NUTRITION LABEL EXTRACTOR

DAVID SHAN & WILBERT GUO

PROBLEM DEFINITION

Problem

- Dietary tracking is a hassle in today's society
- Constant updating of nutrition metrics
- Traditional methods includes keeping handwritten or electronic log
- No current method of comparing different food products based on nutrition facts

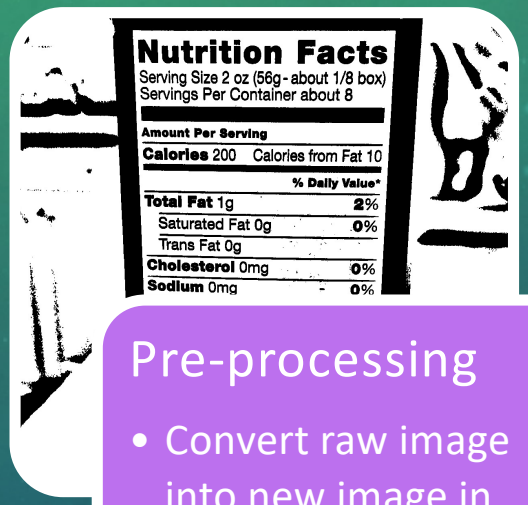
Current Solutions

- Modern applications have made it easier for user to input nutritional info
- MyFitnessPal uses crowd-sourced info to allow users to search for common products
- This approach requires large amounts of collaboration and maintenance or nutrition info will be out of date

Our Solution

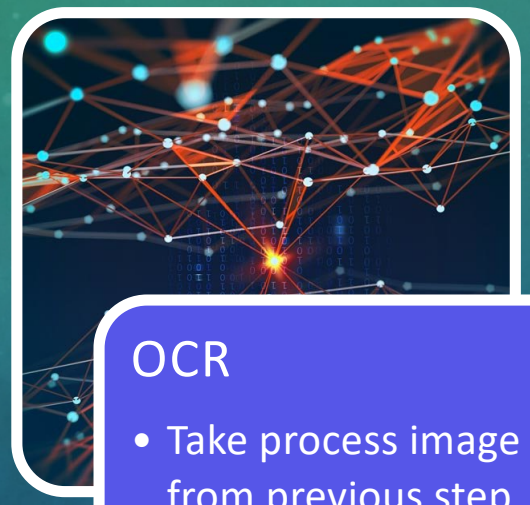
- Use computer vision and leverage widespread use of phone cameras to extract nutrition info
- Pull relevant nutrition data directly from the picture (calorie, fat, protein etc.)
- Allow users to visualize breakdown of the main nutrition categories and compare it to other similar food products

METHOD OVERVIEW



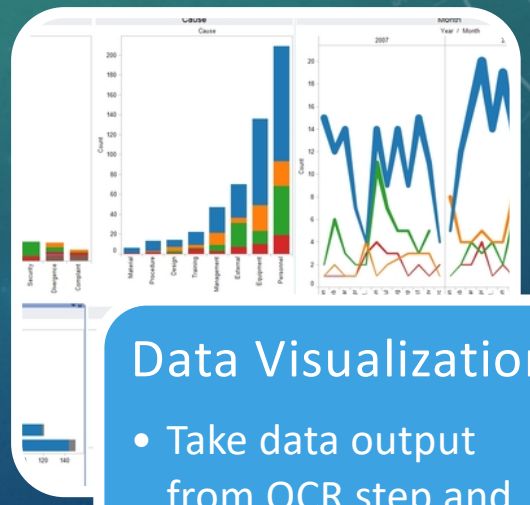
Pre-processing

- Convert raw image into new image in which the English text and numerical values easily identifiable



