Real Time Monitoring of CCTV Camera Images
Using Object Detectors and Scene Classification for Retail and Surveillance Applications – Anand Joshi (aaj1031@stanford.edu)

Introduction
Current surveillance and control systems in retail and elsewhere, still require human supervision and intervention. This work will try to provide a detection system in CCTV videos on real time basis, appropriate for; surveillance and control, inventory tracking, theft deterrence, threat perception and detection etc. and apply Machine Learning/Deep Learning techniques for real world applications. This will try to automate many tasks, which can be error prone otherwise due to human errors and fatigue. This solution can potentially have capability to provide real time alerts, notification on smart phones/tablets and provide rich data for analytics purpose.

Dataset
The Dataset Comprised of Color images in following categories:

a) Every Day Objects found in retail environment, obtained from ImageNet. Over 1.2 million images used for training, divided in over 1000 classes.
b) Guns and Knives: Knives Images Database, which contains 9340 negative examples and 3559 positive examples, Internet Movie Firearms Database, which contains 8557 images
c) Human Hand: Hand Dataset which contains about 14700 hand images from various sources. EgoHands Dataset containing about 120000 images. 90% was used for training and 10% was used for validation.

Some Image Samples

CNN Architecture: Inception-Resnet V2. It is more accurate than previous state of the art models. The Top-1 and Top-5 validation accuracies on the ILSVRC 2012 image classification benchmark based on a single crop of the image is 80.4 & 95.3 respectively.

Deep Learning Framework – TensorFlow, for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them.

REFERENCES
[2] Scalable Object Detection using Deep Neural Networks

Future
Despite Inception-ResNetV2 performing the best, I found that many predictions had a probability of 20% to 40%, even if these predictions were correct. The first step I would like to take is to increase the confidence in these predictions so that the model would be more well trained. This could be done by training it on more data or increasing the epochs when training the CNN. Also after developing an end-to-end Proof Of Concept solution, I strongly feel that it has the potential of becoming a commercially viable product.