Zah Andrei
Tirtan Gabriela

IAPSET, 2nd Year
Machine Learning

Project Proposal: Recipe Generator from Fridge Photos

1. Problem Presentation

Many people struggle with meal preparation, people are often not sure what recipes they could cook based on their available ingridients. This problem often causes people to lean on food delivery apps in order to satisfy their nutritional needs which leads to the food they already have spoiling and forces them to throw it away.

The intersection of computer vision and natural language processing offers a compelling solution: an intelligent system that can identify ingredients from refrigerator photos and generate appropriate recipes. Such a system addresses multiple parts of this problem: reducing food waste by suggesting recipes for existing ingredients, simplifying meal planning, and making cooking knowledge more easily available for individuals with varying culinary experience and skills.

The purpose of this project is to develop a bare bones machine learning based application that combines image classification for ingredient detection with intelligent recipe matching. The system will take photographs of available ingredients as input and output ranked recipe suggestions from an existig recipe database with detailed cooking instructions.

2. Dataset

The project will utilize multiple complementary datasets to address both the vision and language components:

For ingredient recognition, the food subset of the Open Image dataset will serve as the primary training source, containing approximately 9,200 images across 54 food categories and providing bounding box annotations for common ingredients[5]. This dataset provides diverse real-world food imagery of individual ingredients and some dishes. Additionally, the Recipe1M+ dataset [2] will be leveraged, offering over 1 million recipes with associated ingredient lists and cooking instructions, making it ideal for the recipe generation component.

3. Baseline Model & Advanced Models

The project will contain two different modules that will in the end be used in conjunction to accomplish the goal of generating recipes based on indrigients detected:

1. **Ingredient Detection:** The initial ingredient detection module will employ implement YOLO (You Only Look Once) or Faster R-CNN for simultaneous detection and localization of multiple ingredients in a single image [4].
2. **Recipe Matching:** The outputs of the Ingredient detection model will provide the input for the initial recipe recommendation system, which will implement a rule-based matching algorithm in order to connect the ingridients detected with the recipes existing in the database. This approach involves computing ingredient overlap scores between detected ingredients and recipe requirements. Recipes will be ranked based on the percentage of ingredients available, with penalties for missing critical components.

4. Evaluation Metrics

The project will employ multiple evaluation metrics to assess different aspects of system performance:

1. **Ingredient Detection:** Top-1 and Top-5 classification accuracy, mean Average Precision (mAP) for object detection variants, and confusion matrices to identify commonly misclassified ingredients.
2. **Recipe Matching:** Relevance scoring based on metrics such as ingredient coverage percentage, recipe feasibility (can it be made with available ingredients), and diversity of suggestions.

5. Expected Challenges

Several challenges throughout the development process may cause problems: refrigerator photos present significant challenges including occlusion varying lighting conditions, different packaging types, and the need to distinguish between visually similar ingredients. Another part that is likely to cause problems while training the model is the fact that common ingredients like eggs and milk will be overrepresented in training data compared to specialty items, potentially biasing the model toward generic recipes.

References

[1] Bossard, L., Guillaumin, M., & Van Gool, L., Food-101 – Mining Discriminative Components with Random Forests. European Conference on Computer Vision (ECCV), 2014.

[2] Salvador, A., et al., Learning Cross-Modal Embeddings for Cooking Recipes and Food Images. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[3] Kuznetsova, A., et al., The Open Images Dataset V4. International Journal of Computer Vision, 128, 1956–1981, 2020.

[4] Redmon, J., et al., You Only Look Once: Unified, Real-Time Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[5] https://github.com/WuXinyang2012/openimages-food-subset

[6] Raffel, C., et al., Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research, 21(140), 1-67, 2020.