class j

- The likelihood values are used to capture region familiarity.

Identifying easy objects



$$\text{Easiness}(w,\mathcal{C}_t) = Obj(w) + CA(w,\mathcal{C}_t)$$

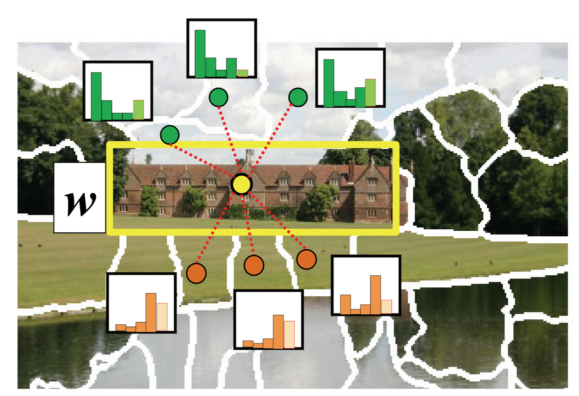


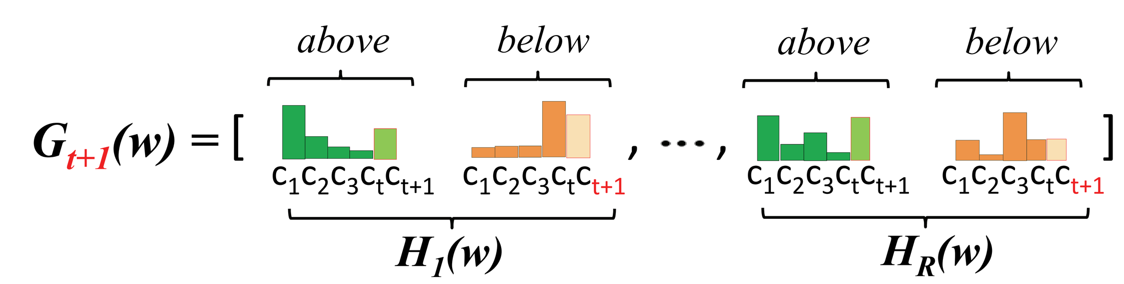
$$CA(w,\mathcal{C}_t) = \sum_{j=1}^{R} \mathrm{w}_j \big(\max_{c_j\in\mathcal{C}_t} P(s|c_j)\big),$$

where $\mathrm{w}_j = R - j + 1$ gives the nearest superpixels most influence.

- $\mathcal{C}_t$: set of familiar categories; w : candidate object window; t : iteration of discovery.
- We select the easiest instances with a threshold computed from the data $\theta_t = 2\sigma - 0.1t$ where σ denotes the standard deviation of all easiness scores.

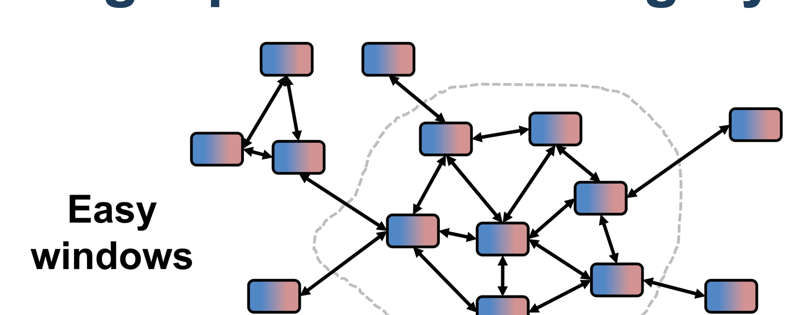
Object-graphs for familiar context





- Model the topology of category predictions with the *Object-Graph* descriptor [Lee and Grauman, 2010].

