Simplifying Grocery Checkout with Deep Learning

Youtube Video Link :https://youtu.be/qEq8ZrTR0UI

Jing Ning, Yongfeng Li, Ajay Ramesh {annaning, liyongfe, ajay_ramesh}@stanford.edu

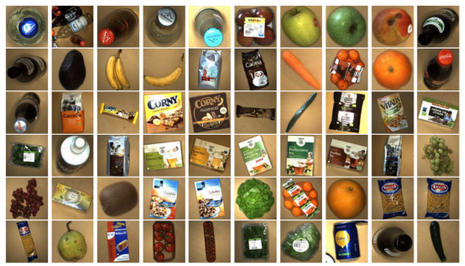
Department of Computer Science, Stanford University

INTRODUCTION

Problem Statement: Fruits and vegetables are usually not packaged in boxes. Most stores use plastic tags with identification numbers on the surface, which presents several problems.

- ✦ Unsustainable and high maintenance cost
- ✦ Not Environment-friendly
- ✦ More label-intensive for store

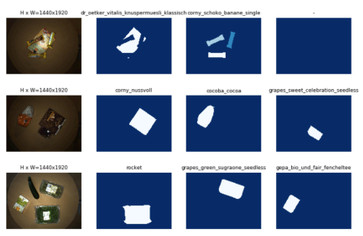
Goal: build a DL system that can efficiently classify and localize fruits and vegetables. We aim to apply the state-of-the-art technique Mask R-CNN.



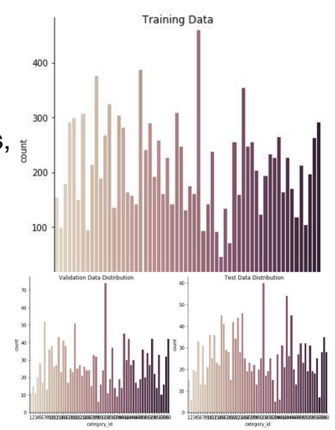
DATASET

MvTec Densely Segmented Supermarket(D2S) Dataset

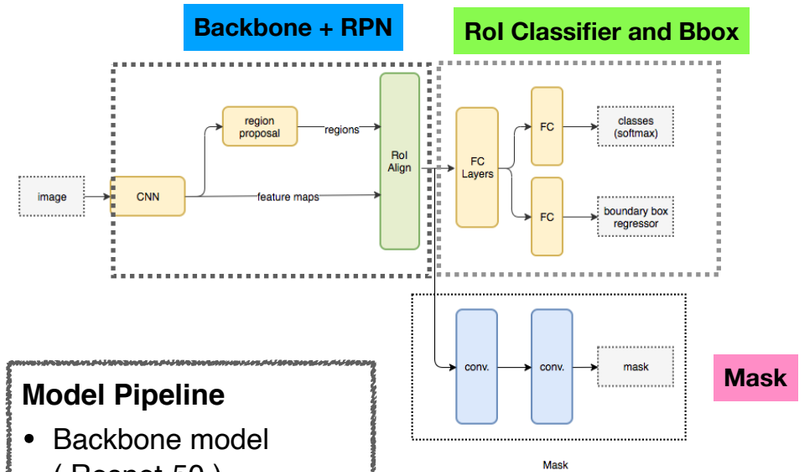
- 60 unique categories
- 1920*1440 RGB images
- Training/Dev/Test split: 2880/360/360
- Annotation format similar to COCO
- With various lighting, rotations, and backgrounds



- **Class Distribution**



MODEL ARCHITECTURE (MASK-RCNN)



Model Pipeline

- Backbone model (Resnet-50)
- Region Proposal Network (RPN)
- Region of Interest Classification and Bounding Box
- Segmentation Masks

Mask

$$L = L_{cls} + L_{box} + L_{mask}$$

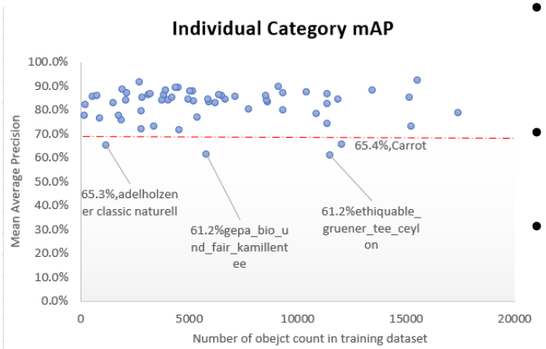
$$L_{loc}(t^u, v) = \sum_{i \in x,y,w,h} smooth_{Li}(t_i^u - v_i)$$

$$smooth_{Li}(x) = \begin{cases} 0.5x^2 & if|x| < 1 \\ |x| - 0.5 & otherwise, \end{cases}$$

$$L_{mask} = -\frac{1}{m^2} \sum_{i \in i,j} [y_{ij} log \hat{y}_{ij}^k + (1 - y_{ij}) log(1 - \hat{y}_{ij}^k)]$$

Loss function for each RoI

ERROR ANALYSIS



- Analysis on mAP on each category and relationship between object count performed
- Carrot, Adelholzener water, and Kamillentee tee have lowest mAP
- Error appear to relate to difference in size of objects from the same category and object occlusion with limited label displaying

RESULTS

Transfer learning was applied and various of parameters were experimented. Model is initiated with COCO pre-trained weights.

Results: mAP of 84 is achieved at IoU 50:95, with tolerance to partial obstruction.

- **Backbone Network**

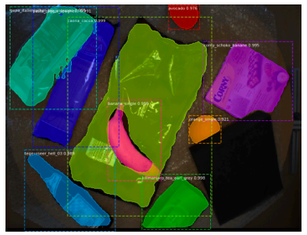
backbone	$AP_{50:95}$	AP_{50}	AP_{75}	AP_M	AP_L
ResNet-50	0.753	0.962	0.962	0.314	0.756
ResNet-101	**0.788**	**0.957**	**0.924**	**0.291**	**0.791**

- **Image Size**

Image Size	$AP_{50:95}$	AP_{50}	AP_{75}	AP_M	AP_L
800*1024	**0.837**	**0.979**	**0.979**	**0.61**	**0.839**
512*512	0.753	0.962	0.962	0.314	0.756
256*256	0.04	0.072	0.042	0	0.058

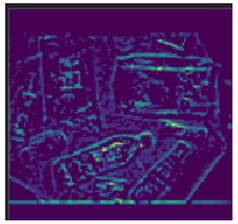
- **Number of layers to re-train**

layers to tune	$AP_{50:95}$	AP_{50}	AP_{75}	AP_M	AP_L
heads	0.625	0.946	0.764	0.147	0.628
3+	**0.84**	**0.98**	**0.97**	**0.451**	**0.843**
4+	0.795	0.976	0.958	0.441	0.798
all	0.753	0.962	0.962	0.314	0.756



Example of model prediction result

- **Feature Maps from Resnet layer**



FUTURE WORK

We show that Mask R-CNN can achieve state-of-art result in object instance segmentation with the right hyper-parameter tuning.

For future work, we could expand the model's application to other tasks such as fruit/vegetable sorting and quality control. To build a robustness model, increase the level of complexity of dataset, including collecting more data with images from different angles, with different background, etc.