Food Calorie and Nutrition Analysis System based on Mask R-CNN

Chiang, el al., 2019, IEEE 5th ICCC

Presentation by: Rocky Y. Gonzalez

Outline

- Introduction / Motivation
- Contributions from Chiang, et al.
- Calorie and Nutrition Analysis System
- Mask R-CNN
- Food Weight Estimation
- Food Calorie and Nutrition Analysis
- Results
- Conclusion and Future Work

Introduction / Motivation

- Obesity associated with many leading causes of death including diabetes, heart disease, stroke, and cancer.
- Most effective way to prevent obesity is through food intake control (this requires understanding food ingestion of nutrients and calories in a meal).
- Mask R-CNN used to analyze the nutrition and calorie content of food from an image.
- This study estimates food weight by linear regression and uses a nutrient table to estimate food calories/nutrients.

Previous Food Detection Models

- *Dehais et al.* proposed dish detection and segmentation system for dietary assessment. Assumes circular dish and approximates dish edges with curves using RANSAC (RANdom Sample Consensus).
- *Pouladzadeh et al.* proposed a system of measuring calories and nutrients using food images captured by phone; average accuracy 86% for 15 foods, food dishes could not overlap.
- *Bolanos and Radeva* presented an approach using GoogleNet-GAP with a recognition rate of 90.90% on *EgocentricFood* dataset and 79.20% on *Food101* dataset.

Contribution of Chiang, et al.

- Development of a food calorie and nutrition system that can analyze the composition of a food based on a provided image using Mask R-CNN.
- Introduced a newly collected dataset of images "Ville Café" for food recognition:
  - 16 categories with 35,842 images.
  - Includes salad, fruit, toast, egg, sausage, chicken cutlet, bacon, French toast, omelet, hash browns, pancake, ham, patty, sandwich, French fries, and hamburger.

Calorie and Nutrition Analysis System

- System broken into 4 main steps:
  1. Food image is resized to 1024x1024 pixel size, I_r (or long edge 1024-pixel, short edge scaled proportionally).
  2. Resized image fed into Mask R-CNN to capture food features and perform food detection and classification.
  3. System estimates weight of the food through the recognized food image.
  4. Food calorie and nutrition estimated according to Ministry of Health and Welfare, and the US Department of Agriculture's Food Nutrition Database

