Rapiscan systems
An OSI Systems Company

Deep Convolutional Object Detection for X-ray Baggage Screening

Dan Strellis[1]
Kevin Liang[2]
ADSA21
November 5, 2019

[1] **Rapiscan Systems; Fremont, California**
[2] **Duke University; Durham, North Carolina**

This research has been funded by the Transportation Security Administration (TSA) under Contract #HSTS04-16-C-CT7020

SO WHAT, WHO CARES?

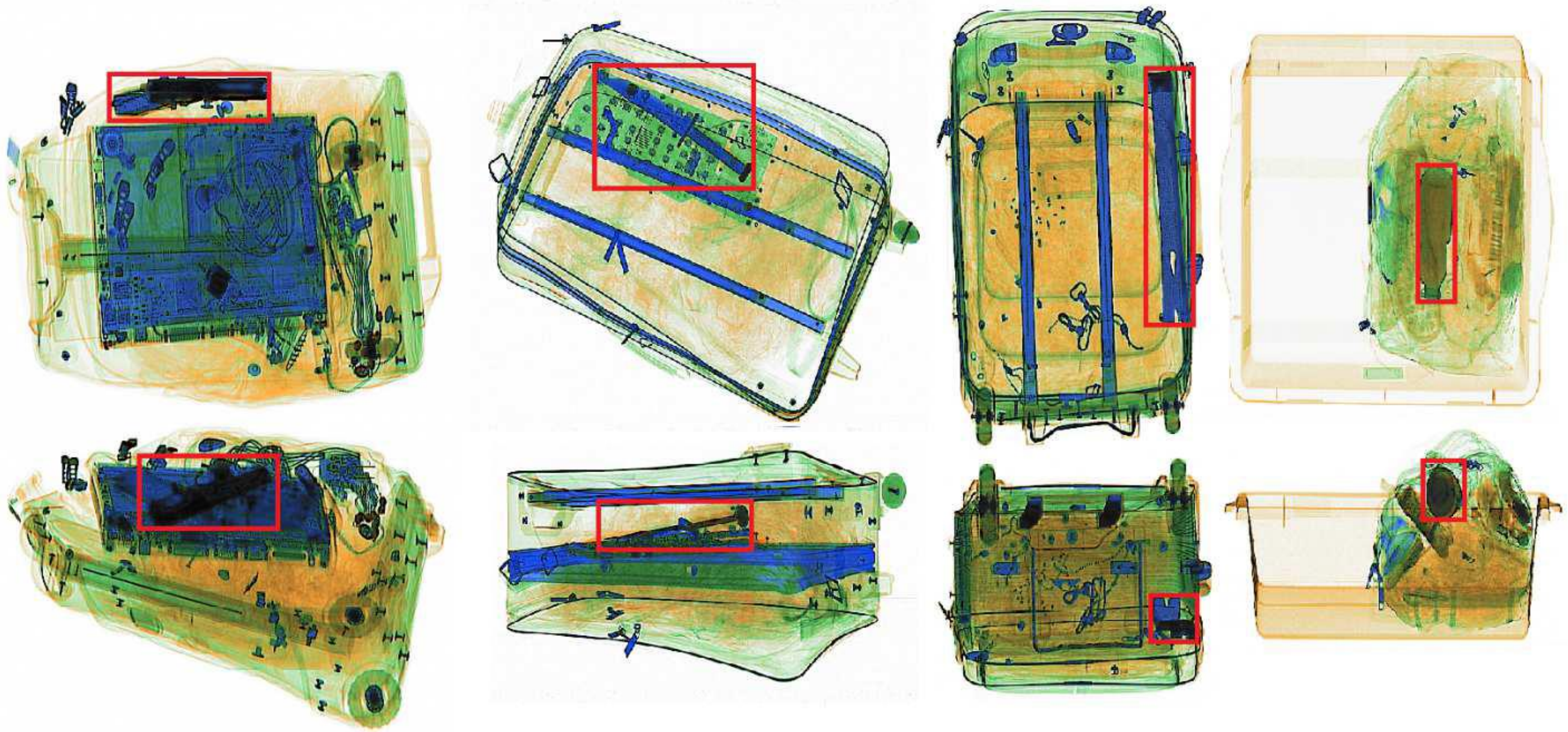
- **Mission space:** Prohibited items detection in carry-on items
- **Problem:** Need to increase detection, reduce cognitive load on TSA screeners while maintaining throughput
- **Solution:** Demonstrate a prototype deep learning based operator assist algorithm for guns, sharp objects, blunts, and non 3-1-1 liquids on board existing X-ray machine
- **Results:** Fieldable model mean average precision, mAP ~ 0.92 and 250ms/image latency [Duke's analysis], across 4 classes
- **Technology Readiness Level:** 7 – Prototype demo in operational environment
- dstrellis@rapiscansystems.com, kevin.liang@duke.edu

