SAND2021-8466C

MLDL

Machine Learning and Deep Learning Conference 2021

Supporting DHS Non-Intrusive Inspection With ML

- Rob Forrest (8716), in collaboration with many others
- DHS S&T

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Big Picture

Our Task: Aid CBP's adjudication of data thereby supporting reduction of risk and the rapid and secure flow of commerce into the United States.

Overview

- We run a portfolio of projects using Non-Intrusive Inspection (NII) for Department of Homeland Security / Customs and Border Protection (DHS/CBP).
- **This talk will be an overview**: an introduction to the problems and how we think about and engineer solutions.
- Please reach out to contributors to find out more.

= MLDL 2021 Talks in this portfolio

Non-Intrusive Inspection (NII): Enables CBP to detect contraband (e.g., narcotics and weapons) and materials that pose potential nuclear and radiological threats.

Ports: Air, Sea, Land
Items: Trucks, Cars, Luggage, Mail, People
Technology: Radiation Detection, X-ray, CT, visual

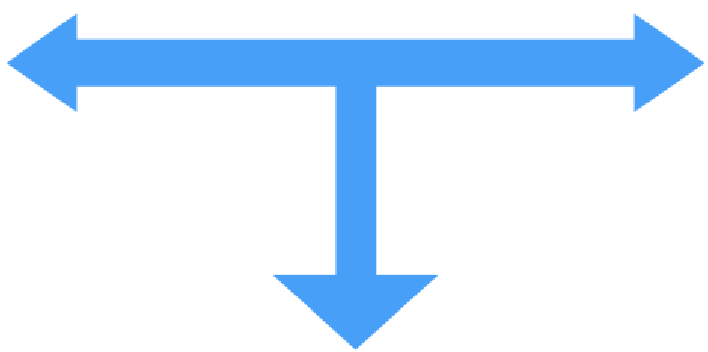
Here, we generally work specifically with X-ray images.

Our Problem

- Tons of data for CBP to adjudicate.
- An assessment by a CBP officer balances the availability of resources and need for the free flow of commerce with the security of the United States.

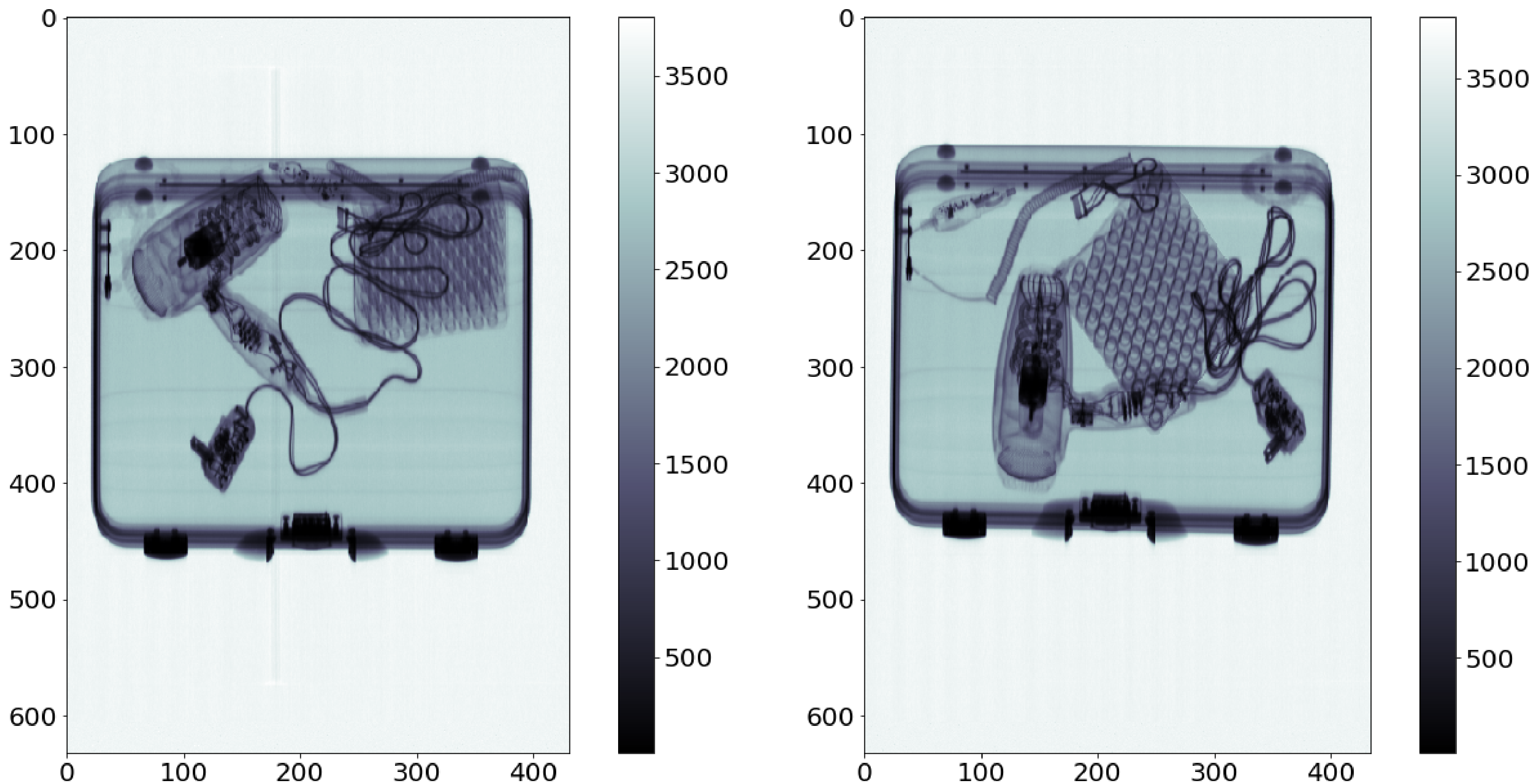
We view the realities of inspection as a real-world **risk tradeoff** between the time an officer has to adjudicate an image and the quality of the adjudication. Project metrics of success are a quantification of the improvement of the quality of adjudication of X-Ray images. We view this as an expansion of the breadth or depth of inspections.

Expanding the breadth of inspections means inspecting for a larger variety of items.



