



DHS SCIENCE AND TECHNOLOGY

Synthetic Data for Efficient Training and Testing

11/5/2019



**Homeland
Security**

Science and Technology

Robert Klueg

Division Chief, Developmental Test and Evaluation
Office of National Labs /
Transportation Security Laboratory
Science and Technology Directorate

William Garret

Test Engineer
Office of Acquisition Program
Management
Transportation Security Administration

So What, Who Cares?

- In the near future, Advanced Imaging Technology (AIT) algorithm testing will require at least 10,000 unique images. AIT algorithm training may require an order of magnitude more image data.
- Synthetically generated image data will likely be necessary for supplying training and test data efficiently but how do we validate it?
 - Industry and Government need to have a discussion about validation of synthetic data – What does valid mean?
 - New expectations (?)
 - Require explainable AI
 - Capability to inject synthetic data into an OEM algorithm
 - More information required from OEM on system specifications
- Government/Industry/Academia development group needed
- Robert.Klueg@st.dhs.gov
- William.Garret@tsa.dhs.gov

Transportation Security Equipment (TSE) Critical Algorithm Training and Testing Concerns

- Advanced Imaging Technology (AIT) training and testing is time consuming and labor intensive
- Testing requirements for body zones, explosives and weapons types, several BMI categories, etc., results in the need of approximately 10,000 unique images
- Hardware changes for the AIT systems result in having to reacquire all images
- Homemade Explosives list is growing and not all are safe to put on mock passengers (MPs)
- Development of physical simulant explosives increases the timeline for completion of training and testing
- Data requirements for training machine learning (ML) algorithms may be in the 100,000 image range
- Pool of MPs limited due to security and availability requirements

It is critical that solutions be developed for efficient and reliable training and testing of TSE

Transportation Security Equipment (TSE) Critical Algorithm Training and Testing Concerns

- Synthetic data generation is a possible solution for the Homeland Security Enterprise's need for efficient and effective technology development.
- A data validation methodology is needed to qualify the image generator for use. How do we do this?
 - Equivalence of Image Quality Metrics
 - Fool a machine learning algorithm trained only on real data
 - Fool experienced human operators
- A validated synthetic data generator must be configurable for each OEM system
- We need to safeguard against systematic biases in the data (e.g. can't train to unreliable features)

A Government/Industry/Academia partnership is key to enabling the optimization of TSE using synthetic data

Image Manipulation and Image Quality Tools

The 5 ANSI N42.59 IQ metrics



“Lateral Resolution”

“Reflectivity Mapping”

“Depth Resolution”

“Illumination Coverage”

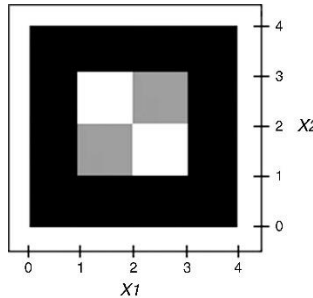
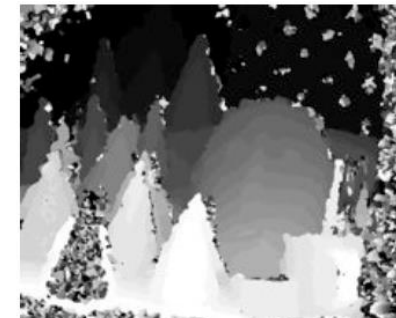
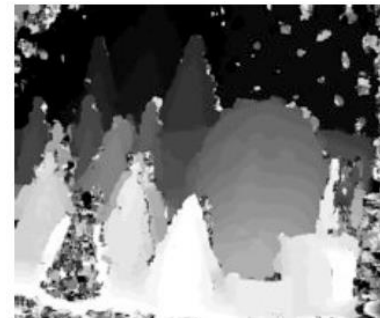


“Dynamic Range at Resolution”



