Ingredient Identification and Recipe Recommendation: How Computer Vision has been used in the culinary world.

Ryan Darrow
Boston University
Darrowry@bu.edu

Introduction

In recent years, the advancement in machine learning models have allowed many industries to leverage these models to improve aspects of the industry that involved so much data that it was hard for a person to quickly or accurately parse through the data to reach appropriate conclusions. One aspect of machine learning, computer vision, has allowed the models to take images or videos as inputs, which contain a vast amount of visual data, and allow humans to parse through this information in a much more efficient manner than before. Computer vision has been used for a variety of applications, including facial recognition, autonomous vehicles, and cancer detection, producing promising results. Additionally, in recent years there have been efforts by people in the culinary world to adapt this technology for ingredient identification and recipe recommendations. In this paper, I will review previous efforts with regards to these two aspects of computer vision, summarizing their main findings and future considerations.

Literature Review

For this literature review, I have found four previous papers that looked to leverage machine learning and computer vision to help identify ingredients from pictures of food or looked to recommend recipes based on identified ingredients from images. Citations to each paper can be found in the References section below:

Deep-based Ingredient Recognition for Cooking Recipe Retrieval [1]:

In this paper, the authors propose deep learning architectures for the simultaneous learning of both ingredient recognition and food categorization by exploiting the mutual relationship between the two. Their main goal is to solve the issue of zero-shot retrieval, which is an issue that arose with other models that focused on the recognition of food categories based on the dish appearance without explicit analysis of the ingredients. Their framework consisted of two modules: the first one focusing on ingredient recognition as a problem of multi-task learning using deep convolution neural network (DCNN), while the second module performs zero-shot retrieval by matching the predicted ingredients against over 60,00 Chinese recipes. To evaluate the performance of their models, the average Top-1 and Top-5 accuracies are adopted for food categorization, whereas micro-F1and macro-F1were used for ingredient recognition.

Looking at the performances of the different architectures that were employed for their experiments, their Arch-D design, which performed the best for both food categorization and ingredient recognition. As opposed to the other architectures employed, Arch-D had a peculiarity

that the shared layer can correspond to the high or mid-level features common between two tasks at the early stage of learning, while the private layer preserves the learning of specialized features useful for optimizing the performance of each task. In terms of quantitative performance, Arch-D showed around 97% Top-5 and 82% Top-1 accuracies with regards to food categorization, to go along with 70% Micro-F1 and 44% Macro-F1 accuracies for ingredient recognition, outperforming other measures in these accuracies by at least 5-10%. Overall, their models showed significant statistical improvements when compared the best single-task model.

There were a couple issues that they identified as areas for future research. Both the cooking method (frying, steaming, etc.) and the way ingredients were cut (chop, slice, mince, etc.) were not considered in the development of the deep learning architecture that was put together for this project. Due to the variance in the appearance of the dishes due to these two factors, their model could not distinguish recipes from these types of dishes. Another issue with their model was that it could not deal with ingredients that were not visible or observable from dishes (honey, oils, etc.). Their consideration for future research involved learning about how bridge the gap between dish pictures and corresponding textual recipes to hopefully deal with the issue of unknown foods and ingredient labels.

From Market to Dish: Multi-ingredient Image Recognition for Personalized Recipe Recommendation [2]:

As opposed to the first paper which focused on ingredient identification and zero-shot retrieval using a deep convolution neural network, this paper focuses on a multi-ingredient recognition solution by using a spatial-regularized multi-label classification network. More specifically, the authors wanted to focus on ingredient identification in non-controlled environments and for non-cooked foods so that they could promote healthier ingredient choices. In order to employ more robust recipe recommendation, they implemented both a Neural network-based Collaborative Filtering (NCF) and Generalized Matrix Factorization (GMF) to capture the interactions between the user and item and leverage the users' preferences.

To evaluate the ingredient recognition experiments, macro/micro precision (P C/P-O), macro/micro recall (R-C/R-O), macro/micro F1 measure (F1-C/F1-O), and Mean Average Precision (mAP) were measured for performance comparison. When looking at their results, their customized Spatial Regularization Network (SRN) network outperformed all the baseline models for ingredient recognition for all metrics listed above on the MV80-Market dataset, which contains over 16000 images for a variety of vegetables from marketplaces in China. For recipe recommendation, the evaluation metrics used were Hit Ratio (HR) and Normalized Discounted Cumulative Gain (NDCG). However, despite their model performing better on both metrics when compared to a base model, the scores were still low due to the sparseness of the data. For future work, they wanted to improve the recipe recommendation system by redesigning the system to consider the combination of food ingredients, nutrition and user special needs in order to achieve the balance between nutrition intake and user preference.

Recipe Retrieval with Visual Query of Ingredients [3]

In this paper, the authors propose a framework that allows a user to take pictures of available ingredients and use their system to identify the potential recipes that can be produced. To learn representations of the ingredients and dishes, the team used a convolutional neural network architecture similar to a query-document matching model. More specifically, ingredient images were treated like a set of query terms and the images of dishes as the collection of documents. The dataset used to evaluate their model was Recipe1M, which at the time consisted of 1,029,720 structured cooking recipes and 887,536 associated images collected from popular cooking websites. However, one thing noted is that the dataset did not include images of the ingredients; to fill this gap, images from Google image search were used.

To quantify the results, two metrics were used: Median retrieval rank (MedR) and recall at top K (R@K). MedR is the median rank position where the right recipe is returned, meaning a lower score indicates higher performance, and R@K represents the fraction of times a correct recipe is found within the top-K retrieved candidates. Comparing the results from a variety of different models, their proposed architecture Img2img showed similar performance to other baseline models such. However, when their model was improved with joint neural embedding and semantic regularization, it performed better than all the other models tested. Overall, they found that their model could capture the visual property of input ingredients and retrieve matching and similar dish images for some well-defined cases. Future improvements mentioned involved incorporating images that vary cooking styles and plating methods, along with using more authentic ingredient photos, rather than the images collected from Google.

Conclusions

Overall, there has been a great deal of research into the use of computer vision models to help users identify ingredients and corresponding recipes from images. Many underlying model types were used (neural network, classification network, etc.), but all produced positive results in terms of proper identification and recommendation. However, there were still a couple of obstacles that each experiment ran into. For one, two of the studies, there were issues with ingredient identification when looking at pictures that have ingredients prepared in different way or the cooking method used for the recipe. Additionally, there were some ingredients that the models found difficult to identify, such as some oils, spices, and honey. Also, none of the studies looked to combine their methods with textual sources, although it was mentioned as a possible future research by one.

References

1. Chen, J. & Ngo, C.W. "Deep-based Ingredient Recognition for Cooking Recipe Retrieval". MM '16: Proceedings of the 24th ACM international conference on Multimedia, 2016, pp. 32-41. https://dl.acm.org/doi/pdf/10.1145/2964284.2964315

2. L. Zhang, J. Zhao, S. Li, B. Shi and L. -Y. Duan, "From Market to Dish: Multi-ingredient Image Recognition for Personalized Recipe Recommendation," 2019 IEEE International Conference on

Multimedia and Expo (ICME), Shanghai, China, 2019, pp. 1252-1257, doi: 10.1109/ICME.2019.00218.

3. Yen-Chieh Lien, Hamed Zamani, and W. Bruce Croft. 2020. Recipe Retrieval with Visual Query of Ingredients. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20). Association for Computing Machinery, New York, NY, USA, 1565–1568. https://doi.org/10.1145/3397271.3401244