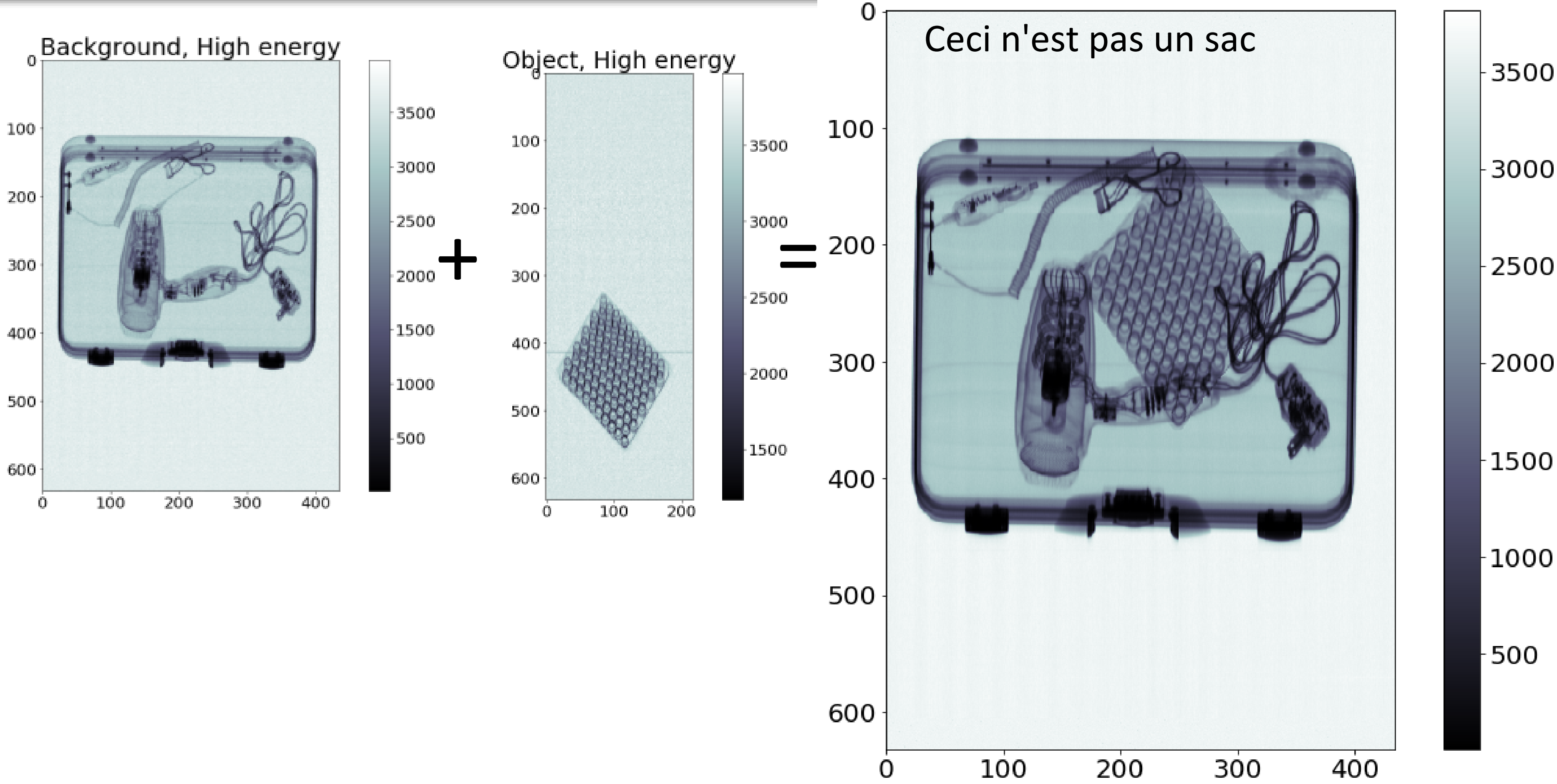
Expanding the depth means searching more extensively the whole image.

Theme: Capability Building with Synthetic Data



One of these bags does not exist

Theme: Capability Building with Synthetic Data



Overview: Three (3) Example Projects

Project (1/3):

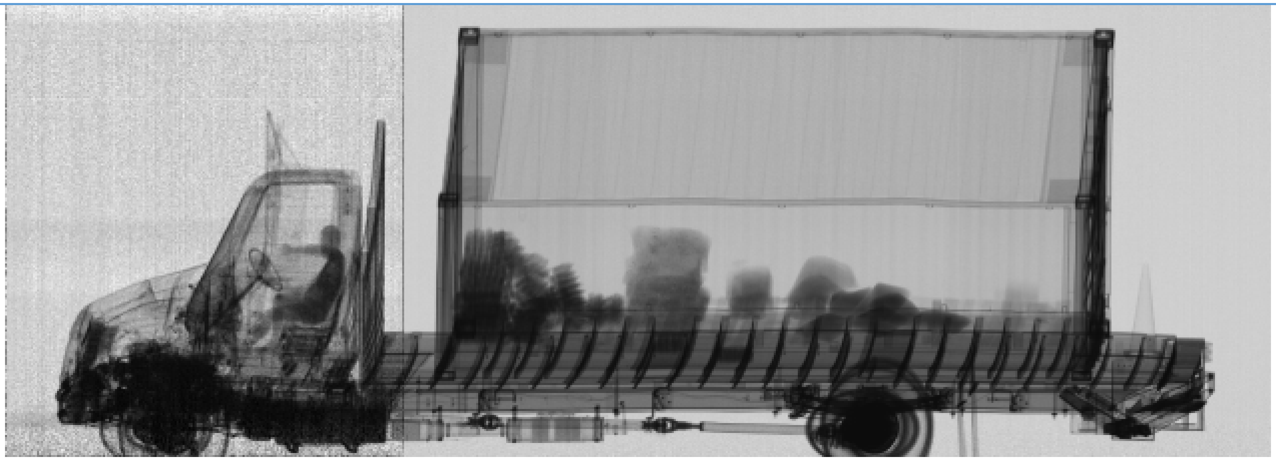
2D X-Ray of passenger bags imitating passenger packed luggage containing Ag/Bio items of interest.

Project (2/3) :

Scans of checked passenger bags.

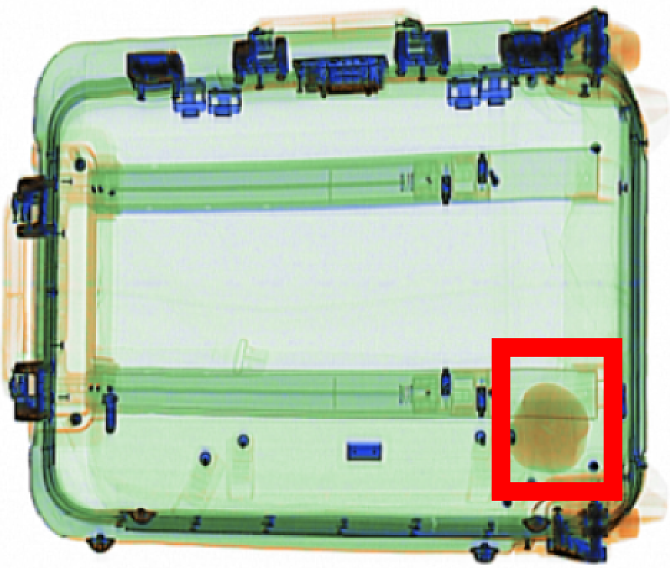
Project (3/3):

