



# DHS SCIENCE AND TECHNOLOGY

## Black-Box Detection Explainability Utilizing RISE (Randomized Input Sampling for Explanation of Black-Box Models)

11 May 2021

Barry Masters

Tyler Karns

DT&E Spectroscopy Branch Manager, TSL

Jr Technician, Battelle



**Homeland  
Security**

Science and Technology

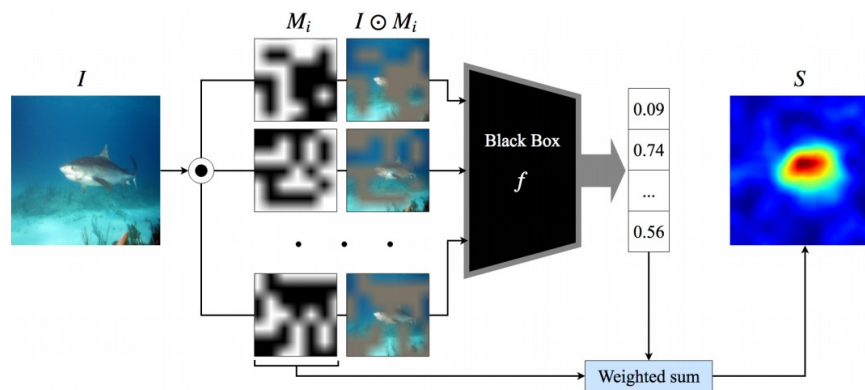
Transportation Security Laboratory  
Science and Technology Directorate  
TSLICC0xxxx

# So What? Who Cares?

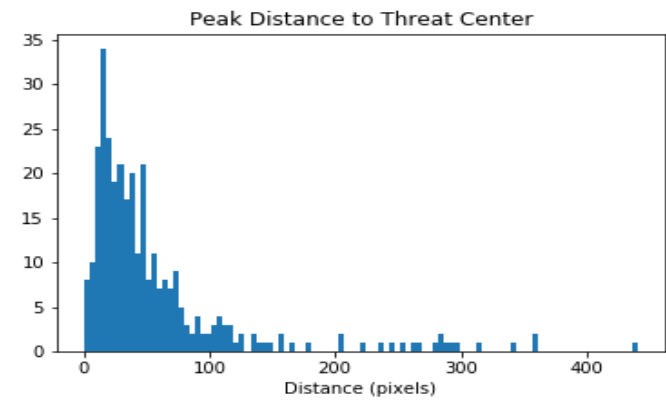
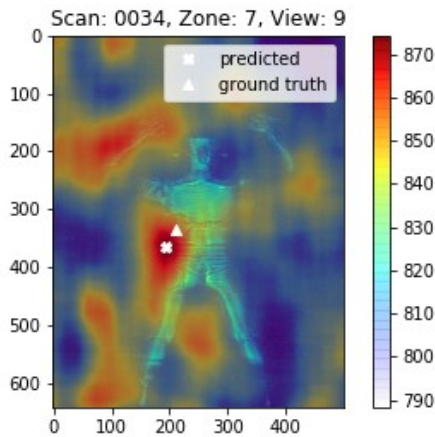
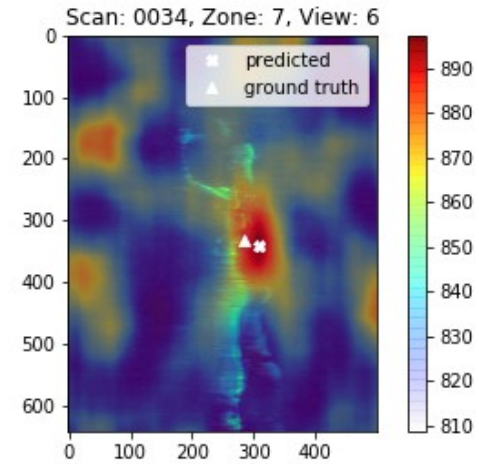
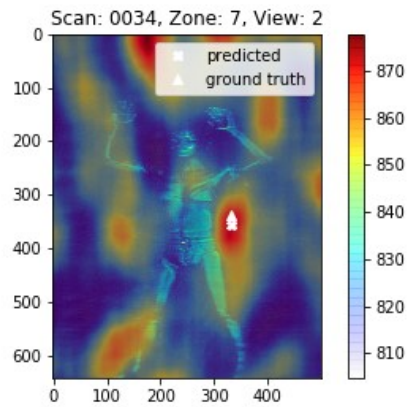
- How do we develop tools to identify biases in detection algorithms?  
How can we be certain that an algorithm is focusing on threat areas in a scan?
  - ATRs are typically black boxes to the testers
- RISE allows us to determine algorithm attention at the pixel level of a scan
  - Works by generating random masks on the input images and leverages the change in detection for each mask
- Preliminary results show we have been able to score algorithms based on whether an ATR's attention in a scan is on features of a threat when it makes a correct detection
- TRL

# RISE

- RISE: Randomized Inter Sampling for Explanation of Black-Box Models
- RISE applies randomized masks to input images and measures the change in prediction confidence
- The prediction confidences are used to weight the masks, and the masks for a given input are pixel-wise summed together to generate attention maps at a pixel level



# Results





# Homeland Security

---

Science and Technology

**DIVERSE PERSPECTIVES + SHARED GOALS = POWERFUL SOLUTIONS**

# Contacts

## **Rob Klueg**

DT&E Division Manager  
Transportation Security Laboratory  
Science and Technology Directorate  
Department of Homeland Security

## **Peter Kenny**

DT&E AIT Test Lead  
Transportation Security Laboratory  
Science and Technology Directorate  
Department of Homeland Security

## **Barry Masters**

DT&E Spectroscopy Branch Manager  
Transportation Security Laboratory  
Science and Technology Directorate  
Department of Homeland Security

## **John Tatarowicz**

Data Scientist III  
Transportation Security Laboratory  
Science and Technology Directorate  
Department of Homeland Security