Adaptive ATR Model Training

Marc Vaillant
MIT Lincoln Laboratory

ADSA Presentation
16 May 2018

This material is based upon work supported by the Department of Homeland Security Science and Technology Directorate under Air Force Contract No. FA8721-05-C-0002. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of Homeland Security.

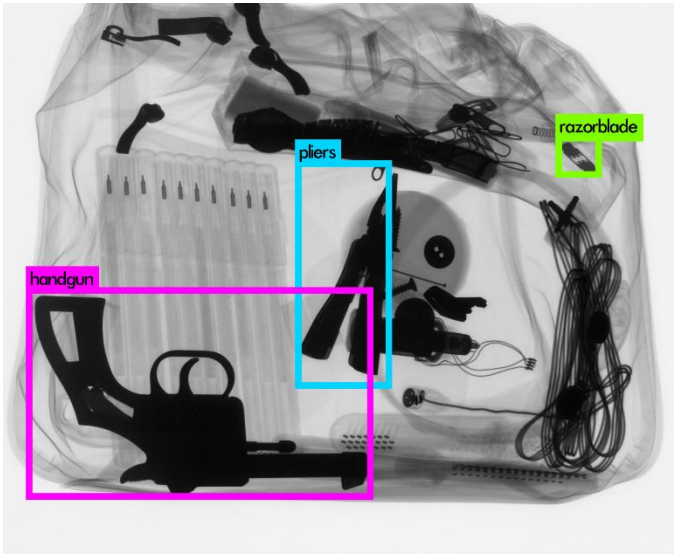
So What, Who Cares?

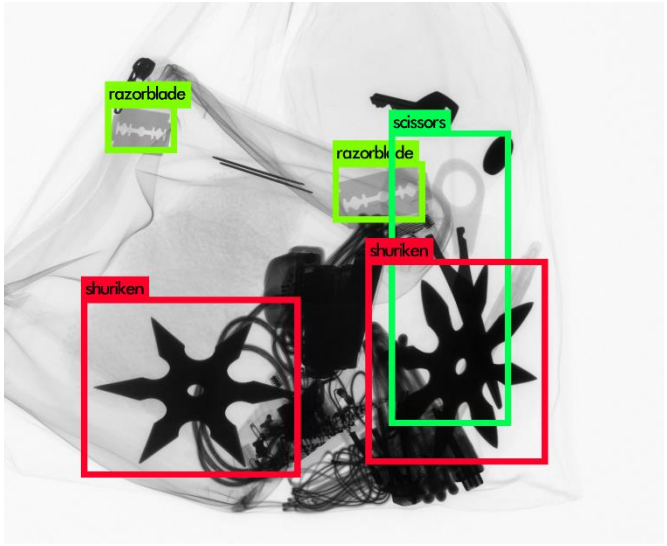
ATR Challenges

- Accurate discrimination difficult when
  - New threat types emerge as security concerns
  - Clutter sources and complexity evolve
  - Training data limited
  - Systems need a mechanism to adapt post deploy

Research Questions

1. Can pre-trained detectors be fine tuned to detect new threats and avoid new clutter sources?
2. How many training samples are needed to get reasonable performance?
3. How detailed do annotations need to be for training sets? How far can we get without annotating bounding boxes?





GDXray open x-ray dataset. Not SSI.

