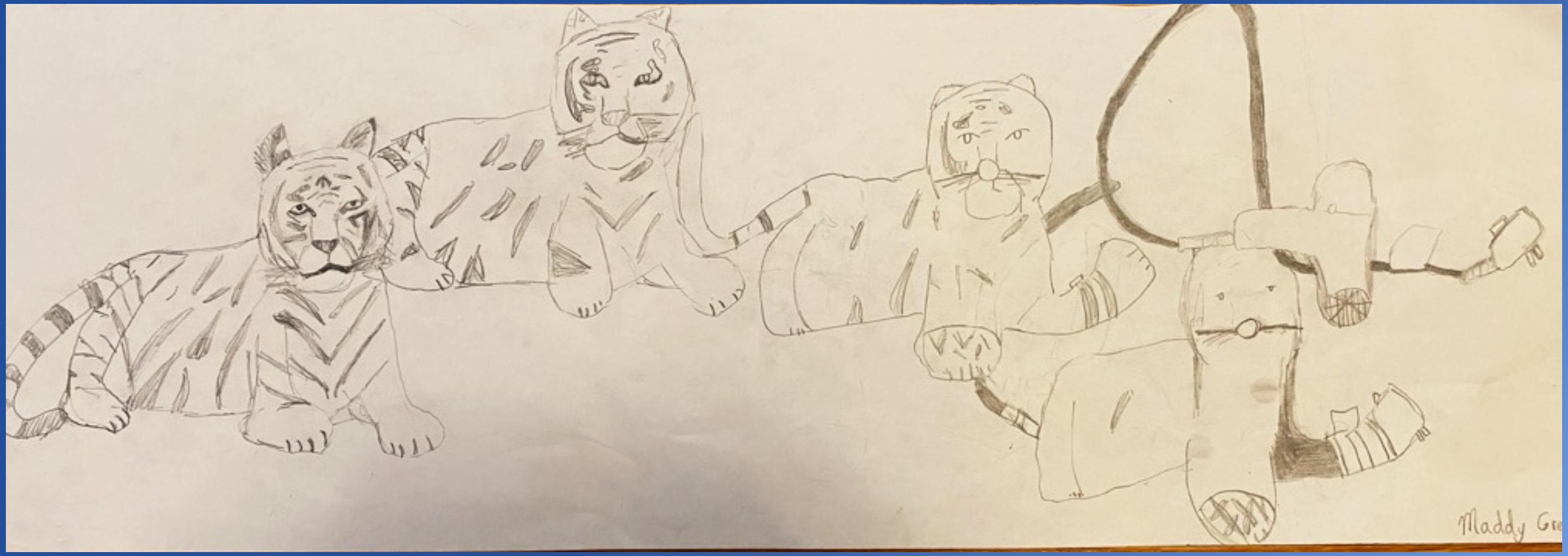
Unique opportunities to sample parameter space using physics-based synthetic data



Joel Greenberg, PhD
President & CEO @ Quadridox
05.03.23

Acknowledgements

Dr. Joel Greenberg
Dr. David Coccarelli
Daniel Pike
Steve Feller

Funding

Funded in part by the UK Department for Transport

Collaborators

iconal

SURESCAN

J.A.G. and D.S.C. have a financial interest in Quadridox, Inc

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agencies.

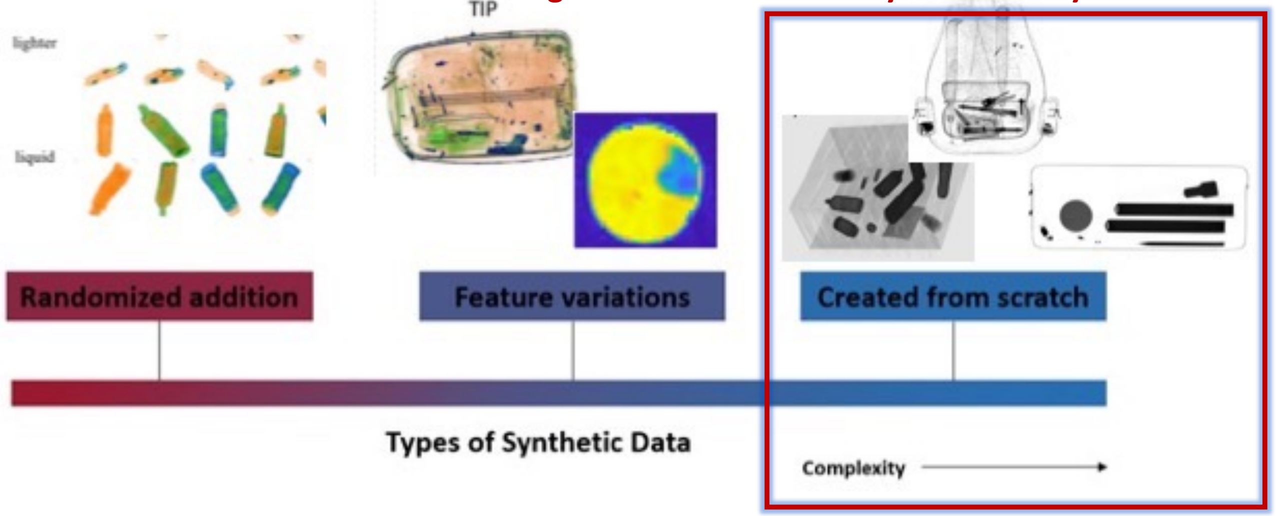
‘Types’ of data

Real data: empirical measurements

Original data

Synthetic data: generated programmatically

GAN TIP & data augmentation Physics-based synthetic



Requires real scanners and bags | Only virtual scanners and bags

R. Krauss, SPIE ADIX Proc 117380L (2021)
https://www.luxresearchinc.com/blog/the-what-and-why-of-synthetic-data

J. Yang et al. IEEE Access, vol. 7, pp. 28894-28902, 2019

Wang, Q. et al. Journal of X-ray science and technology (2020)
J. Greenberg et al., SPIE ADIX 1173807 (2021)

QUADRIDOX

Why use physics-based synthetic data?

- Only way to solve **cold start problem** (when no real data exists)
- Overcomes **data restrictions** (proprietary, sensitive, etc.)
- Naturally includes relevant **physics** (algorithms don't have to 'learn' known physics or have it manually inserted)
- **Fully controlled** and understood data
- **Perfect ground truthing**/labeling for free
- Helpful for **high dimensional data** (especially when not 'images')
- **Safe** (no need to handle energetic materials or ionizing radiation)
- **Cost effective** (less expensive than empirical data collection)

Potential opportunities in security

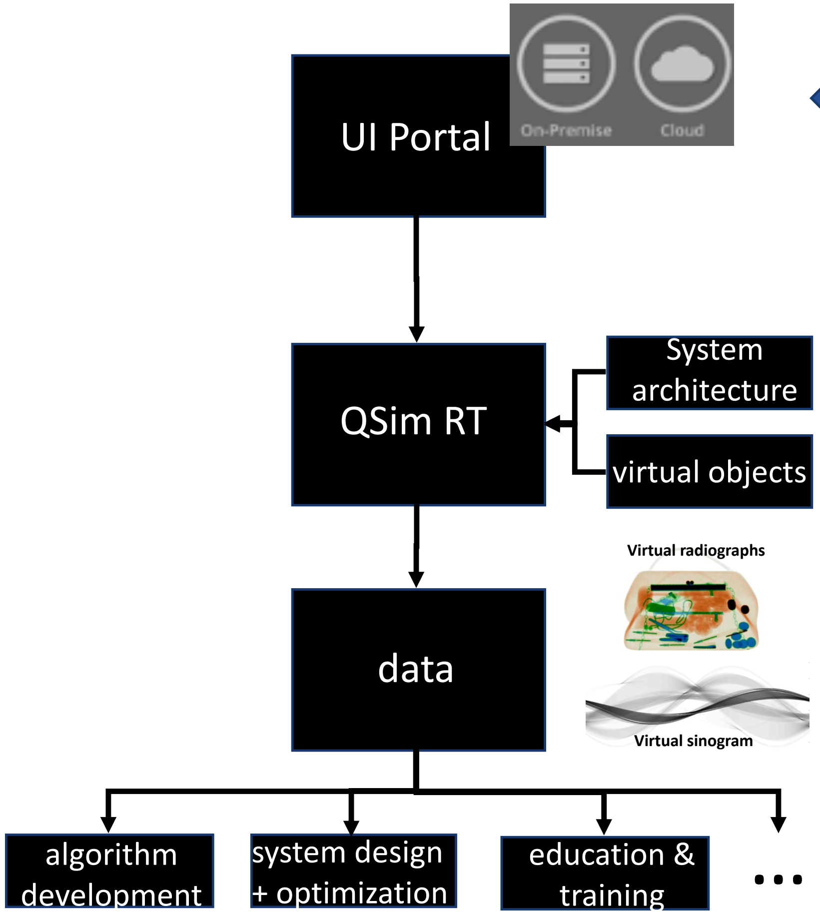
Regulators

- Fast response to new and emerging threats (train and test new algorithms)
- Excursion testing
- Certification testing “at scale” (test Pd)
- Key enabling component to open architecture paradigm (third party ATRs)
- Education/training (e.g., replacing/augmenting TIP libraries)
- Analysis of new technologies and multi-modality/system of systems performance

