SAND2022-2836C

Sandia National Laboratories

Exceptional service in the national interest

Low-shot, Semi-supervised, Uncertainty Quantified Learning with Hyperspectral Imagery Data

Kathryn Gray

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.

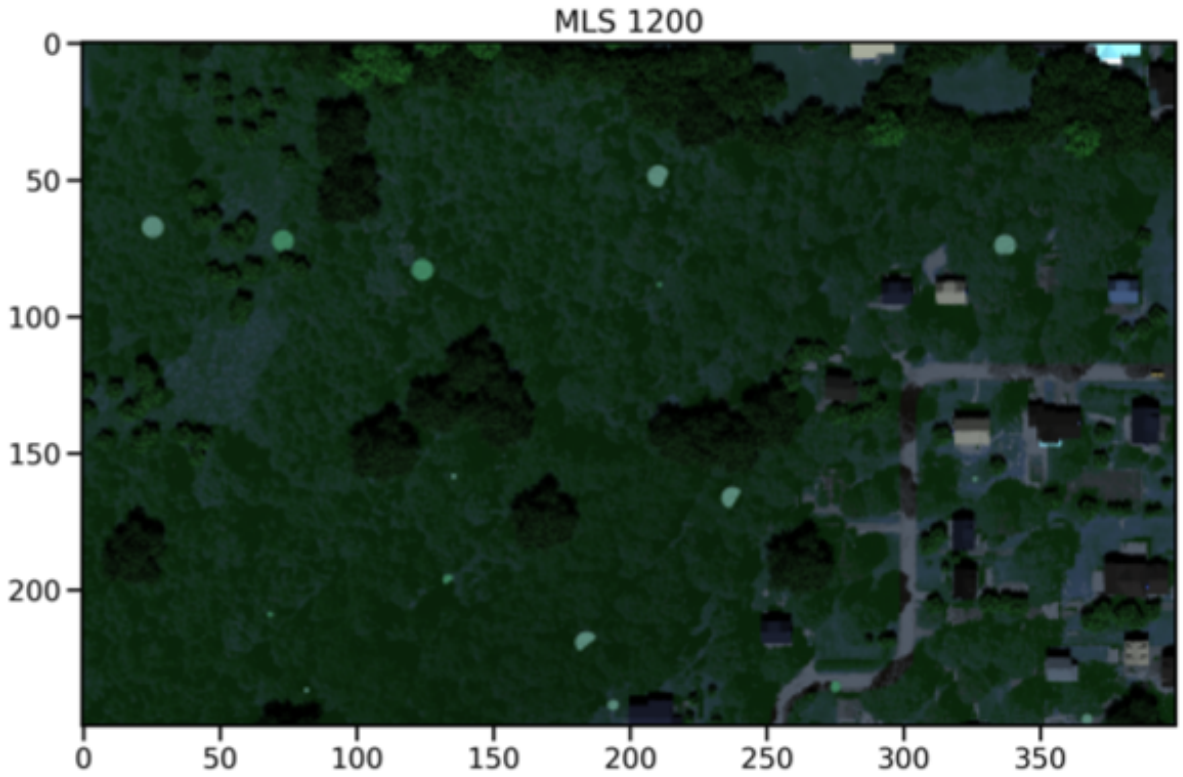
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Hyperspectral Imagery Data

- Contains the following:
  - X, Y data
  - Reflectance
  - Wavelengths

Hyperspectral Imagery Data

- Targets: Train on Green Paint 1
- Split up by x,y values
- Classify each of these by whether it contains target