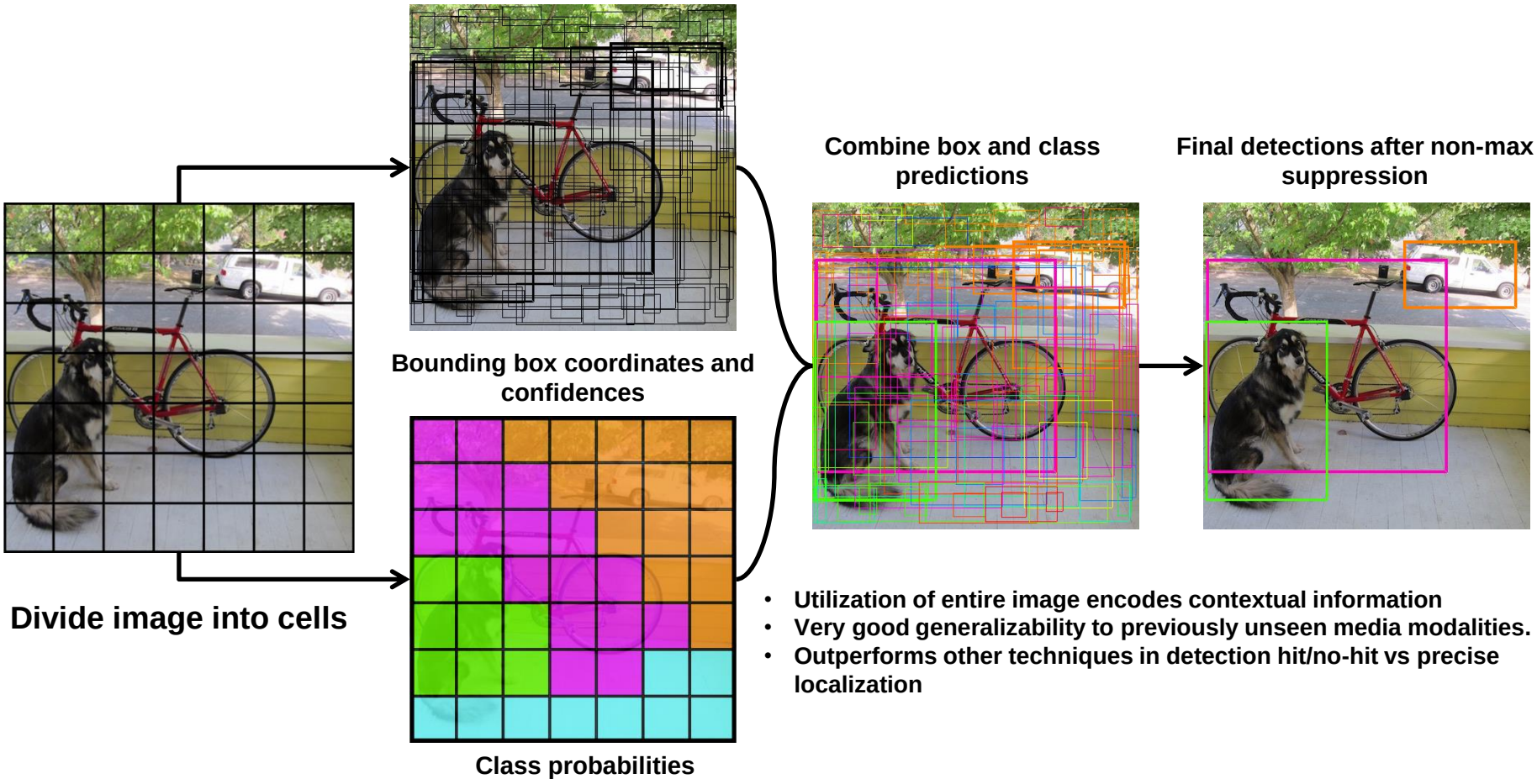
Datasets

- **GDXray, 2D Xray: Universidad Catholica de Chile, Santiago (Open dataset, NOT SSI)**
  - **Object threats (Handguns, Throwing stars (Shuriken), razorblades, scissors, pliers, gun parts, knives)**

- **Open Threat Assessment Platform (OTAP) (TSA, DHS S&T, Sandia and industry partners)**
  - **AT material threats , CT (upcoming collect at Tyndall Air Force Base)**
- **ALERT CT data**
  - **TO4 CT ATR and Segmentation datasets**

YOLO Approach (Joseph Redmon et. al.)