Developers/OEMs

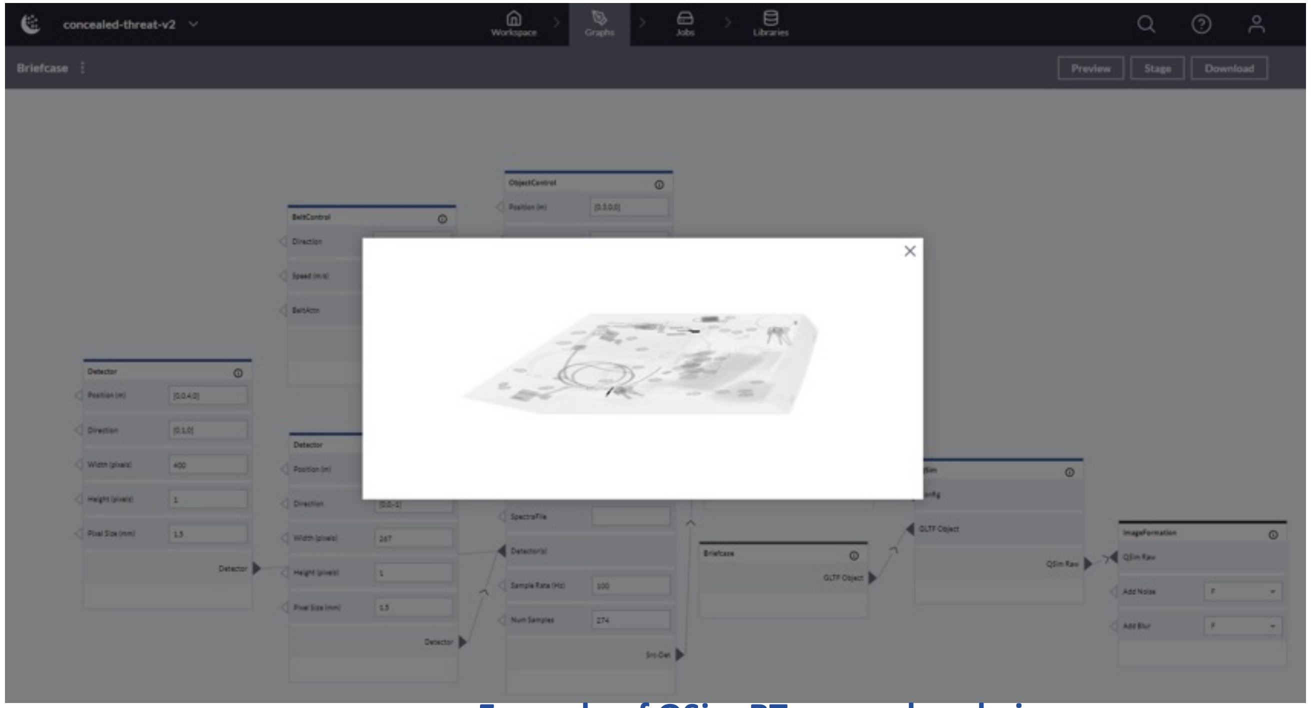
- Shorter, less expensive time to certification/market
- More robust system design/development process
- Improved algorithm performance (better P_d/P_{fa})

Physics-based synthetic data generation



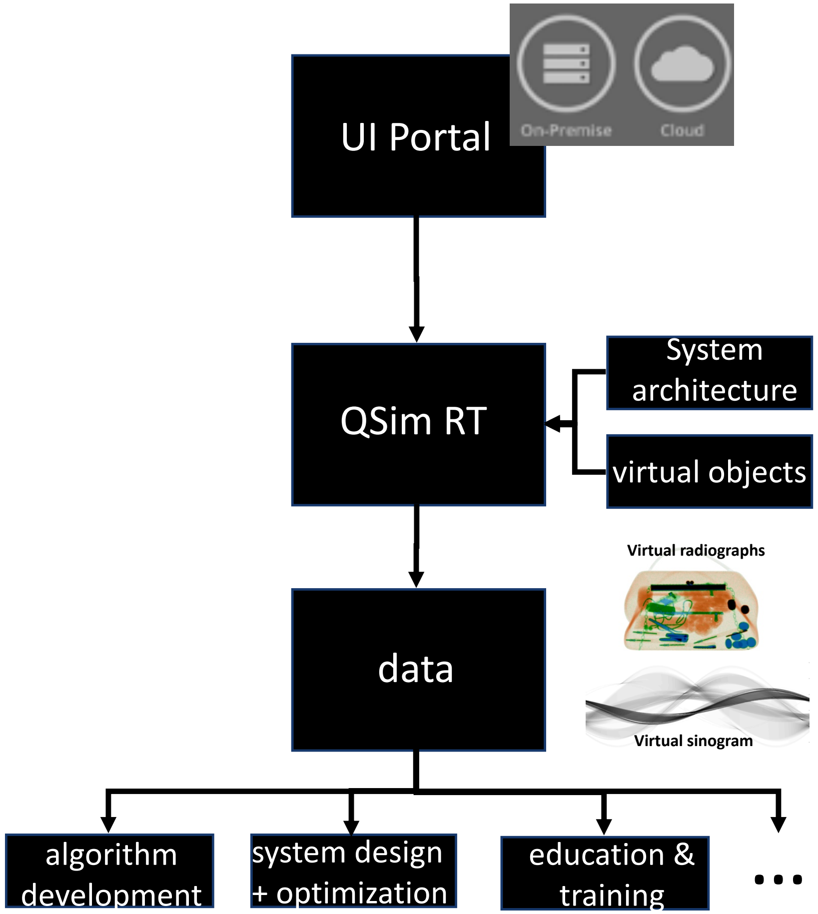
Software to:

- Enable human to design simulation scenario
- Manage jobs
- Manage data

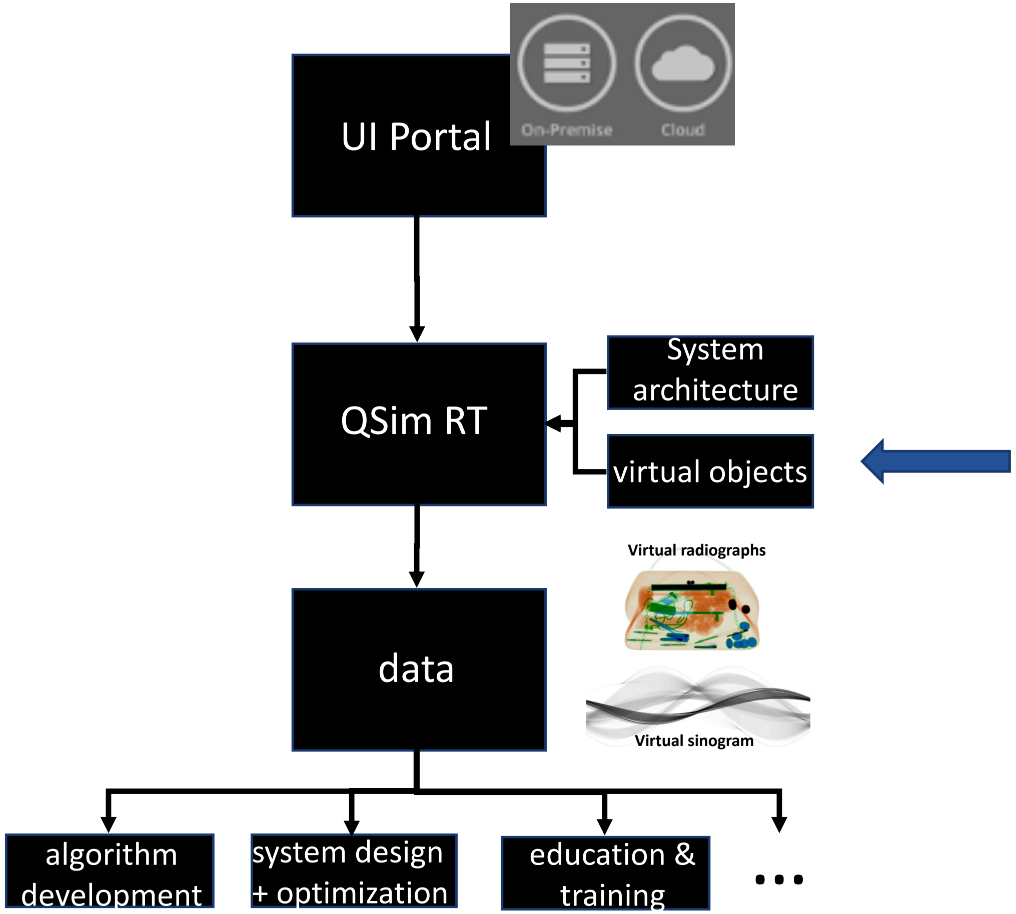


Example of QSim RT on rendered.ai

Physics-based synthetic data generation

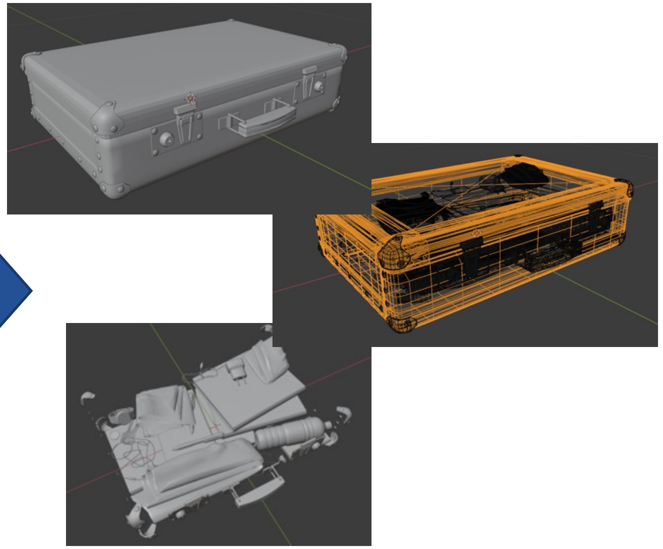


Physics-based synthetic data generation



Define objects

'Pack' objects in enclosures



Physics-based synthetic data generation

UI Portal

QSim RT

System architecture

virtual objects

data

Virtual radiographs

Virtual sinogram

algorithm development

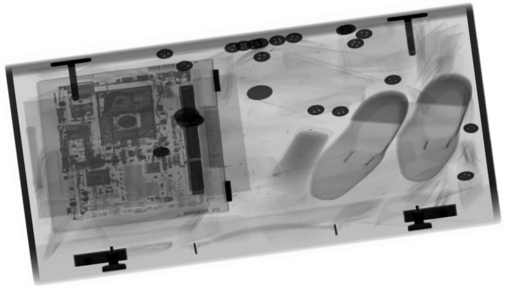
system design + optimization

education & training

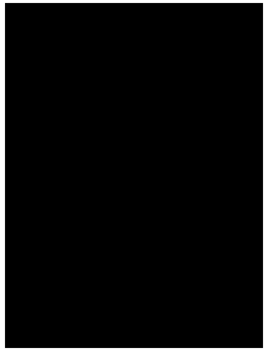
…

'Image' data

AT

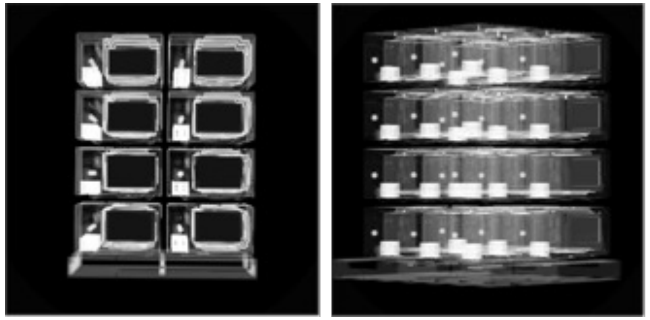


CT

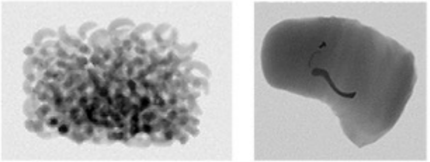


Cargo

Air cargo

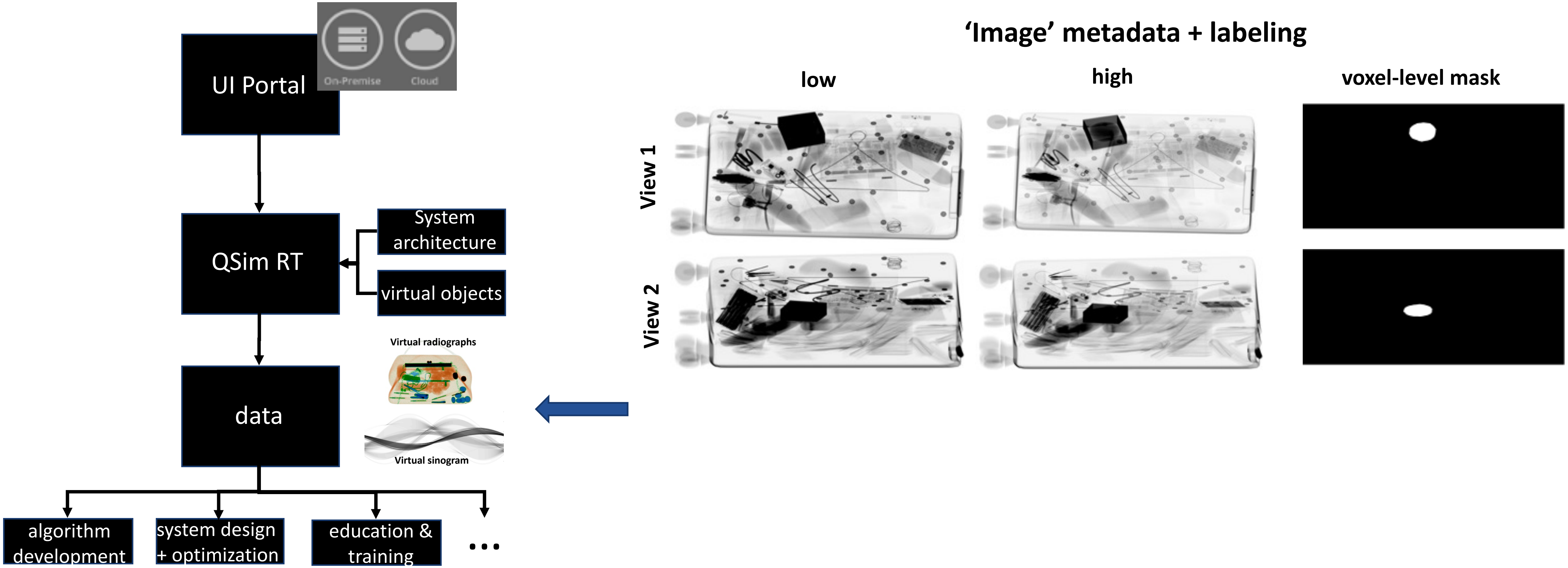


Food/agriculture



Few seconds for a radiograph; couple minutes for a CT volume

Physics-based synthetic data generation



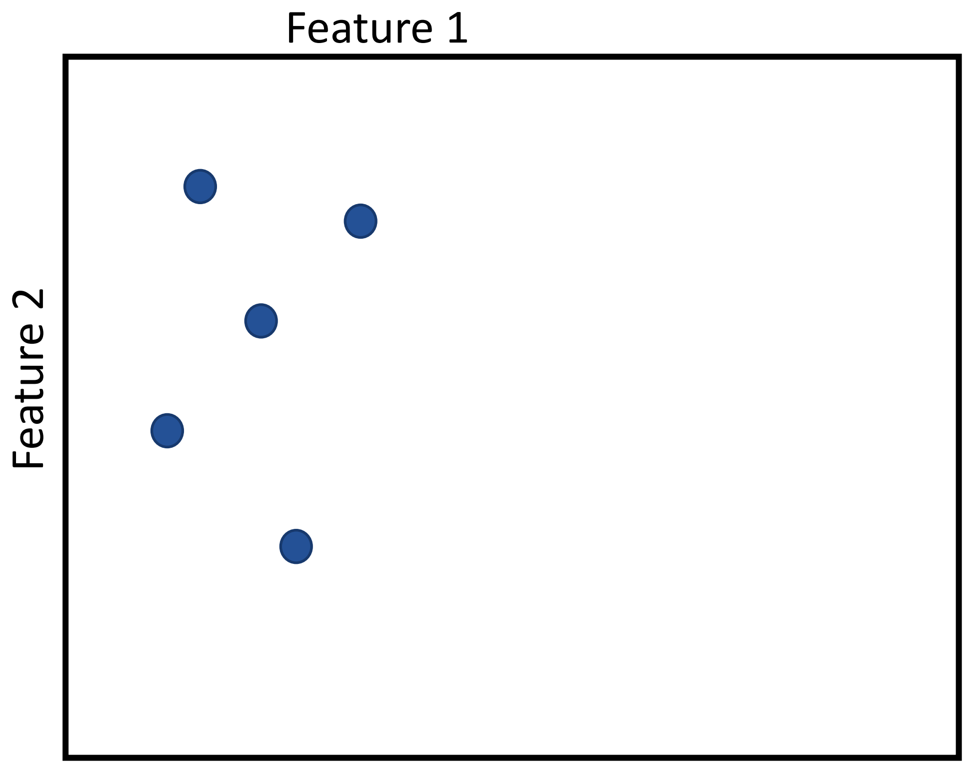
QUADRIDOX

Creating an ensemble

Two approaches to synthetic data ensemble design

1. **Targeted**: targeted sparse, discrete sampling (matches empirical data collect)

2. **Exploratory:** broad, controllable sampling (enables tailored training and testing)



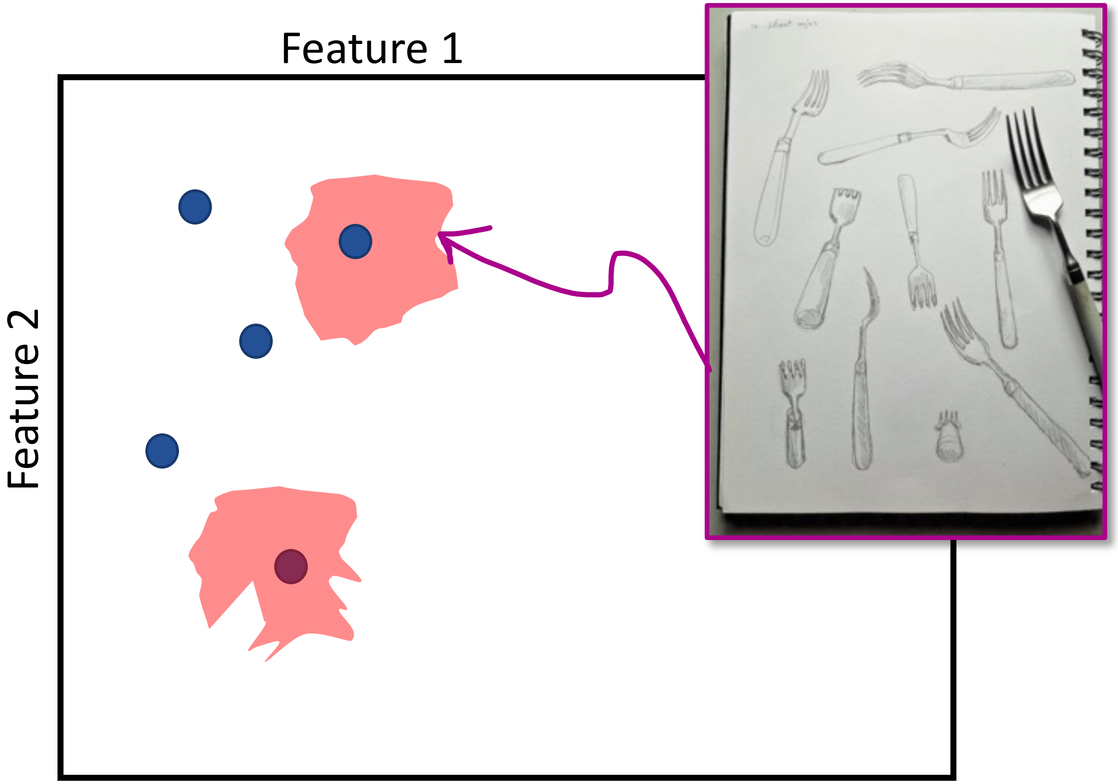
http://2d-preston.weebly.com/hand-transformation.html