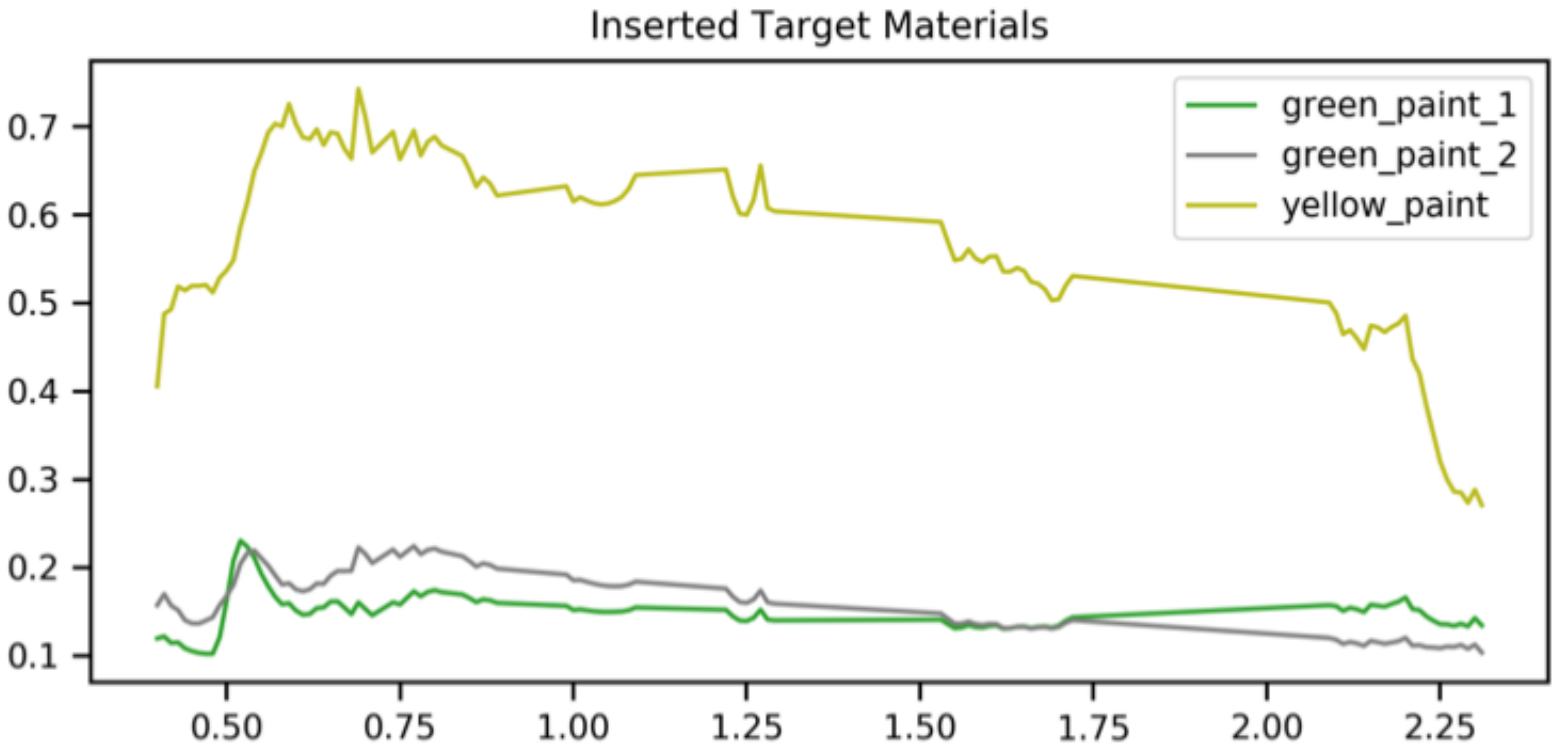
Model Requirements

1. Low-Shot Learning
2. Semi-Supervised Learning
3. Uncertainty Quantification

Low-Shot Learning

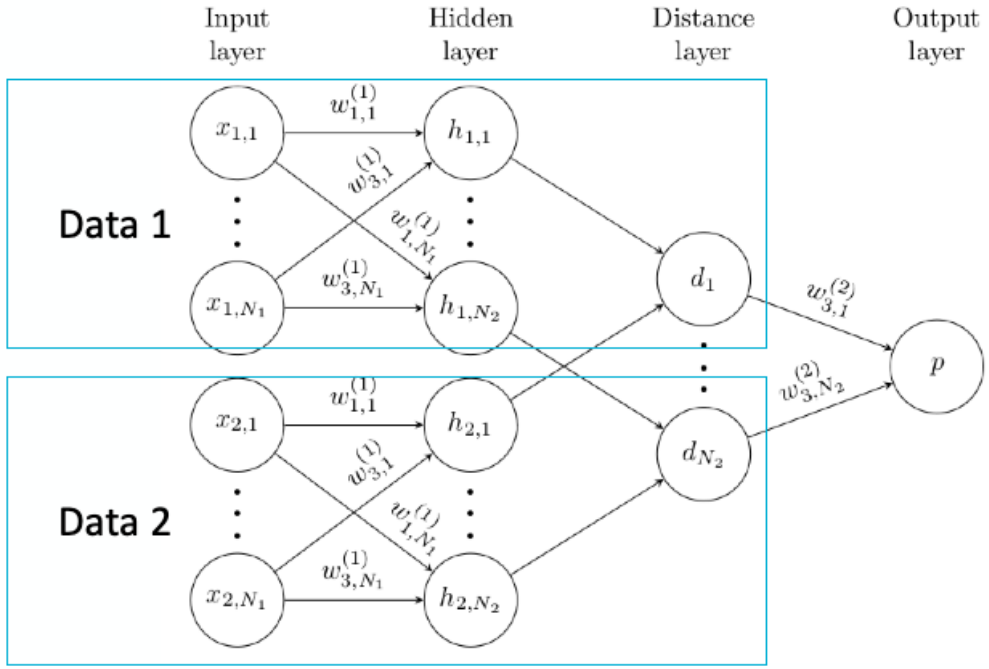
What is Low-Shot Learning?

We want to train on one material and extend to new ones



Siamese Networks

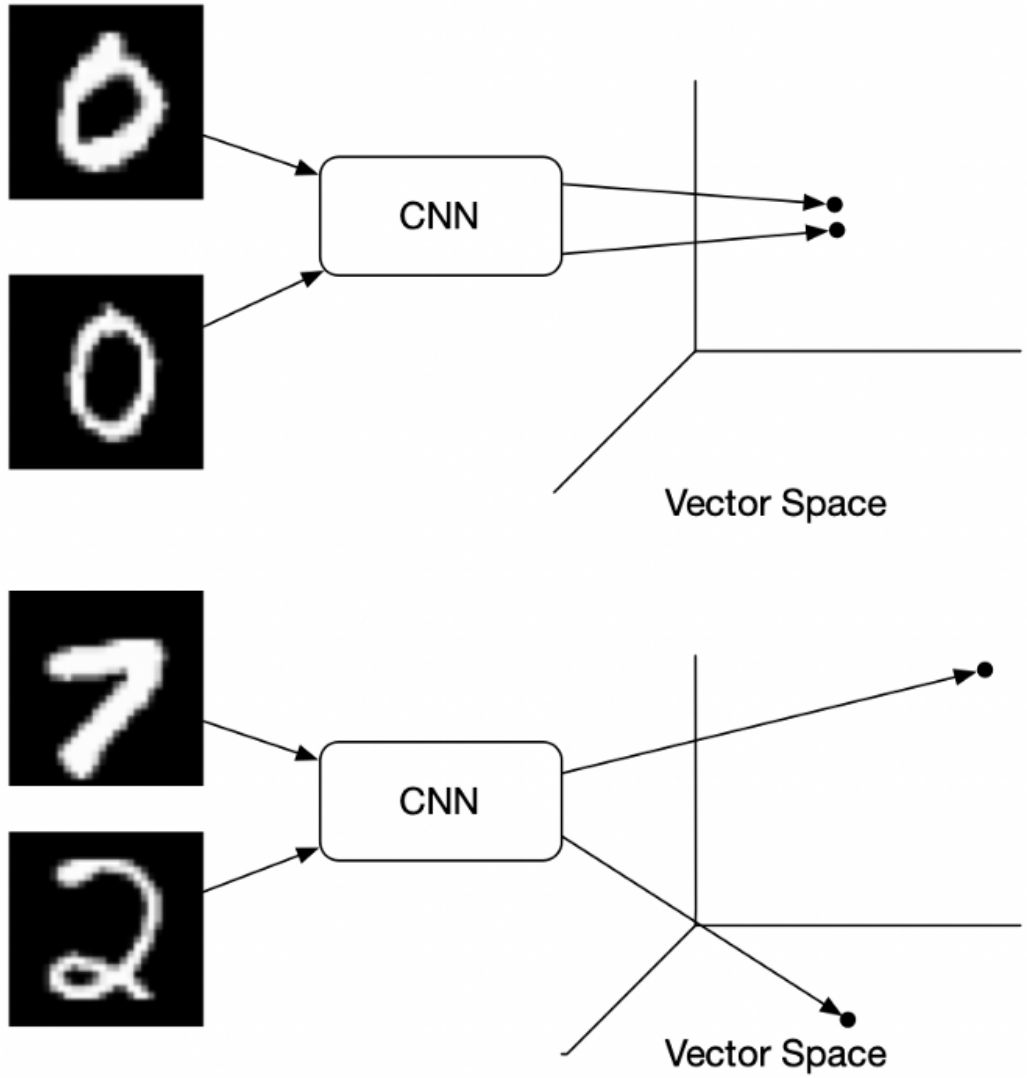
- Learn Distances between datapoints
- Want:
  - Small distances between similar datapoints
  - Large distances from dissimilar



