SAND2022-3884C

Sandia National Laboratories

Exceptional service in the national interest

# Resilient adjudication in non-intrusive inspection with hierarchical object and anomaly detection

**Dan Krofcheck**
Sandia National Laboratories

U.S. DEPARTMENT OF ENERGY NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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.

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.

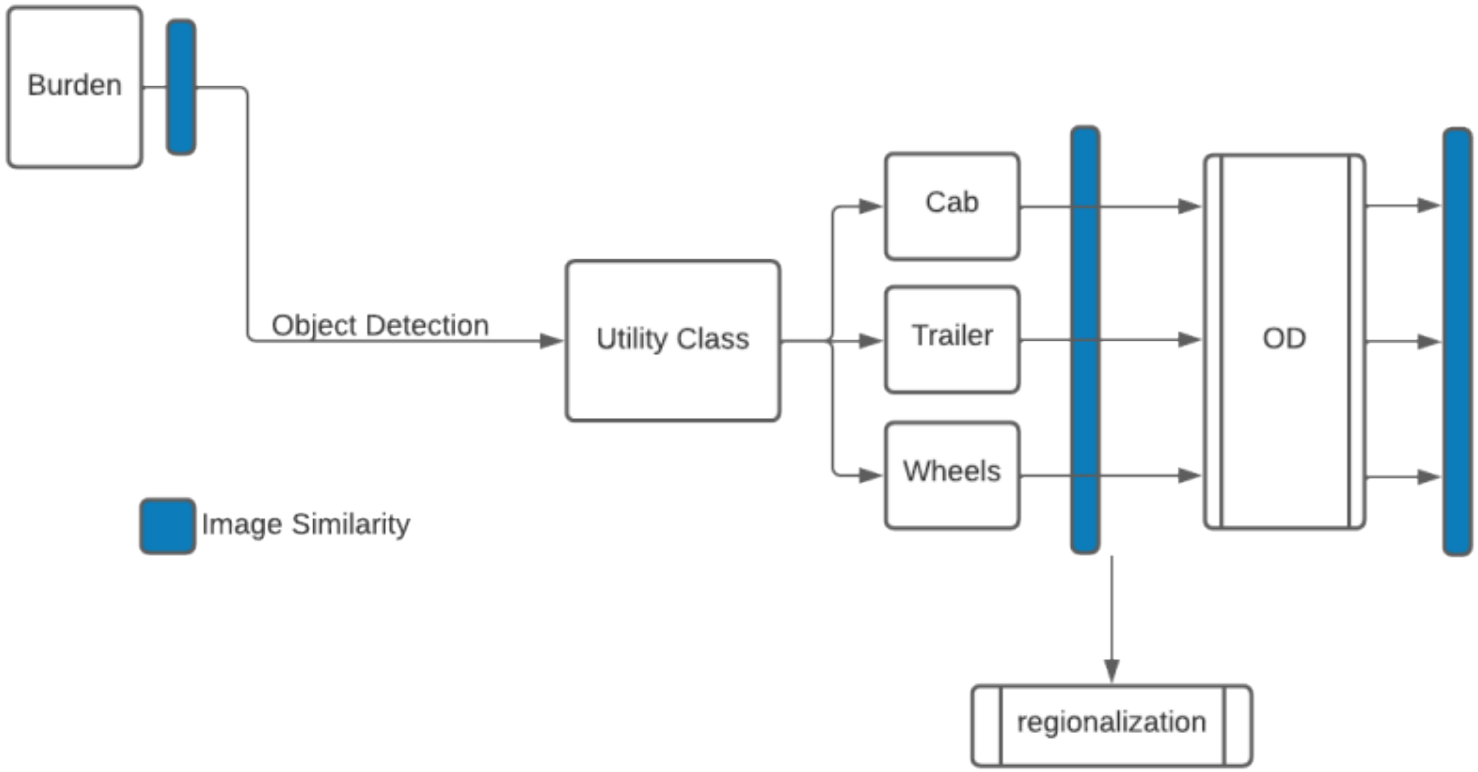
# Trusted and Trust-In AI systems

In high consequence environments, Trust between operators and AI becomes critical

**Trusted systems should be:**

- Resilient
- Robust
- Adaptable
- Ergonomic

# Adjudication from a systems perspective



# Detectron 2 Object Detection

**Facebook object detection API**
Facilitates training collections of object detection models with pre-initialized weights

We trained 8 candidate models on our COCO formatted dataset
Model hyperparameters and training schedule were **NOT** optimized