APPROACH

Rapiscan systems
An OSI Systems Company

- Collected X-ray images with 620DVs
  - Over 13,000 with Firearms + parts, sharp objects, blunt objects, and liquids
  - Over 450,000 Stream-of-Commerce (SOC) from five U.S. airports
- Hand-labeled the threats in the images with tight bounding boxes
  - Split data into 70/10/20 train/validation/test sets
- Trained and compared 4 popular convolutional object detection models
  - SSD-InceptionV2
  - Faster-RCNN-ResNet101
  - Faster-RCNN-ResNet152
  - Faster-RCNN-InceptionResNetV2

GROUND TRUTH

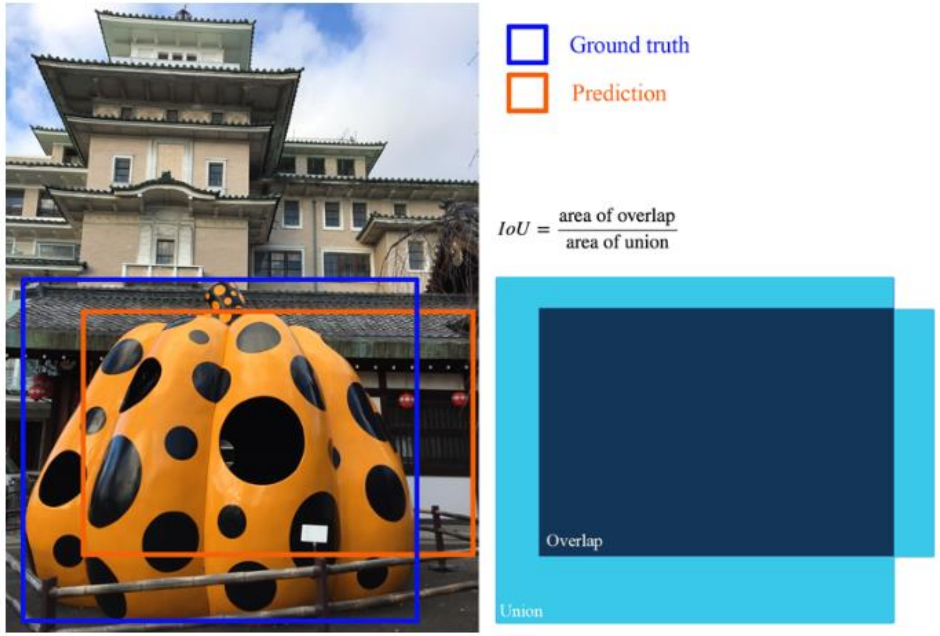


Pistol Knife Crowbar Bottle of Liquid

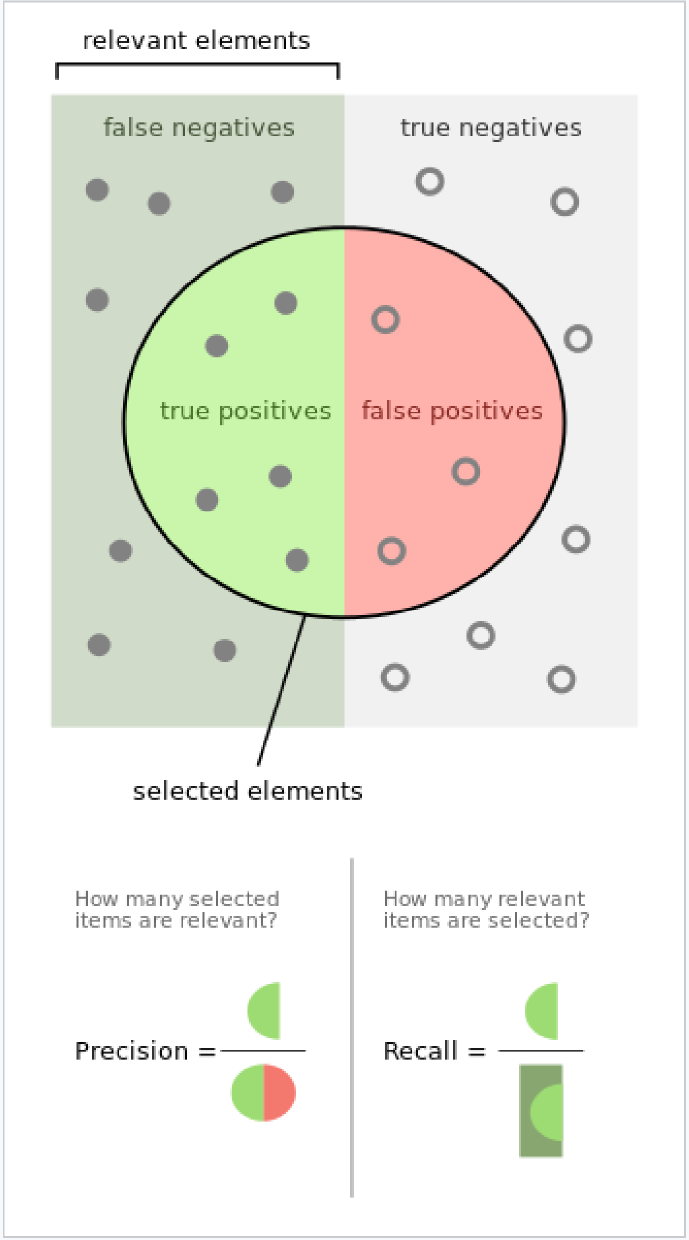
Images shown were obtained from Rapiscan-owned 620DV not in the TSA configuration

