Expanding Training Sets with Unlabeled Samples by Learned Attributes

Jonghyun Choi Mohammad Rastegari Ali Farhadi[†] Larry S. Davis
University of Maryland, College Park [†]University of Washington
{jhchoi,mrastega,lsd}@umiacs.umd.edu ali@cs.uw.edu

1. Introduction

Designing generalizable classifiers for visual categories is an active research area and has led to the development of many sophisticated classifiers in vision and machine learning [9]. Building a good training set with minimal supervision is a core problem in training visual category recognition algorithms [1].

A good training set should span the appearance variability of its category. While the internet provides a nearly boundless set of potentially useful images for training many categories, a challenge is to select the relevant ones – those that help to change the decision boundary of a classifier to be closer to the best achievable. So, given a relatively small initial set of labeled samples from a category, we want to mine a large pool of unlabeled samples to identify *visually different* examples without human intervention.

To expand the boundary of a category to an *unseen* region, we propose a method that selects unlabeled samples based on their attributes. The selected unlabeled samples are not always instances from the same category, but they can still improve category recognition accuracy, similar to [4, 5]. We use two types of attributes: category-wide attributes and example-specific attributes. The category-wide attributes find samples that share a large number of discriminative attributes with the preponderance of training data. The example-specific attributes find samples that are highly predictive of the *hard* examples from a category - the ones poorly predicted by a leave one out protocol.

We demonstrate that our augmented training set can significantly improve the recognition accuracy over a very small initial labeled training set, where the unlabeled samples are selected from a very large unlabeled image pool, *e.g.*, ImageNet. Our contributions are summarized as follows:

1. We show the effectiveness of using attributes learned with auxiliary data to label unlabeled images without annotated attributes.
2. We propose a framework that jointly identifies the unlabeled images and category wide attributes through an optimization that seeks high classification accuracy in both the original feature space and the attribute space.
3. We propose a method to learn example specific attributes with a small sized training set, used with the proposed framework. We then combine the category wide and the example specific attributes to further improve the quality of image selection by diversifying the variations of selected images.

Without modeling the sample distribution and human involvement in the loop, we achieve to find and add samples to categories by attributes, which are helpful for recognition.

For more detailed description of the approach and more results, please refer to our main conference version titled **"Adding Unlabeled Samples to Categories by Learned Attributes."**[1]

2. Approach

Based on recent work on automatic discovery of attributes [7] and large scale category-labeled image datasets [2], we discover a rich set of attributes. These attributes are leaned using an auxiliary category-labeled dataset to avoid biasing the attribute models towards the few labeled examples. The motivation here is similar to what underlies the successful Classemes representation [8] which achieved good category recognition performance by representing samples by external data that consists of a large number of samples from various categories.

Across the original visual feature space and the attribute space, we propose a framework that jointly selects the unlabeled images to be assigned to each category and the discriminative attribute representations of the categories based on either a category wide or exemplar based ranking criteria. Following subsection presents the optimization framework for category wide addition of unlabeled samples to categories. This adds samples that share many discriminative attributes amongst themselves and the given labeled training data.

Categorical Analysis For each category c, we will construct a classifier in visual feature space, w_c^v, using the set $X = \{x_i | i \in \{1, \ldots, l, l+1, \ldots, n\}\}$ that consists of the initially given labeled training images $\{x_i | i \in \{1, \ldots, l\}\} \subset X$ and the selected images from the unlabeled image pool $\{x_i | i \in \{l+1, \ldots, n\}\} \subset X$. The subset of images from the unlabeled set is assigned to a category based on identifying discriminative attribute models. Since the problems of determining the discriminative attributes and selecting the subset of unlabeled data to assign to a category are coupled, we learn them jointly. Additionally, we want to mitigate against unlabeled samples being assigned

[1] The paper is also found in our website: http://umiacs.umd.edu/~jhchoi

to multiple categories, so a term $M(\cdot)$ is added to the optimization criteria to enforce that. The joint optimization function is:

$$
\min_{I_c \in I, w_c^v, w_c^a} \sum_c \Big(\alpha J_c^v(I_c, w_c^v) + \beta J_c^a(I_c, w_c^a) \Big) + M(I)
$$

subject to

$$
\begin{aligned}
& J_c^v(I_c, w_c^v) = \|w_c^v\|_2^2 + \lambda_v \sum_{i=1}^{n} \xi_{c,i} \\
& I_{c,i} \cdot y_{c,i}(w_c^v x_i) \geq 1 - \xi_{c,i}, \quad \forall i \in \{1, \ldots, n\} \\
& J_c^a(I_c, w_c^a) = \|w_c^a\|_2^2 + \lambda_a \sum_{j=1}^{n} \zeta_{c,j} - \sum_{k=l+1}^{n} I_{c,k} \Big(w_c^a \phi(x_k) \Big) \\
& I_{c,j} \cdot y_{c,j}(w_c^a \phi(x_j)) \geq 1 - \zeta_{c,j}, \quad \forall j \in \{1, \ldots, n\} \\
& \sum_{k=l+1}^{n} I_{c,k} \leq \gamma, \quad I_{c,k} = 1, \, \forall k \in \{1, \ldots, l\} \\
& M(I) = \sum\sum_{c1 \neq c2} I_{c1} \cdot I_{c2},
\end{aligned} \tag{1}
$$

