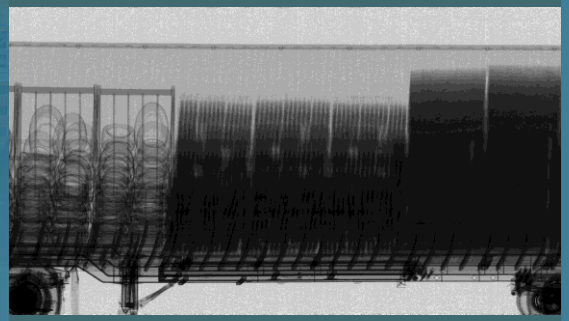
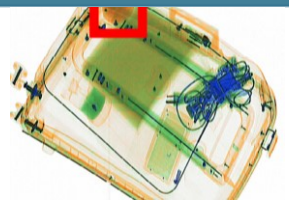
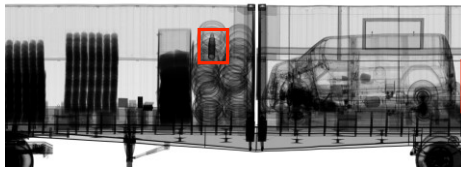
Unclassified Unlimited Release,

Sandia National Laboratories

# Development and Deploying Algorithms and Data for CBP - National Lab Perspective








Rob Forrest

Sandia National Labs, Supporting DHS/S&T

**rforres@sandia.gov**

## Development and Deployment of Algorithms for Enhancing the Interdiction of Contraband

July 25-26, 2023

Unclassified Unlimited Release

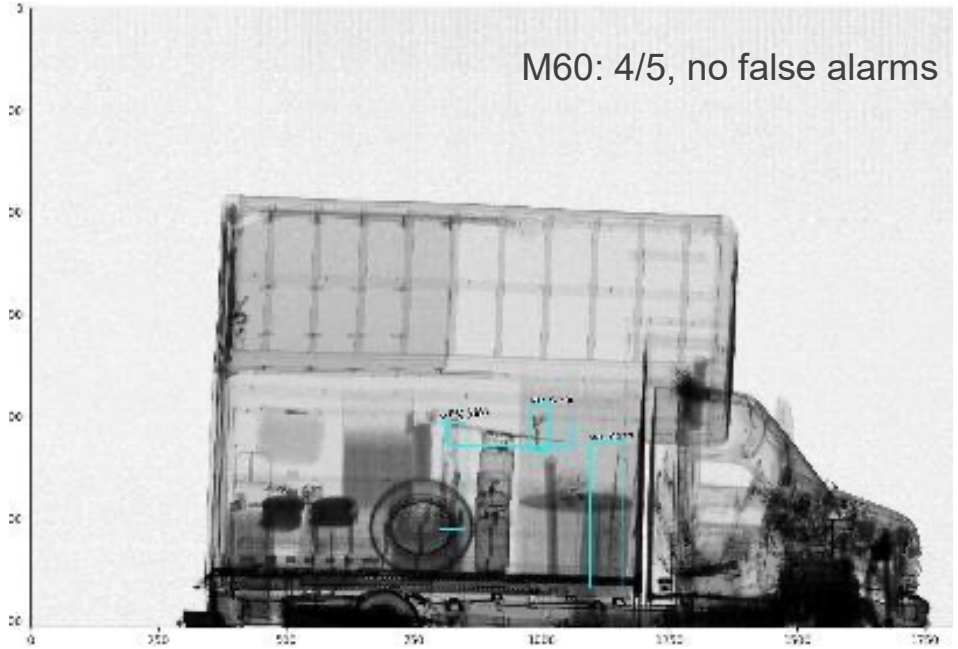
U.S. DEPARTMENT OF ENERGY | NNSA National Nuclear Security Administration

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.

# So What? Who Cares?

- **Goal: Get X-ray Algorithms into production.**

- Issue: Structural barriers exist for **community** to contribute.




- Problems:
  1. No pathway to deployment.

- Solutions:
  1. Cooperation, community, access, risk acceptance.
  2. ?