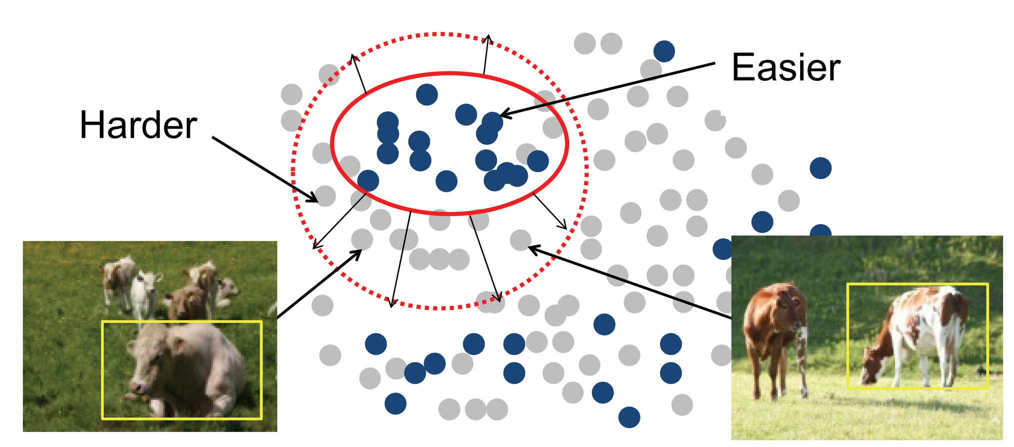
Single prominent category discovery



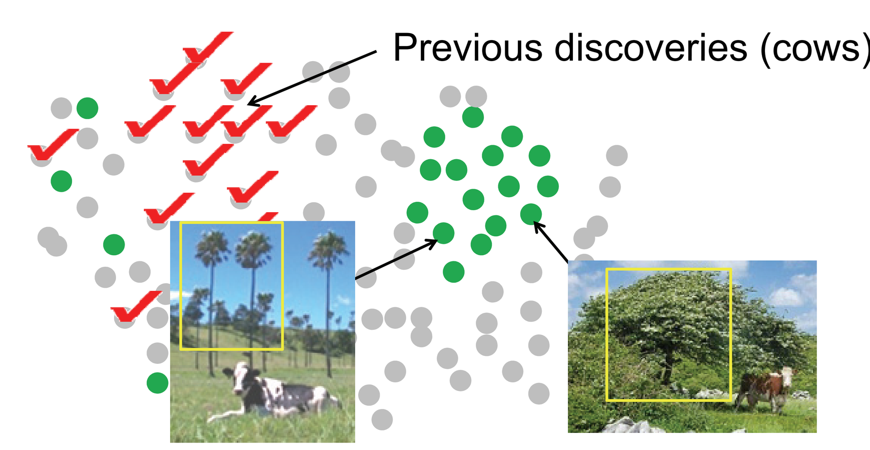
$$K(w_i,w_j) = K_{app}(w_i,w_j) + K_{obj-graph}(w_i,w_j)$$

- Generate candidate groups with agglomerative clustering and select "best" cluster with highest silhouette coefficient. Refine instances with Single-Cluster Spectral Graph Partitioning [Olsen et al., 2005].

