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, one of which is limited knowledge of cooking. Cooking recipe recommendation systems are one technology-based approach to provide information to such populations that can assist them with making more informed choices on nutrition. With the increasing capabilities of Artificial Intelligence (AI) technologies, there are now also AI-based cooking recipe recommendation systems. However, recipe recommendation systems do not provide any guidance on how to cook for a person who has limited experience with cooking. In particular, persons with cognitive disabilities or other mental health disorders can get sidetracked during cooking and forget to return to the cooking activity. 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 in what stage of cooking the user is currently performing.**

Keywords—A.I, 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]. Cooking recipe recommendation systems [3, 4] are one technology-based approach to provide information to such populations that can assist them with making more informed choices on nutrition. 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. In particular, persons with cognitive disabilities or other mental health disorders can get sidetracked during cooking and forget to return to the cooking activity. This can even become dangerous due to fires.

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 is mounted over a stovetop and constantly monitors 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. (Part of the approach is to break down simple recipes into a sequence of steps, each of which has a characteristic step on the stovetop.)

In this paper, we describe in detail 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).

The rest of this paper is organized as follows. Section II reviews the relatively little work on smart cooking assistant systems. Section III describes the system design, including the hardware design (Section III.A) and the image processing approach (Section III.B). We describe the results of system evaluation in Section IV and conclude with a list of possible improvements in Section V.

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, and removing impurities [6].

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 [7, 8]. The system focuses on detecting a set of pre-defined hazardous situations through various sensors, including motion and door sensors, ultrasonic range finders, door openings, flame and high-temperature sensors, and smart switches. The safeCOOK system is designed for enhancing safety when preparing meals, especially for people with cognitive impairments [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 system also integrates multiple object tracking for identifying and managing hazardous situations that may arise during cooking food. 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 [9].

Jelodar *et al*. [10] 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 [11]. In our work, we use MobileNets [12] for image detection because of the need to implement the model in an embedded system.

III. System Design

This section describes the hardware and software design implemented for this project.

III A. Hardware Design

Sensors

The MLX90614 is an infrared thermometer for non-contact temperature readings. It was important to test the accuracy as the trade-off was ease of use and accuracy. The range of the MLX90614 is -70ºC to 382.2ºC. This range is sufficient for the cooking-related application.

The DHT11 is also a non-contact Temperature and Humidity sensor, and it will help check changes in humidity and ambient temperature when cooking. Although the temperature range is only 0 ºC to 50ºC, the humidity measurements from this sensor could augment other measurements and could be useful in disambiguating similar cooking steps (e.g., water in a pot that is boiling or not boiling).

Cameras

The OV5647 Mini Camera Module for Raspberry Pi serves as the main camera for recognizing the various stages of cooking. The OV547 has the capability of capturing photos at 2592x1944 and recording videos in 60FPS at 720P. The quality will suffice for implementing a machine-learning model.

The MLX90640 infrared camera detects temperature changes from -40º to 300º with an accuracy of +- 2º. The MLX90640 will help identify temperature changes while cooking.

Processing Unit

Initial versions of the system used an Arduino UNO R3. However, its computational power proved to be insufficient for image recognition. We therefore use the Raspberry Pi 4 as the main computational platform. Not only were we able to execute multiple scripts controlling the various sensors, it was also capable of running the machine learning image detection models.

Additional Components

For debugging purposes, a 5-inch LCD display was used for portability and testing if the cameras and sensors captured the data. A PCB was used for a cleaner look and use of the GPIO pins that were needed for all the components. Since the LCD display used most of the pins, the display had external pins that could be soldered to a PCB. The overall layout of these components is shown in Fig. 1.

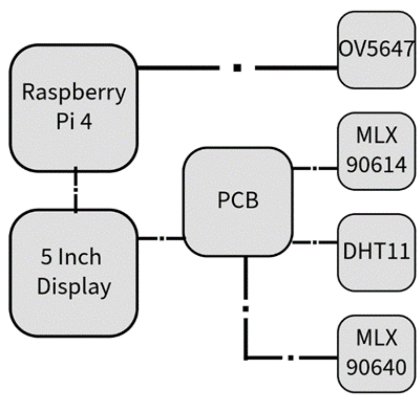

Fig. 1 Hardware Layout

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 (Fig. 2). For testing purposes, a camera tripod was used for easy adjusting and sturdiness. A photograph of the prototype setup is shown in Fig. 3.

