Model Engineering: Grocery Item Detection Used to Determine Product Sales and Inventory Trends

DS 4002 | Group 10 | Oct. 22, 2023
Carson Crenshaw (Lead), Sujith Panchumarthy, Zoe Averill

Data Acquisition

Collecting and finalizing the datasets

Analysis

Run image recognition model, use model on multiple item images, determine success of model

Results

Addressing the hypothesis and new discoveries

Motivation for Study

Modern Technological World

-

Increasingly digital world demands even greater importance of data analytics in every industry

New Competition

-

Grocery retail industry becoming more complex and competitive with introduction of grocery delivery services

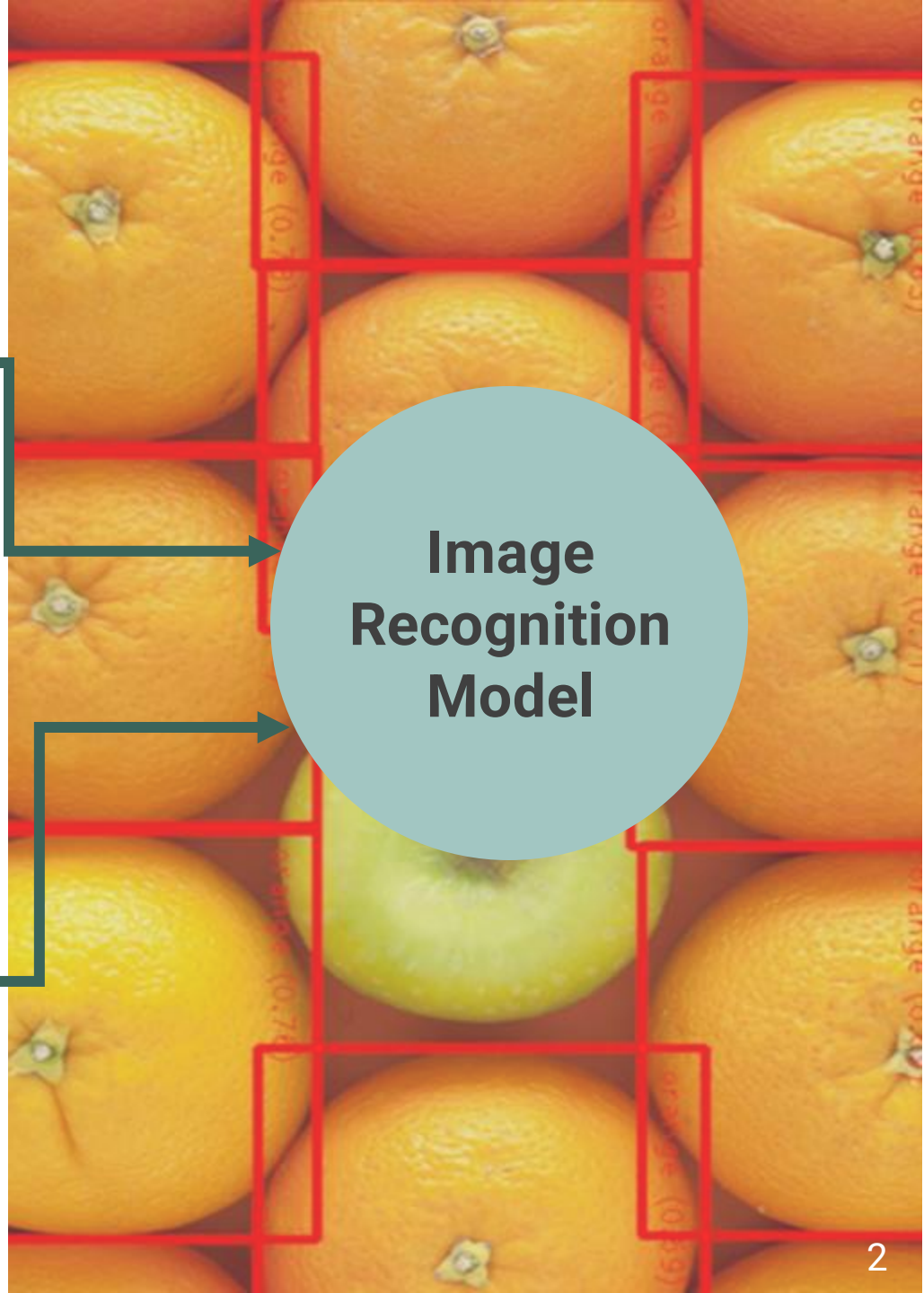
Importance → ***Image recognition = Efficient tool for brick-and-mortar retailers***