Discovered category knowledge expansion



Intra-category model expansion

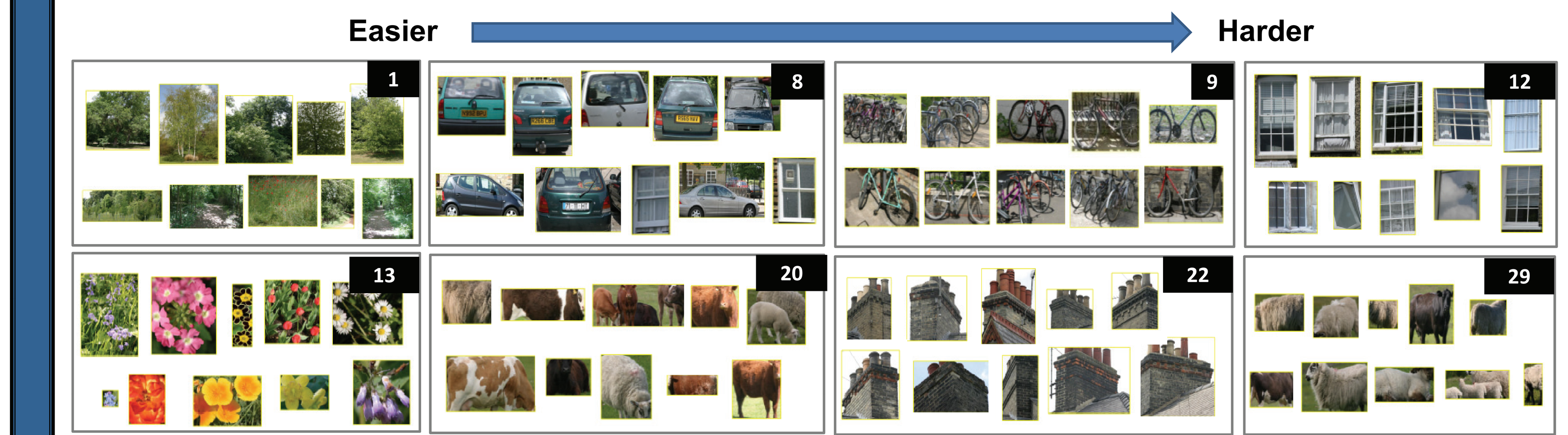


Object-level context expansion

Results

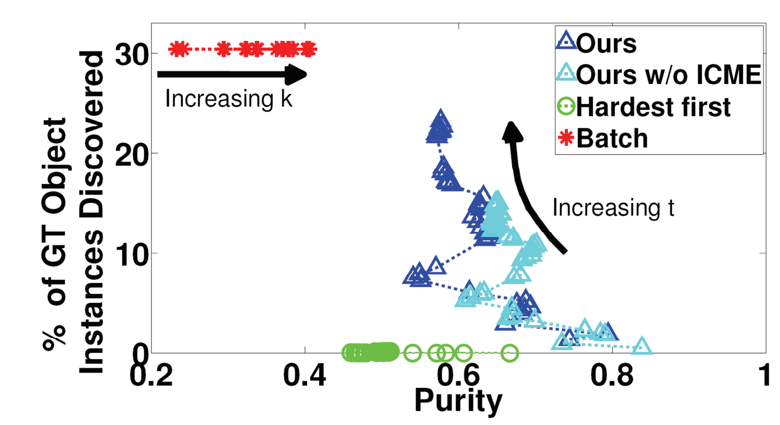
Evaluate on MSRC-v0: 21 categories, 3457 images. Use [Alexe et al., 2010] to generate 50 windows per image, Berkeley seg engine to generate "stuff" segments. Represent regions with SIFT bow, color histograms, pHOG.

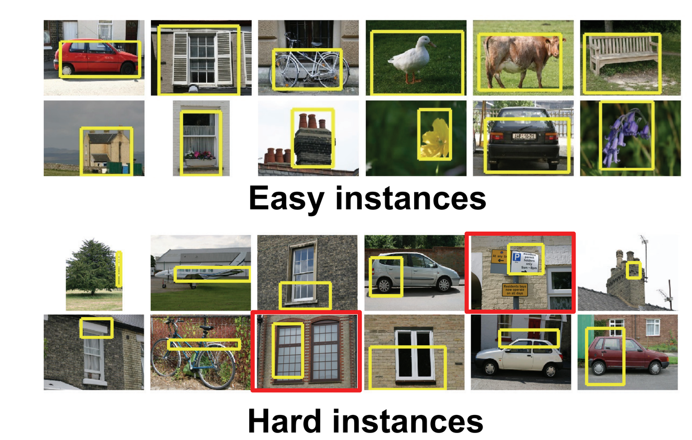
Object discovery accuracy



Example unsupervised discoveries

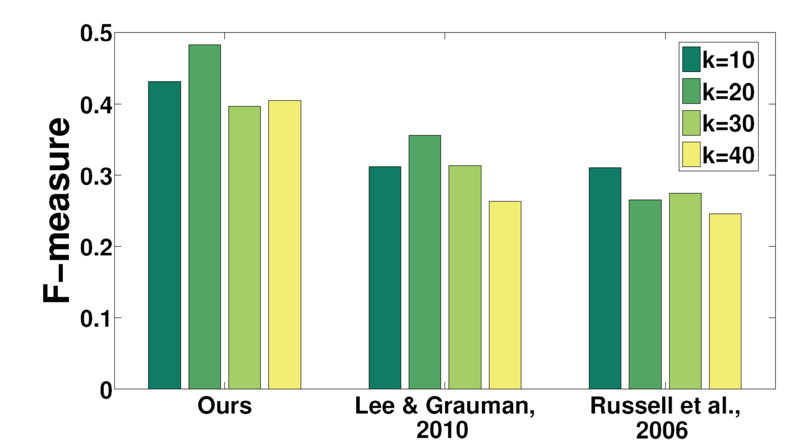
Our method selectively discovers categories in order of predicted difficulty.





