Design of a Multisensor System for a Smart Cooking Assistant

Gustavo Ruiz[1]
gustavoruiz064@csu.fullerton.edu

Sai Chaitanya Kilambi[2]
skilambi@fullerton.edu

Pratishtha Soni[2]
psoni@fullerton.edu

Kiran George[1]
kgeorge@fullerton.edu

Anand Panangadan[2]
apanagadan@fullerton.edu

[1]Electrical and Computer Engineering Department
California State University, Fullerton
Fullerton, California 92831, USA

[2]Computer Science Department
California State University, Fullerton
Fullerton, California 92831, USA

***Abstract*— Homeless people and people who have only recently escaped homelessness are at risk of being malnourished due to several factors. Personalized cooking recipe recommendation systems are one technology-based approach to provide information to such populations that can assist them with making more informed choices about their food. Current recipe recommendation systems do not provide any guidance on how to cook for a person who has limited experience with cooking. This work describes a system that acts as a smart cooking assistant. The main idea is for the system to observe the user perform the steps of cooking based on a recipe, and then provide automatic reminders on when to move to the next step. The system consists of both hardware and machine learning-based software components. The hardware consists of a camera, infrared thermal camera, and temperature sensor. These are integrated around a Raspberry Pi mini-computer. This suite of sensors is mounted over a stovetop and constantly monitors the cooking area, specifically the area where a cooking pot is over the stovetop. Convolutional Neural Network-based image processing algorithms are used to analyze the sequence of images from the cameras to identify the stage of cooking that the user is performing.**

Keywords—IoT, machine learning, Smart Home, supportive housing

I. Introduction

Homeless people and people who have only recently escaped homelessness are at risk of being malnourished due to several factors, such as low income, limited knowledge of nutrition, choice of food, and lack of cooking and storage facilities [1, 2]. Personalized cooking recipe recommendation systems [3, 4] are one technology-based approach to provide information to such populations that could assist them with making more informed choices about their food. With the increasing capabilities of Artificial Intelligence (AI) technologies, there are now also AI-based cooking recipe recommendation systems [5].

However, recipe recommendation systems do not provide any guidance on how to cook for a person who has limited experience with cooking. This paper describes a system that acts as a smart "cooking assistant". The main idea is for the system to observe the user perform the steps of cooking based on a recipe, and then provide automatic reminders on when to move to the next step. The system consists of both hardware and machine learning-based software components. The hardware consists of a camera, infrared thermal camera, and temperature sensor. These are integrated into a mini-computer (Raspberry Pi 4). This suite of sensors will be mounted over a stovetop and will constantly monitor the cooking area, specifically the area where a cooking pot is over the stovetop. Machine learning-based image processing algorithms are used to analyze the sequence of images from the cameras to identify in what stage of cooking the user is currently performing. The algorithms are classification algorithms that assign each frame from the video stream into one of a few classes, each corresponding to a stage in a particular recipe.

In this paper, we describe both the hardware design and the image classification approach used to identify the cooking stages. The contributions of this work are: (1) An approach to assistive cooking based on dividing cooking of simple dishes into a sequence of stages that can be detected based on objects on a stovetop, (2) Hardware design that integrates multiple sensors to collect data to perform the above detection, and (3) image classification using a convolutional deep neural network that is trained using hand-labeled data for detecting the sequence of stages for a simple recipe (cooking pasta).

II. Related Work

AI techniques, especially machine vision and image processing, have been used in certain aspects of food processing. The main tasks are identifying the type and quality of food, food grading, detecting locations of defective spots or foreign objects [6]. The paper presents Chinese recipes in a dataset where different images propose different stages of food. Different models are trained separately for different images in different categories, for example, beginning, intermediate and final stage[7].

However, these tasks do not correspond to the main steps of cooking at home for personal consumption. Relatively little work has been completed on assistive cooking systems. The Cognitive Orthosis for coOKing (COOK) is a stove-connected smart tablet application designed for individuals with cognitive deficits during meal preparation [8, 9]. The system monitors and tracks objects during cooking with real-time object detection and tracking techniques such as YOLO (You Only Look Once) and KCF (Kernelized Correlation Filter). The paper addresses various challenges associated with multiple object tracking, encompassing issues like objects appearing and disappearing, occlusion, and motion blur. Performance evaluations show that combining detection and tracking data significantly enhances the system's ability to identify and trace kitchen utensils [10].

Jelodar *et al*. [11] introduced 11 states that represent the most frequently used cooking objects and created a dataset of cooking-related images containing those objects. They used a Resnet-based deep model for object identification. Such systems require an object detection model that must be re-trained for cooking-specific objects.