Creating an ensemble

Two approaches to synthetic data ensemble design

1. **Targeted**: targeted sparse, discrete sampling (matches empirical data collect)

2. **Exploratory:** broad, controllable sampling (enables tailored training and testing)



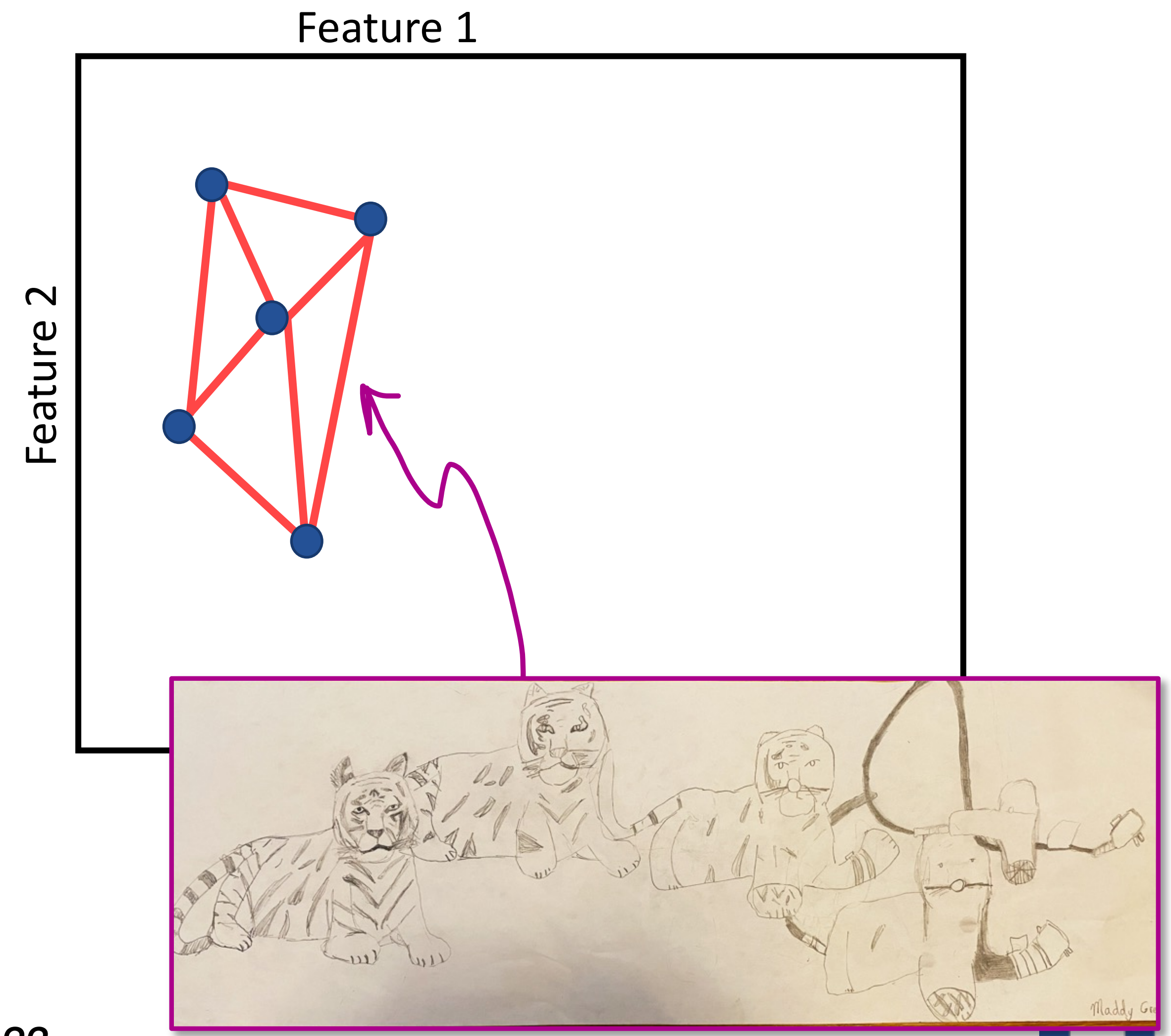
http://2d-preston.weebly.com/hand-transformation.html

Creating an ensemble

Two approaches to synthetic data ensemble design

1. **Targeted**: targeted sparse, discrete sampling (matches empirical data collect)

2. **Exploratory:** broad, controllable sampling (enables tailored training and testing)



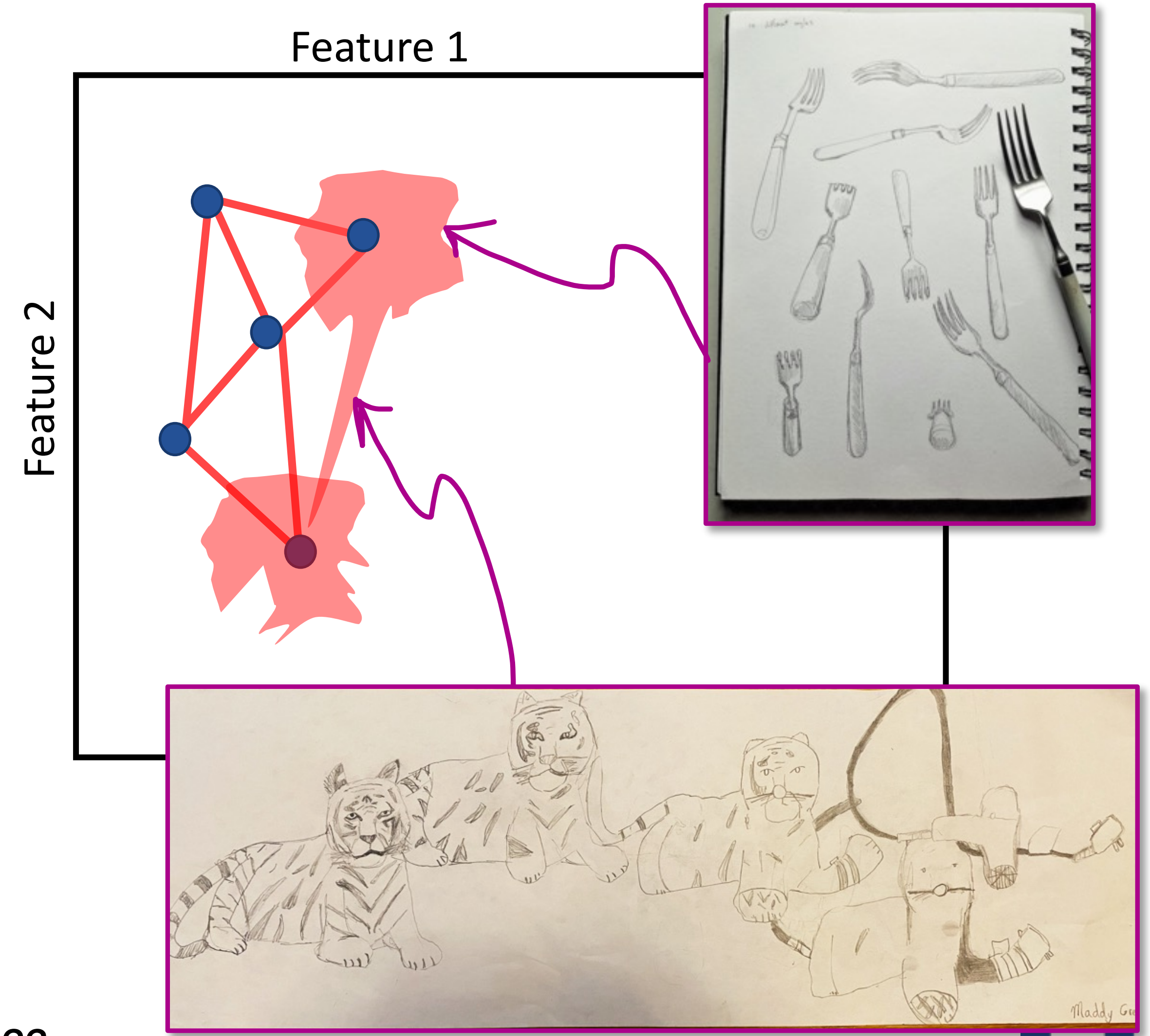
http://2d-preston.weebly.com/hand-transformation.html

Creating an ensemble

Two approaches to synthetic data ensemble design

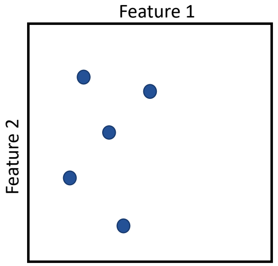
1. **Targeted**: targeted sparse, discrete sampling (matches empirical data collect)

2. **Exploratory:** broad, controllable sampling (enables tailored training and testing)



http://2d-preston.weebly.com/hand-transformation.html

Method 1: match empirical



The UK DfT funded an effort to:

"develop a means by which synthetically created data could be created and validated as to be identical and indistinguishable from data created by scanning real objects in security systems"

QUADRIDOX

- Generated ~2,000 synthetic bag sinograms using QSim RT and the multi-energy, fixed gantry SureScan x1000 EDS model