- TRL: Stuck at 5

# X-Ray Algorithms For X-Ray NII : Sandia's Experience

Worked:

On MEP, Z Portal, CarView Detectors

With most vendors

Manually taking data ~5 weeks @ 3 POEs

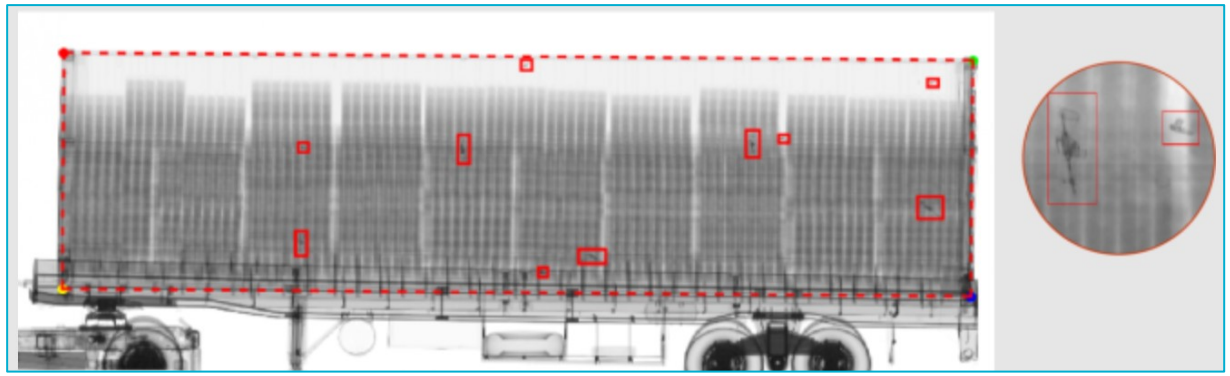
With SOC data, Threat, Synthetic Data

On CBP cloud

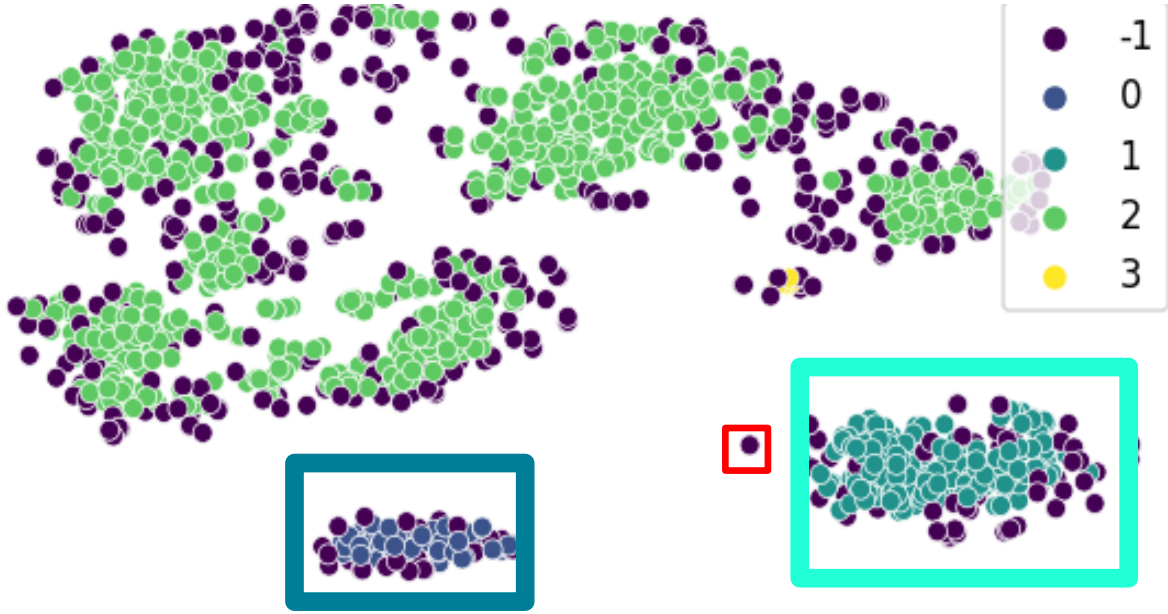
**Developed suites of threat and anomaly detection algorithms for all detector types.**

# The Algorithms are DONE

AI/ML techniques are now far more advanced than what CBP needs.