OCR

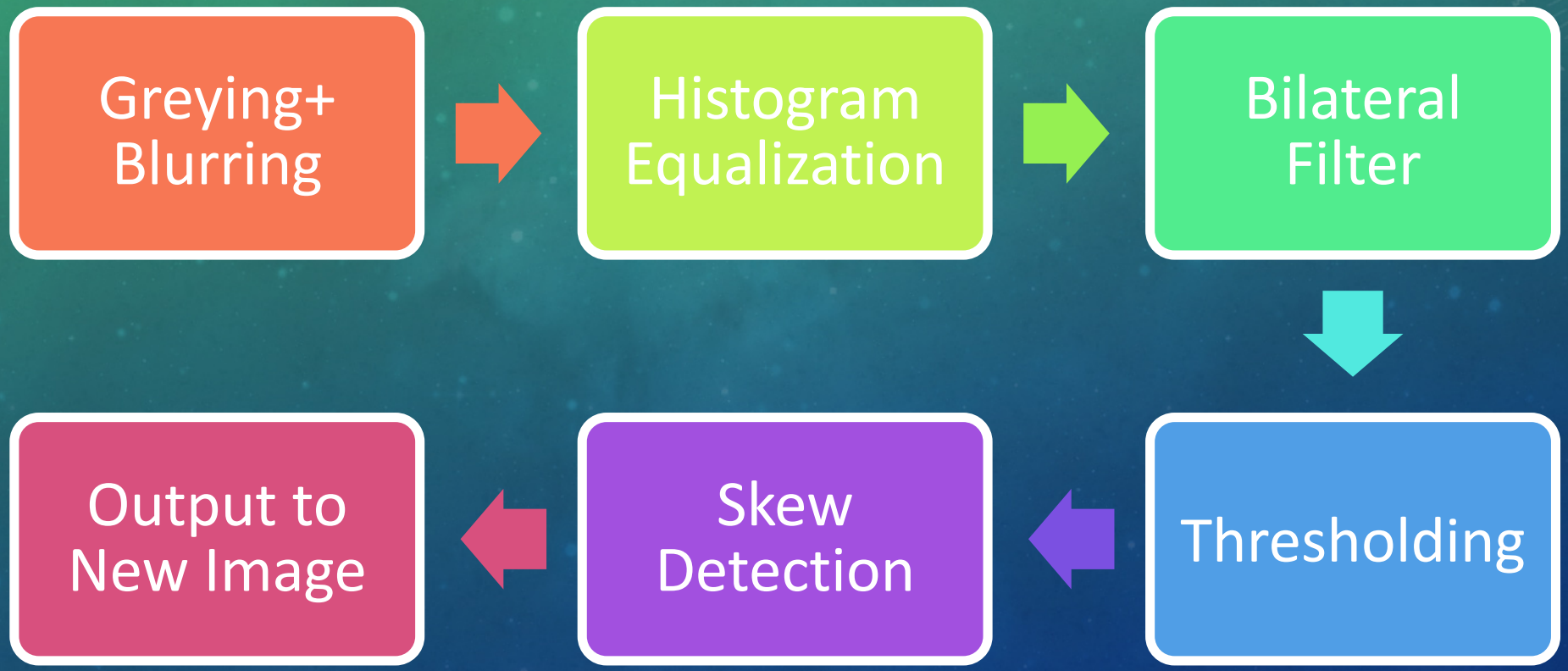
- Take process image from previous step and extract the words and numerical values from image



Data Visualization

- Take data output from OCR step and visualize and output nutrition data using python

PRE-PROCESSING (WILBERT)



OPTICAL CHARACTER RECOGNITION (DAVID)

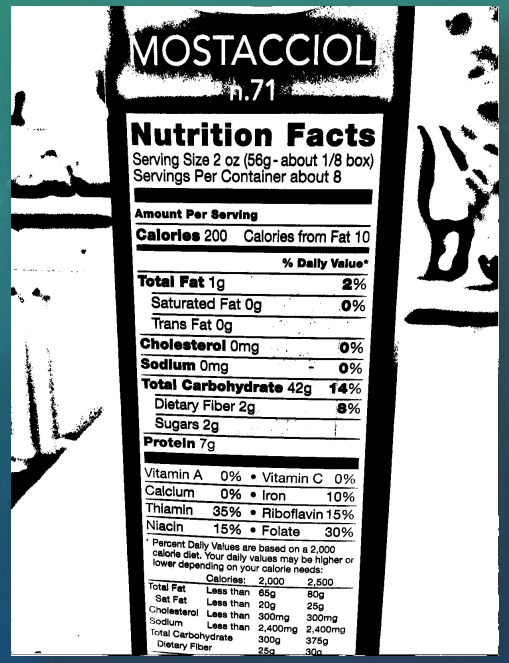
DATA VISUALIZATION (DAVID)

