

Latent Task Adaptation with Large-scale Hierarchies

Cheng Guan

Abstract

Recent years have witnessed the success of large-scale image classification systems that are able to identify objects among thousands of possible labels. However, it is yet unclear how general classifiers such as ones trained on ImageNet can be optimally adapted to specific tasks, each of which only covers a semantically related subset of all the objects in the world. It is inefficient and suboptimal to retrain classifiers whenever a new task is given, and is inapplicable when tasks are not given explicitly, but implicitly specified as a set of image queries. In this paper we propose a novel probabilistic model that jointly identifies the underlying task and performs prediction with a linear time probabilistic inference algorithm, given a set of query images from a latent task. We present efficient ways to estimate parameters for the model, and an open-source tool box to train classifiers distributedly at a large scale. Empirical results based on the ImageNet data showed significant performance increase over several baseline algorithms

1. Introduction

Recent years have witnessed a growing interest in object classification tasks involving specific sets of object categories, such as fine-grained object classification [1, 2] and home object recognition in visual robotics. Existing methods in the literature generally describe algorithms that are trained and tested on exactly the same task, *i.e.* we assume the training data and testing data share the same set of object labels. A dog breed classifier is trained and tested on dogs and a cat breed classifier done on cats, without the use of out-of-task images.

However, two observations may render this one (multiclass) classifier per task approach suboptimal. First, it is known that using images of related tasks is often beneficial to build a better model for the general visual world [4], which serves as a better regularization for the specific task as well. Second, object categories in the real world are often organized in, or at least well modeled by, a nested taxonomical hierarchy (*e.g.* Figure 1), with classification task score corresponding to intermediate subtrees in this hierarchy, and recent efforts on the ImageNet challenge [3, 5] have leveraged the use of large-scale data to learn such information. While

it is reasonable to train separate classifiers for specific tasks, this quickly becomes infeasible as there are a huge number of possible tasks - any subtree in the hierarchy may be a latent task requiring one to distinguish object categories under the subtree.

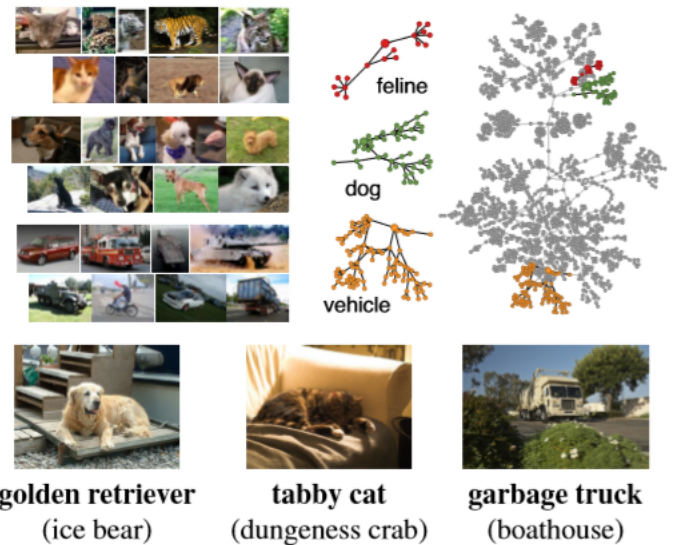


Figure 1. Top: Visualization of specific object classification tasks of interest in daily life, which are often subtrees in a large scale object taxonomy, *e.g.* the ImageNet hierarchy. Bottom: Adapting the ImageNet classifier allows us to perform accurate prediction (bold), while the original classifier prediction (in parentheses) suffers from a higher confusion.

2. Related Work

The problem of task adaptation is analogous to, but essentially distinctive from domain adaptation. While domain adaptation aims to model the perceptual difference of the training and testing images from the same labels, task adaptation focuses on modeling the conceptual difference: different label spaces during training and testing. Additionally, as one is often able to use large amounts of data during training, we assume that the testing tasks involve subsets of labels encountered during training time.

Predicting the intermediate concept in a hierarchy with a set of examples has been discussed in psychology. These methods often make a simplified assumption that labels (leaf

nodes in the hierarchy) are given for the input images. We believe our paper is the first to connect such psychological study with computer vision research by directly taking perceptual inputs, allowing one to perform generalization with images of unknown category.

3. A Generative Model for Task Adaptation

Formally, we define a classification task to be a subset of all the possible object labels that are semantically related (such as all breeds of dogs in ImageNet). During testing time, a number of query images are randomly sampled from the labels belonging to a task, and the learning algorithm needs to give predictions on these images. In this section we propose a probabilistic framework that models the generation of latent tasks and the test time query images.

As stated in the previous section, we are interested in the scenario when the task is latent, *i.e.* only implicitly specified by the query images. We introduce two key components for modeling the generative process of query images: a latent task space that defines possible tasks and their probability, and a procedure to sample query images given a specific latent task. Specifically, we propose the graphical model in the Figure 2 which generates a set of N query images when given T possible tasks and K object categories:

1. Sample a latent task h from the task priors $p(h)$ with hyperparameter α ;
2. For the N query images:
 - (a) Sample an object category y_i from the conditional probability $P(y_i|h; \beta_h)$.
 - (b) Sample a query image from category y_i with $P(y_i|h; \theta_{y_i})$.

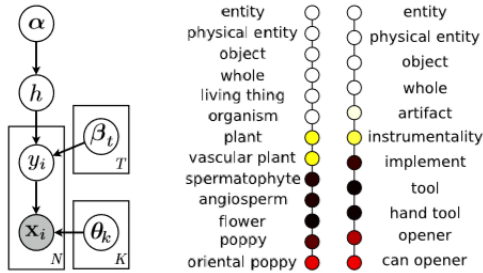


Figure 2. Left: the generative model for the latent task and corresponding query images. Right: the prior probabilities of the latent tasks from psychological study, along the path leading to the synsets oriental poppy and can opener respectively, with darker color indicating higher probability

References

- [1] R. Farrell, O. Oza, N. Zhang, V. I. Morariu, T. Darrell, and L.S.Davis. Birdlets: Subordinate categorization using volumetric primitives and pose-normalized appearance. In *ICCV*, 2011.
- [2] A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei. Novel dataset for fine-grained image categorization. In *CVPRW*, 2011.
- [3] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, 2012.
- [4] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng. Self-taught learning: Transfer learning from unlabeled data. In *ICML*, 2001.
- [5] J. Sanchez and F. Perronnin. High-dimensional signature compression for large-scale image classification. In *CVPR*, 2011.