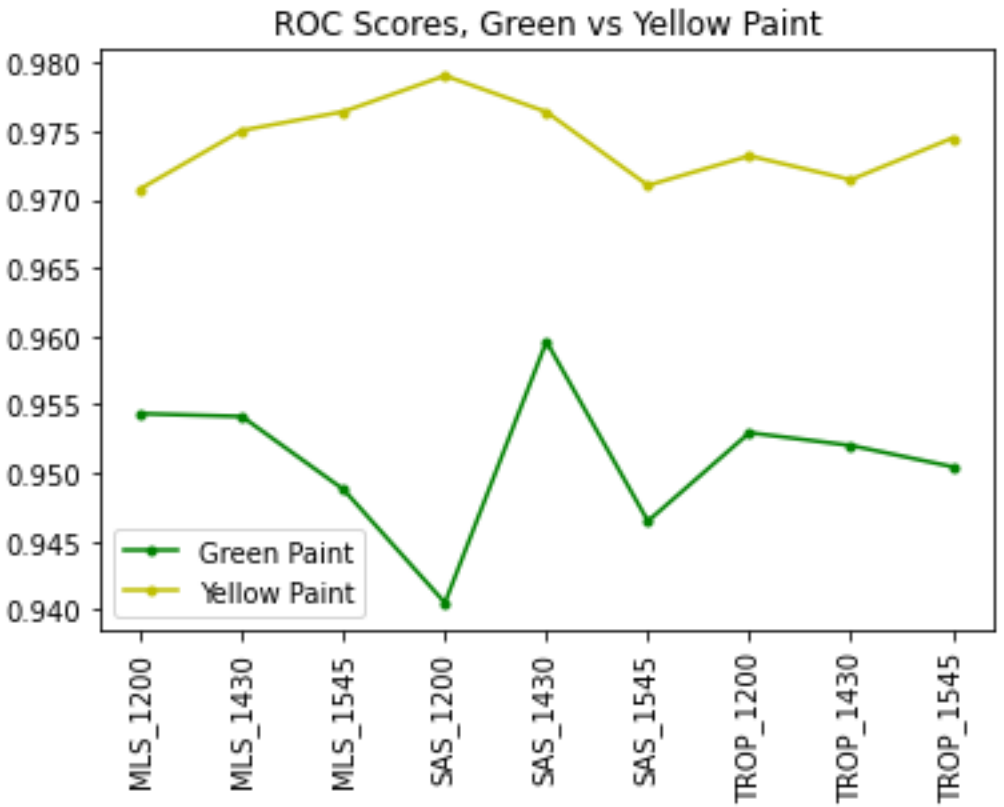
Siamese Networks - Idea

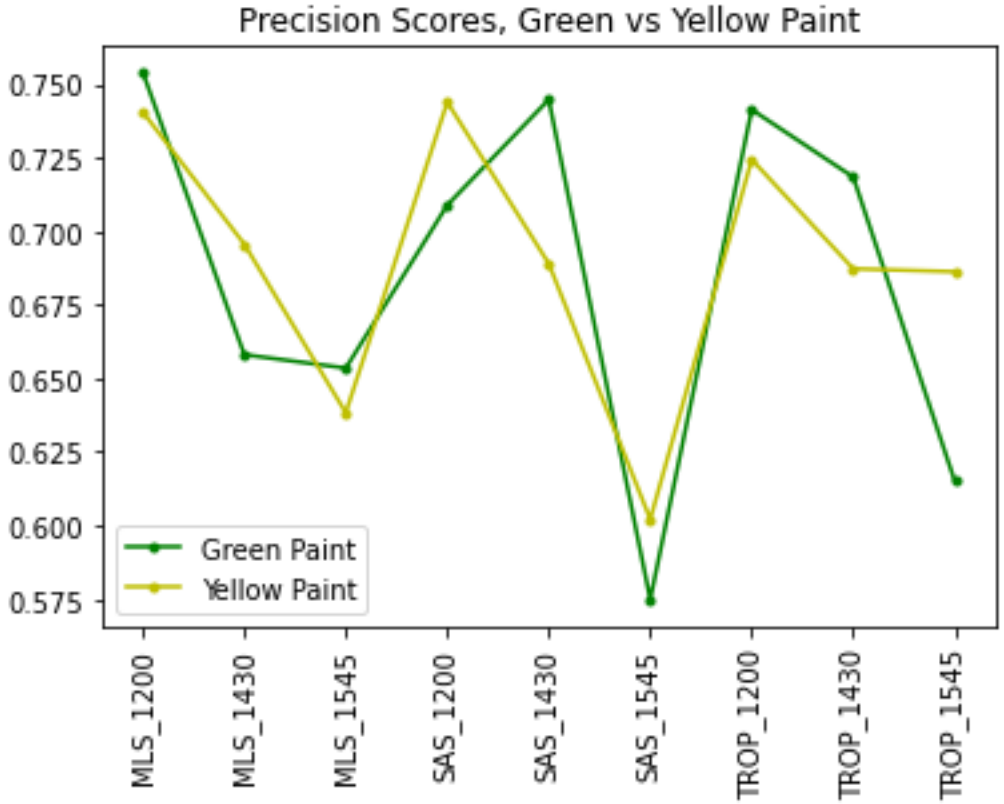
- End up with an embedding space
- Hopefully, we will get any similar datapoints

