End-to-End Vision Based Calories Estimation

By George Lu luq7@rpi.edu @ Computer Science & Aerospace Engineering

Advisor: Dr. Oshani Seneviratne senevo@rpi.edu

Problem Statement

To aid the process of food recommendation, given a food image of any size, efficiently and accurately identify the food items and estimate the calories of the food.

Accuracy is not a strict requirement. The estimated calories result will potentially be used for the user's next meal recommendation.

Efficiency is crucial. User would not want to wait too long.

Method should be interpretable for users.

Related Work

ECUSTFD:

1. achieved fairly accurate result on calories estimation of around 20 types of fruit.
2. Two shots prediction, includes multiple steps.
3. Not practically meaningful.

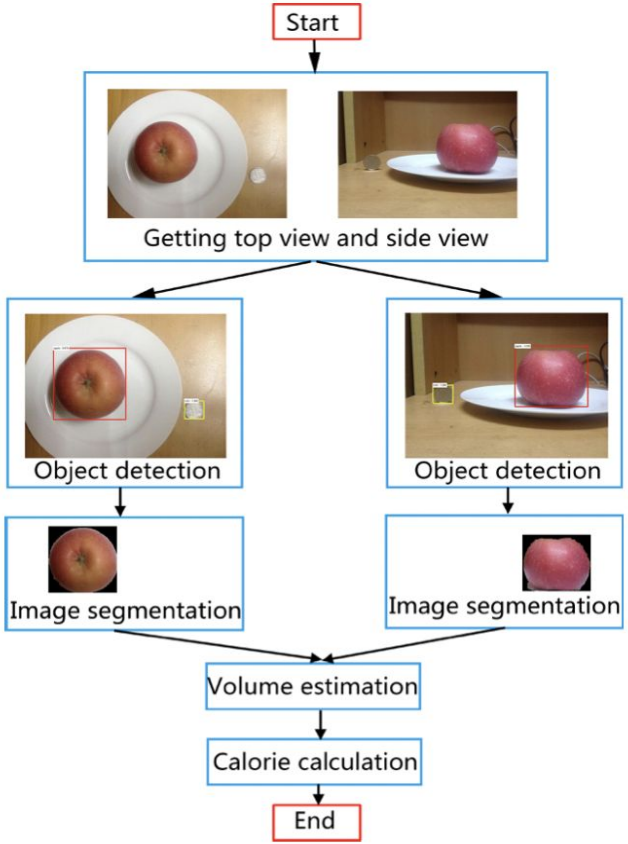


Fig. 3. Calorie Estimation Flowchart

From https://arxiv.org/pdf/1705.07632.pdf

Related Work

Multi-Task Learning for Calorie Prediction on a Novel Large-Scale Recipe Dataset Enriched with Nutritional Information:

1. a multi-task CNN which estimates the calories, average ingredient word embedding, the type of food, and an embedding of the preparation steps of the meal
2. Only accurate on certain types of food.

