Billing system using product image segmentation

Karthik Sriram *Manish Soni*
New York University *New York University*

Abstract:

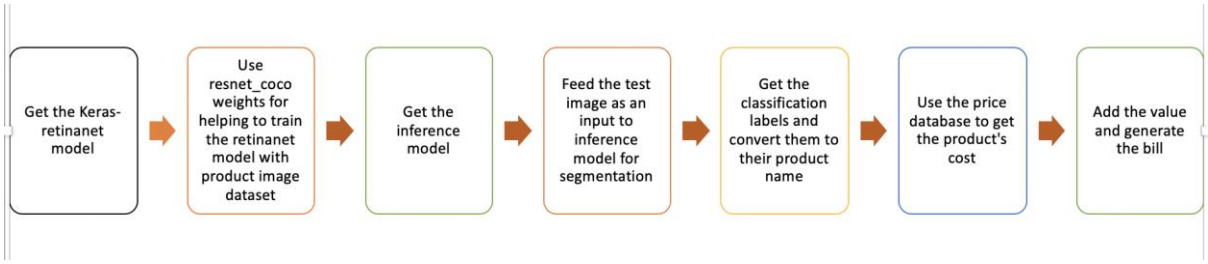
Goal: Minimal human intervention billing and analysis system based on real time object detection

The project we implemented is a real time object detection-based billing system. We wanted to build a solution around object detection and evaluating its price, and analysis of sell for inventory management (can be extended).

The idea is to use a deep neural network approach for various object detection so to evaluate price of all the items. And to use regression analysis for predicting the inventory suggestions.

Approach

We consider that at very basic level image segmentation is required to detect different products available in that image. And once all the products are recognized we can use the price dataset to evaluate the complete price.



Result:

```
Running inference on: test9.jpg
processing time:  0.06533980369567871
```


```
products labels are: [2, 2, 1, 0]
products score are: [0.9851149, 0.9413011, 0.9386572, 0.6491825]
-----------------------
Product Label 2
Product Name: kitkat
Inference score: 0.9851149
=========================
Product Label 2
Product Name: kitkat
Inference score: 0.9413011
=========================
Product Label 1
Product Name: hershey
Inference score: 0.9386572
=========================
Product Label 0
Product Name: pringle
Inference score: 0.6491825
=========================
```

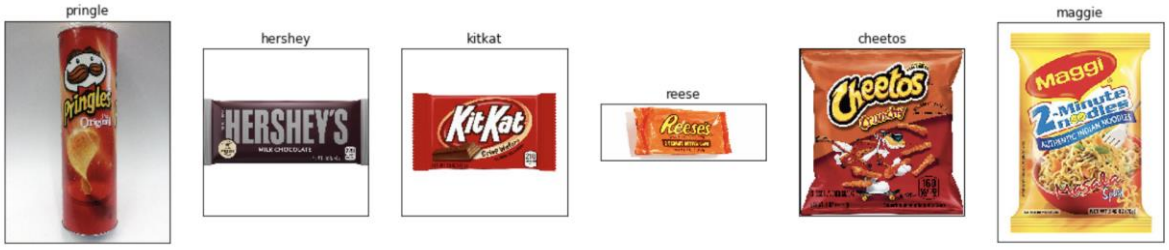
As per shown result we can see that our inference model trained on RetinaNet model works absolutely fine and provides us the list of products segmented in the image (such image could be the result of camera capture input when customer place all the purchased items in a tray or relevant place)

Output Bill Amount:

```
ABC XYZ Limited
-------------------
1 : item: kitkat
item code: 2
price: 2.99 $
---------------------
2 : item: kitkat
item code: 2
price: 2.99 $
---------------------
3 : item: hershey
item code: 1
price: 5.0 $
---------------------
4 : item: pringle
item code: 0
price: 1.67 $
---------------------
=======================
Total Amount= 12.65 $
```

