SAND2022-0397C

Sandia National Laboratories

Exceptional service in the national interest

Automatic anomaly detection in high reliability as-built parts from images

Kevin Potter

LDRD LABORATORY DIRECTED RESEARCH & DEVELOPMENT

Anthony Garland, Emily Donahue

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.

U.S. DEPARTMENT OF ENERGY

NNSA

High consequence components

Why we care

Sandia needs to ensure parts work as intended

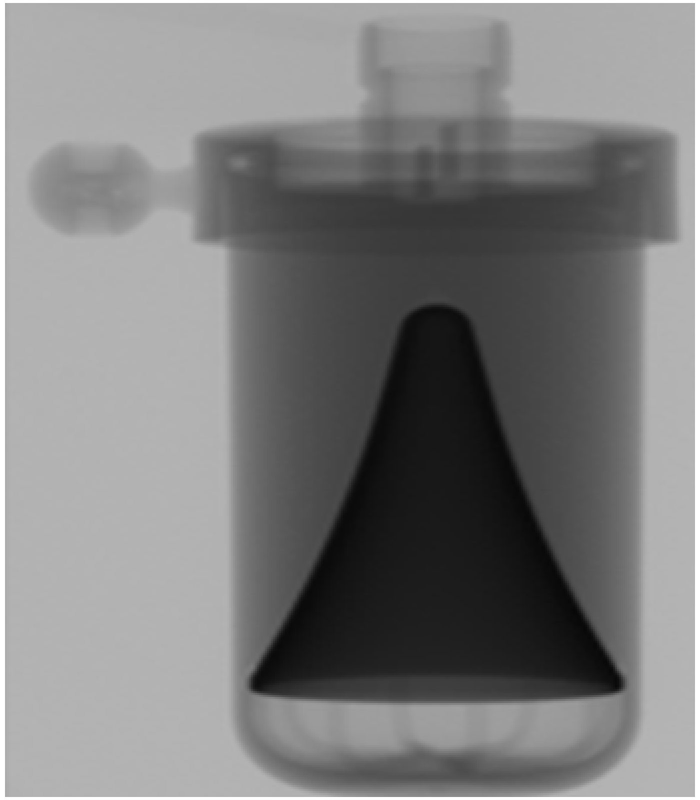
Sandia works with high consequence situations/items

How can we find items that are likely to fail?

- Particularly important for components that are destroyed or damaged when used (explosives, body armor, etc.)
- We need non-destructive means of testing as-built components

X-Ray images

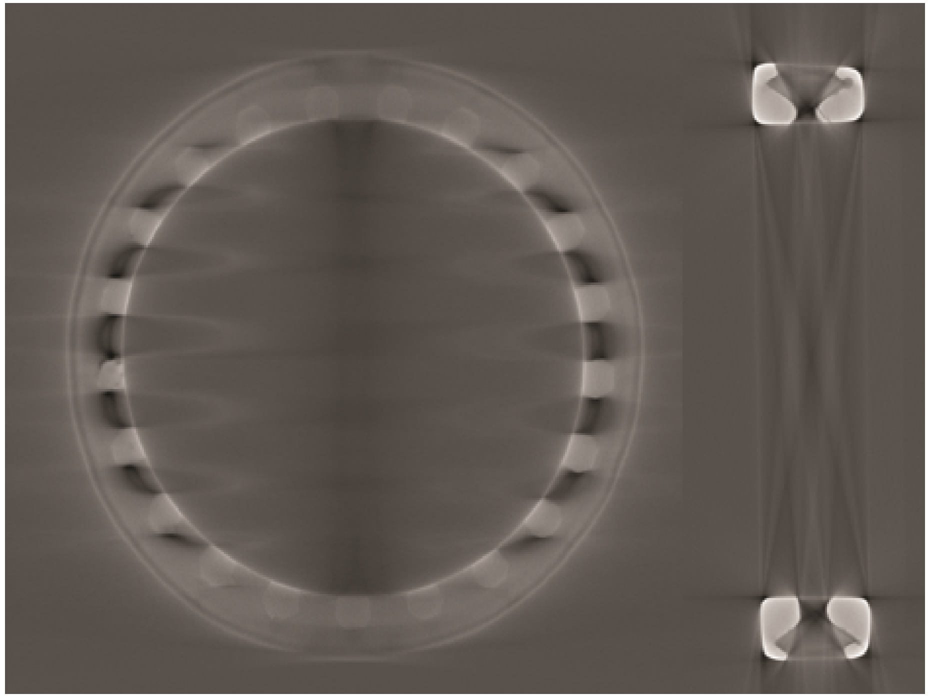
- Non-destructive
- But are just images (grayscale)
- Finding flaws often requires human expertise
- Is a 2-D representation of a 3-D object



X-Ray of shaped charge

X-Ray computed tomography (CT) is a powerful tool but…

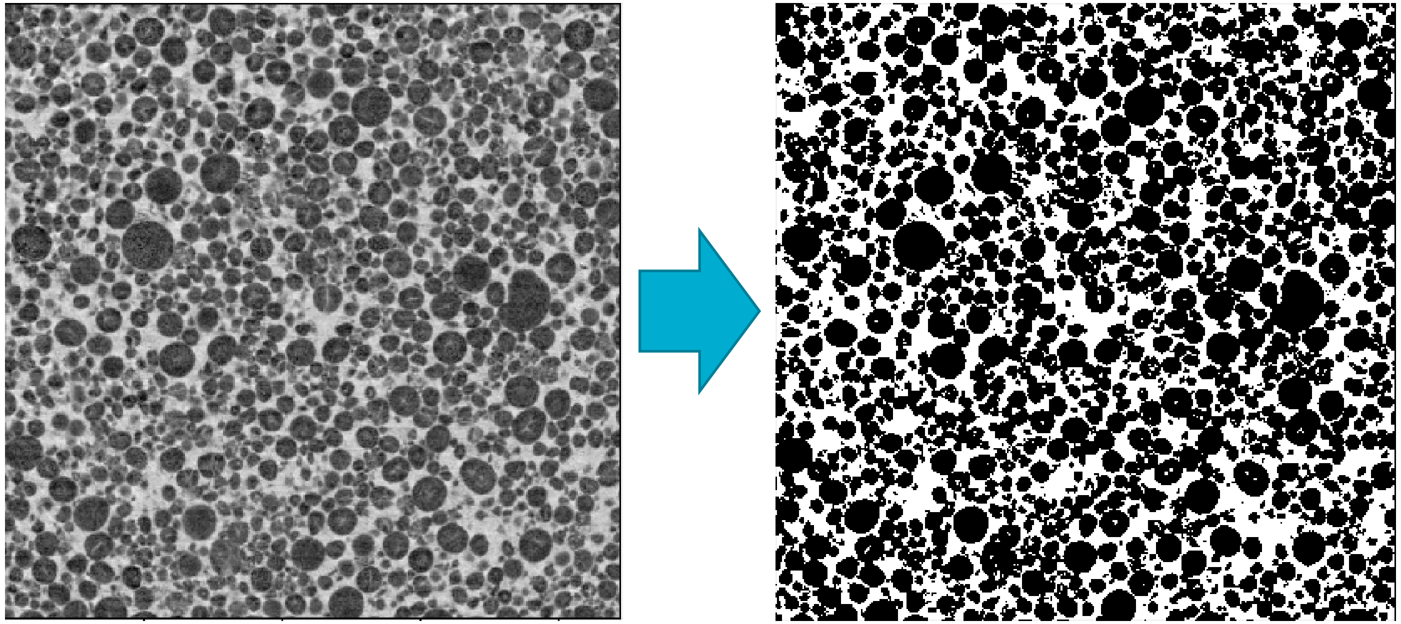
High Z material can cause bright and dark "shadow" artifacts in the CT reconstruction
- Complicates finding anomalies
- Is still an image (grayscale), just now in 3-D



We can apply semantic segmentation to get useful information

Downsides:

- Segmentation requires labelled data (often expensive to acquire)
- Typically there are far more nominal examples than anomalous ones
- Dealing with class imbalance means throwing out some (or most) of the anomaly-free data
- **Only can reliably learn to predict anomalies from the labelled set**



Slice of battery scan and corresponding segmentation

We need an unsupervised solution to utilize data better and handle unexpected anomalies

Two types of unsupervised techniques for anomaly detection

Generative approaches

Generate a close “normal” image to a query image and look at difference to original

- AnoGAN
- PandaNet

Classification approaches

Classify the image as normal or anomalous directly

- Deep one-class classification
- Feature-based Anomaly Detection System (FADS)