Data separated into a 90/10 train/test split, after stratifying to minimize class imbalances

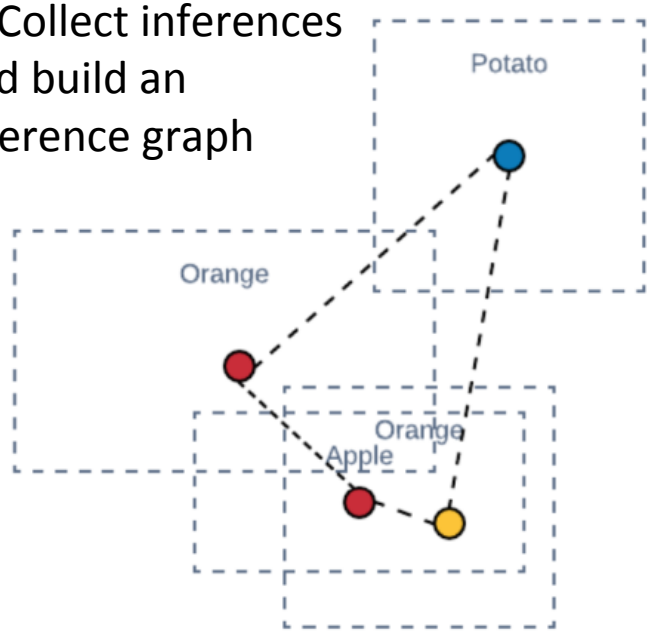
**Dataset**
These whole vehicle scans collected on a Multi-Energy Portal
Hand annotations for 5 classes:

- Trailer
- Cabin
- Tire
- Flatbed
- Target (Mixed material objects of interest)

| Model Name | Model Family | Model Outputs |
|---|---|---|
| R50-FPN-1x | Faster RCNN | Class, Score, Box |
| R50-FPN-3x | Faster RCNN | Class, Score, Box |
| R101-FPN-3x | Faster RCNN | Class, Score, Box |
| RN-R50-1x | RetinaNet | Class, Score, Box |
| RN-R50-3x | RetinaNet | Class, Score, Box |
| R_101_C4_3x | Mask RCNN | Class, Score, Box, Mask |
| R_101_DC5_3x | Mask RCNN | Class, Score, Box, Mask |
| R_101_FPN_3x | Mask RCNN | Class, Score, Box, Mask |