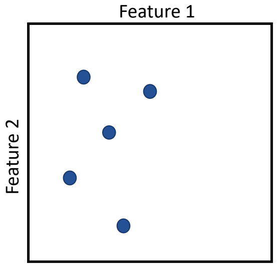
SURESCAN

- Helped process the empirical & synthetic data using the SureScan reconstruction algorithm and ATR

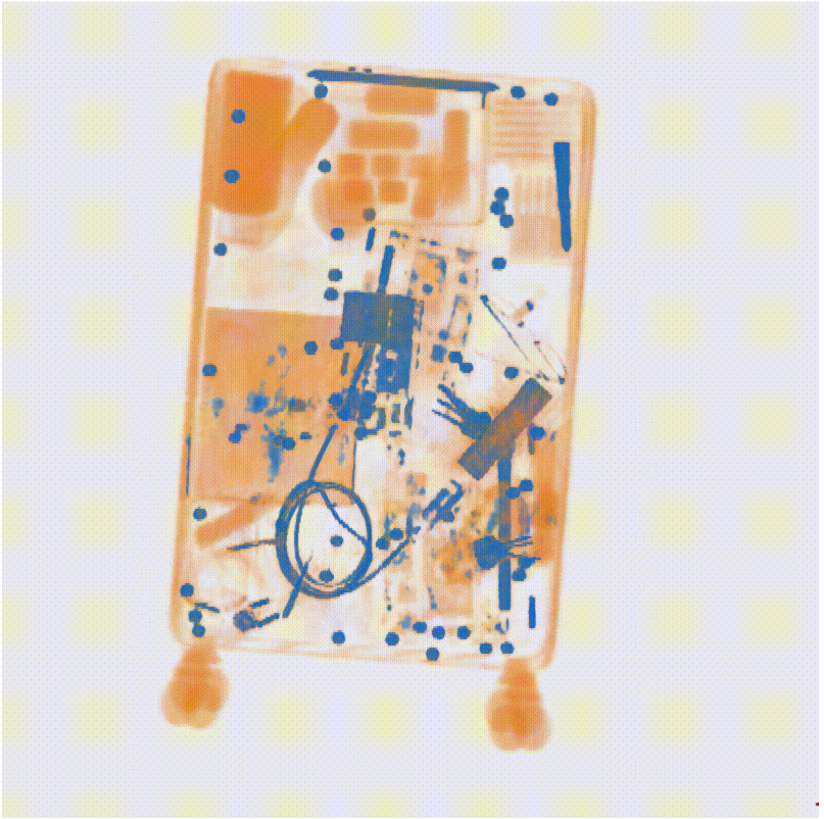
iconal

- Collected ~2000 empirical bag sinograms and processes the data using the SureScan reconstruction algorithm and ATR
- Compared the synthetic and empirical data via
  - Comparison of key properties of the bag data
  - Human observer studies
  - ATR performance metrics

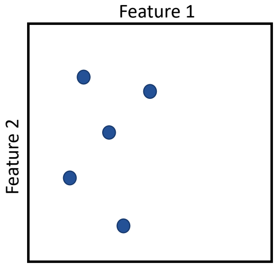
Method 1: match empirical



Empirical vs synthetic CT bag images



Method 1: match empirical

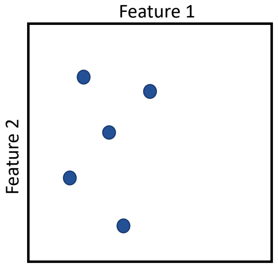


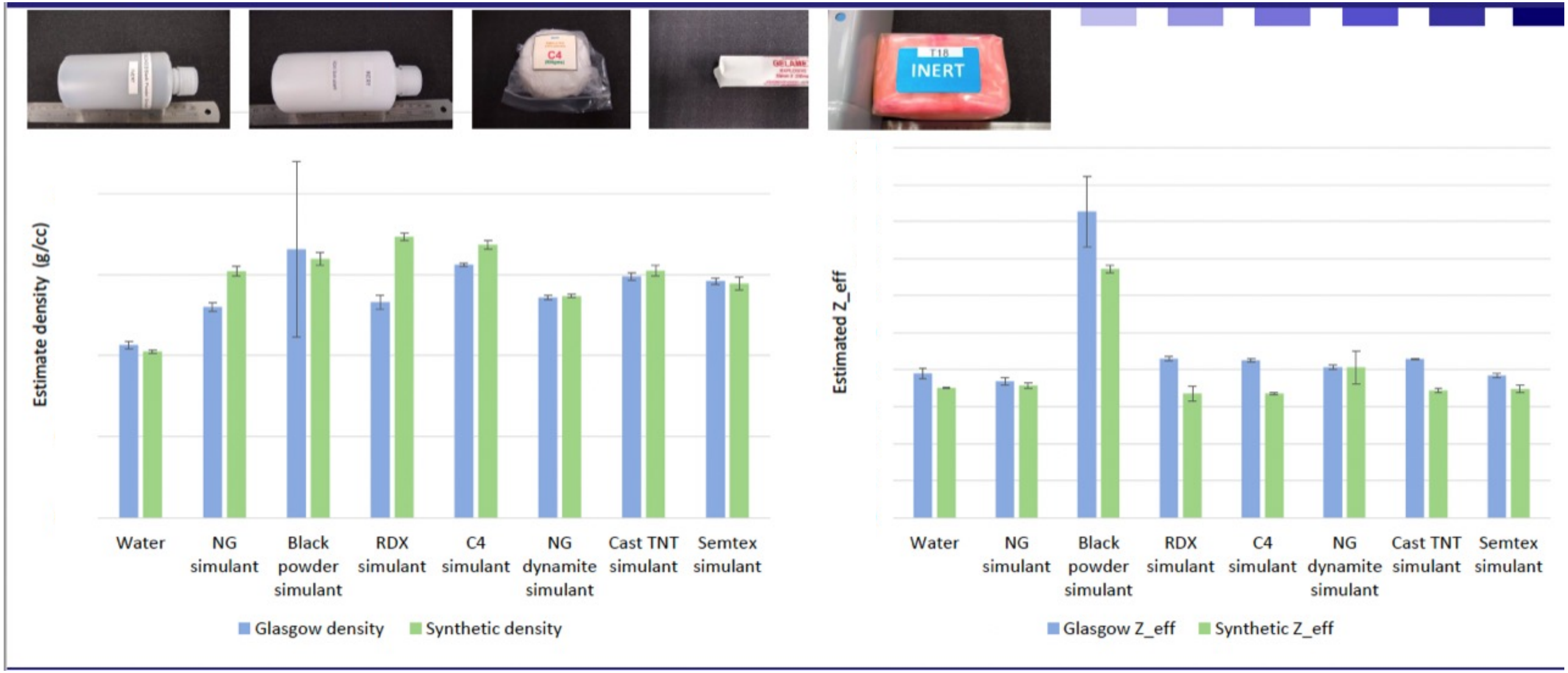
Empirical vs synthetic CT bag images

Empirical

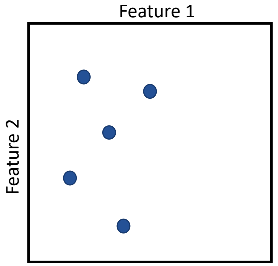
Synthetic

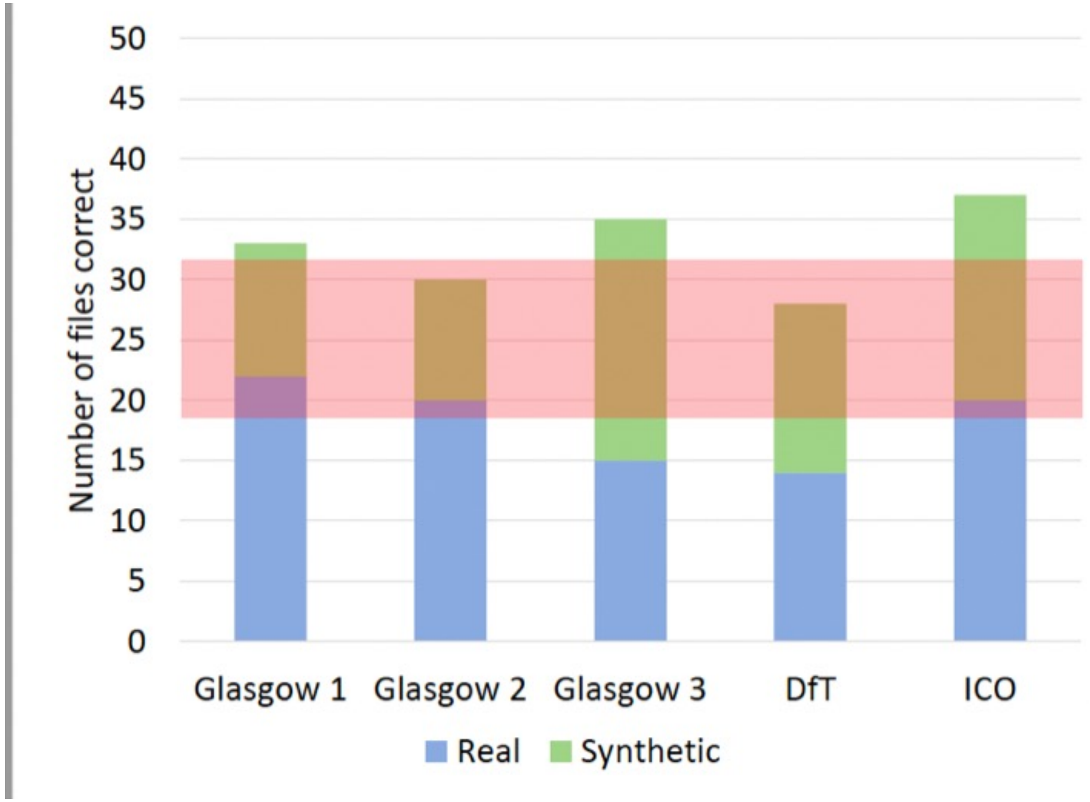
Method 1: matching features





Method 1: human observer study





All screeners evaluated ***50 files*** (25 real, 25 synthetic, randomly selected)

