Package Counting for Shipping Logistics

Jorge Isaac Chang Ortega, Eric Wang, Ruotao Zhang

CSCI 1470/2470 Final Project, Brown University

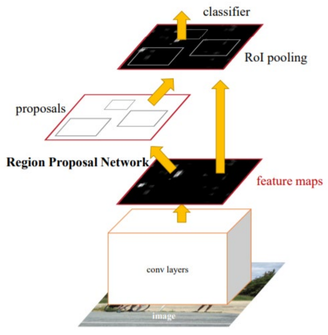
Introduction

Companies with shipping services like Amazon and FedEx require various systems to understand their flow of packages for pipeline analysis and optimization. One important process is counting the number of packages flowing through a particular pipeline, such as a delivery loading window or warehouse storage. Counting packages can be tedious and prone to error when done manually. Thus, we want to leverage deep-learning to create an efficient and accurate computer vision solution for automated package counting. We explore solutions to this problem by researching and augmenting Faster R-CNN for bounding box prediction and the SORT algorithm for the package counting.

Methodology

The complete package counting system consists of two parts, object detection + object tracking.

- Object Detection Model (Faster R-CNN)
  - Obtain feature map from input image
  - RPN computes region proposals
  - Outputs class labels and bounding boxes

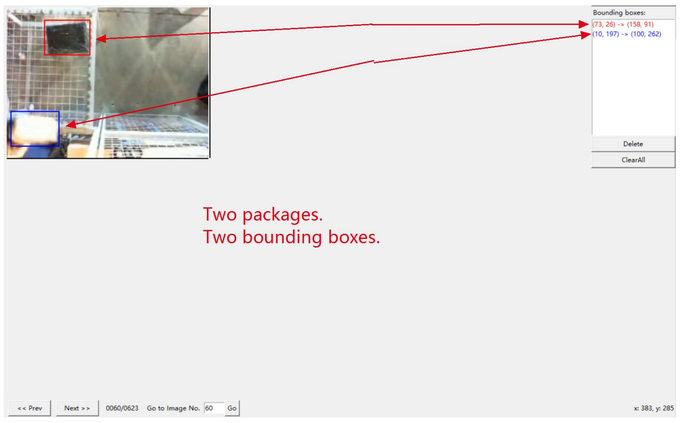


- Object Tracking Algorithm (SORT)
  - Simple Online and Realtime Tracking
  - Keeps track/updates all package states (position and velocity)
  - Predict next states with simple motion model
  - Associates state predictions from (1) to network's detected bounding boxes

Datasets

- **Packages Dataset (original unlabeled):**
  1. Video recordings of people passing packages under a camera on 4 different days: Sep 25, 29, 30 and Oct 2.
  2. Video lengths ranging from 30 seconds to 180 seconds.
  3. 47 Videos in total.
  4. Training set: Sep 30 and Oct 2. Test set: Sep 25 and 29.

- **Packages Dataset (processed and labeled):**
  1. 5,268 labeled video frames.
  2. Each is labeled with information of whether there are packages in the frame. If yes, then bounding box coordinates are also given.



Model Training

- Faster R-CNN model (ResNet-50 backbone) pre-trained on COCO train2017 (downloaded from PyTorch)
- Fine-tune the model with our own Packages dataset (5,268 labeled video frames)
- SORT hyperparameter tuning
- Position of virtual line (L): 0 - top diagonal, 1 - middle horizontal, 2 - bottom diagonal
- Max age for trackers (A): 1, 2, 3

Results

Error rate					
L	A	25-Sep	29-Sep	30-Sep	2-Oct
0	1	81.67	73.21	42.86	25.26
	2	82.14	73.17	42.86	25.26
	3	82.14	73.3	42.86	25.43
1	1	76.63	75.38	26.56	25.23
	2	77.02	75.38	26.56	25.47
	3	77.02	75.51	26.56	25.47
2	1	68.37	67.06	49	20.78
	2	60.35	67.01	50.25	20.78
	3	60.35	66.64	50.25	20.78

- Error was minor in the training set. We consider that the test set error is greater since this set of videos was recorded with a much higher frame rate. This frame rate affects the tracking model since it is more likely that the package's position will be lost.
- It is possible to notice that the horizontal middle line (L=1) gives the best results. Also, the max-age parameter (A) seems not to affect the package count. We consider that to notice a change, we need more frames to store the objects in memory. Three consecutive frames do not have much visual difference.