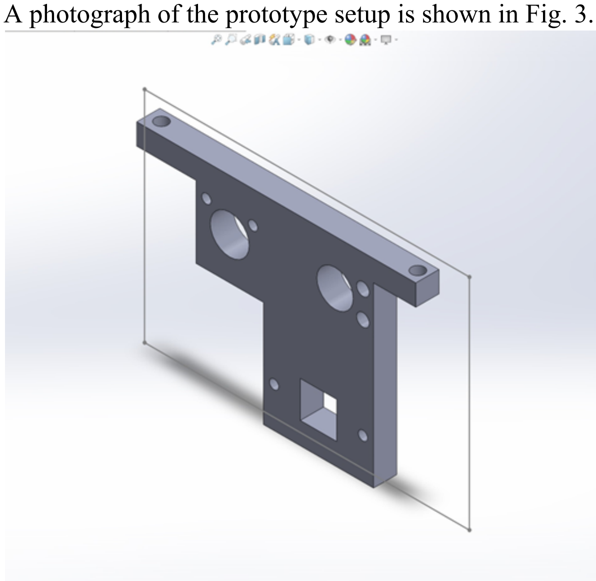

Fig. 2 CAD Design of Bracket using Solidworks

Fig. 3 Prototype hardware

III B. Image Processing

The cooking stage identification problem is an image classification problem and to tackle this problem we are using a pre-trained model called "SSD MobileNet V2 FPNLite 320x320" which is a variant of Tensorflow-2 Model Zoo object detection API [13]. This pre-trained model is a Single Shot Detector with Mobile net architecture and uses the feature pyramid network as feature extractor that is pre-trained on the COCO dataset. The MobileNet is a base network that provides high-level features for classification or detection while the detection layer employs the SSD method which is a feed-forward convolutional network to generate a predetermined set of bounding boxes and associated scores, indicating the presence of object class instances within those boxes which is faster than R-CNN and other models. It is also lightweight and can efficiently perform computation on mobile and embedded devices. It uses depthwise separable convolutions and residual connections to reduce the number of parameters and improve the performance of the network (Fig. 4).

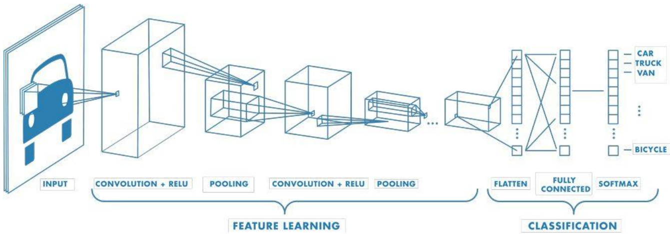

Fig. 4 SSD MobileNet Layout

For this project we are testing the model for a simple recipe: cooking pasta. We divided the cooking steps for this recipe into 6 stages that can be identified by the objects on the stovetop: "Empty burner", "Empty pot", "Pot with water", "Pot with boiling water", "Pot with pasta", and "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. 5.

Fig. 5 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.

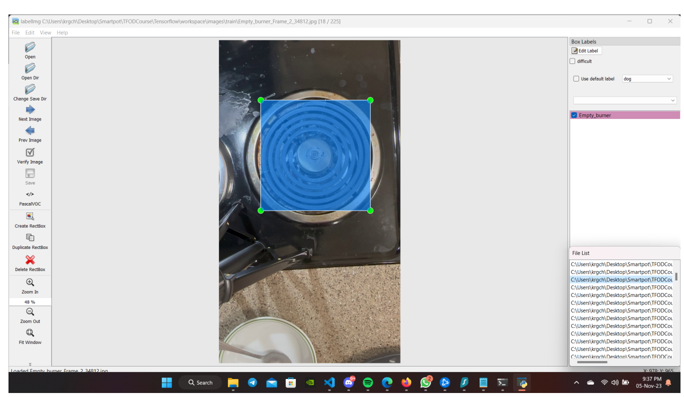

Fig. 6 Sample image depicting labeling using "LabelImg" package

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. 7).

