CS230

What Are You Eating? Food Detection Through CNN

Luis Govea.*
Department of Computer Science
Stanford University
lgovea@stanford.edu

Abstract

As we move towards the future, humanity has observed an increasing trend in the utilization of technology on a daily basis. Nowadays, one of these major technological trends is the tendency for mobile users to upload pictures of food items to social networks. As a way to help the tagging function in multiple social networks, I have decided to develop a neural network that is able to detect multiple kinds of food items. The long-term motivation of this work is to deploy a mobile application that is able to increase the number of labeled pictures in the Internet for facilitation in data collection for further research. I deployed two different convolutional neural networks with different hidden layers in order to access the optimal efficiency of the project. Through the fine-tuning of certain hyper-parameters, we were able to obtain highly accurate labels. Further testing will be done to increase this accuracy.

1 Introduction

Millions of pictures are updated on the Internet on a daily basis. The lack of a tag/label makes the classification of these pictures a hard task for search engines. Through the ability to add labels to uploaded pictures we could facilitate the search process for multiple users as well as researchers looking to create a database. I decided to pursue this topic due to the fact that I, in fact, also have a tendency to upload pictures of my food items to my social pages. I believe that through the ability to provide a label for uploaded images I could increase the user's experiences with social network utilization. In order to solve this problem, I create a convolutional neural network that is able to detect and classify pictures of different food items. The inputs for this neural network are pre-classified images that are 512x512 pixels in size. After feeding this to the network, a possibility distribution is output that can lead to a decision as to what food item is being tested, indeed classifying pictures into various food groups.

2 Related work

The problem of image classification is one of the fundamental example of deep learning algorithms. Because of this, there has been a lot of research regarding optimizations of algorithm that enable image classification. Before starting this project, I decided to investigate multiple research papers regarding image classification. Related Works # 1,2 were utilized as key background components to gain a deeper understanding on how convolutional neural networks operate in the real world and how to make them more efficient. They granted me with multiple strategies to explore the efficiency of my

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

neural network. It was through these papers that I was able to conclude a lot of my hyper-parameters and functions used as I believe they would grant me the biggest recognition percentage. Regarding my topic at hand, there has been a lot of research into food images classifications. I utilized Related Works # 3,4 as a reference for my own project. These discussed several strategies into improving the ability to classify food items into specific categories. Interestingly enough, Related Work #4 utilized the same database that I used during my project. Nonetheless, it was Related Work #5 that surprised me the most. The utilization of a inception module to address the challenge of image classification seemed like a great way to ensure a high percentage of efficiency in the task at hand. GoogLeNet's training framework with MaxPool functions as well as drop-out provided for a pretty elaborate pathway to ensure high accuracy. Although I was not able to fully understand the conceptual ideas behind each node, the overall strategy to this architecture surprised me a lot. I believe that with time and effort a similar pathway could be produced for my work.

3 Dataset and Features

The Food-101 dataset utilized in this project consisted of 101,000 pre-labeled RGB images of 101 different food items. Each class of food item consisted of 1000 pictures in different settings/lighting. As a way to improve the efficiency of this dataset to represent the real world, images were not cleaned and sometimes mislabeled- indeed adding noise to the neural network. Before beginning the coding process, I decided to augment this dataset to include pictures that were rotated/cropped/inverted as a way to supply more data to the network. This augmented data also included the addition of some intentional noise to increase replication in real life. In order to normalize the pictures, I decided to resize all of the samples to a standard size of 128x128 pixels. The final dataset was split into a 75-25 between train set and test set. At first, I tried to implement including a Dev set into the project, but at the end concluded that it was better to simply omit this set altogether.

Figure 1: Example of input data for project.

