Recipe Recognition With Deep Learning & IBM Watson

RAMINEDI SANTHOSH 20BCE1477

raminedi.santhosh2020@vitstudent.ac.in

KRISHNARJUN L 19BLC1134
krishnarjun.l2019@vitstudent.ac.in

Kanuru Uday 19BCE10285
kanuru.uday2019@vitbhopal.ac.in

DHANUSHVARMA R 18BCE10093
dhanushvarma.r2018@vitbhopal.ac.in

Index

1. Introduction

1.1 Overview

Food is an essential component of our individual and social life. Eating habits have a direct impact on our health and wellbeing, while Recipes, flavor and cooking recipes shape specific cuisines that are part of our personal and collective cultural identities. But there are also interesting applications of automatic food recognition to self-service restaurants and dining halls. For instance, accurate detection and segmentation of the different food items in a food tray can be used for monitoring food intake and nutritional information, and automatic billing to avoid the cashier bottleneck in self-service restaurants. This work deals with the problem of automated recognition of a photographed cooking dish and the subsequent output of the appropriate recipe.

1.2 Purpose

Food recognition can be seen as a particular case of fine grained visual recognition, where photos within the same category may have significant variability while are often visually similar to photos from other categories. Effective classification requires identifying subtle details and fine-grained analysis. In the case of restaurants or recipes, the number of categories can explode since the name of dishes in a menu or in a recipe database can be very large. This increases the variability significantly, since the same dish can have very different appearances (due to the particular cooking style, presentation, restaurant, etc). The names also become more elaborated and specific and many of them are signature dishes that can be found only in one place for which only one or even no images are available. In general, this constitutes a long tail of rare dishes with very limited training data, which makes the recognition problem from purely visual appearance very difficult. Retrieving recipes corresponding to given dish pictures facilitates the estimation of nutrition facts, which is crucial to various health relevant applications. The current approaches mostly focus on recognition of food categories based on global dish appearance with explicit analysis of ingredient composition. Such approaches are incapable for retrieval of recipes with unknown food categories, a problem referred to as zero-shot retrieval. On the other hand, content-based retrieval without knowledge of food categories is also difficult to attain satisfactory performance due to large visual variations in food appearance and ingredient composition.

This paper proposes deep learning architectures for simultaneous learning of Recipe recognition and food categorization, by exploiting the mutual but also fuzzy relationship between them. The learnt deep features and semantic labels of Recipes are then innovatively applied for zero-shot retrieval of recipes. By experimenting on a large Chinese food dataset with images of highly complex dish appearance, this paper demonstrates the feasibility of Recipe recognition and sheds light on this zero-shot problem peculiar to cooking recipe retrieval.

2. Literature Survey

2.1 Existing Problem

- Sometimes users cannot identify the food item due to different styles of cooking; this is a major concern for the users.
- When users go to restaurants they like food, when they want to know a food recipe it is difficult for them to meet the chef and know the recipe.
- Lack of availability of resources to know the particular style of cuisine.

