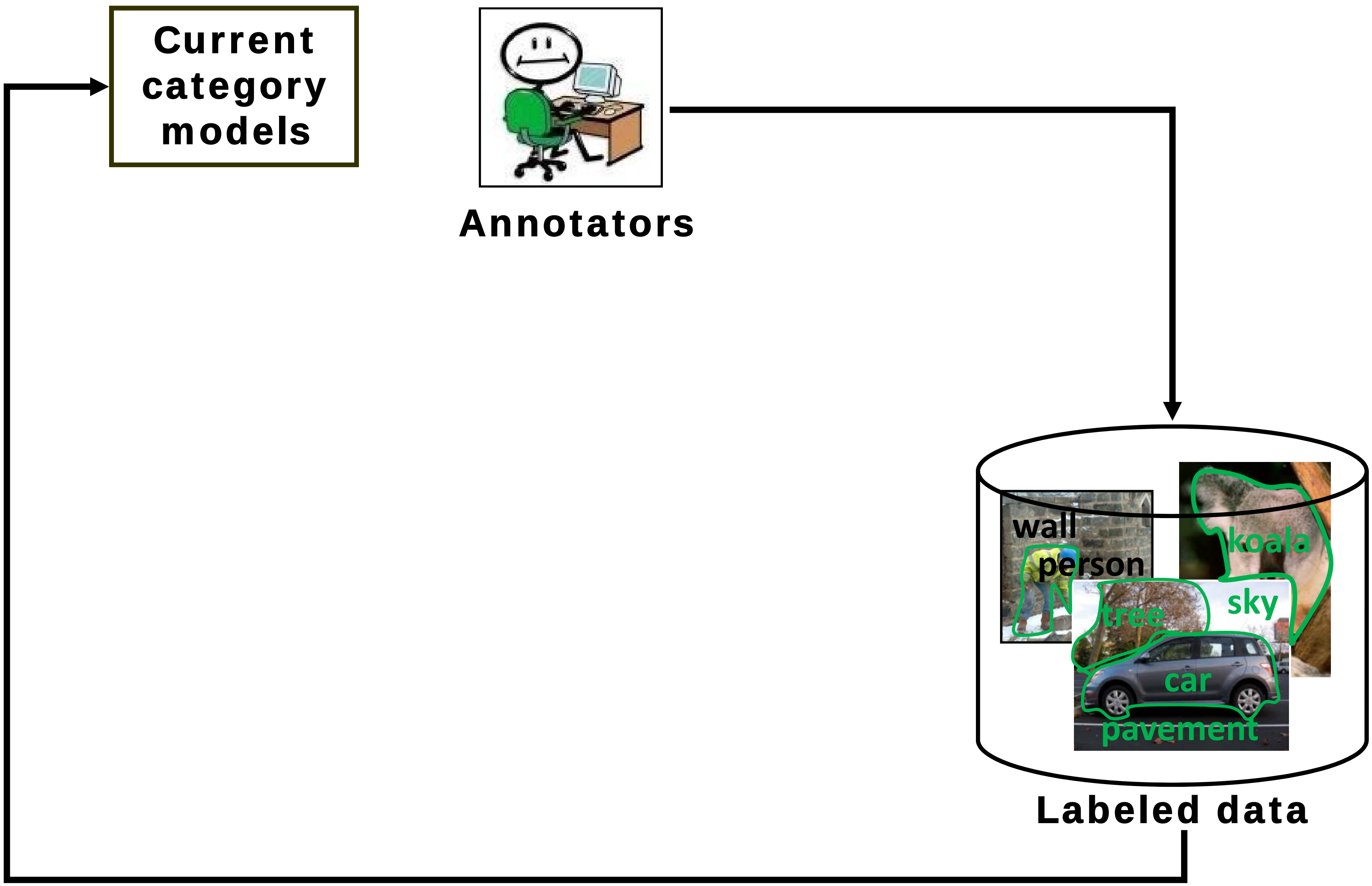
Cost-Sensitive Active Visual Category Learning