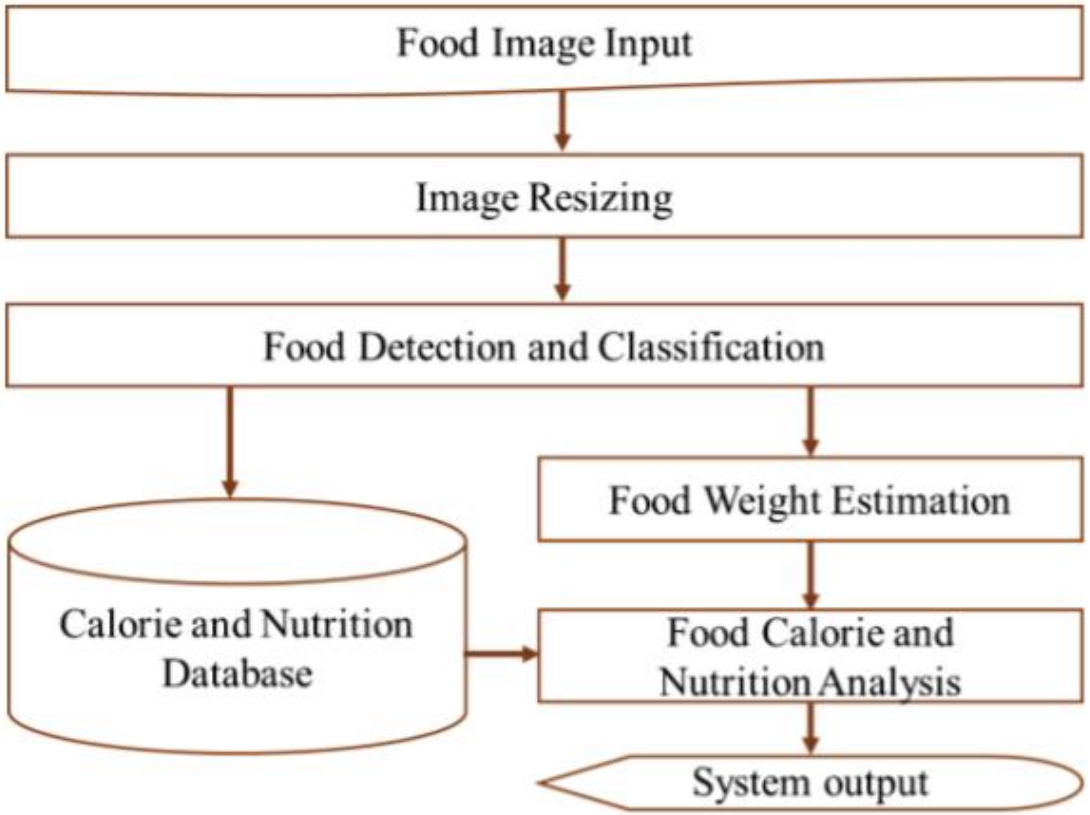

Figure 1 – Flowchart of Calorie and Nutriton Analysis System [Chiand, et al.]

Food Detection and Classification

- The resized food image, I_r, is inputted into Mask R-CNN to obtain prediction results for the food class, food bounding box, and food mask.
- Mask R-CNN is an instance segmentation approach that improves on faster R-CNN.
- Mask R-CNN architecture:
  - Convolutional backbone, RoIAlign layer, and head

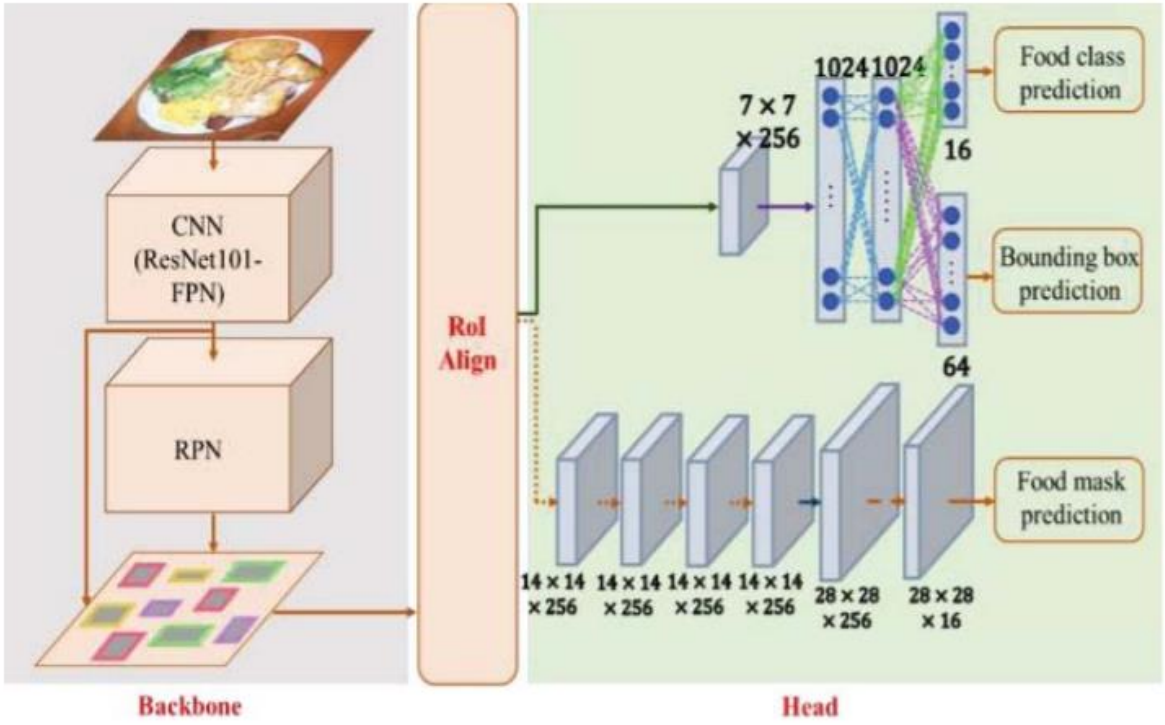

Figure 2 – Architecture of Mask R-CNN [Chiang, et al.]

