Poster: Smart Cook: Making Cooking Easier with Multimodal Learning

Hyoyoung Lim
hyoyoung.lim@colorado.edu
University of Colorado Boulder
Boulder, Colorado, United States

Xiaolei Huang
xiaolei.huang@colorado.edu
University of Colorado Boulder
Boulder, Colorado, United States

Samuel Miller
samuel.j.miller@colorado.edu
University of Colorado Boulder
Boulder, Colorado, United States

Joshua Edelmann
joshua.edelmann@colorado.edu
University of Colorado Boulder
Boulder, Colorado, United States

Timothy Euken
timothy.euken@colorado.edu
University of Colorado Boulder
Boulder, Colorado, United States

Stephen Voida
svoida@colorado.edu
University of Colorado Boulder
Boulder, Colorado, United States

ABSTRACT

Learning how to cook presents at least two significant challenges. First, it can be difficult for novices to find appropriate recipes based on the ingredients available in one's pantry and/or refrigerator. Second, it can be difficult to focus on cooking tasks and following a recipe at the same time. In this poster, we present the design process and implementation of a system that uses deep learning to address the first of these two problems. Our initial design work focuses on streamlining the process of entering and tracking potential ingredients on hand and determining appropriate recommendations for recipes that utilize these ingredients. Here, we present the current state of our project, explaining in particular our contributions to minimizing the overhead of tracking kitchen ingredients and converting this inventory information into effective recipe recommendations using a multimodal machine learning approach.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

KEYWORDS

Recipe recommendation system; ubiquitous computing design; domestic computing; deep learning

ACM Reference Format:
Hyoyoung Lim, Xiaolei Huang, Samuel Miller, Joshua Edelmann, Timothy Euken, and Stephen Voida. 2019. Poster: Smart Cook: Making Cooking Easier with Multimodal Learning. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct), September 9–13, 2019, London, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3341162.3343836

UbiComp/ISWC '19 Adjunct, September 9–13, 2019, London, United Kingdom

ACM ISBN 978-1-4503-6869-8/19/09…$15.00
https://doi.org/10.1145/3341162.3343836

1 INTRODUCTION

Supporting novices in identifying *what to cook* with domestic ubiquitous computing technologies requires developing new methods that model on multiple signal sources, especially in language and vision. In this poster, we present the design process and implementation of a system that uses deep learning to assist novices in finding appropriate recipes based on the ingredients that they have on hand. Our initial design work focuses on streamlining the process of entering and tracking these potential ingredients and determining appropriate recommendations for recipes that utilize them. Here, we focus on the development of image recognition and recipe recommendation models that reduce the overhead to users of "telling" the system what they have in the pantry and/or refrigerator. One possible solution to this challenge is to use multimodal machine learning [1]: the construction of machine learning models that utilize multiple environmental signals or modalities, including voice, images and text.

In this research, we present the design of a recipe recommendation system focused on collecting input about available ingredients quickly and easily. Our approach uses deep learning to facilitate ingredient recognition, streamlining one of the key user interactions in the system. Here, we present the current state of the project and the research activities that we have undertaken, to date:

- First, we brainstormed and sketched a breadth of ideas for technologies to assist novice cooks in the kitchen.
- Second, we presented our sketches to friends and family members, using their feedback to focus in on addressing the most pressing problem(s). In this case, we ended up focusing more tightly on the process of selecting recipes based on available ingredients, and the associated technical problems of identifying and monitoring what ingredients are available in this kitchen at a given moment.
- Finally, we developed a prototype system based on our design explorations and began the process of evaluating our system against the various recipe recommendation sites and apps already on the market.

Here, we present the current state of the project, explaining in particular our contributions to minimizing the overhead of tracking kitchen ingredients and converting this inventory information into effective recipe recommendations using a multimodal machine learning approach.

2 RELATED WORK

We base our research on a significant body of prior ubiquitous computing research literature about augmenting the refrigerator, using computational approaches to access a catalog of recipes and model recipe preferences, and incorporating different kinds of input and feedback from the potential users of such a system.

2.1 Augmenting the Refrigerator

Xie et al. [19] use RFID technology and cluster analysis to identify and locate food that is stored inside the refrigerator. Their system, known as *iFridge*, is an intelligent system that leverages technology to collect food information, understand the user's activity, locate specific foods, and then use "cooking recipe recommendations" to recommend recipes based on a user's eating habits. Floarea et al. [6] present the design of a smart refrigerator that is connected to the Internet of Things. Their smart refrigerator is able to collect information on the items stored inside and process the information into data that can be conveyed to the user through an IoT platform. Nayak et al. [14] proposed an Intelligent Refrigerator that can monitor the quantity of particular food items stored inside. However, the system does not account for the quality or recency/freshness of its contents, limiting its applicability beyond providing simple counts of cold-storage food items. Matthias et al. [16] conducted an ethnographic study of the interactions between humans and refrigerators, and characterized participants' desired refrigerator functionality.

