Identifying Discriminative Features for Fine-Grained Image Classification

Benjamin Poole
Computer Science Department, Stanford University, Stanford, CA
poole@cs.stanford.edu

Abstract

We review the performance of current state of the art fine-grained image classification algorithms using a variety of features on three datasets. We hope to gain a better understanding as to what features are able to capture the fine-grained details of images and how we can best combine these features to achieve the best performance. Additionally, we compare the performance of traditional features with those learned using semi-supervised feature learning techniques to evaluate how well deep architectures can learn highly discriminative fine-grained image details.

1. Introduction

Traditional object classification datasets have focused on objects that are substantially different in their visual characteristics. These datasets generally focus on objects that may be of vastly different sizes, shapes, and colors (e.g. car, plane, chair, person). This focus has led to the development of techniques that are successful at discriminating very different objects, but fail to discriminate similar objects or instances of objects. With the exception of facial recognition, very little work has gone into classifying similar objects such as different types of dogs or cars. These classification tasks rely on very small, fine-grained differences in visual features, such as different ears or tails in dogs. More recent datasets containing humans performing activities and playing instruments has led to new classification techniques, however these techniques tend to rely only on one type of feature (e.g. SIFT or HoG).

In this project, we hope to explore a large set of features for fine-grained image classification, and identify a subset of features that is able to perform well on three datasets: the PASCAL VOC 2010 action classification dataset, the recent People-Playing-Musical-Instrument dataset, and the Caltech-UCSD Birds 200 dataset. Initially, we will look at some standard features used in computer vision: SIFT, shape-based templates, contextual features, HoG, Local Binary Patterns (LBP), wavelets, LLC, and color histograms. We intend to evaluate these features using average precision as the metric on a variety of classifiers, including SPM, Multiple-kernel learning, SVMs, and random forests with discriminative decision trees [2]. Our hope is that a certain subset of these features will provide improved performance across all datasets and classifiers, while some features may provide little or no useful information for fine-grained image classification. Furthermore, we hope to beat the state-of-the-art in activity recognition by utilizing a combination of features (instead of just SIFT found in [2]). Through the analysis and review of this variety of techniques, we hope to gain the intuition to develop a new feature set that is able to capture fine-grained information contained in images. Time permitting, we hope to compare these hand-designed discriminative features to semi-supervised feature learning techniques to determine the effectiveness of semi-supervised feature learning at capturing fine-grained discriminative information [1].

References

[1] J. Ngiam, Z. Chen, P. Koh, and A. Y. Ng. Learning deep energy models. In *International Conference in Machine Learning*, 2011.

[2] B. Yao, A. Khosla, and L. Fei-Fei. Combining randomization and discrimination for fine-grained image categorization. In *The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, June 2011.

2. Appendix

This project is also my course project for CS229 as well as my rotation project. The computer vision component will focus on features while the machine learning component will be focused on classification and semi-supervised feature learning.