In addition to Resnet, another commonly used object detection model is YOLO. Unlike conventional object identification methods, the YOLO model uses an individual network to detect objects over the whole image. The YOLO framework simplifies detection and classification tasks in one model , in comparison to traditional methods [12]. In our work, we use MobileNets [13] for image detection because of the need to implement the model in an embedded system.

III. System Design

III A. Hardware Design

Sensors

The MLX90614, an infrared thermometer, operates in a -70°C to 382.2°C range, suitable for cooking applications, with tested accuracy and ease of use. The DHT11 sensor measures temperature (0°C to 50°C) and humidity, crucial for distinguishing cooking stages like boiling, and complements other sensor data for enhanced system functionality.

Cameras

The OV5647 Mini Camera Module, designed for Raspberry Pi, captures high-resolution images (2592x1944) and 720P videos at 60FPS, essential for a cooking recognition ML model. The MLX90640 Infrared Camera tracks temperature changes from -40°C to 300°C with ±2°C accuracy, ideal for cooking monitoring.

Processing Unit

Initially utilizing an Arduino UNO R3, the system required more computational power for image recognition, leading to the adoption of the Raspberry Pi 4. This platform efficiently handles multiple sensor control scripts and runs ML image detection models.

Display

A 5-inch LCD display, coupled with a PCB, offers portability and facilitates testing of camera and sensor data. The PCB ensures a tidy arrangement and effective use of GPIO pins, compatible with the LCD's pin requirements. The setup is detailed in Fig. 1.

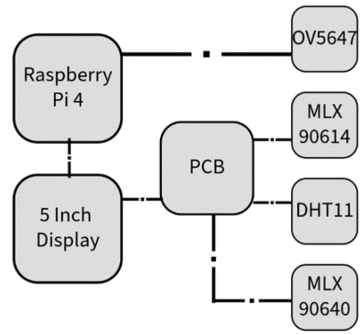

Fig. 1 Hardware Layout

Physical Design

The physical design of the hardware was constructed using SOLIDWORKS. It uses the base of the Raspberry Pi with a bracket for sensors. For testing purposes, a camera tripod was used for easy adjusting and sturdiness.A photograph of the prototype setup is shown in Fig. 3.

Fig. 2 Prototype hardware

III B. Image Processing

The process flow shown in Fig.3 for the cooking assistance system is depicted with a focus on data collection and processing. It integrates temperature data from the MLX90614 infrared sensor, humidity data from the DHT11 sensor, and visual data from the OV5647 Mini Camera Module and MLX90640 Infrared Camera. This data is subsequently processed by an Arduino UNO R3, utilizing the SSD MobileNet V2 FPNLite 320x320 model for image analysis. The processed information is then displayed on an LCD screen, providing real-time cooking guidance. This integration of multiple sensors and processing units forms the core of the cooking assistance application.

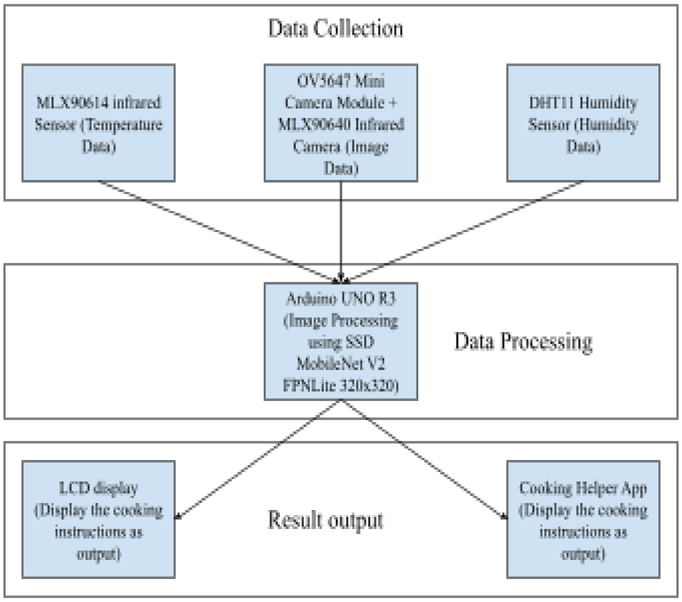

Fig. 3 Process Flow For the cooking assistance system

In this study, we employ the "SSD MobileNet V2 FPNLite 320x320", a pre-trained model from the Tensorflow-2 Model Zoo, to address the challenge of cooking stage identification, a subset of image classification problems. This model, leveraging a Single Shot Detector with MobileNet architecture and a feature pyramid network, is pre-trained on the COCO dataset, optimizing for rapid and efficient object detection on mobile and embedded devices. The model's efficacy is tested on a simplified culinary task—cooking pasta. We categorize the cooking process into six distinct stages, each characterized by specific items observable on the stovetop, ranging from an "Empty burner" to a "Pot with cooked pasta".