EVALUATION METRICS

Intersection over Union (IoU): measures the overlap of two bounding boxes (e.g. ground truth and a detection)



IoU definition



Precision and recall

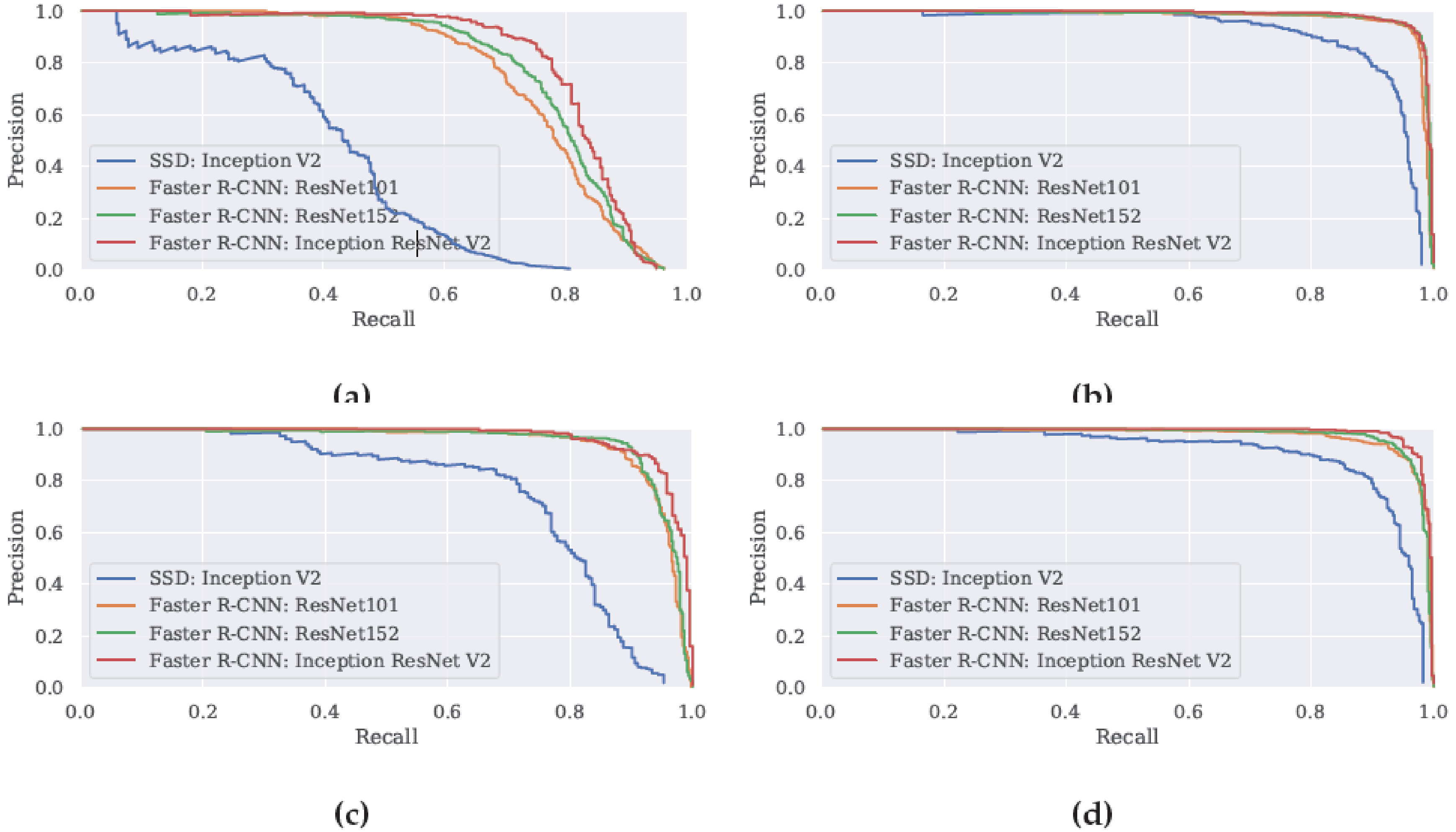
Precision = % of detections that are correct

