Outdoor Scene Classification

Sagar Chordia
Stanford University
sagarc14@stanford.edu

Romil Verma
Stanford University
vermar@stanford.edu

Abstract

This paper aims to improve results of scene classification using Convolutional Neural Networks applied to a mobile setting. We apply transfer learning on pre-trained models in the Caffe framework and evaluate system on mobile context using DeepBelief SDK.

1. Introduction

Outdoor scene classification generally utilizes the global spatial properties of the image without emphasis on specific object detection [1]. There are over 2 million outdoor scene images classified providing a fertile training set for classification systems. This paper focusses on outdoor scene recognition in images using transfer learning from pre-trained models on the Caffe framework [4], further extending the models on newer scene datasets and moving the evaluating/test criteria to mobile setting using DeepBelief SDK [5].

2. Problem Statement

2.1. Baseline

We intend to use pretrained Places-CNN model from MIT [2] on Model-Zoo [6] as our baseline. The authors of Places-CNN have published two models, one trained on Places database and other on combination of Places database and ImageNet. The published accuracy on validation set of Places-only model is 56.2% and for hybrid model is 52.3%.

2.2. Dataset

The dataset we will use for further training of the models will be on subset of the scene-centric SUN database [3]. All train, validation and test images will be appropriately pre-processed to match the inputs to the baseline learning model. The Places database contains 477 single-level categories that are subset of the SUN database categories. The SUN database has 908 scene categories but is in a hierarchical structure, for our purposes we will collapse the SUN hierarchy into the 477 top level categories to match the places dataset on which baseline has been trained. Backup approach would be to only choose the matching 477 categories of SUN dataset for further model learning.

3. Technical Approach

3.1. Implementation

As a first iteration we used the CNN architecture provided in [1] and tried to train it using caffe on local mac machine in CPU only mode (no GPU available on local machine). But time taken to train the network in CPU mode was very large and hence infeasible for development. So we shifted to rye machines with GPU support. We evaluated the performance of given model snapshot and got similar accuracy as reported in paper.

3.2. Features

We intend to train new features on the baseline to improve accuracy. Things we plan to try include HOG, Color histogram, PCA, Prewitt filtering processing. We plan to evaluate effect of drop-out and data augmentation (flips, crop, contrast, tint) on validation accuracy.

Classification Hierarchy Variation: Given SUN dataset has 908 categories in hierarchical structure while Places dataset has only 477 flat categories, one approach could be to collapse the SUN category hierarchies to its top levels to match the places dataset. This would give more robust dataset for each class and the collapsing will be valid since the baseline training was on Places, which has a subset of the classes derived from SUN. Another benefit of collapsed hierarchy would be improved usability in the mobile application setting (as less classification) and more varied hidden features extracted per class given more training images.

Model Ensembles: Performing model ensembles of various CNNs we train on top of provided models for these datasets by Caffe framework would help give consistent classification labels and higher accuracy.

3.3. Evaluation

Given this validation result, our test accuracy will be obtained by transfer learning the model parameters to mobile setting DeepBelief SDK and applying on real images. We are currently exploring the efficacy of transfer learning and testing on mobile setting and via DeepBelief.

The evaluation dataset should have same number of images per class, if unfeasible to take mobile images for all such classes we can partition the SUN dataset initially and extract subset of images per class and set them aside as test images.

For evaluating the performance on mobile application, we add Softmax classifier layer on the learned model to make real-time predictions.

4. Preliminary Results

Our current milestones have been to achieve baseline accuracy using Caffe framework models on Places dataset and get DeepBelief Android application set for extension. We observed about 50% accuracy with CNN available in caffe on Places database as reported in [2].

The next milestones would be to train with suggested features on SUN dataset, improve validation accuracy, evaluate the test accuracy and transfer the learned model to the mobile application to perform real-time image classification.

References

[1] A. Quattoni, and A.Torralba. Recognizing Indoor Scenes. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[2] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. "Learning Deep Features for Scene Recognition using Places Database." Advances in Neural Information Processing Systems 27 (NIPS), 2014..

[3] "SUN Database: Large-scale Scene Recognition from Abbey to Zoo". J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba. IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[4] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. `http://caffe.berkeleyvision.org/`, 2013

[5] DeepBelief SDK `https://github.com/jetpacapp/DeepBeliefSDK`

[6] Caffe Model Zoo `https://github.com/BVLC/caffe/wiki/Model-Zoo`