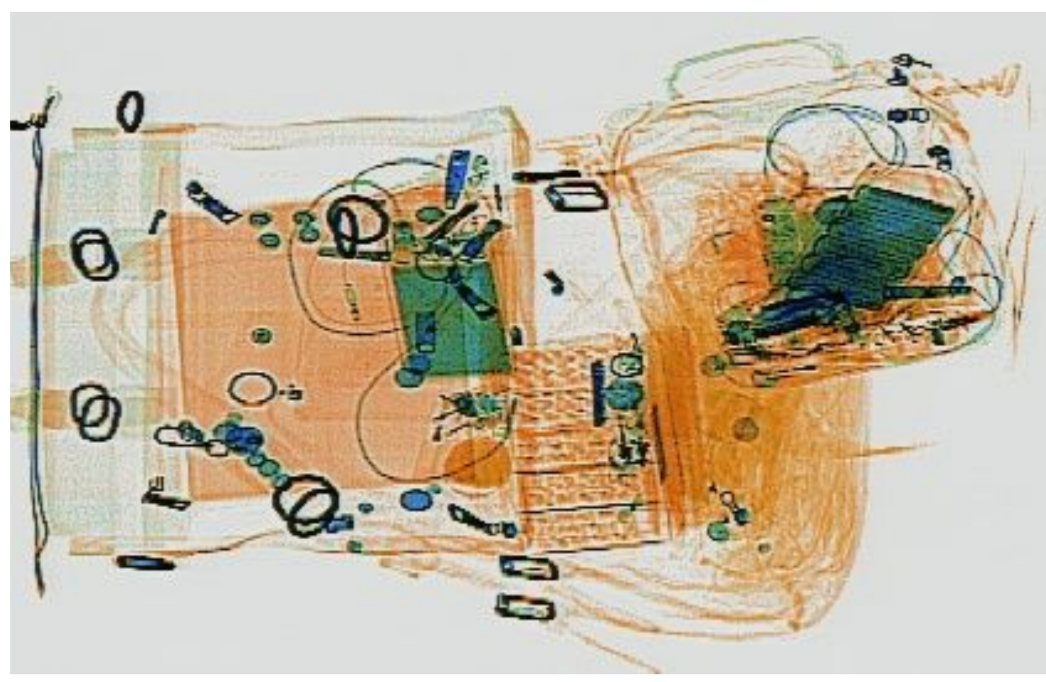
Dangerous Object Detection In X-Ray Images

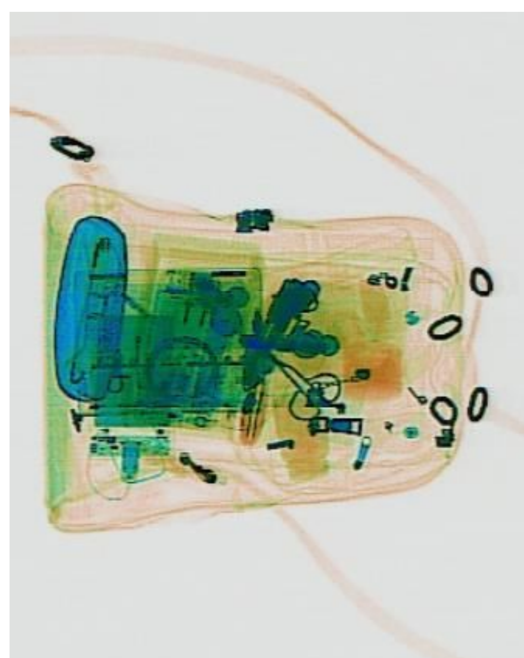
Edgar Suritis, Rashik Azad, and James Njoroge

COLGATE UNIVERSITY

Introduction

- **The Problem:** Airport and public transit security relies heavily on x-ray baggage screening, but accurately spotting dangerous items (knives, guns, etc.) is difficult. Some of the challenges involved include identifying items which are overlapping, at odd angles, and lacking color/texture in x-rays. Our aim is to design more accurate and faster x-ray image detection systems to improve public safety.
- **Our Question:** Can we fine tune a pre-trained object detection or vision model meant for regular images, to detect dangerous items in x-ray scans of baggage?
- **Why It Matters:** If classical object detection models show high performance when adapted to classifying x-ray images, future research and improvements in these models should also translate to x-ray image classification. Improved automated detection means safer public spaces and potentially smoother, faster security lines.