# Graph Theoretic Ensembling

1) Collect inferences and build an inference graph




2) Split inference graph into cliques, and score each clique

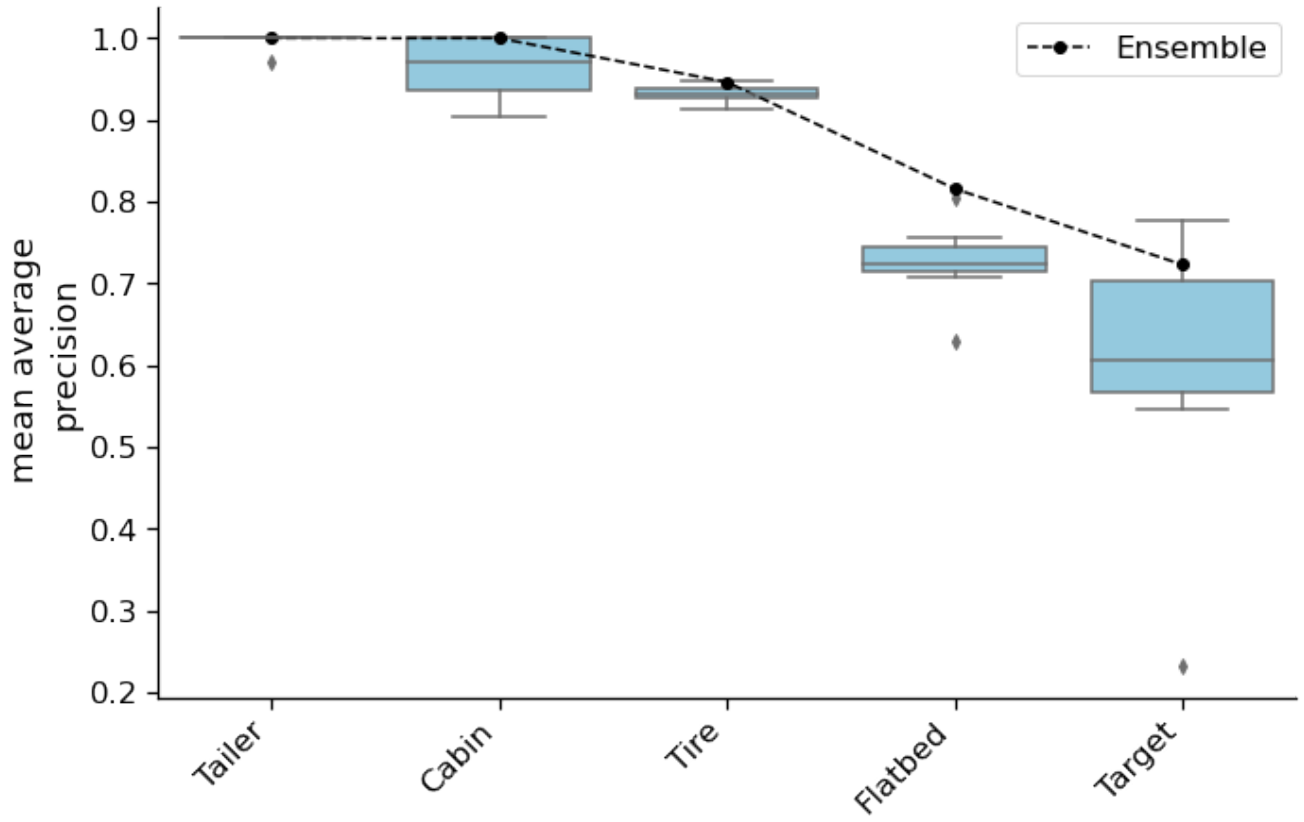
$$1 - \prod_{s} (1 - s)$$

3) If models disagree on a classification, calculate a categorical confidence score to weight clique selection

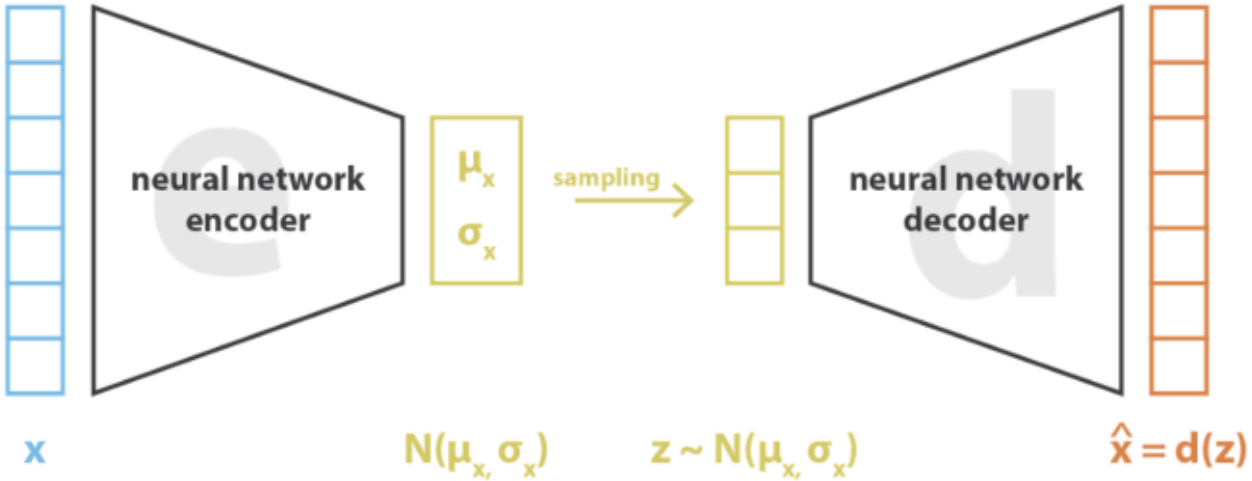
$$\left(1 - \prod_{s \in C} (1 - s)\right) \prod_{s \notin C} (1 - s)$$

