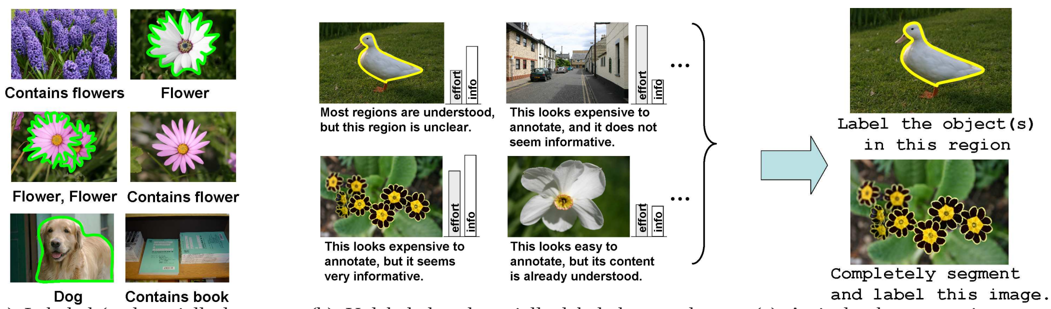
Cost-Sensitive Active Visual Category Learning

Sudheendra Vijayanarasimhan and Kristen Grauman
Department of Computer Sciences
University of Texas at Austin



(a) Labeled (and partially labeled) examples to build models

(b) Unlabeled and partially labeled examples to survey

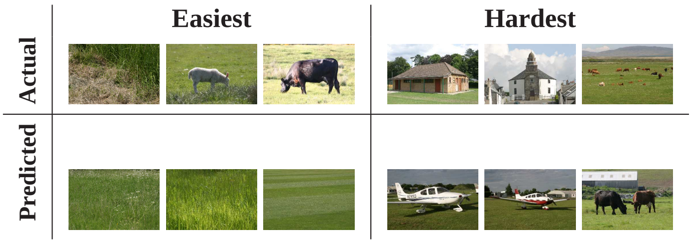
(c) Actively chosen queries sent to annotators

Figure 1: Overview of the proposed approach. (a) We learn object categories from multi-label images, with a mixture of weak and strong labels. (b) The active selection function surveys unlabeled and partially labeled images, and for each candidate annotation, predicts the tradeoff between its informativeness vs. the manual effort it would cost to obtain. (c) The most promising annotations are requested and used to update the current classifier.

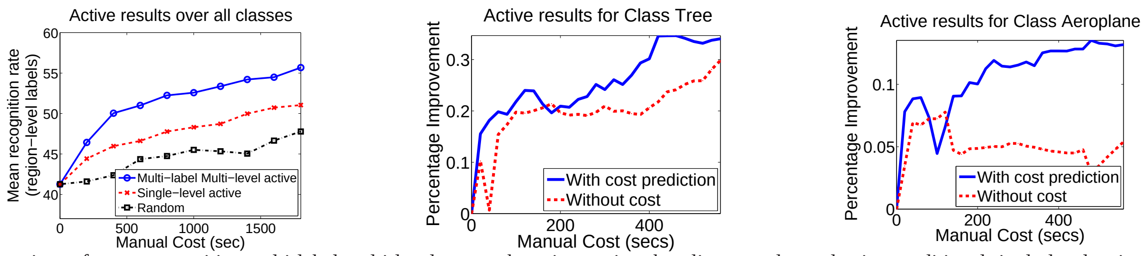
Are larger image training sets necessarily better for recognition? The accuracies of most current object recognition methods steadily improve as more and more labeled training data is made available. However, this requires manually collecting and possibly further annotating image examples, which is an expensive endeavor. Though the protocol of learning models from carefully gathered images has proven fruitful, it is too expensive to perpetuate in the long-term. Active learning strategies have the potential to reduce this burden by generally selecting only the most informative examples to label.