Vendor Threat Detection Algorithm
~2019




**Sandia Anomaly Detection Algorithm (ADA) for MEP**

Complete circa 3/2022

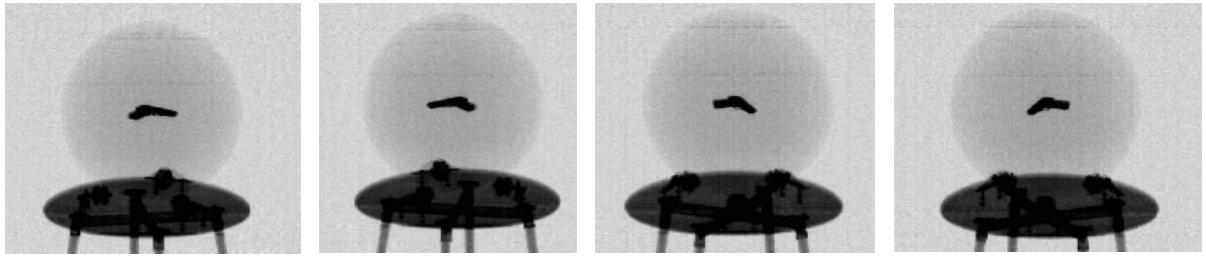
# Data is Not the Problem

(Data access may be)

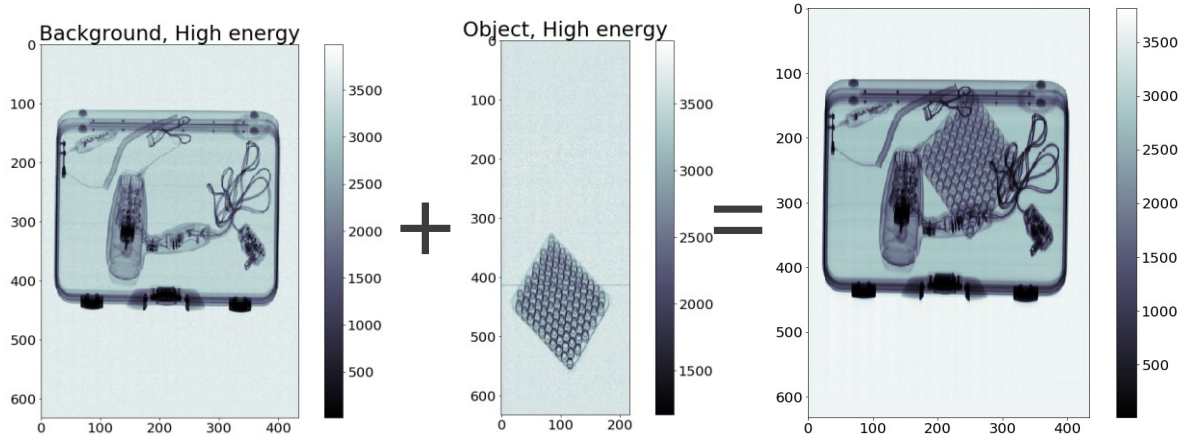
We have:

- Stream of Commerce data
- Threat data
- Synthetic data
- Labeled data

## Threat Data



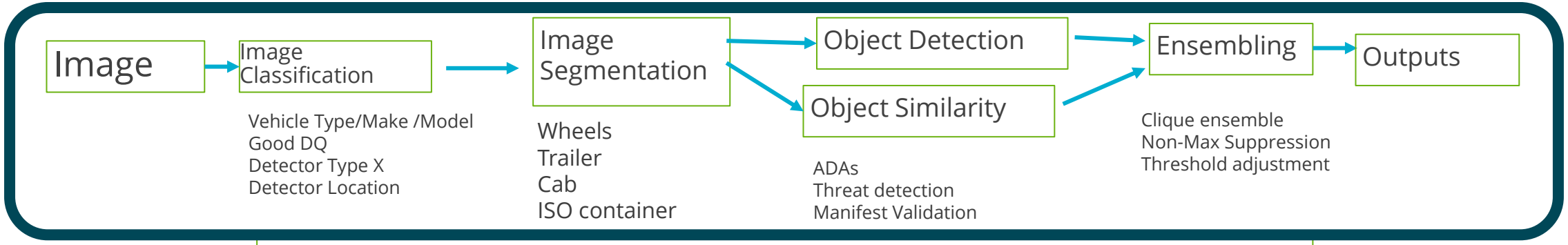
## Synthetic Data: passenger luggage




"The future is already here – it's just not evenly distributed."