Recall = % of objects that you find

Average Precision is the area under the Precision vs Recall (PR) curve

Images courtesy of Wikipedia and Medium.com

PR RESULTS



(a) Sharps
(b) Blunt objects
(c) Firearms
(d) Liquids

Note: Analysis is performed by Duke and not confirmed by TSA.

INFERENCE TIME RESULTS

- Inference time with NVIDIA GeForce GTX1080 GPU

Model	Average Latency (ms/image)	mAP across 4 classes
SSD-Inception V2	42	0.752
Faster-RCNN-ResNet101	222	0.917
Faster-RCNN-ResNet152	254	0.924
Faster-RCNN-InceptionResNetV2	812	0.941

- Algorithm needs to run with minimal latency to keep up with passenger flow. Average of <750ms was our measure.
- We chose the ResNet152 option for our prototype implementation

IMPLEMENTATION

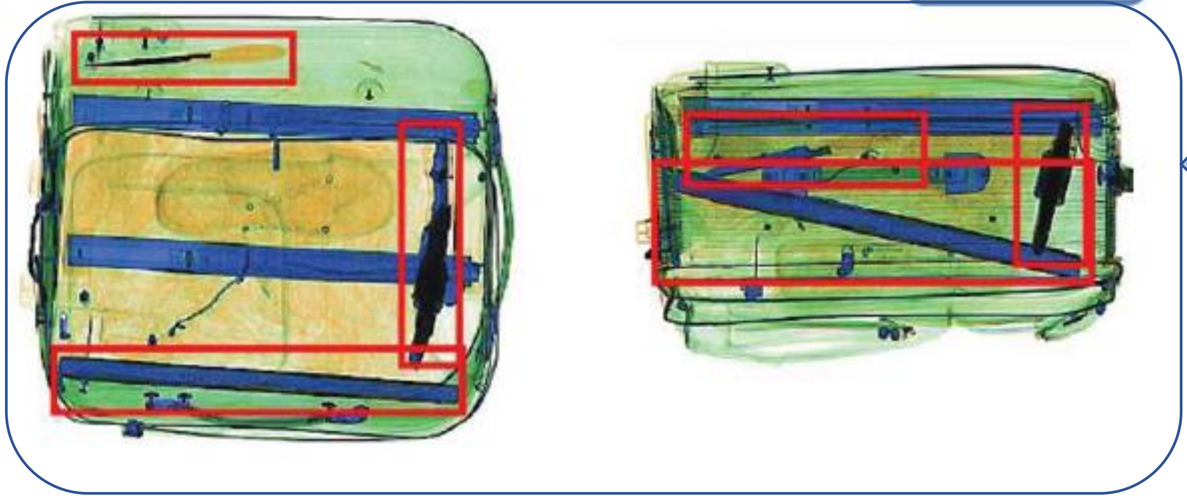
Data Acquisition
- Produce X-ray data from dual-energy transmission in two complementary views

Material Characterization
- Calculate Z_{eff} per pixel from dual energy attenuations