**Clique: A fully induced subgraph, or a *complete graph***

# Notional Object Detection Results



# Image Similarity



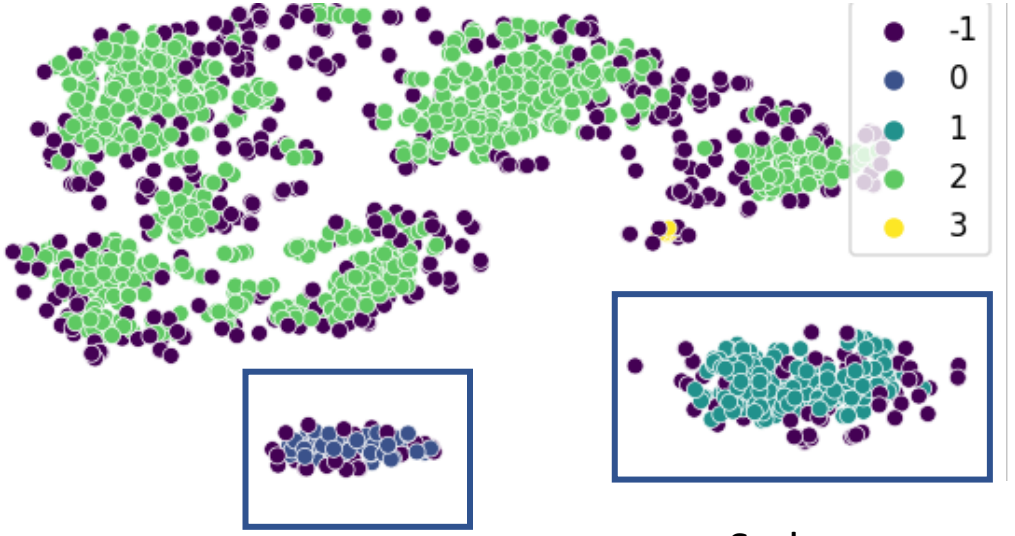
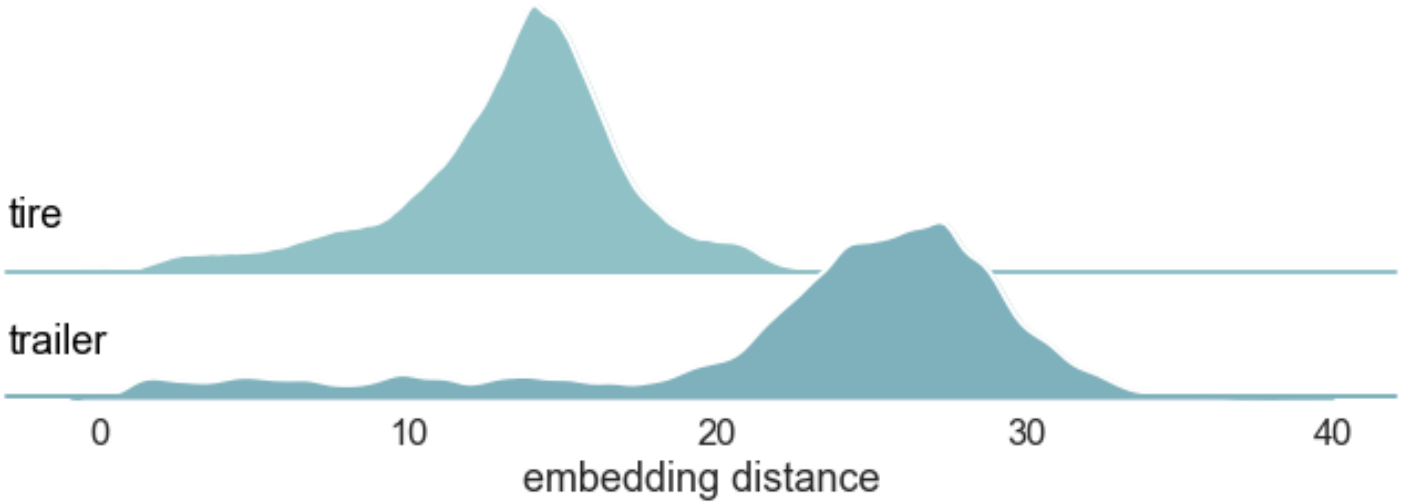
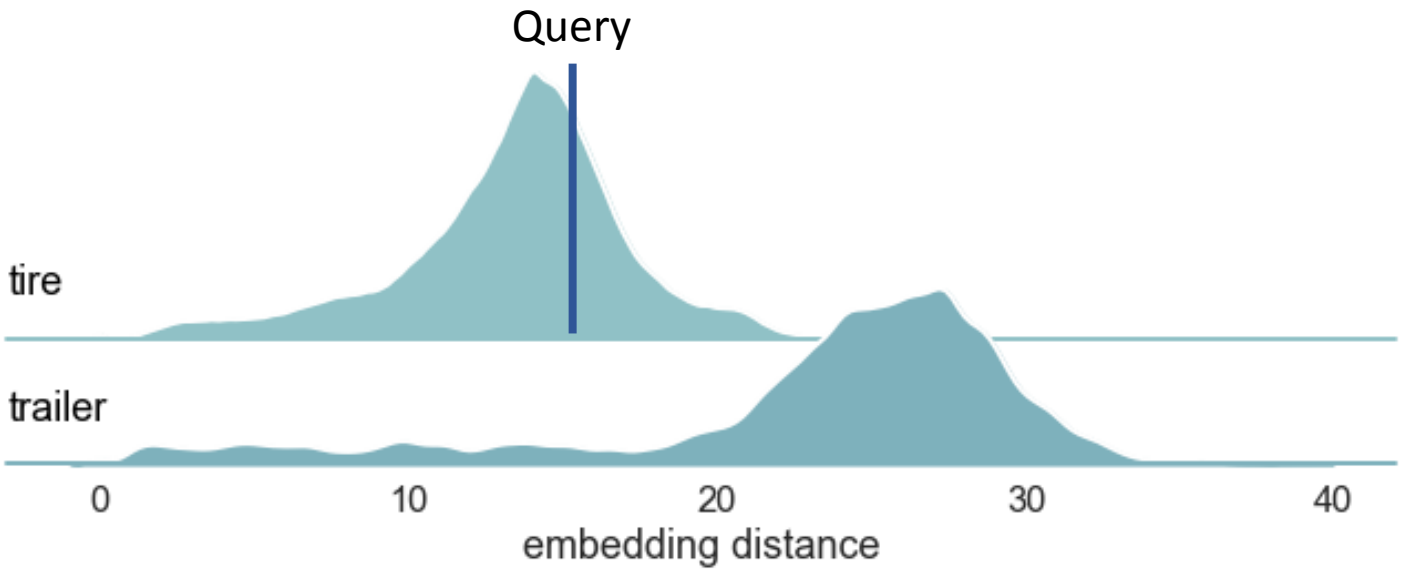