# Obstacles

- *Algorithm developers* need to understand POEs.
- *Users* need to understand: AI models need **investment.**
- Move away from: “Buy it and were done.”
  Understand: Building a **system** that needs **curation**.
  This IS the work.




What part are you going to ‘buy’ here?

# Solutions?

- Ongoing relationship between CBP, POEs and developers
- Community of trusted developers. Access.
- Algorithm governance structure
- Open standards (API, Metrics, processes)
- Risk acceptance
- Look and learn from other models of success

# Backup

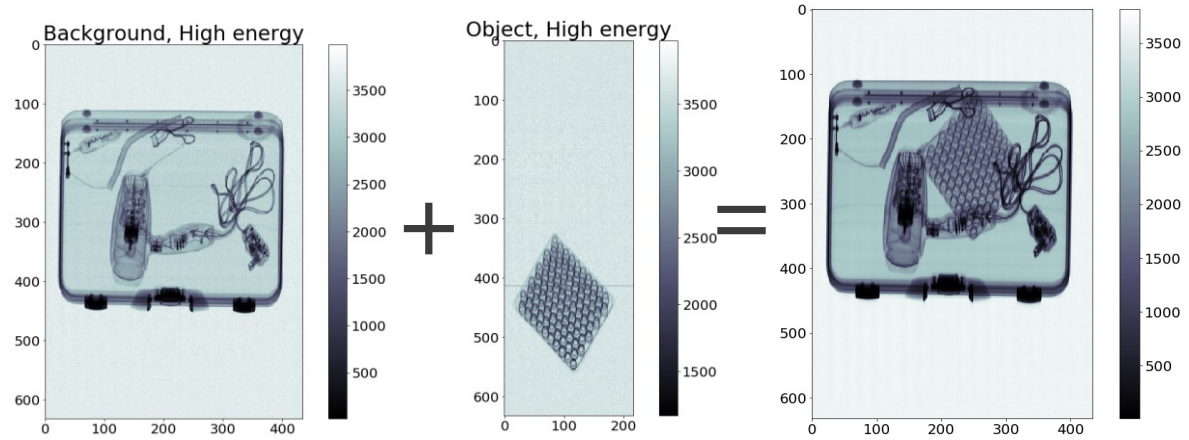
# Synthetic Data

Fundamental problem: need lots of labeled data. Solution: Synthetic Data

Everyone has different dataset access. Manual data generation untenable.

- “Photoshop for X-Ray images” – but much more powerful
- Start with dozens of threat images – end with thousands of labeled training images.
- Generate/Share with community as needed.
- **Results: We have millions of images of conveyances with embedded threats.**

**Example from packed passenger luggage:**




Example Result:

| | 100% Syn / 0% Real | 0% Syn / 100% Real | 100% Syn / 100% Real |
|---|---|---|---|
| **Acc (%)** | 61. | 72. | 77. |

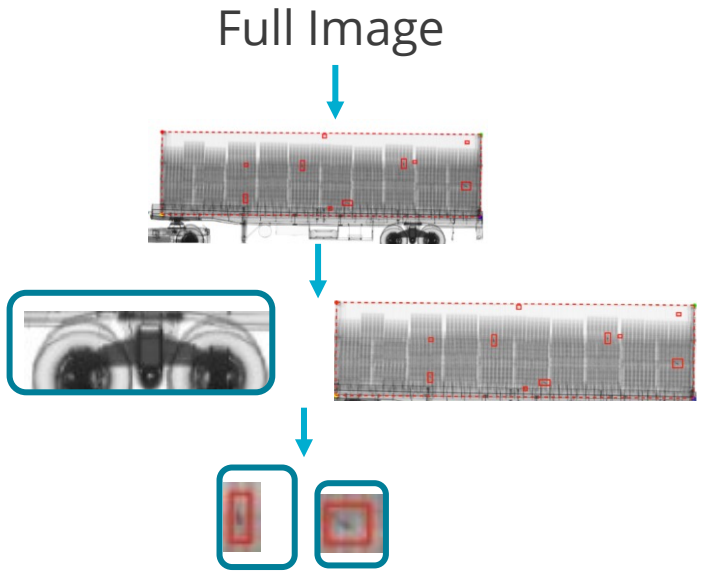
250 Images → Synthetic Data Engine → 5,000 Images for Algorithm Development