  to a similar location in embedding space



Low-Shot Learning Results

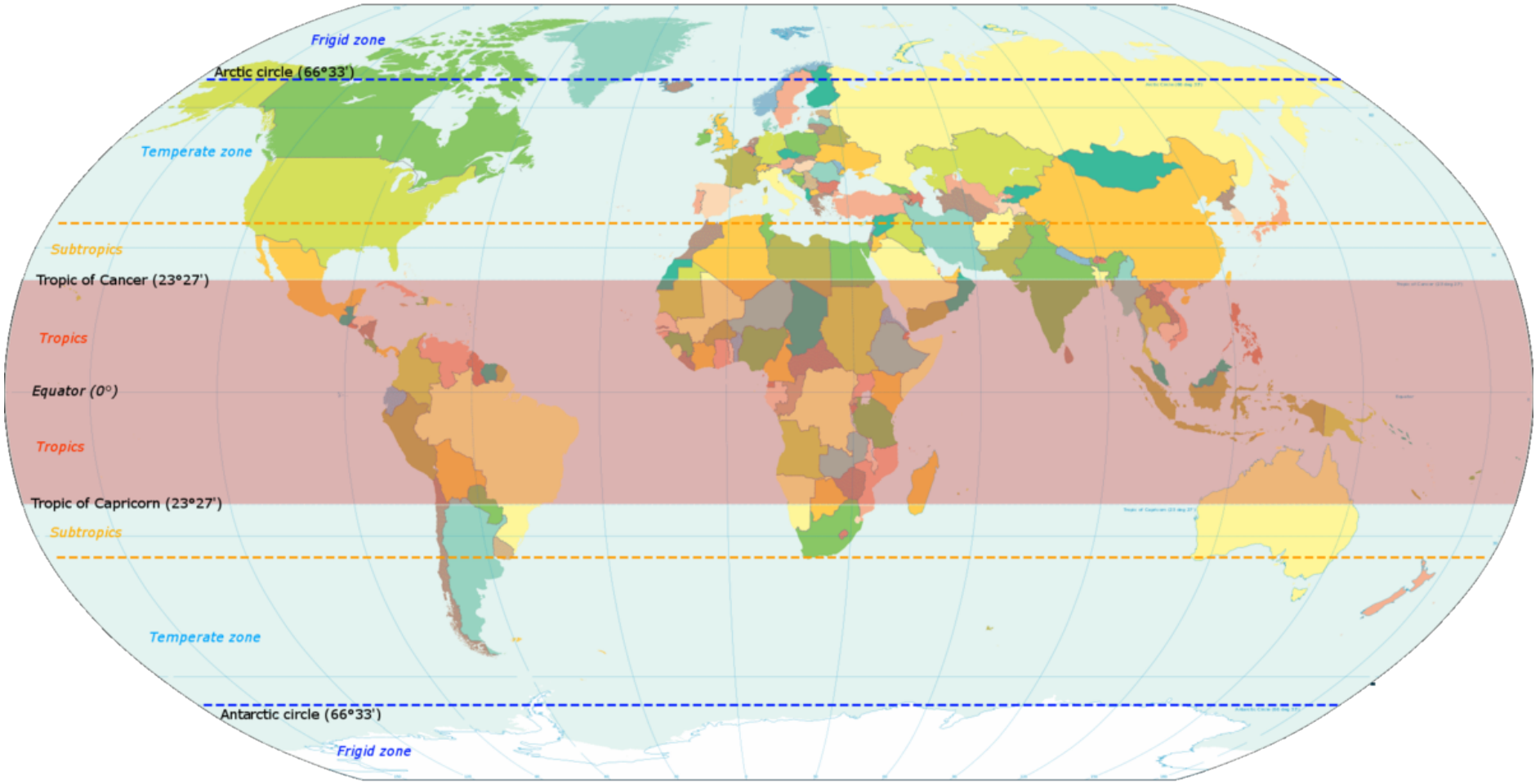




Semi-Supervised Learning

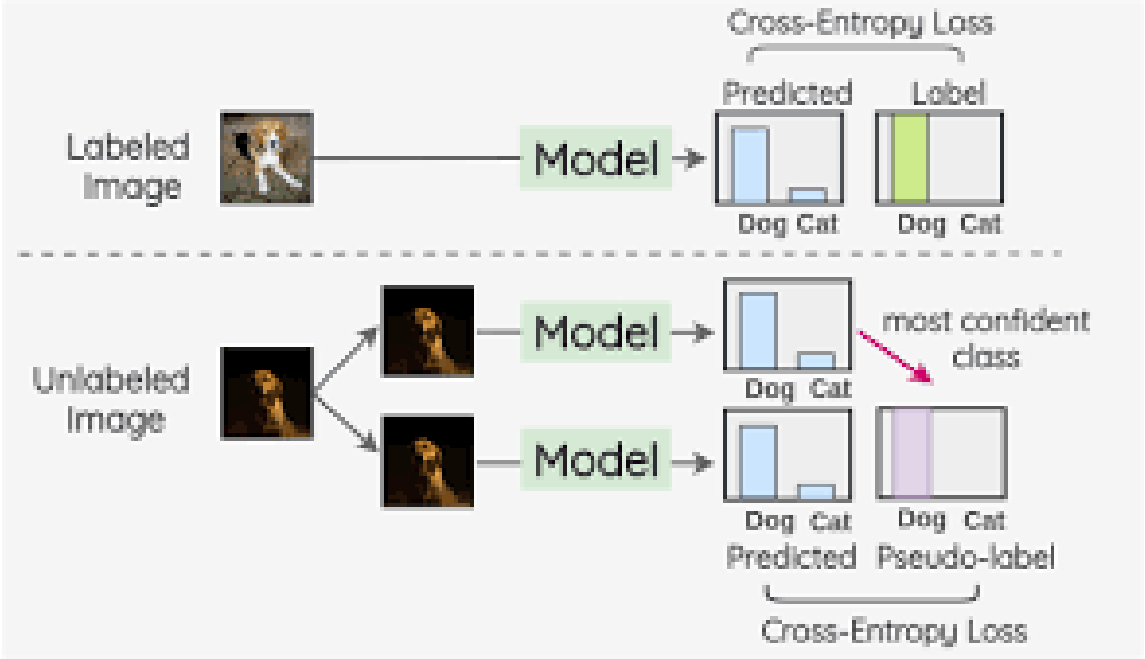
What is Semi-Supervised Learning?