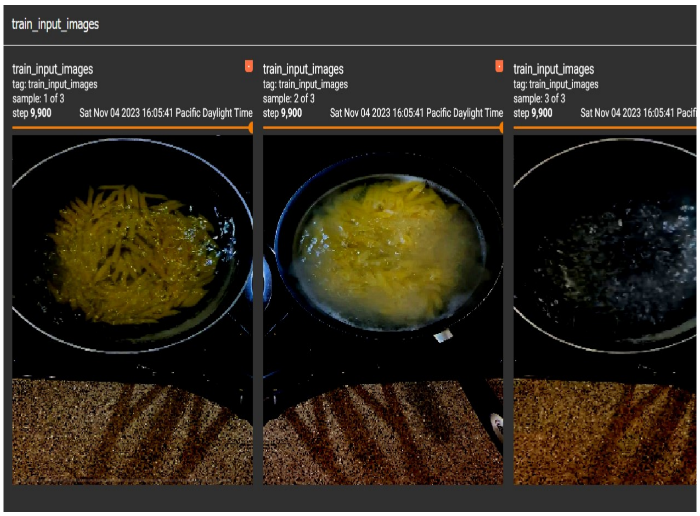

Fig. 7 Sample image depicting train input images after addition of noise

IV. Results and Discussions

MLX90614/DHT11

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

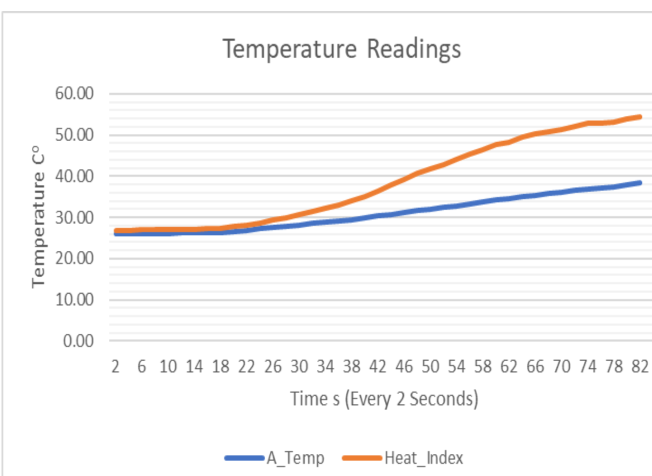

Fig. 8 Temperature readings recorded every 2 seconds

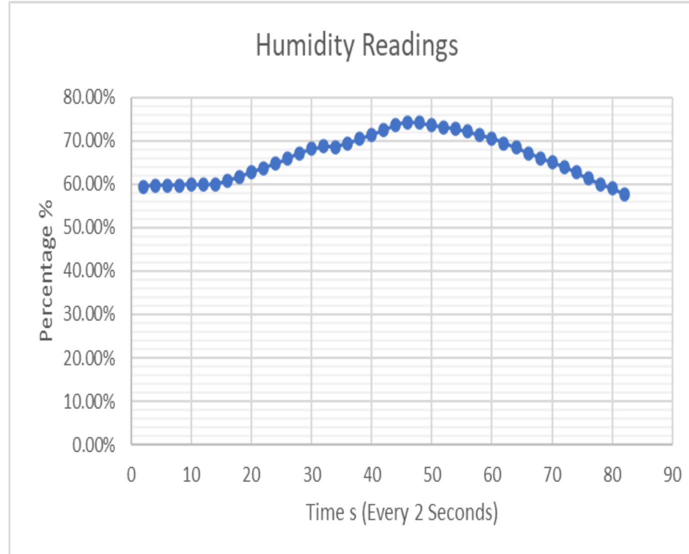

Fig. 9 Humidity readings recorded every 2 seconds

Figs. 8-9 shows the ambient temperature around the DHT11 and Heat_Index shows the combined index of the ambient temperature and the relative humidity. Compared to Fig 8 and Fig 9, we can use as the temperature rises, the humidity also rises until the 50-second mark. This is when the stove turned off and a big rush of steam was released, increasing the temperature while decreasing the humidity.

