SeeFood

CS231N (Spring 2022)

Wai Tong Chung (wtchung@stanford.edu)
Mahesh Saboo (mssaboo@stanford.edu)
Chiin-Zhe Tan (chiin@stanford.edu)

Introduction

Objective

- Food detection from a photo
- Identify food objects present
- Feed ingredients directly into recipe database
- Automate and augment decision process

Background

- Universality of shopping and cooking experiences
- Significant potential for technological facilitation
- Existing models have a more general scope

Problem Statement

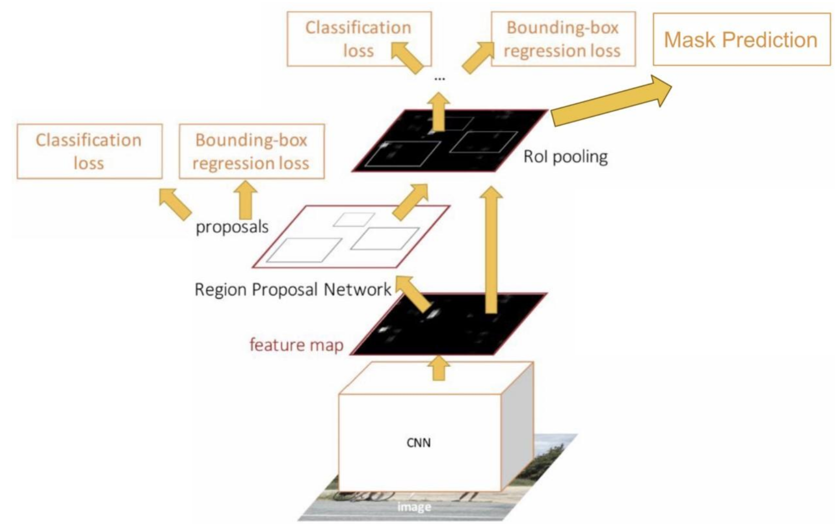
- Using computer vision techniques, build a food detector
- Train a model on a dataset of food objects
- Compare with baseline model using qualitative and quantitative metrics

Quantitative metric: Mask Average Precision

- Average Precision (using IoU and a threshold)
- Compute for thresholds from 0.5 to 0.95 (at 0.05 increments)
- Average over classes and thresholds

Mask R-CNN

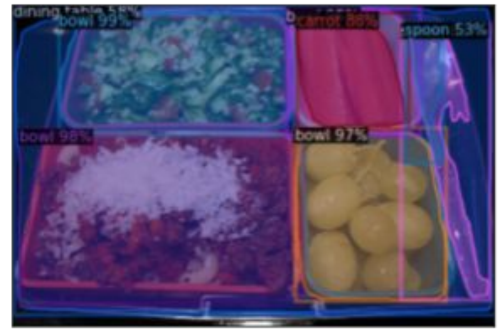
- Mask R-CNN for Instance Segmentation
- Extension to Faster R-CNN
- Additional loss term for predicting segmentation mask on each RoI
- Used the model mask_rcnn_R_50_FPN_3x from detectron2
- Backbone of ResNet-50-FPN
- ResNet-FPN backbones for feature extraction with Mask R-CNN have been known to give excellent gains in both accuracy and speed



Overview of Mask R-CNN

Results: Baseline (Pretrained on MSCOCO)

Class	banana	apple	sandwich	orange	broccoli	carrot	pizza	donut	cake	hotdog
seg AP	25.6	24.3	42.3	32.5	28.9	24.7	56.9	55.1	37.5	28.6
box AP	29.7	25.1	42.2	32.9	30.7	27.9	58.6	55.1	37.6	31.7



Results: Our Model (Trained with a processed LVIS)

Class	banana	apple	sandwich	orange	broccoli	carrot	pizza	donut	cake	hotdog
seg AP	11.1	13.3	11.0	14.6	15.2	11.5	4.1	28.6	22.5	7.8
box AP	12.7	12.8	10.8	14.0	11.2	15.3	8.6	28.4	21.7	7.5

Class	lettuce	mushroom	nut	onion	pastry	potato	strawberry	tomato	wine
seg AP	1.8	1.0	0.6	0.5	0.6	0.4	1.3	4.7	13.4
box AP	1.8	1.0	0.9	0.5	0.6	0.3	1.1	4.7	13.8

Class	beer	bell pepper	blueberry	bread	crumb	cupcake	grape	lemon
seg AP	2.7	0.1	0.6	4.7	0.3	9.6	1.4	0.2
box AP	2.1	0.1	0.5	4.9	0.3	8.5	1.4	0.2

Conclusions

- Assembled comprehensive food database
- Used transfer learning to train model to detect 17 new classes
- Qualitative results: promising, with failures on similar food objects
- Quantitative results: less promising
- Future work: further annotation, focusing on mispredicted classes