High energy scans of trucks entering through land borders



Project 1/3

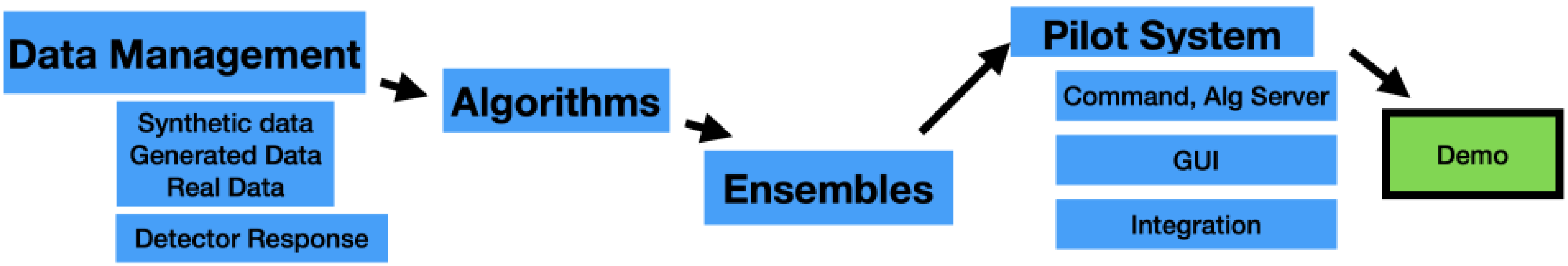
- 2D X-Ray images lab generated, imitating passenger packed luggage containing Ag/Bio items of interest.
- 8 Items to detect: Apples, oranges, bananas, potato, sausage, petri dishes, saline solution, vials.
- **Goal: Detect these items.**
- X-Ray provides Z effective and density information in both top and side view.
- Very well labeled and annotated data.
- Easy to generate data: Detector onsite. Lots of data takes a LONG time to generate.



Pitfalls: Not much variability in the data. Idealized data. SOC has many more complications. Manual data taking is onerous.

We have a **strategy**

- Don't just 'go': Think about big picture
- Initial framework (below)
- Can't possibly present everything here.



Focus on ML here, but first, some best practices:

How do you compare algorithms against one another?

Version your datasets. Have a project training and test set.

How do you measure performance?

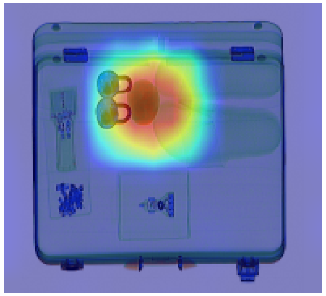
Establish project wide metrics. ONE relevant number to compare. (It's never this easy)

Remember errors (both systematic and statistical) and look at all training curves.

(1/3) Approaches

Try complimentary approaches to triangulate what works.

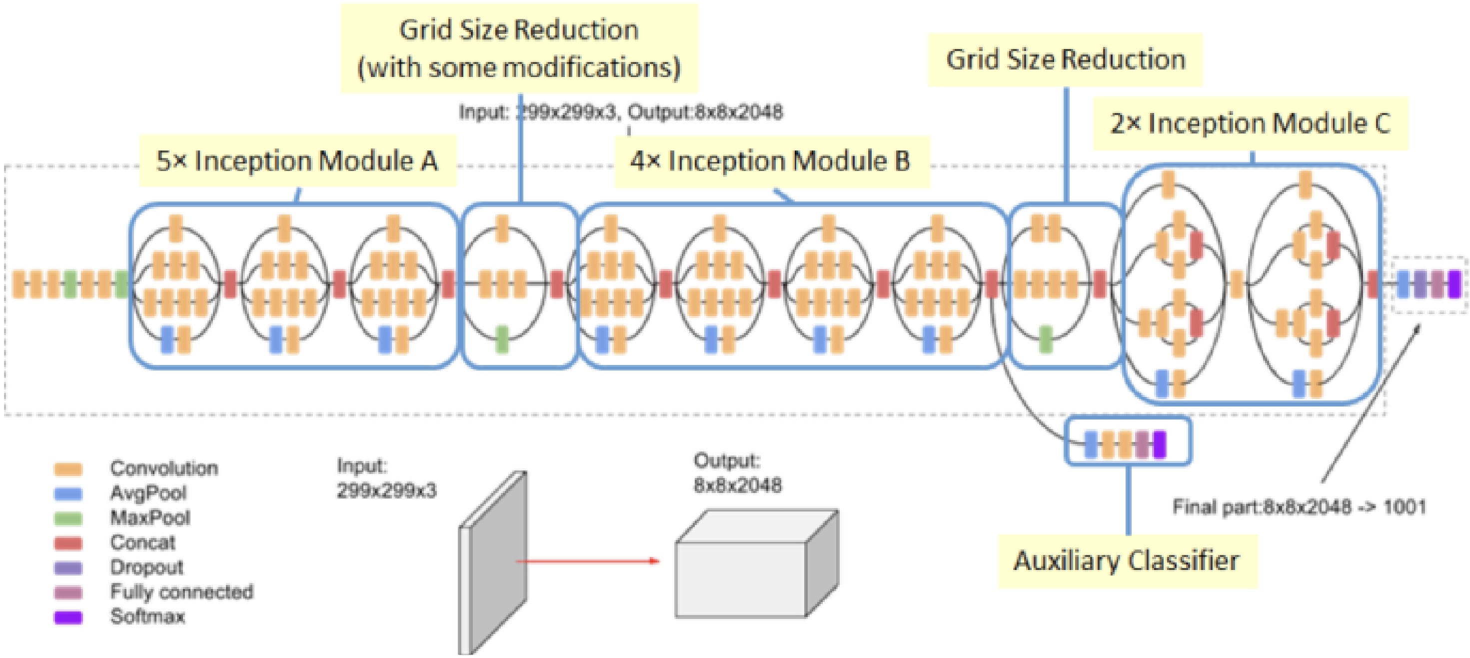
DL Convolutional NN Methods - Brute Force, popular and powerful.	Inception V3 (Eric Goodman)
Semi-Supervised Methods - Improve performance using unlabeled data.	Mean Teacher (Jenny Galasso)
Transformer Architectures Improve performance using attention.	DEtection Transformer (EJ Guillen)
Region Proposal Methods- Point to subsection of an image and classify.	Detectron2 (Dan Krofcheck)
Understanding, optimizing results - Thresholding, Visual Explainers, Ensembling	Optimal Thresholding (Goodman, Salaudeen), Clique Ensembles (Krofcheck, Blanchard)



Many others not listed, these are examples.