MLX90640/OV5647

The experiment begins by placing the cameras near a pot filled with water. Then, it captures pictures at various points in the cooking stages (Figs. 10-11).

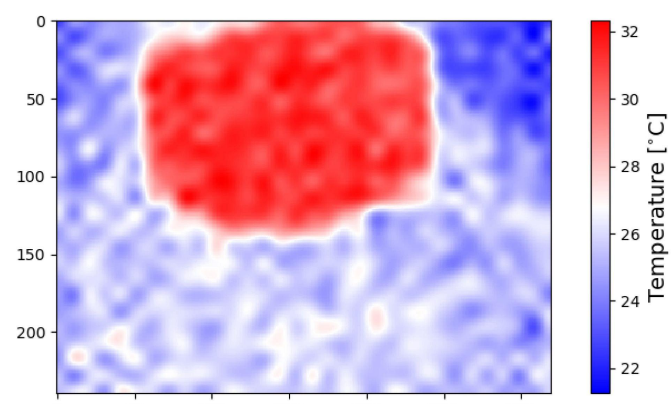

Fig. 10 MLX90640 Thermal Camera Starting Point

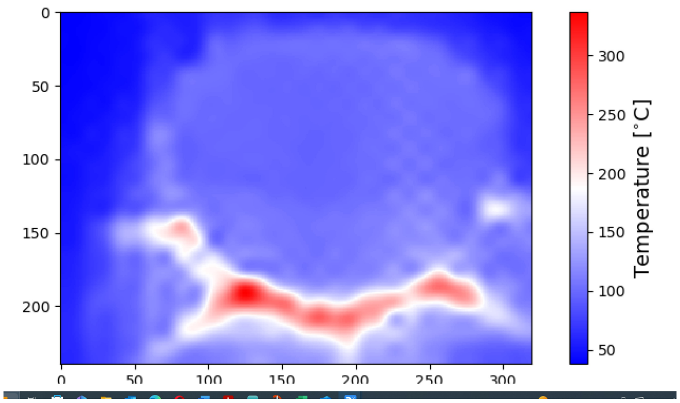

Fig. 11 MLX90640 Thermal Camera Boiling Point

As shown in Fig. 10, The "highest" temperature the camera could detect was the water in the pan at 32 ºC. As the water starts to boil, we can see in Fig. 11 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. 12-15.