Image Creation
- Assign color to every pixel based on Z_{eff}

Object Detection Inference
- Run trained forward model on *RGB* image

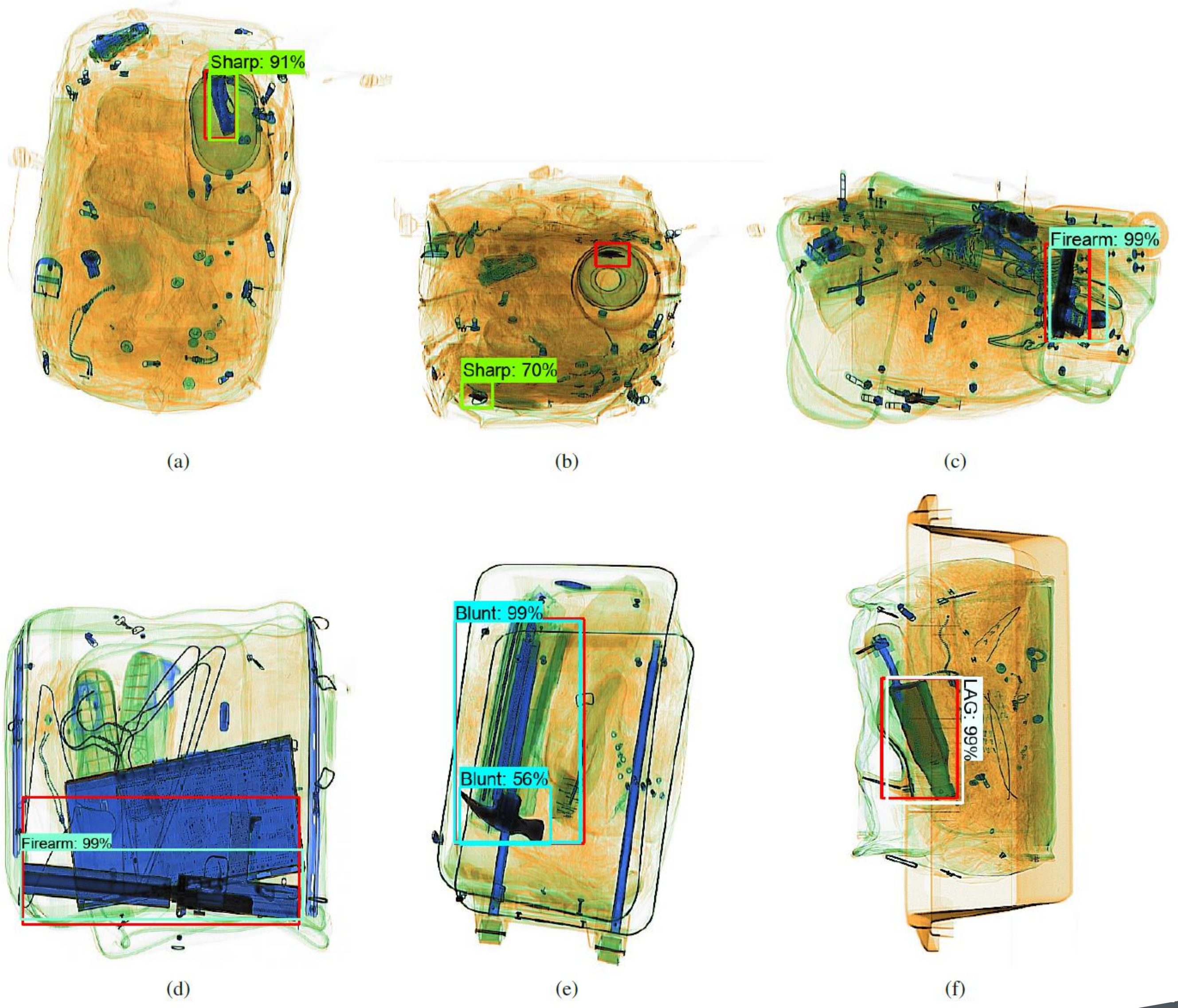
Image and Alarm Display
- Display *RGB* image with alarms for TSO



- Weapon alarms displayed on Operator Workstation monitors
- Operates with no significant latency with NVIDIA GeForce GTX1080 GPU

Images shown were obtained from Rapiscan-owned 620DV not in the TSA configuration

