# CS7643 Final Project Proposal

Sai Chen schen721@gatech.edu
Ruiyang Li rli386@gatech.edu
Shiya Song ssong331@gatech.edu
Bing Yang byang322@gatech.edu

October 20th, 2021

## 1 Team Name

CS7643 Food Images to Recipe Group (FIRG)

## 2 Project Title

***Use Joint Embedding to Transform Food Images to Recipes***

## 3 Project Summary

In this project, our goal is to build a deep learning model for transforming food images to the matching cooking recipes. The recipe here contains both the ingredients as well as the instructions to cook the food. We will implement the model described in [1]. We plan to make some modifications listed below and evaluate the performance on several related datasets.

## 4 Approach

The model we are going to implement is from [1]. The basic idea is to build a joint embedding model to project the cooking recipes and the associated food images into a common space during training. We will use the Recipe1M dataset to train our model. Specifically, we will do the following during training:

1. Use transfer learning to extract features from food images using existing image classification models.

2. Use various Long Short-Term Memory (LSTM) models to extract representations of ingredients and cooking instructions in cooking recipes.

3. Matching the above two learned representations by minimizing the cosine differences between them.

Besides what have been done in the paper [1], we plan to make following modifications

1. Try different state-of-art image classification models to evaluate the influence on the performance.

2. Quantify the performance on different recipe datasets (Food-101, Chefkoch.de) to evaluate the generalizability of the model.

# 5 Related work

Recipe extraction from food images has been formulated as an image classification problem before [2]. It is found that the limiting factor in this problem is the size of the dataset used for training deep learning models [1]. We mainly use ideas and dataset from [1, 3].

# 6 Dataset

The main dataset for training is Recipe1M (`http://pic2recipe.csail.mit.edu/`). The two additional datasets we want to use are listed below.

- Food-101: `https://data.vision.ee.ethz.ch/cvl/datasets_extra/food-101/`

- Chefkoch.de: `https://towardsdatascience.com/this-ai-is-hungry-b2a8655528be`

# 7 Group Members

- Sai Chen, schen721

- Ruiyang Li, rli386

- Shiya Song, ssong331

- Bing Yang, byang322

# References

[1] Amaia Salvador, Nicholas Hynes, Yusuf Aytar, Javier Marin, Ferda Ofli, Ingmar Weber, and Antonio Torralba. Learning cross-modal embeddings for cooking recipes and food images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[2] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101 – mining discriminative components with random forests. In *European Conference on Computer Vision*, 2014.

[3] Javier Marin, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019.