Mask R-CNN Backbone

- Contains ResNet101-FPN feature extraction and RPN.
- Bottom-up pathway extracts food features from low to high levels in neural network (ResNet-101).
- Lateral connections output feature maps of the bottom-up pathway to the number of channels.
- Top-down pathway transfers high-order food features from neural network to lower order.
- RPN uses feature maps of the ResNet101-FPN output to determine foreground block in image.

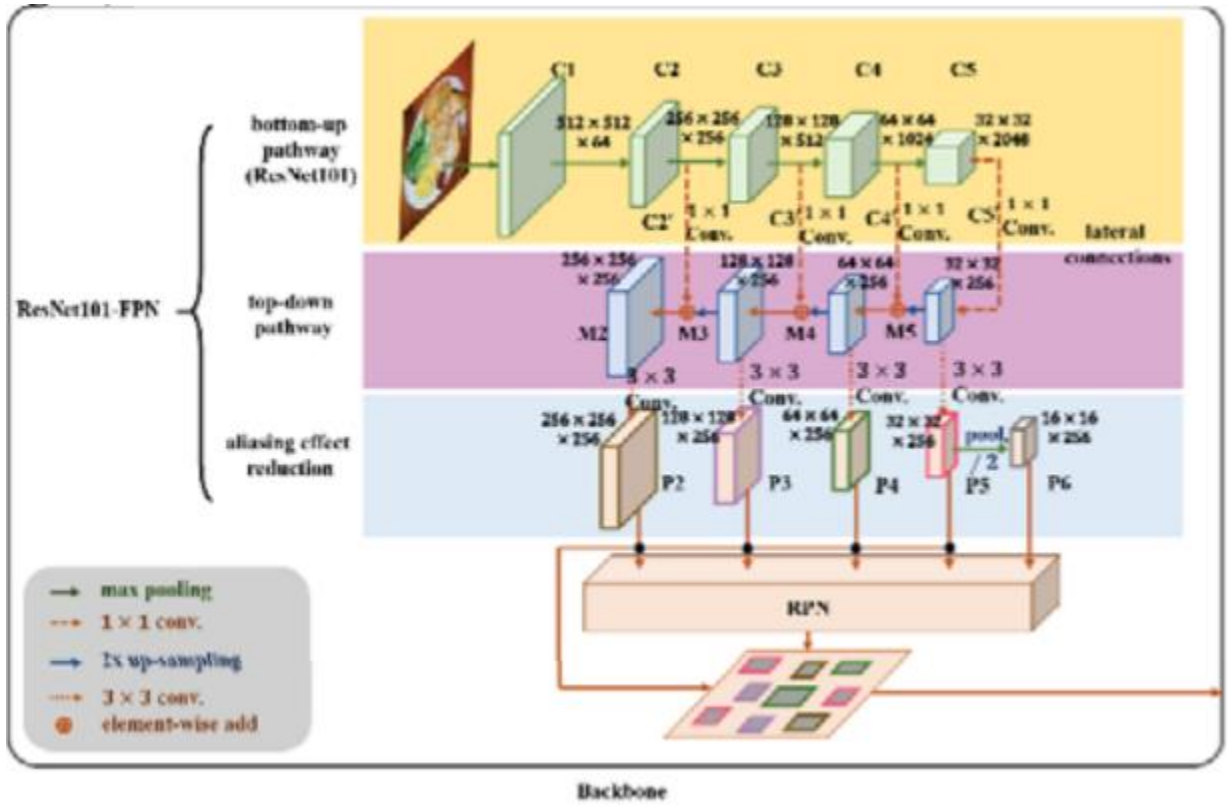

Figure 3 – Architecture of backbone Mask R-CNN [Chiang, et al.]