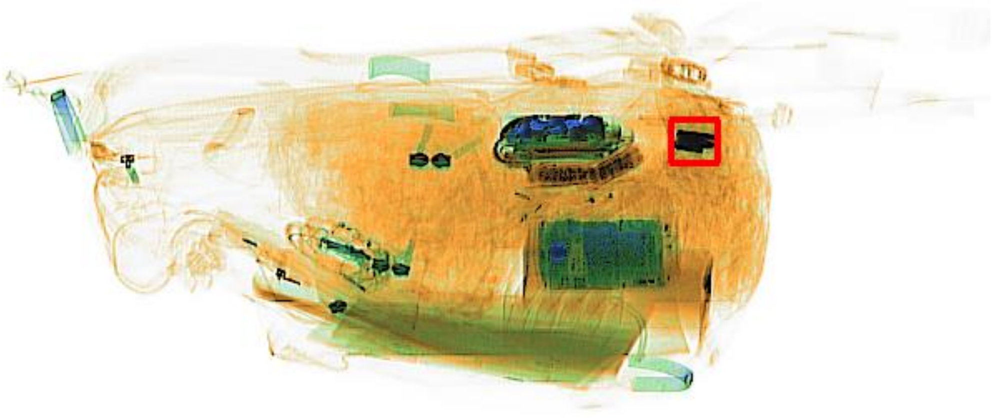
EXAMPLE DETECTIONS

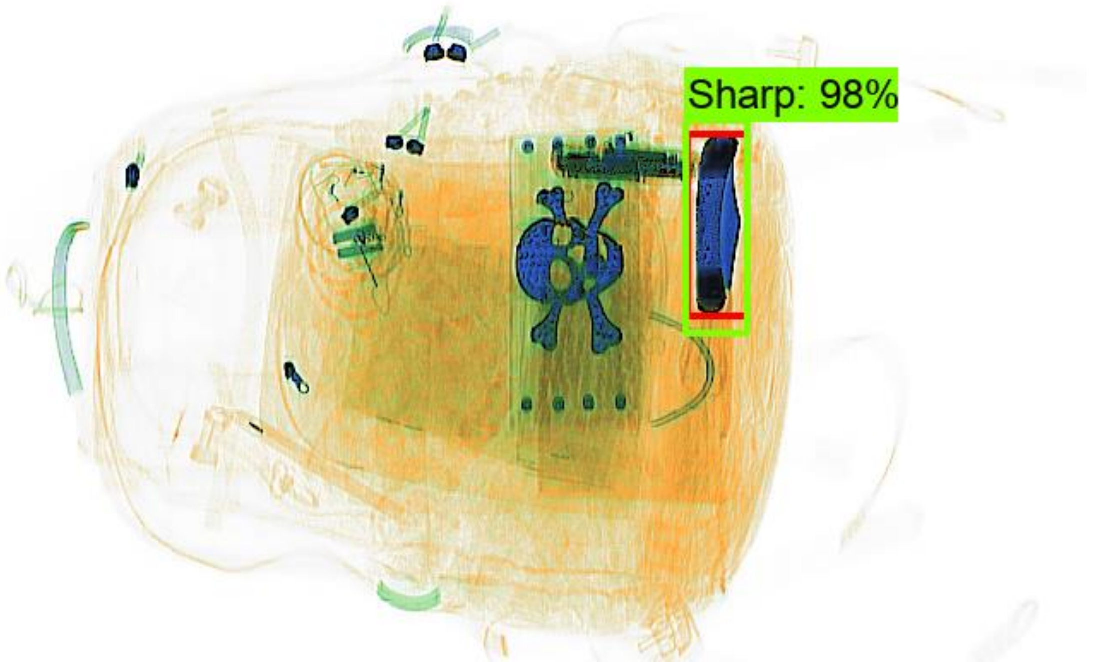


- Example detections with Faster-RCNN-ResNet152.
- Ground truth boxes are in red, while color denotes predicted class.
    - (a-b): Sharps
    - (c-d): Firearms
    - (e): Blunt
    - (f): Liquids

Images shown were obtained from Rapiscan-owned 620DV not in the TSA configuration

EXAMPLE DETECTION (MULTI-VIEW)