Hypothesis

An accurate model will be able to correctly identify all products within each image of a sample grocery basket.

Research Questions

To what extent does the image recognition model **accurately** determine the grocery products within a given image? Will the **similarity of appearance** between some grocery items obscure their identification? How can data generated from check-out cameras be used to learn more about **customer preferences** and even track **revenue generation?**



Data Acquisition/Explanation

Grocery Item Photos

Column Name	Description	Data Type
apple	Single item images of apples	Folder
orange	Single item images of oranges	Folder
banana	Single item images of bananas	Folder
broccoli	Single item images of broccoli	Folder
carrot	Single item images of carrots	Folder
group	Multiple item images of grocery products; images featuring multiple of the single produce items (apple, orange, banana, broccoli, carrot)	Folder

apple

Column Name	Description	Data Type
apple_01	Single item image of an apple	JPEG
apple_02	Single item image of an apple	JPEG
apple_03	Single item image of an apple	JPEG
apple_04	Single item image of an apple	JPEG
apple_05	Single item image of an apple	JPEG

Images collected and taken by Carson Crenshaw.
Image file structure above adapted from Marcus Klasson [4].

Example of apple_01 image data

Data Acquisition/Explanation

Dataset constructed at the conclusion of the model testing.
Accuracy for each test run on each single item image (apple, banana, orange, carrot, and broccoli)

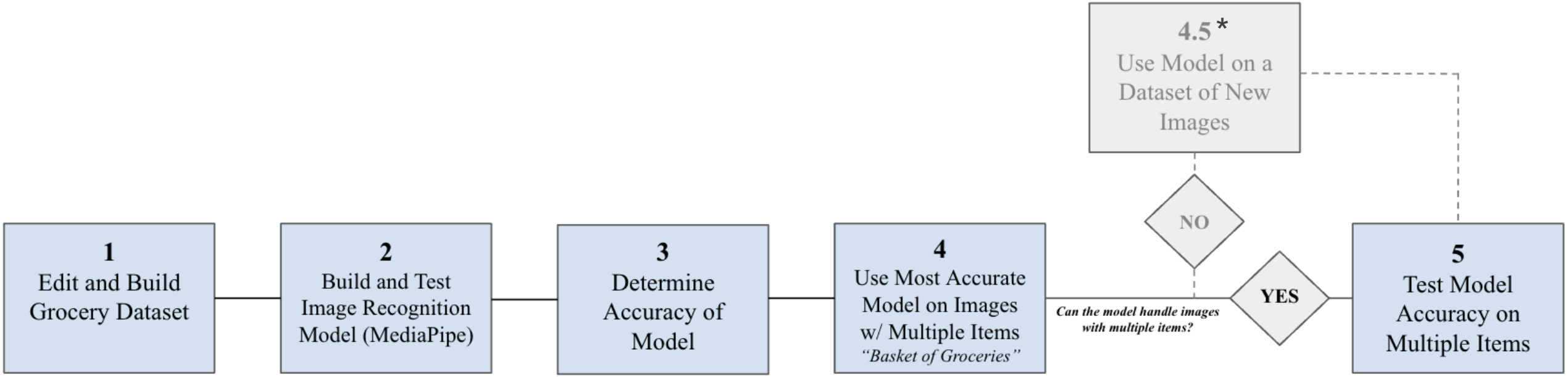
accuracy.csv

Column Name	Description	Data Type
File Name	Name of image file used	String
Object	Name of object detected in the file	String
Detected?	Has the object in question been detected?	Binary (Yes/No)
Confidence	Numerical quantification of object detection model accuracy	Float
Accepted?	Is the confidence value greater than or equal to 75%?	Binary (Yes/No)

