Deep Learning Food Recognition Model Proposal

Hou, Chuyi Li, En Xu Shentu, Chengnan

June 30 2019

Contents

1 Introduction

Our team is pleased to submit this paper to propose a machine-learning food-recognition model. This project is motivated by the (shocking statistics behind the junk food consumption in North America.) real consequences caused by having unhealthy diet.

"Unhealthy diet contributes to approximately 678,000 deaths each year in the U.S., due to nutrition- and obesity-related diseases, such as heart disease, cancer, and type 2 diabetes. In the last 30 years, obesity rates have doubled in adults, tripled in children, and quadrupled in adolescents."[1]

The goal of this model is to identify what kinds of food are consumed by the person from a overview photo of the meal. In particular, we will use the model to target the western-style breakfast meals, for example, eggs, tomatoes, and lettuce. Supervised learning approach will be the taken in this project as we use labelled images with corresponding categories to train the model. The model's weights and biases will be tuned by the optimizer according to the cross entropy loss function. We expect the trained model to recognize various kinds of foods in the photo and report them back to the user. Machine learning is a reasonable approach to this project since the model could be fed with the RGB values of each pixel of the image and output how likely each kind of food appears to be inside.

Simply recognizing kinds of food in an image is not enough to solve the unhealthy diet problem. Essentially, by having this base functionality, future developments will implement recommendation feature to accompany the food recognition feature. Thus, the model would be able to make diet recommendations for meals later in the day based on the person's current meal.

As for the project, our goal is mainly building the machine learning model which can achieve the food recognition feature. The recommendation feature is proposed and it should be implemented in the future.

2 Background and Related Work

Background research has been conducted prior to proposing this model. We were able to find some food recognition models on the market; however, most of them could only make a decent prediction when there was only one dish in the photo. They could not handle the situation where multiple foods are present in a single picture frame. In particular, we will discuss about two of the related projects that have similar functionality.

2.1 NutriNet

NutriNet is a deep convolutional neural network architecture that has been used for food and drink image detection and recognition. This model was developed in 2017 and had been trained to recognize 520 different categories. The authors provided that the model achieved a classification accuracy of 86.72% on the recognition dataset, along with an accuracy of 94.47% on a detection dataset. With a real-world dataset acquired by smartphone cameras, the model was able to achieve a 55% accuracy[2].

Currently, this model is being used for the PD Nutrition Dietary-assessment application for Parkinson's disease patients[2].

2.2 Image AI

Image AI is a python library developed by DeepQuest AI regarding deep learning and computer vision. This library is a very powerful tool in terms of detecting objects and extracting features from the image[3]. This library uses the following 4 architectures: SqueezeNet, ResNet, InceptionV3 and DenseNet. Among these 4 models, there exists trade-offs between the processing time and accuracy. For example, SqueezeNet gives a moderate accuracy with the fastest prediction time, while DenseNet outputs the highest accuracy with a slower processing time [3].

3 Data Processing

Data that are used to train, validate and test the model will be in the forms of RGB values of each pixel in an image. In this section, the methods of data collecting, cleaning, and splitting will be discussed.

3.1 Data Collection

A list of 67 categories[4] of popular western style breakfast foods are used for out recognition task, including muffin, oatmeal, omelette, pancake, etc. Then a python script will go through this list, search and download the first 200 images from Google Custom Search API for each category. Afterwards, 67*200 images will be saved into 67 folders to be prepared for data cleaning. This method of data collection provides flexibility in terms of the number of categories, specific categories, number of images, and more.

3.2 Data Cleaning

Prior to data splitting, we have to make sure each image has the same size. We will be using an python image resizing function[5] to resize all images to 512 by 512 pixels.

3.3 Data Splitting

We will be randomly selecting 150*67 collected and processed images as our training dataset for computing loss functions and tuning parameters of the model. In addition, data augmentation techniques are applied to generate more training data to avoid overfitting. In particular, we will apply rotation, vertical and horizontal flip, and vertical and horizontal shift to all images.

The rest of the data collected will be equally divided into the validation and test dataset. The purpose of the validation data is to tune hyperparameters, such as the batch size, learning rate, and number of layers. At the last phase of the project, the test dataset will be used to compute the model's test accuracy.

4 Architecture

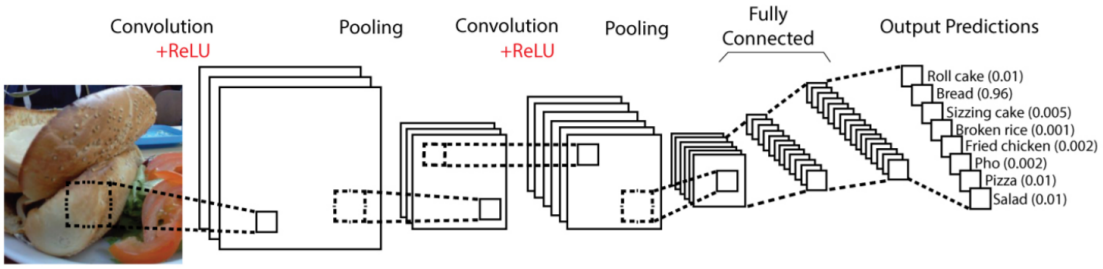


Figure 1: The idea of our model's architecture, taken from [6]

Supervised Machine Learning with convolutional layers and fully-connected layers

Input 512 x 512 image file or (some dimension of the embedding produced by a pretrained neural network). An example image input is shown in Figure 2

Computer Vision We would use the help from computer vision to cut the image file into many smaller pieces containing possible categories of food. Then the model would take in those smaller pieces with labels as actual input to the network.

