FACULTY OF COMPUTING AND INFORMATICS

TDS 3651 Visual Image Processing

TRIMESTER 2 2016/2017

Assignment 2

Report

Lecture Section : 01
Tutorial Section : 01

for:

Dr. John See Su Yang

from:

Student ID	Name	Email Address	Phone No
1141125087	Hii Yong Lian	yonglian146@gmail.com	0164111005
1122702848	Lee Zhen Yong	bruceoutdoors@gmail.com	0163188854

Abstract

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Such a system for food would be useful as one can search for similar food without knowing it's name. In this work, we utilise a deep convolutional neural network to perform retrieval, and demonstrate that it performs exceptionally well in the provided dataset.

Introduction

In this work, we take advantage of the availability of image labels to perform supervised learning. To do this, we chose SqueezeNet, a recently introduced convolutional neural network (CNN) for classification. SqueezeNet has been shown to perform slightly better than AlexNet while using 50x less parameters. Without pruning or weights compression, this equates to a weight file of just 5mb, while still providing excellent results.

Description of Methods

Data Preprocessing

- Images were first resized to a standard resolution of 224x224
- The mean along each of the color channels R,G,B was subtracted from each image

Image Classification

Overall Architecture

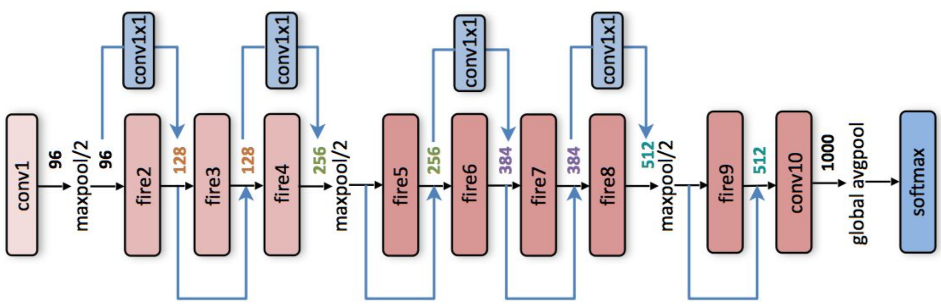


Image classification was performed end to end using Squeezenet, a convolutional neural network.

The network is first pre-trained on ImageNet.

Then, it is fine tuned on the food dataset provided for 10 epochs, using the Adam optimiser with an initial learning rate of 0.001. We found that loss tend to spiral out of control if the learning rate is set too high, and opted to use the highest learning rate that did not exhibit such problems.

Image Retrieval

After training the model for classification, we first precompute the 10-class softmax output for every image in the dataset via a feedforward pass. So for each image, we will have a score for each of the 10 categories (a float between 0 to 1). These will be our features; we store them in a pickle that will be used for retrieval later.

To perform retrieval, the input image is first classified using the model to calculate our features. To compare similarity between our database of images, this features will be compared to all the precomputed features for every image in our food database via L2 distance, given by:

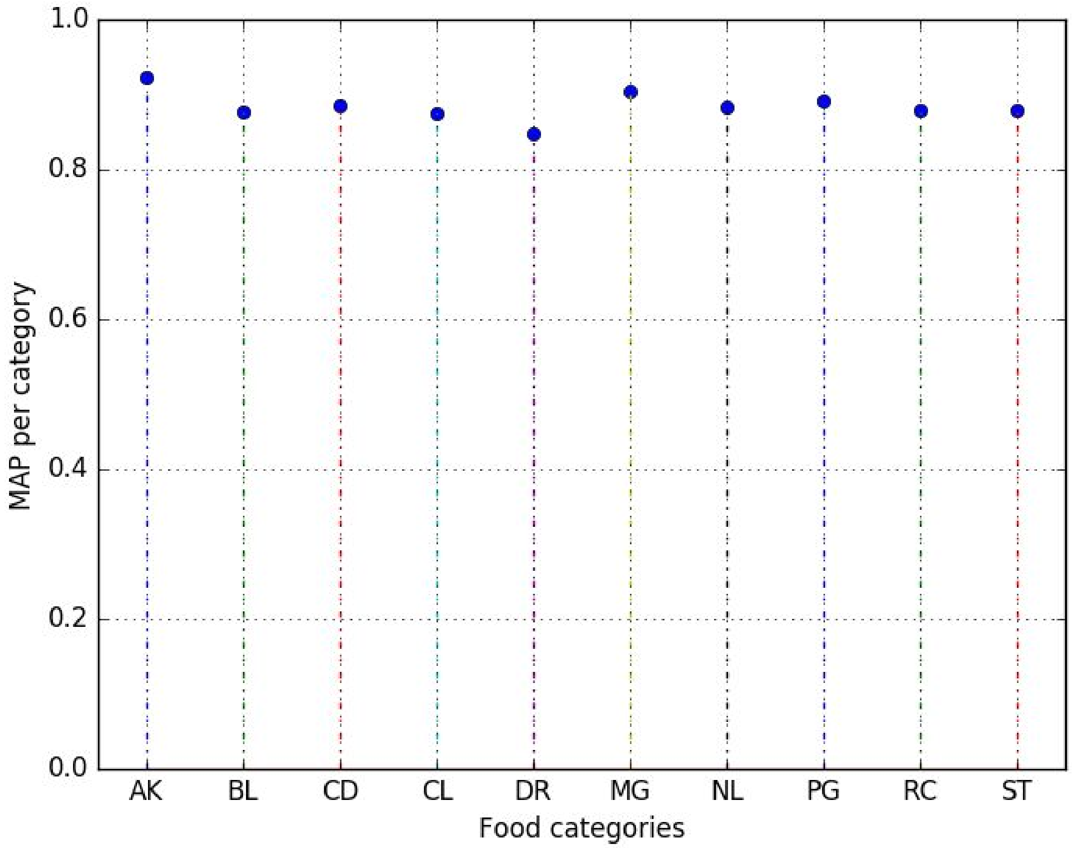
$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