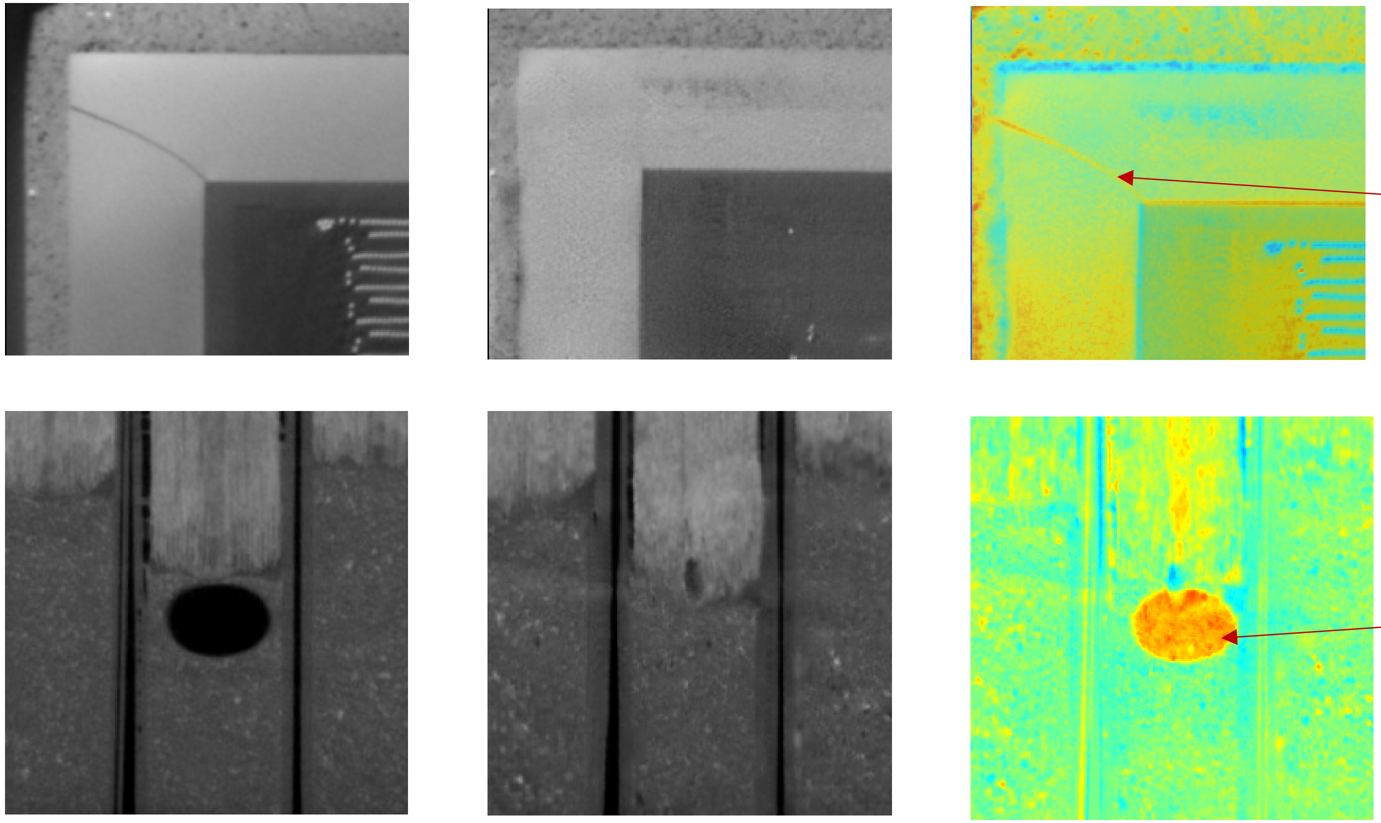
Anomaly Detection

Generative approaches

AnoGAN results highlight cracks and voids in as-built components

Emily Donahue

Real scan image | Generated image | Anomaly detection



Detects cracks in ferrite core

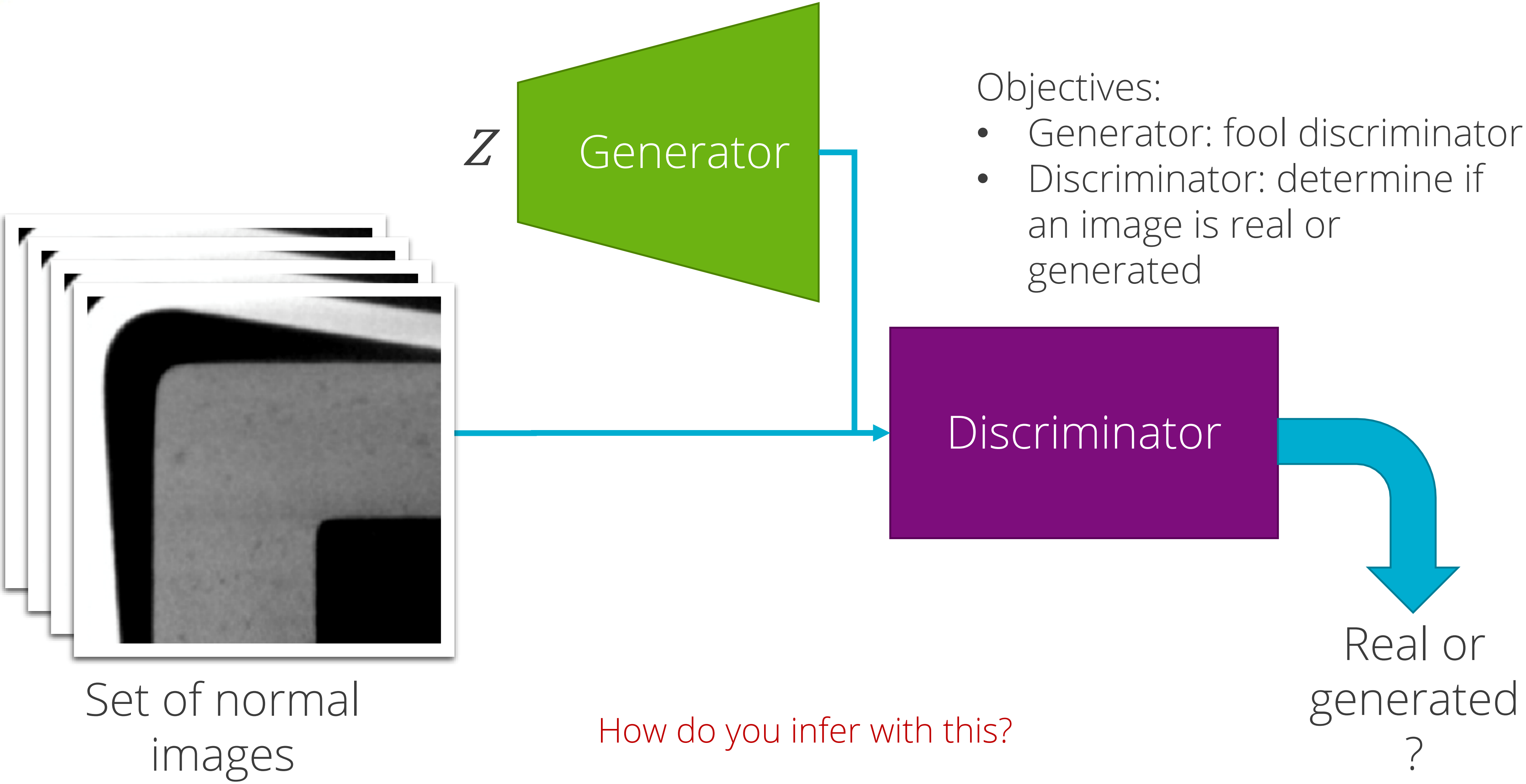
Detects void in capacitor

AnoGAN succeeded in identifying rare features where supervised learning is not possible

AnoGAN from https://arxiv.org/abs/1703.05921

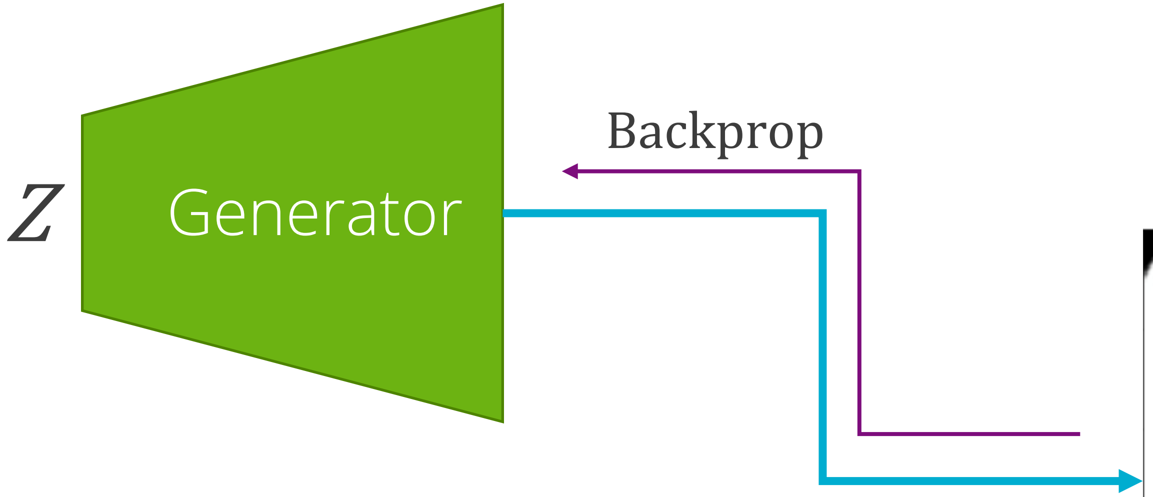
AnoGAN architecture

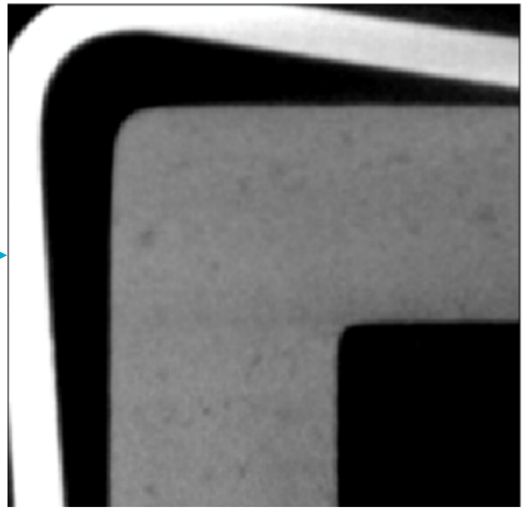
Objectives:
- Generator: fool discriminator
- Discriminator: determine if an image is real or generated



How do you infer with this?

AnoGAN inference uses backpropagation

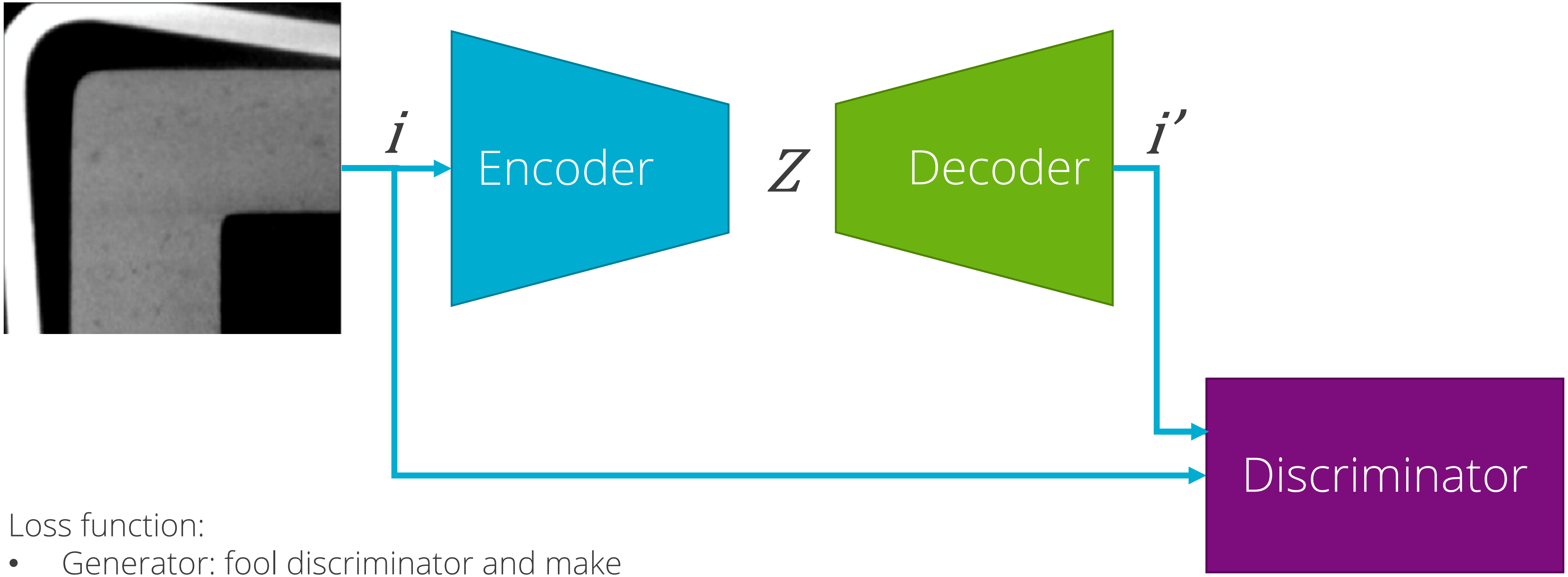




- Start with a random guess (*Z*)
- Backpropagate to change *Z* using the image as if it was label
- Repeat until *Z* changes are small

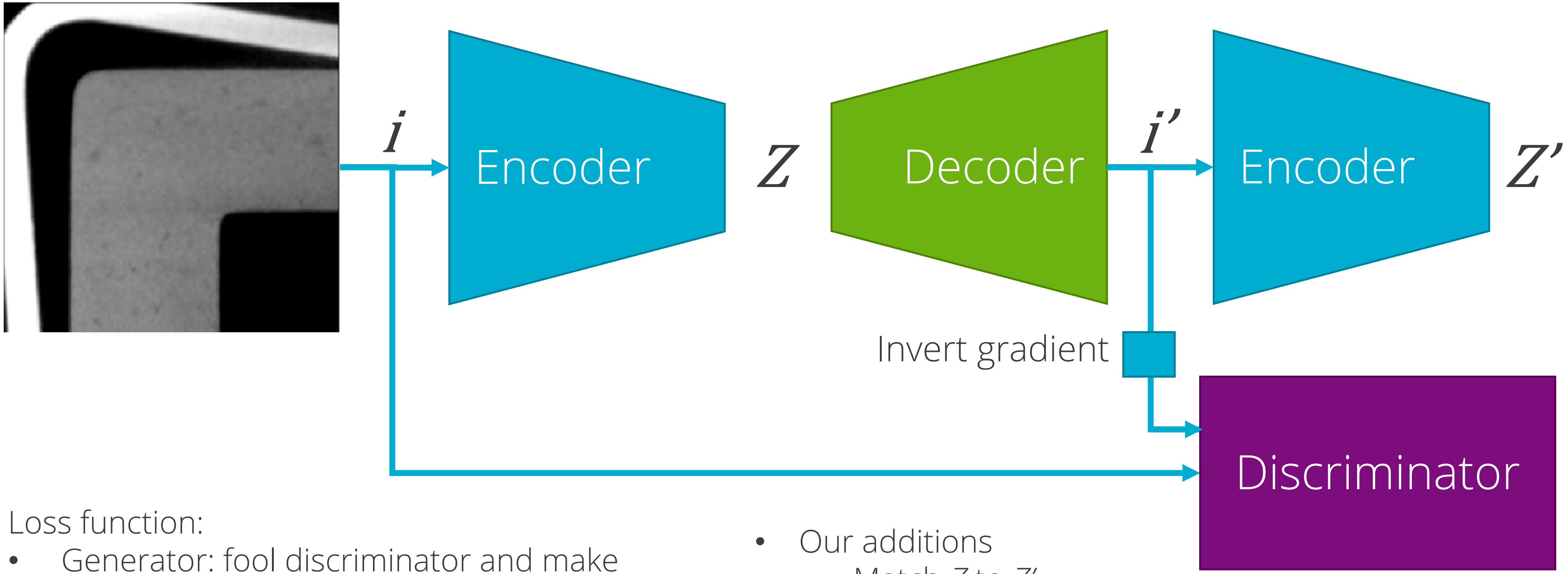
This works but is slow (seconds per image)

AnoGAN's successor: f-AnoGAN



Loss function:
- Generator: fool discriminator and make sure decoded image is similar to input
- Discriminator: determine if an image is real or generated
- Match *i* to *i'*

Introducing PandaNet



Loss function:

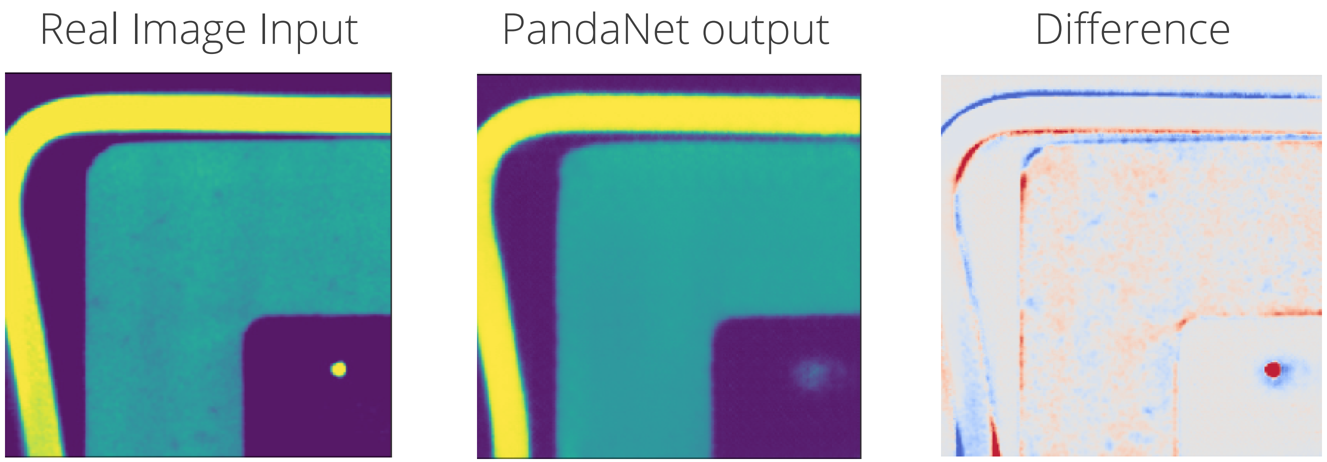
- Generator: fool discriminator and make sure decoded image is similar to input
- Discriminator: determine if an image is real or generated
- Match *i* to *i′*

- Our additions
  - Match *Z* to *Z′*
  - 2-D and 3-D support
  - Support for images too large to fit at once into GPU
  - Single pass training using gradient inversion