Output An 1D tensor containing the probabilities corresponding to each possible categories. An example output is shown in Figure 3

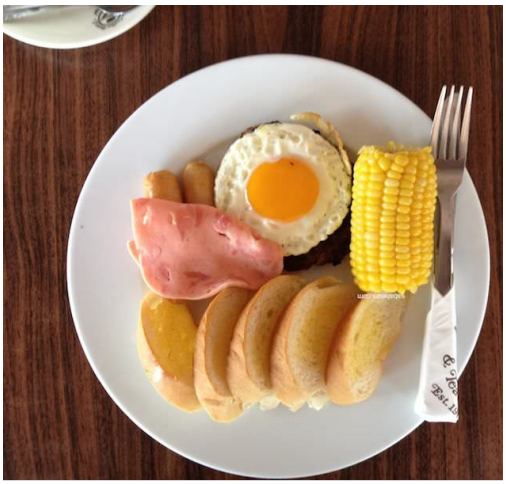


Figure 2: A typical input

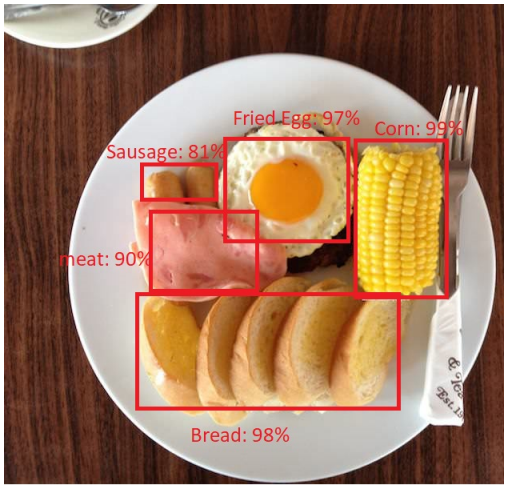


Figure 3: A typical output

All categories of food with probabilities higher than a fix amount (e.g. 80%) will be produced as the food recognized for the images.

In the above example, the model should also print "The food that are most likely be presented in the picture is the following: corn, fried egg, bread, meat, sausage"

5 Baseline Model

We will apply a simple algorithm which takes the average RGB value of the input image and compares to the values of categories from the training set, and take the closest categories as its prediction.

6 Ethical Considerations

Figure 4: Boiled Egg

Figure 5: Tang Yuan

Our target food is particularly appeared in western breakfasts. In this case, our model would unlikely recognize food in other continents. For instance, our model would likely classify the food in figure 5 as egg, while it is never the food shown in figure 4, it is a traditional Chinese food called Tang Yuan.

Such bias can be reduced by expanding our data set to include more categories of foods, but we are working with the limited data set given the scope of this project.

Other than the recognition range of our model, we believe that there is no other major biases contained in our model. Moreover, the task itself, recognizing food, should be seen neutrally by people.

7 Project Plan

This project will start on July 1, 2019 and is expected to be complete by August 10, 2019. Guided by the Critical Path Method, the project is divided into 4 phases: Data pre-processing, model training, testing, and GUI building.