Image Collection

The collection of images for the dataset and labeling them is non trivial and when creating labels we need to consider scalability of the model for the future. For our present study we tried to detect several stages in cooking pasta and create labels for the collected images accordingly. Since our study focuses on creating an object detection model for supportive housing, the images are collected from a video recording that shows how to cook pasta and the parameters like orientation of the camera, type of pans used are kept identical to those of the supportive housing to better emulate the desired conditions and extract the frames using a python script. The dataset consists of 300 images which are categorized based on the stages of cooking pasta and what the camera sees at each and every instance namely “Empty burner”, “Empty pot”, “Pot with water”, “Pot with boiling water”, “Pot with pasta”, “Pot with cooked pasta”. Sample images from these classes are shown in Fig. 4.

Fig. 4 Sample collected images for Empty burner, Pot with boiling water, Pot with Cooked pasta, Empty pot, Pot with pasta, Pot with water respectively.

Labeling Images

The collected images are then annotated using the “LabelImg” package that allows us to manually encapsulate the portion of the image we want to label and give the encapsulated portion its appropriate label. This step is illustrated in Fig. 6. This process creates an XML file of the image we labeled. The dataset is then split into test and train along with their annotation.

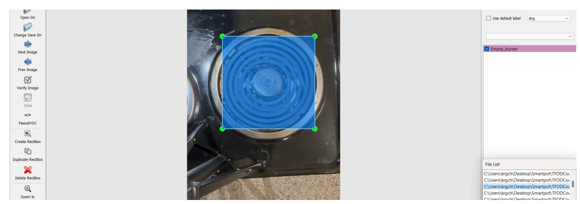

Fig. 5 Sample image depicting labeling using “LabelImg” package

Transfer Learning

MobileNetSSDv2 (MobileNet Single Shot Detector) is an object detection model with 267 layers and 15 million parameters pre-trained on COCO dataset, was used for training the stage identification model.The model was trained for 10,000 steps and was tested using both real time detection and test image dataset. The MobileNetSSDv2 pre-trained model has inbuilt image abstraction and adds additional noise to the dataset when using them for training thereby increasing the actual performance of the model even under different lighting conditions (Fig. 6).

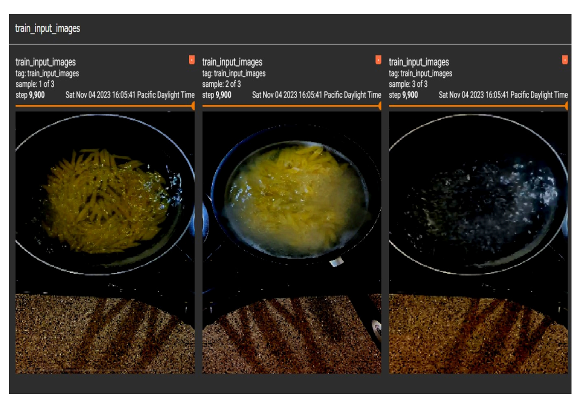

Fig. 6 Sample image depicting train input images after addition of noise

IV. Results and Discussions

The experiment begins by placing the sensors near a pot filled with water and recording data until the water boils.

