Visually Impaired Grocery Shopping

Undergraduate Research Opportunity (UROP)

Laura Arias Fernandez
arias048@umn.edu
University of Minnesota

April 13, 2022

1 Summary and Results

The overall goal of the project was to help the visually impaired individuals who struggle when grocery shopping. To achieve this, a software to enable visually impaired individuals gain independence when grocery shopping was built. Specifically, the software consists of recognizing the objects the camera points to (object recognition software), alongside scanning the barcode of the products (product information software) to obtain more insights of the product, all through a verbal interface.

In order to build the product information software via barcode scanner. A deep learning based framework, yolov5[4], was used to design a barcode object detector, and a convolution neural network was trained on a dataset of approximately 1500 annotated barcode images from a publicly available dataset in Kaggle[6]. The barcode detector was used to provide for a more robust barcode scanner software. In order for the camera to accurately scan a barcode, the barcode needs to be very close to the camera, which can be very difficult for visually impaired individuals to do. This object detector allows the camera to locate a barcode in a muddled area. It minimizes the need for the visually impaired user to be very close to a product in order to recognize the barcode. The system is designed to detect barcodes and assists the user by directing them through verbal instructions on where and how to move their camera to successfully scan the barcode.

Figure 1: Barcode Object Detector

The proposed model was able to detect barcodes (class = 1) accurately with 100% recall (percentage of barcodes correctly identified) and a 98% precision (percentage of correct predictions), as demonstrated in Figure 2.0

```
Model Summary: 213 layers, 7015519 parameters, 0 gradients, 15.8 GFLOPs
               Class     Images     Labels          P          R     mAP@.5 mAP@.5:.95: 100% 5/5 [00:02<00:00,  2.10it/s]
                 all        142        163      0.862      0.818      0.833      0.596
                   0        142         22      0.737      0.636      0.674      0.535
                   1        142        141      0.988          1      0.993      0.657
Results saved to custom_yolov5/exp
CPU times: user 25.8 s, sys: 3.42 s, total: 29.2 s
Wall time: 40min 51s
```

Figure 2: Barcode Object Detector training result

The software integrates a web-scraping technique using the Python Library BeautifulSoup[7] to collect product information data. This data is gathered from an open source database called Open Food Facts[3] which has information in regards to ingredients, nutritional information, name, possible allergies and quantity information of over 2 million products across the globe. Thereby, once the software was able to obtain the EAN number, web scraping was used on the Open Food Facts to gather information about the product. To provide assistance to the blind, this information is read by the software out loud to them.

Even though this write up describes my project and its results, I believe it is very important to see the software created in action. Therefore, a small demonstration of the software can be found here: https://youtu.be/z4ixJT4SuLI. Moreover,the source code has been uploaded to my github[1].

The second part of this software is object recognition. To recognize many different types of objects, to provide "visual assistance" to the blind, i.e. identify objects in their surroundings, and then tell them what they are. The neural network was built using yolov4[2], and trained to classify the following object classes:

Fruit	Pastry	Pasta	Cheese
Coffee	Bread	Salad	Cookie
Tea	Flower	Popcorn	Cooking spray
Milk	Fish	Toilet paper	Sushi
Ice cream	Juice	Sandwich	Candy
Vegetable			

For each of the classes, I gathered a dataset from Open Images[5], a dataset of 9M images annotated with image-level labels from Google. I then converted the annotations to the desired format, and used it to start training. Training a neural network of such dimensions is very challenging, and in order to do so I made use of Google codelab, a platform that allows to train the neural networks in a GPU, rather than the CPU of the computer. By leveraging Codelab resources, I was able to accelerate training by saving checkpoints every hour. This task can still take a lot of time, and by the time writing this report training has still not finished, but as figure 3.0 shows we can see slight performance of the neural network.