Inception V3 (Eric Goodman)

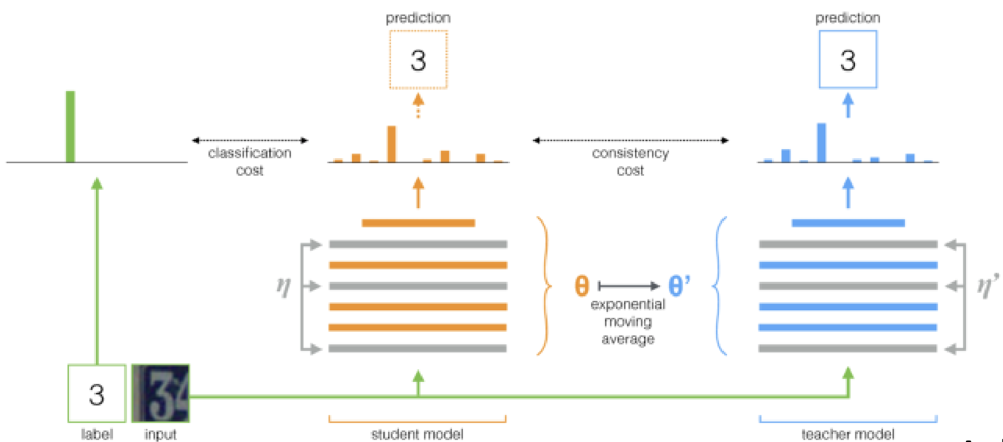
- Deep CNN – Many designs.
- Binary image classification
- Workhorse to answer many useful questions about data.



	apple	orange	banana	potato	sausage	petri	vials	contact solution
Average Precision	0.87	0.97	0.88	0.91	0.92	0.98	**1.00**	0.69

Semi-Supervised Learning (SSML) (Jenny Galasso, Tim Draelos)

Mean Teacher
Resnet-50



Baseline classifier is trained on labeled data only
SSML classifier is trained on labeled and unlabeled data

Labeled Data: 10% of packings
Unlabeled Data: 85% of packings
Validation Data: 5% of packings

- **Split Accuracies**
  - **Baseline: 86.3%**
  - **SSML: 93.9%**

Baseline

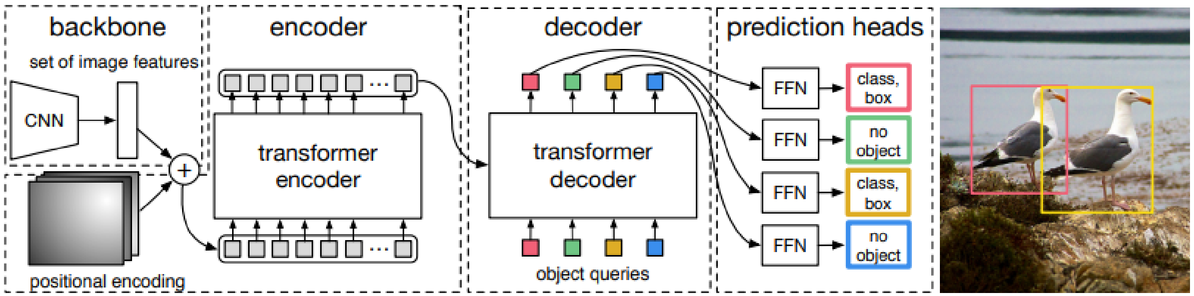
Apple	[52	0	0	1	0	2	5	0]
Banana	[0	47	0	1	0	4	8	0]
Contact soln	[0	0	41	0	1	1	17	0]
Orange	[2	0	0	51	0	1	3	0]
Petri dishes	[0	0	0	0	50	2	6	0]
Potato	[0	0	0	0	1	57	2	0]
Sausage	[0	1	2	0	1	1	55	0]
Vials	[0	0	1	0	0	1	1	57]

SSML

Apple	[57	0	0	2	0	0	1	0]
Banana	[1	55	0	2	1	0	0	1]
Contact soln	[2	3	51	1	1	2	0	0]
Orange	[2	0	0	55	0	0	0	0]
Petri dishes	[0	0	1	0	55	0	2	0]
Potato	[0	0	0	0	0	60	0	0]
Sausage	[1	1	1	1	1	1	54	0]
Vials	[0	0	0	0	1	0	0	59]

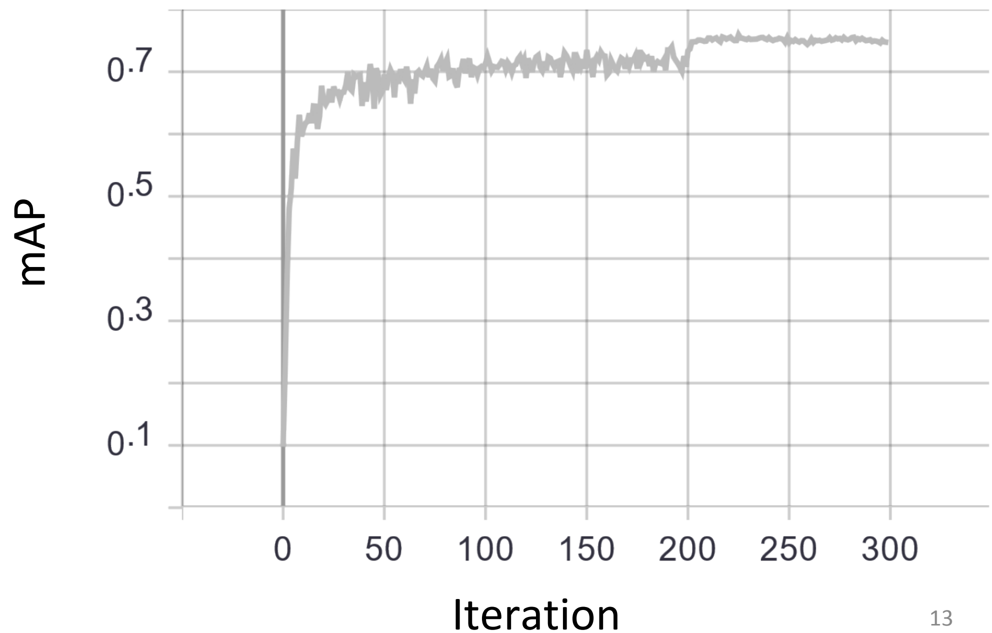
Object Detection with Transformers (EJ Guillen)

Algorithm: DEtection Transformer (DETR)