Mask R-CNN RoI Align

- Technique used for adjusting the size of RoIs provided inputs of different sizes.
- When the food image I_r is input into the Mask R-CNN backbone, the feature map of the food block RoI in the image can be obtained.
- The RoIs are obtained by dividing the image into 14x14 grids and taking four sample points for each grid.
- After obtaining four sampling points, maxpooling is performed on the sampling points in the cell and outputs are RoIs of size 14x14.

Mask R-CNN Head

- Neural Network Architecture used to predict food class, bounding box, and mood masks.
- First branch used to perform convolution operation using 1024 kernals for RoIs. The food class of RoI and probability of belonging to that class is obtained.
- Second branch constructed of Fully Convolution Network. Feature extraction and feature upsampling performed (upsampling captures pixels in image belonging to a food class).

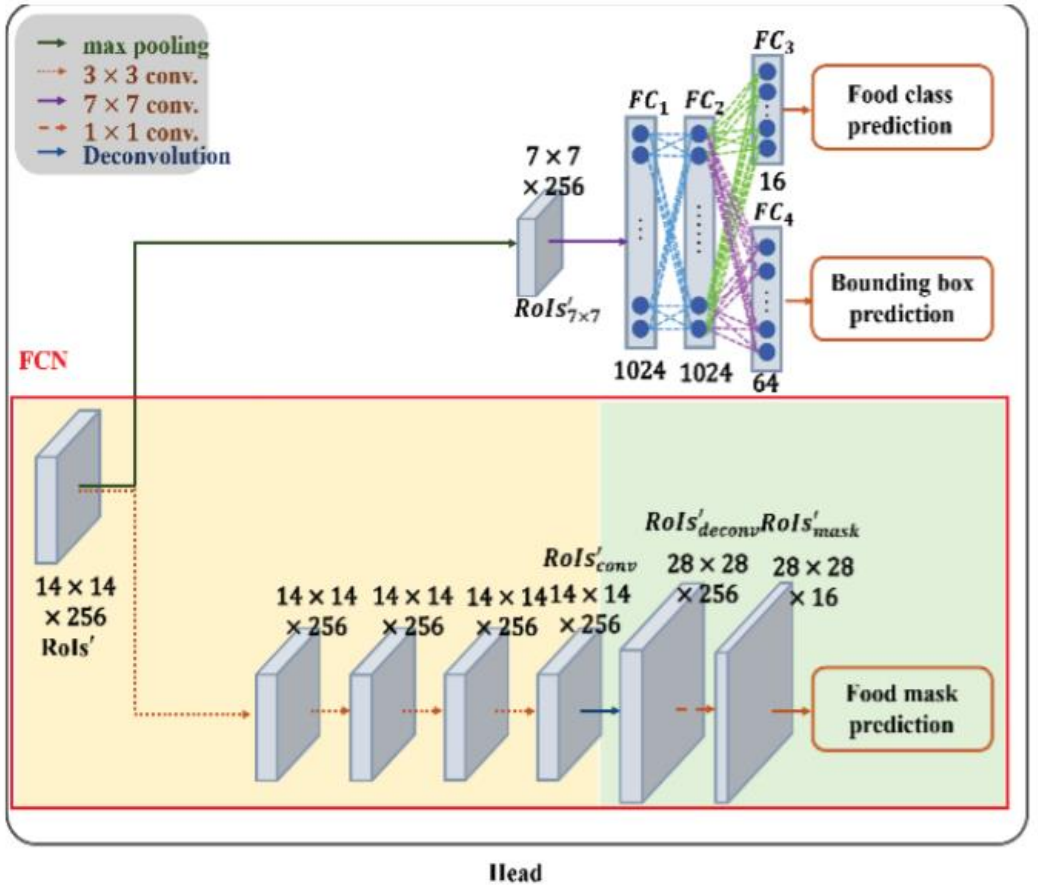

Figure 4 – Architecture of Mask R-CNN Head [Chiang, et al.]