$I_c \in \{0, 1\}$ is the sample selection vector for category c, and indicates which unlabeled samples are selected for assignment to the training set of category c. $I_{c,i} = 1$ when the i^{th} sample is selected for category c. $x_i \in \mathbb{R}^D$ is the visual feature vector of image i. $y_{c,i} \in \{+1, -1\}$ indicates whether the label assigned to x_i is c $(+1)$ or not (-1). $\phi(\cdot) : \mathbb{R}^D \to \mathbb{R}^A$ is a mapping function of visual feature to the attribute space that is learned from auxiliary data, where $\mathbb{R}^D$ and $\mathbb{R}^A$ denote visual feature space and attribute space, respectively. α and β are hyper-parameters for balancing the max margin objective terms for both the visual feature and attribute based classifiers. γ is a hyper-parameter for specifying the number of selected images.

Exemplar Analysis The discriminative attributes learned by the Categorical Analysis capture commonality among all examples in a category. We refer them as *categorical attributes*. Each example, however, has its own characteristics that may help to expand the visual space of the category by identifying images based on example-specific characteristics. To discover *exemplar attributes*, a straightforward solution would be to learn exemplar-SVMs [6]. The exemplar-SVM, however, requires many negative samples to make the classifier output stable. For our purposes, though, we can accomplish the same thing by analyzing how the ranks of unlabeled samples change when a single sample is eliminated from the training set of the attribute SVM. If an unlabeled sample sees its rank drop sharply from its rank in the full-sample SVM, then the training sample dropped should have strong attribute similarity to the unlabeled sample.

3. Results

We demonstrate the effectiveness of our method by improvements in average precision (AP) of category recognition. We compare to baseline algorithms which are applicable to the large unlabeled data scenario. The first baseline algorithm is to select nearest neighbors. The second baseline selects images by an active criterion that finds examples close to a learned decision hyperplanes [3]. Both baseline algorithms selects images based on analysis in the visual feature space.

Category Name	Init.	NN	ALC	Cat.	E+C
Mashed Potato	45.03	34.02	51.15	61.39	63.92
Orange	29.84	16.29	26.97	40.61	41.05
Lemon	32.21	27.58	32.43	35.37	34.23
Green Onion	25.06	16.50	19.66	38.57	40.20
Acorn	13.09	11.05	15.41	19.35	20.10
Coffee bean	58.29	43.89	56.62	64.65	66.54
Golden Retriever	14.54	15.57	12.61	17.54	18.61
Yorkshire Terrier	29.62	13.62	27.63	41.41	45.65
Greyhound	15.24	15.73	15.64	14.75	15.22
Dalmatian	43.84	27.97	37.91	54.42	57.23
Miniature Poodle	26.10	12.50	21.16	28.87	30.21
Average	30.26	21.34	28.84	37.90	39.36

Table 1. Comparison of average precision (AP) (%) for each category with 50 added examples by various methods. 'Init.' refers to initial labeled training set. 'NN' refers to addition by 'nearest neighbor' in visual feature space, 'ALC' refers to addition by 'active learning criteria (ALC)' that finds the examples close to the current decision hyperplanes [3]. 'Cat.' refers to our method of select examples using categorical attributes only. 'E+C' refers to addition using categorical and exemplar attributes. The size of the unlabeled dataset is roughly 3,000 from randomly chosen categories out of 1,000 categories.

As shown in Table. 1, the two baseline strategies decrease mean average precision (mAP). However, our method identifies useful images in the unlabeled image pool and significantly improves mAP by 7.64%. Except for the category *Greyhound*, we obtain performance gain from 2.77% - 16.36% in all categories. The added examples serve not only as positive samples for each category but also as negative samples for other categories. The quality of the selected set can change the mAP significantly in both ways.

References

[1] T. L. Berg, A. Sorokin, G. Wang, D. A. Forsyth, D. Hoiem, A. Farhadi, and I. Endres. It's All About the Data. In *Proceedings of the IEEE, Special Issue on Internet Vision*, 2010. 1

[2] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009. 1

[3] P. Jain, S. Vijayanarasimhan, and K. Grauman. Hashing Hyperplane Queries to Near Points with Applications to Large-Scale Active Learning. In *NIPS*, 2010. 2

[4] J. Kim and K. Grauman. Shape Sharing for Object Segmentation. In *ECCV*, 2012. 1

[5] J. J. Lim, R. Salakhutdinov, and A. Torralba. Transfer Learning by Borrowing Examples for Multiclass Object Detection. In *NIPS*, 2011. 1

[6] T. Malisiewicz, A. Gupta, and A. A. Efros. Ensemble of Exemplar-SVMs for Object Detection and Beyond. In *ICCV*, 2011. 2

[7] M. Rastegari, A. Farhadi, and D. Forsyth. Attribute Discovery via Predictable Discriminative Binary Codes. In *ECCV*, 2012. 1

[8] L. Torresani, M. Szummer, and A. Fitzgibbon. Efficient Object Category Recognition Using Classemes. In *ECCV*, 2010. 1

[9] W. Zhang, S. X. Yu, and S.-H. Teng. PowerSVM: Generalization with Exemplar Classification Uncertainty. In *CVPR*, 2012. 1