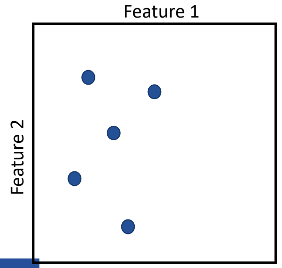
Average:

Guessed 30 out of 50 files correctly (58% of synthetic correct, 73% of real correct)

For random guesses at 95% confidence, number of correct guesses should be 18-32

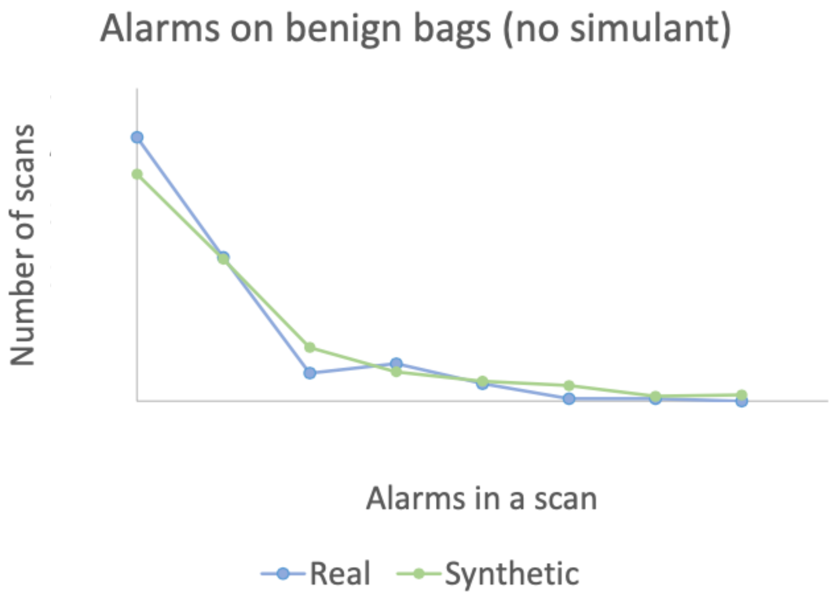
Our synthetic data is effectively indistinguishable from empirical data to the human observer

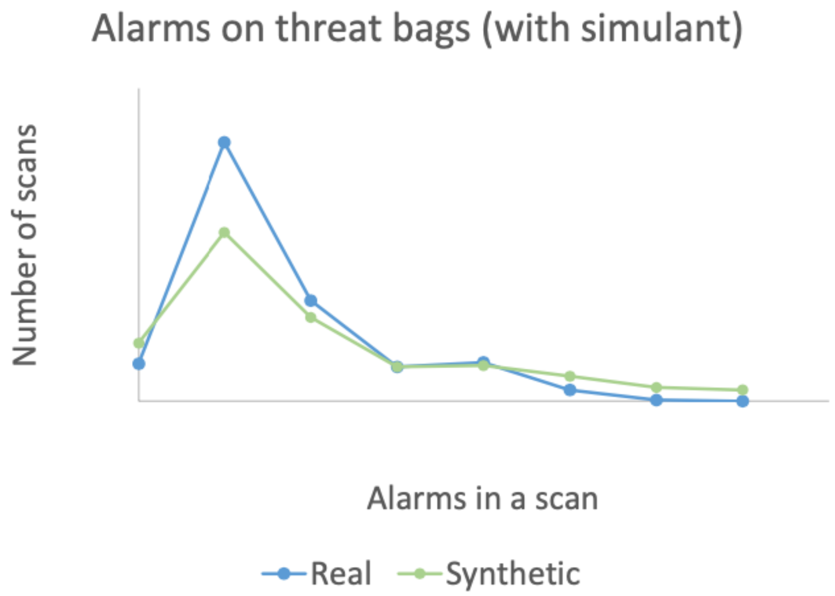
Method 1: validation via ATR



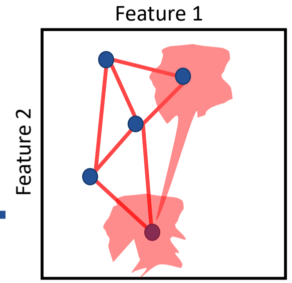
A certified checked baggage EDS ATR study showed that the algorithm performs similarly on synthetic and empirical data across the ensemble

Algorithm study with a certified ATR





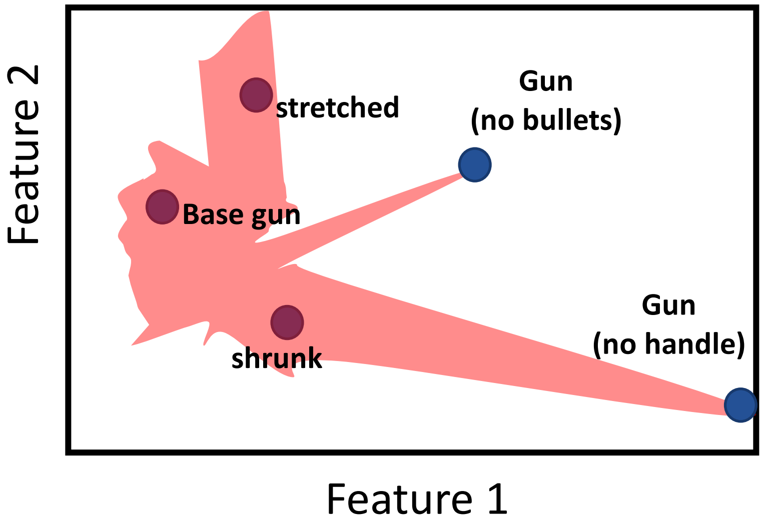
Method 2: beyond empirical



Don't stop at simply trying to create a digital copy of an empirical ensemble – **create the data that you need for your task.**

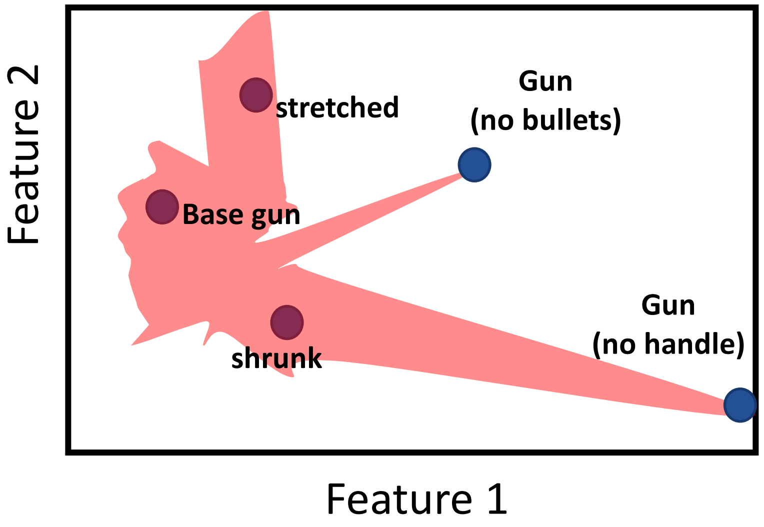
Method 2: Shape perturbations

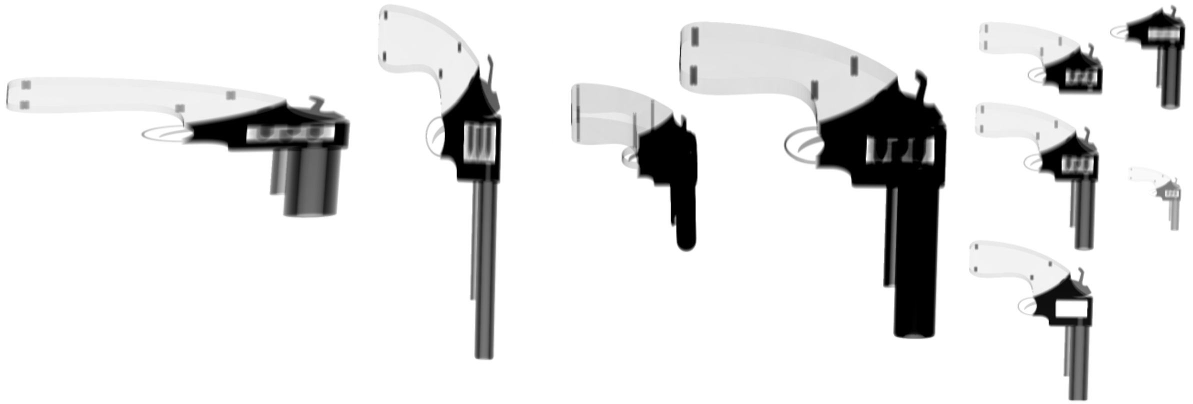
- **Scenario:** arbitrarily vary the geometry (external and internal) of an object
  - Scaling (up/down and uniform/non-uniform)
  - Remove/add components
  - Modify shape
- **Examples:** gun



