Computer Engineering Department

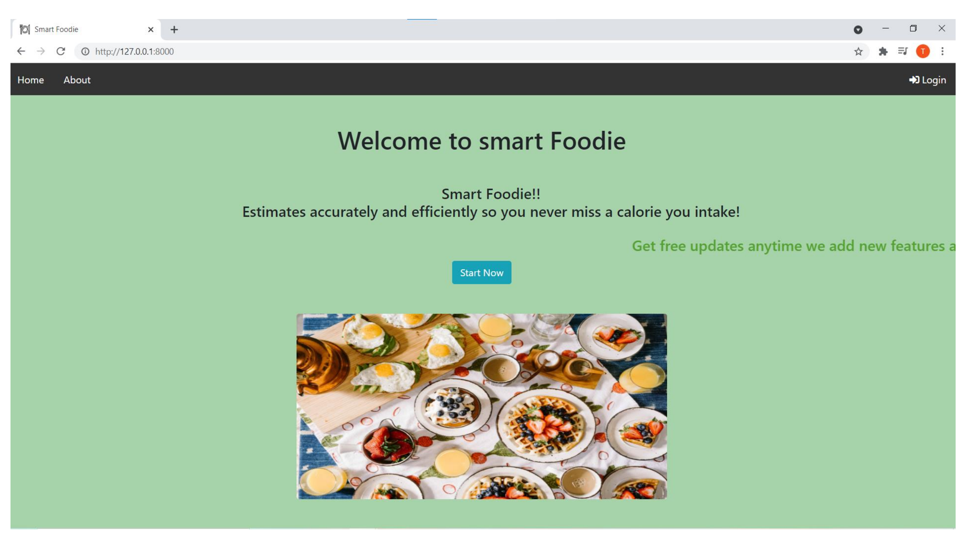
Smart Foodie

Project Advisor: Vijay Eranti

Baslan, Samer (MS Software Engineering)
Liang, Jimmy (MS Software Engineering)
Kumar, Rohan (MS Software Engineering)
Gorla, Tripura Chandana Gayatri
(MS Computer Engineering)

Introduction

With food being an essential part of human life and a large contributor to the health of a population, it is important to make efforts that will improve our understanding of what we are eating and how it fits into our daily dietary needs. But, due to our hectic lifestyles regular applications that require manual entry of nutritional values and give poor personalized recommendations and alternative options are not preferred by users. As a result, the average user retention of current dietary applications is less than three months. Our dietary application called "Smart Foodie" is fast and automated that fits the needs of today's gen.

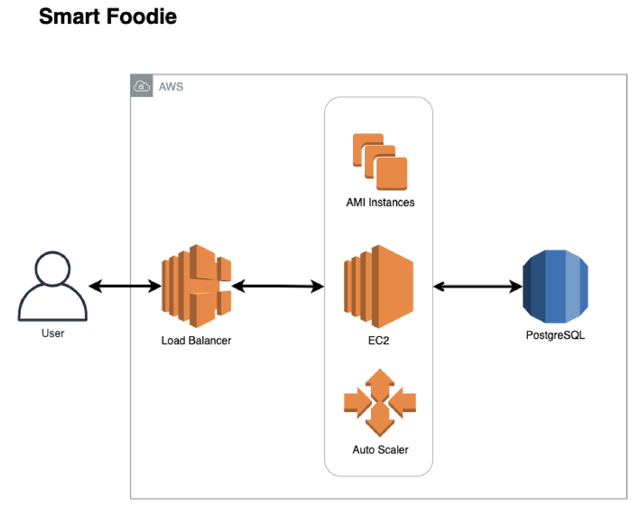


Using machine learning, "Smart Foodie" can identify the nutrition values from commonly prepared dishes, individual food items, and Nutritional Facts labels. This is done simply by uploading a photo. Smart Foodie will provide users with analytical dashboards to track nutritional intake and actively encourage users to maintain their dietary goals.

Methodology

Frontend Web Implementation

The front end web interface portal that users will be interacting with. A user will be able to arrive at the application and create an account, using only their name and email address. This will allow the user to maintain a log of their meals throughout time, and to further aid in the calculation of the accurate dietary suggestions that would fit the user's unique needs.



For the frontend, along with basic HTML and CSS, we used Django, along with Bootstrap and Javascript. Django is a high-level Python web framework that enables rapid development and easy to read, pragmatic design.