2.2 Big Data and Cooking

Javier et al. [13] present the large-scale *Recipe1M* dataset, a corpus of one million structured cooking recipes and associated images. Many researchers are already working to incorporate this dataset into cooking tools and platforms (e.g., [3, 8]). *Recipe1M* data can also be used to evaluate the accuracy and effectiveness of cooking guidance and recommendation systems like ours.

2.3 Multimodal Machine Learning

Modeling the recipe selection process can be viewed as a multimodal machine learning problem. Here, *modal* refers to the basic human sensory channels, such as visual and language signals; thus, multimodal machine learning is an approach in which machine learning models are developed based on multiple heterogeneous inputs [1]. Yet, most existing "deep" machine learning research only focuses on unimodal channels; that is, language-only [4] or vision-only [9] models. In this study, we focus on both inputs from language and visual signals and train the heterogeneous inputs jointly. The advances of deep learning, especially convolutional neural networks [12] and recurrent neural networks [10], bring both challenges and opportunities in jointly modeling different signals, in which image and text inputs are encoded using fixed-length vectors. Further, aligning heterogeneous embedding spaces of images and text is also a challenge. Recent studies have proposed a joint training framework on language and image inputs through semantic similarities and co-training regularization [17]. We apply this approach in our model development, as well.

2.4 Subjective Preference and Decision Models

Ueda et al. [18] propose a system in which recipes can be searched according to user mood. In their system, six aspect of a user's mood are computed, and the resulting mood characterization is applied when making decisions about a potential menu. Of course, there is a limitation that it is difficult to satisfy everyone, particularly because of cultural and religious differences.

3 SYSTEM DESIGN

3.1 Design Rationale and Initial Sketches

Our team began our design exploration using a variety of sketches and storyboards suitable for eliciting informal feedback from friends and family members about the high-level problem of supporting novice cooks. We initially focused on addressing the primary issue of deciding *what to cook* and *how to cook it efficiently*. In order to foster conversations about these challenges, we created a suite of sketches for a refrigerator interface that would gather information about the contents of the refrigerator using cameras and/or sensors then display available ingredient and recipe recommendation information on a device-mounted display.

Discussions around these sketches helped us to refine the scope and focus of our project. For instance, despite our initial intuition about instrumenting the refrigerator as the primary locus of food preparation, our informants expressed concern that not all potential ingredients are stored in the refrigerator—some food is stored in the pantry, cabinets, and other places. As a result, we realized that a comprehensive solution would require designing multiple devices, one for each area of the kitchen; creating a single device that is somehow able to keep track of everything in the kitchen; or to pursue a combination of mobile and fixed technologies.

During our second round of design iteration, we created sketches for a mobile phone application (Figure 1) that would, among other things, be able to scan images of food items or their UPC barcodes as a means of entering information into the system's food inventory database. We also explored technical issues/errors that various sensing and data collection platforms might raise, and we prototyped the user interactions related to such events. Finally, we sketched the

Figure 1: A mock-up of the mobile phone interface designed to elicit data collection about the contents of an individual's refrigerator.

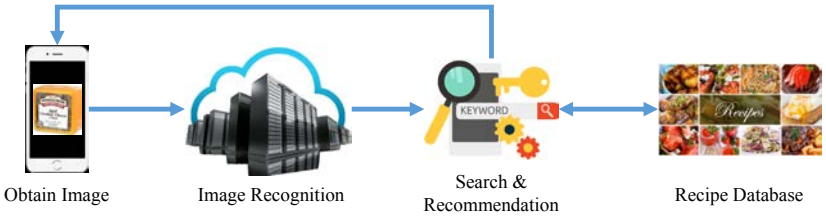

Figure 2: Our proposed system design aims to connect different user interfaces with a multi-purpose back-end, including food ingredient recognition model, search engine, recommendation system, and database.

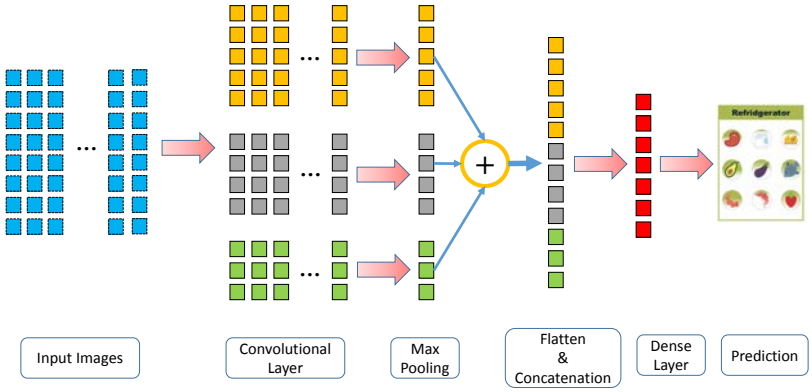

Figure 3: The proposed convolutional neural network model to identify potential ingredients.

concept of a main interface that would be mounted in the kitchen to provide a persistent display of the recipe and/or instructional tutorials without having to juggle a handheld device (i.e., phone or tablet) or have to wipe their hands off to touch a screen.