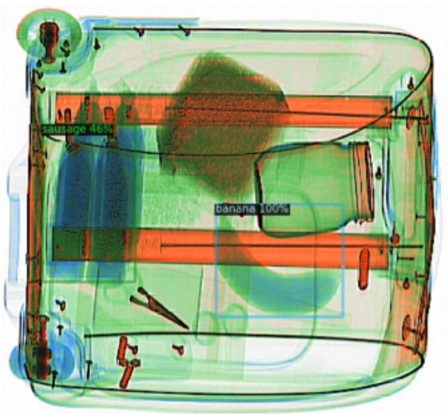
Average Precision @[IoU=0.50:0.95 | area= all | maxDets=100

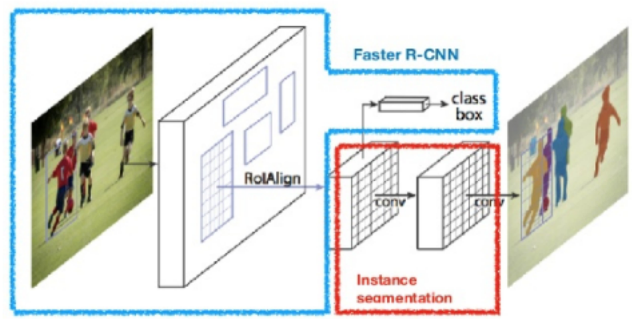
mAP: 0.7593
Among all classes.



Detectron2 (Dan Krofcheck)

- Presented at MLDL last year.
- Idea: why not try everything using a segmentation and detection suite?
- Mask R - CNN and many derivative improvements.

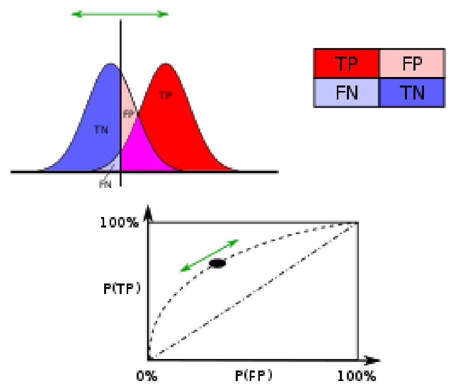




AP[0.50:0.95]	orange	potato	sausage	banana	apple	vials	petri dishes	c. lens sol.	mAP
Faster R-CNN 1	0.757	0.841	0.801	0.655	0.770	0.776	0.736	0.587	0.740
Faster R-CNN 2	0.758	0.880	0.770	0.662	0.759	0.823	0.756	0.487	0.737
Faster R-CNN 3	0.744	0.852	0.814	0.700	0.806	0.823	0.756	0.466	0.745
RetinaNet 1	0.834	0.850	0.823	0.684	0.816	0.834	0.768	0.567	0.772
RetinaNet 2	**0.841**	0.880	0.835	0.695	0.831	0.811	0.796	0.603	0.787
RetinaNet 3	0.835	**0.882**	0.814	0.723	**0.836**	0.837	0.773	0.592	0.787
Mask R-CNN 1	0.782	0.769	0.770	0.686	0.797	0.808	0.766	0.537	0.739
Mask R-CNN 2	0.795	0.869	0.818	0.691	0.743	0.816	0.780	0.543	0.757
Mask R-CNN 3	0.730	0.852	0.831	0.654	0.795	**0.848**	0.742	0.563	0.752

Independent Multiclass Optimal Thresholding (Salaudeen, Goodman) (1/3)

- Understanding fundamentals of results and how it influences operations – Ability to optimize performance and adjust confusion matrix.



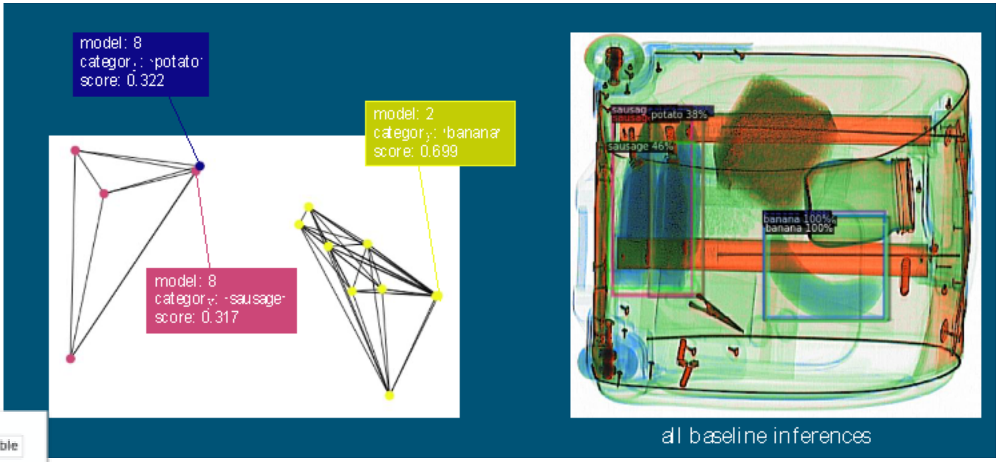
- What happens as detectable objects grow to hundreds, each with a certain FPR?
- How do we control individual class TPR/FPR based on risk and human factors?

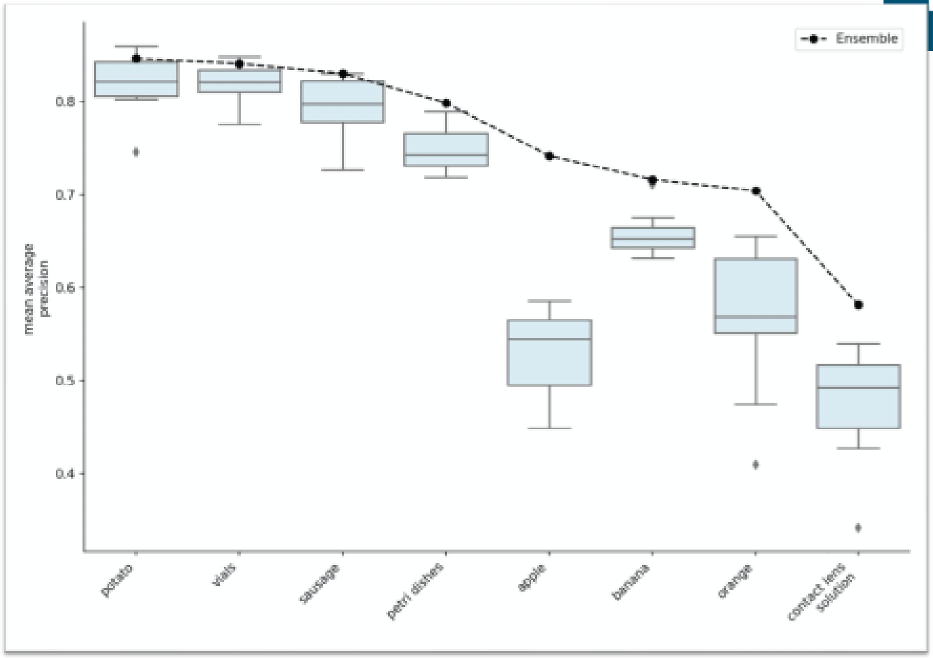
See: Independent Multiclass Optimal Thresholding (Olawale Salaudeen)

Ensembling Results (Dan Krofcheck , Eion Blanchard)

- Use all analysis, hope they present orthogonal information – Most powerful methods use a host of methods.