4 Methods

I decided to explore the functionality of the project in convolutional networks with different numbers of hidden layers. Each test consisted of a multi-layered neural network. Each layer consisted of a different number of filters to apply to the image along with the inclusion of a padding function. I decided the optimal activation function to be utilized was a normal ReLU function for each layer. Furthermore, I decided to implement a Max-Pooling function in order to increase the processing of the images. The size of the pool function was 3x3 as I believed this would be enough to gather the main details of the picture. Moreover, I decided to implement a dropout function in each layer. After some analysis, I was able to conclude that the best value of this dropout function was 0.25. This enable me to obtain the higher accuracy possible for the given neural network. The final layer of the neural networks was a Softmax function that enabled for classification of images into different categories to see into which bin the picture would belong.

$$L(X,Y) = -\frac{1}{n}\sum_{i=1}^{n} y_i \ln(x_i)$$

Figure 2: Loss Function utilized in this project

$$\Delta\theta_t = -\frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t$$

Figure 3: Optimizer Function utilized in this project

The loss function utilized in this project was a fundamental categorical cross-entropy function in order to ensure that the multiple categories we're well represented. Moreover, I decided to utilize a RMS prop optimizer since, after analysis, it provided me with the biggest values as compared to other forms of optimizers.

5 Experiments/Results/Discussion

Throughout this whole process, I decided to try out multiple hyper-parameters to see which one would give me the biggest accuracy levels. I ended up deciding to utilize a 128 mini-batch size since this would provide me with the biggest precision in image detection. Furthermore, I decided accuracy to be my primary metric of search since I believed this would make the model more applicable to real-life applications and detection of food items. After testing the model in different number of hidden layers, I concluded that a 15-hidden layer created the highest accuracy level for the classification problem. Nonetheless, further testing on the program could revel contradicting information.

	Training Accuracy	Test Accuracy
7 Conv CNN	0.74	0.62
15 Conv CNN	0.70	0.64

Figure 4: Accuracy Table

We were able to see a decrease in training accuracy as the number of hidden layers increased. We believe this is due to the fact that feature extraction became more and more abstract as the number of hidden layer increased. Nonetheless, this lead to a higher ability in recognition during the testing phase. Throughout this whole experiment, some images were mislabeled by the program. I found that most of the pictures that were mislabeled consisted of similar-colored food items and extreme zoom-ins.

Figure 5: Mislabeled image. Confused with chocolate cake

Furthermore, I believe that a small dataset per food category represented a big problem for the accuracy in food item detection. The ability to expand on the number of items per categories could further improve the number of true labels the program outputs.

6 Conclusion/Future Work

Through the utilization of a convolutional neural network, we were able to obtain a high efficiency in image classification. I discovered that a fifteen layered convolutional neural network was more efficient at the task than any number less that this value. I believe that this was due to the trend that I experienced while working on this project such that a higher number of layers led to a higher degree of accuracy. Since we are processing the images through more and more filters, several different aspects of the foods are better recognized leading to a better neural machine capable of predicting the

food items. Although accuracy levels were not super high (60%) I believe that further modifications to the pathway could ensure a higher value for this project. For future work regarding this project, the main course of action is to recruit more data in the provided categories and new categories as a way to ensure that our model is applicable to real life. Moreover, I believe that there could be a further fine-tuning in hyper-parameters as a way to increase the efficiency of my project. There also exist the possibility of exploring different architectures to see if another architecture could be more fitting for the necessary goal. Furthermore, I would like to create a mobile application that is able to utilize this same algorithm so that image classification can be utilized as people utilize their smart-phones to post food pictures on the social networks.

References

[1] Krizhevsky, Alex, et al. “ImageNet Classification with Deep Convolutional Neural Networks.” Communications of the ACM, vol. 60, no. 6, 2017, pp. 84–90., doi:10.1145/3065386. [2] Giacinto, Giorgio, and Fabio Roli. “Design of Effective Neural Network Ensembles for Image Classification Purposes.” Image and Vision Computing, vol. 19, no. 9-10, 2001, pp. 699–707., doi:10.1016/s0262-8856(01)00045-2. [3] Kagaya, Hokuto, and Kiyoharu Aizawa. “Highly Accurate Food/Non-Food Image Classification Based on a Deep Convolutional Neural Network.” New Trends in Image Analysis and Processing ICIAP 2015 Workshops Lecture Notes in Computer Science, 2015, pp.350–357. [4] Singla, Ashutosh, et al. “Food/Non-Food Image Classification and Food Categorization Using Pre-Trained GoogLeNet Model.” Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management - MADiMa 16, 2016, doi:10.1145/2986035.2986039. [5] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9. doi: 10.1109/CVPR.2015.7298594