The breadth of sketches allowed our group to converge on a concrete direction for our final design: a system with two parts. The first part is the system's display, a kitchen-mounted recipe display system that can also serve as a hub for other useful information like ingredient inventory, recipe suggestions, and eating/cooking statistics, suggestions based on early-stage feedback. The second part of the system is a mobile application, which serves as the mechanism for capturing images of potential ingredients as an input to the recommendation engine. Here, we detail some of our initial experiments in developing this second component of the system—the portion that most directly leverages and embodies "deep" machine learning approaches.

3.2 Technology Development

The architecture of our ingredient recognition and recipe recommendation system is shown in Figure 2. Our system is based around four main modules: a series of user-facing interfaces, an image recognition model, a search and recommendation engine, and a database. We formalized and implemented the communication among the components of our system using RESTful services [15].

Our architecture supports multiple user interfaces, including a web-based portal accessible via a fixed, kitchen-based terminal (the main display, described above) or via users' mobile phones. Users can use the mobile interface to send refrigerator and pantry images to the image recognition model, while both interfaces can be used to review recipe recommendations from our recommendation engine.

Our image recognition module identifies potential recipe ingredients and feeds a list of candidate ingredients onward to our recipe search and recommendation engine. We designed and developed our image recognition model using deep learning techniques. Convolutional neural networks (CNNs), in particular, have shown promising results in the field of image recognition [1]. As a result, we designed a neural model using a CNN approach (Figure 3). We implemented the model using Keras [5]. It applies three convolutional layers with the same filter size (128) and different kernel sizes $(3 * 3, 4 * 4, 5 * 5)$ on the input images, which enables the system to extract features at multiple levels of details. We apply max pooling layers on the three convolutional layers and then concatenate the outputs. We learn a fixed-length representation via the dense layer. The dense layer learns the input with ReLU [7] and feeds it to a 128-node output layer. Finally, we feed this representation to a softmax function to make a final prediction. We used cross entropy and an Adam [11] optimizer to optimize our model. We set the learning rate as .0001 and left the other parameters set to Keras' defaults.

By rotating and zooming images, we created a random training set for the image recognizer. We trained our model with 2000 epochs and a batch size of 64. We tested our proposed model on the Food 101 data set [2], holding out 20% of the training set as a development dataset. We tested the classifier that performed best on the development set, instead of heavily tuning the model's parameters. Finally, when running the algorithm on the test set, we obtained an accuracy score of 52.63%, indicating that our model outperforms the original's accuracy (50.76%) [2] by 2%.

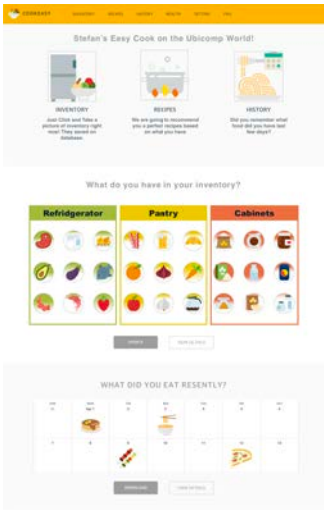

Figure 4: A "slideware" sketch of a proposed refrigerator interface for supporting novice cooks.

After receiving a list of recognized ingredients from the neural model, the search and recommendation system retrieves recipe information and generates a suite of cooking recommendations that are displayed to the user (a more refined mock-up appears in Figure 4). Our recipe database was derived from a previously-published dataset[1]. Our search and recommendation engine is based on Elasticsearch[2], with recipe data indexed by both *recipe name* and *ingredients*.

4 CONCLUSIONS AND FUTURE WORK

We designed our system based on the idea that the majority of users of our prototype would have little prior experience cooking. With this in mind, we focused on providing computational support for one of the most essential questions faced by novice cooks: *What can I cook?* Our design work prioritized minimizing the overhead of maintaining an accurate ingredient inventory within our system and leveraging deep learning techniques to assist with food identification and recipe selection tasks—all connected to interfaces that are intended to be simple enough for even novice cooks to use. We are also optimistic that these tools would also be of interest to more experienced chefs, enabling them to hone their skills with more advanced recipes, as well. Eventually, we would be interested in adding functionality to our system to enable tracking of left-over or older ingredients that are wasting away in the refrigerator and pantry, prioritizing recipe selection to use up these items or communicating issues of potential food safety.