Sudheendra Vijayanarasimhan
Kristen Grauman

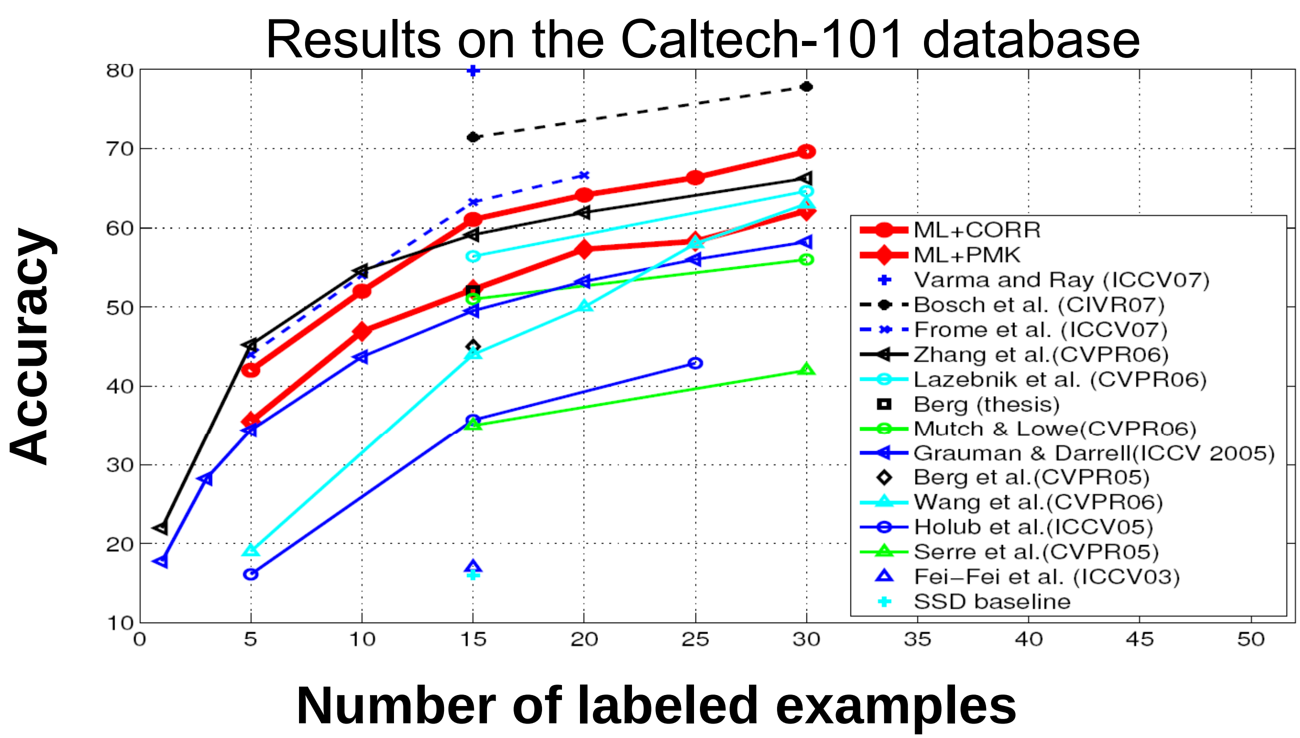
University of Texas at Austin

Learning visual categories



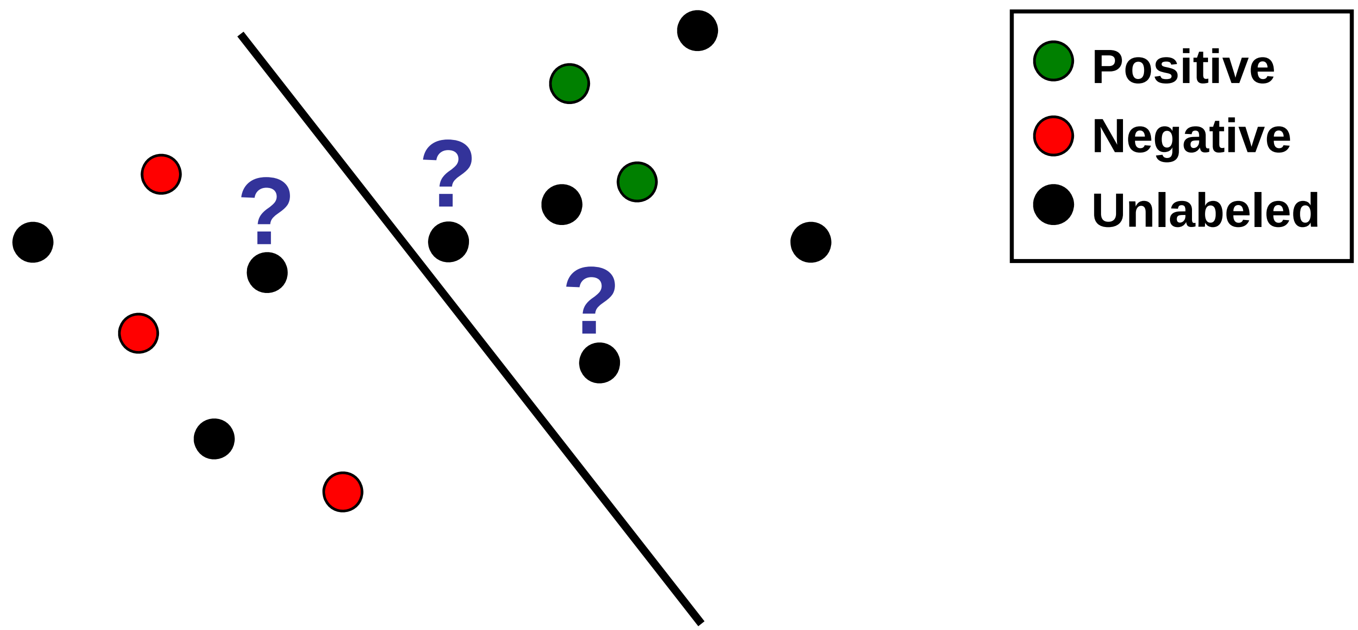
More supervision → better learning?

Access to more labeled examples (and “strongly” labeled examples) often leads to more accurate recognition results.



Active learning

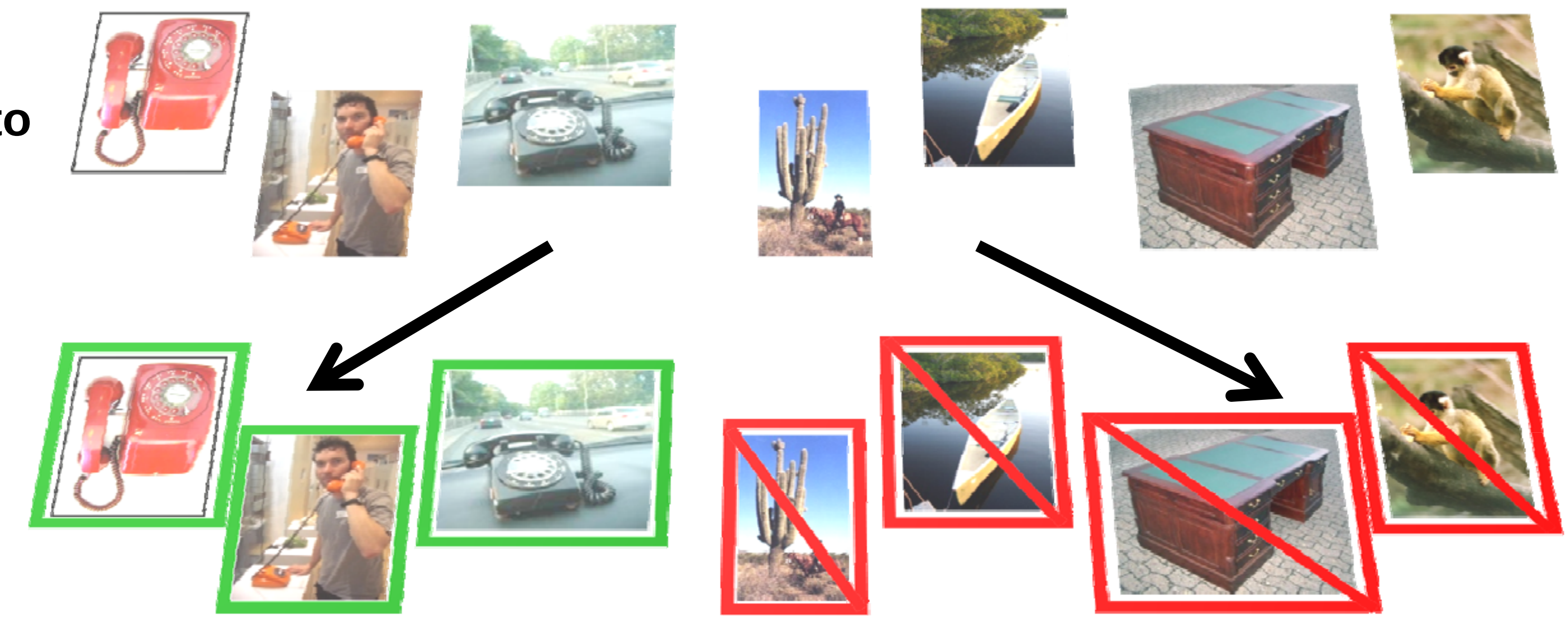
- **Traditional active learning** reduces supervision by obtaining labels for the most informative or uncertain examples first.

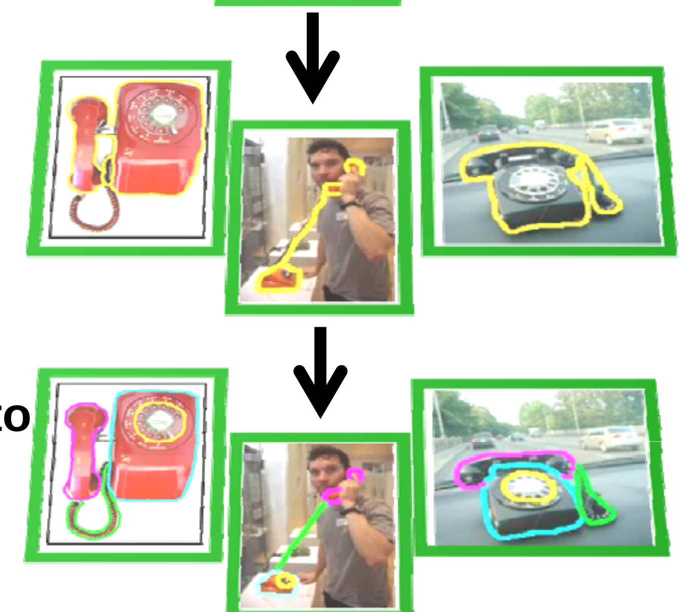


[Mackay 1992, Freund et al. 1997, Tong & Koller 2001, Lindenbaum et al. 2004, Kapoor et al. 2007, Collins et al. 2008, Holub & Perona 2008,...]