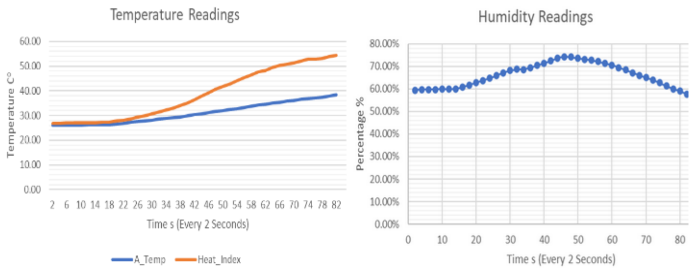

Fig. 8 Temperature and Humidity readings recorded every 2 seconds

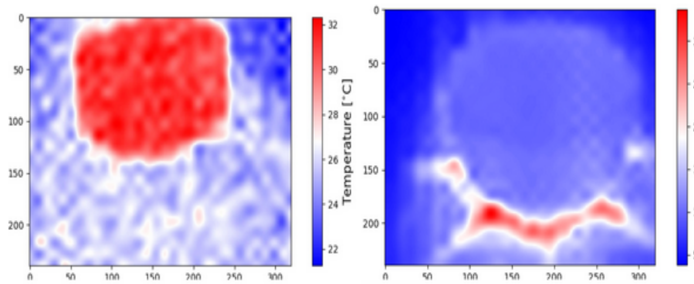

Fig. 7 MLX90640 Thermal Camera Starting Point

Figs. 8 shows the ambient temperature around the DHT11 and Heat_Index shows the combined index of the ambient temperature and the relative humidity. As shown in Fig. 7 The "highest" temperature the camera could detect was the water in the pan at 32 ºC. As the water starts to boil, we can also see that the highest temperature the thermal camera recorded was the flame below the stove above 300 ºC while the water ranged from 100 ºC to 170ºC.

Average Precision (AP)

The Average Precision at IoU thresholds from 0.50 to 0.95 for all object sizes with a maximum of 100 detections is 0.802. This means that, on average, the model correctly identifies and localizes objects in images 80.2% of the time.

Average Recall (AR)

The Average Recall at IoU thresholds from 0.50 to 0.95 for all object sizes with a maximum of 100 detections is 0.827, which means that the model consistently recalls 82.7% of the objects when considering a larger number of detections.

Loss Values

The Localization loss is 0.050686, which represents the error in localizing objects in images. The Classification loss is 0.147525, indicating the error in classifying detected objects. Regularization loss is 0.096303, representing the regularization term in the model. Total loss is 0.294514, which is the combined loss including localization, classification, and regularization components. The decrease in these losses are shown in Figs. 11-14.

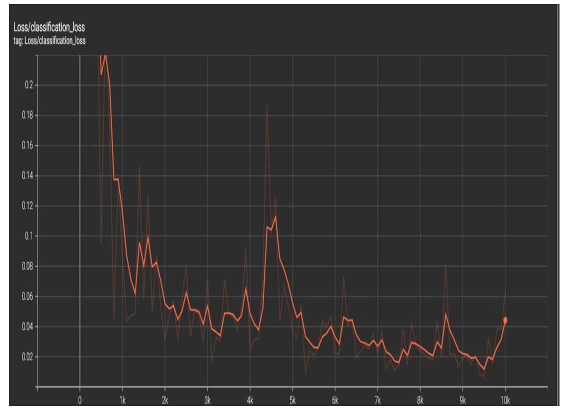

Fig. 11 Decrease in classification loss

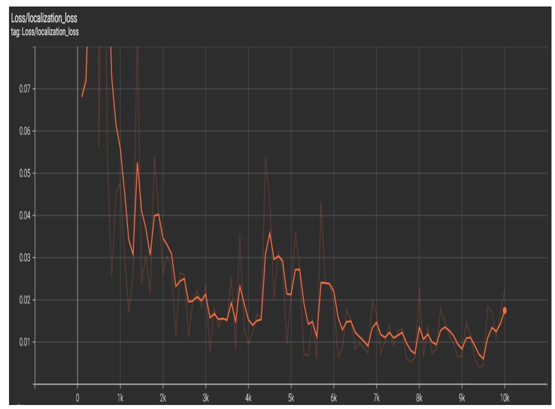

Fig. 12 Decrease in localization loss

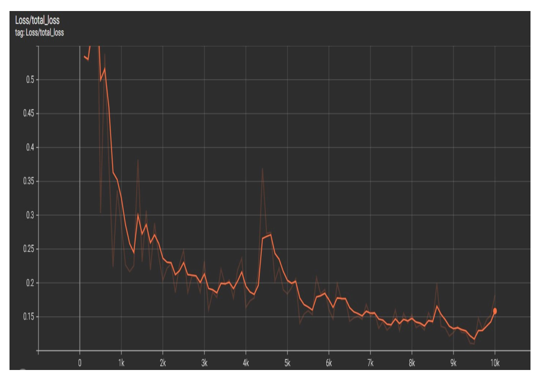

Fig. 13 Decrease in total loss

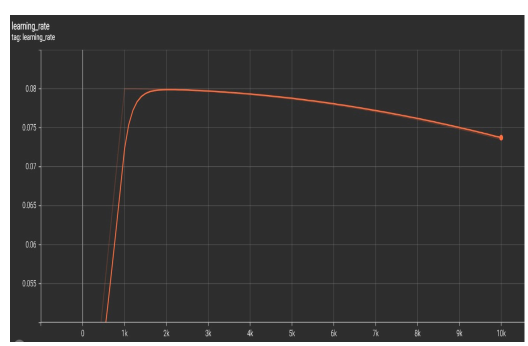

Fig. 14 Varying the Learning rate

Fig. 15 shows the detection of objects in a sample image. The model exhibits good overall performance with high precision and recall scores for medium to large objects. However, it performs less accurately with smaller objects. The model achieves 100% precision at a lower IoU threshold (0.50), indicating that it can identify objects accurately, even if their predicted bounding boxes slightly differ from the ground truth. The reported loss values also provide insights into the model's training and regularization.