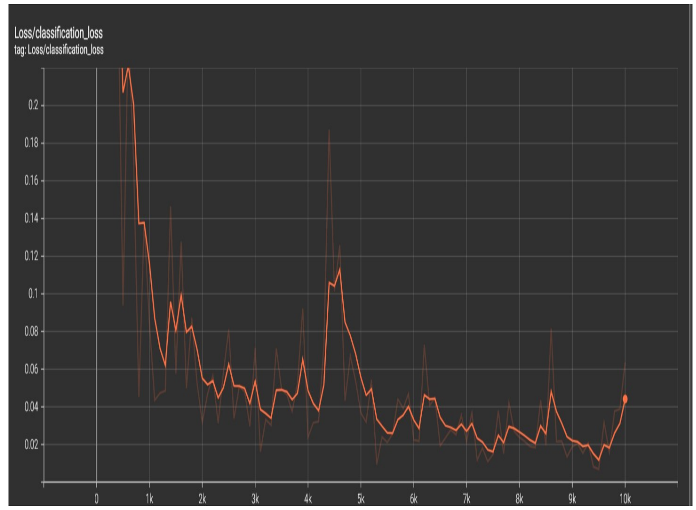

Fig. 12 Decrease in classification loss

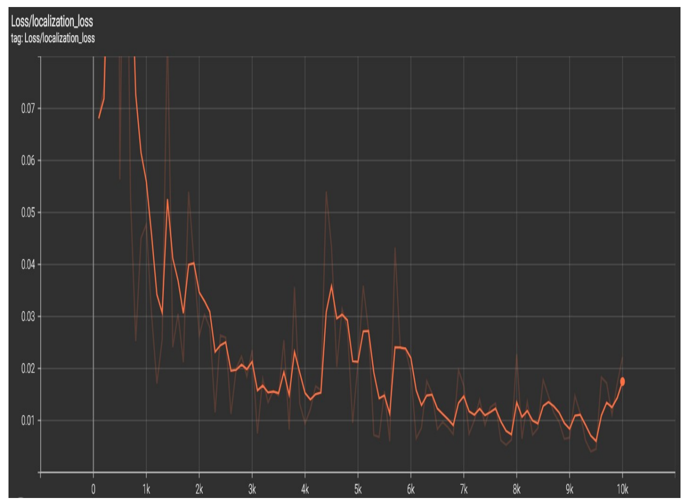

Fig. 13 Decrease in localization loss

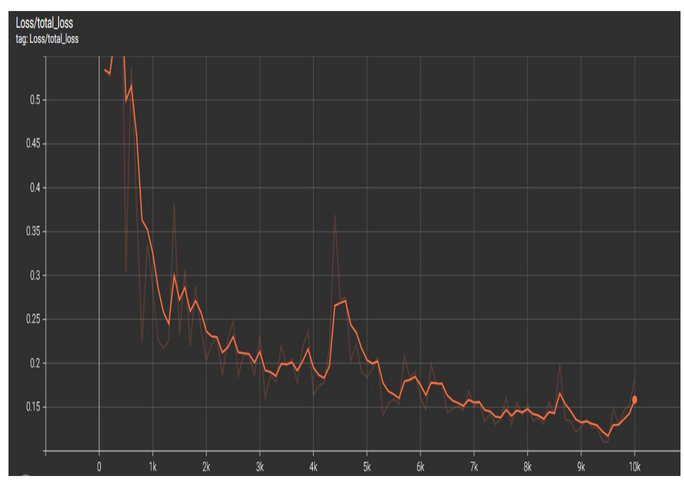

Fig. 14 Decrease in total loss

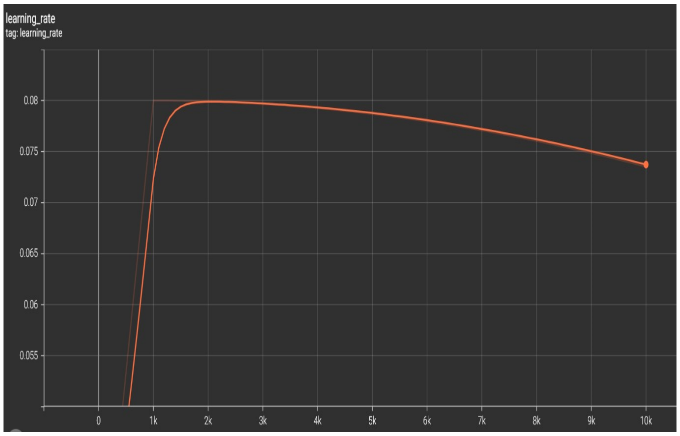

Fig. 15 Varying the Learning rate

Fig. 16 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. 16 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. People may cook with oil or with other substances that could splash onto the device. Thus, the system must be protected from such substances. One solution is to switch from the bracket design and print an enclosure using ABS printing material.
2. Another 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.
3. The Object detection model works accurately and can detect the stages of cooking even in real time and when only supplied with one image. The performance of the model is sufficient for this application as it has a high recall and precision values but the object detection fails when the type of cooking pot is changed (a steel pot from a dark-colored pot for cooking pasta). In this case, the system fails to detect the pot with water and wrongly detects it as an empty pot. This can be countered by re-training the model on different varieties of pots.
4. The temperature and humidity data collected will be integrated in the object detection model.
5. 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.

Acknowledgements

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

[8] 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)

[9] 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.

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

[11] You Only Look Once: Unified, Real-Time Object Detection Joseph Redmon, Santosh Ross Girshick, Ali Farhadi University of Washington, Allen Institute For Ai, Facebook Ai Research.

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

[13] Sharrab, Yousef O., et al. "Performance comparison of several deep learning-based object detection algorithms utilizing thermal images." 2021 second international conference on intelligent data science technologies and applications (IDSTA). IEEE, 2021.