Problem

Less expensive to obtain





More expensive to obtain

- Multi-label examples
- Multiple levels of annotation are possible
- Variable cost depending on level and example

Our approach: Cost-sensitive "multi-level" active learning

Main idea:

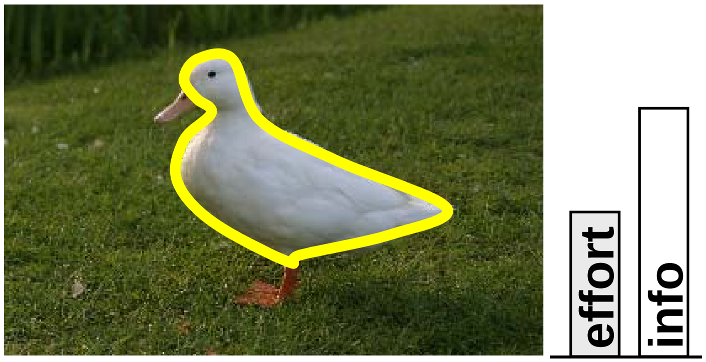
Compute decision-theoretic active selection criterion that weighs both:

- which *example* to annotate, and
- what *kind* of annotation to request for it

as compared to

- the predicted *effort* the request would require

Our approach: Cost-sensitive "multi-level" active learning



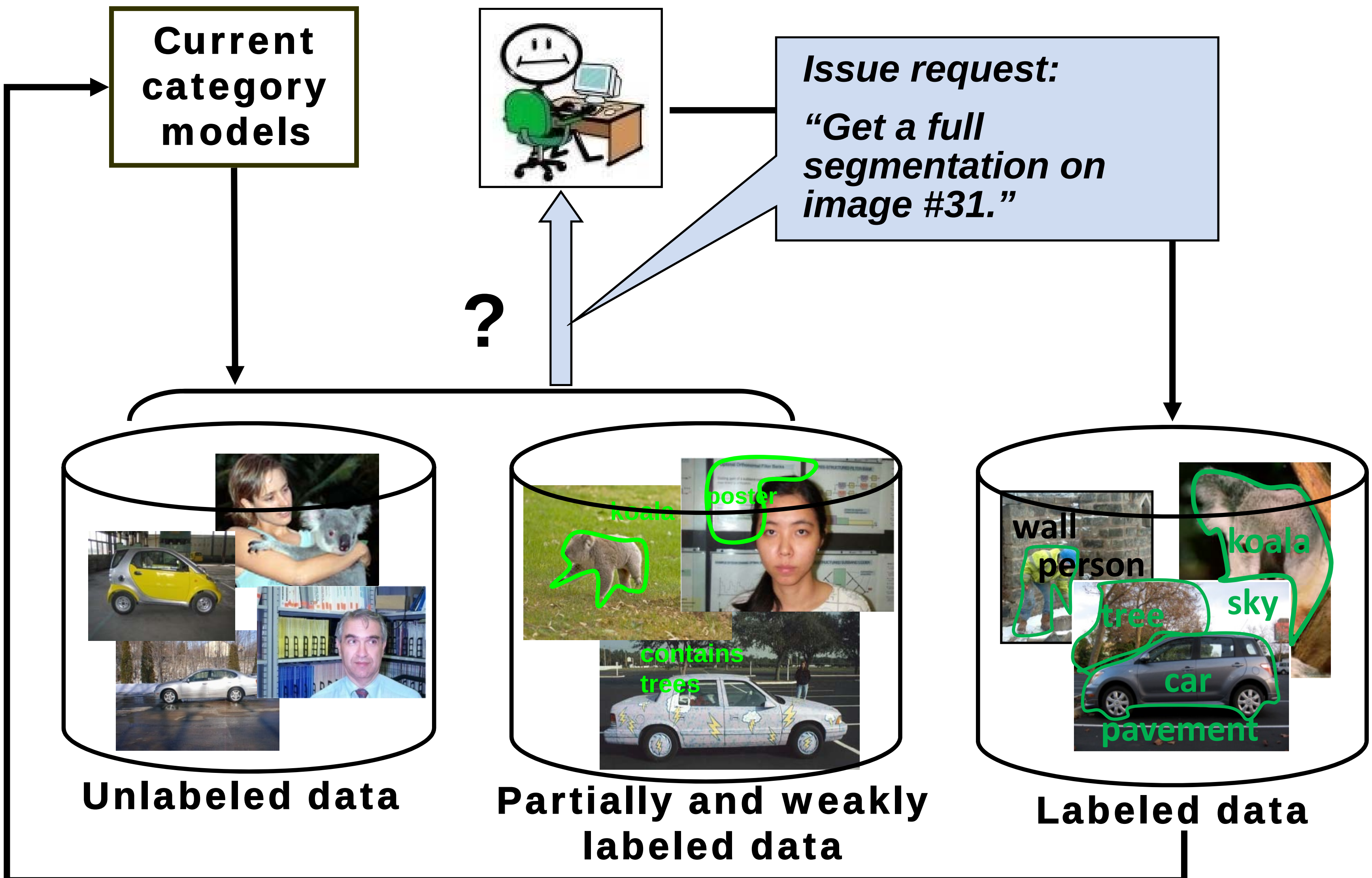
Most regions are understood, but this region is unclear.

This looks expensive to annotate, and it does not seem informative.

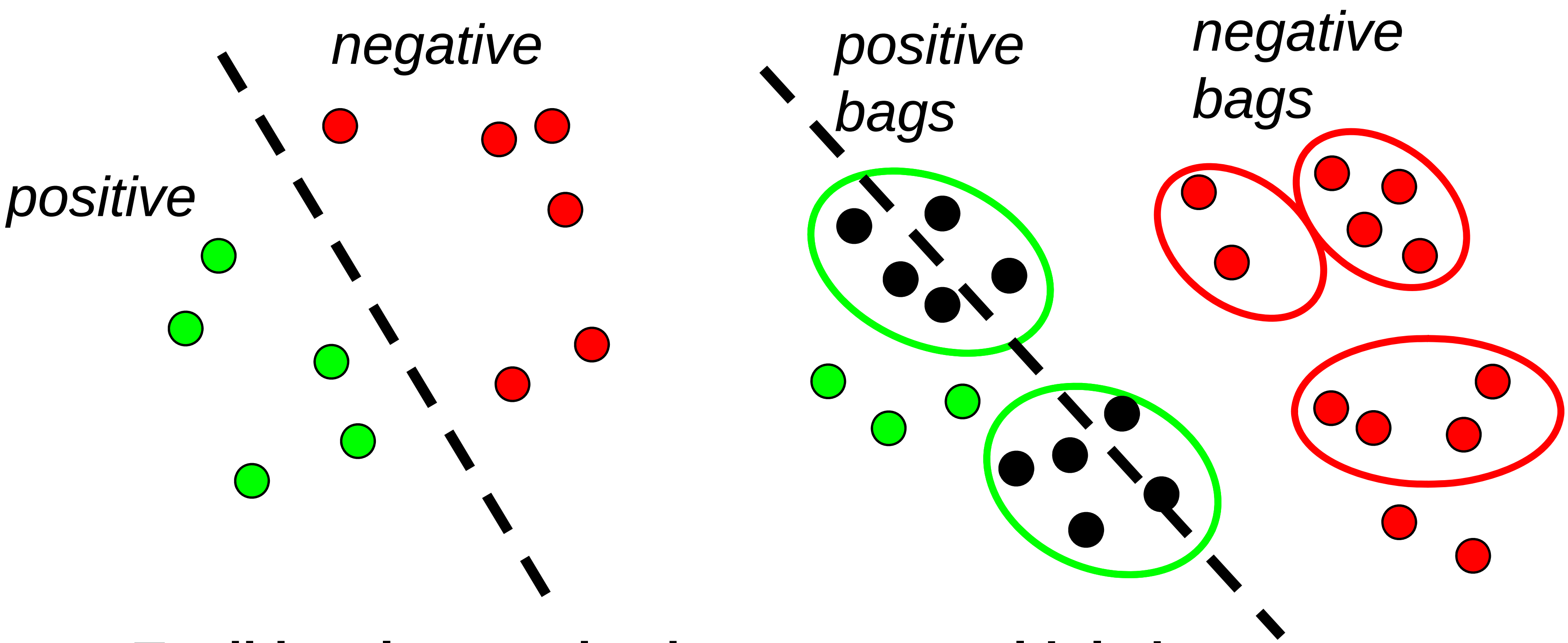
This looks expensive to annotate, but it seems very informative.

This looks easy to annotate, but its content is already understood.

