Deep Learning Based on Food Image Recognition And Portion Size Determination

Shreyaa Sridhar
School of Computing and Engineering
University of Missouri-Kansas City
ssn2m@mail.umkc.edu

Geovanni Nicque West
School of Computing and Engineering
University of Missouri-Kansas City
gwrn2@mail.umkc.edu

Khushbu Kolhe
School of Computing and Engineering
University of Missouri-Kansas City
kkvcm@mail.umkc.edu

Naga Venkata Satya Pranoop Mutha
School of Computing and Engineering
University of Missouri-Kansas City
nmgfg@mail.umkc.edu

ABSTRACT

Good nutrition is an important part of leading a healthy lifestyle. Food choices that we make each day affect our health - how we feel today, tomorrow and in the future. In addition to that, food portion also plays an important role because it allows for you to have a tight handle on how many calories you are presumably taking in. We propose a new deep learning based model which recognizes food, predict calories and determines the portion size of the food image recognized.

Keywords: Deep learning, food image recognition, portion size determination

1. INTRODUCTION

Accurately recognizing food images is a particularly challenging task based on the nature of images, which is why approaches like Spark API models in the field have achieved a low classification accuracy. Deep neural networks have outperformed such solutions, and we present a novel approach to the problem of food image detection and recognition and also portion size determination that uses a Convolutional Neural Network based system. We are comparing the accuracy of our model with that of SoftMax Regression Model.

2. RELATED WORK

NutriNet model [2] was built to recognize food and drinks. NutriNet+ was built on 6 convolutional layers when compared to AlexNet it has one extra layer. This convolutional layer was added to gain additional knowledge about the features in the higher resolution images. NutriNet model uses UNIMIB 2016 Food image dataset. NutriNet with the NAG solver was the best-performing model with a classification accuracy of 94.47%. This model majority of images are correctly classified in the recognition dataset, but not necessarily all of them. As a consequence, this could lower the classification accuracy for real-world images. Finally, since image segmentation was not performed, irrelevant items present in the images made the recognition difficult. When comparing NutriNet to NutriNet+, we can see that the extra convolutional layer did not yield any performance increase, as NutriNet+ models achieved results that are almost identical to the results by NutriNet models.

A food image recognition system that uses the multiple kernel learning method was introduced, which tested different feature extractors, and their combination, on a self-acquired dataset [5]. This proved to be a step in the right direction, as the authors achieved an accuracy of 26% to 38% for the individual features they used and an accuracy of 61.34% when these features were combined; the features include color, texture and SIFT information. Upon conducting a real-world test on 166 food images taken with mobile phones, the authors reported a lower classification accuracy of 37.35%, which was due to factors like occlusion, noise and additional items being present in the real-world images. The fact that the combination of features performed better than the individual features further hinted at the need for a more in-depth representation of the food images.

Pairwise local features method, which applies the specifics of food images to their recognition [6], analyzes the ingredient relations in the food image, such as the relations between bread and meat in a sandwich, by computing pairwise statistics between the local features. The authors performed an evaluation of their algorithm on the PFID dataset and achieved an accuracy of 19% to 28%, depending on which measure they employed in the pairwise local features method. However, they also noted that the dataset had narrowly-defined food classes, and after joining them into 7 classes, they reported an accuracy of 69% to 78%. This further confirmed the limitations of food image recognition approaches of that time: if a food image recognition algorithm achieved a high classification accuracy, it was only because the food classes were very general (e.g., "chicken").

4. PROPOSED WORK

In this project, we have decided to use deep learning technologies to design a model which can predict food and determining portion size.

3.1. Model Architecture

The architecture diagram in figure 1 explains how our model works. Images from Food-101 dataset is provided as input to our model. Here we have compared how our dataset works with different models such as Clarifai API model, Spark API model and Deep Learning models(SoftMax Regression Model and Convolutional Neural Network Model) . The next step is running our model to obtain annotations for each image, check accuracy and obtain confusion matrix.

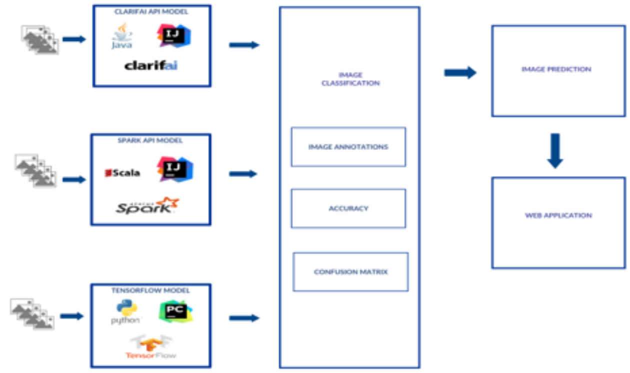

Figure 1

3.2 Features, workflow, technologies

3.2.1 Activity Diagram

The activity diagram figure 2 shows the workflow of our model.