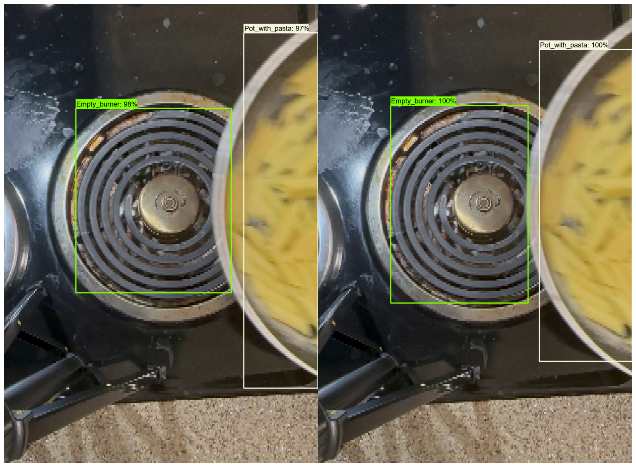

Fig. 15 Sample image depicting detection

V. Conclusion and Future Work

The hardware prototype is designed for rapid development and testing. The physical design will be redesigned for more harsh environments. In particular, the following issues will be explored.

1. An issue to consider is how to power the device. For testing purposes, a wall outlet was located near the stove. A battery pack did not last the time needed to start a cooking recipe.

2. The object detection model demonstrates high accuracy in real-time stage recognition in cooking, even with a single image input. It exhibits notable recall and precision, affirming its suitability for the intended application. However, the model's performance is compromised when encountering variations in cookware, such as changes from a dark-colored to a steel pot. This issue manifests in misidentifications, like mistaking a pot with water for an empty one. To mitigate this limitation, retraining the model with a more diverse dataset encompassing various pot types is recommended

3. The temperature and humidity data collected will be integrated in the object detection model.

4. It is observed that when the camera is placed directly above the pot the steam generated by the boiling water obstructs the vision and may fail in properly detecting the stage. This problem can be solved by also including the temperature and the humidity data.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. 2125654.

References

[1] Sprake, Eernest F., John M. Russell, and Maria E. Barker. "Food choice and nutrient intake amongst homeless people." Journal of Human Nutrition and Dietetics 27.3 (2014): 242-250.

[2] Wiecha, Jean L., Johanna T. Dwyer, and Martha Dunn-Strohecker. "Nutrition and health services needs among the homeless." Public Health Reports 106.4 (1991): 364.

[3] Pecune, Florian, Lucile Callebert, and Stacy Marsella. "A Recommender System for Healthy and Personalized Recipes Recommendations." HealthRecSys@ RecSys. 2020.

[4] Ge, Mouzhi, Francesco Ricci, and David Massimo. "Health-aware food recommender system." Proceedings of the 9th ACM Conference on Recommender Systems. 2015.

[5] Yera, Raciel, Ahmad A. Alzahrani, and Luis Martínez. "Exploring post-hoc agnostic models for explainable cooking recipe recommendations." Knowledge-Based Systems 251 (2022): 109216.

[6] Zhu, Lili, et al. "Deep learning and machine vision for food processing: A survey." Current Research in Food Science 4 (2021): 233-249.

[7] ZHANG, Yixin, Yoko YAMAKATA, and Keishi TAJIMA. "Stage-Aware Recognition Method for Foodstuffs Changing in Appearance in Different Cooking Stages on Chinese Recipe."

[8] Giroux, Sylvain, et al. "Cognitive assistance to meal preparation: design, implementation, and assessment in a living lab." 2015 AAAI Spring Symposium Series. 2015.

[9] Yaddaden, Amel, et al. "Using a cognitive orthosis to support older adults during meal preparation: Clinicians' perspective on COOK technology." Journal of Rehabilitation and Assistive Technologies Engineering 7 (2020): 2055668320909074.

[10] Ngankam, Hubert, et al. "Real-Time Multiple Object Tracking for Safe Cooking Activities." International Conference on Smart Homes and Health Telematics. Cham: Springer Nature Switzerland, 2023.

[11] Jelodar, Ahmad Babaeian, Md Sirajus Salekin, and Yu Sun. "Identifying object states in cooking-related images." arXiv preprint arXiv:1805.06956 (2018).

[12] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[13] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).