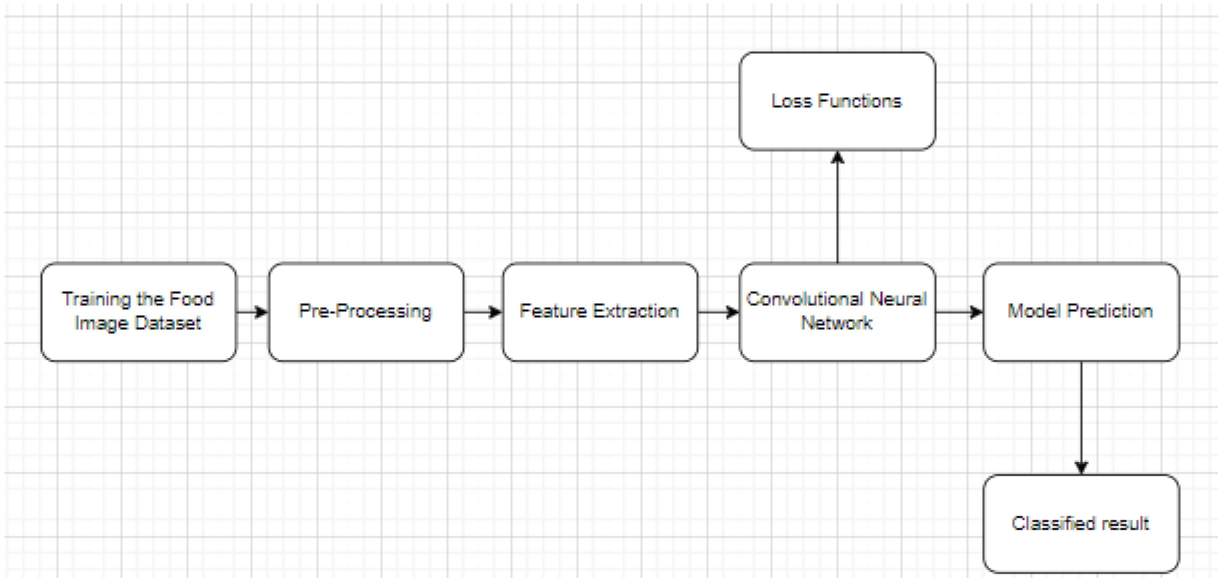
2.2 Proposed Solution

The objective of this work is to propose a Recipe Recognition system which assists a user to decide what is the recipe and how to cook using generic object recognition technology. We assume that the proposed system works on a smartphone which has built-in cameras and Internet connection such as Android smartphones and iPhones. We intend a user to use our system easily and intuitively during shopping at grocery stores or supermarkets as well as before cooking at home. By pointing food ingredients with a mobile phone built-in camera, a user can receive a recipe list which the system obtained from online cooking recipe databases instantly. With our system, a user can get to know the cooking recipes related to various kinds of food ingredients unexpectedly found in a grocery store including unfamiliar ones and bargain ones on the spot. To do that, the system recognizes food ingredients in the photos taken by built-in cameras, and searches online cooking recipe databases for the recipes which need the recognized food ingredients.
As an object recognition method, we adopt bag-of-features with SURF and color histogram extracted from not single but multiple images as image features and linear kernel SVM with the one-vs-rest strategy as a classifier.

To give good experience we are coming with a user-friendly interface with accurate results, with more datasets to recognize the image with more accuracy, and we are using good analytics tools to predict recipes using images, with more data it has the ability to deliver high-quality results.
First it will recognise the food image we are given then it will recognise the food and start predictions, first it will predict the ingredients used in food, then it will predict the method of making the food using our datasets.

3. Theoretical Analysis

3.1 Block Diagram



Recipe Recognition block diagram

3.2 Software designing

In order to develop this project we need to install the following software/packages:

3.2.1 Anaconda Navigator

Anaconda Navigator is a free and open-source distribution of the Python and R programming languages for data science and machine learning related applications. It can be installed on Windows, Linux, and macOS.Conda is an open-source, cross-platform, package management system.

3.2.2 Tensor flow

TensorFlow is an end-to-end open-source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries, and community resources that lets researchers push the state-of-the-art in ML and developers can easily build and deploy ML-powered applications.

3.2.3 Keras

Keras leverages various optimization techniques to make high-level neural network API easier and more performant.

3.2.4 Flask

Web framework used for building Web applications

3.2.5 Installing Necessary libraries

While using anaconda navigator, below steps are to be followed to download the required packages

Open the anaconda prompt.

Type “pip install numpy” and click enter.
Type “pip install pandas” and click enter.
Type “pip install matplotlib” and click enter.
Type “pip install scikit-learn” and click enter.
Type "pip install tensorflow” and click enter.
Type "pip install keras and click enter.
Type "pip install opencv-python==4.4.0” and click enter.
Type “pip install Flask” and click enter.