Our approach: Cost-sensitive "multi-level" active learning



Multiple-instance learning (MIL)

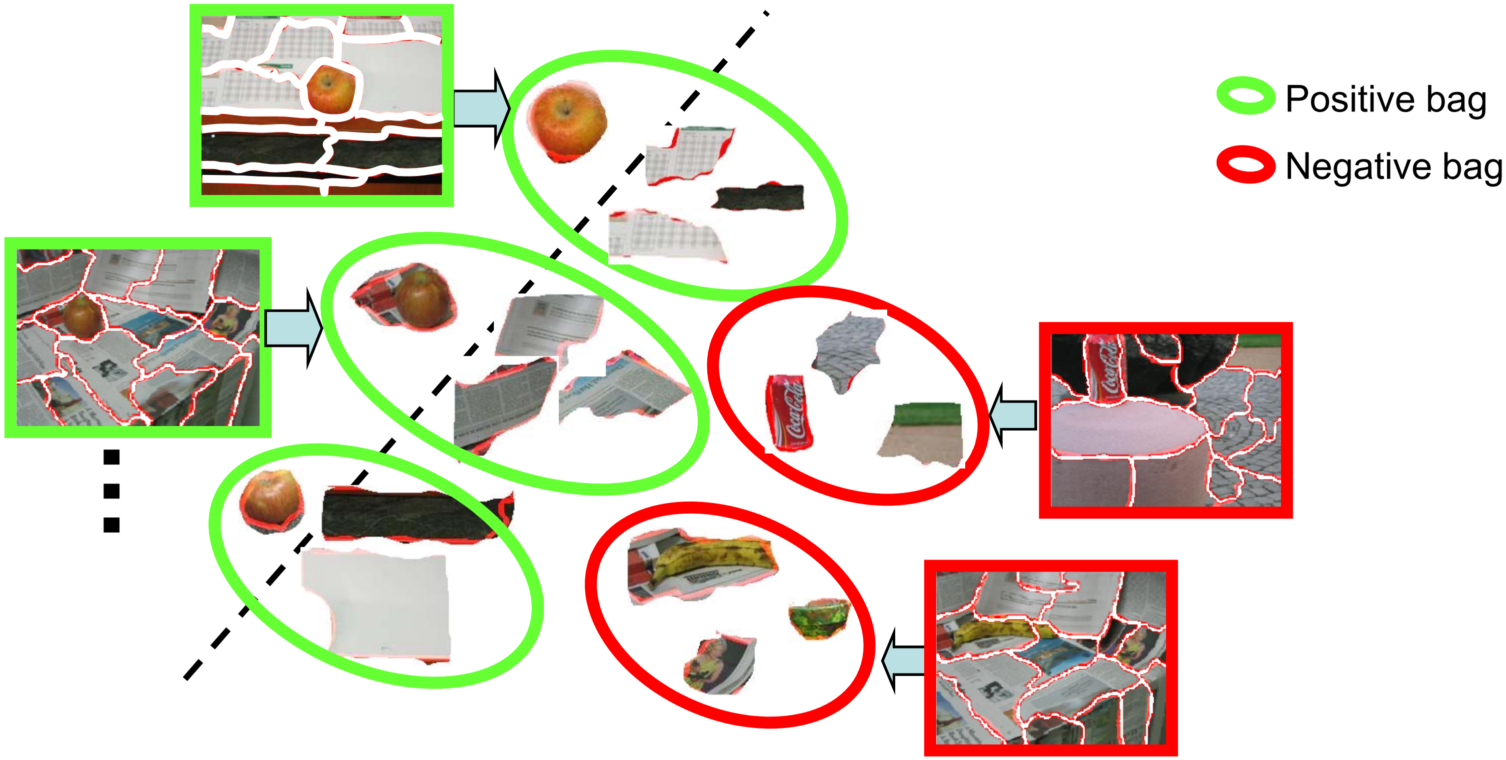


Traditional supervised learning

Multiple-instance learning

[Dietterich et al. 1997]

MIL for visual category learning

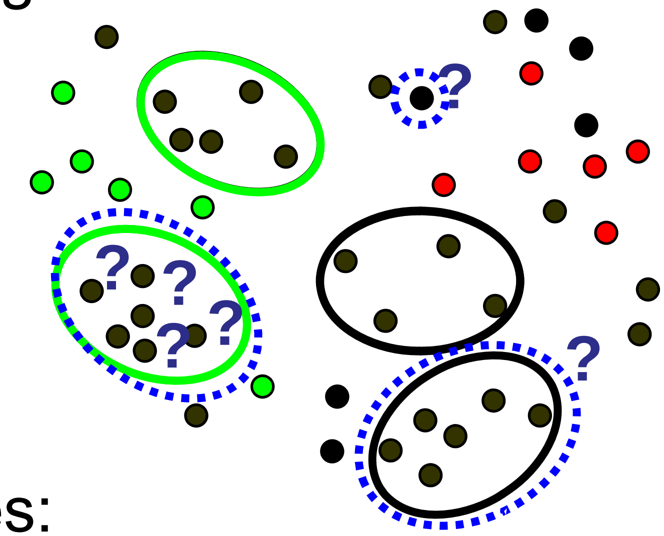


- **Positive instance**: Segment belonging to class
- **Negative instance**: Segment not in class
- **Positive bag**: Image containing class
- **Negative bag**: Image not containing class

[Maron & Ratan, Yang & Lozano-Perez, Andrews et al.,…]

Multi-level active queries

Predict which query will be most informative, given the cost of obtaining the annotation.



Possible queries:

1. Label an unlabeled instance

2. Label an unlabeled bag

3. Label all instances in a bag

Decision-theoretic multi-level criterion

We measure the value of information (VOI) for choosing a potential query $\mathbf{z}$ by the expected reduction in total cost:

$$VOI(\mathbf{z}) = T\underbrace{(\mathcal{X}_L, \mathcal{X}_U)}_{\text{Current dataset}} - T\left(\underbrace{\mathcal{X}_L \cup \mathbf{z}^{(t)}, \mathcal{X}_U \setminus \mathbf{z}}_{\text{Dataset after } \mathbf{z} \text{ is labeled with true label } t}\right),$$

$$= Risk(\mathcal{X}_L) + Risk(\mathcal{X}_U)$$

$$- \left(Risk(\mathcal{X}_L \cup \mathbf{z}^{(t)}) + Risk(\mathcal{X}_U \setminus \mathbf{z})\right)$$

$$-\mathcal{C}(\mathbf{z})$$

= **Risk under the current classifier** - **Risk after adding *z* to the labeled set** - **Cost of obtaining annotation for z**

Decision-theoretic multi-level criterion

=	**Risk under the current classifier**	-	**Risk after adding *z* to the labeled set**	-	**Cost of obtaining annotation for z**

Decision-theoretic multi-level criterion

=	**Risk under the current classifier**	–	**Risk after adding *z* to the labeled set**	–	**Cost of obtaining annotation for z**

To estimate the risk of incorporating **z** into labeled set before knowing its true label *t*, compute expected value:

$$Risk\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + Risk\left(\mathcal{X}_U \setminus \mathbf{z}\right)$$

$$\mathbb{E} = \sum_{\ell \in \mathbb{L}} \left(Risk(\mathcal{X}_L \cup \mathbf{z}^{(\ell)}) + Risk(\mathcal{X}_U \setminus \mathbf{z})\right) p(\ell|\mathbf{z}),$$

where $\mathbb{L}$ denotes all possible labels for $\mathbf{z}$.

✓ Easy if we are considering an unlabeled instance or bag.

Decision-theoretic multi-level criterion

= **Risk under the current classifier** – **Risk after adding *z* to the labeled set** – **Cost of obtaining annotation for z**

But if we are considering a positive bag $\mathbf{z} = \{z_1, \ldots, z_M\}$, then $\mathbb{L} = \{1, \ldots, C\}^M$

We compute the expected cost using Gibbs sampling:

$$\mathbb{E} = \frac{1}{S} \sum_{k=1}^{S} \Big(Risk(\mathcal{X}_L \cup \underbrace{\{z_1^{(a_1)_k}, \ldots, z_M^{(a_M)_k}\}}_{k^{\text{th}} \text{ sample: a label assignment for all instances in the bag}}) + Risk(\mathcal{X}_U \setminus \mathbf{z}) \Big)$$

k^{th} sample: a label assignment for all instances in the bag

Decision-theoretic multi-level criterion

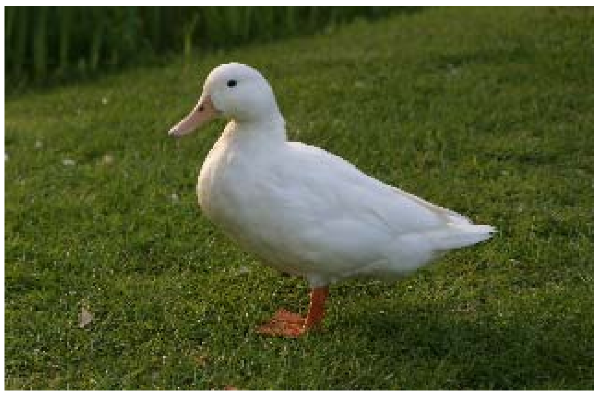
= **Risk under the current classifier** - **Risk after adding *z* to the labeled set** - **Cost of obtaining annotation for z**

We learn a function to predict the cost (effort) required to obtain any candidate annotation.