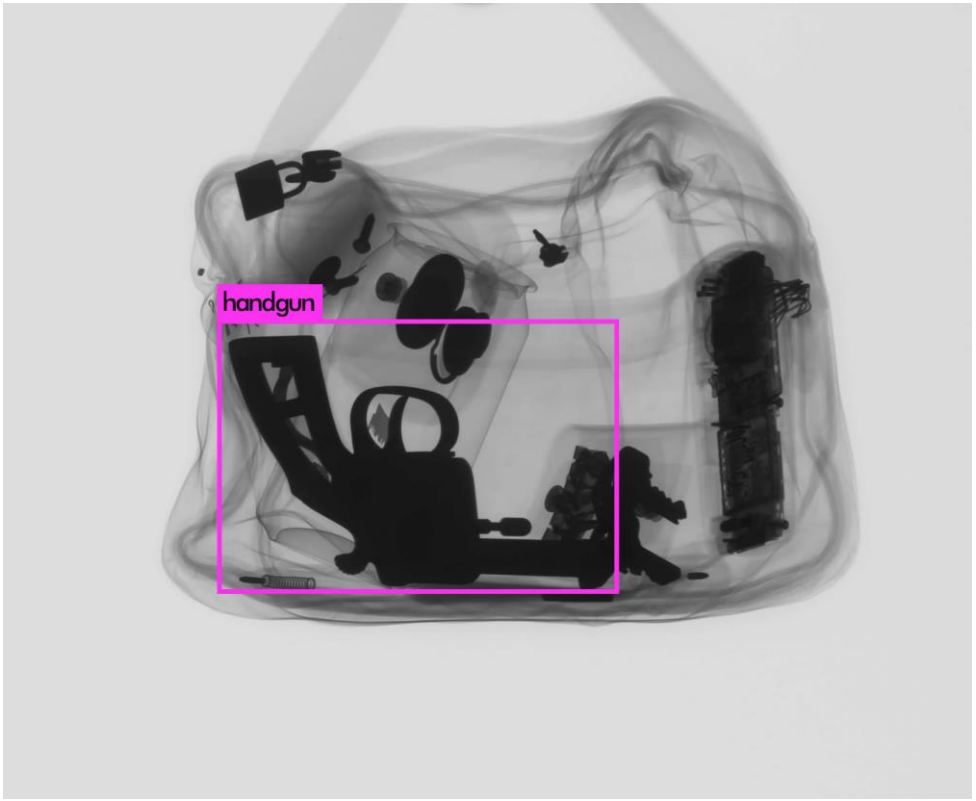
Predict bounding box coordinates (regression) and object categories (classification) directly from the image pixels.



- Utilization of entire image encodes contextual information
- Very good generalizability to previously unseen media modalities.
- Outperforms other techniques in detection hit/no-hit vs precise localization

Initial Detection Results

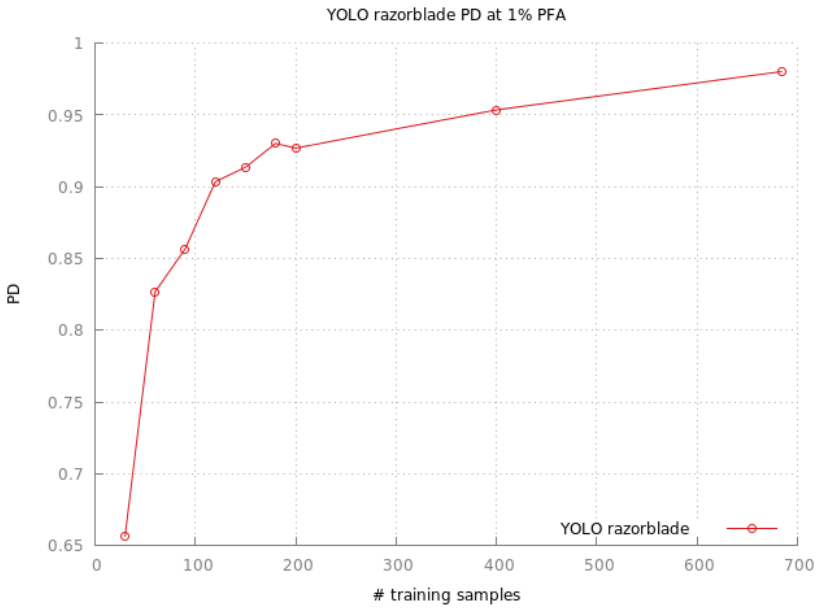
- **ImageNet/VOC pretrained YOLO V2 fine tuned on 300 Training images (207 Handguns, 180 Shuriken, 157 Razorblades, 47 knives, 23 scissors, 25 pliers).**
- **100 Test images.**