However, the active selection task for visual category learning has important distinctions from traditional active learning. First, most real-world images consist of multiple objects, and so should be associated with *multiple* labels simultaneously. Second, annotations in visual recognition can be provided at multiple levels (for example, full segmentation vs. present/absent flag). To use a fixed amount of manual effort resources most effectively the active learner must therefore be allowed to choose from multiple levels of annotation depending on how confident it feels about an example. Finally, traditional active learning methods implicitly assume that all annotations cost the same amount of effort and thus aim to minimize the total number of queries. On the other hand, in visual recognition, the actual manual effort required to label images varies both according to the annotation type as well as the particular image example (e.g., a complicated scene vs. an image with few objects).

We introduce an active learning framework where the expected informativeness of *any* candidate image annotation is weighed against the predicted cost of obtaining it [6, 5]. We explicitly account for the fact that image annotations can exist at multiple levels of granularity by expressing the problem in the *multiple instance learning* (MIL) setting [2], and we design a decision-theoretic criterion that balances the variable costs associated with each type of annotation versus the expected gain in information. Thus, our active learner chooses both which image example as well as what *type* of annotation to request: a complete image segmentation, a segmentation of a single object, or an image-level category label naming one of the objects within it.



(a) Which image would you rather annotate? Whereas traditional active learning methods assume each query it makes will require equal effort on the part of the annotator, we learn a classifier to predict the difficulty per image. This figure shows the easiest and hardest images to annotate based on actual users' timing data (top), and the predictions of our cost function on novel images (bottom). Our method learns to predict annotation times directly from image features.



(b) Comparison of our cost-sensitive multi-label multi-level approach against various baselines: random selection, traditional single-level active selection, and active selection with a flat cost. All curves are plotted against the true cost associated with the selected annotations. Our approach produces the best improvement per unit manual cost compared to the baselines. **Left:** Region-level accuracy for the 21-class MSRC dataset for our approach, single-level active selection, and random selection. **Middle, Right:** Representative learning curves when using active selection with the learned cost predictor, as compared to a baseline that makes active selections using a flat cost.

Figure 2: Results obtained on the MSRC dataset.

We generalize our earlier model ([6]) to the multi-label setting where multiple categories can exist in a single image, and devise a *multiple-instance*, *multi-label learning (MIML)* formulation for both the classifier and the decision-theoretic criterion for choosing examples. Since different images might require different amounts of manual effort to annotate, we allow the active learner to estimate an image's difficulty before requesting the user to provide the annotation. Based on the observation that humans (especially vision researchers) can easily glance at an image and roughly gauge the difficulty, we train a regression function to predict the annotation cost of an image given its features.

In contrast to the proposed approach, previous active learning methods for recognition only consider which examples to obtain a class label for [3, 1, 4], or else are limited to binary and/or single-label problems [3, 6]. This is the first work to learn from both multi-label image-level and region-level annotations, and the first to consider predicting the cost of an unseen image annotation in order to strengthen active learning choices.

Based on experiments on the MSRC dataset, a multi-label dataset containing 21 classes, we show that we can predict the annotation costs for unlabeled images directly from image features by learning with data collected from anonymous users through Amazon's Mechanical Turk interface (Figure 2(a)). Further, by incorporating the cost predictor in the active learning framework we show that, in contrast to traditional active learning, one needs to account for the variable cost and multiple types of annotations to truly reduce manual effort (Figure 2(b)).

References

[1] B. Collins, J. Deng, K. Li, and L. Fei-Fei. Towards Scalable Dataset Construction: An Active Learning Approach. In *ECCV*, 2008.

[2] T. Dietterich, R. Lathrop, and T. Lozano-Perez. Solving the Multiple Instance Problem with Axis-Parallel Rectangles. *Artificial Intelligence*, 89(1-2):31–71, 1997.

[3] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell. Active Learning with Gaussian Processes for Object Categorization. In *ICCV*, 2007.

[4] G. Qi, X. Hua, Y. Rui, J. Tang, and H. Zhang. Two-Dimensional Active Learning for Image Classification. In *CVPR*, 2008.

[5] S. Vijayanarasimhan and K. Grauman. What's It Going to Cost You?: Predicting Effort vs. Informativeness for Multi-label Image Annotations. To appear in *CVPR, 2009*.

[6] S. Vijayanarasimhan and K. Grauman. Multi-Level Active Prediction of Useful Image Annotations for Recognition. In *NIPS*, 2008.

Topic: Visual Processing and Pattern Recognition, Learning algorithms