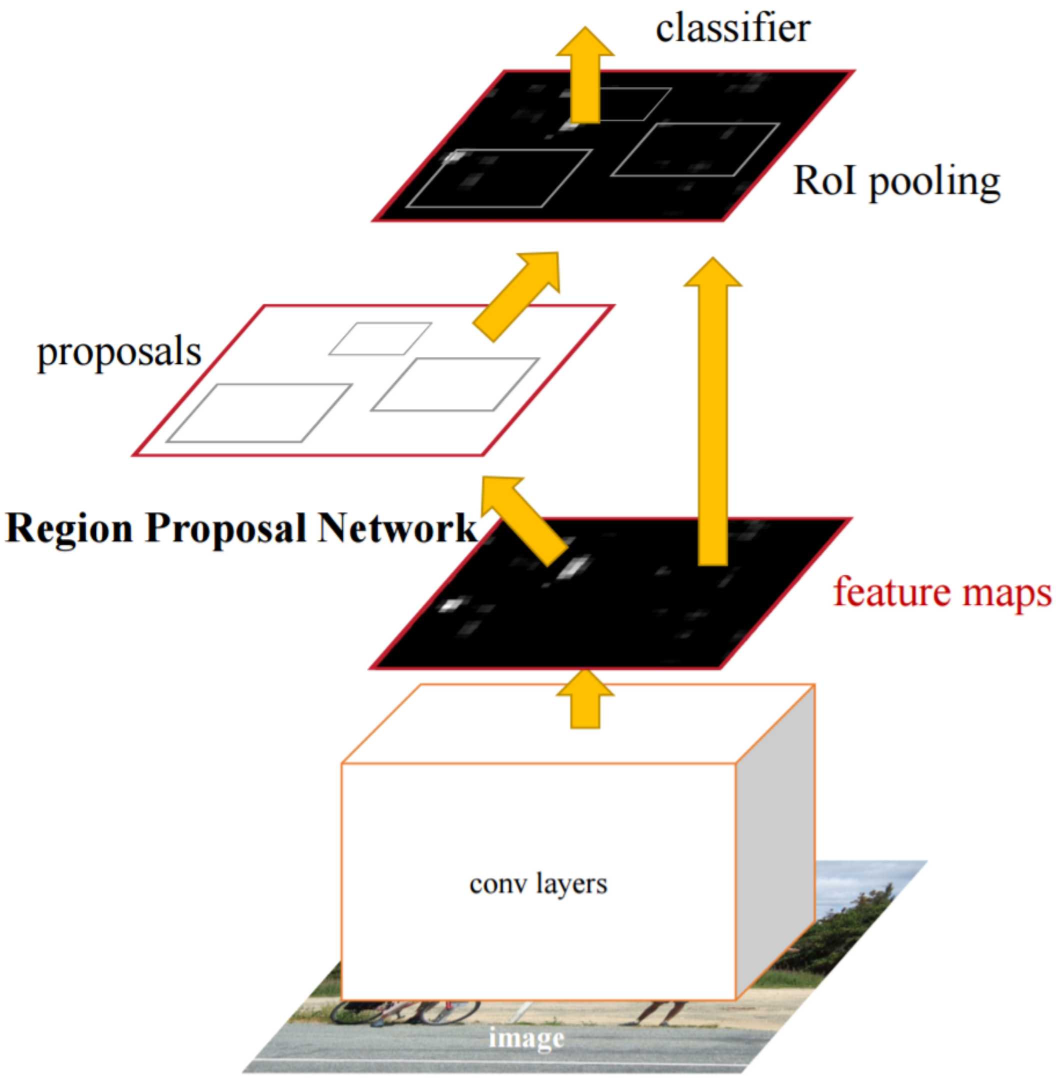


- Top and side views of a bag containing one knife. Detection is missed on the side view but detected on the top view.
- Demonstrates a benefit of multiple views.

Images shown were obtained from Rapiscan-owned 620DV not in the TSA configuration