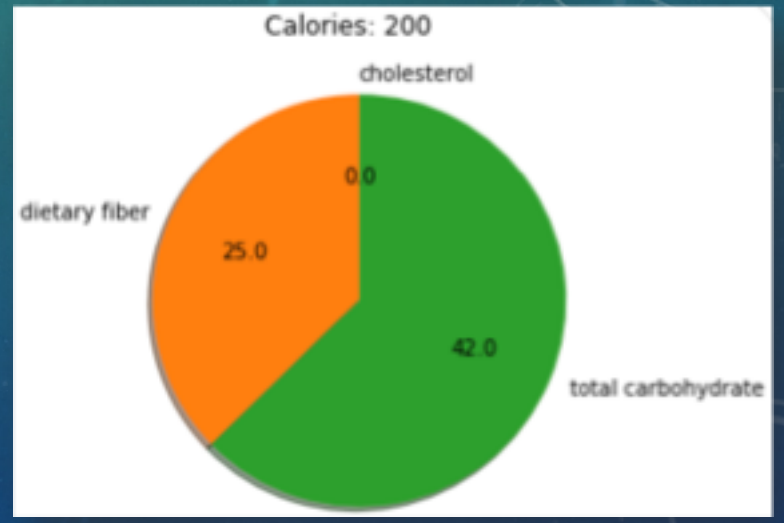
REFERENCES

- Pre-Processing
  - https://www.researchgate.net/publication/283727396_Image_enhancement_by_Histogram_equalization
  - https://www.cs.jhu.edu/~misha/ReadingSeminar/Papers/Tomasi98.pdf
  - https://pdfs.semanticscholar.org/3796/3f1672c4f53f57a1774c381aaf86695f2087.pdf
  - https://www.researchgate.net/publication/220494200_Adaptive_Thresholding_using_the_Integral_Image
  - https://www.pyimagesearch.com/2017/02/20/text-skew-correction-opencv-python/
- OCR
  - https://pypi.org/project/pytesseract/
  - https://arxiv.org/pdf/1507.05717.pdf

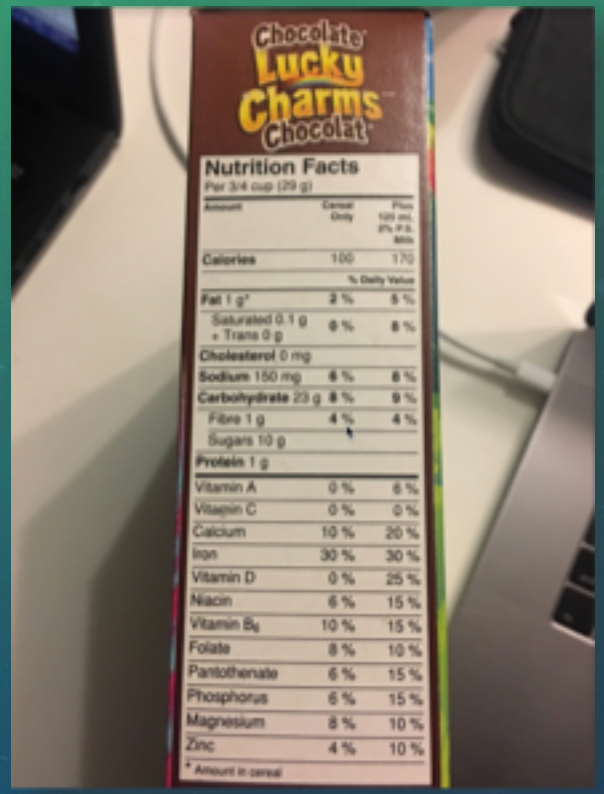
RESULTS

MOSTACCIOLI
n.71
Nutrition Facts
Serving Size 2 oz (56g - about 1/8 box)
Servings Per Container about 8
Amount Per Serving
Calories 200 Calories from Fat 10
% Daily Value*
Total Fat 1g 2%
Saturated Fat 0g 0%
Trans Fat 0g
Cholesterol 0mg 0%
Sodium 0mg 0%
Total Carbohydrate 42g 14%
Dietary Fiber 2g 8%
Sugars 2g
Protein 7g
Vitamin A 0% • Vitamin C 0%
Calcium 0% • Iron 10%
Thiamin 35% • Riboflavin 15%
Niacin 15% • Folate 30%
* Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs:
Calories: 2,000 2,500
Total Fat Less than 65g 80g
Sat Fat Less than 20g 25g
Cholesterol Less than 300mg 300mg
Sodium Less than 2,400mg 2,400mg
Total Carbohydrate 300g 375g
Dietary Fiber 25g 30g