PandaNet applied on transformer data

PandaNet learns from a set of defect free images

- When predicting an image with an anomaly, such as the solder ball in the top image, it fails to reproduce it accurately
- The difference highlights the defect.
- The learned encoder significantly speeds up the process



Similar results to AnoGAN in ms

Moving to full CT image anomaly detection

CT volumes tend to be quite large (typically we deal with scans on the order of a billion voxels)

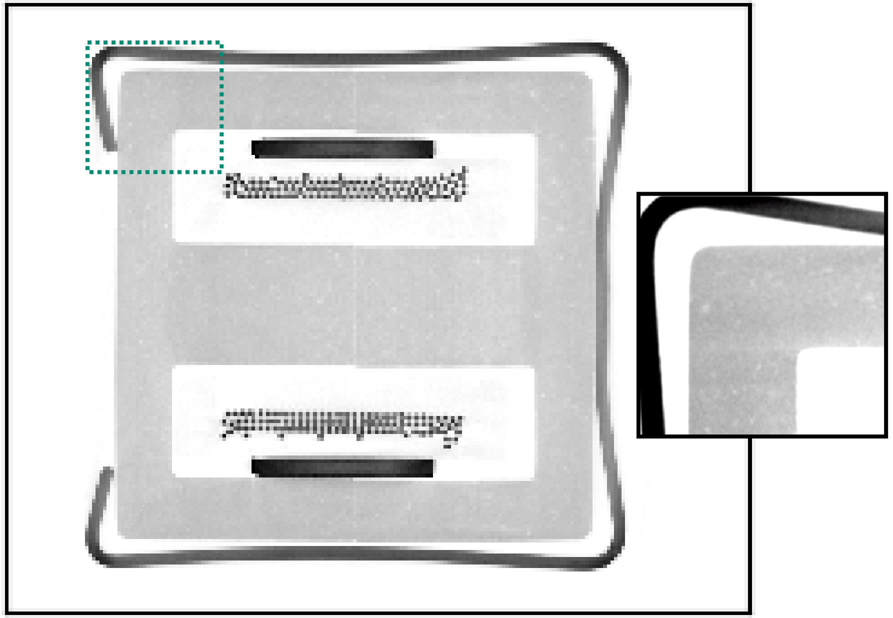
- Slice of overall image to right and there are almost 1000 slices in that scan's volume
- Seconds per image for AnoGAN would mean days for a **single** full 3-D image to be inferred!

Labels for testing are difficult

- Labelling all anomalous voxels by hand is impractical
- Thin flaws can be practically invisible in a slice by slice view
- Solution: Synthetic anomalies

Cannot fit in a GPU alongside a reasonable model

- Solution: Split into smaller chunks of the whole



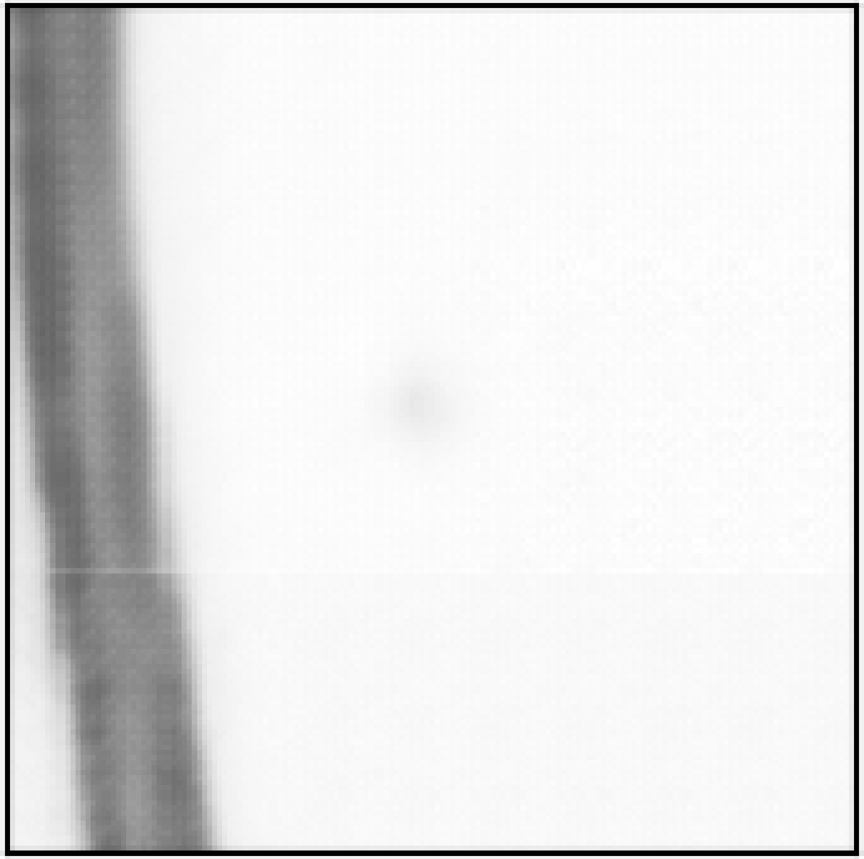
Slice of whole image

Small inset is region used in prior work

Stitching chunks has flaws

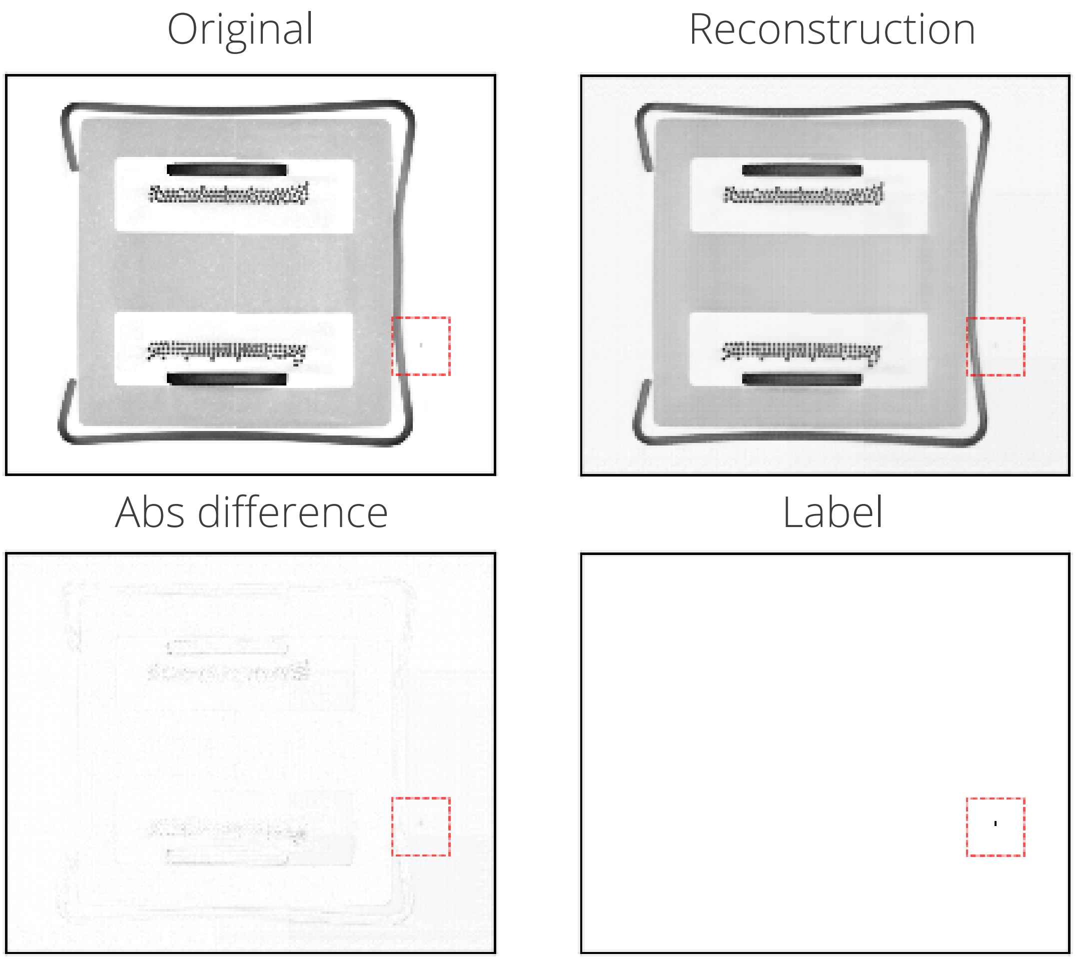
Solutions bring their own problems

- Chunks have to be stitched back together during inference
- Can create artifacts along edges from lack of context
- Overlap and weighted average to reduce stitching artifacts

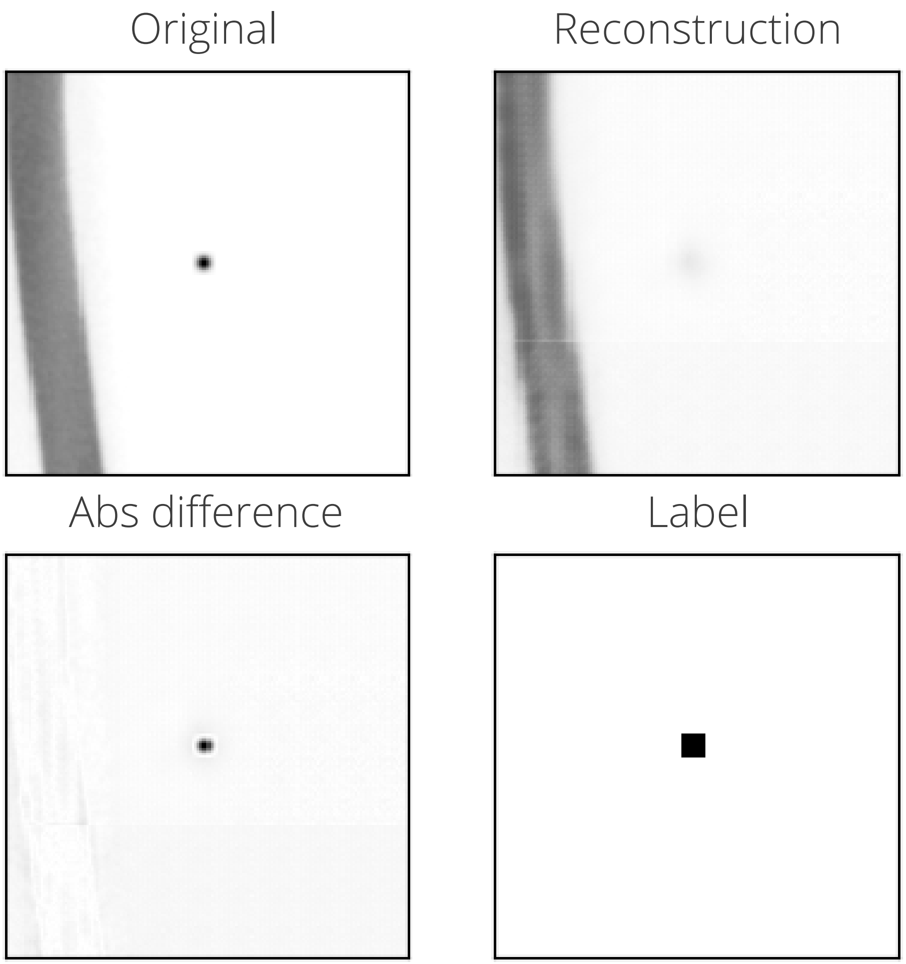


Zoomed in slice of reconstruction

Full size 3-D anomaly detection on transformer



Slice of image



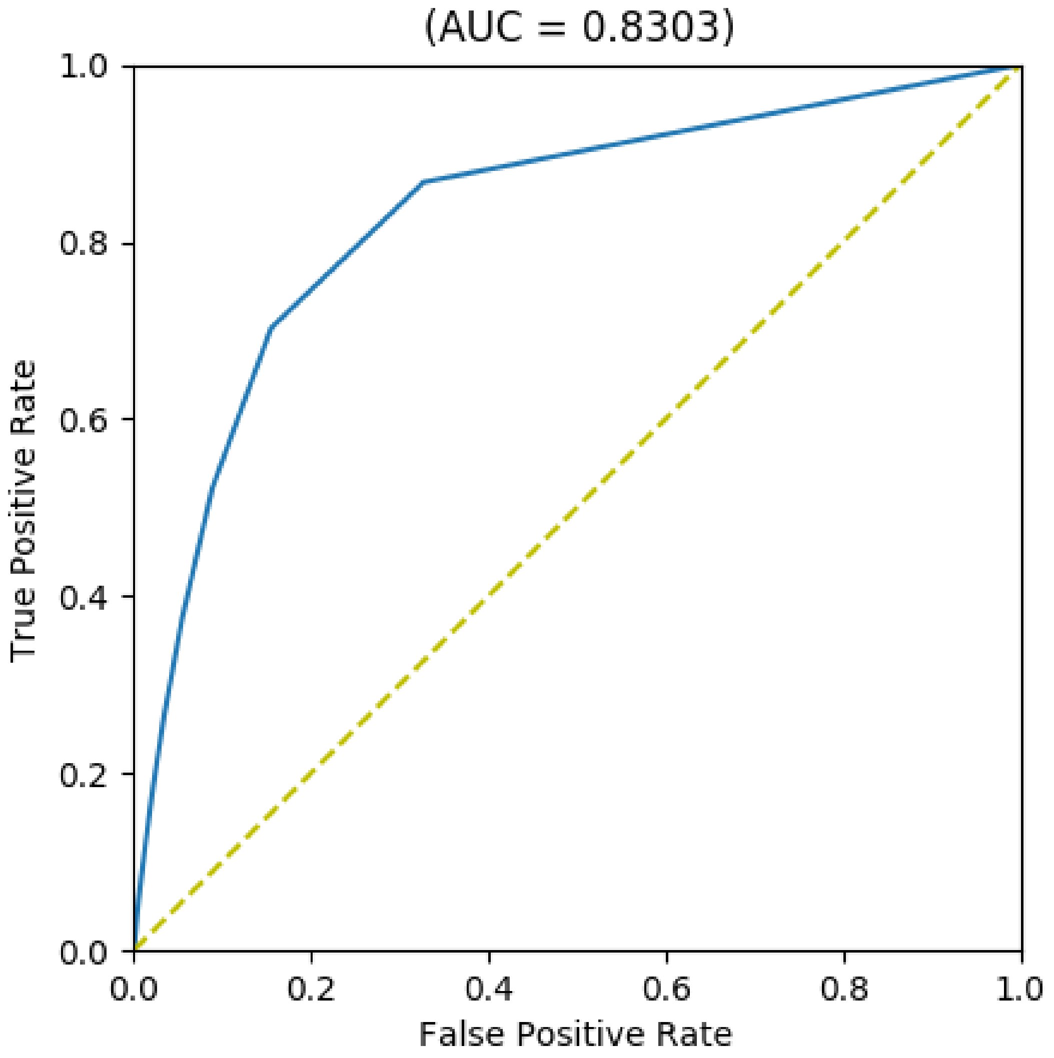
Zoomed in on synthetic anomaly

Full size 3-D anomaly detection on transformer (cont)