Finally, the top-n images with lowest differences in output was selected as the images to be retrieved.

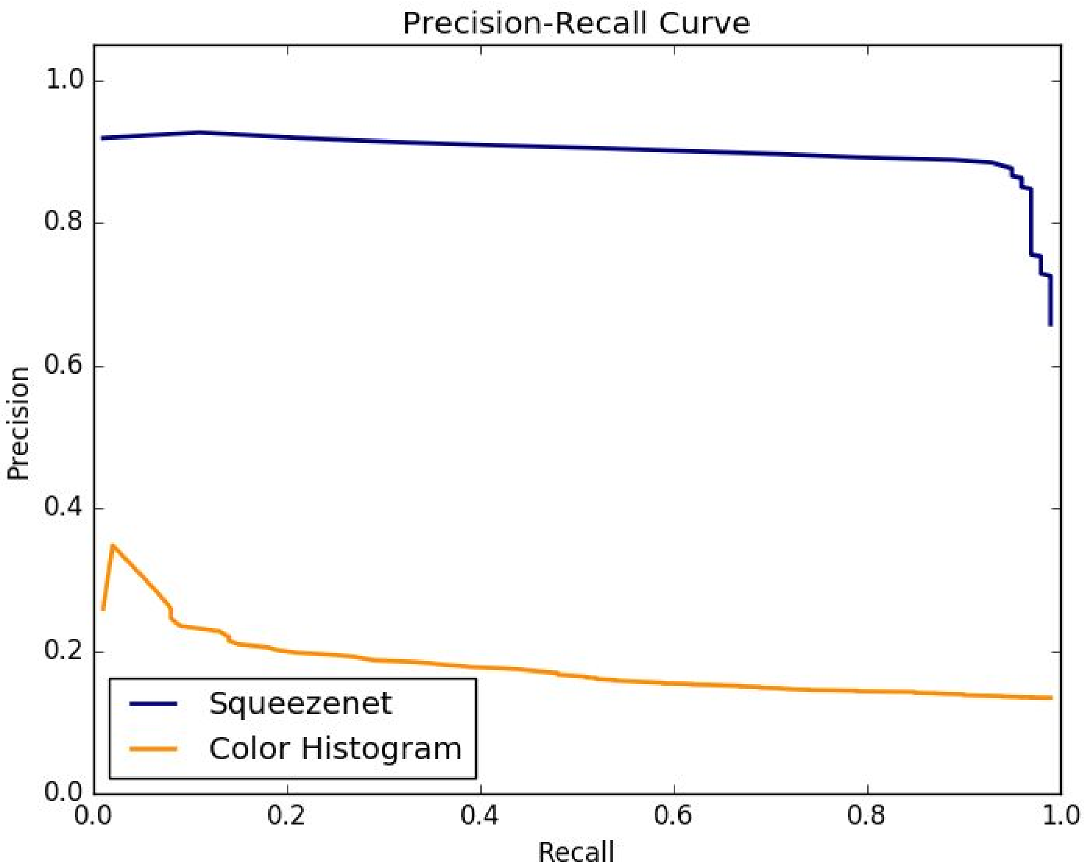
Results & Analysis

Mean Average Precision, MAP@100: 0.8847
Recall Rate@100: 0.9200



The approach taken obtains excellent performance on all food categories. The model obtains a mean average precision score of 0.8847 for a retrieval size of 100, as well as a recall rate of 0.9200.

Precision Recall Curve



It can be seen that squeezenet outperforms the baseline of color histogram considerably.

Suggestions for Improvement

Tweaking Learning Rate

- A better minima can probably be found with a better initial learning rate.

Image Classification

- **More Labeled Data**
  As with all problems involving convolutional neural networks, the more data points there are the better the overall performance.
- **A larger model**
  With more computational overhead, a larger model could be used to obtain even better accuracy

Unsupervised Learning

- Currently we are limited to 10 categories of images. This would not prove to be flexible should we introduce new categories (i.e. more exotic foods, or images other than food).
- Unsupervised learning methods would allow the model to retrieve similar images for input images of unknown categories and would therefore be more robust.