- Graph based method uses Cliques to resolve overlaps in baseline inferences.



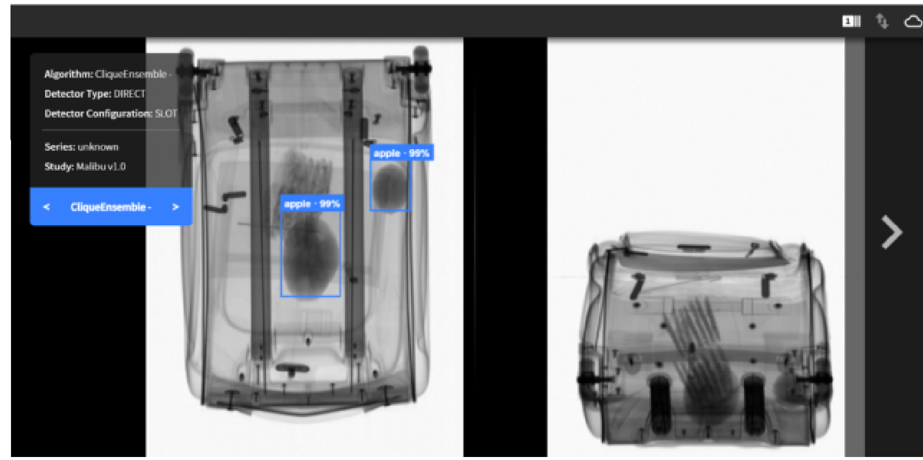


- Result is superior performance in all classes.
- Allows addition of models to improve performance.

Presented at MLDL last year

Back to Customer

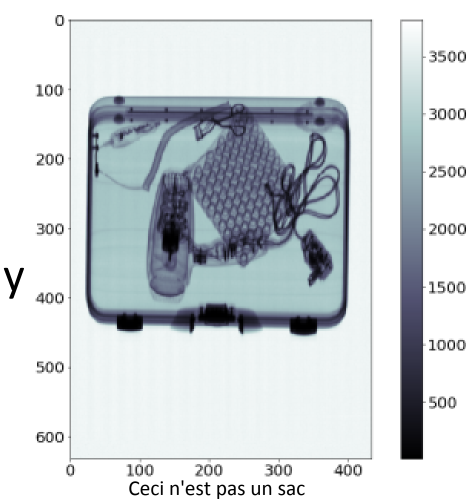
- **Successful Demo Of ATR. Ability to detect 8 items!**



Customer Response: “What about X?”, “Can you add Y?”
Our Response: Synthetic Data as a Lab Capability

Synthetic Data:

Strategic Capability
Annotated, labeled and plentiful
Rapid Response to emerging threats – Train algorithms quickly
Ideal: Move to cloud
Ideal: Threat feed of images
Ideal: Develop for several imaging platforms



Benefits of Synthetic Data

See Talk: Machine Learning with Synthetic Data:
What Approaches are Most Effective? (Jenny Galasso)

- Many lessons learned about synthetic data usage.
- Quantity and Quality matter – tradeoff.
- Remember: statistical error and don't pollute your test set!

Quick Check Analysis

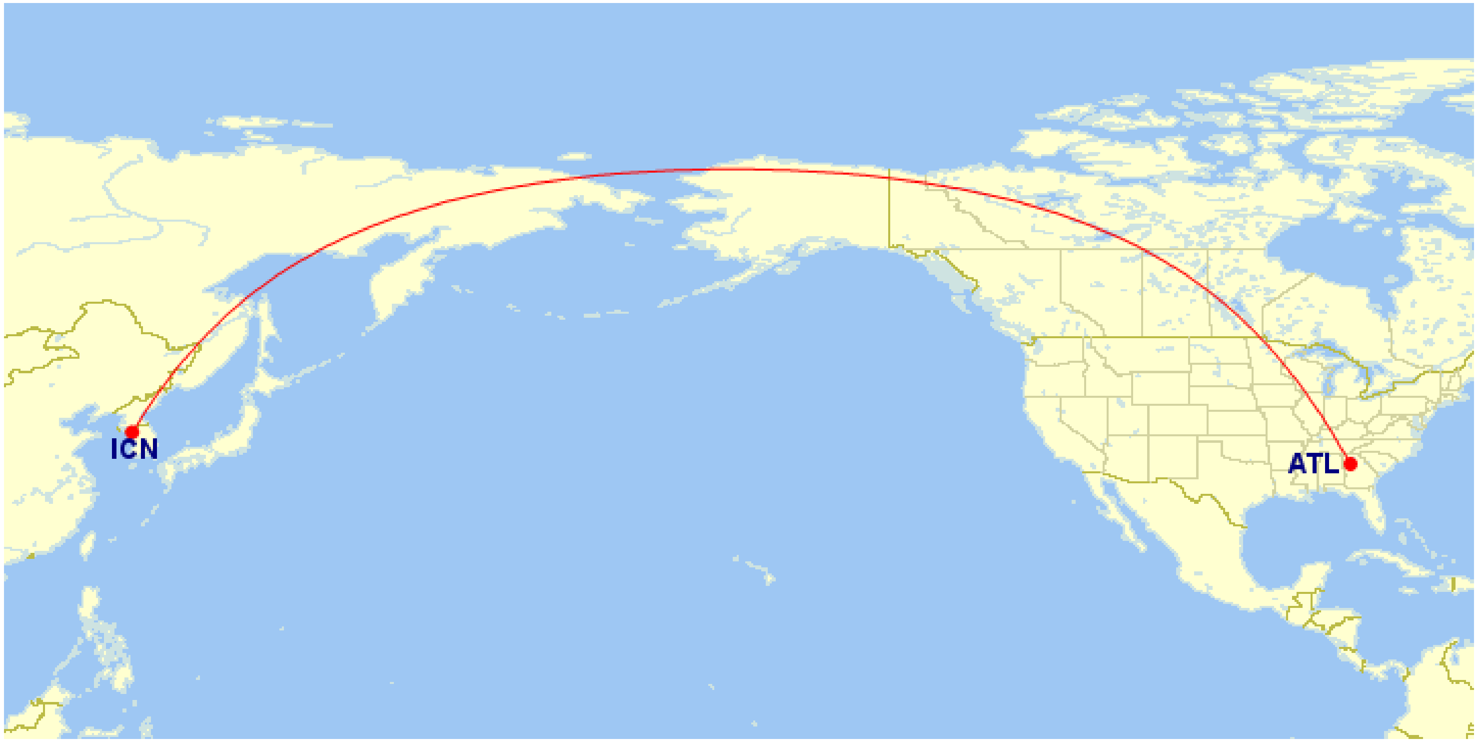
- Defined a Test Set of about 1000 images from Organics dataset
- Overall Process:
  - Select % of Original Images
  - Select % of Synthetic Images
  - Train on Original + Synthetic images
  - Evaluated trained model on Test Set