This looks expensive to annotate, and it does not seem informative.

Predicting effort

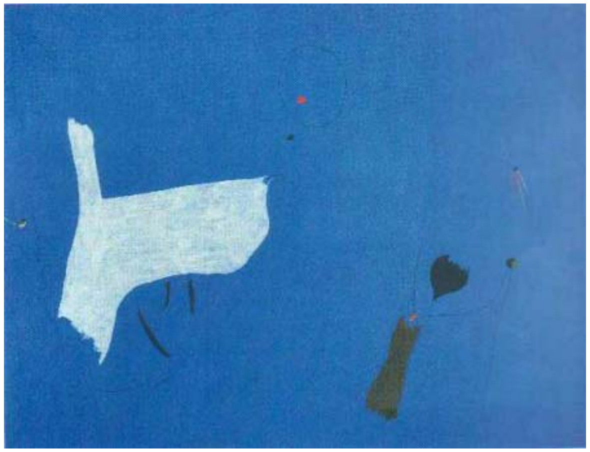
- What manual effort cost would we expect to pay for an unlabeled image?




Which image would you rather annotate?

Predicting effort

- What manual effort cost would we expect to pay for an unlabeled image?

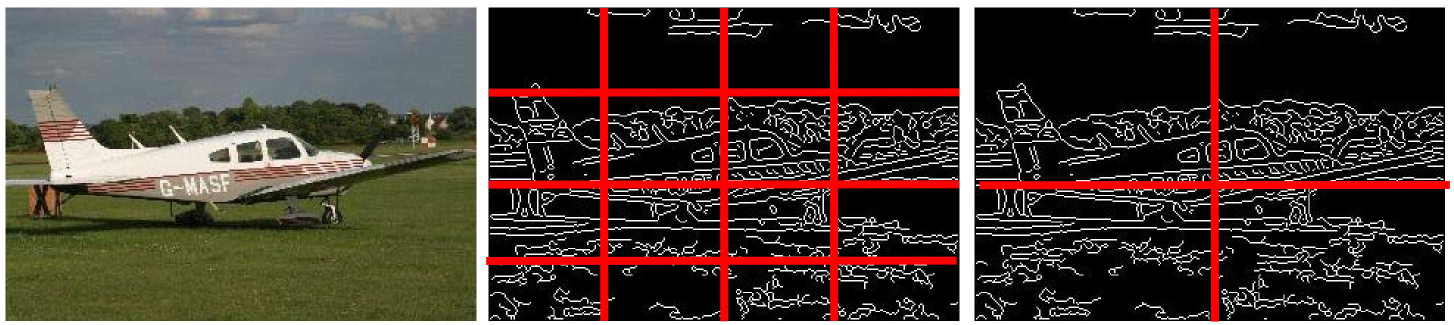


Which image would you rather annotate?

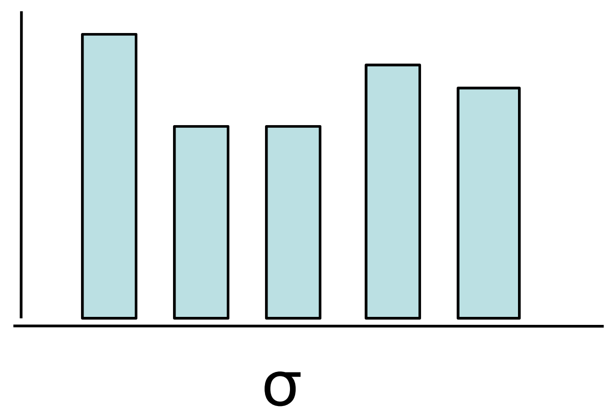
Learning from annotation examples

Extract cost-indicative image features, and train a support vector regressor to map features to times.