Union Postprocessing

- After obtaining food class, bounding box, and mask prediction, results are then integrated.
- Output state divided into three steps:
  - Filtering out low confidence prediction boxes and food masks if class probability too low.
  - Non-maximum suppression used to retain bounding box of same class with highest class probability.
  - In the apply mask step, union operation performed for RoIs of the same class.
- Inputs: $I_r, class, class_{prob}, box, RoRs_{mask}$
- Outputs: $class_{refine}, box_{refine}, mask_{refine}$

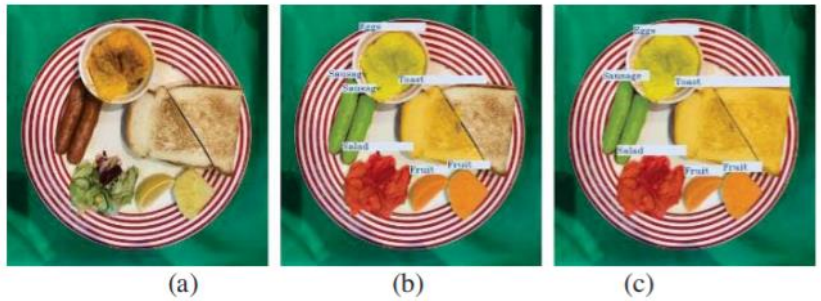

Figure 5 – (a) Input image, (b) Result of Mask R-CNN w/ NMS, (c) Result of Mark R-CNN w/ Union Postprocessing [Chiang, et al.]

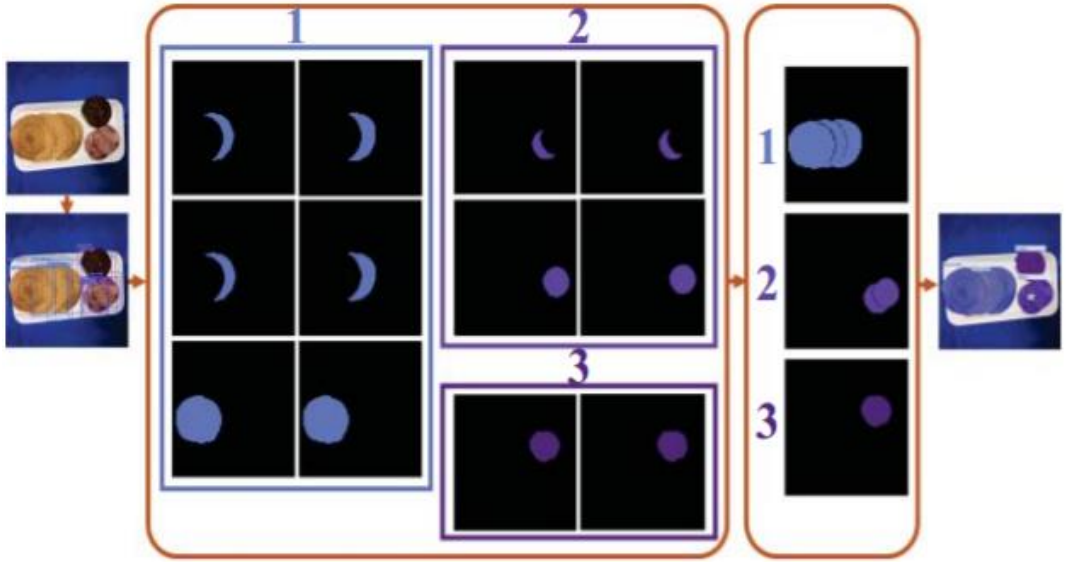

Figure 6 – Example of Union Postprocessing [Chiang, et al.]

Food Weight Estimation

- In this study, fixed angle photographs of the same food of different portions and weight/amount of food of image were recorded.
- Least Squares Method (LSM) used to find error between estimated food weight obtained from linear regression, $\overline{f(x_i)}$, and the actual food weight, y.
- For each class, $pixel_{food}$ can be obtained and substituted into $f(x_i)$ to estimate the calories and nutrients of food in an image.