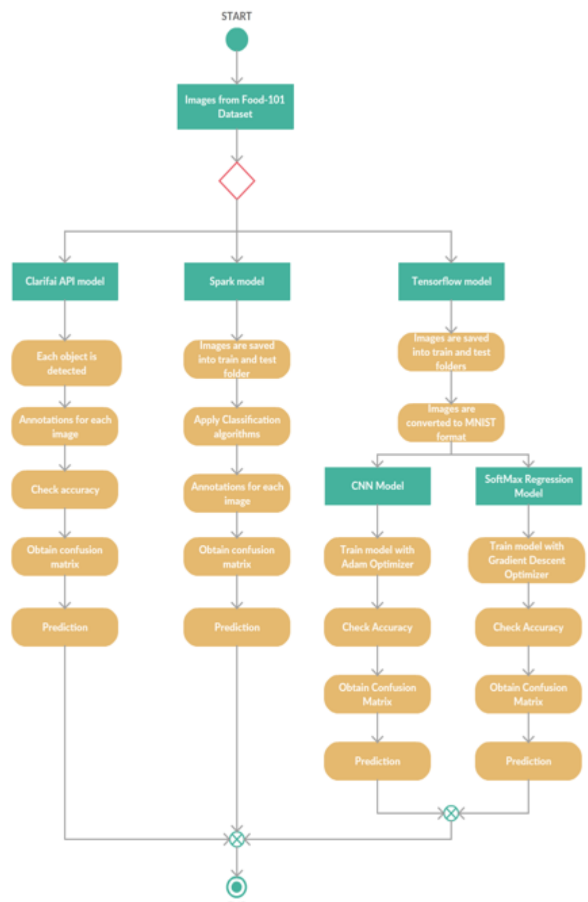

Figure 2

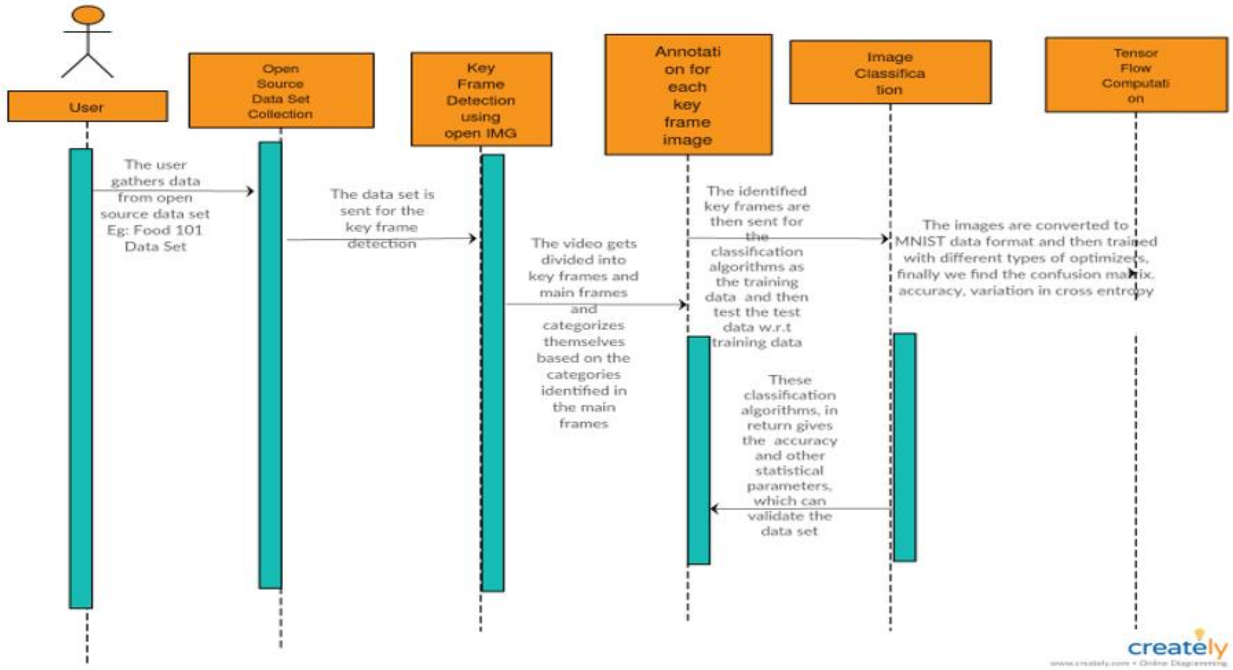

Figure 3

3.2.2 Sequence Diagram

The sequence diagram figure 3 shows the interaction or collaboration within our model.

3.2.3 Feature Specification

- Food Image Recognition
- Portion Size Determination

3.2.4 Operation Specification (Input/output, exceptions)

Input - Images from Food-101 dataset

Output - Image Categorization, Image Prediction, Statistical parameters like accuracy, precision, confusion matrix etc.