Background

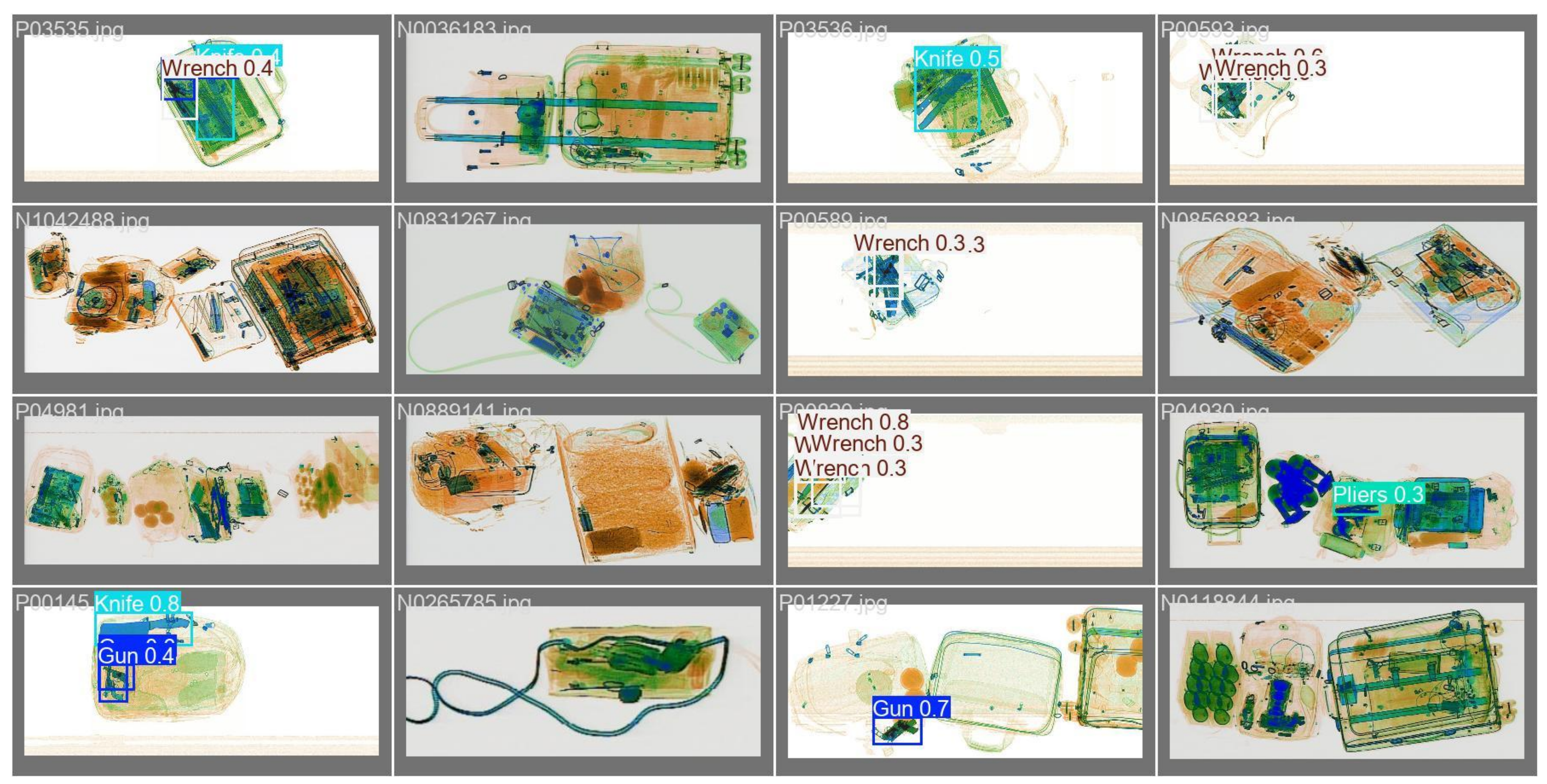
- **X-Ray Challenges:** Standard computer vision models often struggle with x-ray images due to:
  - Low contrast and minimal texture.
  - Objects frequently overlapping and blending.
  - Variations in item appearance and orientation.
  - Varying backgrounds making it difficult for models to learn representations of dangerous items.
- **Learning from Others:**
  - **YOLO-T (Wang et al., 2022):** Modified YOLO using Transformers for x-ray detection, aiming to improve speed and accuracy on hidden items.
  - **YOLO (Redmon et al.):** A fast, effective general object detector, but needs adaptation for the specific challenges of x-ray data.
  - **LightRay (Ren et al., 2022):** Built on a MobileNetV3 backbone with hard-swish nonlinearity, enabling real-time multi-class prohibited-item detection in x-ray scans with a minimal compute footprint.