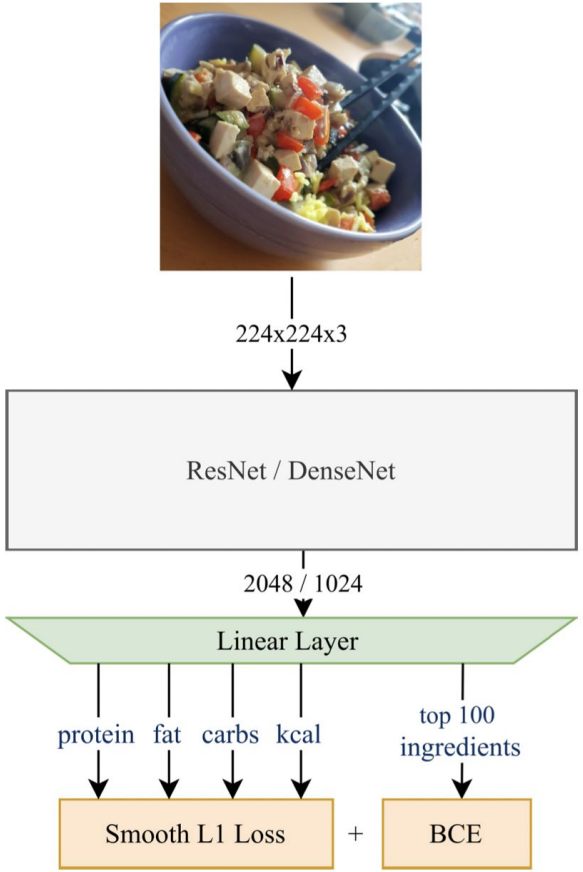
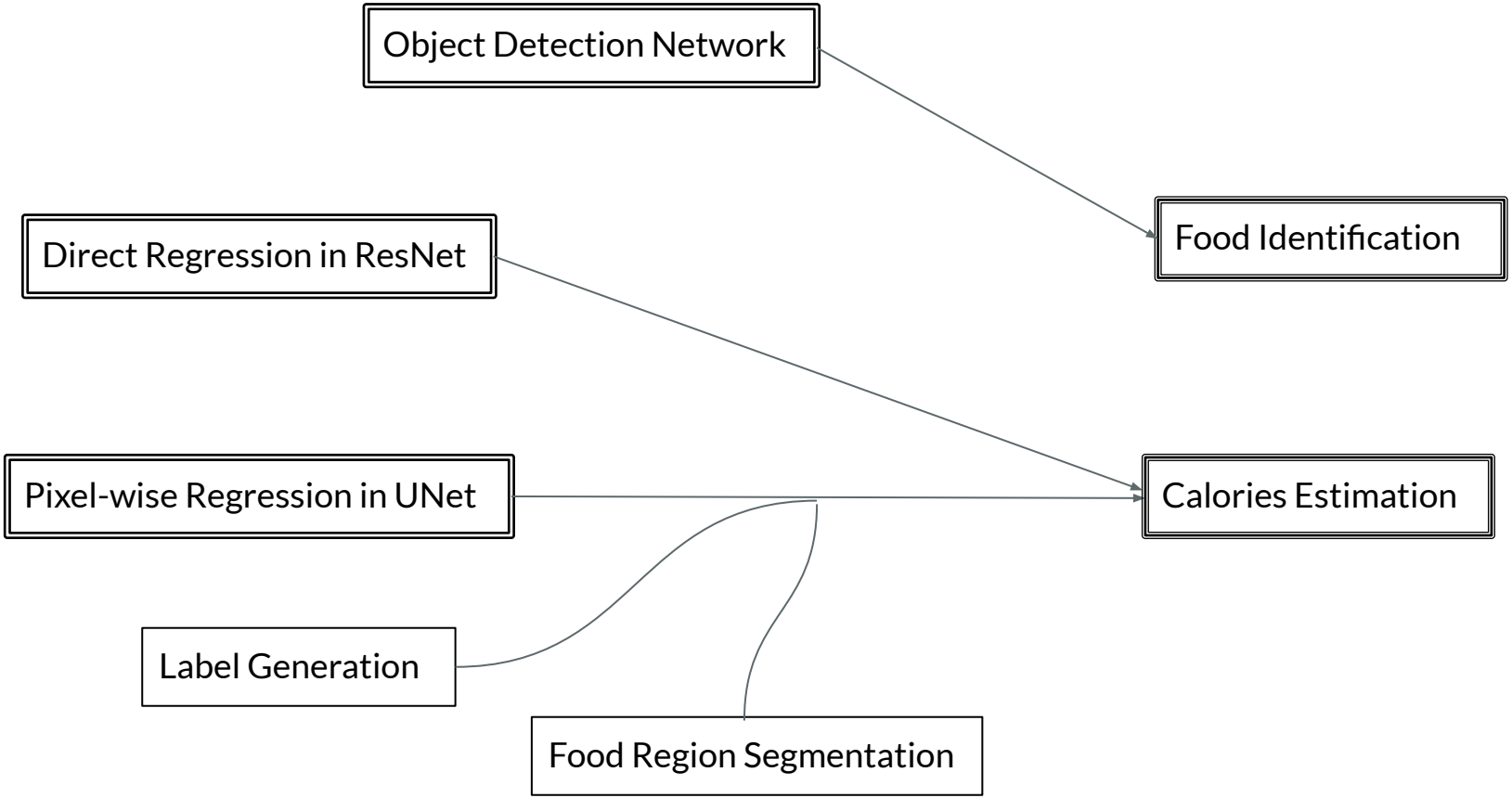


Fig. 2: Overview of our proposed architecture. Our model is trained in a multi-task setup inferring the calories, the ingredients, and the macronutrients (e.g. protein).

From https://arxiv.org/pdf/2011.01082.pdf


Object Detection Network
Food Identification
Direct Regression in ResNet
Pixel-wise Regression in UNet
Calories Estimation
Label Generation
Food Region Segmentation

Dataset : Nutrition 5k

Published by google

Direct Regression

- Rather “trivial”, serve as a comparison benchmark
- ResNet with appropriate modification of parameters to fit a regression task of RGB image
- Input images resized to the same size
- Hyper parameter tuning

Pixel-wise Regression

Before training:

Step 1: run a image segmentation model to mask out food regions

Step 2: create the calories label maps so that calories are only distributed to pixels belonging to a food region

Step 3: modify the original UNet model

Food region segmentation

Open source pretrained model from
https://tfhub.dev/google/seefood/segmenter/mobile_food_segmenter_V1

Calories Label Creation

resize to 512*512

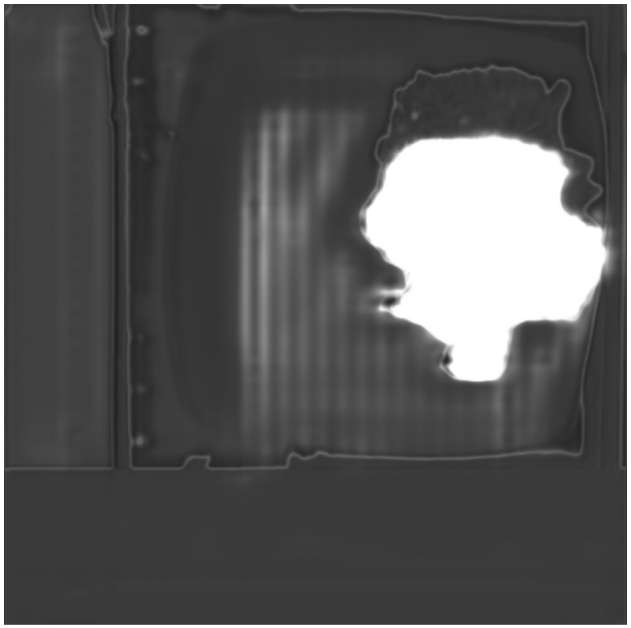
Modifying Unet

- ReLu Layers everywhere possible:
  - The calories should always be a positive value
- Number of classes = 1, output the logits map, so this becomes a 2-D Regression
- MSE over MAE

Explainable Output

True: 57.5000 kcal

Prediction: 63.3885 kcal



Result

	Mean Absolute Error(in kcal)
ResNet50	76.18592637160728
Unet + pixelwise loss	71.36310889437738
Unet + joint loss	77.9916347123654

Future Work

- Attention based model for better accuracy
- Use model pre-trained on food knowledge and apply it to downstream vision task