3.3 Existing Applications/Services Used

- We have used **Clarifai API service** which automatically tags the images or video file in such a way ensuring a quick organization, management and searching throughout the content.
  URL: https://www.clarifai.com/
- **Spark API** offers easy-to-use APIs, for operating on large datasets, across languages: Scala, Java, Python, and R.
- We have used the **Tensorflow** for Deep learning models such as Convolutional Neural Network Model and SoftMax Regression Model.

4. IMPLEMENTATION & RESULTS

4.1 Experimentation results of Clarifai API, Spark API and Tensorflow Model

4.1.1 Clarifai API Model

In Clarifai API, the type of prediction is based on what model you run the input through. For example, if you run your input through the 'food' model, the predictions it returns will contain concepts that the 'food' model knows about. Clarifai API has predicted an accuracy of 96.4%. Dumplings predicted in figure 4.

Figure 4

4.1.2 Spark API Prediction

Spark API works well on large dataset and by using the dataset on different models like Random Forest Model, Decision Tree Model and Naive Bayes Model, we found that Random Forest Model works well for our model. The accuracy of our model using Spark API is 44.04%. Image predicted as French fries in figure 5.

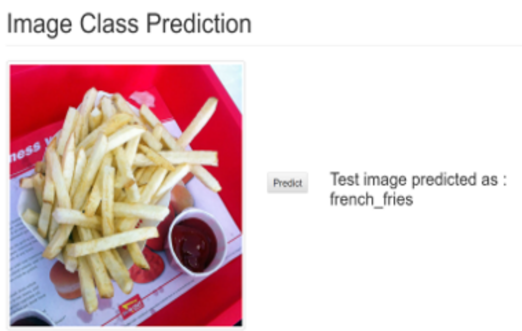

Figure 5

4.1.3 Tensorflow Model

In our project, we are using SoftMax Regression Model and Convolutional Neural Network Model for food image recognition. We have trained our model with images in the ratio 70% for train and 30% for testing. Our model has predicted an accuracy of 41.85% for SoftMax Regression Model and 51.3% for CNN Model.

4.1.4 Compare Accuracy of Clarifai API, Spark API and Tensorflow Model

4.1.4 Compare Accuracy of Clarifai API, Spark API and Tensorflow Model

Model	Accuracy
Clarifai API Model	96.4%
Spark API Model(Random Forest Model)	44.04%
SoftMax Regression Model	41.85%
CNN Model	51.3%

Figure 6

Based on the dataset considered, the Clarifai API model was the best model based on the accuracy. Whereas, the CNN model was found to be the best Deep Learning Model with an accuracy of 51.3%.

5. DISCUSSION & LIMITATIONS

This data set contains of 101 food types,each food type consisting of 1000 images. In comparison among different models used for our dataset, we have got the best accuracy for Clarifai API Model and least for SoftMax. Though, the accuracy looks very low except on Clarifai, it is mainly because we restricted input to only five food types, which is minute compared to the data set we have taken for reference. We tested with multiple food types, but taking five or six food types at a time. We observed that the accuracy varies for the models (i.e not one model stays low every time, they keep altering). The future work will include more food types and hope for a better accuracy. We also would include food types which look alike (like donuts and bagels), for a better efficient application, detecting exact food type. In addition, we also try to display the calories present in each food type, and the accuracy comparison for each of the models.

The solution to lost data is still under study. The provisions for editing or making changes in the app are not yet to be figured out. The model cannot interpret between different food items which look almost similar and will be provided in the next increment. The accuracy with the SoftMax and CNN in this case is low and this can be resolved by increasing the categories.

6. REFERENCES

(1) Yuzhen Lu, “Food Image Recognition by Using Convolutional Neural Networks (CNNs)”, Cornell University Library, Computer Vision and Pattern Recognition, Dec 2016, arxiv:1612.00983

(2) Simon Mezgec and Barbara Korousic Seljak, “NutriNet: A Deep Learning Food and Drink Image Recognition System for Dietary Assessment” NCBI, Nutrients, v.9(7), 2017 July, ,PMC5537777

(3) Chen M., Dhingra K., Wu W., Yang L., Sukthankar R., Yang J. PFID: Pittsburgh Fast-Food Image Dataset; Proceedings of the ICIP 2009; Cairo, Egypt. 7–10 November 2009; pp. 289–292.

(4) Lowe D.G. Object Recognition from Local Scale-Invariant Features; Proceedings of the ICCV'99; Corfu, Greece. 20–21 September 1999; pp. 1150–1157.

(5) Joutou T., Yanai K. A Food Image Recognition System with Multiple Kernel Learning; Proceedings of the ICIP 2009; Cairo, Egypt. 7–10 November 2009; pp. 285–288.

(6) Yang S., Chen M., Pomerlau D., Sukthankar R. Food Recognition using Statistics of Pairwise Local Features; Proceedings of the CVPR 2010; San Francisco, CA, USA. 13–18 June 2010; pp. 2249–2256.

(7) Artifact from the Future: GOTTA EAT 'EM ALL http://www.iftf.org/future-now/article-detail/artifact-from-the-future-gotta-eat-em-all/