This CVPR2015 extended abstract is the Open Access version, provided by the Computer Vision Foundation.

Model Recommendation: Generating Object Detectors from Few Samples

Yu-Xiong Wang and Martial Hebert
Robotics Institute, Carnegie Mellon University

All of the modern detectors have in common the same supervised training framework in which a large annotated dataset is required and in which training from scratch is restarted for a new task. In practice, however, it might be difficult to produce enough data for a new task. Moreover, in many applications it is desirable to *rapidly* train a new detector for a new task, something that is typically not possible in any of these current approaches which require expensive training iterations. In this paper, we explore an approach to generating detectors on novel classes *with* few samples *without* conventional supervised training involved.

Our approach is based on the observation that, while it may be hard (impossible) to generate enough training data for a new task, it may be easy to generate a large library of models and evaluate them *off-line* on a large set of tasks. While any *specific* detector cannot generalize well across tasks, in a large-scale library, however, it is likely that one of the library models happens to be tuned with the similar conditions as the new target task. Combining multiple such models into a single new model may perform well on the new task. This would be true especially when considering the shared properties across instances and categories.

This is not sufficient, however, as we are still faced with the problem of selecting the right models out of the library. A naïve approach, which directly evaluates each model from the library on the input task and select the one(s) that perform the best, typically performs poorly because of the limited data available in the input task. The second observation then is that, if the library and the set of tasks are large enough, there might be enough correlation between the models that it is possible to predict the performance of the models on the target tasks by using the *entire* combined experience with the model library and the tasks. This is similar to the approach taken in recommender systems in which a matrix of ratings of items (the analog of our models) by users (the analog of our tasks) is used to predict the ratings that a new user (the target task) would generate, i.e., predict the top ranked items (the analog of the best models for the target task).

The general setup is thus as follows: We first build a large library of object detectors and we record their performance or "ratings" on a large set of detection tasks, which we call the "ratings store". In order to distinguish this phase from the traditional training phase, and by analogy with training hyper-parameters, we call the process of generating and evaluating large library of models *off-line* as "hyper-training". Given a new target task, recommendations are made by trying, or rating, a small subset, called the probe set, of detectors on the input task with few samples, and then using collaborative filtering techniques, e.g., matrix factorization [2], based on that small set of ratings along with the ratings of all the detectors in the ratings store to predict the ratings of all the models on the target task. We then select models based on the recommendations and use them for the new task. This thus becomes our new formulation of object detection, which we term as *model recommendation for object detection* as shown in Fig. 1.

Naturally, the first and most important question to answer is whether collaborative filtering could successfully recommend *correct* models. A second question is to elucidate the role of the different design choices involved in this approach. To answer these questions, we designed a controlled experiment with real detection tasks, large-scale data (namely, PASCAL VOC 2007) and full-scale ratings store so that for any of the target tasks one or several of the n models in the library is the correct model to use or at least a reasonable enough approximation. Experimental results validate our claim that model recommendation is able to select useful models.

The other important issue remains how to generate a large collection of potentially "expressive" models, which could be achieved via *unsupervised hyper-training*. We show how a universal detector library, based on combining *unsupervised* predictable discriminative binary codes (PBCs) classifiers [3] and convolutional neural networks (CNNs) features [1], is generated without a bias to a particular set of categories, and we show how it is particularly effective for populating the ratings store in this setup. We then conducted large-scale detection evaluation across different datasets following the typical R-CNN pipeline [1]. Specifically, we used the pre-trained CNNs structure on ImageNet, generated model library and ratings store on PASCAL VOC 2007, and finally tested the system on SUN 09.

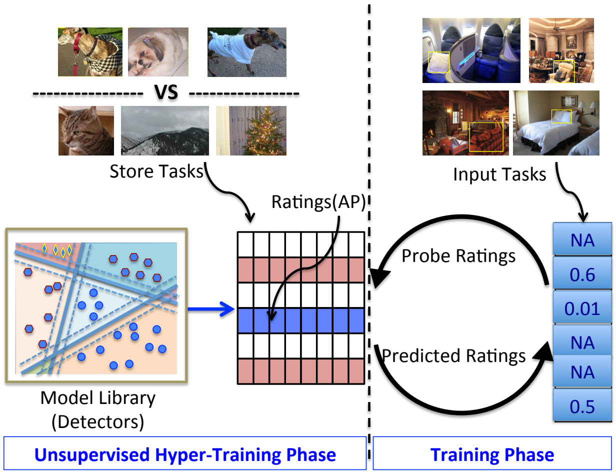

Figure 1: During unsupervised hyper-training phase, a large library of object detectors informative across categories is generated. Their ratings on different detection tasks are recorded to form a ratings store. For a new target task or category and using ratings of a small probe set of detectors on its input task with limited samples, recommendations are made by collaborative filtering. A usable object detector for this new task is thus rapidly generated as single or ensemble of the recommended models.

Experimental results show that ensemble of PBC models works consistently well as the size of the recommended model increases. The result is quite significant if one notices the difference in numbers of training samples: Using only 10 images to select models generated from an out-of-domain dataset our approach is not only better than R-CNN directly trained from few samples, but it is also better than supervised DPM trained from *lots of* in-domain data (hundreds to thousands), comparable to DPM with additional contextual models. Besides, in our case because we have only few target samples and because the large-scale generic sources are weak, direct transfer learning is very noisy. When visualized, the PBC models demonstrate attributes-like behaviors, and a continuous category space is discovered by large-scale model sharing.

We have shown the feasibility of generating new detectors for a new detection task given a large store of ratings of detectors on tasks. This approach has three key advantages of great interest in practice: 1) generating a large collection of expressive models in an unsupervised manner is possible; 2) a far smaller set of annotated samples is needed compared to that required for training from scratch; and 3) recommending models is a very fast operation compared to the notoriously expensive training procedures of modern detectors. (1) will make the models informative across different categories; (2) will dramatically reduce the need for manually annotating vast datasets for training detectors; and (3) will enable rapid generation of new detectors.

[1] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.

[2] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.

[3] M. Rastegari, A. Farhadi, and D. Forsyth. Attribute discovery via predictable discriminative binary codes. In *ECCV*, 2012.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.