$$\mathrm{LSM} = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \overline{f(x_i)})^2. \quad (1)$$

$$arg\min_{a,b}\left(\frac{1}{2n}\sum_{i=1}^{n}(y_i - ax_i - b)^2\right). \quad (2)$$

$$\frac{\partial \mathrm{LSM}}{\partial b} = \frac{1}{n}\sum_{i=1}^{n}(y_i - ax_i - b)(-1). \quad (3)$$

$$\frac{\partial \mathrm{LSM}}{\partial a} = \frac{1}{n}\sum_{i=1}^{n}(y_i - ax_i - b)(-x_i). \quad (4)$$

Food Calorie and Nutrition Analysis

- Study uses the *Ministry of Health and Welfare Nutrition* Database and *US Department of Agriculture's Food Nutrition* Database as a reference for calories and nutrients.
- Food calorie and nutrition analysis step recognizes the food class and food weight; and estimates the calories and nutrients of the food in an image.
- Nutritional components include crude protein, crude fat, saturated fat, trans fat, carbohydrates, dietary fiber, sugar, and sodium.
- Calories and nutrients listed in database for every 100 grams which are used as reference for estimating calories and nutrients from a food image based on weight.

Ville Café Dataset

- Variety of foods placed on a dish, some covering one another or placed beyond the dish.
- Resolutions of 4000x3000 and 4608x3456.
- Study takes a variety of different servings of the same food and records the weight of each serving.
- Linear regression used to analyze relationship between image ratio in food, food weight based on proportion of food, and actual weight of previous record.
- 16 Classes in Ville Café dataset. 5 different food items from 5 restaurants. 35,842 images and 9,776 food items.

Breakfast Food Detection & Recognition

- Food-256 dataset and Ville Café dataset used for training and testing.
- Three parameters in experiment:
  - Step per epoch (how many images trained each epoch)
  - RPN train anchors per image (number of RoIs output by the RPN)
  - RoI per image (number of anchors used in RPN training)
- One parameter adjusted at a time with remaining parameters fixed to calculate optimal rate for each parameter.
- Average precision rate 98.48%, average recall rate 96.31%, average IoU (Intersection over Union) 97.17%, and average accuracy rate for 16 classes of food 99%. Of 3,680 testing food images, 5 misrecognized and 112 undetected.

No.	Class	# of Training Images	# of Training Food Items	# of Validation Images	# of Validation Food Items
1	Salad	267	286	107	107
2	Fruit	236	516	104	266
3	Toast	274	561	142	265
4	Egg	196	196	64	64
5	Sausage	119	222	57	120
6	Chicken Cutlet	51	153	15	45
7	Bacon	240	618	51	124
8	French Toast	67	179	23	62
9	Omelet	64	67	21	21
10	Hash Browns	90	142	50	68
11	Pancake	130	253	43	86
12	Ham	126	219	134	363
13	Patty	261	405	146	255
14	Sandwich	81	107	36	47
15	French Fries	105	131	33	37
16	Hamburger	80	81	48	48
-	Total	850	4,118	428	1,978

Table 1 – Numbers of Food Image of 16 Classes for Training and Validation Data [Chiang, et al.]

Food Detection and Recognition Model Improvement

- Food images places too close to each other result in non-detected foods.
- Calculating results after union operation show improvement in accuracy, recall, and F1 measure.

(a) (b) (c) (d)

Figure 7 – (a) Input Image, (b) Toast Detection Failure, (c) Input Image, (d) Pancake and Ham Detection Failure [Chiang, et al.]

