Team: DeepChef

Team Members: Bilal Mawji, Franz D Williams, Jiangqin Ma

Project Title: Deep Image-to-Recipe Translation

Project Summary:

Just like the saying goes, "You Are What You Eat," the meaning behind this phrase could not be truer in more ways than one. While it is true that our bodies are composed of the nutrients we consume, our identities are also shaped by the ingredients, the preparation, and the cooking of food we make daily. Recipes that have been passed down generation to generation have meaning in our lives – they are the fond memories of parents, grandparents, aunts, and uncles who have shared meals with us. For someone who has lost their old recipe or for someone else starting life in a new city or a new country, they may not have access to a recipe of their favorite dish. All that they have is a picture of them smiling with a relative eating their favorite food.

With our Image-to-Recipe Translation, we hope to give those who have pictures of their favorite dishes a chance at creating a dish that they have enjoyed eating. We would like to be able to provide ingredients, preparation, and cooking instructions for a recipe that would closely match the food in an image. We would hope to give someone a chance to enjoy the same bite of food that their mother made for them years ago when they were a child.

The task of generating ingredients and cooking instructions from an image is an intersection of two machine learning fields: computer vision (CV) and natural language generation (NLG). Deep CV methods have improved considerably since the advent of convolutional layers. Deep NLG has traditionally used recurrent layers, but more recently, attention-based methods have proven to be more powerful. Transfer learning also allows for reuse of previously trained networks to extract bottleneck features as part of a deep learning pipeline. We draw inspiration from and seek to reproduce the architecture found in [2].

Approach:

Our approach to solving Image-to-Recipe Translation is composed of three stages. Stage 1 is our primary goal and involves predicting ingredients from a food image. Stage 2 and 3 are stretch goals, with stage 2 involving generation of a full set of recipe steps and recipe title from the food image and predicted ingredients list. Stage 3 focuses on deployment and user feedback.

Stage 1: Ingredient Prediction from Food Images

1. Data collection and preprocessing:
   Gather a diverse dataset of food images with corresponding ingredient labels. Preprocess the images by resizing, normalizing, and augmenting to improve model robustness. Extract textual ingredient labels from the dataset. Preprocessing will also include cropping, mirroring, and whitening images. We will also be preprocessing the ingredients [2] to merge ingredients that share "*N*" words at the start or end of their name. This will reduce ingredients to their more basic forms (e.g., "finely grated cheese" and "coarsely grated cheese" become "grated cheese").
2. Develop ingredient model:
   Utilize Convolutional Neural Networks (CNNs) for image classification. Fine-tune a pre-trained CNN (e.g., ResNet, Inception) on the food image dataset for image feature extraction. Add a custom output layer for ingredient prediction.
3. Ingredient model training methodology:
   Split the dataset into training, validation, and test sets. Then, utilize image features generated from the pre-trained CNN with a fully connected multi-label classifier to predict ingredients. Train the CNN-classifier pipeline to predict ingredients from food images using an appropriate loss function (e.g., binary cross-entropy).

4. Ingredient prediction and evaluation:
   Input food images to the trained CNN model and obtain lists of predicted ingredients and associated confidence scores. Apply a threshold to filter out low-confidence predictions, and use appropriate evaluation metrics (e.g., accuracy, F1-score, Hamming score, or intersection over union).

Stage 2: Recipe Generation from Predicted Ingredients and Image Features

5. Develop and train recipe instruction generation model:
   Implement a text instruction generation model utilizing Recurrent Neural Network (RNN) techniques or Transformer-based architectures. Train the instruction generation model on the recipe dataset using predicted ingredients and image bottleneck features as input and recipe instructions as output.
6. Post-processing and evaluation:
   Refine the generated recipe instructions, ensuring coherence and clarity. Format the instructions into a readable recipe format. Evaluate the quality of generated recipes by human judgment and automated metrics (e.g., BLEU, METEOR).

Stage 3: Validate that application interface is user-friendly, and recipe generated is convincing

7. Deployment and Integration:
   Create a user-friendly interface or application where users can upload food images. Implement the ingredient prediction and recipe generation pipeline as a service. Provide users with generated recipes based on their uploaded images. Collect user feedback to improve the recipe generation system. The user feedback would come from friends or family members experienced with cooking who can provide a sanity check and validate whether or not they would follow the instructions created by the model.
8. Fine-tuning and Continuous Learning:
   Periodically retrain the ingredient prediction model and recipe generation model with new data to improve accuracy. Fine-tune models based on user feedback and usage patterns.

Resources/Related Work:

[1] "Image-to-Recipe Translation with Deep Convolutional Neural Networks" (Muriz, 2018)

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

Here, Muriz presents a retrieval-based method for translating food images into recipe instructions. One interesting aspect of this article is the use of TF-IDF (term frequency – inverse document frequency) to identify important terms in recipe names when exploring the data and as a preprocessing step for dimensionality reduction.

[2] "Inverse Cooking: Recipe Generation from Food Images" (Salvador et al., 2019)

https://research.facebook.com/publications/inverse-cooking-recipe-generation-from-food-images

In "Inverse Cooking: Recipe Generation from Food Image," Salvador et al. present a system that generates a cooking recipe given an image as input. Their system is a multi-step process that involves a pretrained image encoder (ResNet-50) that extracts bottleneck features from input images. Those features are then passed to a model that generates ingredients. These ingredients are then encoded and combined with the image bottleneck features and passed to a transformer architecture that generates the recipe title and instructions.

Salvador et al. utilize the Recipe1M dataset and emphasize the importance of data cleaning and preprocessing. For example, the raw ingredient information in their dataset contained over 400 different kinds of cheese. To reduce the number of unique ingredients, Salvador et al. Merge ingredients that share the first or last two words and remove ingredients that appear less than 10 times in the dataset.

The authors compare this new method to retrieval-based methods such as in [3] and find that it both improves ingredient prediction performance and generates more realistic recipes.

[3] "Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images" (Salvador et al., 2017)

http://pic2recipe.csail.mit.edu

"Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images" is an earlier article by Salvador et al., where a large-scale recipe image and text dataset is introduced. The authors explore a retrieval-based method that utilizes a learned joint embedding between recipe ingredients, instructions, and images.

[4] Evaluation metrics references:

- https://mmuratarat.github.io/2020-01-25/multilabel_classification_metrics
- https://stackoverflow.com/questions/69853180/tensorflow-multi-label-accuracy-metrics
- https://stats.stackexchange.com/questions/233275/multilabel-classification-metrics-on-scikit

Datasets:

Kaggle Food Ingredients and Recipes Dataset with Images
(https://www.kaggle.com/datasets/pes12017000148/food-ingredients-and-recipe-dataset-with-images)

Recipe1M (Original web article: http://pic2recipe.csail.mit.edu/, dataset currently available from http://wednesday.csail.mit.edu/temporal/release/)