ROC curve created against abs difference of each voxel

Small, high-Z Gaussian anomalies added to images to make a test dataset

- Similar to solder ball flaws
- Label is the whole volume of the Gaussian cube (voxels labelled anomalous when not)
- Placed randomly in the images
  - Can be over already high-z areas
  - There may be existing flaws that are not labelled
- Very small portion by volume (1 in 6 million voxels is part of a synthetic anomaly)



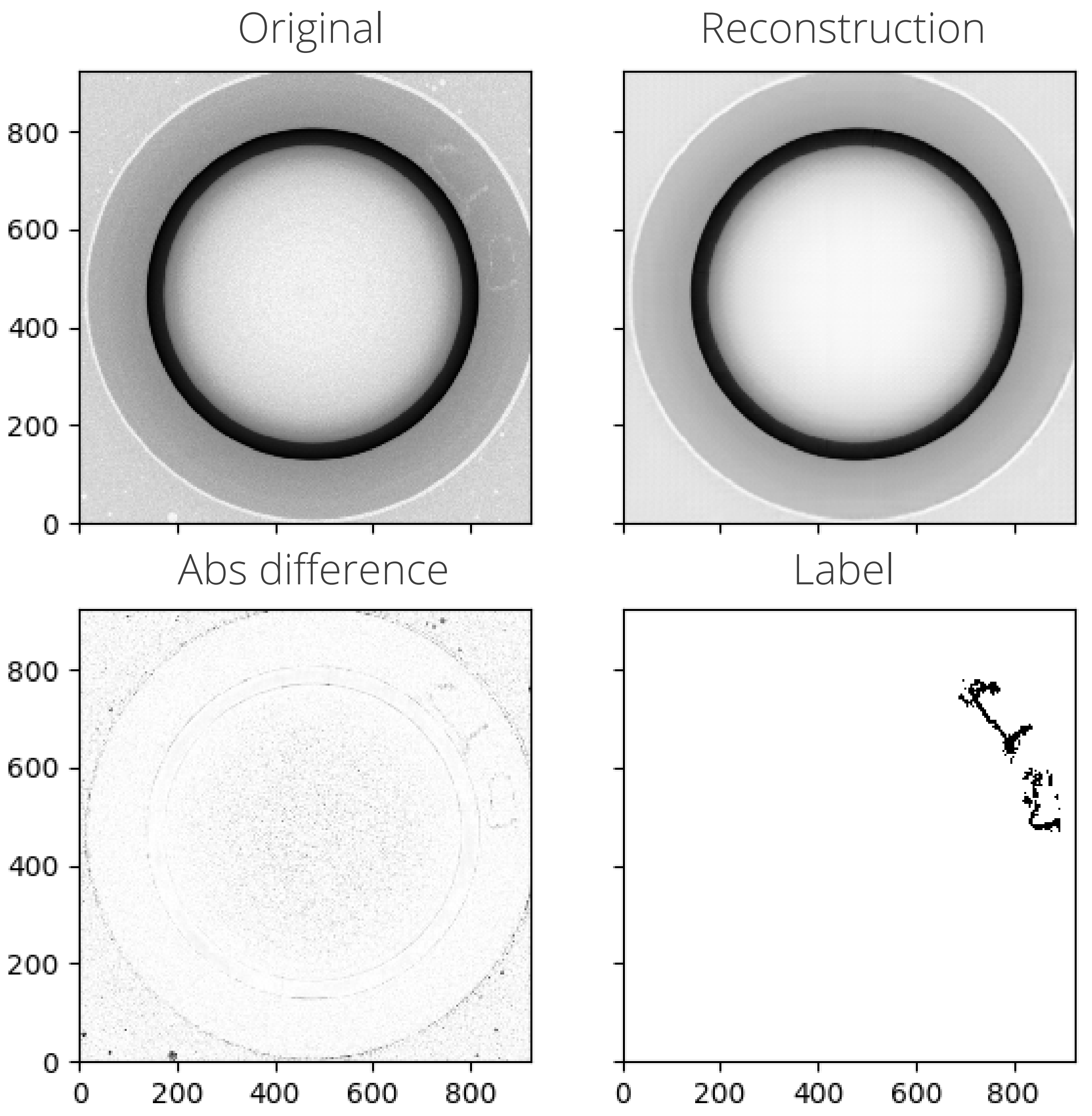
Shaped charge detection hampered by unimportant anomalies

Challenges:

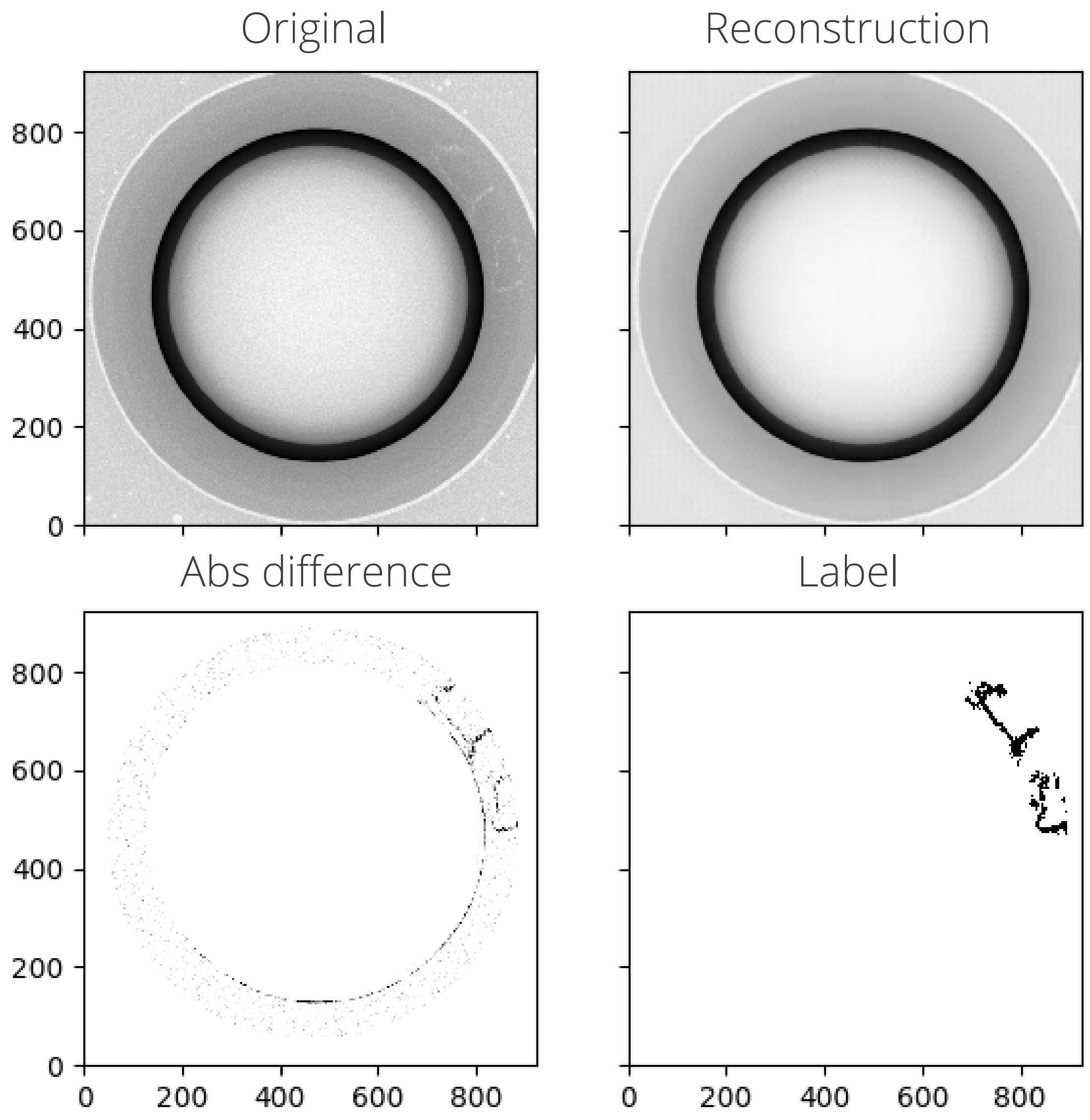
- The charge has approximately 1 billion voxels with roughly 30% being the region that we care about and 0.5% defects
- There were no flaw free charges for training (model learned to output a clean image anyway)
- Anomalies are a **very** small portion of the scan
- Anomalies had low contrast to surroundings

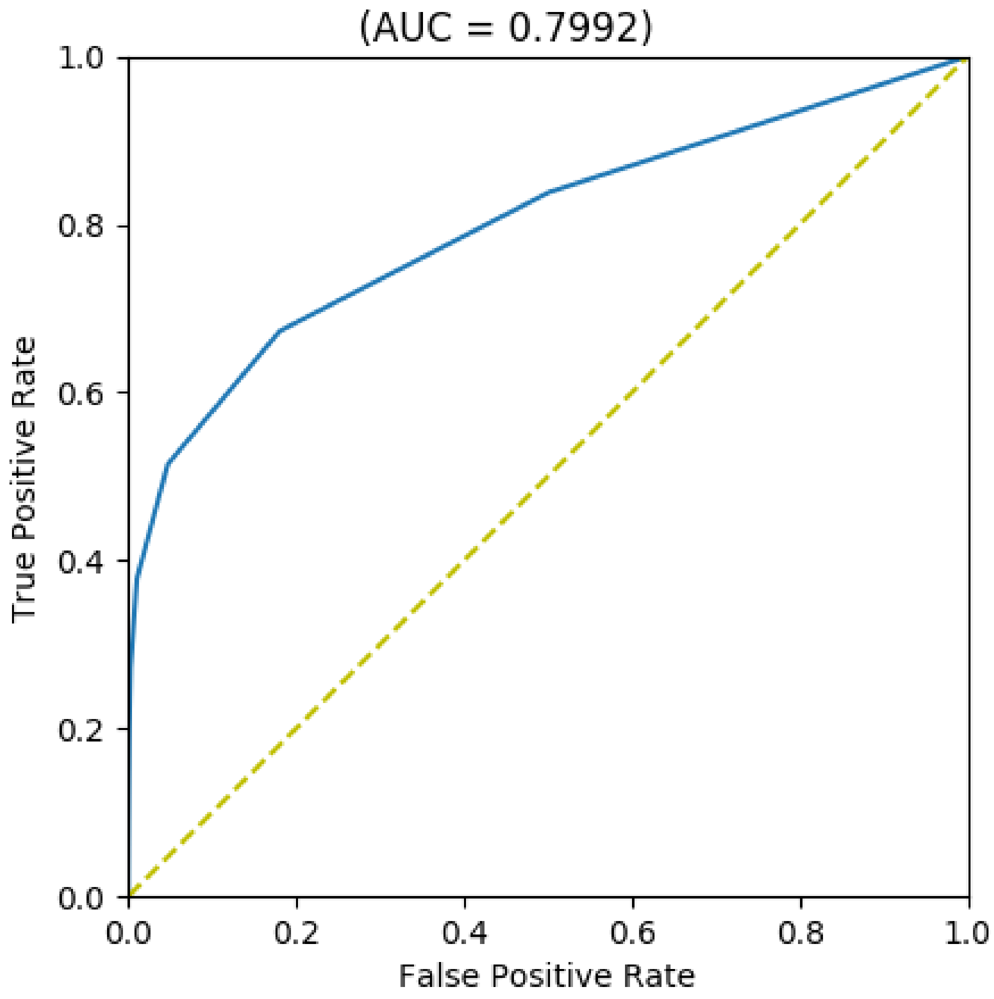
Despite learning a clean output, there was too much noise outside the region of interest to get a clear anomaly signal

This was part of another effort were we do have segmentations available and we were able to focus our view on the explosive



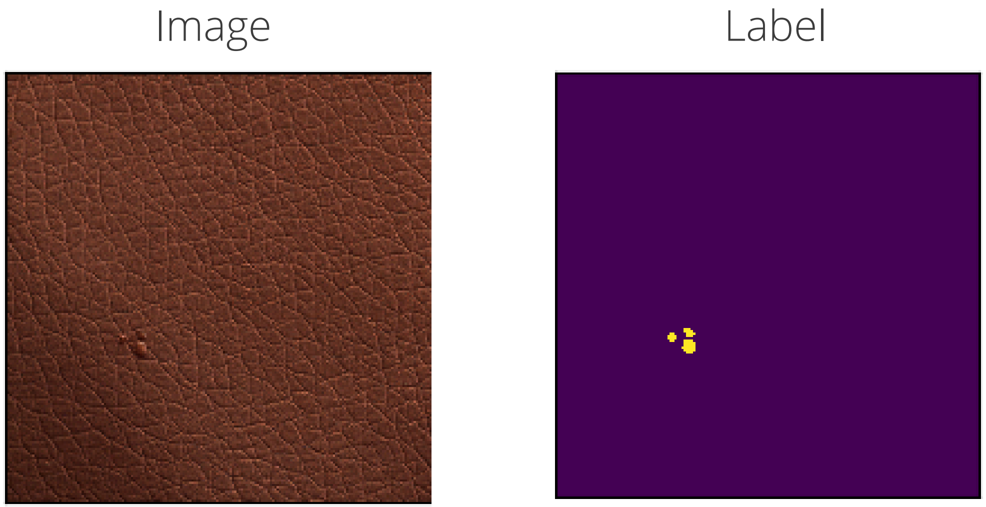
When we look only at the volume of interest