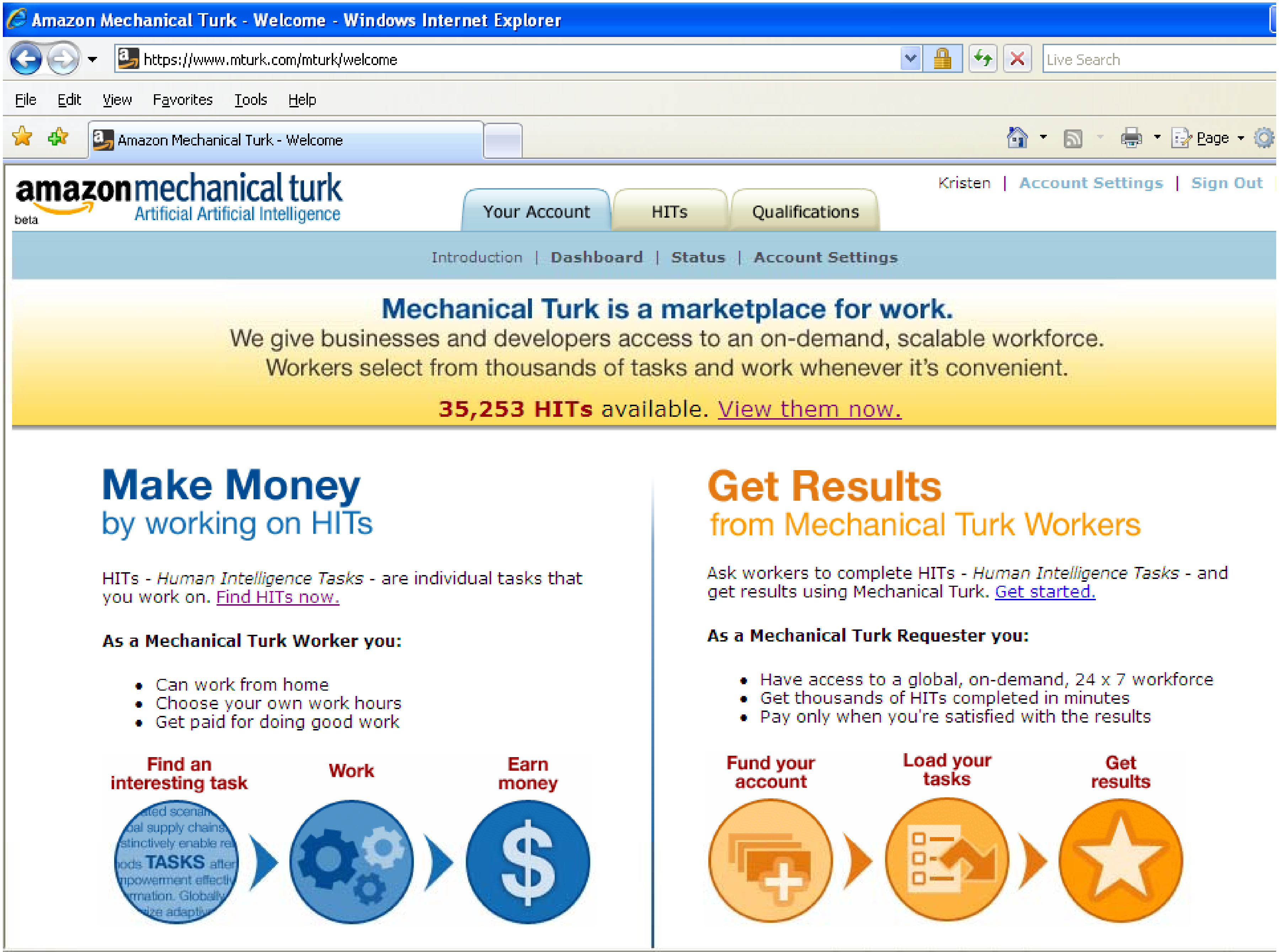
Localized measures of edge density



Measure of how fast color changes locally



K. Grauman, Learning Workshop, April 2009


Amazon Mechanical Turk - Welcome - Windows Internet Explorer
https://www.mturk.com/mturk/welcome
Live Search
File Edit View Favorites Tools Help
Amazon Mechanical Turk - Welcome
Page
amazon mechanical turk
Artificial Artificial Intelligence
beta
Kristen | Account Settings | Sign Out
Your Account
HITs
Qualifications
Introduction | Dashboard | Status | Account Settings
Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.
35,253 HITs available. View them now.
Make Money
by working on HITs
HITs - Human Intelligence Tasks - are individual tasks that you work on. Find HITs now.
As a Mechanical Turk Worker you:
• Can work from home
• Choose your own work hours
• Get paid for doing good work
Find an interesting task
Work
Earn money
TASKS
Get Results
from Mechanical Turk Workers
Ask workers to complete HITs - Human Intelligence Tasks - and get results using Mechanical Turk. Get started.
As a Mechanical Turk Requester you:
• Have access to a global, on-demand, 24 x 7 workforce
• Get thousands of HITs completed in minutes
• Pay only when you're satisfied with the results
Fund your account
Load your tasks
Get results

Learning from annotation examples

Interface on Mechanical Turk

…

…
32 s
24 s
48 s

Collect about 50 responses per training image.

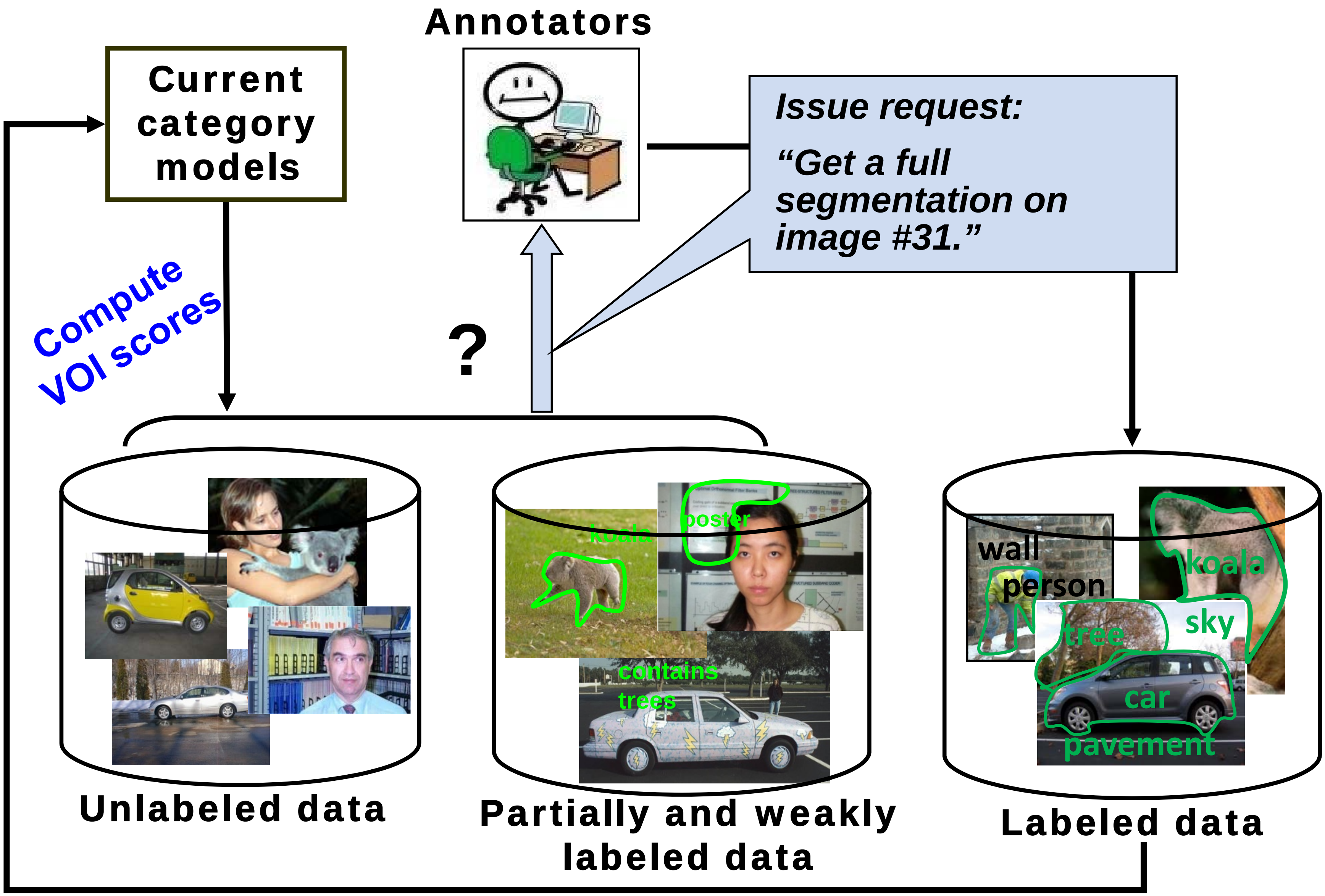
Decision-theoretic multi-level criterion

= **Risk under the current classifier** – **Risk after adding *z* to the labeled set** – **Cost of obtaining annotation for z**

We learn a function to predict the cost (effort) required to obtain any candidate annotation.

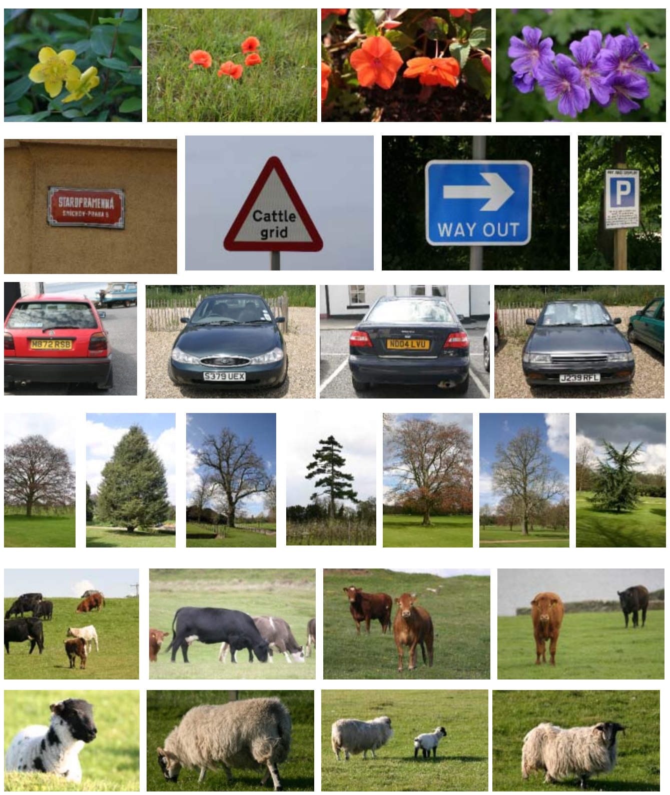
This looks expensive to annotate,
and it does not seem informative.