**Algorithms should be developed in weeks, not years**

# Open Standards

**Benefit: Allows contributors to understand how to easily add their algorithms into the CBP-owned algorithm system.**

Power lies in a **system** of capabilities working together.

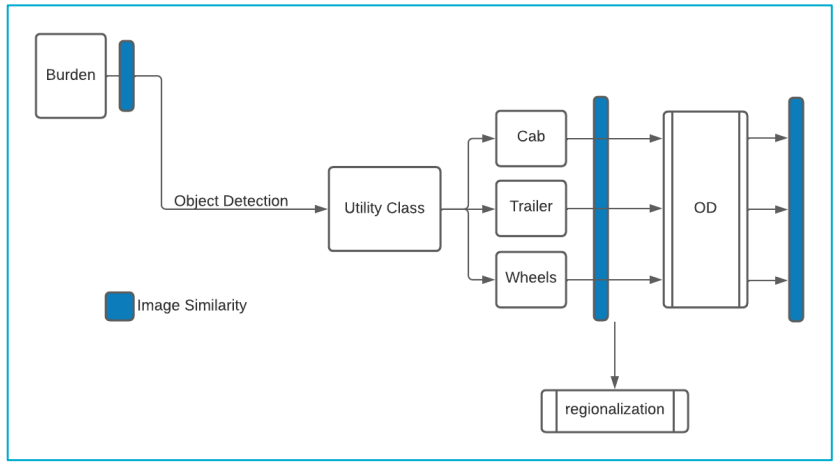



## Solutions

1. CBP defined APIs –How algorithms talk to each other.
2. Single CBP Ontology – Common algorithm language.
3. Sanctioned Cloud Client Environments (See PNNL talk)
4. Common execution Environment…

## Ontology

| Conveyance Type | Trailer Type |
|---|---|
| Fuel Truck | Dry Van |
| Dump Truck | Refrigerated Trailer |
| Cement Truck | Flatbed/Lowboy |
| Motorcycle | Tanker |
| Bus | Vehicle Carrier |



# Open Standards: Running Operational Algorithms

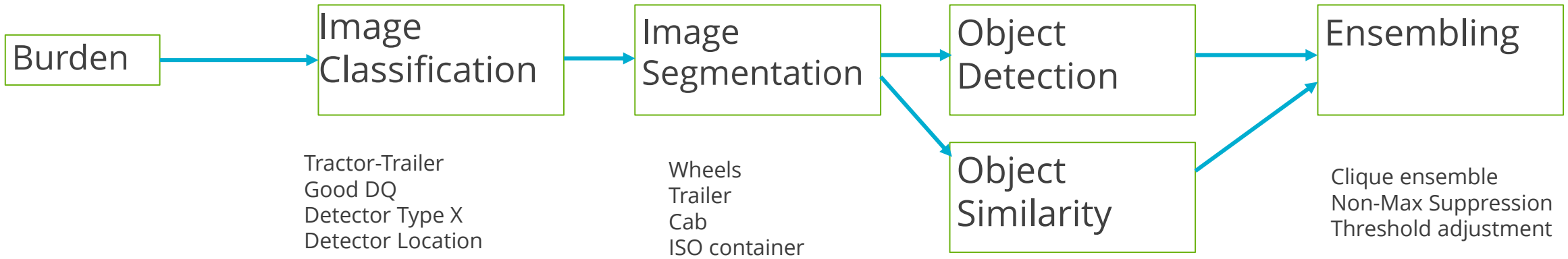
Good: Simplify. Adding additional algorithms should be easy.