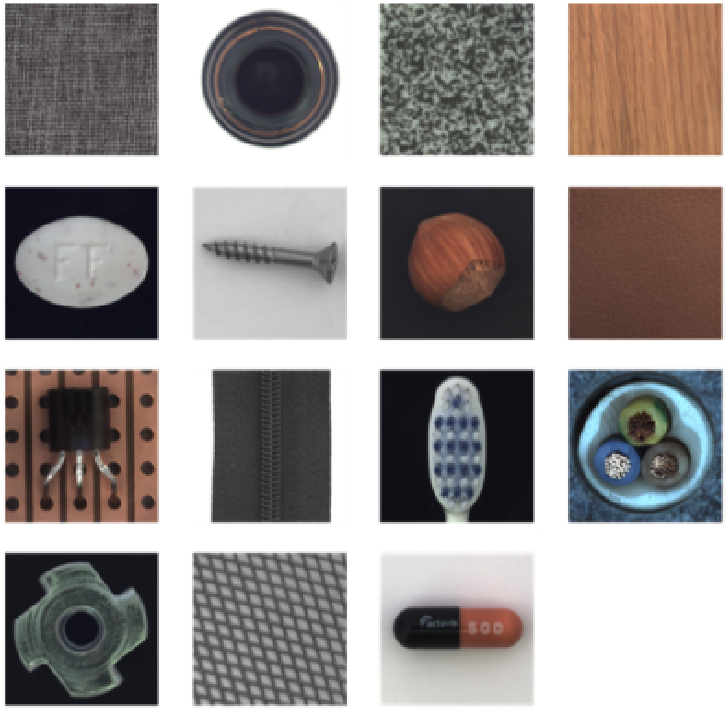


MVTec Anomaly Detection (MVTec AD) dataset

The MVTec AD dataset is comprised of 5000 high-resolution color images of fifteen different categories with corresponding pixel-wise anomaly labels



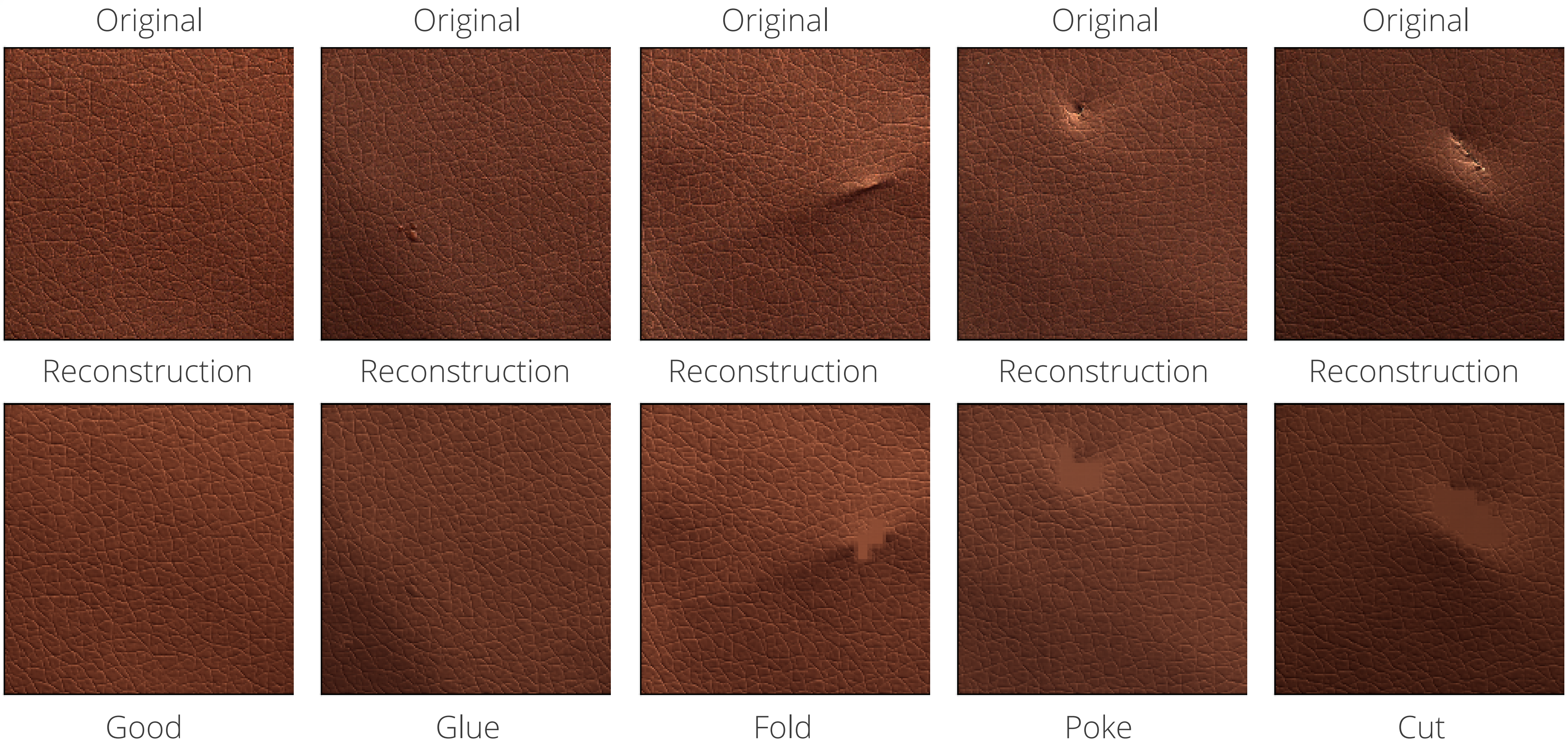
Leather category example with a glue defect



Examples for each category

Bergmann, Paul, et al. "MVTec AD--A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection." CVPR 2019.

MVTec AD leather examples



Examples from MVTecAD leather with PandaNet reconstructions

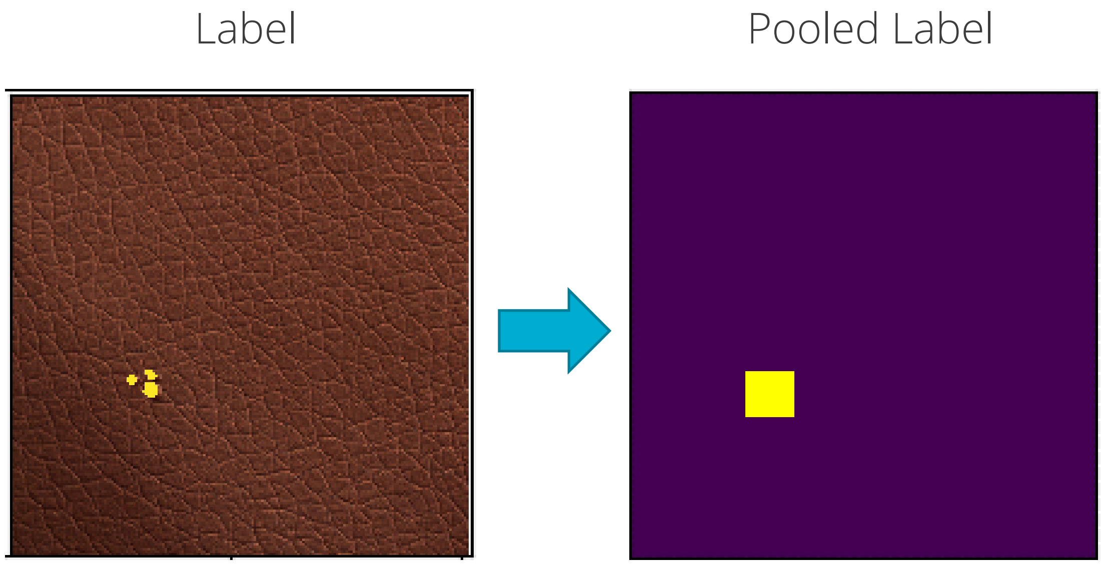
Relaxing the localization of anomalies

Many times, defects do not need to be identified at the pixel level

- Regional for human comparison
- Whole image for rejection

Max pooling to test for coarse details in anomaly detection

- Pixels are thresholded as normal
- Both prediction and label are max pooled at various scales
- Gives “credit” for getting close



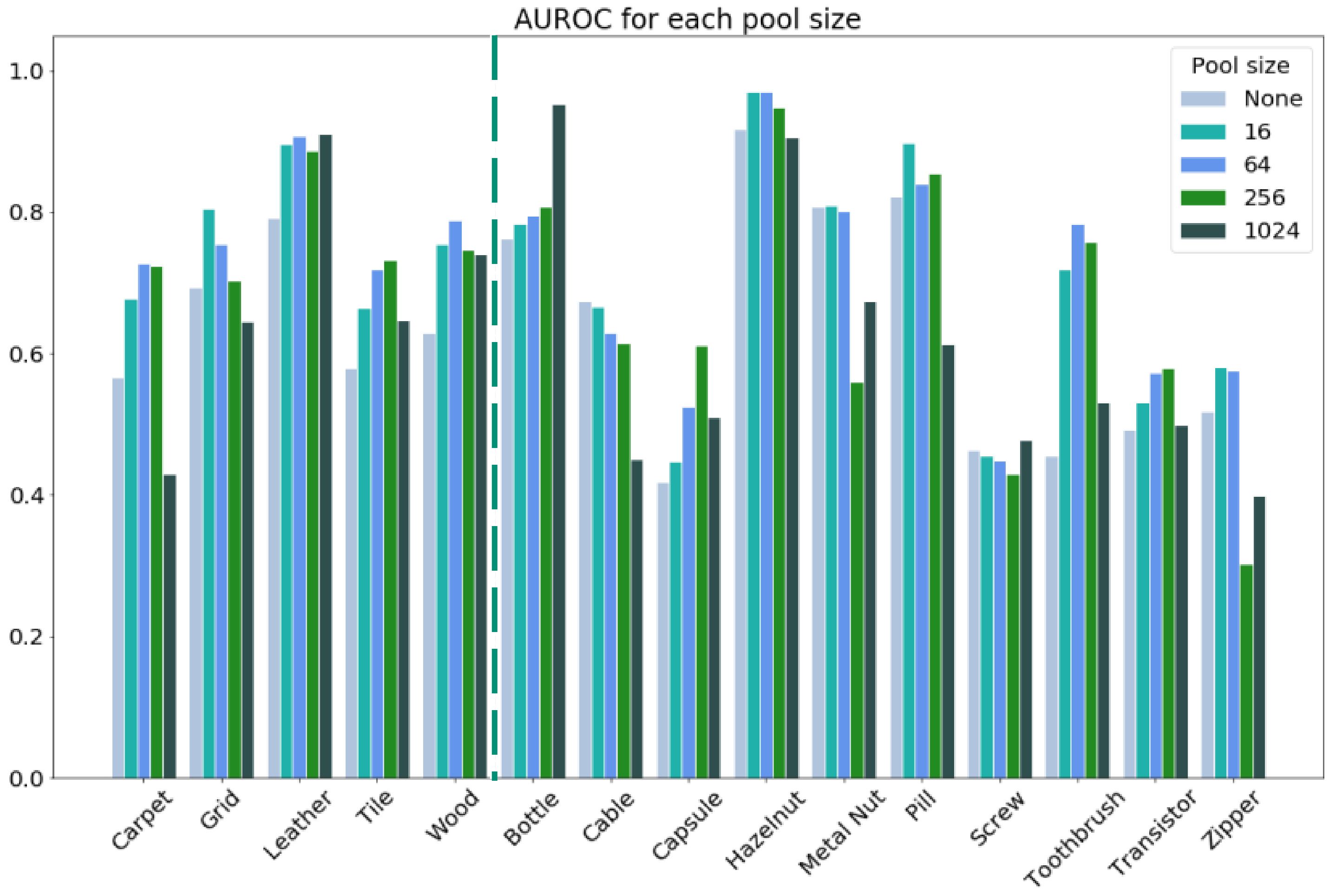
PandaNet applied naively against all the categories in MVTec AD

Only tuned using leather and bottles for structured categories

All other models use the same hyperparameters (as appropriate for texture or structure) with no tweaks

Categories to left of the dashed line are dominated by texture; those to the right are dominated by structure

At this point, while color was output, it was not considered when testing for anomalies

