OPTIMOL: A Framework for Online Picture Collection via Incremental Model Learning

Li-Jia Li
University of Illinois
jiali3@uiuc.edu

Juan Carlos Niebles
Universidad del Norte, University of Illinois
jnieble2@uiuc.edu

Li Fei-Fei
Princeton University
feifeili@cs.princeton.edu

Abstract

OPTIMOL (a framework for Online Picture collecTion via Incremental MOdel Learning) is a novel, automatic dataset collecting and model learning system for object categorization. Our algorithm mimics the human learning process in such a way that, starting from a few training examples, the more confident data you incorporate in the training data, the more reliable models can be learnt. Our system uses the Internet as the (nearly) unlimited resource for images. The learning and image collection processes are done via an iterative and incremental scheme. The goal of this work is to use this tremendous web resource to learn robust object category models in order to detect and search for objects in real-world scenes.

Motivation

Type the word "airplane" in your favorite Internet search engine, say Google (or Yahoo!, flickr.com, etc.), what do you get? Of the thousands of images these search engines return, only a small fraction of them would be considered good airplane images (15%). It is fair to say that for most of the average users surfing the web for images of generic objects, the current commercial state-of-the-art results are far from satisfying. This problem is closely related to the problem of learning and modeling generic object classes, a topic that has captured the attention of computer vision researchers for the past few years. In order to develop effective object categorization algorithms, researchers rely on a critical resource - object class dataset. A good dataset serves as important training data as well as evaluation benchmark. So far the story is a frustrating one: Users of the web search engines would like better image returns when looking for, say, objects; developers of these search engines would like more robust visual models to improve these results; computer vision researchers are developing the models and algorithms for this purpose; but in order to do so, it is critical to have large and diverse object datasets for training and evaluations; this, goes back to the same problem that the users face.

Currently, there is no optimal solution for this problem. Thus, researchers have to manually select the desire images, which is extremely time consuming. For example, in Caltech 101 (Fei-Fei, Fergus, & Perona 2004), one of the well

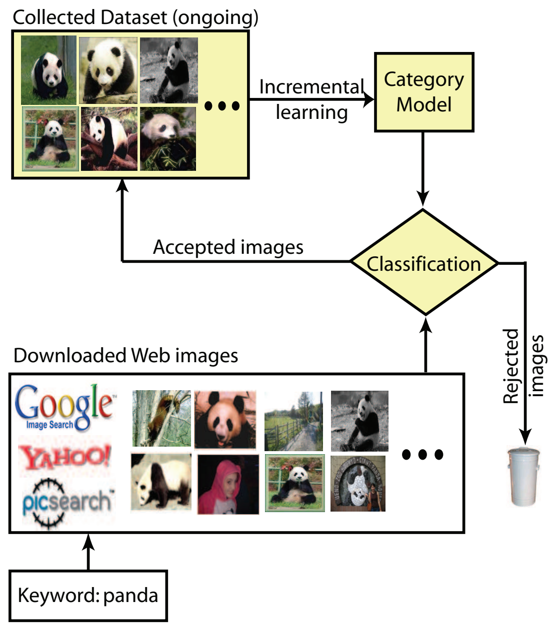

Figure 1: Illustration of the framework of OPTIMOL. Our goal is two-fold: we want to automatically collect an object category dataset from images on the web; to achieve this, we will also train a more and more robust object category model via an incremental learning scheme. Our system starts with a handful of seed images to initialize the object category dataset, say, "panda". This can be done both automatically. With this small number of images, we learn a model of the pandas. Given this model, we classify a subset of images downloaded from the web. If an image is accepted as a "panda" based on some statistical criteria, then we append our existing panda dataset with this new image. We then update our panda model with all newly accepted images – note that the already existing images in the dataset no longer participate in this round of learning. We repeat this process till a satisfying dataset is collected or we have exhausted all downloaded images from the web. In the mean time, the model learnt for the "panda" class can be used as a robust classifier for this object category.

known datasets with 101 categories, the images in each category were selected manually from the downloaded web images. On average, it costs a researcher 2 hours to select 100 images from the downloaded raw dataset. But, what will happen if we want 10,000 or 100,000 images? Recently LabelMe offered an alternative way of collecting datasets of objects by having people uploading their images and labeling them (Russell *et al.* 2005). This dataset is much more varied than Caltech 101, potentially serving as a better benchmark for object detection algorithms. But since it relies on people uploading pictures and making uncontrolled annotations, it is difficult to use it as a generic object category dataset. In addition, while some classes have many images (such as 8897 "car"), others have very few (such as 6 for "airplane"). A few other object category datasets are also used by the researchers, e.g. (Agarwal, Awan, & Roth 2004; CMUFaceGroup). In short, all of them, however, have rather limited number of images and offer no possibility of expansion other than extremely costly manual labor. In addition, while serving as research datasets for researchers, they are not suitable or usable for general search engine users. Our proposed work offers a unified way of automatically gathering data both as a research dataset as well as user queries.

Overview

We would like to tackle simultaneously the problem of automatic dataset collection and model learning for object categories, taking advantage of the vast but highly contaminated data from the web. The sketch of our idea is described as the following. Given a very small number of seed images (in the order of a few to a dozen) of an object class (collected automatically), our algorithm learns a model that best describes this class. Serving as a classifier, the algorithm can pull from the web images that belong to the object class. The newly collected images go to the object dataset, serving as new training data to update and improve the object model. With this new model, the algorithm can then go back to the web and pull more relevant images. This is an iterative process that continuously gathers a highly accurate image dataset while learning a more and more robust object model (Li, Wang, & Fei-Fei 2007). Here, we summarize our approach using the following algorithm:

Algorithm 1 Incremental learning, classification and data collecting

Download from the Web a large reservoir of images obtained by keyword(s)
Initialize the object category dataset with seed images (manually or automatically)
repeat
  Learn object category model with the latest input images to the dataset
  Classify downloaded image using the current object category model
  Append the dataset with accepted images
until user satisfied or images exhausted

In this scenario, an incremental way of visual model learning is essential to make the whole process more efficient. We come up with our OPTIMOL framework as illustrated by Fig.1. We would like to emphasize here that our proposed framework is not limited to any particular object model. Any model that can be cast into an incremental learning framework is suitable for our protocol.

Contribution

In this section, we summarize our contribution. The significance of our approach lies in 2 major aspects:

- We propose an iterative framework that simultaneously collects object category datasets and learns the object category models. The framework has its theoretical base on Bayesian incremental learning.
- We have developed an incremental learning scheme. A major advantage of this learning scheme is to use only the newly added data points (i.e. images) for training a new model. This memory-less learning scheme is capable of handling any arbitrarily large number of images, a vital property for large image datasets.

By introducing the iterative framework of Bayesian incremental learning, we are able to embed the process of image collection and model learning into a mutually facilitating process. In this way, a framework is proposed that can replace much of the current human efforts in dataset collection.

References

Agarwal, S.; Awan, A.; and Roth, D. 2004. Learning to detect objects in images via a sparse, part-based representation. *PAMI* 26(11):1475–1490.

CMUFaceGroup. http://vasc.ri.cmu.edu/idb/html/face/.

Fei-Fei, L.; Fergus, R.; and Perona, P. 2004. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. In *Workshop on Generative-Model Based Vision*.

Li, L.-J.; Wang, G.; and Fei-Fei, L. 2007. Optimol: automatic online picture collection via incremental model learning. In *Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Russell, B.; Torralba, A.; Murphy, K.; and Freeman, W. 2005. Labelme: a database and web-based tool for image annotation.