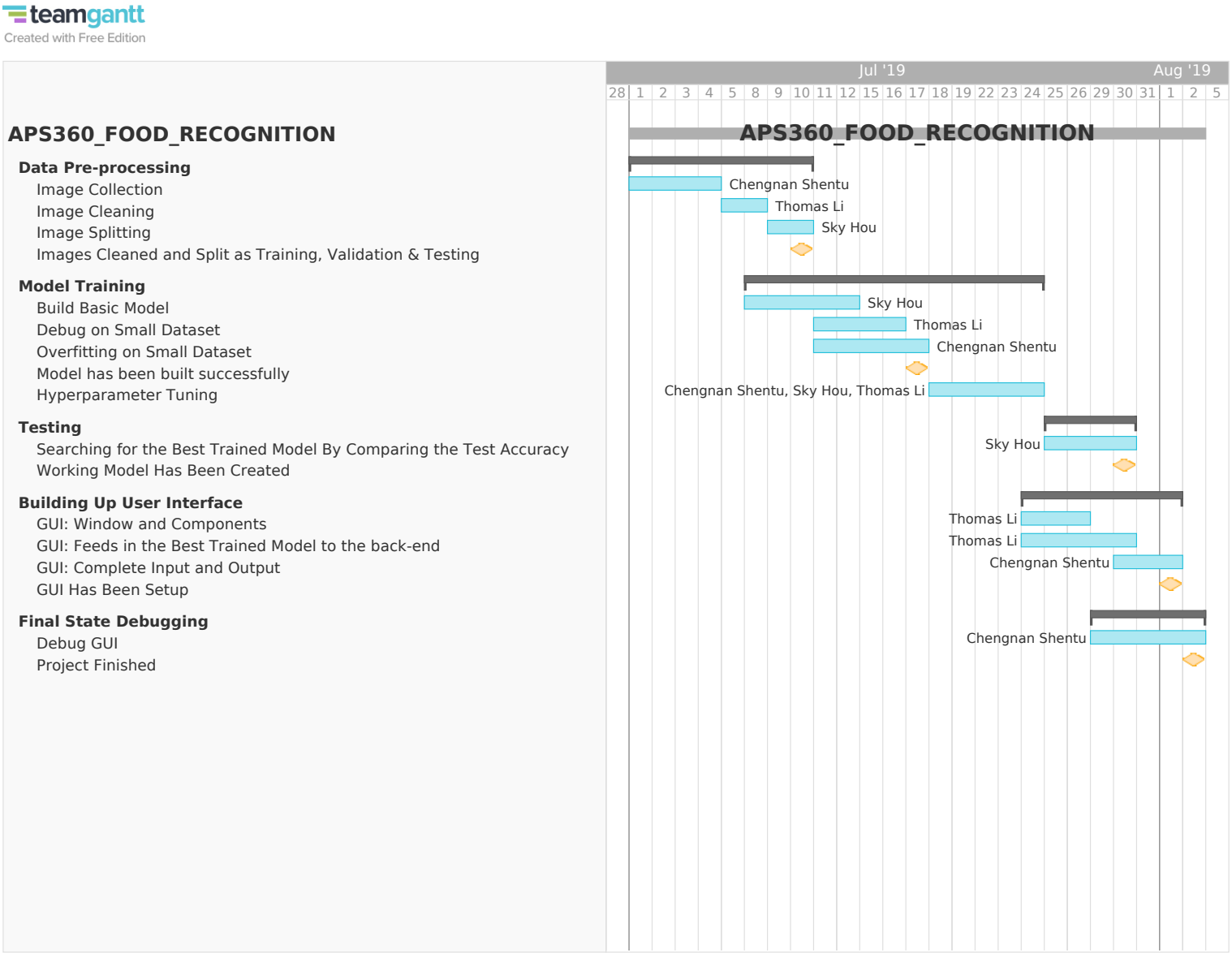
Specifically, we break down the project into the following steps:

- Image Collection
- Image Cleaning
- Splitting Data into train/validation/test sets
- Building the neural network
- Debugging neural network using small data set
- Hyperparameter tuning including modification to the neural network architecture.
- Testing and verification of the model

7.1 Project Schedule

We will be having weekly meetings from 7pm to 9pm on Thursdays and using Wechat to communicate with each other. A team Gantt chart acts as a detailed scheduled plan and is presented below:

Figure 6: Gantt Chart



8 Risk Register

Possible risks within this project will be discussed in this section as well as the associated likelihoods, impact and solutions.

8.1 Inefficient Dataset

Inefficient dataset has low likelihood of occurrence and high impact on the project. Settings from the data collection section can modify the dataset easily.

8.2 Long Training Time

Long training time has a high likelihood of occurrence, and it has high impact on the project. The solution would be to apply techniques such as drop-out to reduce training training time. In the worst case scenario, transfer learning will be considered.

8.3 Unsatisfactory Performance

Unsatisfactory performance has a moderate likelihood of occurrence and high impact on the project completion. Hyperparameter tuning is supposed to improve the performance. Modifying the model architecture or using transfer learning can also be considered to improve performance.

9 Miscellaneous

We will be using Github to store and sync our project. The link to the repository is: https://github.com/nbjameslee/APS360_FOOD[4]

References

[1] "Why good nutrition is important." [Online]. Available: https://cspinet.org/eating-healthy/why-good-nutrition-important

[2] S. Mezgec and B. K. Seljak, "Nutrinet: A deep learning food and drink image recognition system for dietary assessment," *Nutrients*, vol. 9, no. 7, p. 657, 2017.

[3] OlafenwaMoses, "Olafenwamoses/imageai," Jun 2019. [Online]. Available: https://github.com/OlafenwaMoses/ImageAI

[4] E. Li, C. Hou, and C. Shentu, "Aps360food." [Online]. Available: https://github.com/nbjameslee/APS360_FOOD/blob/master/food_list.txt

[5] P. Canuma and P. Canuma, "Image pre-processing," Oct 2018. [Online]. Available: https://towardsdatascience.com/image-pre-processing-c1aec0be3edf

[6] B. T. Nguyen, D.-T. Dang-Nguyen, T. X. Dang, T. Phat, and C. Gurrin, "A deep learning based food recognition system for lifelog images," *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*, 2018.