Method 2: Shape perturbations

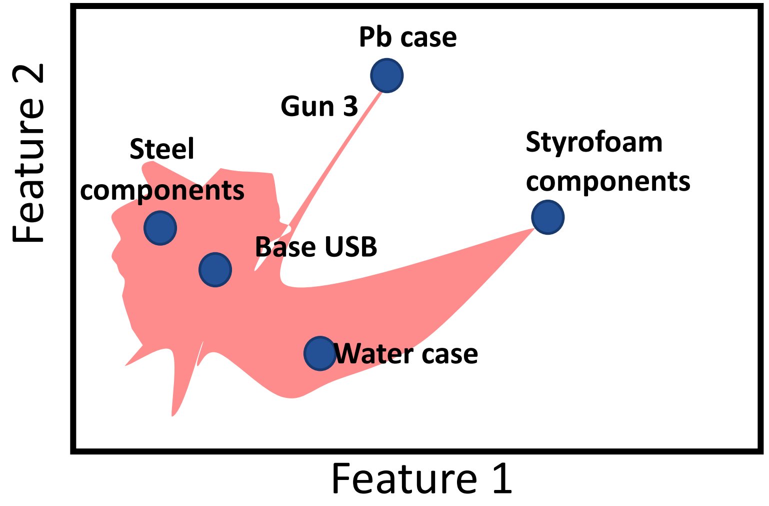
- **Scenario:** arbitrarily vary the geometry (external and internal) of an object
  - Scaling (up/down and uniform/non-uniform)
  - Remove/add components
  - Modify shape
- **Examples:** gun
  - Longer barrel or handle
  - Missing handle
  - No bullets
  - 2x larger or smaller





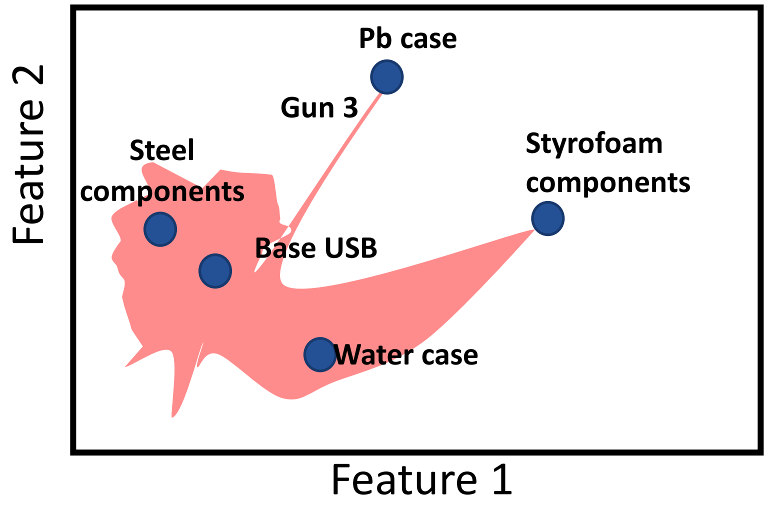
Method 2: Material substitutions

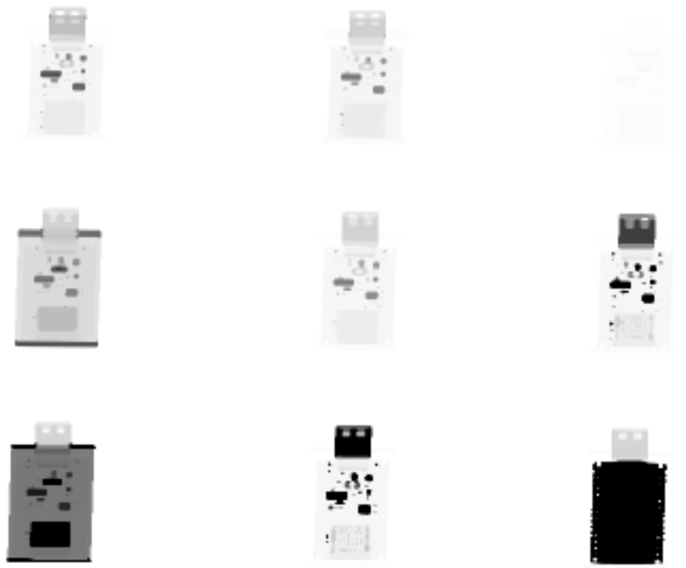
- **Scenario:** vary the material composition of objects/components
  - Assign different specific materials
  - Design arbitrary materials
  - Vary density, Zeff, k-edges, scatter strengths
  - Control shield, artifacts, decouple shape/material
- **Example:** USB dongle with fixed geometry



Method 2: Material substitutions

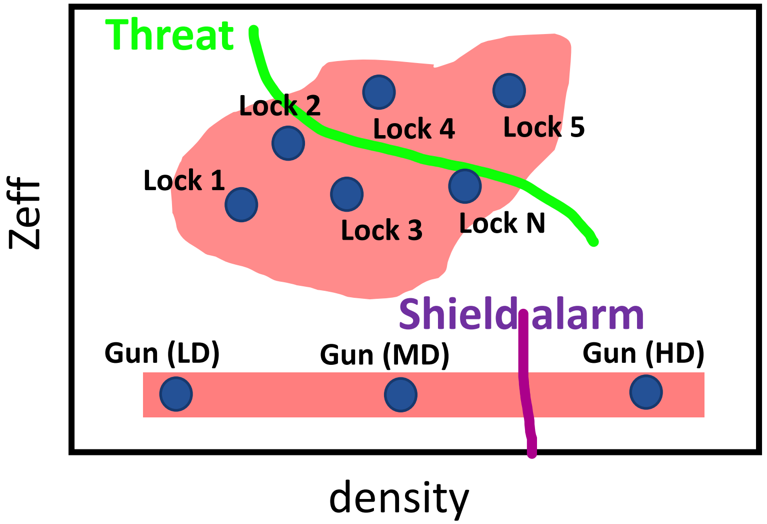
- **Scenario:** vary the material composition of objects/components
  - Assign different specific materials
  - Design arbitrary materials
  - Vary density, Zeff, k-edges, scatter strengths
  - Control shield, artifacts, decouple shape/material
- **Example:** USB dongle with fixed geometry
  - Different exterior material
  - Different metal connector
  - Different internal electronic materials
  - “impossible” scenarios (e.g., water enclosure)





Method 2: Bracketed studies

- **Scenario:** sweep material parameters
  - Arbitrarily vary density
  - Arbitrarily vary energy-dependent attenuation (~Zeff)
  - Map out ATR decision boundaries
  - Explore corner cases and packing dependences
- **Example:** USB dongle with fixed geometry
  - Gun made of steel with various densities
  - Lock with a continuum of density/Zeff value

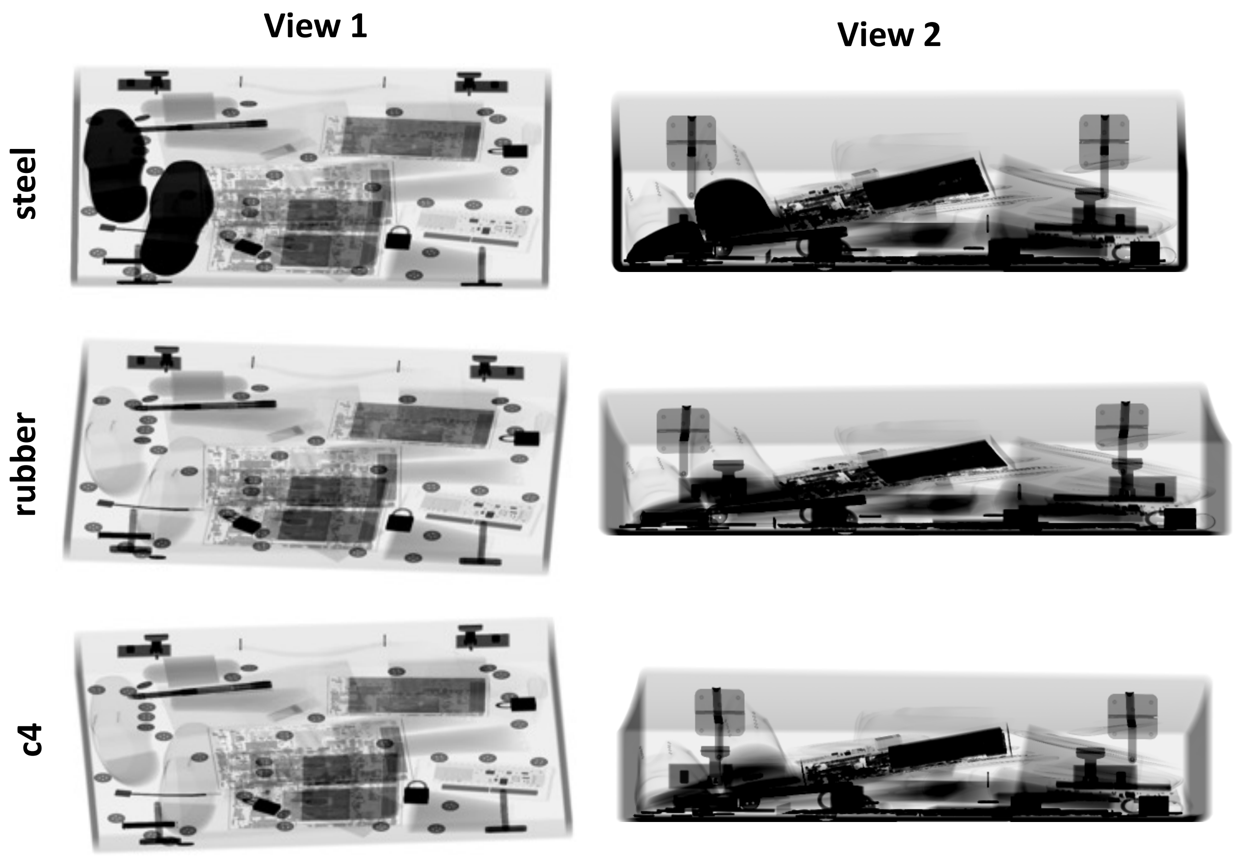


Locks

Guns (steel with densities 2, 5, and 8 g/cm³)

