Deep Dish : Deep Learning on a Platter

Abhishek Goswami
Microsoft
Redmond, WA
agoswami@microsoft.com

Haichen Liu
haichen.sjtu@gmail.com

Abstract

We consider applying deep convolutional neural networks on images of food dishes. Food items have unique characteristics - they come in different colors and shapes, can be clustered into groups (e.g. fruits, vegetables), and food items can be combined in several ways to prepare a meal etc. This makes images of food dishes particularly suitable for visual recognition tasks.

1. Introduction

This project aims to use deep learning on images of food dishes. Currently there are model zoos e.g. Caffe Model Zoo [2] where people submit deep learning models and data from a wide range of application domains. We want to apply the principles of deep convolutional networks on images of food. Food images are unique: there are multiple cuisines from around the world, each food item has a unique color, size, shape and texture, and food items can be combined in several ways to prepare a meal. Being able to use artificial intelligence (and Deep Learning in particular) on food images has the potential to revolutionize the field of dining, promote healthy eating, prevent food waste etc.

To that effect, we are working on two problems. In the first problem, we want to classify food dishes using Convolutional Neural Networks. We formulate this problem as a standard classification task with one class per image, i.e. given an image of a food dish, we want to predict what dish it is. For the second problem, we want to detect the different food items in the image, i.e. given an image of a dish with chicken wings and celery (see Figure 1), we want to detect and tag each item separately. We formulate the second problem as a CNN based detection task.

2. Related Work

Deep Convolutional Neural Networks have been shown to be very useful for Visual Recognition tasks. AlexNet [4] won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012, spurring a lot of interest in using Deep Learning to solve challenging problems. Since then Deep Learning has been used successfully in several fields like Machine Vision, Facial Recognition, Voice Recognition, Natural Language Processing etc.

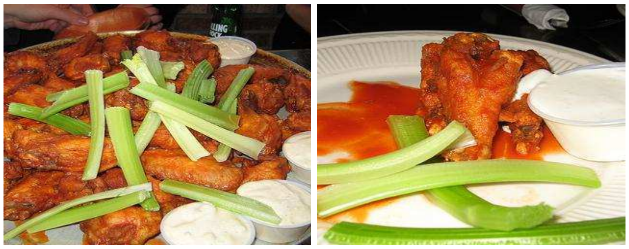

Figure 1. Images of chicken wings dish with celery.

3. Methods

We plan to use Convolutional Neural Networks for an application based project.

4. Dataset and Features

We provide some details about our dataset below.

4.1. Dataset Details

For the milestone, we are using a dataset of food dishes collected from ImageNet [3]. We plan to collect more images using the Bing Image Search API [1] in the future.

4.2. Class Label Distribution

Currently our dataset has 15 classes. Table 1 shows the class label distribution of the dataset. We do note that currently the distribution of images across the classes is not uniform. We plan to address this in the following weeks.

4.3. Preprocessing Steps

We re-sized all of our images to 32*32*3, by cropping the image if it was larger than the standard size, or

Food Name	Num of Images
Clam Chowder	723
Jambalaya	588
Fish Stick	175
Cannelloni	230
Buffalo Wing	340
Lamb Curry	347
Fish And Chips	803
Pepperoni Pizza	773
Eggs Benedict	1124
Scotch Egg	388
Sashimi	818
Tempura	1008
Spaghetti	523
Beef Wellington	279
Lobster Thermidor	76

Table 1. Class Label Distribution

adding black padding if it was smaller. We found that ImageNet [3] tends to contain spurious images; it has images which clearly do not belong to the class (e.g. images of human beings in the ’Fish And Chips’ class). We removed some of these spurious images as part of data cleanup. Also, care must be taken for images which do not contain the RGB channels. Our training pipeline hit some issues when we encountered images without channel information. We subsequently used a fix to filter out such images.

5. Experiments/Results/Discussion

In this section we discuss our experiments and results. While we ran multiple experiments, we are only providing details for the experiments which look best so far.

5.1. Classifying food dishes using CNNs

In the first problem, we want to classify food dishes using Convolutional Neural Networks. We formulate this problem as a standard classification task with one class per image. We use accuracy as our evaluation metric.

We split our dataset randomly into 3 disjoint sets: Train(70% approx.), Validate(20% approx.) and Test(10% approx.). Table 2 provides a count of the number of images in each set.

5.1.1 Experiment

We are currently using a very simple model which has the following architecture:

- 7x7 Convolutional Layer with 32 filters and stride of 2
- ReLU Activation Layer
- Affine layer with 15 outputs

Dataset	Num of Images	Accuracy
Train	5756	0.18
Validate	1619	0.2
Test	819	0.0842

Table 2. Dataset Details

Also, we are using the Hinge loss function in conjunction with the Adam optimizer to minimize loss.

5.1.2 Results

Figure 2 shows the reduction in the loss over multiple iterations in the first epoch. We see the loss reduces very sharply in the beginning, and then flattens out gradually. Currently we are running only a single epoch to train our model.

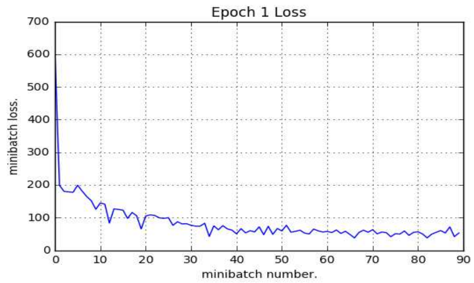

Figure 2. Reduction in loss over several mini-batches in the first epoch

Table 2 shows the accuracy metric for train, validation and test sets.

5.1.3 Discussion

We observe that train (and validate) accuracy is much higher than the test accuracy. This suggests that we are probably overfitting the training set. We plan to investigate this in depth going forward, including the use of (a) dropout (b) L2 weight regularization (c) increasing size of the dataset etc. Any feedback on how to improve the test accuracy is highly welcome.

5.2. Detecting individual food items in an image.

For this problem we are planning to use some standard CNN based detection methods such as SSD [5], Fast-RCNN [7] and YOLO [6]. These approaches typically detect bounding boxes of interest in an image, and we want to leverage these techniques to detect the individual food items in an image. For this milestone we were not able to make much progress on this problem. We plan to work on this in the coming weeks.

6. Conclusion/Future Work

For the first problem of classifying dishes, we have made some progress in our workflow : data collection, preprocessing and building simple models. As future work, we need to investigate why we are seeing low accuracies in the test set, and explore techniques to improve the same.

For the second problem of detecting individual food items from images, we plan to start looking into CNN based detection methods.

References

[1] Bing image search api. https://azure.microsoft.com/en-us/services/cognitive-services/bing-image-search-api/.

[2] Caffe model zoo. http://caffe.berkeleyvision.org.

[3] Image-net. http://www.image-net.org/.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In *European Conference on Computer Vision*, pages 21–37. Springer, 2016.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.

[7] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.