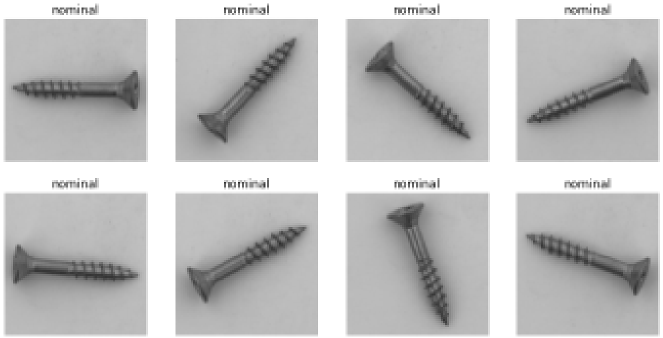


Anomaly Detection

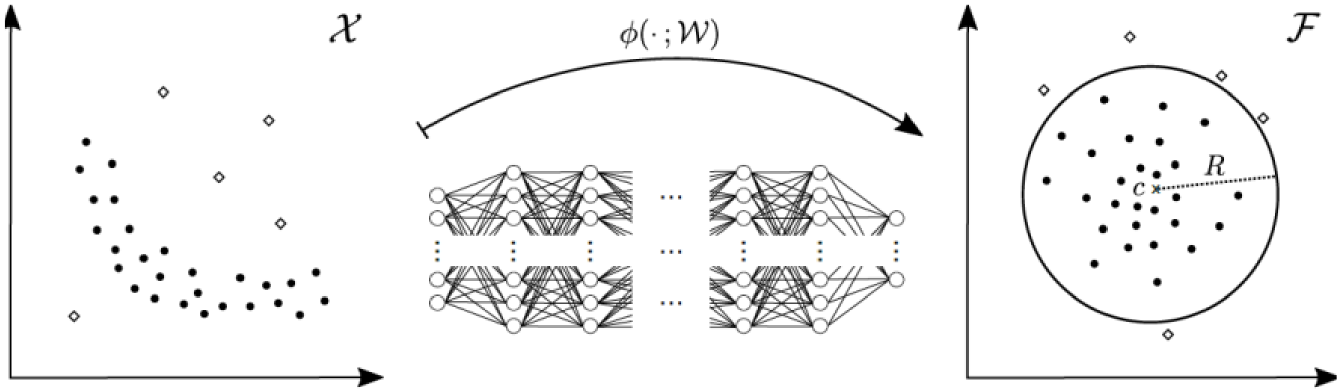
Categorization approaches

Deep one-class classification

- An out of distribution detection machine learning algorithm
  - Map high dimensional images to some reduced order space
  - Map all the nominal images to the same region in the reduced order space
  - A machine learning model provides the mapping
  - Use some notion of 'distance' from c as the measure of how 'anomalous' an image is



Nominal images are input



ML model provides the mapping

Far from c = Anomaly

Ruff, L. et al. (2018). Deep One-Class Classification. *Proceedings of the 35th International Conference on Machine Learning*, in *Proceedings of Machine Learning Research* 80:4393-4402

Feature Based Anomaly Detection System (FADS)

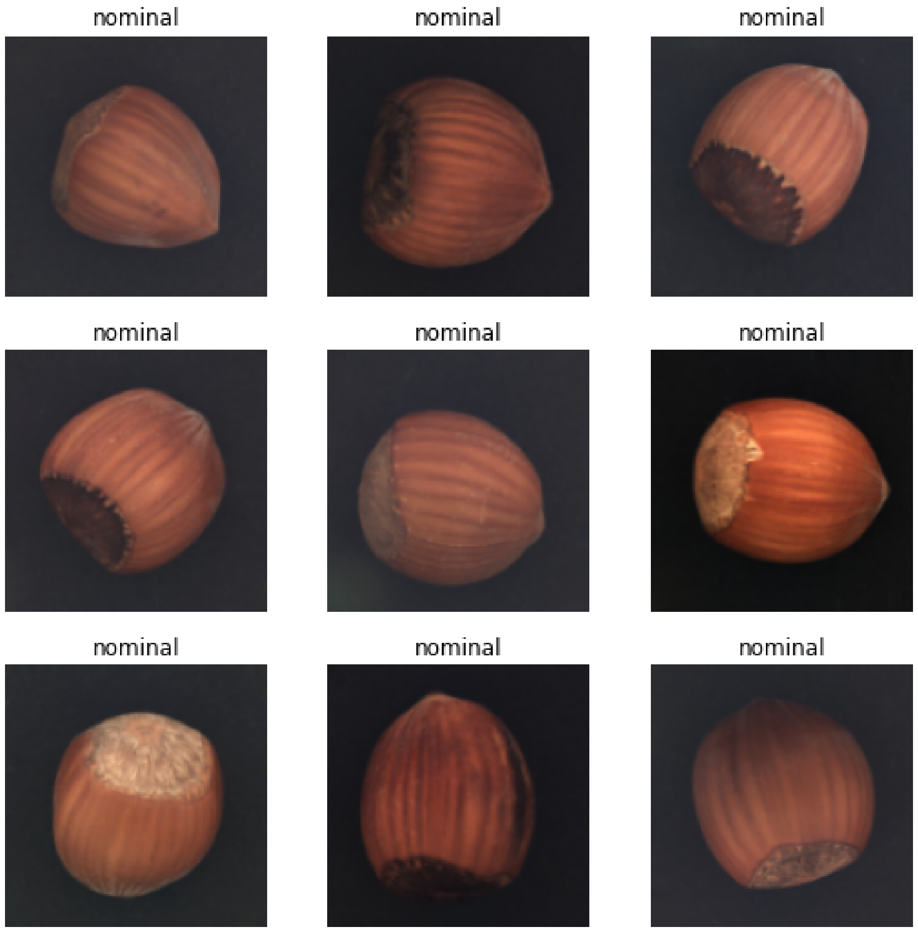
Anthony Garland

Extends the deep-one-class classification idea

Works via a **pretrained network** to provide the mapping

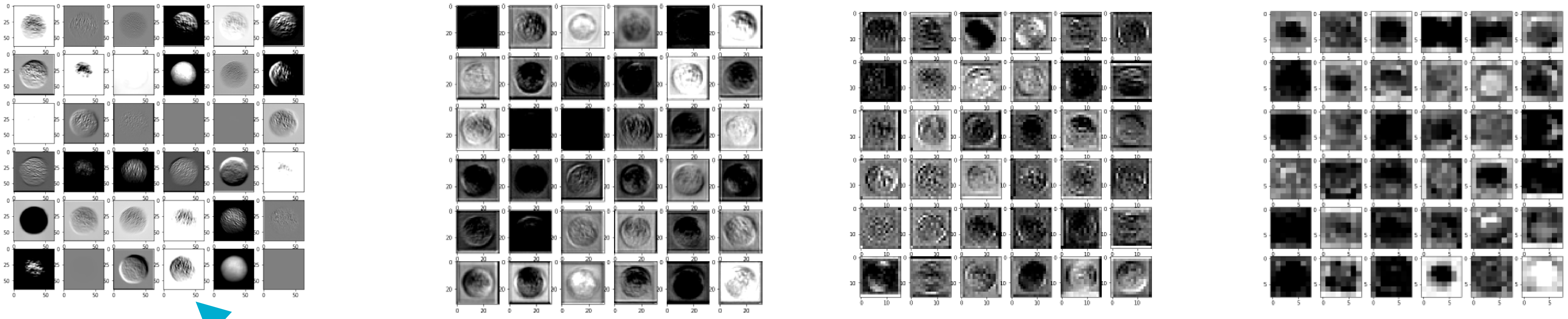
- **Record** the activations from the model's convolutional filters
- **Aggregate** each filter to a single value (max, min, mean, etc.)
- Develop a **statistical** model of expected convolutional activations ($\bar{x}$ and σ) based on the nominal images' activations
- At inference time, measure activation of the input image from the same model and **normalize** based on the nominal statistics

Example nominal data (hazelnuts)

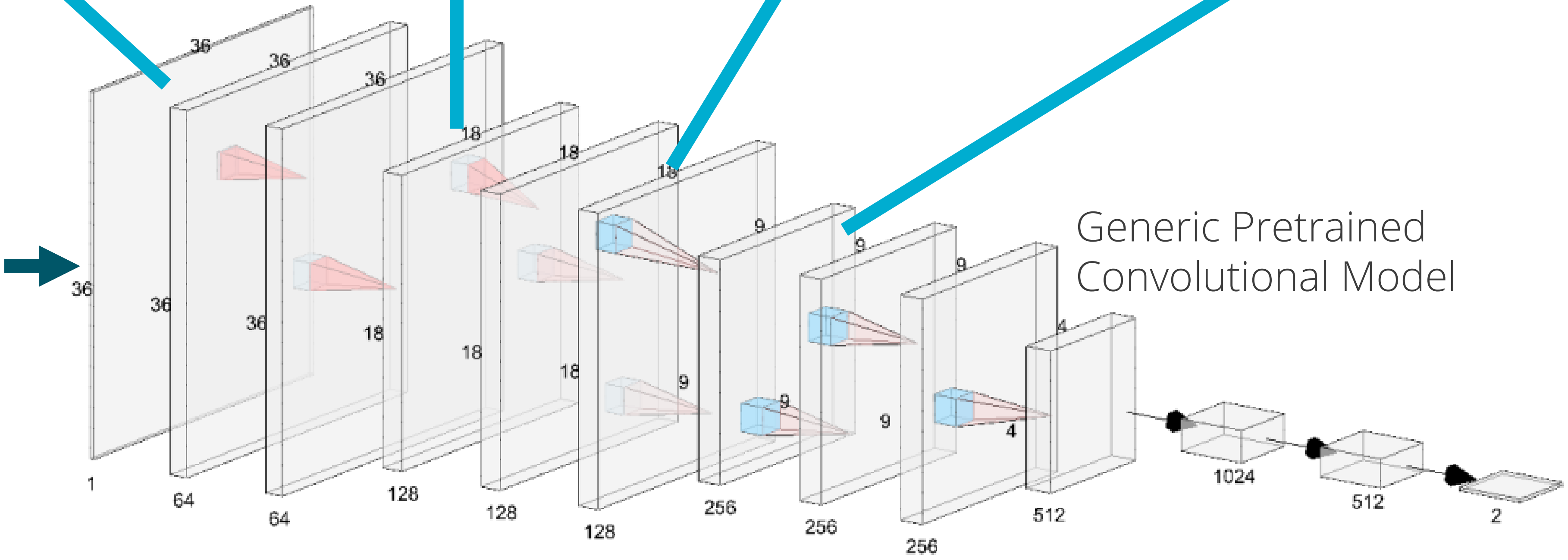


CNN activations

Example CNN activations

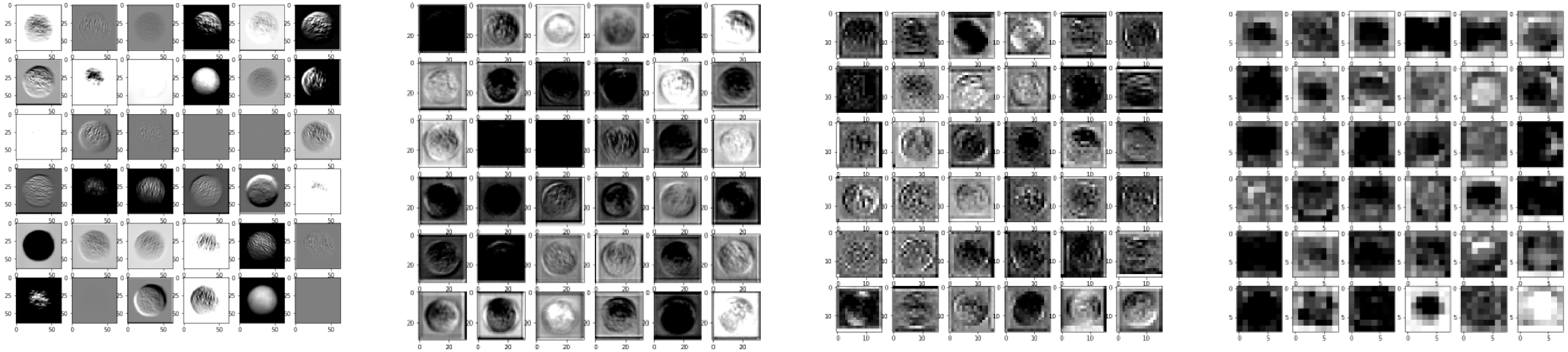


Input Image

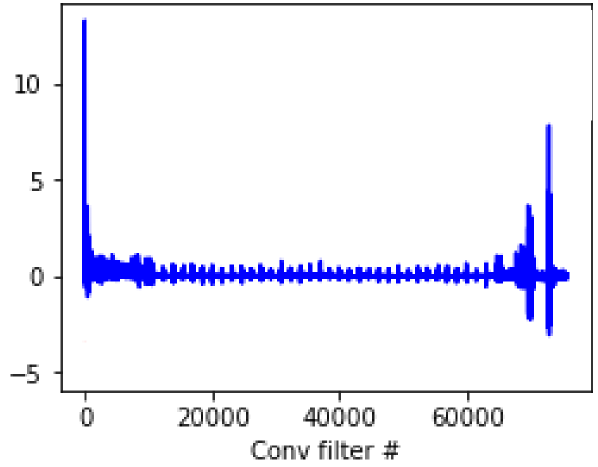


Generic Pretrained
Convolutional Model

Aggregate activations



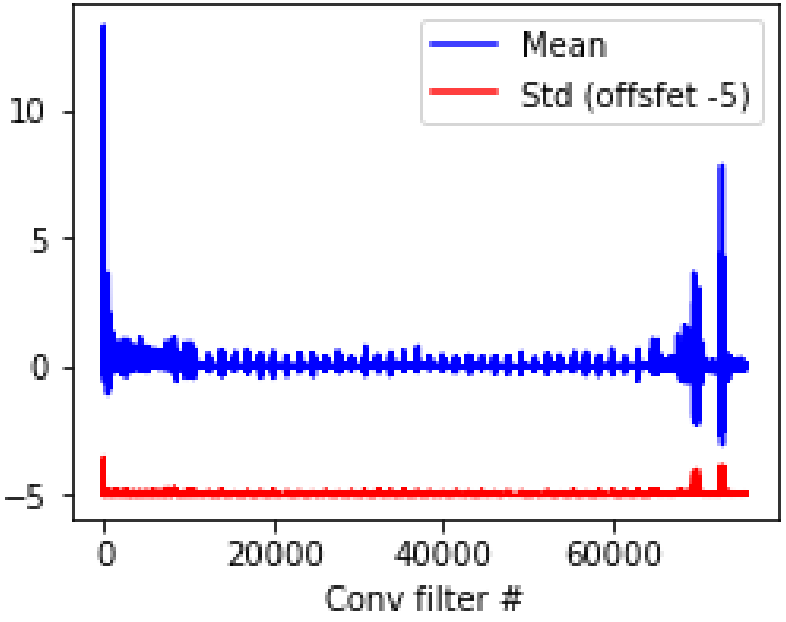
Collapse each filter's activations to a single value and stack