	100% Syn / 0% Original	0% Syn / 100% Original	100% Syn / 100% Original
Accuracy (%)	60.99	72.18	77.43

(Eric Goodman, Carter Jameson)

Project 2: Common Viewer Air System

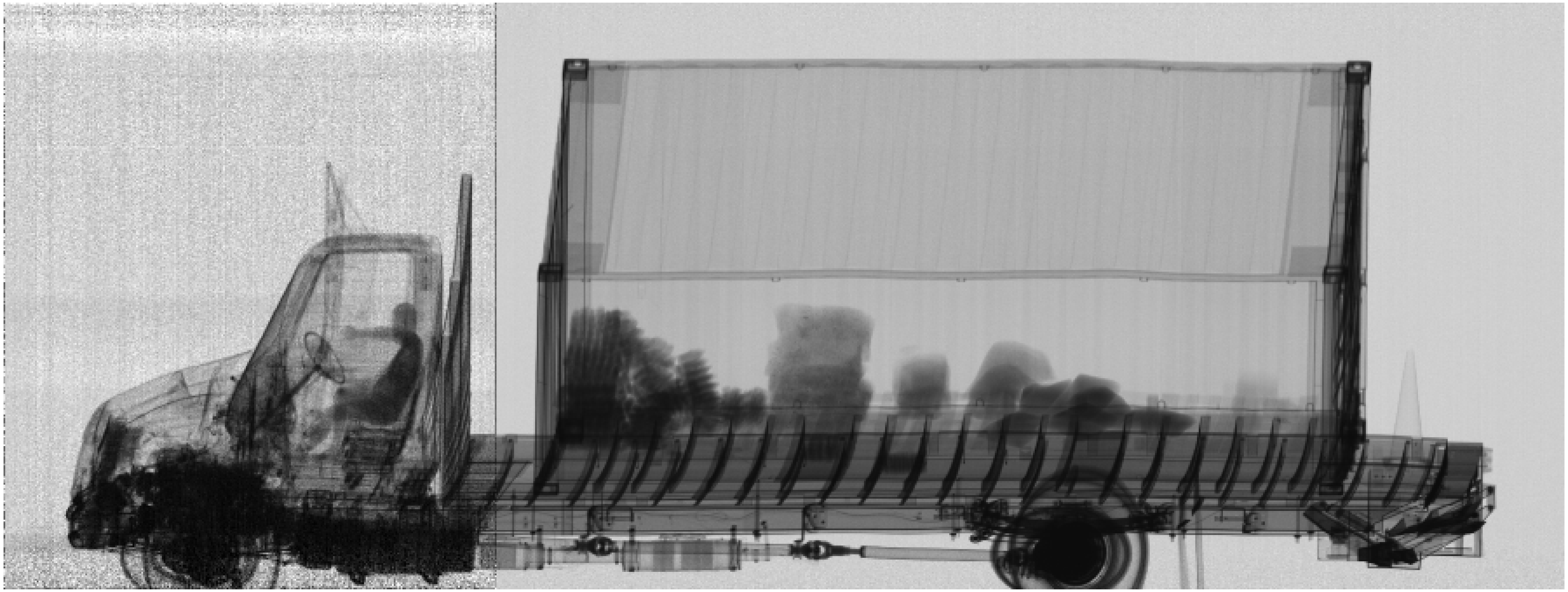
- Field demonstration will pilot the Common Viewer Air System, a cloud-based baggage pre-screening software system
- remotely screen checked baggage before arriving in the U.S. and landing at ATL.
- allow us to pilot the feasibility of screening images of every checked bag on an arriving flight before it lands in the U.S.
- pilot program will run in the summer of 2021.



See: https://www.dhs.gov/science-and-technology/news/2021/05/21/news-release-dhs-partners-south-korea-aviation-security

Project 3: Multi Energy Portal (MEP)

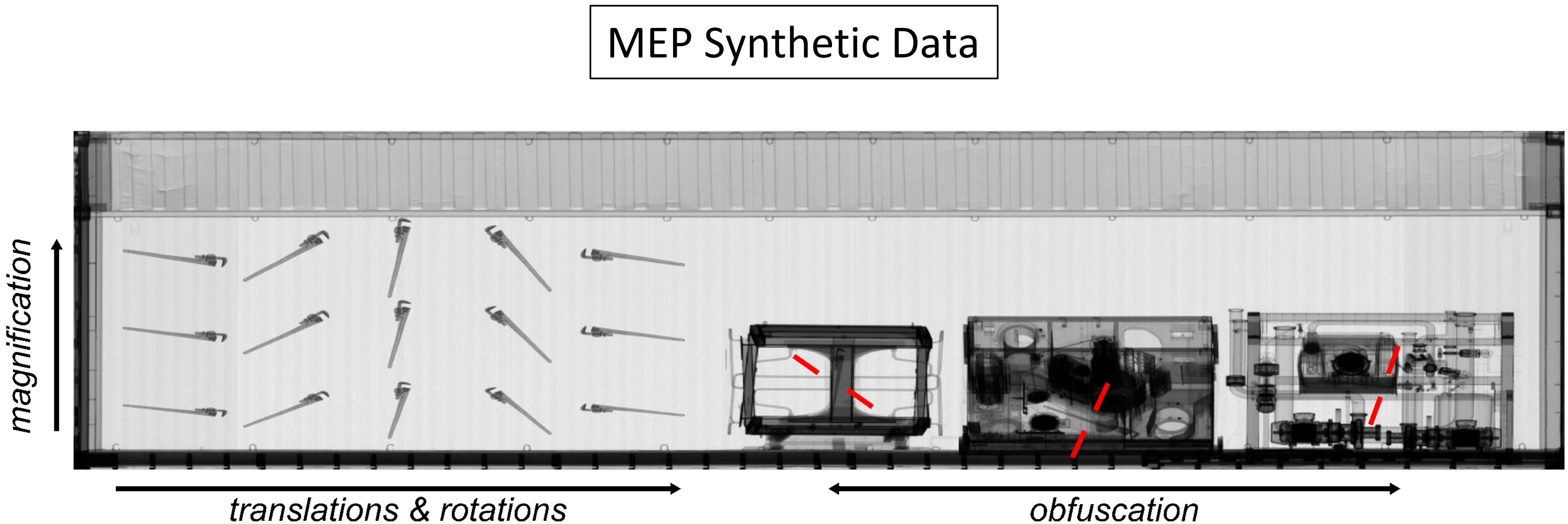
- High Energy scans of trucks entering through land borders



- Data: More sensitive, more rare, more expensive therefore more valuable

MEP Algorithms from Synthetic Data

We have small scale experience.