3.2.6 Library methods used in the project

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Convolution2D, MaxPooling2D, Flatten, Dense
import numpy as np
from tensorflow.keras.models import load_model
from tensorflow.keras.preprocessing import image
```

4. Experimental Investigation

Advances in the classification of individual cooking ingredients are sparse. The problem is that there are almost no public edited records available. This work deals with the problem of automated recognition of a photographed cooking dish and the subsequent output of the appropriate recipe. The distinction between the difficulty of the chosen problem and supervised classification problems is that there are large overlaps in food dishes (aka high intra-class similarity), as dishes of different categories may look very similar only in terms of image information.

The combination of object recognition or cooking court recognition using Convolutional Neural Networks (short CNN) and the search for the nearest neighbors (Next-Neighbor Classification) in a record of over 2500 images. This combination helps to find the correct recipe more likely, as the 3 categories of the CNN are compared to the next-neighbor category with ranked correlation. Rank correlation based approaches such as Kendall Tau essentially measure the probability of two items being in the same order in the two ranked lists.

The exact pipeline looks like the following:

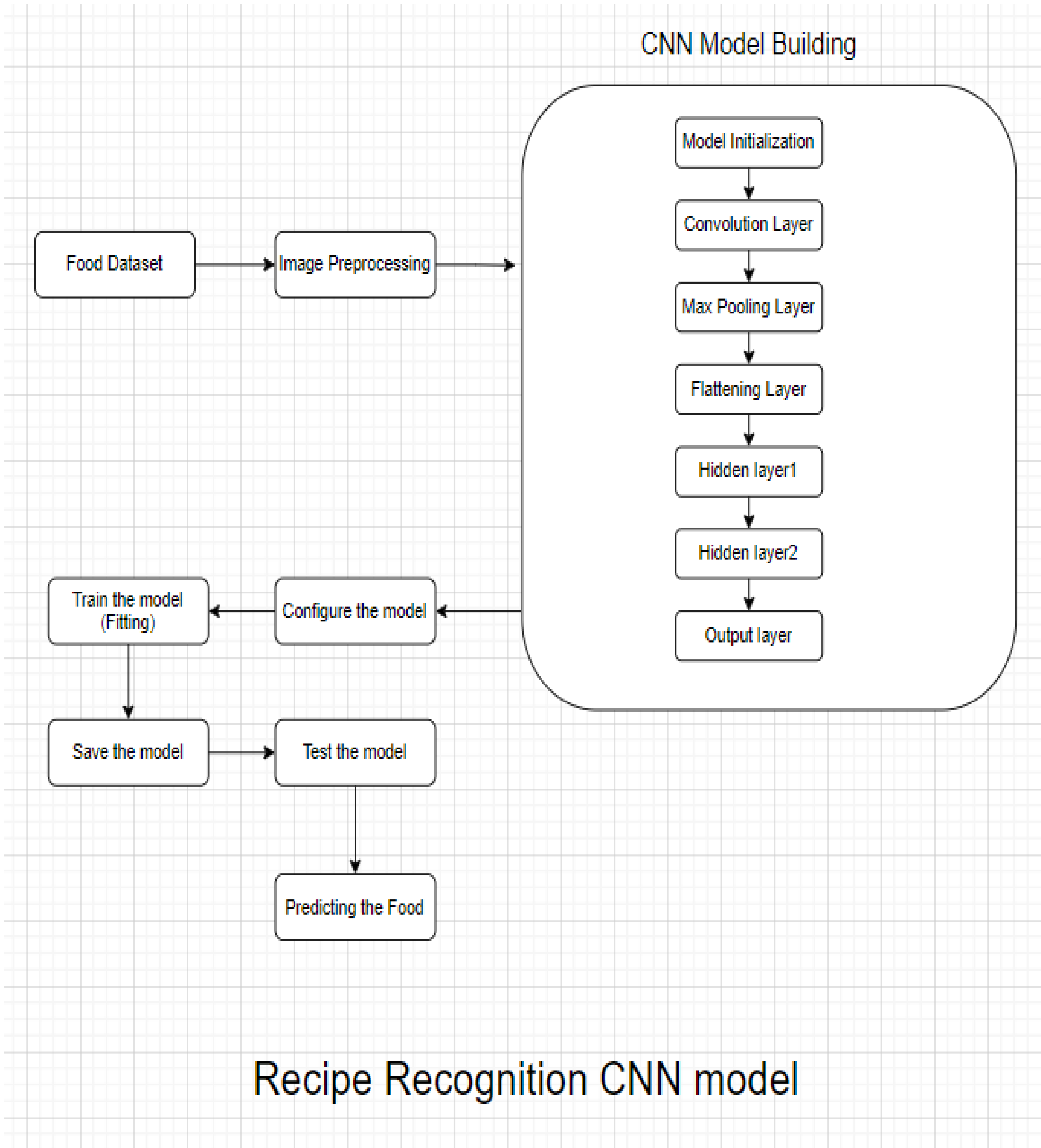
For every recipe W there are K numbers of pictures. For each of these images feature vectors are extracted from a pre-trained Convolutional Neural Network trained on 1000 categories in the ILSVRC 2014 image recognition competition with millions of images. The feature vectors form an internal representation of the image in the last fully connected layer before the 3-category Softmax Layer, which was removed beforehand. These feature vectors are then dimensionally