Abstract to be presented at the CVPR Workshop on Fine-Grained Visual Categorization (FGVC), June 2011.

Sharing Features Between Visual Tasks at Different Levels of Granularity

Sung Ju Hwang[1], Fei Sha[2] and Kristen Grauman[1]
[1]Department of Computer Science
University of Texas at Austin
{sjhwang,grauman}@cs.utexas.edu
[2]Computer Science Department
University of Southern California
feisha@usc.edu

1. Introduction

We explore how learning visual features shareable between tasks of different levels of granularity may enhance recognition accuracy. Our idea is that by accounting for the multiple semantic labels applicable to objects during feature learning (i.e., basic-level, fine-grained, or attributes), one may recover more robust representations. In particular, by simultaneously learning features well-suited to discriminate both at the basic-level (dog, cat, mouse) as well as at the fine-grained level (collie, dalmatian, greyhound), we seek a representation that allows better fine-grained predictions than would be possible if learning features based on the fine-grained classes alone. Similarly, by targeting features shareable between the objects and generic descriptive attributes (furry, white, has-legs), we aim to recover features that more reliably discriminate the objects themselves.

Why should this work? The assumption is that tasks at different levels of granularity rely on some shared structure in the original image descriptor space. In effect, we expect that human-defined semantics as revealed by attributes or basic-level "supercategory" labels can help regularize the training process, providing generic information about which low-level cues are valuable to finer-grained recognition.

To this end, we propose an approach to discover such structure and learn a shared representation amenable to discriminative models for object categories at some target granularity (see Figure 1). During training, we require that the classifiers for each task share a lower-dimensional feature space, using a multi-task feature learning approach developed in [1]. Given a low-level visual feature space together with the tasks—which are either *main* fine-grained tasks on which we want to improve, or the *auxiliary* basic-level or attribute tasks from which we want to benefit—we learn a feature subspace based on a joint loss function that favors common sparsity for all labeling tasks.

A side benefit of this strategy is that in some cases the auxiliary levels of labels are easily transcribed from instances originally labeled for the target space, making efficient use of external knowledge. For example, a semantic hierarchy (like WordNet or others) allows us to transfer labels of an object's superclass to any of its subclass's image instances.

Previously, shared representations and transfer learning have been shown to improve recognition at a single level of granularity, particularly when working with scarce labeled examples [2, 8], amortizing feature extraction costs during detection [9, 10], or leveraging auxiliary tasks from text [7, 5]. In contrast, we aim to improve recognition accuracy by leveraging tasks at different abstraction levels. Furthermore, work with visual hierarchies or attributes has largely focused on issues in scalability or mid-level representations, respectively, and typically provides supervision to separately learn each task (e.g., [6, 4]). Our goal, in contrast, is to study how the relationships between tasks at different levels can strengthen a joint learning process.

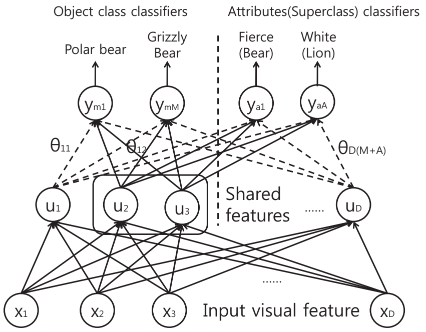

Figure 1. In our model, object class classifiers share a lower dimensional representation (dashed lines indicate zero-valued connections) with either visual attribute classifiers or superclass classifiers, thereby allowing supervision for object recognition at different levels of granularity to regularize the learned object models.

We briefly sketch the approach, and then provide example results based on our findings for both sharing with attributes (which will appear in CVPR 2011 [3]) as well as newer results for sharing with supercategories.

2. Learning Shared Features

Suppose we have M *main* tasks for which we want to improve accuracy, and A auxiliary tasks that we want to leverage for feature learning. Let $\boldsymbol{x}_n \in \mathbb{R}^{\mathsf{D}}$ denote the n-th feature vector in the training data and y_n its target fine-grained class label. For the A auxiliary tasks, let y_{na} denote the label for the a-th auxiliary task, which is either a supercategory or attribute task. Conventionally all $\mathsf{T} = \mathsf{A} + \mathsf{M}$ classifier parameters $\{\boldsymbol{w}_m\}_{t=1}^{\mathsf{T}}$ would be learned independently, but we want to learn them jointly.

First, we transform the original features to a shared feature space $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_n \in \mathcal{U}$ for all tasks [1]. Then, we learn models in the space of $\mathcal{U}$ and promote a common sparsity pattern in the new parameters. Denote the linear discriminants as $\{\boldsymbol{\theta}_t\}$, such that $\boldsymbol{w}_t = \boldsymbol{U}\boldsymbol{\theta}_t$. We jointly optimize all loss functions, regularized with $\boldsymbol{\Theta}$'s $(2,1)$-norm:

$$\boldsymbol{\Theta}^*,\ \boldsymbol{U}^* = \arg\min \sum_t \sum_n \ell(\boldsymbol{\theta}_t^{\mathrm{T}} \boldsymbol{U}^{\mathrm{T}} \boldsymbol{x}_n, y_{nt}) + \gamma \|\boldsymbol{\Theta}\|_{2,1}^2$$

where $\ell(\cdot)$ denotes the classifier loss using the learned features, and $\boldsymbol{\Theta} \in \mathbb{R}^{\mathsf{D}\times\mathsf{T}}$ contains the discriminants $\{\boldsymbol{\theta}_t\}$ as its columns. The $(2,1)$ norm regularization term favors choosing the $\boldsymbol{\Theta}$ with the *smallest* number of *non-zero rows*, so that it yields solutions that use a subset of features that are commonly effective for all tasks. We use a kernelized form of the above, and optimize with the alternating minimization algorithm proposed in [1]. Essentially, the optimization process alternates between regularizing towards shared features, and retraining task-specific classifiers based on those features.