BACKUP SLIDES

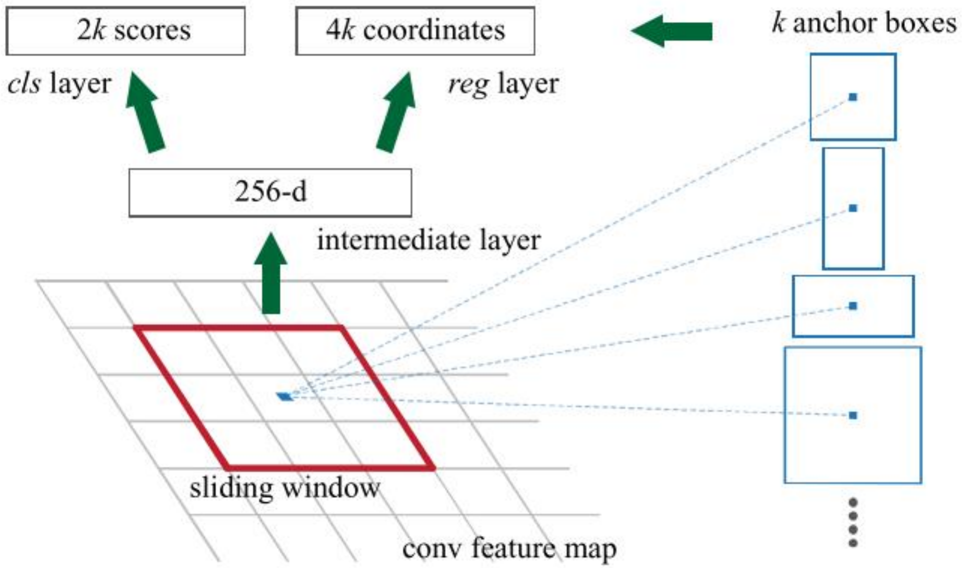
Faster R-CNN

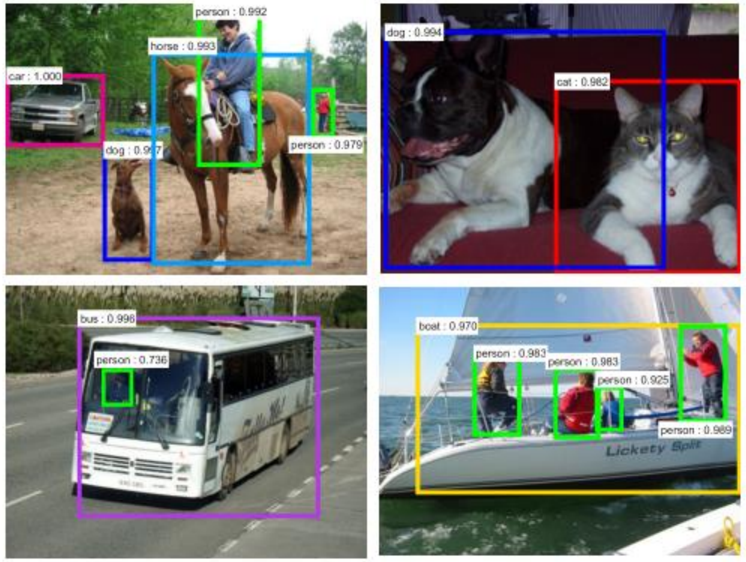


- Two-stage detection paradigm:
    1. Convolutional neural network (CNN) acts as a feature extractor, generating a set of feature maps
    2. **Stage 1**: Region Proposal Network (RPN) produces a set of region proposals from the feature maps
    3. Feature regions corresponding to the mostly likely proposals are cropped
    4. **Stage 2**: Proposed regions are refined and classified with a neural network

- Entire network can be trained jointly

Faster R-CNN

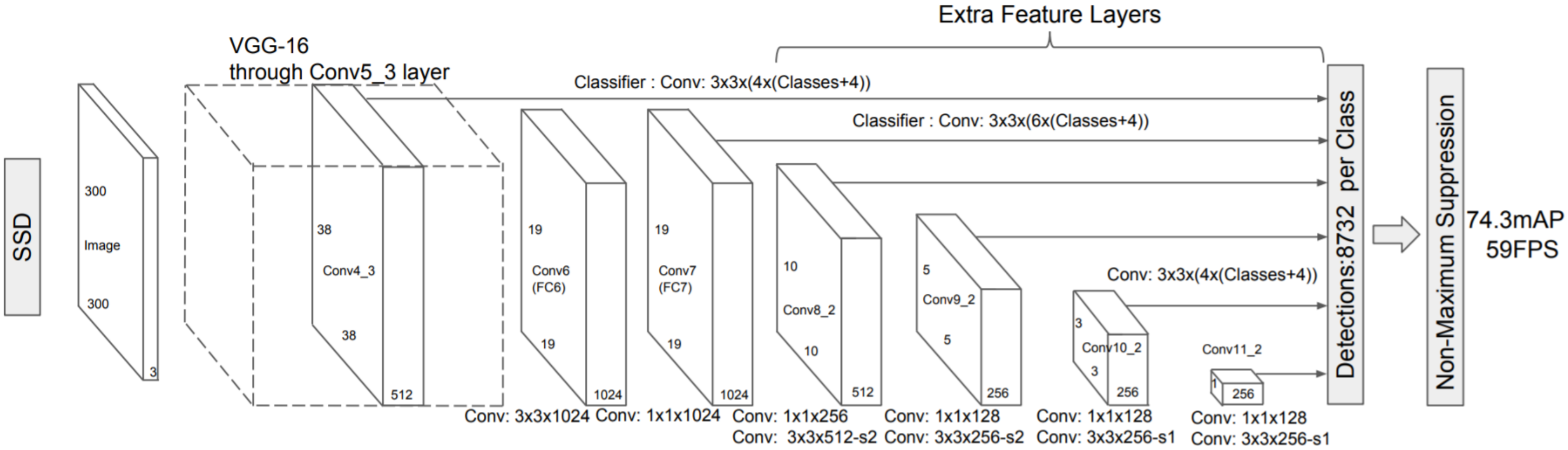




- Proposals are made relative to references: “anchor boxes”
- Diverse anchor box sizes help the model capture objects of many sizes

Single-Shot MultiBox Detector (SSD)

Rapiscan systems
An OSI Systems Company



- Single-stage detection paradigm:
  - Classifications and bounding box prediction are performed once
  - Different scales are capture at different layers of the network