We have several datasets, not all of them are labeled. We want to be able to use this data.



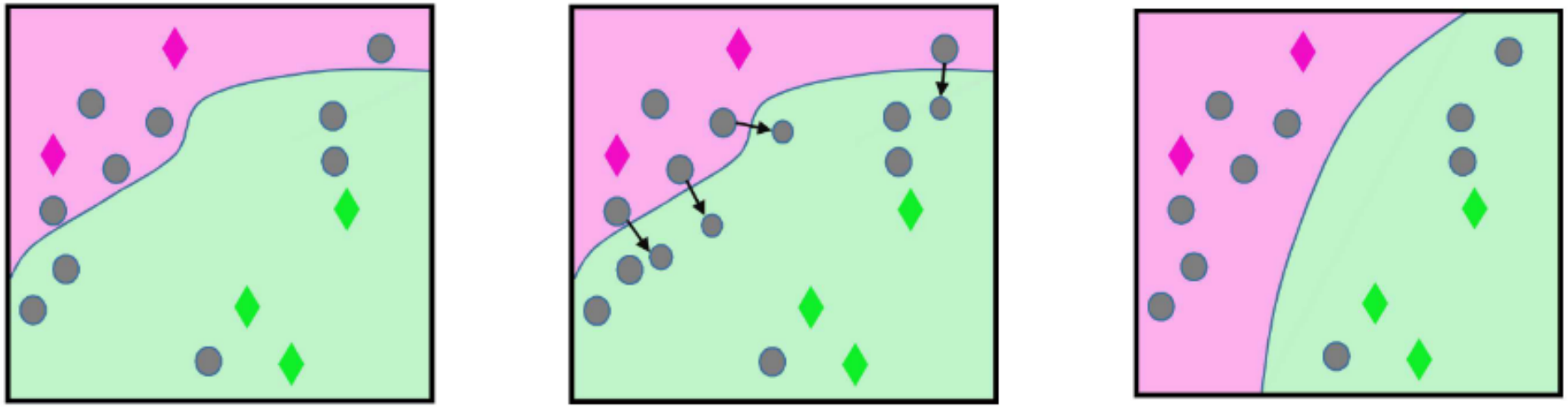
Mean Teacher

- Student and Teacher models
- Teacher updates each round as mean of students



Virtual Adversarial Training

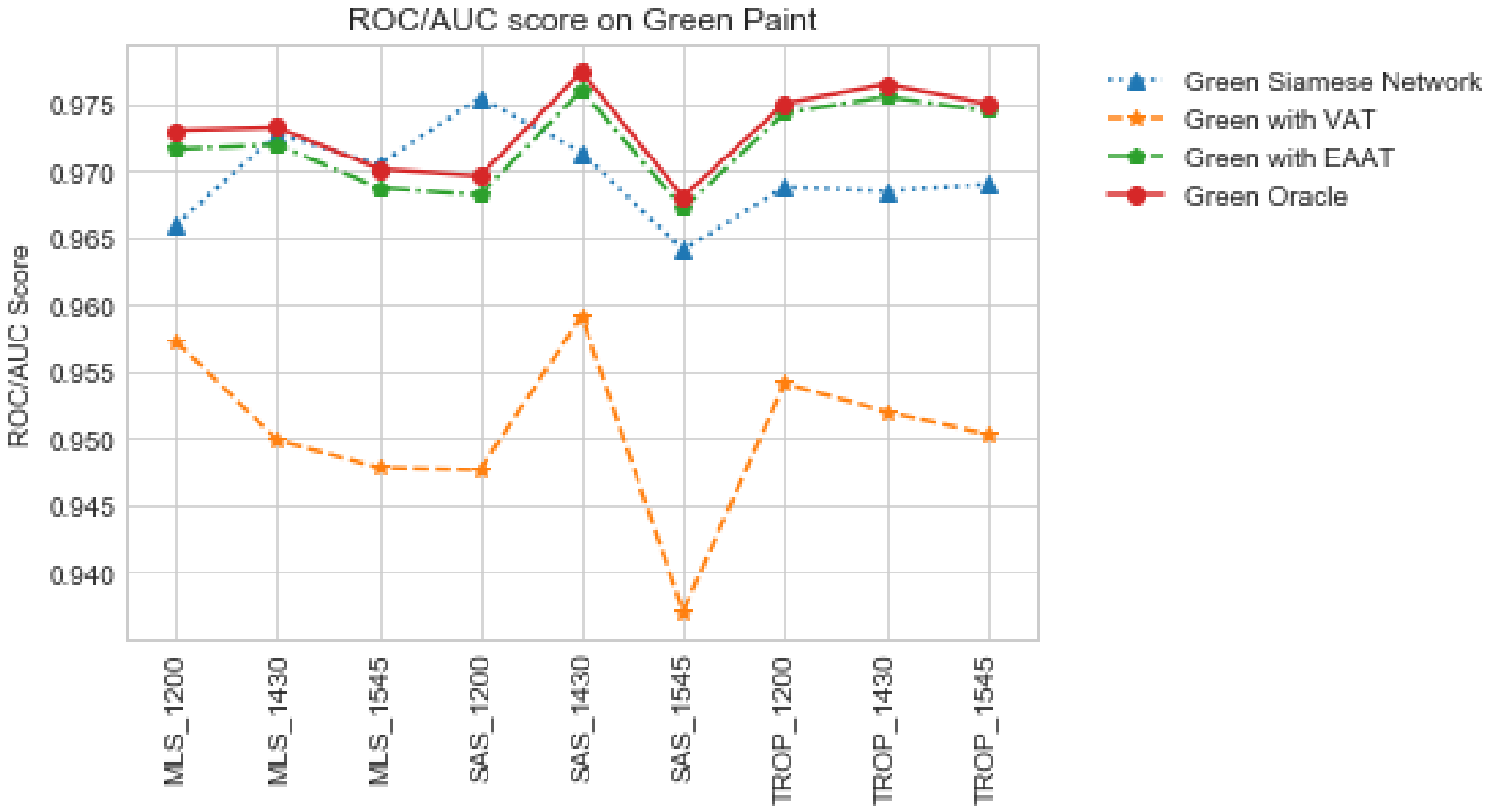
Perturb all datapoints towards the boundary and penalize if this changes the prediction