Learn the nominal datasets activation stats for each filter

Nominal images
(“training” set)

Generic Pretrained
Convolutional Model

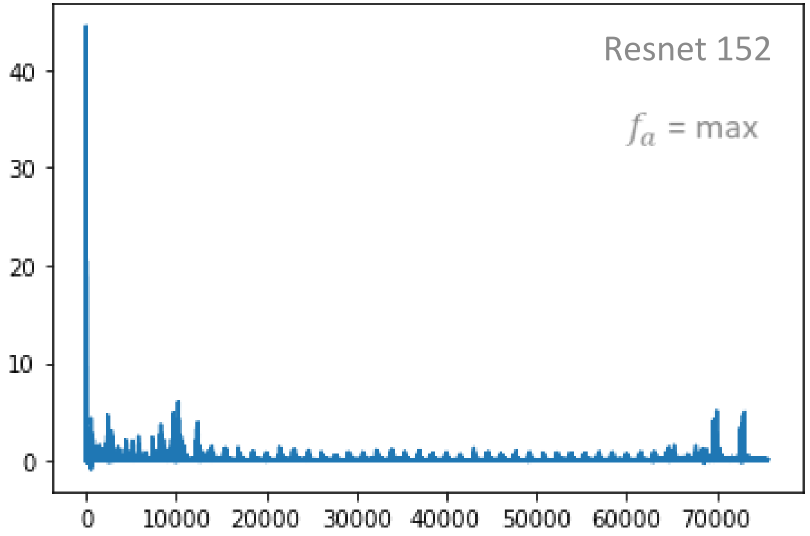


2-D FADS example – inference normalization

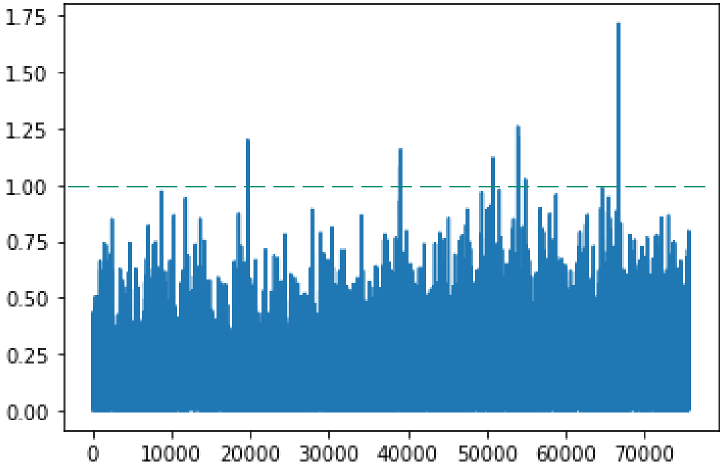
Input image

Generic Pretrained Convolutional Model

Feature activations

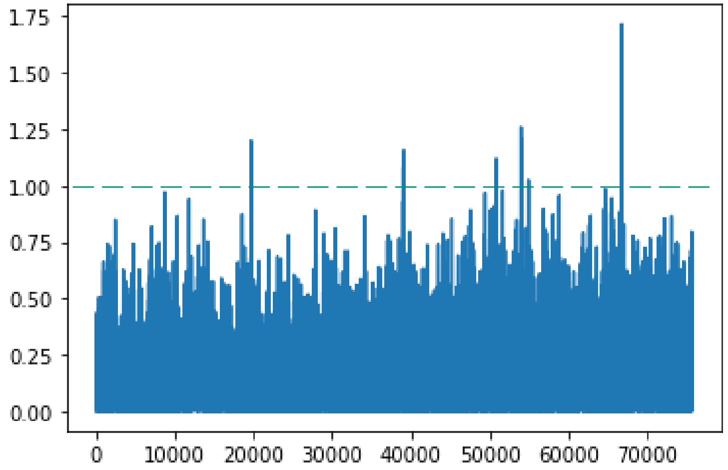


Normalize against “training” set



R-vector (standard deviations from nominal mean)

2-D FADS example – converting to a threshold



R-vector (standard deviations from nominal mean)

Scoring Function

$$s_m = \max(r)$$

Or

$$S_p = percentile(r, 90)$$

Anomaly Scores

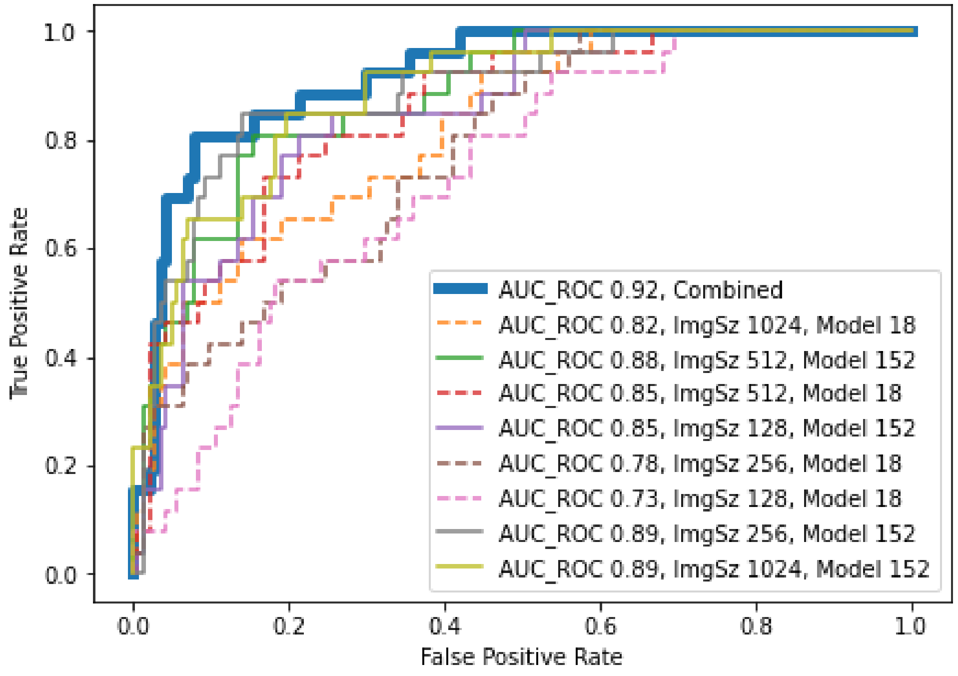
1.75

Or

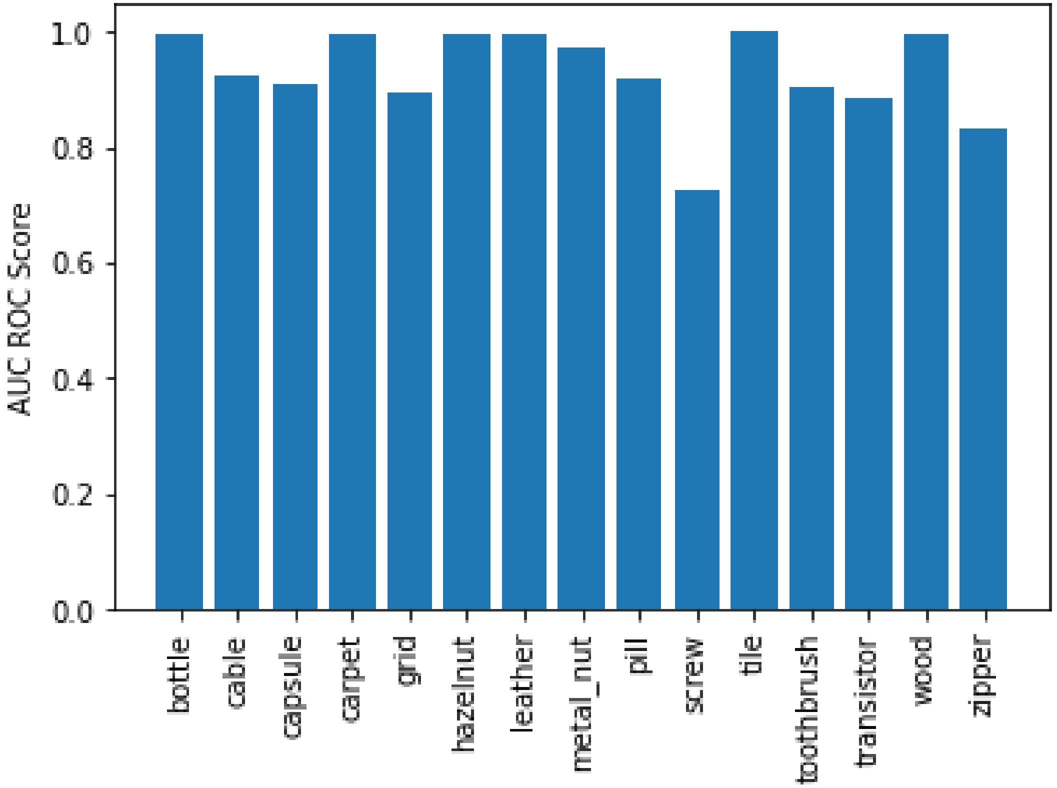
1.1

FADS on MVTec AD dataset (whole image)

ROC for pill category



FADs AUC score on each class

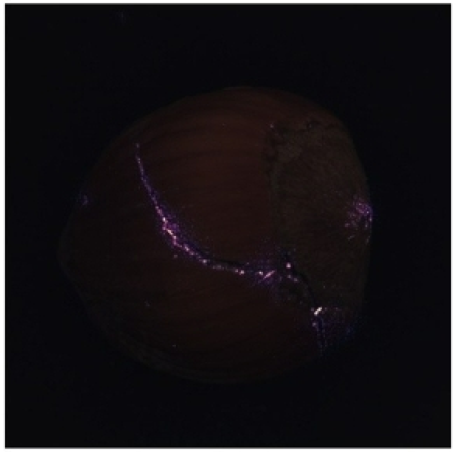


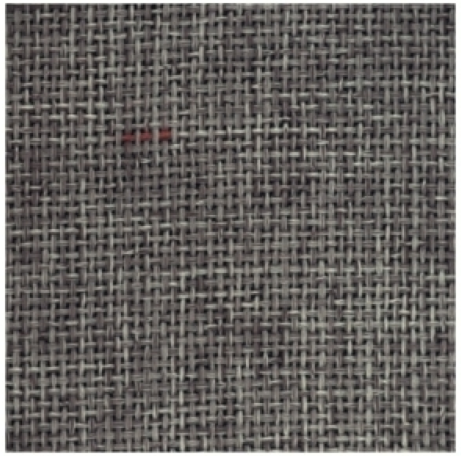
FADs achieves an average AUC of 0.93

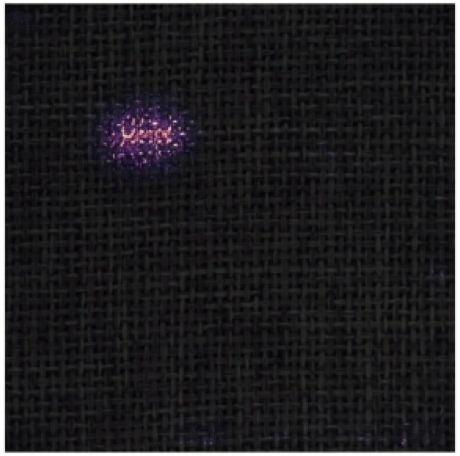
FADS can also highlight the anomalies

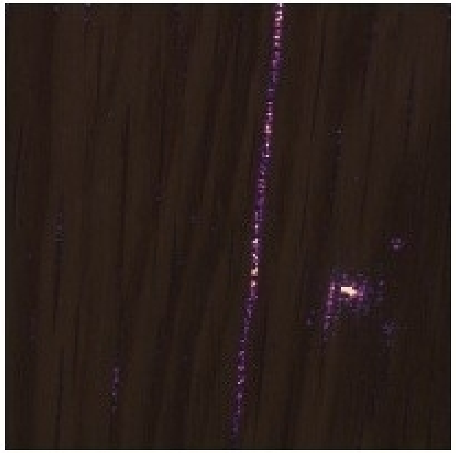
By taking the gradient to minimize the anomaly score with respect to the input, the pixels that contribute to anomalousness are highlighted

- Very sensitive – wood image picks up a scuff that is barely visible for instance
- Still fast as it uses a single backward pass