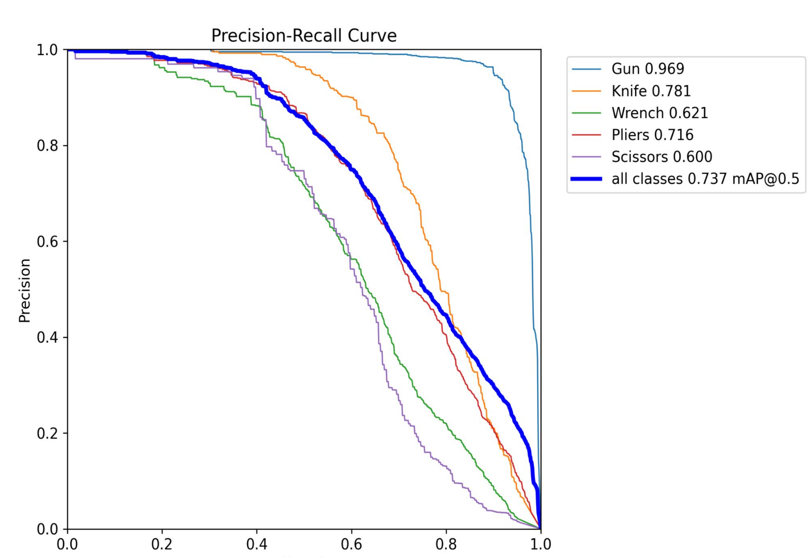
Experimental setup

- **Dataset:** SIXray (Publicly available x-ray baggage scan images)
  - 16,000 image subset was used, half containing dangerous items
  - 80/20 Train/Test split
  - Preprocessing: Size normalization to 416x416
- **Models:**
  - **Safe/Dangerous Classification:** ResNet-50
  - **Multi-Class Classification:** Inception-V3, ViT-b16
  - **Bounding Boxes:** LightRay, YOLO-v8 and YOLO-v11
- **Task:** Supervised Object Detection / Classification (training with image labels and bounding boxes).
- **Evaluation Metrics:**
  - Primary: F1-Score, mAP, Accuracy, Precision, Recall
  - Qualitative analysis (visualizing detections)
    - Bounding box mapping
    - Confusion matrices

Results

YOLO-v11: 0.737 mAP-50



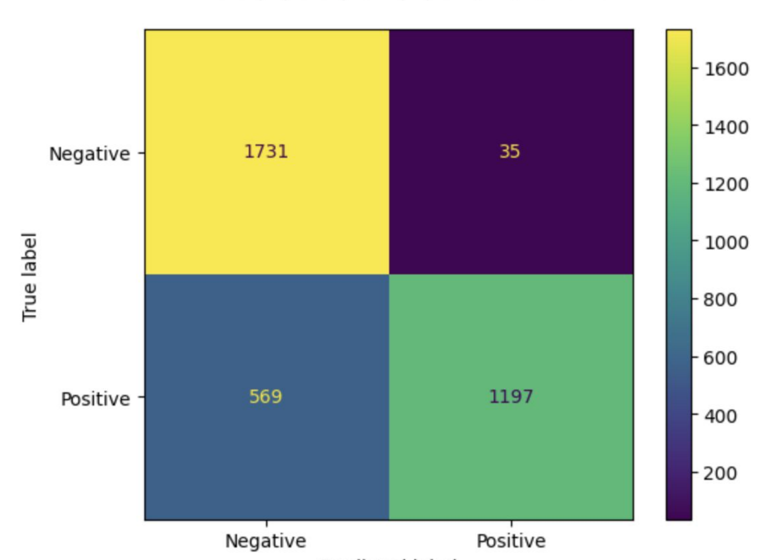


ViT-b16:

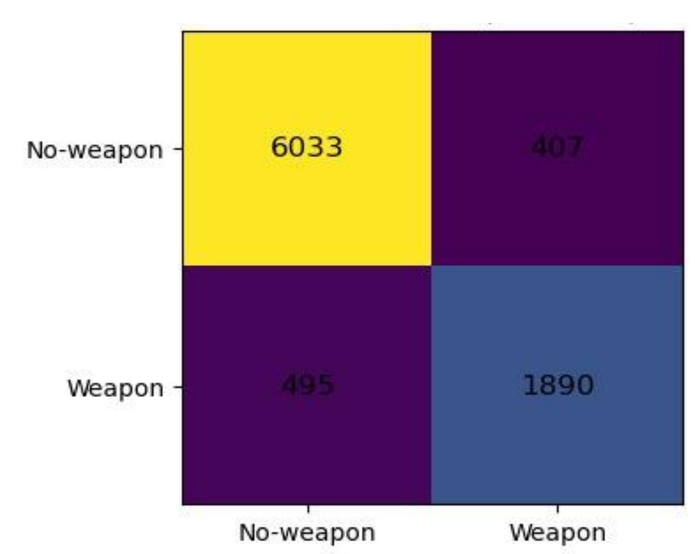
Overall accuracy: 0.8978

	precision	recall	f1-score
knife	0.98	0.84	0.90
gun	0.98	0.95	0.97
wrench	0.65	0.55	0.59
pliers	0.76	0.84	0.79
scissors	0.73	0.63	0.68

ResNet-50: 0.80 F1



ViT-b16: 0.84 F1



Summary

- **Summary:** Our fine-tuned transformer (ViT-b16) achieved a **Micro F1 of 0.84** on the SIXray x-ray dataset, while ResNet-50 achieved an **F1 score of 0.80**. The YOLO-v11 fine tuned model achieved an **F1 score of 0.70.**
- **Interpretation**
  - Our fine-tuned ViT's high accuracy (90%) and F1 (0.84) show that it captures the signatures of all five weapon classes well, demonstrating its effectiveness as a baggage-screening aid. While a model with this level of accuracy shouldn't be left unattended, it is useful in aiding humans at this task.
- **Limitations:**
  - Performance of our models is likely to be dataset-specific to SIXray. If these models were implemented in real world baggage scanning, they would have to contend with less than ideal conditions across varying x-ray machine models / configurations.
  - It's difficult to know why these models make the classification decisions they do. The discovery of specific shortcomings or misclassification patterns by bad actors is a significant risk of over-reliance on this technology in such a sensitive application.
  - ViTs often demand careful tuning of learning rates, weight decay, and dropout; poorly chosen settings can lead to unstable training or suboptimal convergence on domain-specific data.

Conclusion & Future Directions

- **Conclusion:** Transformer architectures like ViT-b16, when fine-tuned, can effectively overcome challenges of low contrast, object overlap, and varied orientation, yielding high performance for automated dangerous object detection in x-ray baggage scans.
- **Future Directions:**
  - Incorporate meta-learning approaches to adapt to new weapon classes or emerging threats with only a handful of annotated examples, without full retraining.
  - Testing and adjusting the architecture of all of the models used to prevent against SIXray dataset specificity.

References

Wang, M., Yang, B., Wang, X., Yang, C., Xu, J., Mu, B., Xiong, K., & Li, Y. (2022). YOLO-T: Multitarget Intelligent Recognition Method for X-ray Images Based on the YOLO and Transformer Models. *Applied Sciences, 12, 11848.*

Joseph Redmon, Santosh Divvala, Ross Girshick, & Ali Farhadi. (2016). You Only Look Once: Unified, Real-Time Object Detection.

Yu Ren, Haigang Zhang, Hongxing Sun, Guanglin Ma, Jin Ren, & Jinfeng Yang (2022). LightRay: Lightweight network for prohibited items detection in X-ray images during security inspection. *Computers and Electrical Engineering, 103, 108283.*