There are a few notable differences between our system and existing, off-the-shelf recipe recommendation tools. For starters, our solution is more personalized and adaptive, as the system analyzes what (and how much) you have in the refrigerator and pantry and creates recipes tailored to the ingredients on hand; off-the-shelf sites can not do that, only providing recipe instructions that assume that all ingredients are available. Second, our system can tailor recipes based on the amount of ingredients currently available. For example, imagine that a user has one pound of ground beef in the freezer. Our system would allow the user to select a category of food, such as Mexican food, and our recipe recommendation tool can recommend recipes that are Mexican style, contain ground beef, and can be scaled up or down in volume to use up to one pound of meat. As a result, our system stands to reduce the amount of food waste that is generated each week and reduce how much money is spent at the grocery store. Finally, our decision to integrate a large and popular recipe database that is also at the center of several contemporary research efforts means that we will likely be able to augment our system to include step-by-step video instructions during meal preparation once other chefs create and share tutorial content to enrich this existing and open recipe data set.

[1]https://github.com/fictivekin/openrecipes
[2]https://www.elastic.co/

REFERENCES

[1] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2019. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Analysis and Machine Intelligence* 41, 2 (2019), 423–443.
[2] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101–mining discriminative components with random forests. In *European Conference on Computer Vision*. Springer International, 446–461.
[3] Micael Carvalho, Rémi Cadène, David Picard, Laure Soulier, Nicolas Thome, and Matthieu Cord. 2018. Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings. In *Proc. SIGIR 2018*. ACM, 35–44.
[4] Minsuk Chang, Vivian M Hare, Juho Kim, and Maneesh Agrawala. 2017. Recipescape: Mining and analyzing diverse processes in cooking recipes. In *Ext. Abstracts CHI 2017*. ACM, 1524–1531.
[5] François Chollet et al. 2015. Keras. https://github.com/fchollet/keras.
[6] Aurel-Dorian Floarea and Valentin Sgârciu. 2016. Smart refrigerator: A next generation refrigerator connected to the IoT. In *Proc. ECAI 2016*. IEEE, 1–6.
[7] Richard HR Hahnloser, Rahul Sarpeshkar, Misha A Mahowald, Rodney J Douglas, and H Sebastian Seung. 2000. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature* 405, 6789 (2000), 947.
[8] Luis Herranz, Weiqing Min, and Shuqiang Jiang. 2018. Food recognition and recipe analysis: integrating visual content, context and external knowledge. arXiv preprint. arXiv:1801.07239.
[9] Yoshiyuki Kawano and Keiji Yanai. 2014. Food image recognition with deep convolutional features. In *Ext. Abstracts Ubicomp 2014*. ACM, 589–593.
[10] Chloé Kiddon, Luke Zettlemoyer, and Yejin Choi. 2016. Globally coherent text generation with neural checklist models. In *Proc. EMNLP 2016*. Association for Computational Linguistics, 329–339.
[11] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. arXiv preprint. arXiv:1412.6980.
[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Proc. NIPS 2012*. Neural Information Processing Systems Foundation, 1097–1105.
[13] Javier Marin, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. 2018. Recipe1M: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images. arXiv preprint. arXiv:1810.06553.
[14] G Subramanya Nayak, C Puttamadappa, et al. 2011. Intelligent Refrigerator with monitoring capability through internet. *Int. J. Comput. Appl.* 2 (2011), 65–68.
[15] Leonard Richardson and Sam Ruby. 2008. *RESTful web services*. O'Reilly.
[16] Matthias Rothensee. 2008. User acceptance of the intelligent fridge: empirical results from a simulation. In *Proc. IOT 2008*. Springer, 123–139.
[17] Amaia Salvador, Nicholas Hynes, Yusuf Aytar, Javier Marin, Ferda Ofli, Ingmar Weber, and Antonio Torralba. 2017. Learning cross-modal embeddings for cooking recipes and food images. In *Proc. CVPR 2017*. IEEE, 3020–3028.
[18] Mayumi Ueda, Yukitoshi Morishita, Tomiyo Nakamura, Natsuhiko Takata, and Shinsuke Nakajima. 2016. A recipe recommendation system that considers user's mood. In *Proc. iiWAS 2016*. ACM, 472–476.
[19] Lei Xie, Yafeng Yin, Xiang Lu, Bo Sheng, and Sanglu Lu. 2013. iFridge: an intelligent fridge for food management based on RFID technology. In *Ext. Abstracts Ubicomp 2013*. ACM, 291–294.