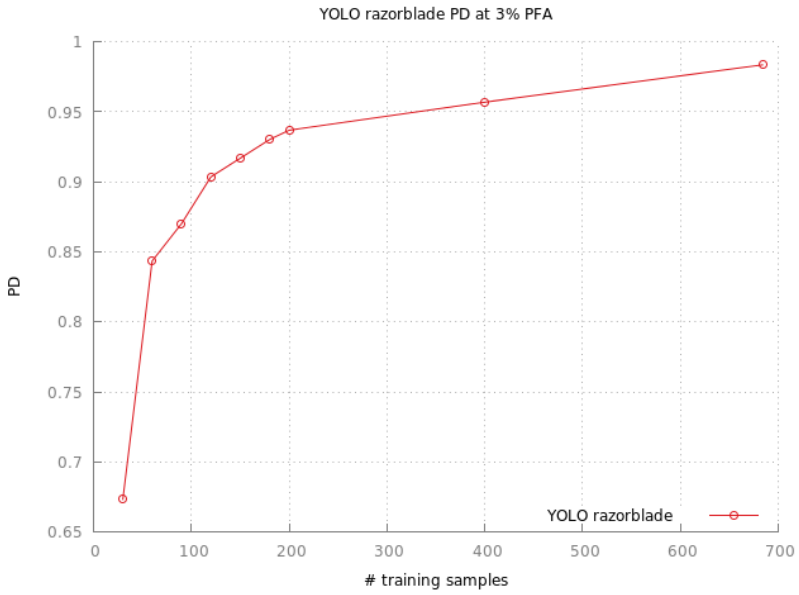


GDXray open x-ray dataset. Not SSI.

Small Sample Performance

- **Training samples ranged from 30 samples to approximately 700 samples, performing 3 fold cross validation. Test set of 200 samples, half positive half negative samples.**
- **Reasonable performance with just 60 samples, Diminishing returns around 200-400 samples.**



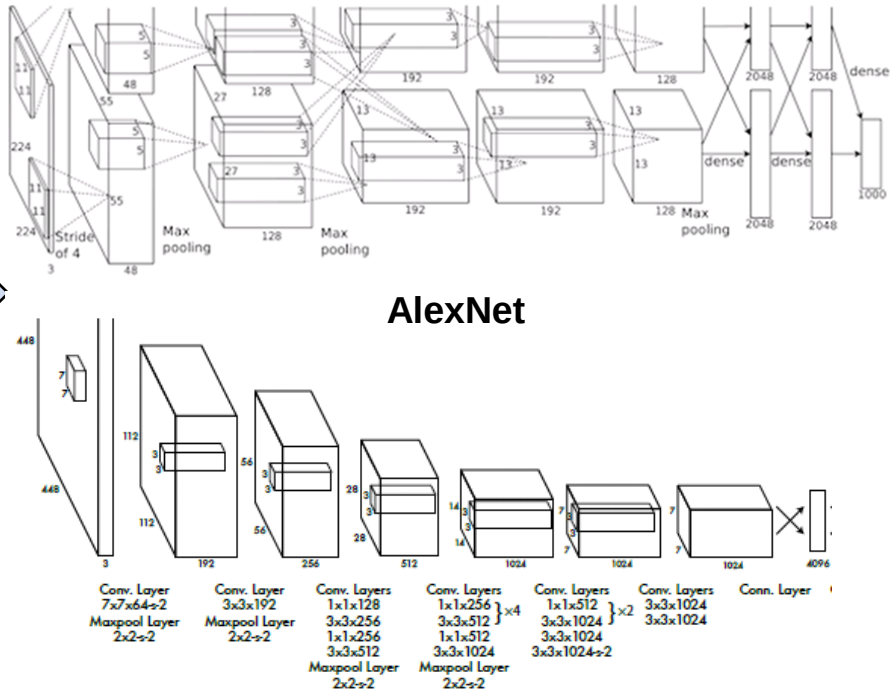


Holistic Approach "What Threats Are Contained in the Image?"