Method 2: Static clutter

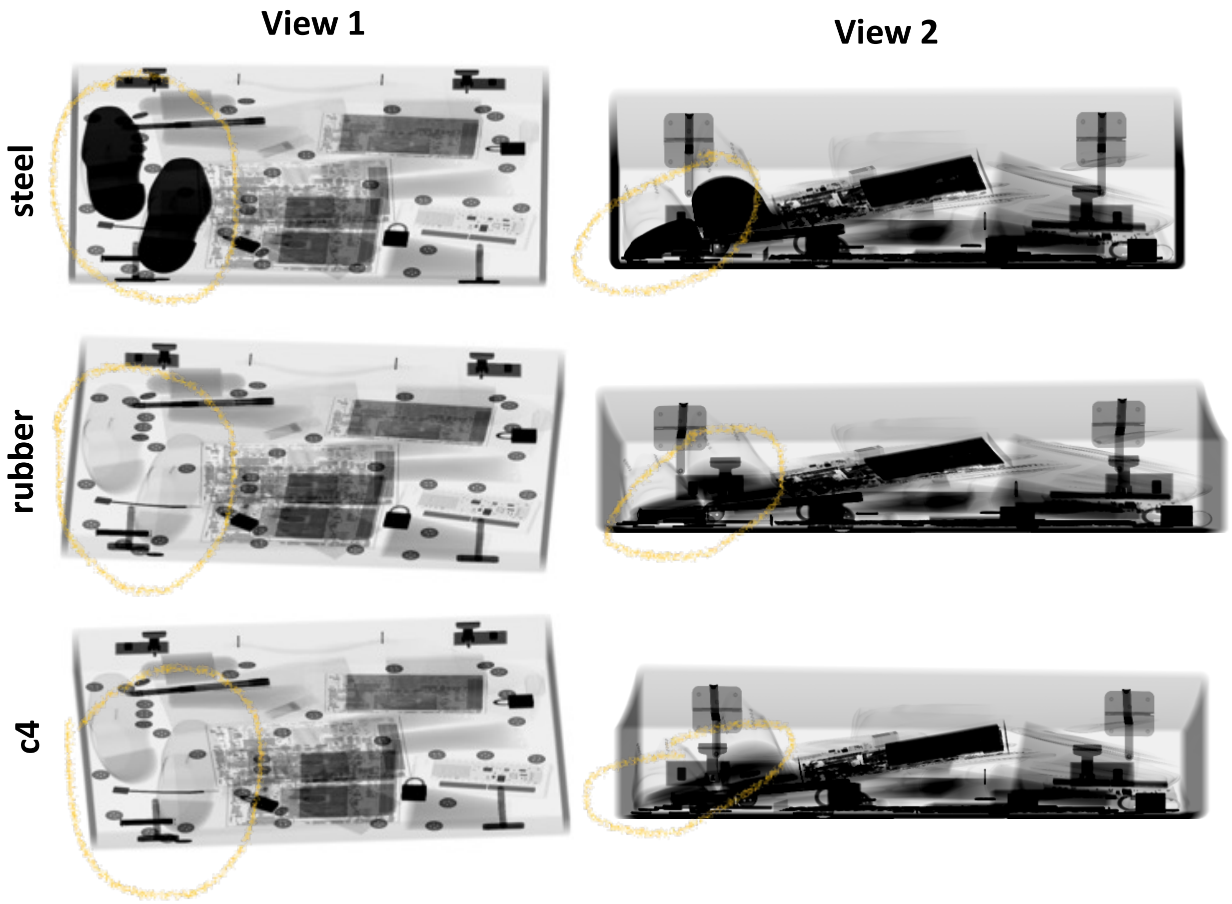
- **Scenario:** keep clutter fixed and vary only the item of interest
  - Modify target size/shape
  - Modify target material
  - Modify concealment
  - Rotate/shift entire bag
- **Example:** briefcase with shoes
  - Modify shoe sole material
  - Rotate bag



QUADRIDOX

Method 2: Static clutter

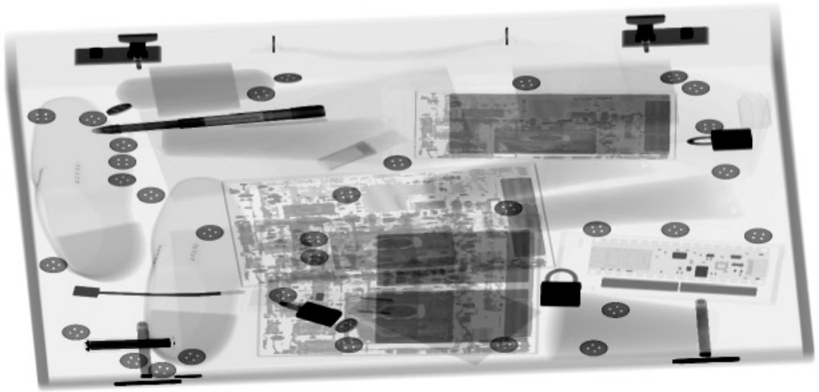
- **Scenario:** keep clutter fixed and vary only the item of interest
  - Modify target size/shape
  - Modify target material
  - Modify concealment
  - Rotate/shift entire bag
- **Example:** briefcase with shoes
  - Modify shoe material
  - Rotate bag



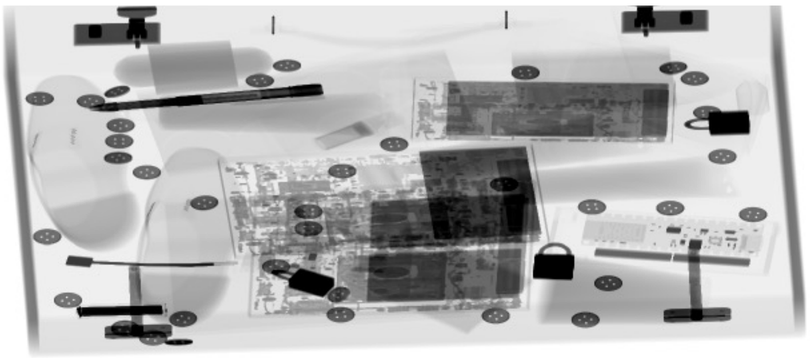
Method 2: Static clutter

- **Scenario:** keep clutter fixed and vary only the item of interest
  - Modify target size/shape
  - Modify target material
  - Modify concealment
  - Rotate/shift entire bag
- **Example:** briefcase with laptop
  - Modify shoe material

c4



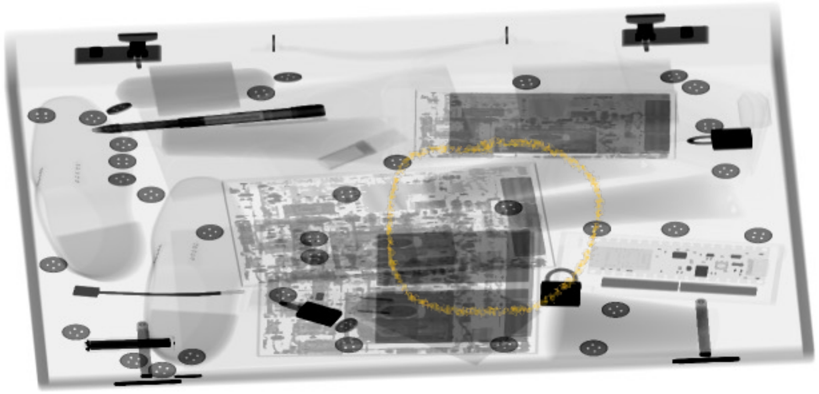
battery



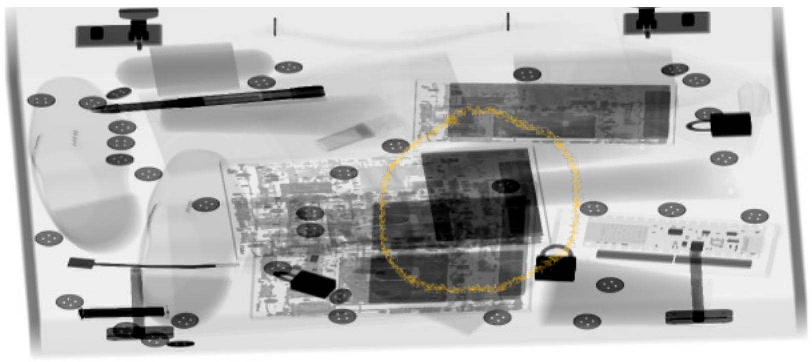
Method 2: Static clutter

- **Scenario:** keep clutter fixed and vary only the item of interest
    - Modify target size/shape
    - Modify target material
    - Modify concealment
    - Rotate/shift entire bag
- **Example:** briefcase with laptop
    - Modify shoe material

c4

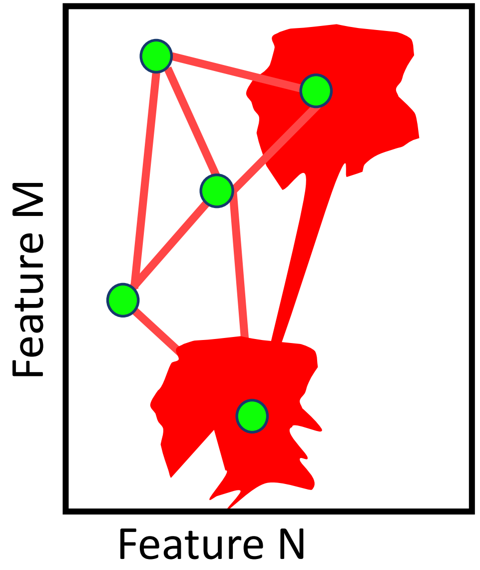


battery



Summary

- Physics-based synthetic data is critical to algorithm development and testing
- Our synthetic data framework rapidly produces high-fidelity X-ray data
- Validation of synthetic data against empirical data is critical **BUT the real benefits of synthetic data stem from ways that it can differ from empirical data**
    - **Bracketed/continuous feature studies to map decision boundaries**
    - **Geometric perturbations**
    - **Material variations**
    - **Perfect labeling**
    - **Consistent clutter**
    - Identical bag in different systems
    - Controllable noise/detector response



Questions?