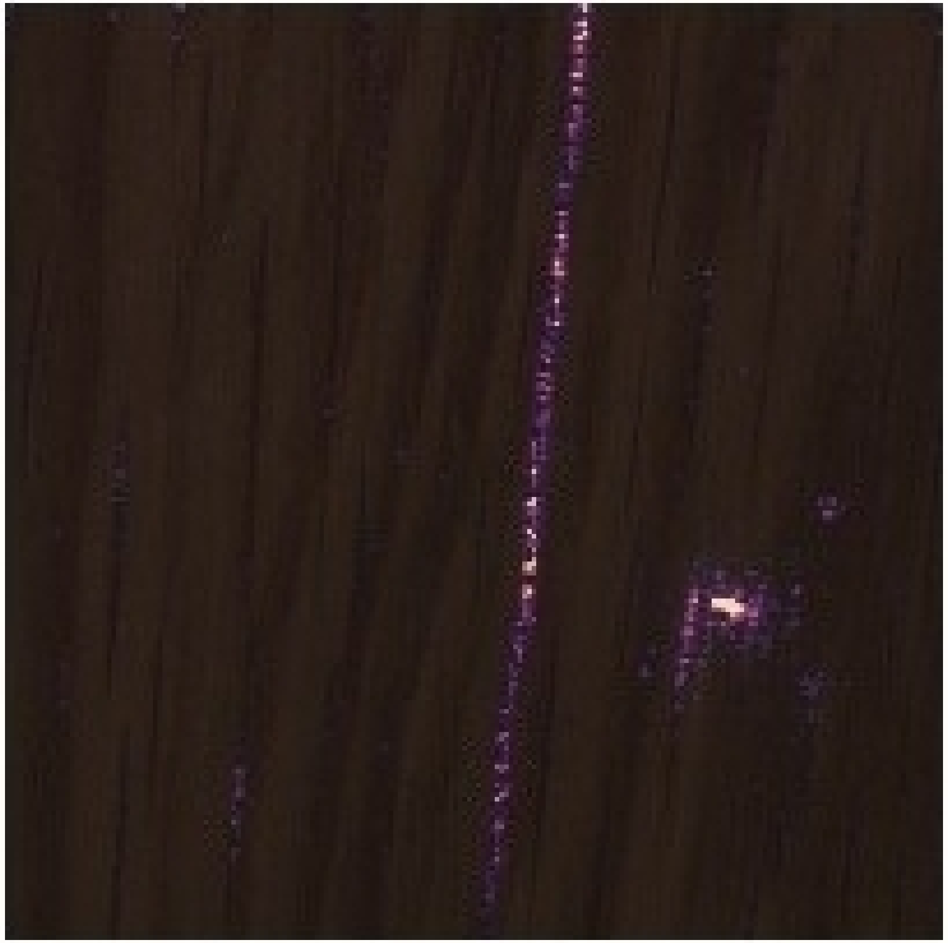








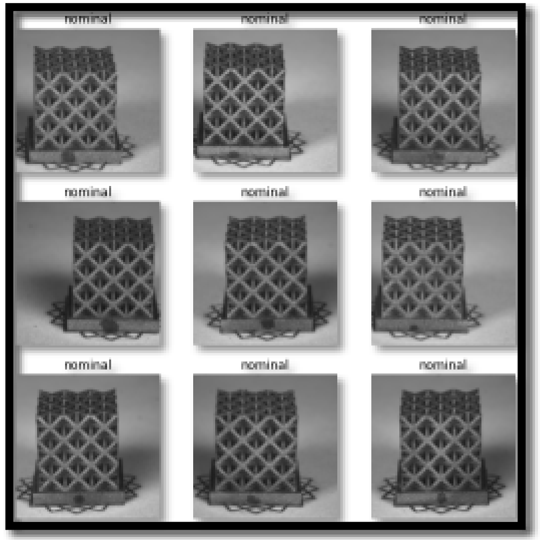
FADS can also highlight the anomalies



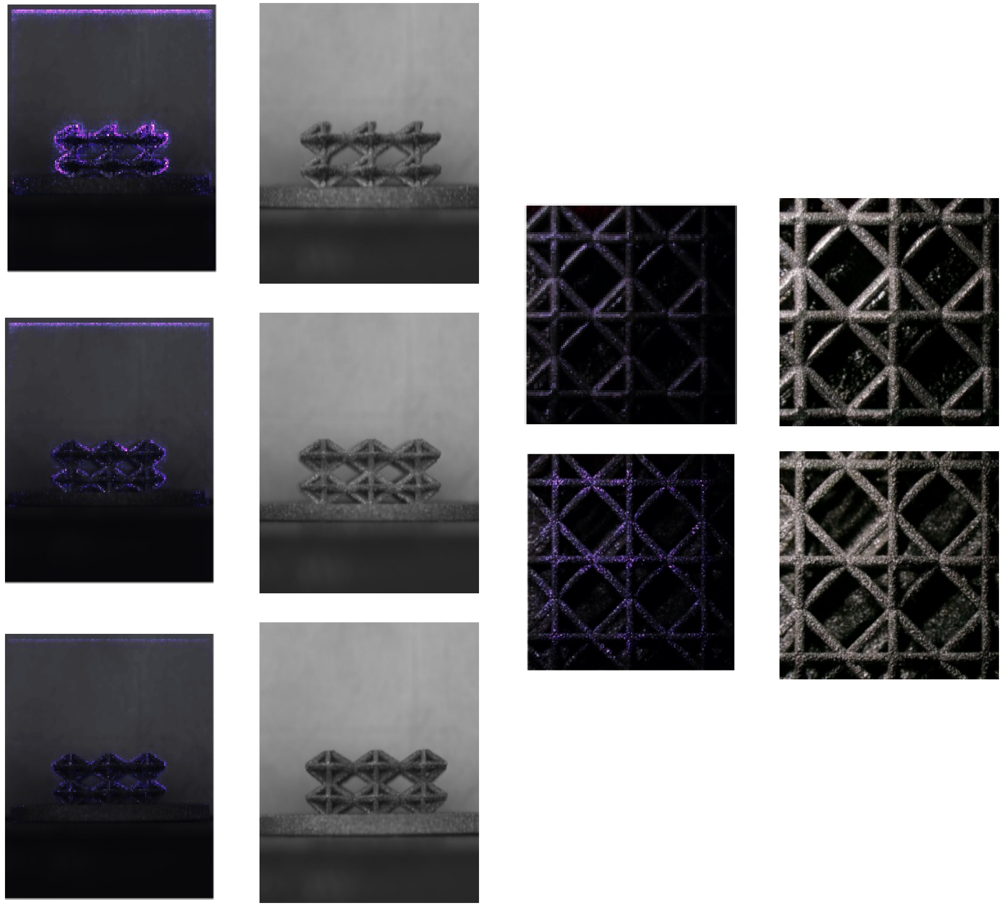
FADS against a real world application

Additive Manufacturing

- (right) Localizing flaws in real prints
- (below) Dedicated print testing: Using just images, identify defective parts with incorrect print process settings
  - "Trained" on 18 lattices
  - Result: Avg AUC of 0.99

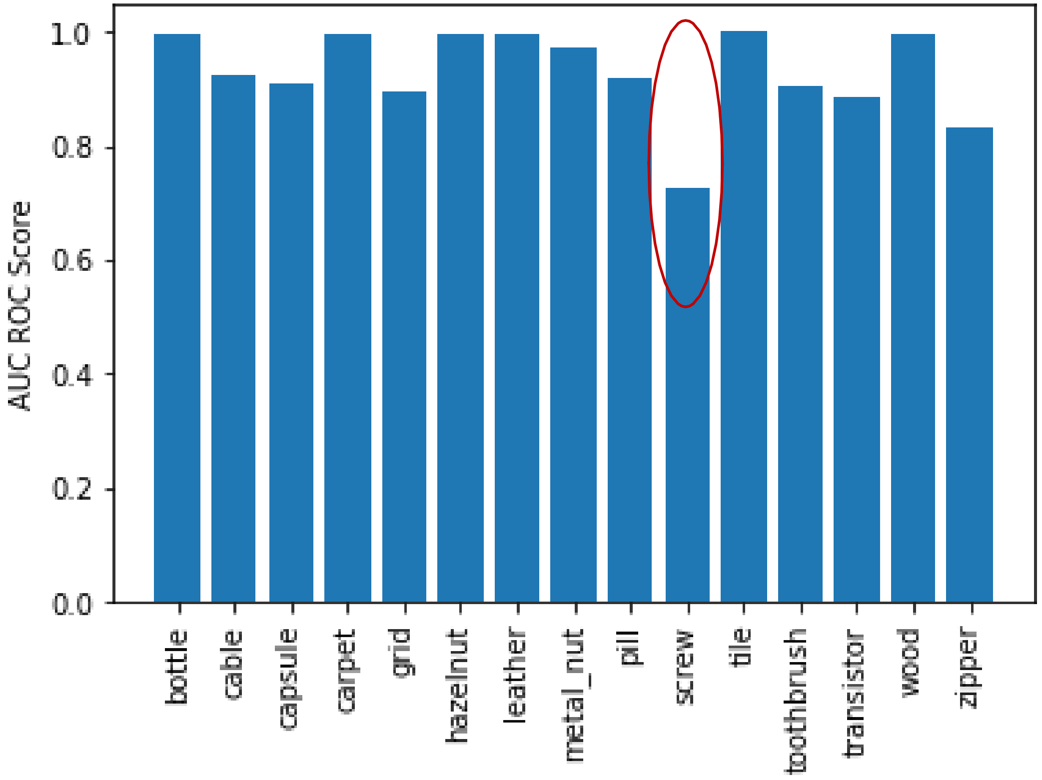


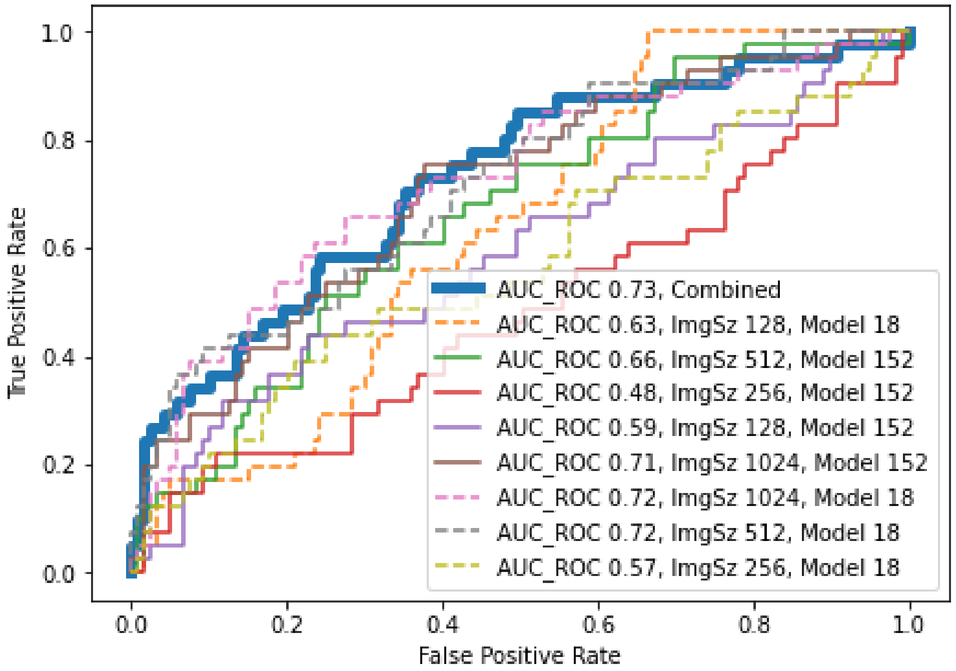
3-D Printed Lattice



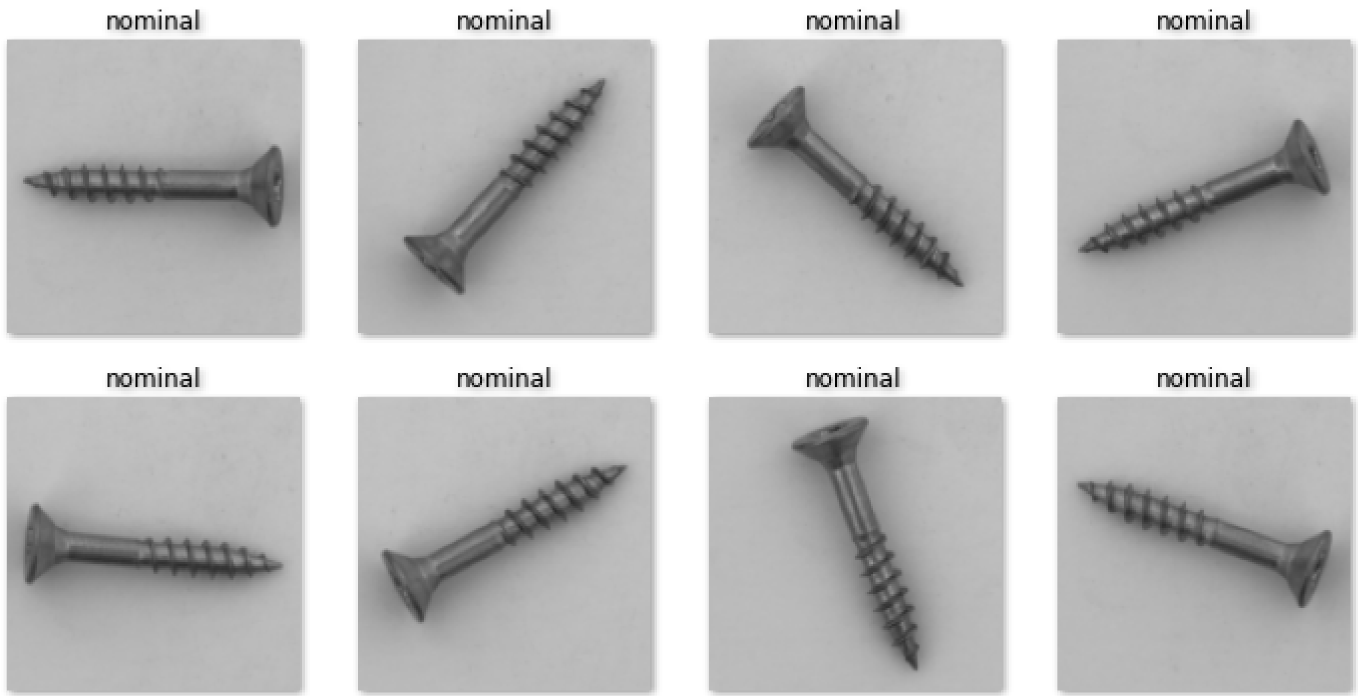
Visualization of the regions causing high anomaly scores

Open challenges remain





Localization gives clues as to the problem



Nominal or abnormal?

In summary

Made models that can detect and localize anomalous regions within images

- Work independently of image source/modality
- 2 and 3-D support

A new method that operates using pretrained networks

Questions?

simlr@sandia.gov