- ConOps may care more about whether or not image contains threats, rather than precise location of where they appear in the image.
- Can we train a CNN to determine, from an entire image input, which threats it contains? Only requires content labels vs bounding box annotations for detection

Full Image Input

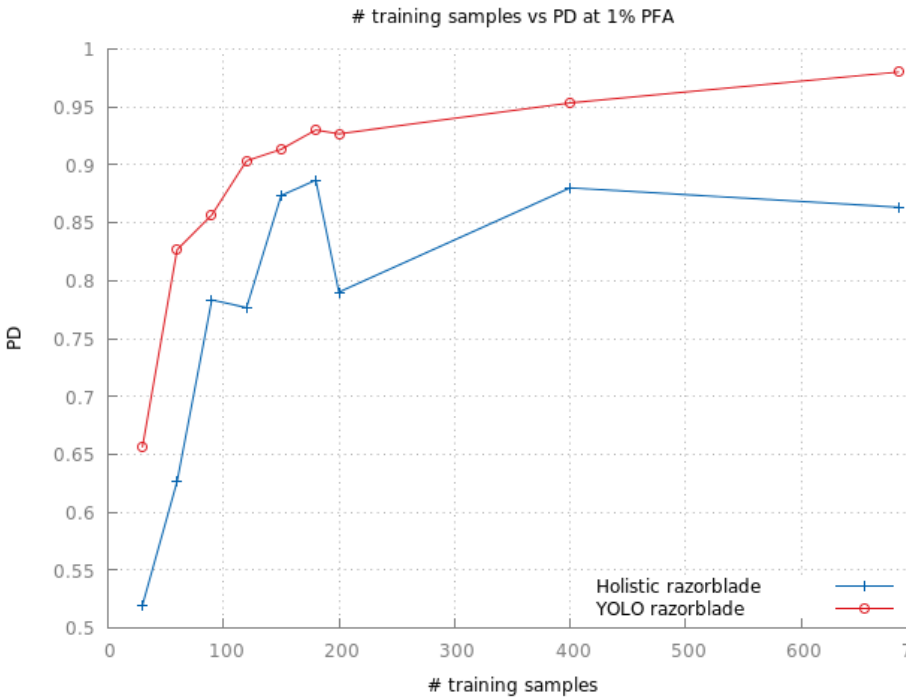
GDXray open x-ray dataset. Not SSI.

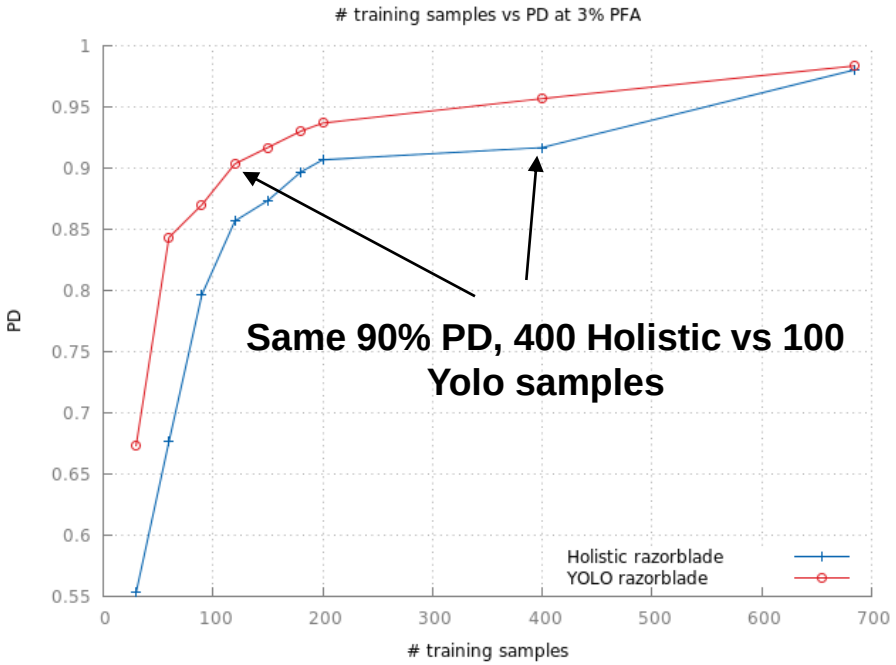


Handgun: NO
Shuriken: YES
Razorblade: YES
Knife: NO
Scissors: YES
Pliers: NO

PD at 1% and 3% PFA

- **Performance at a deployable operating point (E.g. 1% or 3% PFA) shows that YOLO out performs Holistic by about 5-10%.**