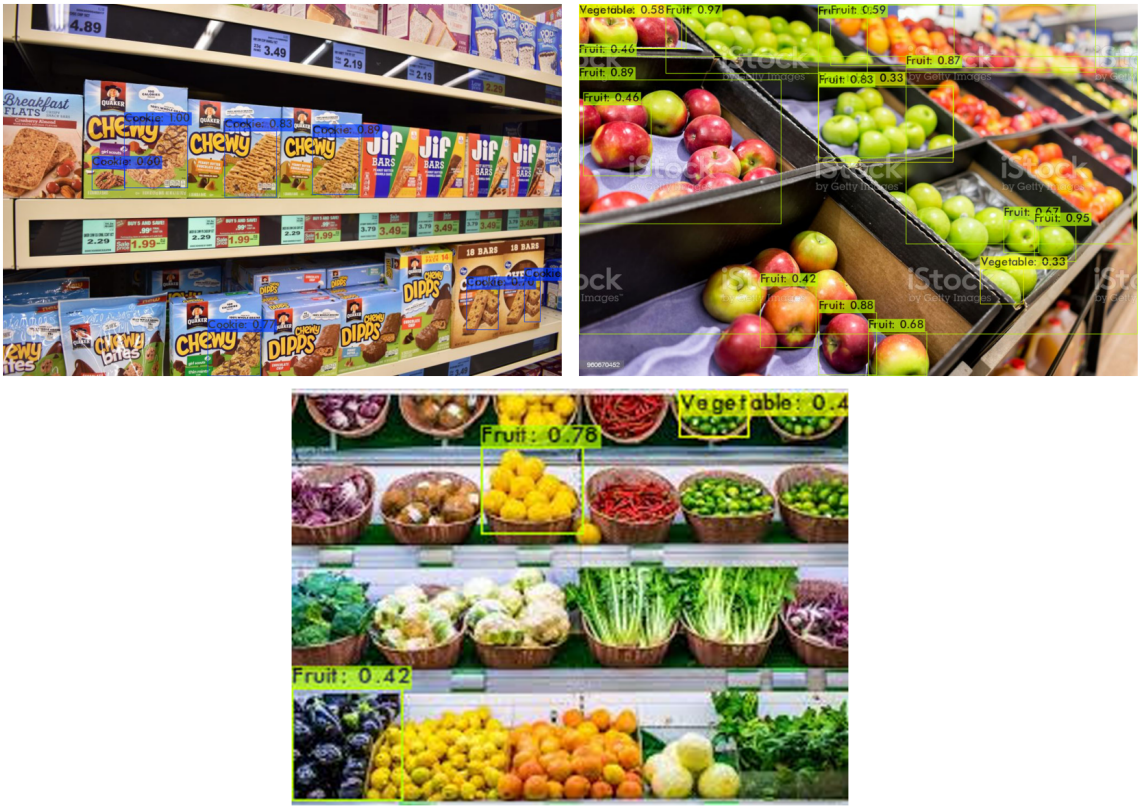

Figure 3: Object Recognition of Grocery Items

2 Objectives Reflection

The aim of the project is to develop a robust program that aids visually impaired individuals shop at a grocery store. More specifically, the goal is to develop two different tools: (1) an object recognition interface that is able to distinguish between different grocery items, and indicates to the user which product is in front of them., and (2) a barcode scanner, such that when a product is scanned the user can hear the different ingredients of the product through a speech interface.

Although the design of the product information software was very robust, a lot of room for improvement remains. The experimental results proved that the computation could be better made to tell the user where to move the camera to in order to scan the barcode. Furthermore, although through aid, it is still necessary to bring the barcode very close to the camera to be scanned effectively. The barcode object detector could be improved so that it would not only detect but also scan the barcode when the object is further away.

Moreover, the results for the object recognition were not as intended. I used a convolutional neural network to automatically learn a representation of each object class. The dataset was very large, with thousands of images per class. Training a neural network of such dimensions is very challenging, and such tasks can still take a lot of time. After hours of training we can see that the neural network has started recognizing some objects (Figure 3.0). Yet, as seen from figure 4.0 there are still some errors in the object recognition, and more training time needs to occur in order to accurately evaluate the performance of the object recognition.

Figure 4: Inaccuracy in object recognition of grocery items

Furthermore, another improvement that can be made is to combine web scraping with a local database. Right now my project only uses web scraping, a technique that requires internet connection to extract information from a public website. In order to make the program more robust a dataset could be implemented to save products the user has already scanned, since it is most probable that he/she will scan it again. If all necessary information is stored locally, then there is no need for a reliable internet connection and overhead time cost could be reduced.

Finally, it is to note that my program does not acquire all information, for instance it does not gather the products' prices. This is due to the fact that the program is intended to work for all different supermarkets, and even though the same product has the same identifiers, the price varies from grocery to grocery.

3 UROP Evaluation

Research is crucial for innovation, yet little is known about the entry into this field as an undergraduate student. This undergraduate Research Opportunity Program (UROP) allowed me to gain some research background and get an idea of what a job in research is like. Participating in research as an undergraduate student is very rewarding as you get to learn how to think independently about new projects. UROP helped me understand what it is like to start a project from scratch, the research background needed, the need to be comfortable and proactive in uncertainty, as well as allowed me to become more competent and confident in my ability to express technical ideas in public and in writing.

References

[1] Laura Arias Fernandez. https://github.com/lauraAriasFdez/barcodeDetector.

[2] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection, 2020.

[3] Open food facts. https://world.openfoodfacts.org/.

[4] Glenn Jocher. ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements. https://github.com/ultralytics/yolov5, October 2020.

[5] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, Tom Duerig, and Vittorio Ferrari. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *IJCV*, 2020.

[6] Sana Riaz. Barcode detection annotated dataset, Dec 2020. https://www.kaggle.com/datasets/whoosis/barcode-detection-annotated-dataset.

[7] Leonard Richardson. Beautiful soup documentation. *April*, 2007.