Description about Products in Image dataset

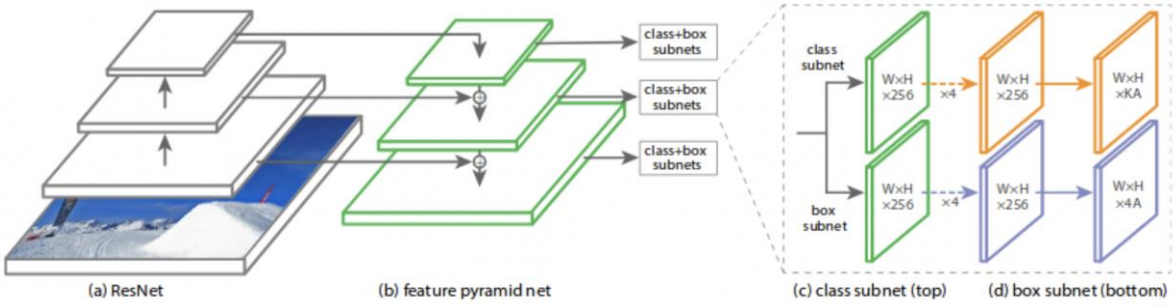
Consider that abc xyz store has 6 different products as shown below that are used for training retinanet model for product image segmentation.



Description

Formulation: The billing system designed solves the problem of reducing the time spent standing in queue, be it register or self-checkout. Statistically, by multiple surveys, it has been identified that the average wait time at a store like Target or Trader Joe's during a busy weekday is anywhere between 8-10 minutes and requires assistance on multiple levels to guide customers through queues, thereby leading to more employment intake. Further, the change in customer behaviour has been

noticed to change into frustration (primarily due to impatience) as well as customer return numbers tend to decrease due to the increased wait times at all types of stores. As per a survey done at a reliable retailer, 43% of its consumers said long lines would affect their decision to shop at a retailer in the future. Out of those consumers: • 21% said they would avoid the store if they knew the checkout lines would be long at the time. • 19% said they would only go to the store to pick up specific items they couldn't find at other stores. 3% said they would stop going to the store all together. This data analysed provides an insight that we have to model a billing system that identifies objects and the customers who picked them up, and then presents the total value of the objects. The problem is clearly mathematically stated, and the RetinaNet takes the dataset from image acquisition and performs object detection using the following mathematical
model:



The RetinaNet basically leads to 2 models: The classification model or class subnet which is a fully convolutional network (FCN) given to every FPN network. The model uses convolution layers and then sets up RELU activations. Then, a convolutional layer with K×AK×A filters are used which leads to sigmoid activation. The parameters then become (W,H,KA)(W,H,KA), in which

WW: Width proportionality of the input parameters.

HH: Height proportionality of the input parameters.

KK: shape of classes.csv

AA: shape of annotations.csv.

The loss here is:

$$Lclass=-K\sum i=1(yilog(pi)(1-pi)\gamma\alpha i+(1-yi)log(1-pi)p\gamma i(1-\alpha i))$$

The regression model or box subnet is applied to the FPN simultaneously to the classification model. The only difference is that the last convolutional layer uses filters that are 4 times that of A. This leads to a shape of (W,H,4A). The loss here is:

$Lreg=\sum j\in\{x,y,w,h\}$ smooth$(Pij-Tij)$ where Pij and Tij are the prediction and target parameters.

The total loss is the sum of regression and classification loss. In our model, it is low which leads to higher accuracy of prediction based on the target. Therefore, this formulation addresses the key aspects to the problem.

Future- Extension: This project can be extended with a sophisticated network of high-resolution cameras to evaluate the price of objects as soon the customer picks or drops a product item from certain area. Which makes the whole shopping experience for the customer smooth. While in parallel the same system can recognize the different customers and help them with their shopping experience.

Note: Shown products in images and trademark are the property of their respective companies. these images are used for experiment purpose.