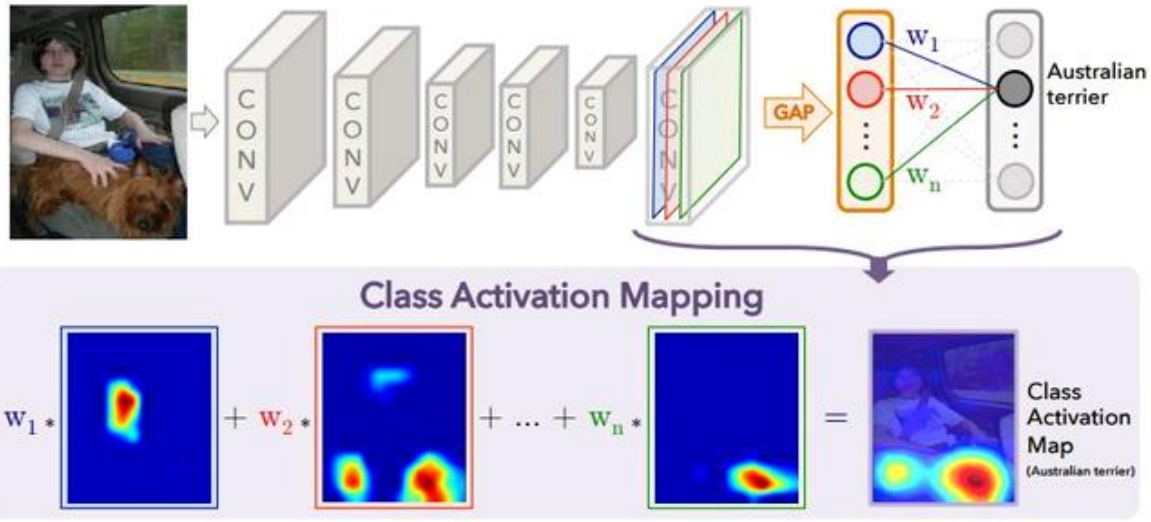


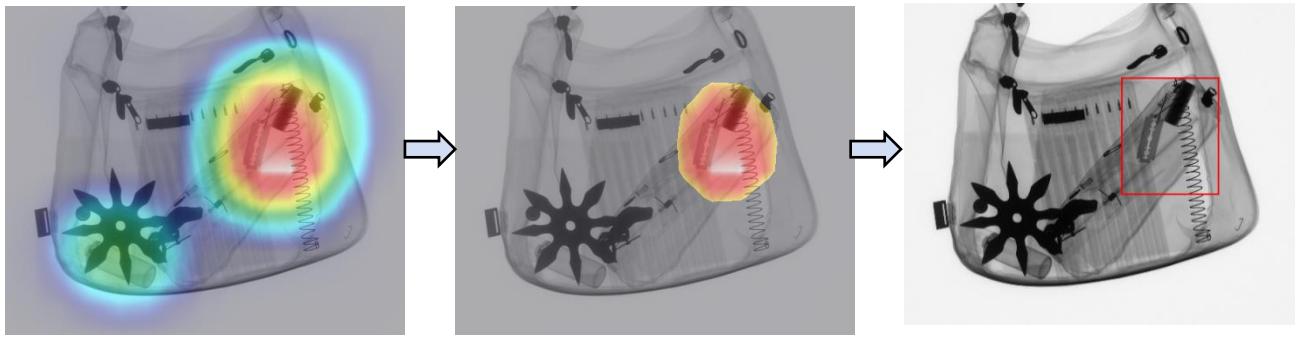
Yolo better, but requires bounding box truth vs simple content label for holistic

Localization of threats in Holistic Network

- **Identify where networks have directed their attention, and use that information to localize the classified object.**