Common Integration Platform: Run all algorithms together on **CBP systems**.

Measure performance of whole system (PD, PFA). Test offline. Don't break anything.

Professional and custom platforms enable this.

**Benefit to CBP: Upgrade and plug in a new algorithm to existing system easily.**



# Algorithms, Adversaries are not stagnant.

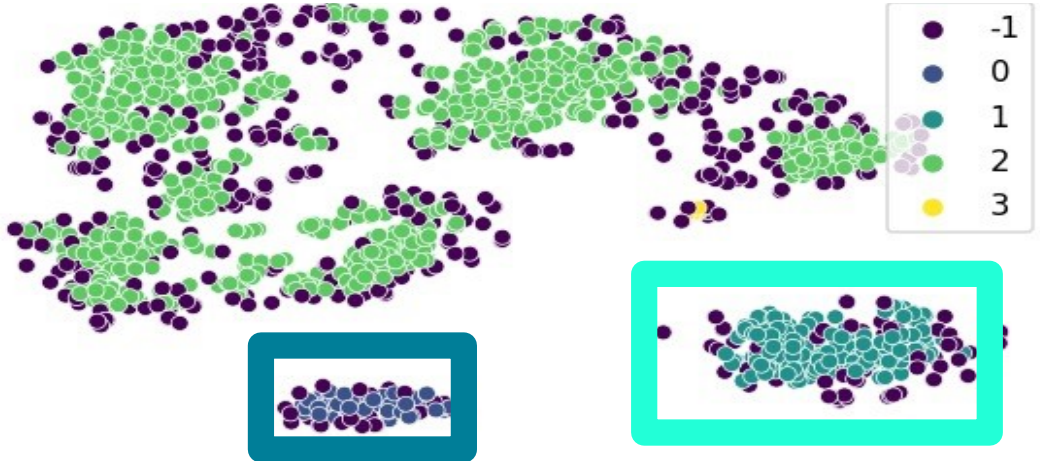
Every algorithm gets smarter with more data and types of data.

Example: Officer adjudications, seizures, analytics.

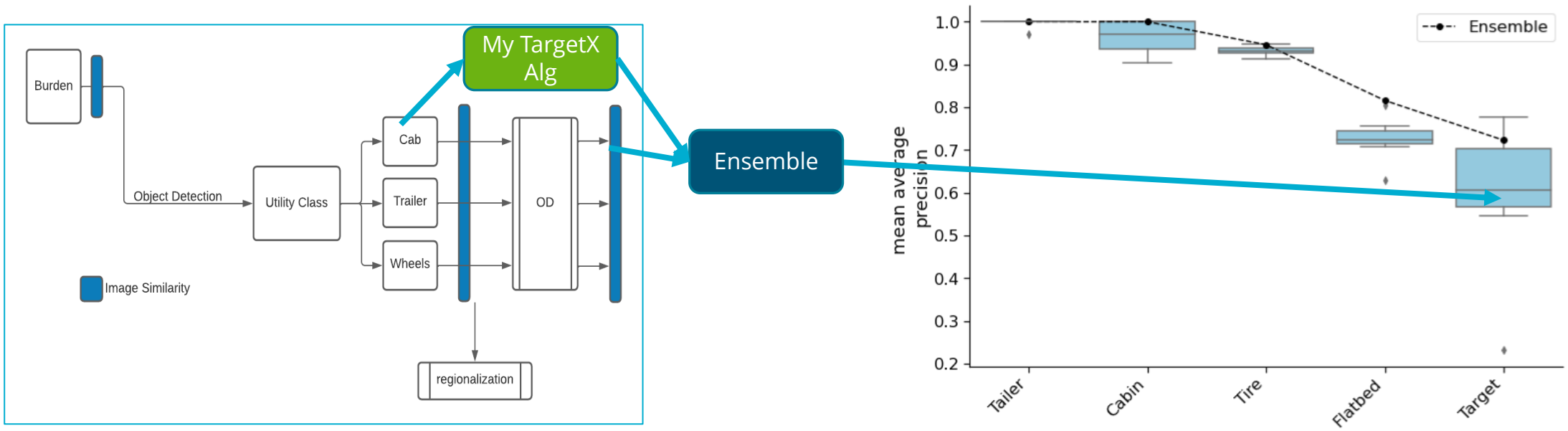
Adversary adapts, threats change.