Methodology

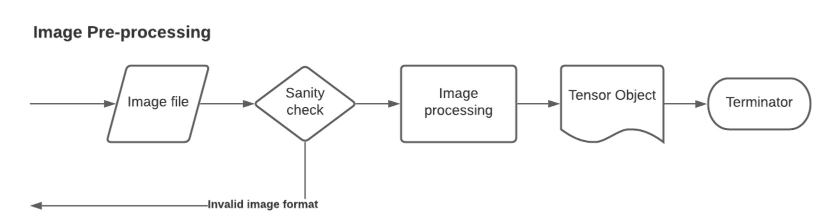
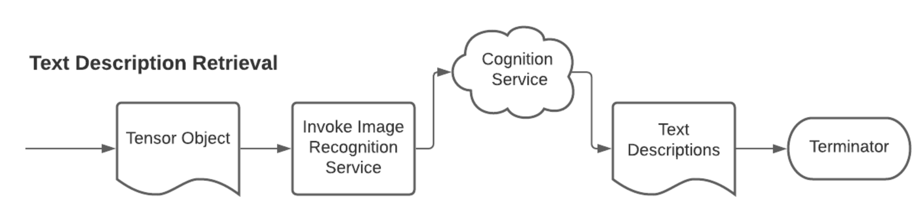


Image Recognition Technologies

The Image Recognition component is the most important component of the app and will be a huge factor that determines its success. We hope to obtain a high classification accuracy such that the app will recognize the users' food often, but that alone won't guarantee the degree of accuracy required for the application to be successful. After every classification attempt by the Image Recognition module, the user will be prompted to ensure that the information output by the model is correct (Is this a picture of X? Does this look like Y grams? Is there anything in the food item that isn't shown on the photo?). The model used for inference in this project is an InceptionV3 convolutional neural network architecture, which is perfect for image analysis and object detection.

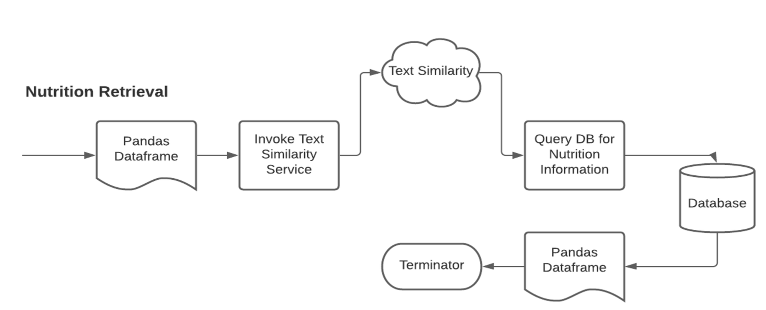


Nutrition Retrieval Technologies

This component is the nutrition retrieval layer of the Smart Foodie application. The purpose of this layer is to return back the nutrition values for a given food item and serving size. Full text search and cardinality matching will be used to return the best possible results. If a user uploads a Nutrition Facts label, the values parsed by the image recognition layer will be upserted into the database. This will ensure that the database is up to date, thus eliminating data maintenance. The database must be comprehensive and fast to meet service level agreements (SLA).

The nutrition retrieval component takes as input a Pandas dataframe to invoke the text similarity service to retrieve the most likely food item prediction. It then queries the food nutrition database for the nutrition information. The output produced by this component is also a Pandas dataframe that contains the nutrition information.

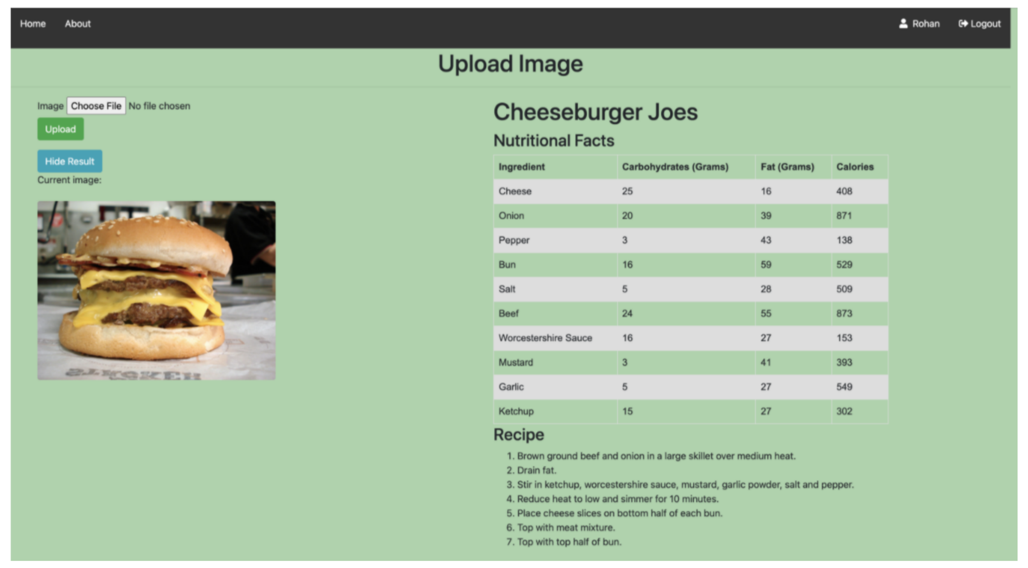
For data persistence, including nutrition retrieval, we used SQLite database. SQLite is a C-language library that implements a small, fast, self-contained, high-reliability, full-featured, SQL database.