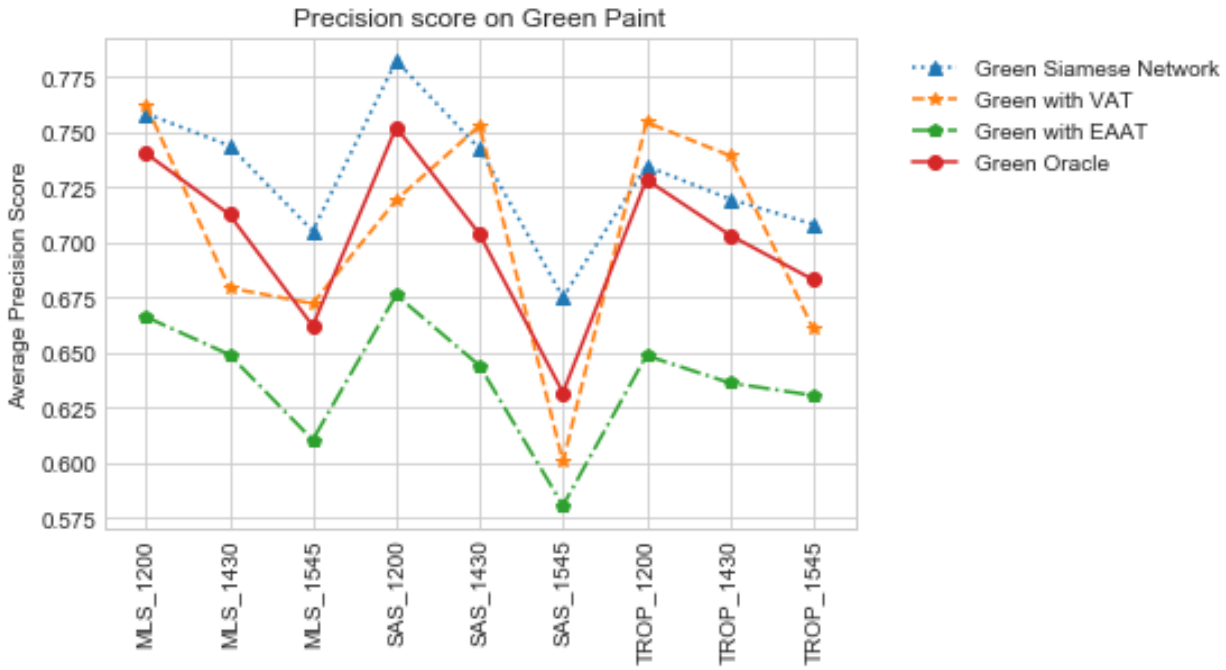
Exponential Averaging Adversarial Training

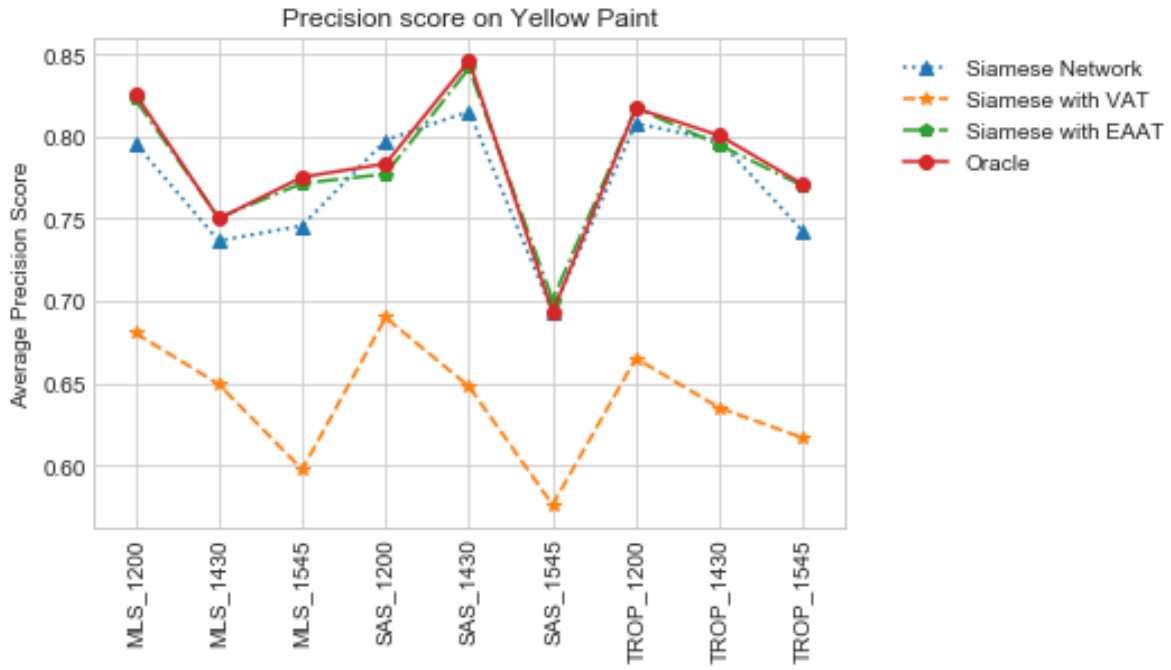
- Combination of MT and VAT
- Students are trained with VAT
- Teacher remains the same

Results


ROC/AUC score on Green Paint
ROC/AUC Score
0.975
0.970
0.965
0.960
0.955
0.950
0.945
0.940
MLS_1200
MLS_1430
MLS_1545
SAS_1200
SAS_1430
SAS_1545
TROP_1200
TROP_1430
TROP_1545
Green Siamese Network
Green with VAT
Green with EAAT
Green Oracle