MOSTACCIOLI
n.71
Nutrition Facts
Serving Size 2 oz (56g - about 1/8 box)
Servings Per Container about 8
Amount Per Serving
Calories 200 Calories from Fat 10
% Daily Value*
Total Fat 1g 2%
Saturated Fat 0g 0%
Trans Fat 0g
Cholesterol 0mg 0%
Sodium 0mg 0%
Total Carbohydrate 42g 14%
Dietary Fiber 2g 8%
Sugars 2g
Protein 7g
Vitamin A 0% • Vitamin C 0%
Calcium 0% • Iron 10%
Thiamin 35% • Riboflavin 15%
Niacin 15% • Folate 30%
* Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs:
Calories: 2,000 2,500
Total Fat Less than 65g 80g
Sat Fat Less than 20g 25g
Cholesterol Less than 300mg 300mg
Sodium Less than 2,400mg 2,400mg
Total Carbohydrate 300g 375g
Dietary Fiber 25g 30g


Calories: 200
cholesterol
0.0
dietary fiber
25.0
42.0
total carbohydrate

RESULTS


Chocolate
Lucky
Charms
Chocolat
Nutrition Facts
Per 3/4 cup (29 g)
Amount
Cereal Only
Plus 100 mL 2% P.S. Milk
Calories 100 170
% Daily Value
Fat 1 g* 2 % 5 %
Saturated 0.1 g + Trans 0 g 0 % 8 %
Cholesterol 0 mg
Sodium 150 mg 6 % 8 %
Carbohydrate 23 g 8 % 9 %
Fibre 1 g 4 % 4 %
Sugars 10 g
Protein 1 g
Vitamin A 0 % 6 %
Vitamin C 0 % 0 %
Calcium 10 % 20 %
Iron 30 % 30 %
Vitamin D 0 % 25 %
Niacin 6 % 15 %
Vitamin B6 10 % 15 %
Folate 8 % 10 %
Pantothenate 6 % 15 %
Phosphorus 6 % 15 %
Magnesium 8 % 10 %
Zinc 4 % 10 %
* Amount in cereal

Chocolate
Lucky
Charms
Chocolat
Nutrition Facts
Per 3/4 cup (29 g)
Amount
Cereal Only
Plus 100 mL 2% P.S. Milk
Calories 100 170
% Daily Value
Fat 1 g* 2 % 5 %
Saturated 0.1 g + Trans 0 g 0 % 8 %
Cholesterol 0 mg
Sodium 150 mg 6 % 8 %
Carbohydrate 23 g 8 % 9 %
Fibre 1 g 4 % 4 %
Sugars 10 g
Protein 1 g
Vitamin A 0 % 6 %
Vitamin C 0 % 0 %
Calcium 10 % 20 %
Iron 30 % 30 %
Vitamin D 0 % 25 %
Niacin 6 % 15 %
Vitamin B6 10 % 15 %
Folate 8 % 10 %
Pantothenate 6 % 15 %
Phosphorus 6 % 15 %
Magnesium 8 % 10 %
Zinc 4 % 10 %
* Amount in cereal

Calories: 100
carbohydrate
saturated
sugars
fibre
23.0
150.0
sodium

METHOD COMPARISON

CRNN

- Works well for single line OCR
- Doesn’t capture text for the entire nutrition label
- Relatively slow compared to Tesseract

Tesseract

- Works well in obtaining accurate keywords
- Handles multiple lines of text
- Extremely fast

GitHub Project

- **Github:** sgrpanchal31
- More accurate information extraction
- Pre-processing stage uses Single Shot Detector (SSD) object detection model

PROJECT TOPIC COVERAGE

- Image Processing
  - Converting to grayscale + Correlation/ Convolution
  - Histogram Equalization
  - Bilateral Filter
  - Thresholding
  - Skew Detection
- Neural Network (CRNN)

GROUP WORK DIVISION

David Shan

- OCR
- Python Data Visualization
- GitHub Repo Manager

Wilbert Guo

- Pre-processing Steps
- Slideshow Presentation
- Video Editing

THANK YOU FOR LISTENING