Analysis and Results

In the 1st part of the implementation, MobileNet V2 and Inception V3 are the two pretrained models that we have implemented to compare the performances. The model was trained using 18,750 food images, allowing it to identify various different types of the most popular foods accurately which yielded a classification accuracy of approximately 90% using Inception V3 and around 68% for MobileNet V2.

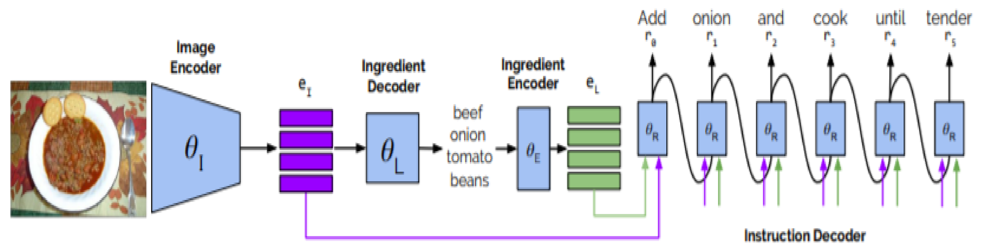
Generating a recipe (title, ingredients and instructions) from an image is a challenging task, which requires a simultaneous understanding of the ingredients composing the dish as well as the transformations they went through, e.g. slicing, blending or mixing with other ingredients. Instead of obtaining the recipe from an image directly, we argue that a recipe generation pipeline would benefit from an intermediate step predicting the ingredients list.



```
RECIPE 1

Title: Pizza margherita

Ingredients:
tomato, cheese, basil, oil, pizza_sauce, salt, yeast, pepper
```

```
RECIPE 1

Title: Cheeseburger

Ingredients:
pepper, onion, bun, cheese, beef, salt, tomato, garlic, lettuce
```

Recipe generation model. We extract image features eI with the image encoder, parametrized by θI . Ingredients are predicted by θL, and encoded into ingredient embeddings eL with θe.



The cooking instruction decoder, parametrized by θR generates a recipe title and a sequence of cooking steps by attending to image embeddings eI , ingredient embeddings eL, and previously predicted words (r0, ..., rt−1).

```
Instructions:
-In a large skillet, cook beef, onion, and garlic over medium heat until meat is no longer pink; drain.
-Stir in salt and pepper.
-Reduce heat; simmer, uncovered, for 10 minutes.
-Place cheese on bottom half of each bun; top with meat mixture.
-Cover with top halves of buns.
=====================
RECIPE 2

Title: Cheeseburger
```

Summary/Conclusions

This project acts as a proof of concept for our vision of the intersection between health eating and deep learning and demonstrated a flexible framework to bring deep learning into the health and nutrition field. By combining deep learning vision models with a distributed computing environment, we integrated the latest trend in software engineering and artificial intelligence that were integral in this masters program. We improved usability and ease of adoption to one of the most time consuming and manual tasks of nutrition tracking.

Key References

[1] Y. He, C. Xu, N. Khanna, C. J. Boushey and E. J. Delp, "Food image analysis: Segmentation, identification and weight estimation," 2013 IEEE International Conference on Multimedia and Expo (ICME), 2013, pp. 1-6, doi: 10.1109/ICME.2013.6607548.

[2] Wu, Xiongwei, Xin Fu, Ying Liu, Ee-Peng Lim, Steven CH Hoi, and Qianru Sun. "A Large-Scale Benchmark for Food Image Segmentation." arXiv preprint arXiv:2105.05409 (2021).

[3] Liu, Chang, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma. "Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment." In International Conference on Smart Homes and Health Telematics, pp. 37-48. Springer, Cham, 2016.

[4] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101–mining discriminative components with random forests. In ECCV, 2014.

Acknowledgements

The authors are deeply indebted to San Jose State University Professors.