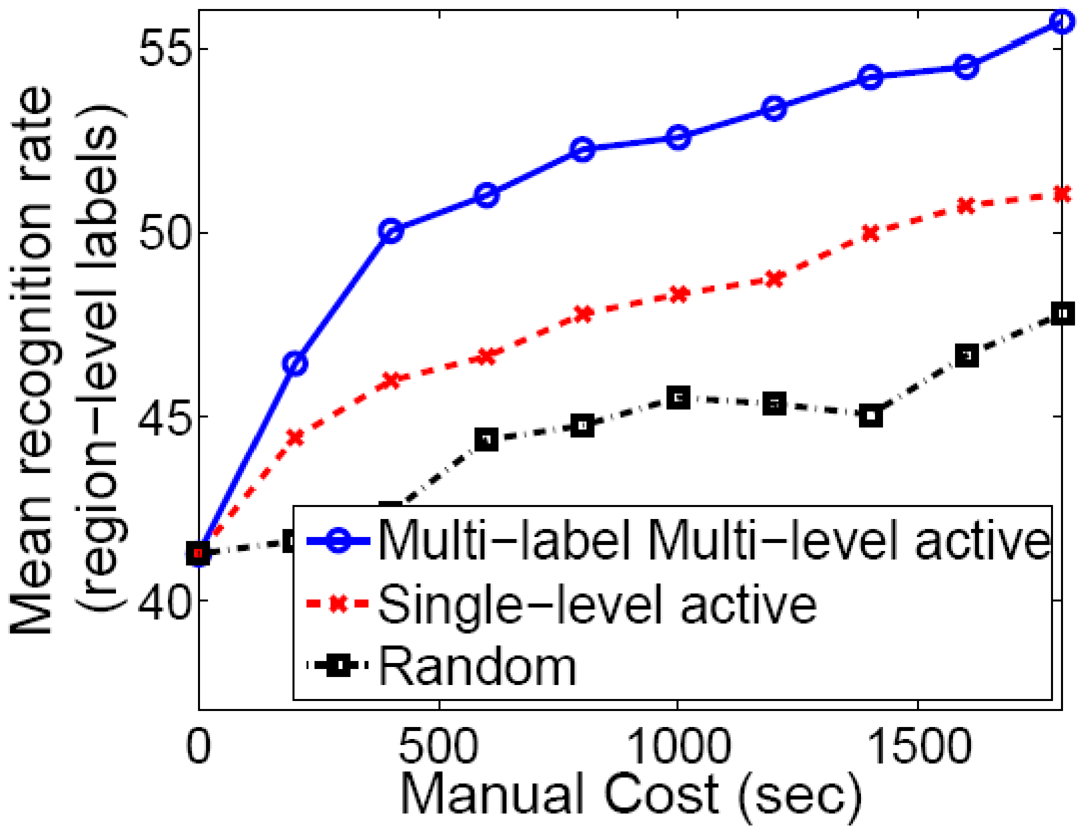
Recap: actively seeking annotations



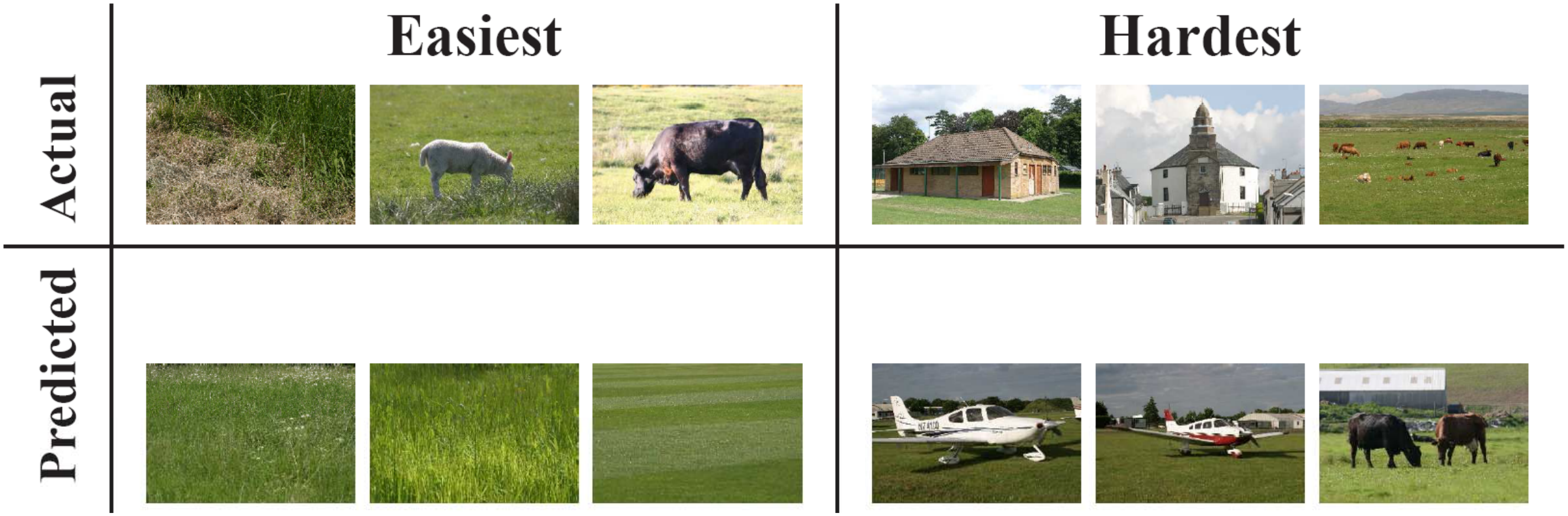
Results: MSRC dataset

- 21 classes, 591 images
- Multi-label data



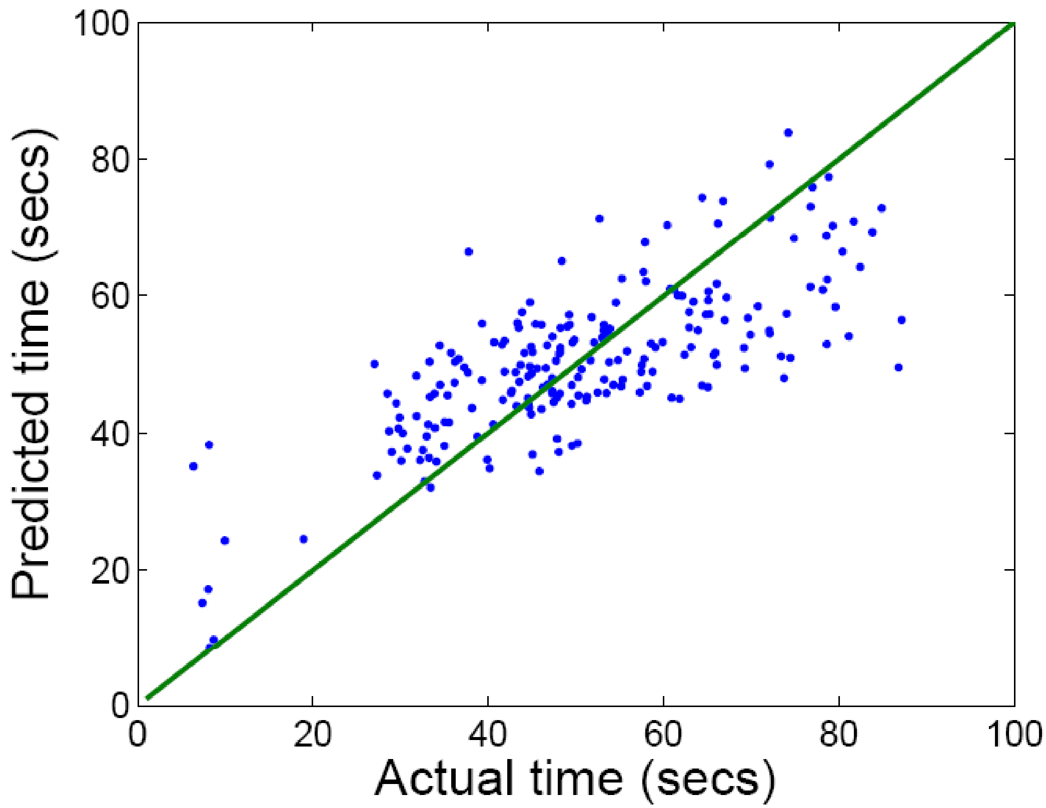


Results: predicting effort



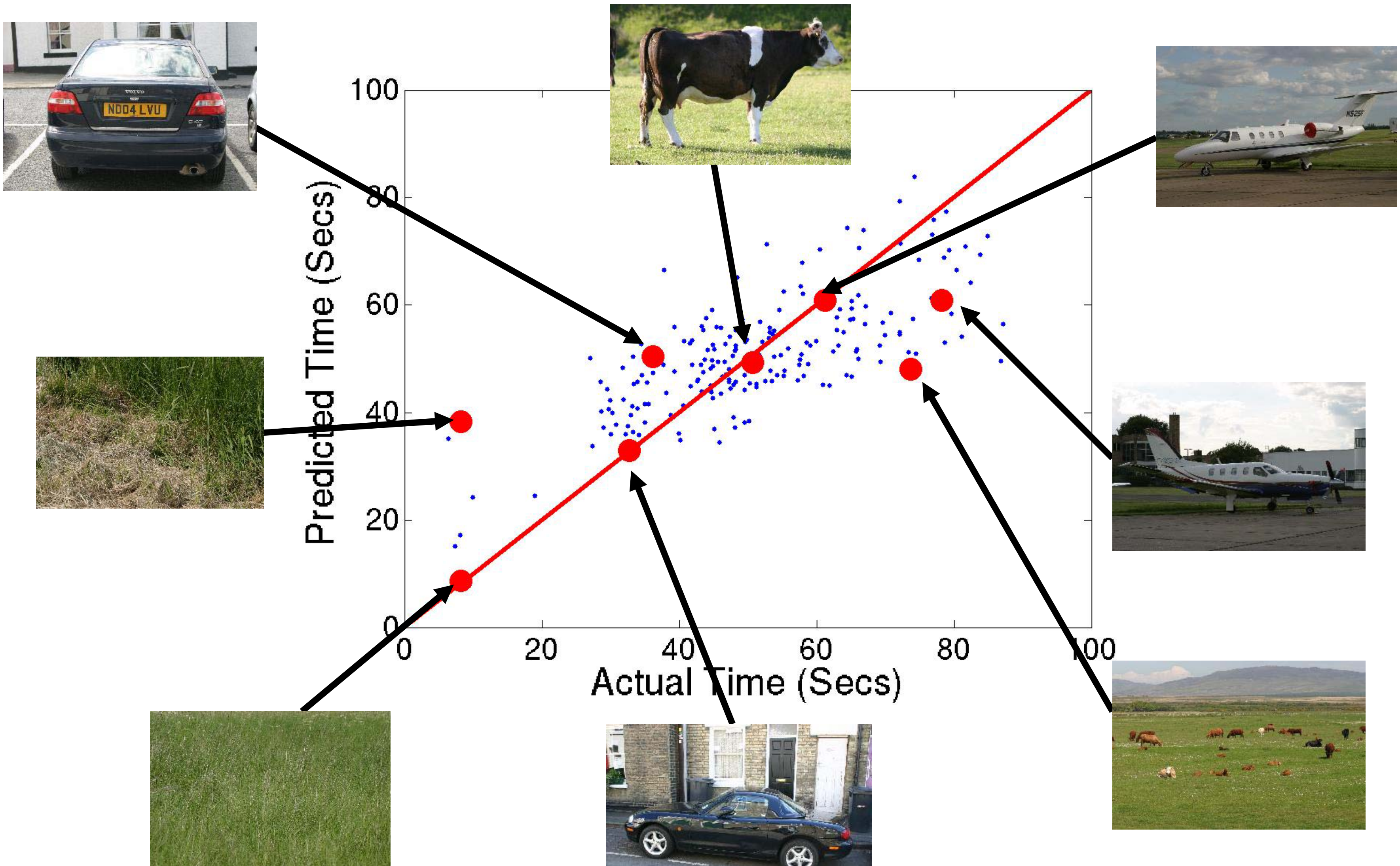
- Predicted examples are from a novel test set

Results: predicting effort

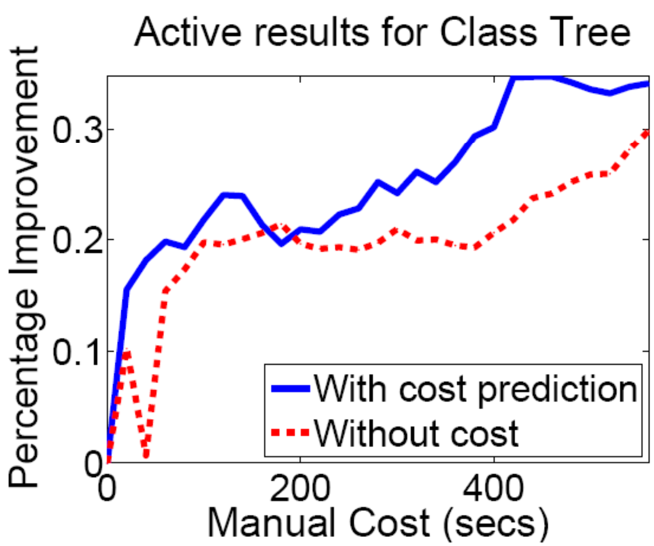


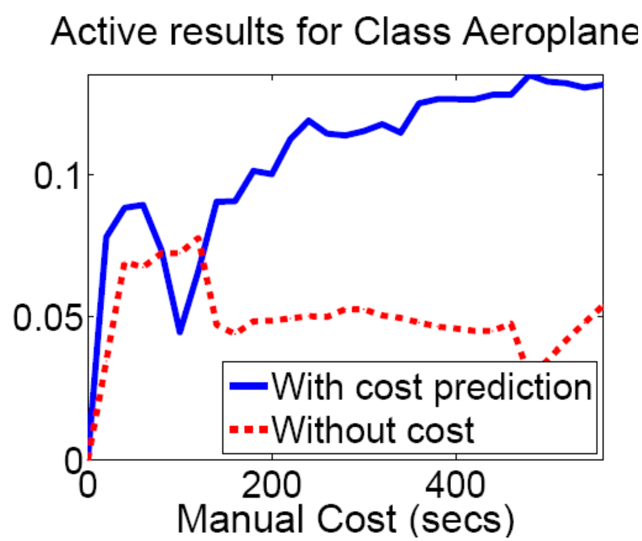
K. Grauman, Learning Workshop, April 2009

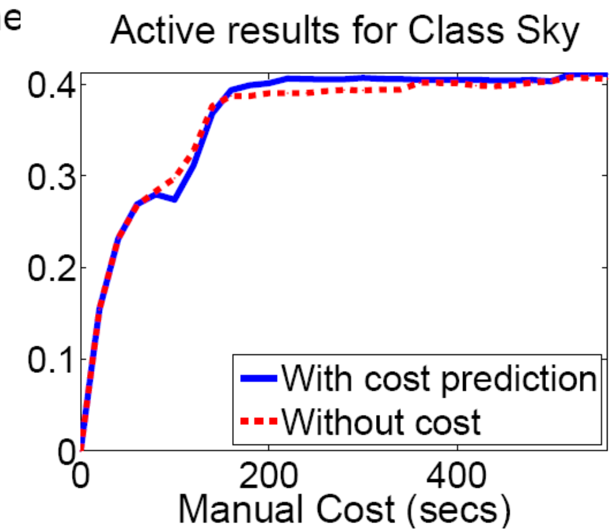
Results: predicting effort



Results: impact of cost predictions







Predicting the amount of effort entailed leads to wiser choices during active selection.





Summary

- Multi-level active learning formulates annotation requests that specify the example *and* the task.
- Balance cost and effort to use human attention most efficiently: learn more with less!
- Predict which examples are hard/easy to annotate.
- **References**:
  – Vijayanarasimhan & Grauman. Multi-Level Active Prediction of Useful Image Annotations for Recognition. In *NIPS* 2008.
  – Vijayanarasimhan & Grauman. What's It Going to Cost You? : Predicting Effort vs. Informativeness for Multi-Label Image Annotations. To appear, *CVPR* 2009.