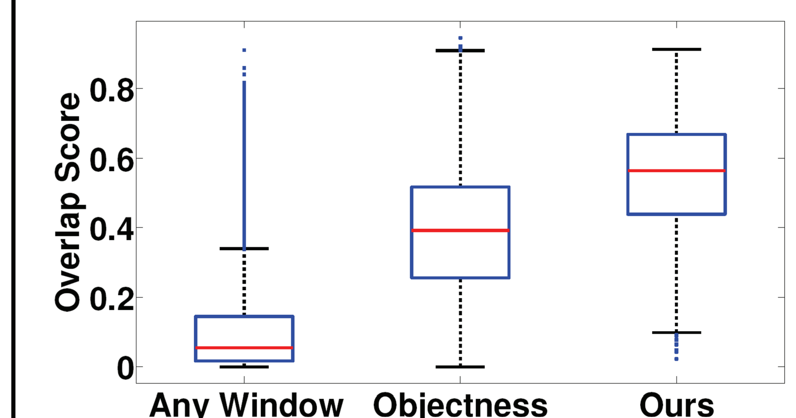
Ours produces more accurate clusters, while selectively ignoring instances that cannot be grouped well.

Comparison w/ state-of-the-art



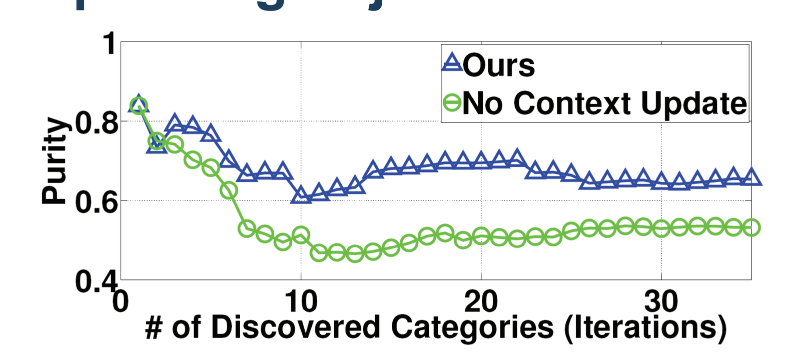
More accurate summarization than the state-of-the-art.

Segmentation accuracy



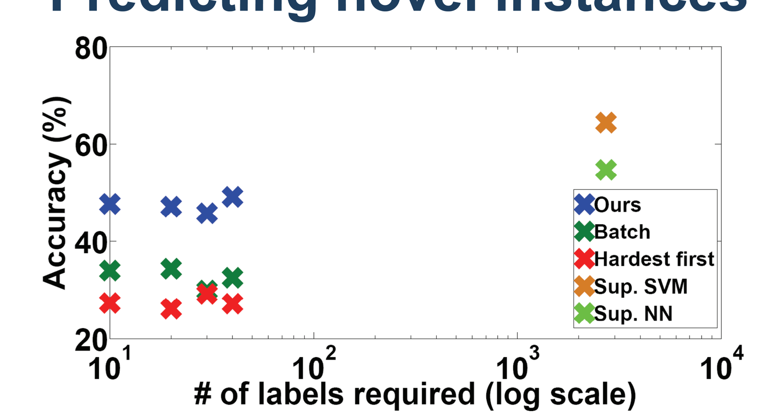
Ours produces strongest segmentations, showing impact of easiness scoring and self-paced grouping.

Expanding object-level context



Revising object-level context w/ each new discovery leads to better clustering accuracy.

Predicting novel instances



Ours yields good prediction with minimal human effort.

Conclusions

- Self-paced discovery framework that progressively accumulates object models from unlabeled data.
- Clear advantages over traditional batch approaches and representative state-of-the-art techniques.