reduced by PCA (Principal Component Analysis) from an N x 4096 matrix to an N x 512 matrix. As a result, one chooses the top 3 images with the smallest Euclidean distance to the input image (Approximate nearest neighbor), i.e. the top 3 optical, just from the picture information, similar pictures to the Input image.

Furthermore, a CNN is trained with C number of categories with pictures of W recipes. C has been determined dynamically using topic modeling and semantic analysis of recipe names. As a result we obtain for each category a probability to which the input image could belong.

The top-k categories from CNN (2.) are compared with the categories from the top-k optically similar images (1.) with Kendall Tau correlation.

The goal of this procedure is to divide all recipe names into n-categories. For a supervised classification problem, we have to provide the neural network with labeled images. It is only with these labels that learning becomes possible.

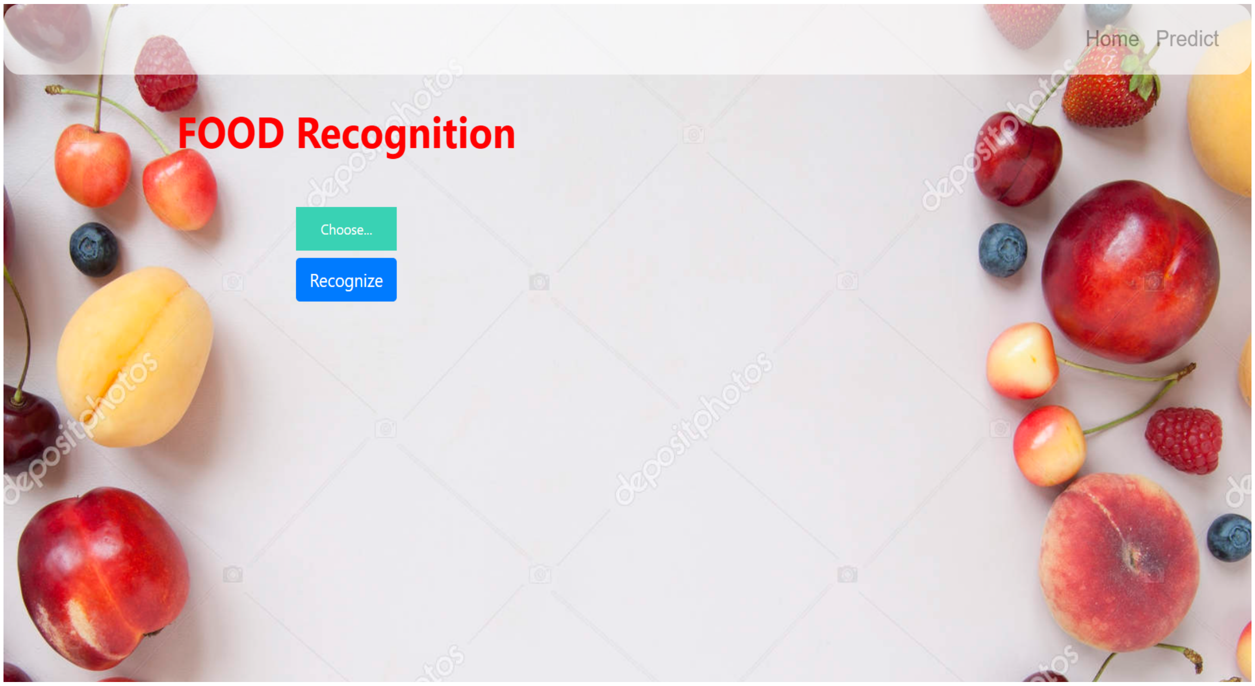
5. Flow chart


CNN Model Building
Food Dataset
Image Preprocessing
Model Initialization
Convolution Layer
Max Pooling Layer
Flattening Layer
Hidden layer1
Hidden layer2
Output layer
Configure the model
Train the model (Fitting)
Save the model
Test the model
Predicting the Food
Recipe Recognition CNN model

6. Result

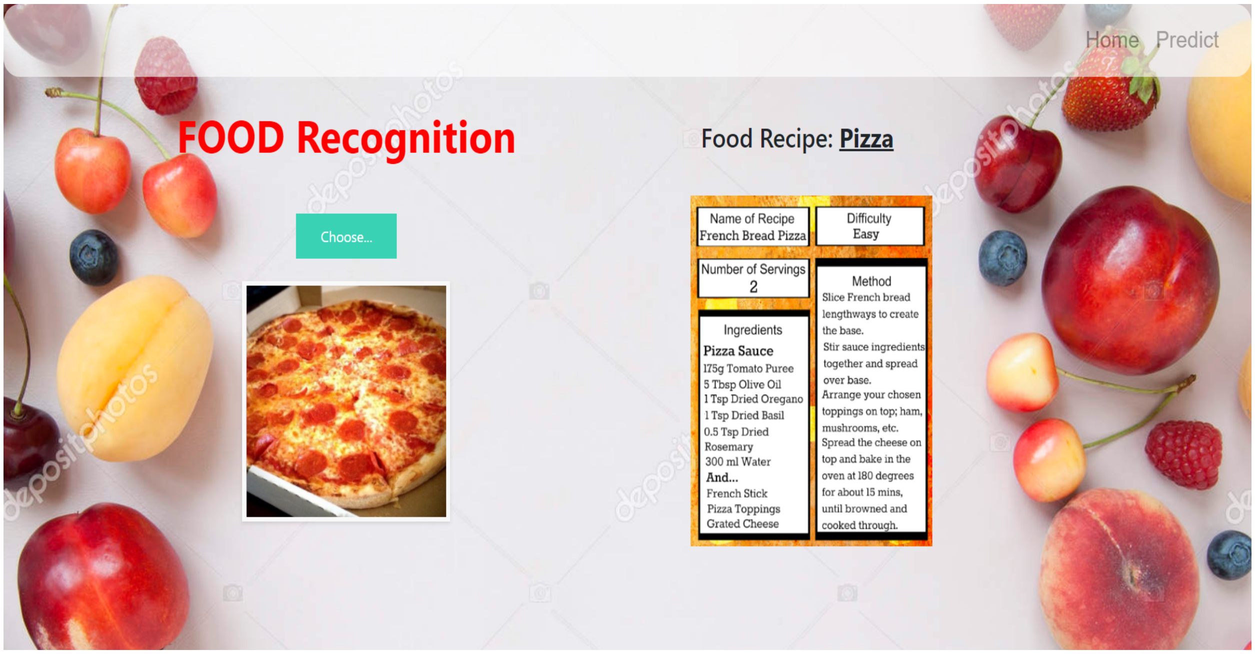
The application runs on a local host which is developed using flask integration .

When the “predict” button is clicked it will redirect to the prediction page.
When the image is uploaded, it predicts whether the uploaded image is French fries or pizza or samosa. If the image predicts value as 1, then it is displayed as “Pizza”. If the image predicts value as 0, then it is displayed as “French Fries”. If the image predicts value as 2, then it is displayed as “Samosa”.

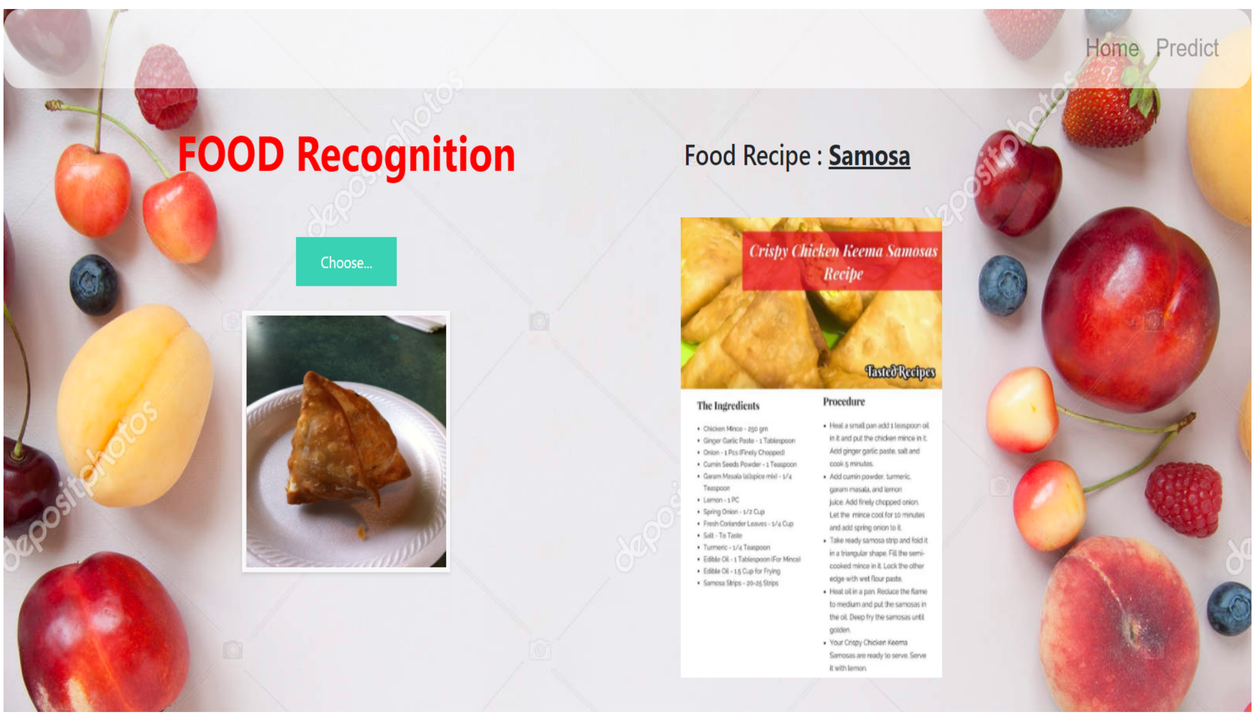


6.1 Prediction 1

6.2 Prediction 2