	File Name	Object	Detected?	Confidence	Accepted?
0	apple_01.jpeg	apple	Yes	0.792969	Yes
1	apple_02.jpeg	apple	No	NaN	NaN
15	carrot_01.jpeg	carrot	Yes	0.691406	No

Example of final table

Analysis Plan



***Note:** The original analysis plan shown above was built to include all possible contingencies. Although the model was able to handle images with multiple items, the project prepared for a necessary alternative.

Tricky Analysis Decision: ***DATA COLLECTION***

Problem	***Solution***
M. Klasson dataset (GitHub) • Lack of detectable object images • Too many images, long runtime	Create ***own*** dataset
Multiple Image Detection • Background pattern disruption	***Recreate*** dataset with new images

Bias and Uncertainty Validation

Machine learning dataset is small

- Only **25** images used for single item images
- Deemed necessary due to time constraint + image upload time
- Mitigated slightly by variation in the sample (5 different classes)
- When detected, single item images demonstrated high confidences

Detection Inaccuracies (next slide)

- Carrots had lower confidence in detection
- Some instances of inaccurately detected and labeled apples

Accuracy of the 25 images

88% of items were detected

77% of images had a confidence of ≥ 75%

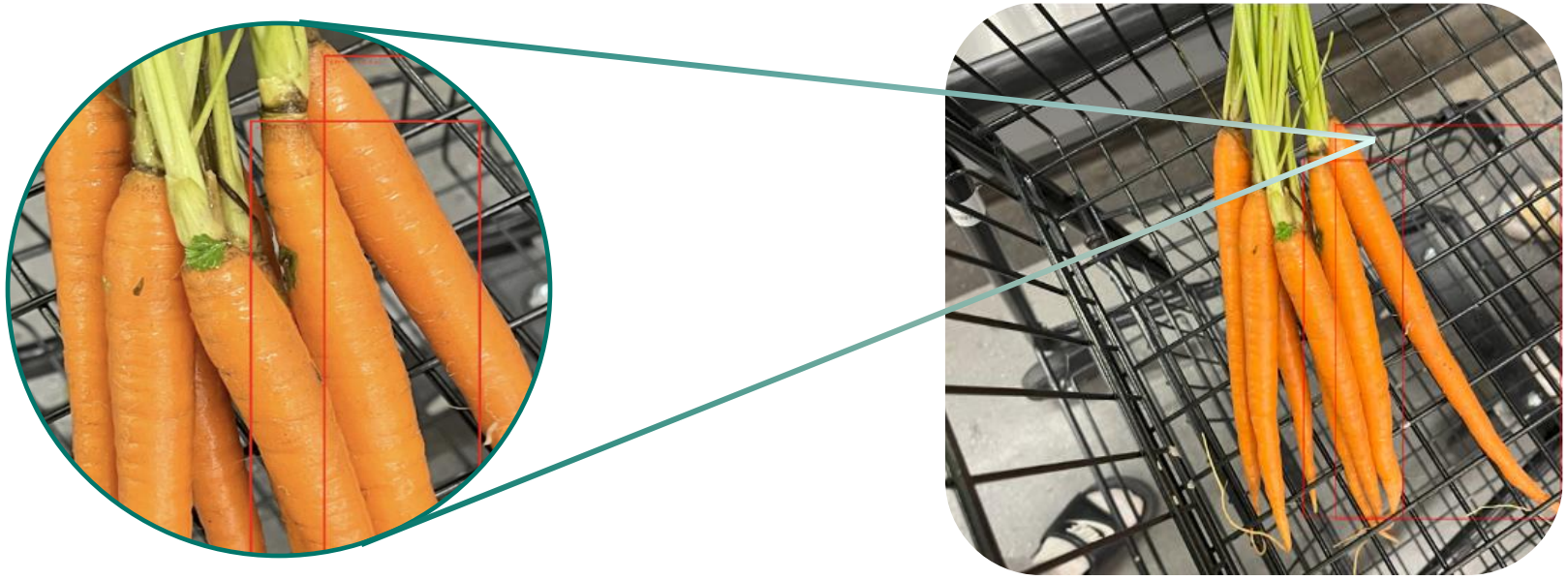
Av. Confidence of **78%**

Example of banana_01 correct detection

Model inaccurately detected apples as cake on multiple occasions

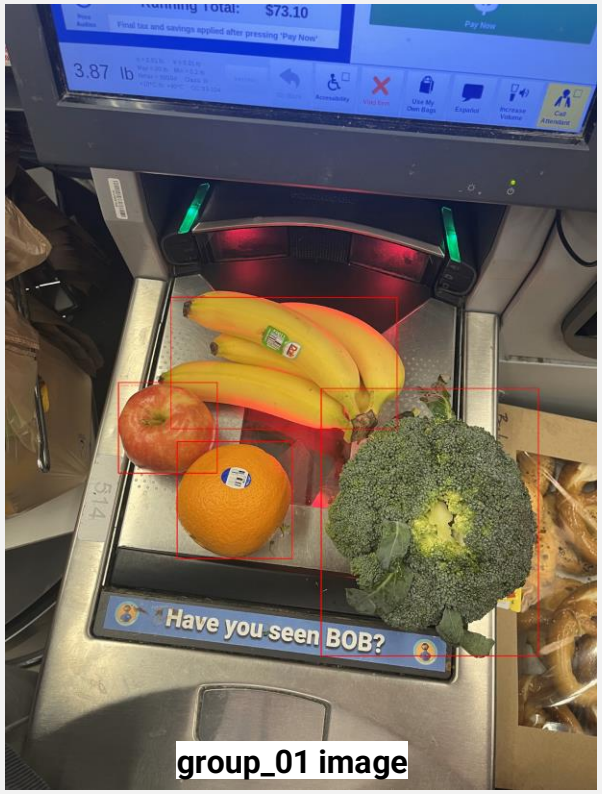
Note: *Although not within the scope of the project, the person was accurately detected*