No.	Class	Mask Color	Precision (With NMS)	Precision (Union)	Recall (With NMS)	Recall (Union)	F1 Measure (With NMS)	F1 Measure (Union)
1	Salad		100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
2	Fruit		100.00%	100.00%	98.19%	98.36%	99.09%	99.17%
3	Toast		**95.92%**	**100.00%**	**92.17%**	**100.00%**	**94.01%**	**100.00%**
4	Egg		100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
5	Sausage		100.00%	100.00%	**96.93%**	**100.00%**	**98.44%**	**100.00%**
6	Chicken Cutlet		100.00%	100.00%	**94.65%**	**95.88%**	**97.25%**	**97.90%**
7	Bacon		98.98%	98.98%	99.49%	99.49%	99.23%	99.23%
8	French Toast		100.00%	100.00%	**97.03%**	**99.26%**	**98.49%**	**99.63%**
9	Omelet		100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
10	Hash Browns		99.60%	99.60%	100.00%	100.00%	99.80%	99.80%
11	Pancake		**97.63%**	**100.00%**	**95.38%**	**100.00%**	**96.49%**	**100.00%**
12	Ham		**93.38%**	**94.08%**	**87.58%**	**88.82%**	**90.38%**	**91.37%**
13	Patty		93.71%	93.71%	88.16%	88.16%	90.85%	90.85%
14	Sandwich		96.97%	96.97%	94.12%	94.12%	90.52%	95.52%
15	French Fries		97.44%	97.44%	100.00%	100.00%	98.70%	98.70%
16	Hamburger		95.16%	95.16%	90.77%	90.77%	92.91%	92.91%
-	Total		98.48%	99.09%	96.31%	97.91%	97.38%	98.50%

Table 2 – Accuracy Rates of 16 Classes w/ NMS and Union Operation [Chiang, et al.]

Food Detection and Recognition Model Improvement (Cont.)

Figure 8 – Example of Input Image and Improved Results [Chiang, et al.]

Food Weight Estimation

- Linear regression performed in food weight estimation experiment.

	Absolute Error	Relative Error
Salad	2.71	0.34
Fruit	8.45	0.11
Toast	15.98	0.19
Sausage	8.00	0.11
Bacon	2.50	0.08
Ham	1.79	0.07
Patty	6.53	0.06
French fries	9.83	0.04
Total	8.22	0.13

Table 3 – Absolute Error and Relative Error for Food Weight Estimation [Chiang, et al.]

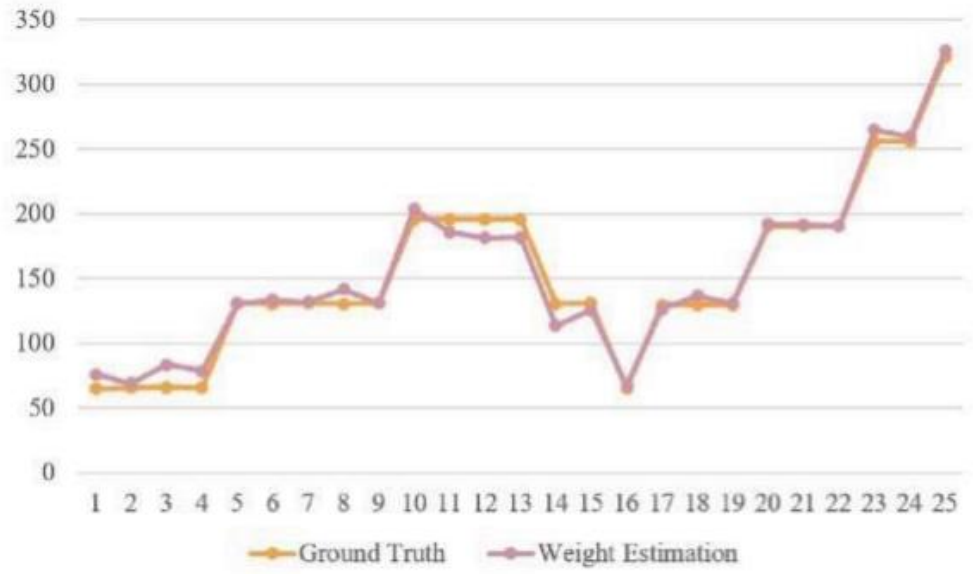

Figure 9 – Weight Estimation Results for patty where x-axis is number of pixels in food image and y-axis is estimated weight of food [Chiang, et al.]

Conclusion and Future Work

- Conclusion:
  - Mask R-CNN based System aimed to help users manage their diet through food recognition and calorie nutrient analysis.
  - Ville Café dataset proposed.
- Limitations (possible future work):
  - System still needs to be enhanced to allow users to record meals not limited to breakfast foods.
  - Does not provide dietary advice for patient with different conditions.

References

- M. Chiang, et al. "Food Calorie and Nutrition Analysis System based on Mask R-CNN." *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*.