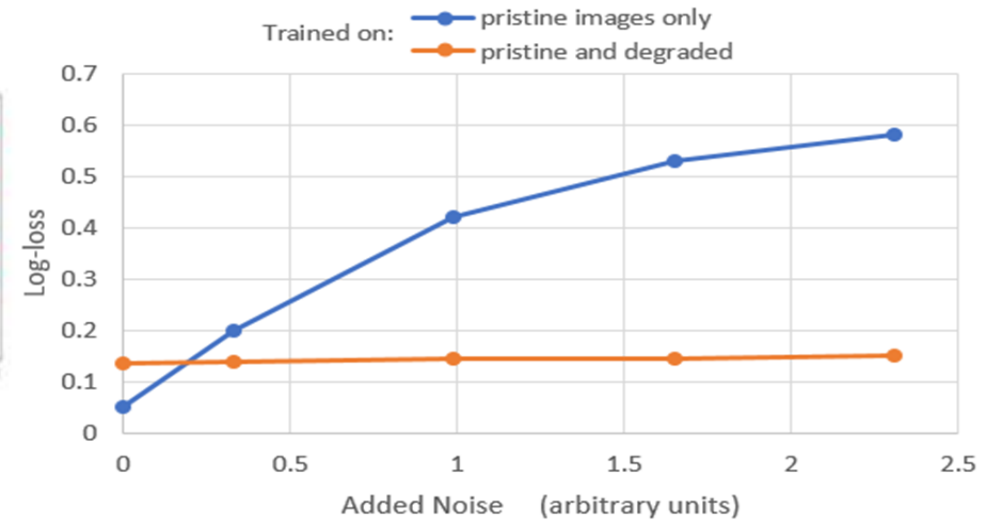
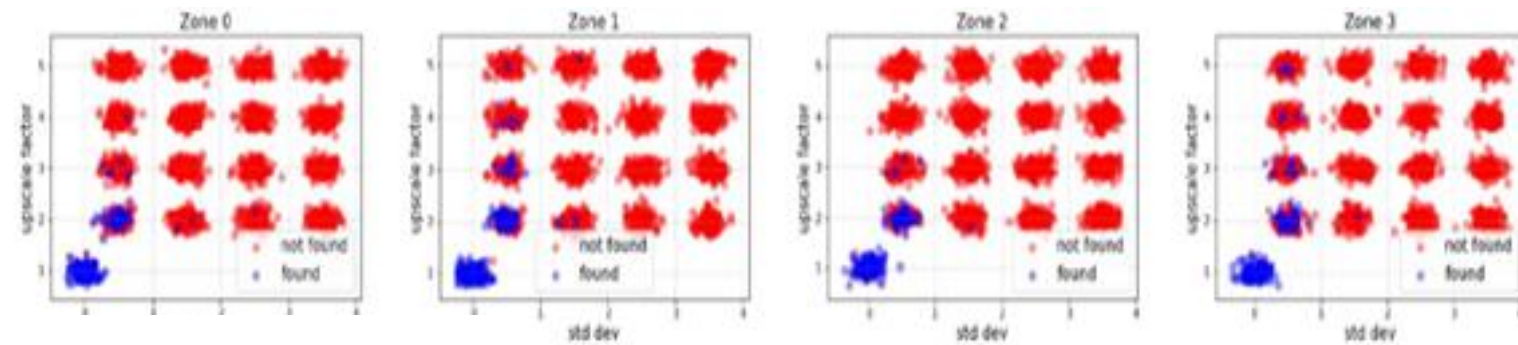
(with styro sandwich mod)

1



DIVERSE PERSPECTIVES + SHARED GOALS = POWERFUL SOLUTIONS

Image Quality in Practice / Compensating for Noise

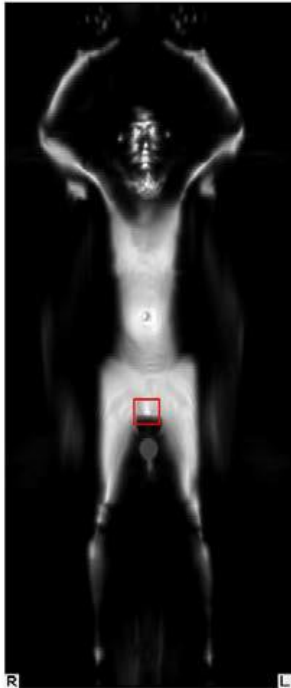


NIST

Left Fig: Detection performance as a function of body zone and image quality. Blue = threat found. Red = not found.

Right Fig: Performance of an ATR algorithm as a function of image noise as measured by its log-loss score (higher log loss means poorer detection). ATR algorithms can be made more robust against noisy images by including noisy images in the training set.

Strategic Advantage for Test and Evaluation, Own the Algorithms and Realize Synthetic Test



Manipulate the Image, Increase Test Efficacy, Control the Algorithms, Synthetic Test in Practice



NIST

- Use dithering, rotation and translation to create synthetic test
- When do the algorithms not care anymore
- Create thousands of instances, fig on right never existed



Homeland Security

Science and Technology

DIVERSE PERSPECTIVES + SHARED GOALS = POWERFUL SOLUTIONS