Model detects carrots individually and inconsistently

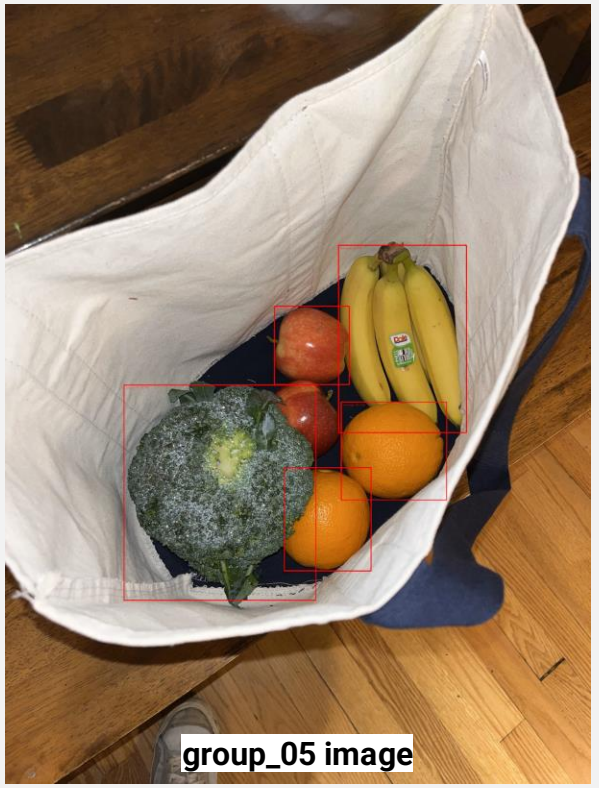


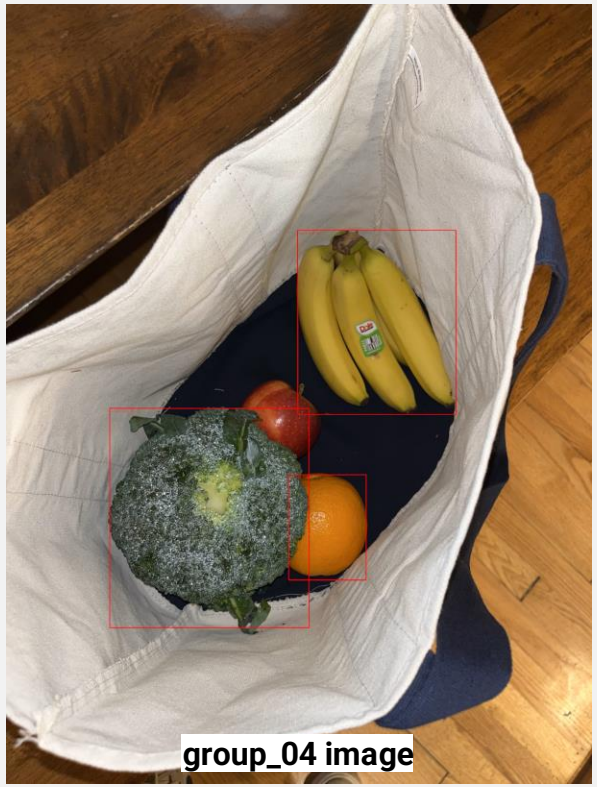
Results and Conclusions

group_03 image



group_01 image


group_05 image


group_04 image

Accuracy of Single Item

Competitive confidence overall (av. 78%)

Accuracy of Multiple Items

Accurately detects all items except for those w/single item image issues (carrots, apples)

Next Steps…

Build and test a model *specialized* in identifying grocery store items

- Increased accuracy and confidence of detections
- Detect other common grocery store items besides produce

Once **accuracy and confidence are high**, calculate basket *prices*, *trends* in inventory

Weight-based and **brand** items will still be a limitation; figure out how to address

References & Acknowledgements

[1] “Data-Driven Customer Personality Analysis of the Grocery Store,” www.linkedin.com. Available: https://www.linkedin.com/pulse/data-driven-customer-personality-analysis-grocery-store-zamohylna/ [accessed Oct. 09, 2023].

[2] M. Hickins, “10 Ways Grocers Can Use Data Analytics,” Oracle.com, Jul. 31, 2023. Available: https://www.oracle.com/retail/grocery-data-analytics/ [accessed Oct. 09, 2023].

[3] M. Klasson, C. Zhang, and H. Kjellström, “Using Variational Multi-view Learning for Classification of Grocery Items,” Patterns, vol. 1, no. 8, p. 100143, Nov. 2020, doi: https://doi.org/10.1016/j.patter.2020.100143. [Accessed Oct. 16, 2023].

[4] M. Klasson, “Grocery Store Dataset,” GitHub, Oct. 05, 2023. [Online]. Available: https://github.com/marcusklasson/GroceryStoreDataset/tree/master. [Accessed Oct. 16, 2023].

[5] "Object detection guide for Python," developers.google.com. Available: https://developers.google.com/mediapipe/solutions/vision/object_detector/python [accessed Oct. 16, 2023].

Link-to-MI1-Doc
Link-to-MI2-Doc


СКИДКИ
18+
ЭКО УПАКОВКА