Figure from Joseph Rocca: https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

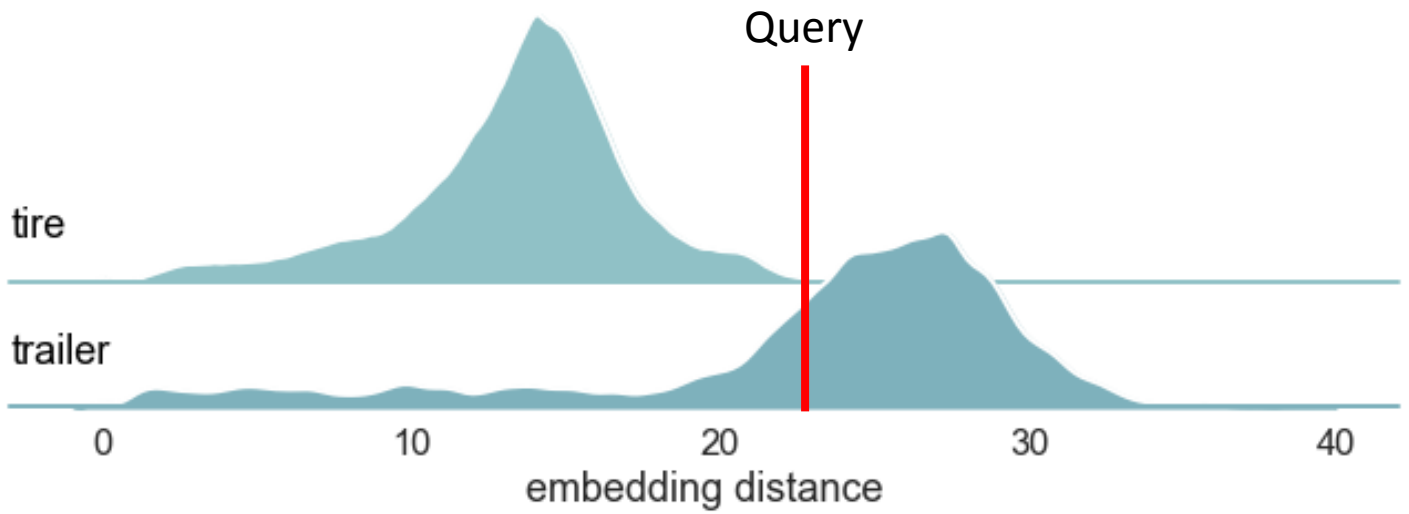
# Image Similarity



# Image Similarity



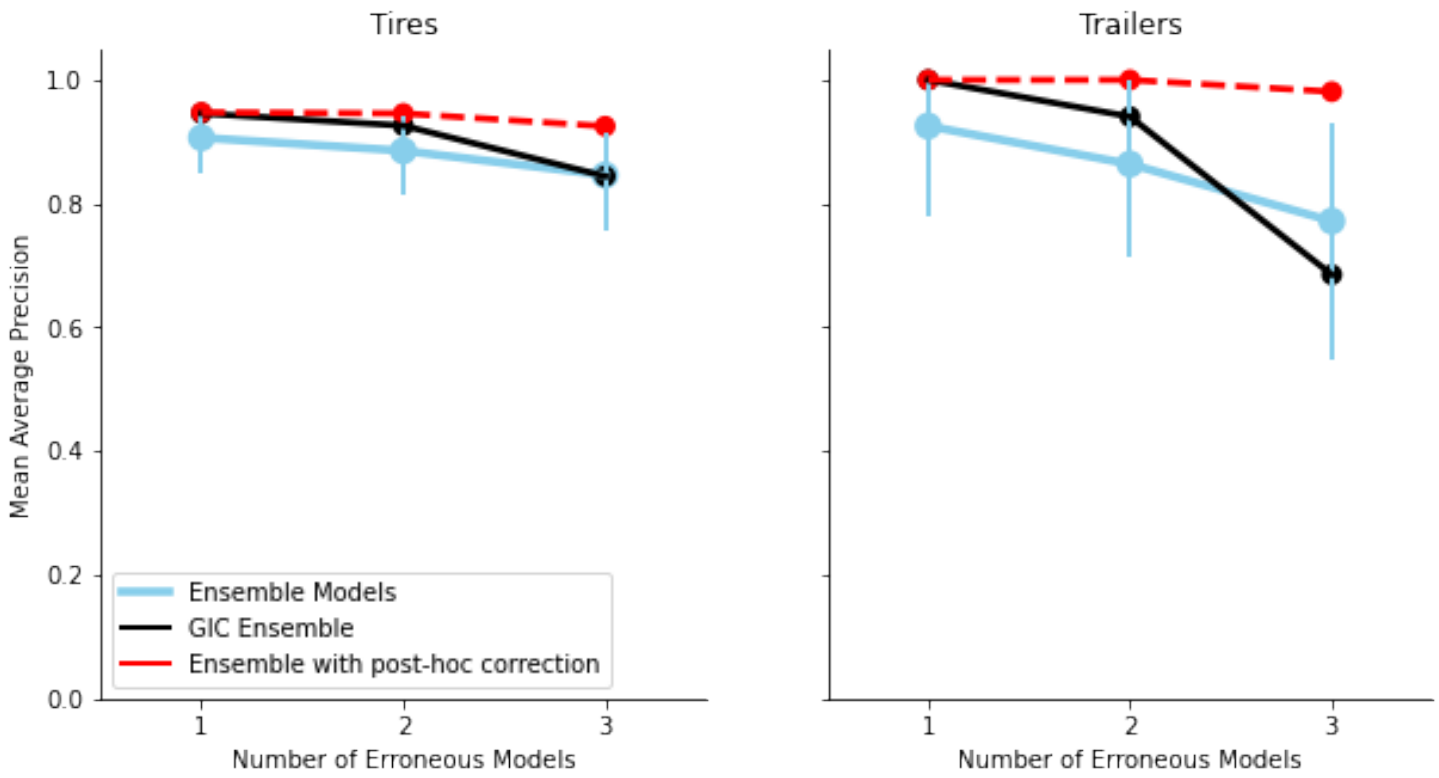
# Image Similarity



# Post-hoc Correction

## Simple experiment:

- Run inference on our test dataset (Blue)
- For each ‘tire’ detection, swap the label to ‘trailer’ (probability 25%), and recalculate mAP (Black)
- Repeat the experiment, but before running the ensemble, expose trailer and tire inferences to the post-hoc similarity correction. Deprecate the confidence of potentially erroneous inferences prior to ensembling (Red)



# Conclusions

**Key Takeaways**

- We can develop trusted and resilient systems for NII
- using off the shelf capabilities
- The judicious combination of tools in scalable ways
- will be the key to widespread adoption
- Ensembles of object detection models are fast to train and implement
- Post-hoc correction using image similarity or any other distribution style test may provide orthogonal checks and balances