3. Results

We validate our approach in two scenarios: 1) where binary attributes serve as auxiliary tasks, and 2) where superclass basic-level categories serve as auxiliary tasks. For both, we use the *Animals with Attributes* dataset (AWA) [4] which contains 30,475 images and 50 animal classes associated with 85 different attributes. The classes are a mix of basic and fine-grained (e.g., Persian cat, Siamese cat, Polar bear, etc.). We use the features provided with the dataset, and average their χ^2 kernels to form the original feature space. We form splits of 60% of the images to train, 20% for validation, and 20% for testing. Due to limited space, we only briefly summarize some outcomes here.

As baselines, we consider 1) a "No sharing" approach, which is a traditional multi-class SVM using the same kernel with which our method begins, and 2) a "Sharing-Same level only" approach, which uses our method but shares features only among the main task's classes, i.e., at the same level of granularity.

Sharing Features with Attributes Table 1 shows the impact of feature sharing with attributes, where the main task is to categorize the animal present in the image. We compare our "Sharing+Attributes" approach to the two baselines defined above, as well as a "No sharing+Attributes" baseline that learns attribute classifiers separately, and then uses them as mid-level features to categorize the animal with the Direct Attribute Prediction model of [4]. We see consistent improvements over the baselines that lack feature sharing. Furthermore, the proposed approach gives the best results when sharing with attributes, showing the value of regularizing according to a label space from a separate level of granularity.

Method / % train data	10%	20%	40%	60%
No sharing	31.96	38.12	44.08	48.03
No sharing+Attributes	31.03	35.61	41.12	43.59
Sharing-Same level only	**37.08**	41.01	46.46	49.15
Sharing+Attributes	36.73	**42.60**	**47.70**	**50.94**
% gain over No sharing	14.92%	11.75%	8.21%	6.06%

Table 1. Accuracy when sharing features with attributes. Regularizing the classifiers with auxiliary attribute tasks improves predictions for the fine-grained and basic-level animal categories—especially when training with fewer labeled examples.

Method	Main=subclasses	Main=superclasses
No sharing	35.97	**47.92**
Sharing-Same level only	36.88	47.01
Sharing+Superclass	**37.43**	-
Sharing+Subclass	-	46.87

Table 2. Accuracy when sharing features between different semantic levels. While a finer-grained recognition task benefits from feature sharing with the coarser-level animal categories (middle column), the reverse is not true (rightmost column). See text.

Sharing Features across Abstraction Levels Now we explore generating auxiliary tasks from the semantic taxonomy over object classes. As an initial proof of concept, we build a two-level class hierarchy out of 40 of the 50 animals, by grouping them into 16 superclasses according to the WordNet hierarchy (e.g., felines, bears). We consider two variants, where the main task is to recognize either the subclasses (i.e., predict the original AWA labels) or the superclasses.

Table 2 shows the results, for 10% training data.[1] Interestingly, we see that the finer-grained categorization tasks benefit from sharing a representation with their superclasses (middle column), whereas the coarser-grained categorization tasks do *not* benefit from sharing with their subclasses (right column). These results are fairly intuitive. For the former, we obtain a regularization benefit similar to the attribute-based results above: the features for discriminating the finer-grained classes are better focused by accounting for the auxiliary superclass tasks in parallel. For the latter, however, the finer-grained classes introduce a distraction that harms the superclass predictions, forcing those models to account for variability that is not relevant for the desired main task.

4. Conclusion

We find that by learning a common feature space suitable to tasks at multiple levels of granularity, we obtain noticeably stronger object recognition results. In ongoing work, we are developing ways to automatically select tasks to share, and exploring the super/subclass sharing behavior more deeply.

References

[1] M. P. A. Argyriou, T. Evgeniou. Convex Multi-task Feature Learning. *Machine Learning*, 73(3):243–272, 2008.

[2] L. Fei-Fei, R. Fergus, and P. Perona. A Bayesian Approach to Unsupervised One-Shot Learning of Object Categories. In *ICCV*, 2003.

[3] S. J. Hwang, F. Sha, and K. Grauman. Sharing Features Between Objects and Their Attributes. In *CVPR*, 2011.

[1] Baseline results differ in tables due to number of classes.

[4] C. Lampert, H. Nickisch, and S. Harmeling. Learning to Detect Unseen Object Classes by Attribute Transfer. In *CVPR*, 2009.

[5] N. Loeff and A. Farhadi. Scene Discovery by Matrix Factorization. In *ECCV*, 2008.

[6] M. Marszalek and C. Schmid. Constructing Category Hierarchies for Visual Recognition. In *ECCV*, 2008.

[7] A. Quattoni, M. Collins, and T. Darrell. Learning Visual Representations Using Images with Captions. In *CVPR*, 2007.

[8] M. Stark, M. Goesele, and B. Schiele. A Shape-based Object Class Model for Knowledge Transfer. In *ICCV*, 2009.

[9] A. Torralba and K. Murphy. Sharing Visual Features for Multiclass and Multiview Object Detection. *PAMI*, 29(5), 2007.

[10] X.-T. Yuan and S. Yan. Visual Classification with Multi-Task Joint Sparse Representation. In *CVPR*, 2010.