Example: find a threat at one port, look for it everywhere.

**Benefit to CBP: New threat? Algorithms updated quickly to find it.**



# Open Standards, Ensembling:

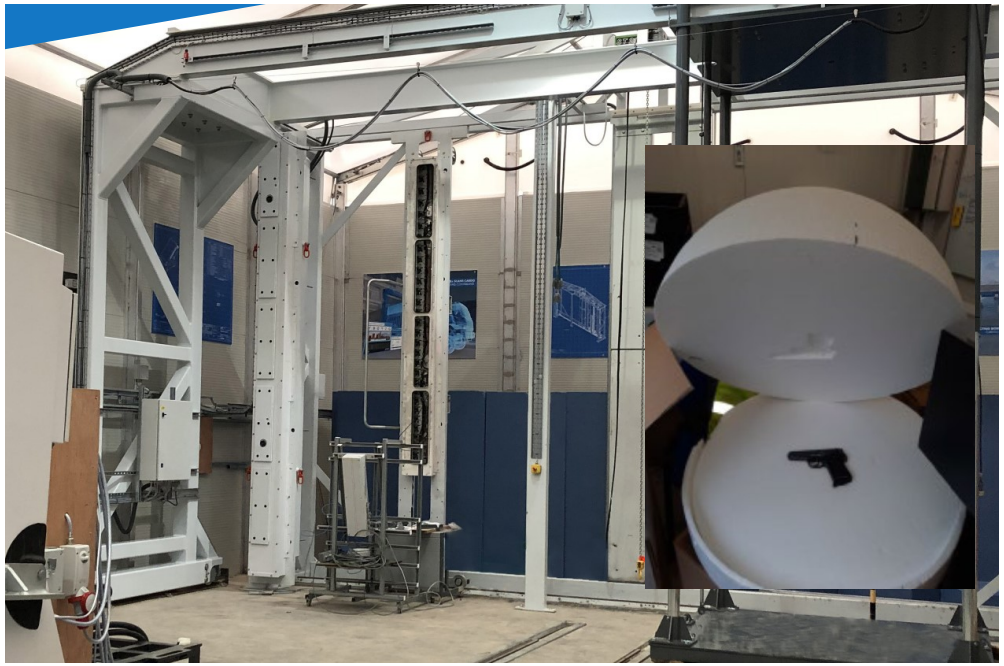



Open Standards Enables: Ensembling Results

# Solution: Synthetic Data

Enables:

- Ability to respond and adapt to emerging threats quickly.
- Less dependency on fielded systems.
- Huge savings of resources. Less human labeling.
- Capability: Applies to many NII detectors, large and small

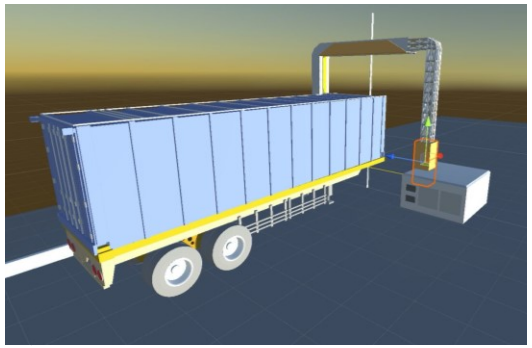


Real-world validation: Using hardware in the field.

One step further:

Imaging threats in the lab

Digital twin models



**Algorithms should be developed in weeks, not years**

# Problem 2: Integrating Algorithms Together

With quality data, algorithm development is greatly simplified.
Lots of developers have similar capabilities.

- Object segmentation : Dissect image into parts (automatic labeling: Wheels, cab)
- Threat detection : Can we find an item in an image? (Narcotics, weapons)
- Anomaly detection : What does “normal” look like? (Empty verification)
- Image similarity : What other images look like this one? (Manifest verification)