Results

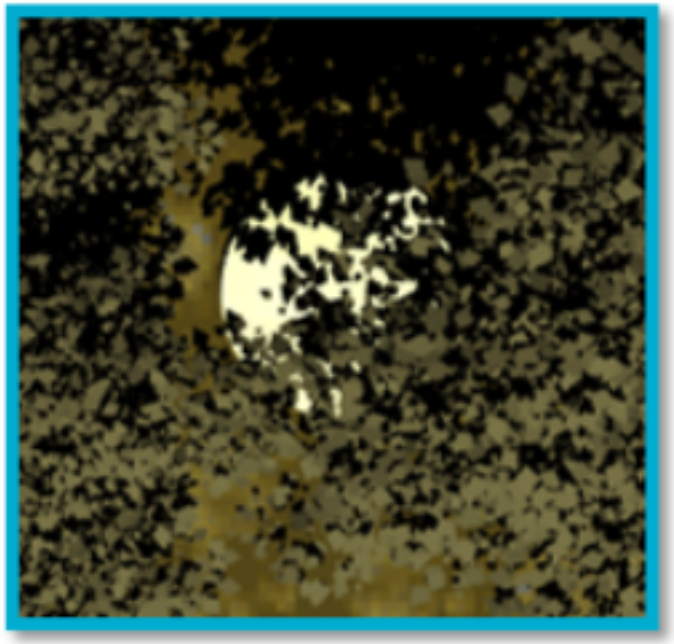

Precision score on Green Paint
Average Precision Score
0.775
0.750
0.725
0.700
0.675
0.650
0.625
0.600
0.575
MLS_1200
MLS_1430
MLS_1545
SAS_1200
SAS_1430
SAS_1545
TROP_1200
TROP_1430
TROP_1545
Green Siamese Network
Green with VAT
Green with EAAT
Green Oracle


Precision score on Yellow Paint
Average Precision Score
0.85
0.80
0.75
0.70
0.65
0.60
MLS_1200
MLS_1430
MLS_1545
SAS_1200
SAS_1430
SAS_1545
TROP_1200
TROP_1430
TROP_1545
Siamese Network
Siamese with VAT
Siamese with EAAT
Oracle

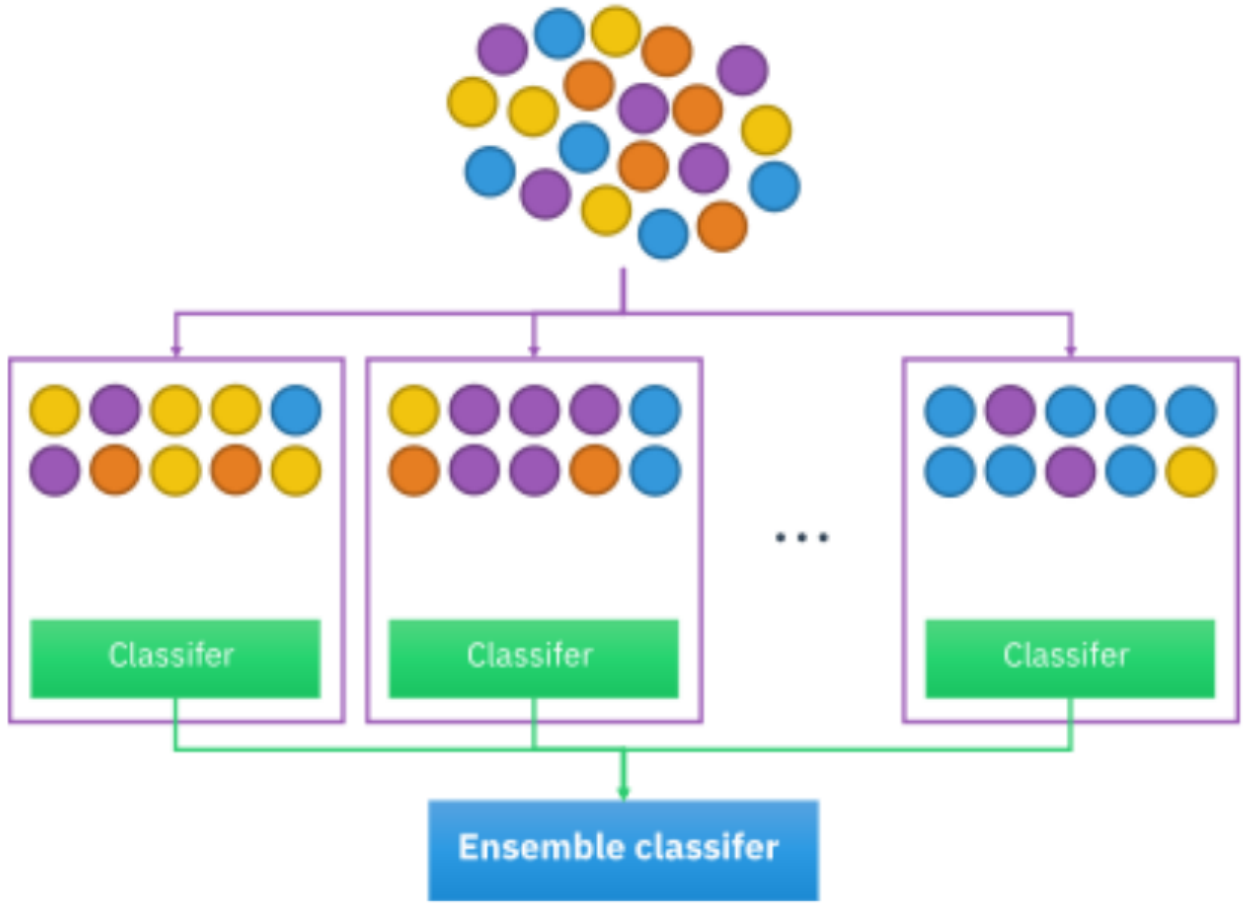
Uncertainty Quantification

Uncertainty Quantification

We have targets of different sizes and visibility. We want a way to categorize how certain the model is the target exists