Vendors have shown it can be done and it helps.
Different energy levels, detector concpets.

More Complex Approach

(Synthetic) Data Generation

Physics-based TIP tuning (Arvind)

SinGAN Single-Packing Image Generation, (Jameson)

Tension Cycle Networks (Sorensen)

- Threat Image Projection (TIP) into X-ray images of cargo containers for training humans and machines (TBD)

(Synthetic) Data Refinement

CycleGAN for Threat Image Projection (Goodman)

GANs for machine Interchangeability (TBD)

Algorithms

Existing Algorithms (Mean Teacher, Detectron, Others.)

Utility Algorithms (Krofcheck)

Vision Transformer (Guillen)

Evaluation

Quick-Turnaround Synthetic image analysis (Sorenson, Galasso)

Structure Recognizes the importance of:

Cultivation and refinement of data

Quick evaluation of results

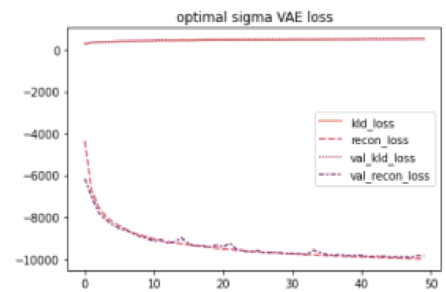
Appropriate approaches beyond 'latest and greatest'

Other Approaches

- Other approaches
  - Convolutional Variational Autoencoder leverages image similarity.
  - Topological Data Analysis (TDA) – Fingerprints structures in X-ray scans

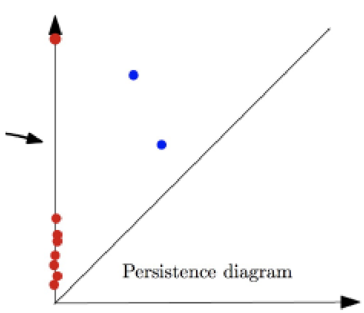
Image Similarity for Large Scale Non-Invasive Inspection

(Hugh Galloway, Dan Krofcheck)



Topological Data Analysis as a Descriptive Tool for Structural Similarity in Non-Invasive Inspection

(Derek Kielty, Dan Krofcheck)



What we didn't cover:

- Variety of possible algorithms. Risk Reduction is goal, image analysis in one part of that.
- Importance of data engineering and infrastructure.
- Proprietary systems and data.
- System compatibility. Can we only develop for one make/model of detector at a time?
- How to field? Getting to TRL 9 and the valley of death.
- Working with customers.

Additional Work Performed By:

- University Partners
- Industry Partners
- Two Future LDRDs ?

Thanks to all contributors

Chris Lamb
Tim Draelos
JD Doak
Joe Bertino
Chris Stork
Eion Blanchard
Steve Morrison
Rajeev Sahay
Olawale Salaudeen
Nathan Stewart
Mario Morford-Oberst
Kiril Taskov
Jackson Gittinger

Jenny Galasso
Jacob Fenger
Eric Goodman
Dan Krofcheck
Carter Jameson
Baylee Brown
Ace Sorenson
Sean Timm
Ryan Birmingham
Hugh Galloway
EJ Guillen
David Hannasch
Connor Frost
Connor Aubry
Arvind Prasadan

Sponsor: DHS/S&T

Backup

Abstract

- Inspection of commerce entering the country is currently an onerous, bespoke and extremely labor intensive process requiring specialized skills developed by CBP officers over a career. Non-Intrusive inspection (NII), specifically the imaging of commerce with X-Rays, aids CBP's ability to interdict illicit items entering the country, reducing risk and increasing national security. Although vast amounts of imaging data exist, they are not fully leveraged. Sandia has a portfolio of projects using ML techniques to use this vast amount of data to detect threats entering the country. This talk will be a review of the CBP NII portfolio including several modalities of inspection and projects using a variety of data to enhance NII for CBP. Subsequent talks will then give detailed descriptions of specific analysis supporting this portfolio.

- Specifically, we highlight synthetic data as an enabling lab capability to rapidly respond to threats, as well as several other techniques that leverage this capability to reduce risk at the border.

- These capabilities ultimately serve to aid CBP officer's adjudication of images thereby supporting reduction of risk and the rapid and secure flow of commerce into the United States.

Problem you are trying to solve

NII is fraught with challenges. Ultimately, the adjudication of X-Ray images is a practice in the acceptance of risk. An assessment by a CBP officer balances the availability of resources and need for the free flow of commerce with the security of the United States. We view any improvement in this balance as a win. Supplementing the adjudication of images with ML techniques augments the breath and depth of inspection, and therefore nudges this balance in the favor of reducing risk.

Algorithmic approach of your solution

- We employ several algorithmic techniques to solve the problem. Generally, we have synthetic data generation techniques that can rapidly spin up datasets that are then employed to train ML models. However, this is not enough for a systematic approach. GANs are used to refine the data; thresholding analysis objectively improves S/N; semi-supervised methods are used to handle data sparsity; similarity and change detection along with topological data analysis leverages historical data to investigate anomalies.
- Overall, a systems analytical approach views all the above as a study in risk, and both technical and systems knowledge is required to employ capabilities to increase security a the border.

Description of the data used

Currently the NII data we use is generated with X-Rays collected at the border by CBP or manually by SNL. It comes in many forms. We handle both manually generated and real, stream of commerce (SOC) data. Modalities include hand carried luggage, airport checked bags, as well as large scale tractor trailer cargo. We handle 2D transmission X-Rays in lower (~100KeV) as well as higher (~10MeV) energy ranges, optical and backscatter images and computed tomography (CT) data.

Results

We will present results for each of the individual component projects as subsequent talks. Generally, we think you will see individual results as a systematic approach to reducing risk for CBP.

Conclusions

- CBP NII has a surplus of unused data, fleetingly used to adjudicate risk of commerce entering the country. By leveraging a host of ML techniques supported by synthetic data generation capabilities, we can bend the risk equation to the benefit of national security.

- We will present a 20 minute overview of the CBP portfolio that will set the stage for subsequent technical talks.