6.3 Prediction 3



7. Advantages and Disadvantages

Advantages:
- Recipe Recognition helps the users to categorize the food items.
- It will help the user to recognize the food item if it is in different styles
- Key ingredients help to know the nutritional value of the food.
- Procedure helps the users to make food on their own.
- Nutrition can be analyzed with the help of Recipe Recognition.

Disadvantages:
- It requires a very large amount of data in order to perform better than other techniques.
- It is extremely expensive to train due to complex data models. Moreover deep learning requires expensive GPUs and hundreds of machines. This increases cost to the users.
- There is no standard theory to guide you in selecting the right deep learning tools as it requires knowledge of topology, training method and other parameters. As a result it is difficult to be adopted by less skilled people.
- It is not easy to comprehend output based on mere learning and requires classifiers to do so. Convolutional neural network based algorithms perform such tasks.

8. Application

1. Users who go to restaurants can use our application to know the recipe of the food they ordered.
2. Our deep learning Recipe Recognition model can help the users to cook the food on their own.
3. Our Application acts as not only limited to a Recipe Recognition but also it can be extended to the nutrition analysis so that users can know the nutrition value of the particular cuisine.

9.Conclusions

This project revealed that artificial intelligence modules have the ability to be applied in the various business processes in order to improve the quality of provided services. The final decision about integrating modern technologies is possible after comparing the results of the impact of the human recognition time of a customized recipe on the entire process of production of products and service quality with artificial intelligence. We also show that