Class Activation Mapping (CAM)
Zhou et. al



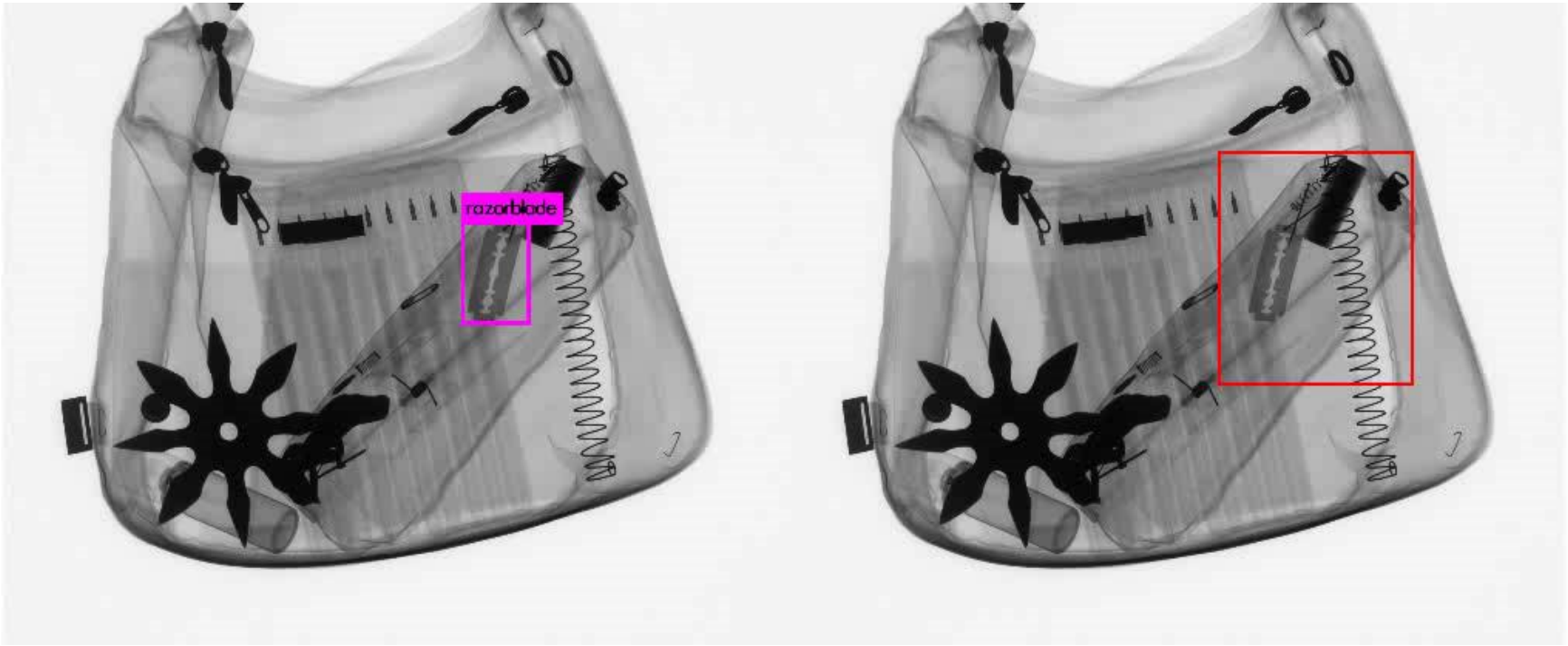


Localization surprisingly accurate

GDXray open x-ray dataset. Not SSI.

Class Activations Results

- **But, not as good as YOLO**



Video shows detection comparison for YOLO vs Holistic at approximately 2% PFA. 100 positive + 100 negative samples.

GDXray open x-ray dataset. Not SSI.

Results Summary

- **Able to achieve 90-95% PD with fewer than 500 training samples.**
- **With respect to most metrics, YOLO outperformed Holistic.**
- **Reasonable performance of holistic results, and localization via post processing show utility of “weakly” (no bounding boxes) labelled training data.**
- **Results suggest an equivalence in performance between fewer bounding box annotated samples and a larger number of simple content based training samples.**

Current and Future Work

- **Currently working toward extending Deep Neural Network study to CT and OTAP material threat AT data**
- **DNNs that utilize projections of 3D data perform quite well on 3D object recognition tasks, including MMW ATR**
  - **tractable in terms of compute requirements and labelling requirements.**
- **Investigating Multi-view CNN (MVCNN) approach (and variants), as well as YOLO extensions to the 3D domain**