Bootstrapping



Original Data

Bootstrapping

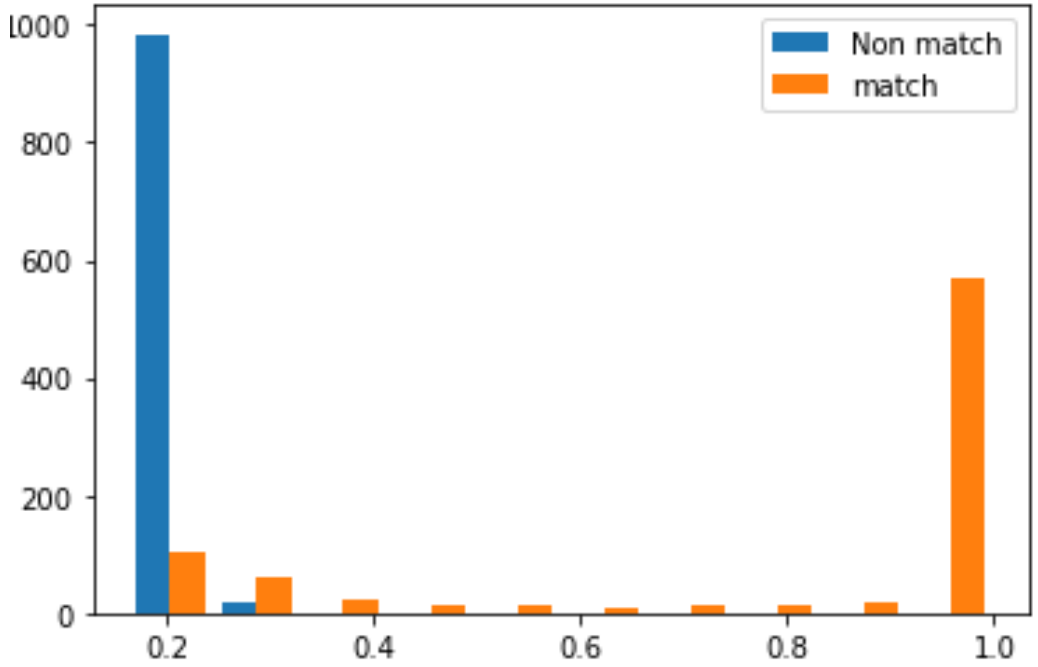
Aggregating

Bagging

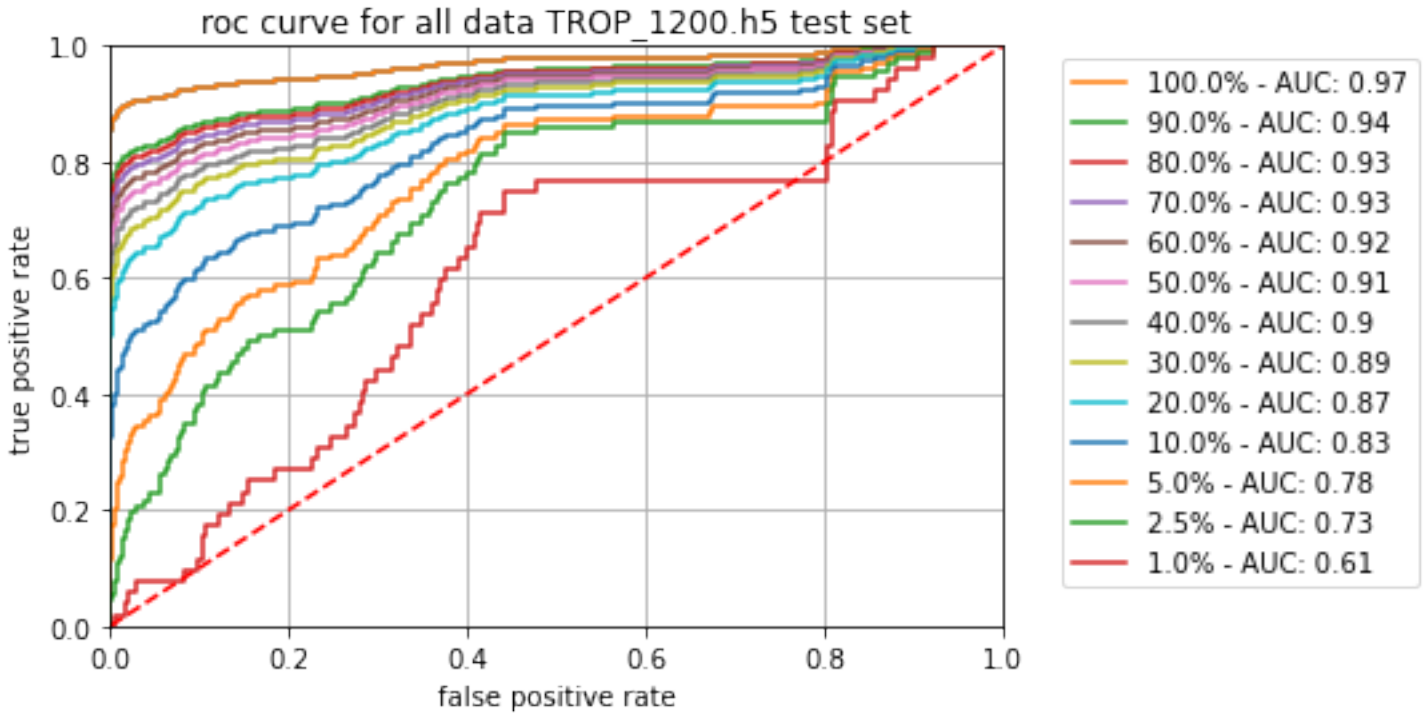
Problems – Hyperparameters

How do we define which values are classified with high confidence?

- Currently defining 3 hyperparameters, alpha, beta, gamma
  - Alpha: Upper cut off
  - Beta: Lower cut off
  - Gamma: Standard Deviation cut off



Results



CFAR Score	0-0.25	0.25-0.75	0.75-1
Total	0.68	0.995	0.998
Good	0.233	1	1
Bad	0.808	0.986	0.826

Model Requirements

1. Low-Shot Learning
2. Semi-Supervised Learning
3. Uncertainty Quantification

Questions?