unsupervised tree topologies can improve food cross-modal retrieval abilities by adding structural information to cooking instruction representations. We used quantitative and qualitative analyses for the food retrieval task, and the results show that the model with tree structure representations outperformed the baseline model by a large margin. Our proposed algorithms are based on Convolutional Neural Network (CNN). Our experimental results on two challenging data sets using our proposed approach exceed the results from all existing approaches. In the future, we plan to improve performance of the algorithms and integrate our system into a real word mobile devices and cloud computing-based system to enhance the accuracy of current measurements of dietary caloric intake.

10. Future Scope

With the increase of individuals having an interest in the culinary world, the demand for recipe and lifestyle applications have increased. As we adapt to the changes around us during these trying times, many have also taken an interest in home-cooking. However, it may be challenging, especially for beginners to brainstorm recipes for cooking as they may not be equipped with the proper ingredients to do so. In this paper, we propose Feast In, a platform for web which aims to meet a user's needs for home-cooking. The platform focuses on three unique features which make Feast In more than just the average recipe platform. Firstly, an improved search algorithm which goes beyond searching for keywords would help users narrow down recipes which they can use in their kitchen. Next, customization features which would create a personalized experience, specifically towards recipes results. This would provide individuals who may face allergies or dietary restrictions an improved experience as they would not have to browse through recipes which do not meet their needs. Lastly, the search-by-image function which utilizes image recognition and