Team Name: Convolutionalists

Project Title: Food for Deep Thought

Project Summary:
Project summary (4-5+ sentences). Fill in your problem and background/motivation (why do you want to solve it? Why is it interesting?). This should provide some detail (don't just say "I'll be working on object detection")

The goal of this deep learning group project is to develop a comprehensive food recognition and information systems. The model will perform object detection and classification of fruits, vegetables, and other dietary components. Upon learning this information, the model will then perform information retrieval to provide various types of information such as nutrition facts, allergies, health benefits, and recipes associated with these foods. The primary motivation is to develop an application that allows individuals to make healthier, safer dietary choices by identifying nutritious foods and their ingredients, increases individuals' general understanding on the types of foods and drinks consumed, and reduces waste by creating suggestions for recipes for those with leftover ingredients in the kitchen. This serves as an interesting problem because it presents a practical real-world application that addresses a major area for improvement in people's lifestyles: making healthier food choices. From a deep learning perspective, this project poses an interesting challenge as it requires integrating object detection, multi-category classification, and information retrieval into a single pipeline, presenting a complex technical task. Additionally, the model must be robust enough to accurately identify a wide variety of foods under diverse real-world conditions, such as varying lighting, angles, and presentation styles.

Approach:
What you will do (Approach, 4-5+ sentences) - Be specific about what you will implement and what existing code you will use. Describe what you actually plan to implement or the experiments you might try, etc. Again, provide sufficient information describing exactly what you'll do. One of the key things to note is that just downloading code and running it on a dataset is not sufficient for a description or a project! Some thorough implementation, analysis, theory, etc. have to be done for the project.

For real-time object detection and object classification, we will implement a model that follows state-of-the-art techniques such as YOLO or Faster R-CNN, which would utilize existing code adapted for these models but applied to our selected datasets. Specifically, we will implement and tune different types of state-of-the-art models (YOLO, Faster R-CNN, EfficientDet, etc.) and compare their performances against one another, analyzing the strengths and weaknesses of each model for this specific use-case. The models are intended to be trained on a combination of multiple existing food datasets to enable them to classify a wider variety of different food types. For the retrieval of nutrition facts, health benefits, recipes, and more, pre-trained large language models (LLMs), such as BERT or GPT, will be used. Additionally, we will analyze our model using visualization techniques such as GradCAM and Saliency Maps to detect any biases within the model or identify any additional methods of improvement to model performance.

Furthermore, to improve the robustness of our model, we will also explore the use of data augmentation techniques, such as flipping, rotation, scaling, cropping, and color distortion, and fooling images and style transfer techniques, which will help increase the size, diversity, and reliability of our training data.

Resources:

Resources / Related Work & Papers (4-5+ sentences). What is the state of the art for this problem? Note that it is perfectly fine for this project to implement approaches that already exist. This part should show you've done some research about what approaches exist.

There are various state of the art models that we will leverage for the object recognition and object classification portions of this problem: YOLO, Faster R-CNN, and EfficientDet,. Each of these three models work well against food detection and classification. Firstly, YOLO (You Only Look Once) is a single-stage object detection model that has been shown to achieve high accuracy in real-time speed, which may be an ideal choice. It has shown exceptional real-time performance and accuracy in food detection tasks. Faster R-CNN is a two-stage object detection model that has been shown to achieve higher precision and accuracy than the other two models (which would be particularly useful in detecting multiple types of foods in a single image or overlapping foods) but is the slowest among the three models. EfficientDet uses EfficientNet as its backbone with a Bidirectional Feature Pyramid Network (BiFPN) and achieves a better balance between speed and accuracy; specifically, it achieves more accurate results than YOLO, all while constituting a model size much smaller than Faster R-CNN. As each model possesses its own strengths and weaknesses, it will be interesting to compare the performance and accuracy results among these three types of models. Lastly, the state of the art for the information retrieval portion of the project includes the use of transformer-based models such as BERT and GPT, one of which will be used as the information retrieval model for this use-case. Recent studies (Stojanov R et al., 2021) have shown the effectiveness of fine-tuning these models for food-specific tasks, such as ingredient identification and nutritional analysis. The same study proposed integrated approaches combining visual and textual information for comprehensive food analysis. The FoodNER model uses a fine-tuned BERT for food information extraction, representing more accurate and context-aware food recognition systems.

Relevant Papers:

- YOLO - https://ar5iv.labs.arxiv.org/html/1909.05994
- Food detection with YOLO - https://www.irjmets.com/uploadedfiles/paper//issue_10_october_2023/45302/final/fin_irjmets1697522755.pdf
- Fine-tuned BERT - https://www.jmir.org/2021/8/e28229/

Datasets:

Datasets (Provide a link to the dataset). This is crucial! Deep learning is data-driven, so what datasets you use is crucial. One of the key things is to make sure you don't try to create and especially annotate your own data! Otherwise, the project will be taken over by this.

Primary Datasets:

- Food-101: https://paperswithcode.com/dataset/food-101
- Food 2k: https://paperswithcode.com/dataset/food2k

These datasets will serve as our primary datasets for this project. Food-101 consists of 101 food categories, with 750 training and 250 test images per category, totaling 101,000 images. The Food-2k dataset is a large-scale food image dataset with 2,000 categories, expanding the range of food types that our model can recognize. This combined dataset will offer a wide variety of food types and real-world images. Each dataset has different image qualities and styles, which will help our model perform well in various real-world scenarios. We plan to use these datasets for training, validation, and testing our models. The diversity in these datasets will help ensure that our final model is robust and generalizable across different food types and image conditions. Additionally, we may explore other secondary datasets for further validation and testing such as the following below.

Secondary Datasets:

- Fruits-360 dataset: https://www.kaggle.com/datasets/moltean/fruits
- Food 11: https://www.kaggle.com/datasets/trolukovich/food11-image-dataset
- Open Food Facts Images: https://blog.openfoodfacts.org/en/news/open-food-facts-images-on-aws-open-dataset-the-ultimate-food-image-database
- Food Seg 103: https://paperswithcode.com/dataset/foodseg103
- Deepfruits: https://www.kaggle.com/datasets/mexwell/deepfruits-dataset-of-fruits-images

Group Members:

- Joe Granados <joe.granados@gatech.edu>
- Peter Park <ppark46@gatech.edu>
- Praveen Ravishankar <pravishankar6@gatech.edu>
- Scott Schmidl <sschmidl3@gatech.edu>