machine learning technologies. Users will be able to upload an image of food that they have come across and Feast In will return a list of results which matches the image uploaded. By conducting this, we were able to propose a unique lifestyle and recipe application which would aid users in searching for the perfect recipe.

11. Bibliography

- Wang, X., Kumar, D., Thome, N., Cord, M., & Precioso, F. (2015, June). Recipe recognition with large multimodal food dataset. In *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 1-6). IEEE.
- Chen, J., & Ngo, C. W. (2016, October). Deep-based ingredient recognition for cooking recipe retrieval. In *Proceedings of the 24th ACM international conference on Multimedia* (pp. 32-41).
- Wang, X., Kumar, D., Thome, N., Cord, M., & Precioso, F. (2015, June). Recipe recognition with large multimodal food dataset. In *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 1-6). IEEE.
- Maruyama, T., Kawano, Y., & Yanai, K. (2012, November). Real-time mobile recipe recommendation system using food ingredient recognition. In *Proceedings of the 2nd ACM international workshop on Interactive multimedia on mobile and portable devices* (pp. 27-34).
- Herranz, L., Min, W., & Jiang, S. (2018). Food recognition and recipe analysis: integrating visual content, context and external knowledge. *arXiv preprint arXiv:1801.07239*.
- Salvador, A., Gundogdu, E., Bazzani, L., & Donoser, M. (2021). Revamping cross-modal recipe retrieval with hierarchical transformers and self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15475-15484).
- Zhu, B., Ngo, C. W., & Chan, W. K. (2021). Learning from Web Recipe-image Pairs for Food Recognition: Problem, Baselines and Performance. *IEEE Transactions on Multimedia*.

Appendix

A. Source Code

Project source code and dataset required for the project is uploaded in the following link
https://github.com/smartinternz02/SI-GuidedProject-49318-1652766389