HYDRA

CLAS Collaboration Meeting
11/7/23

TORRI JESKE

roark@jlab.org

THOMAS BRITTON

tbritton@jlab.org

DAVID LAWRENCE
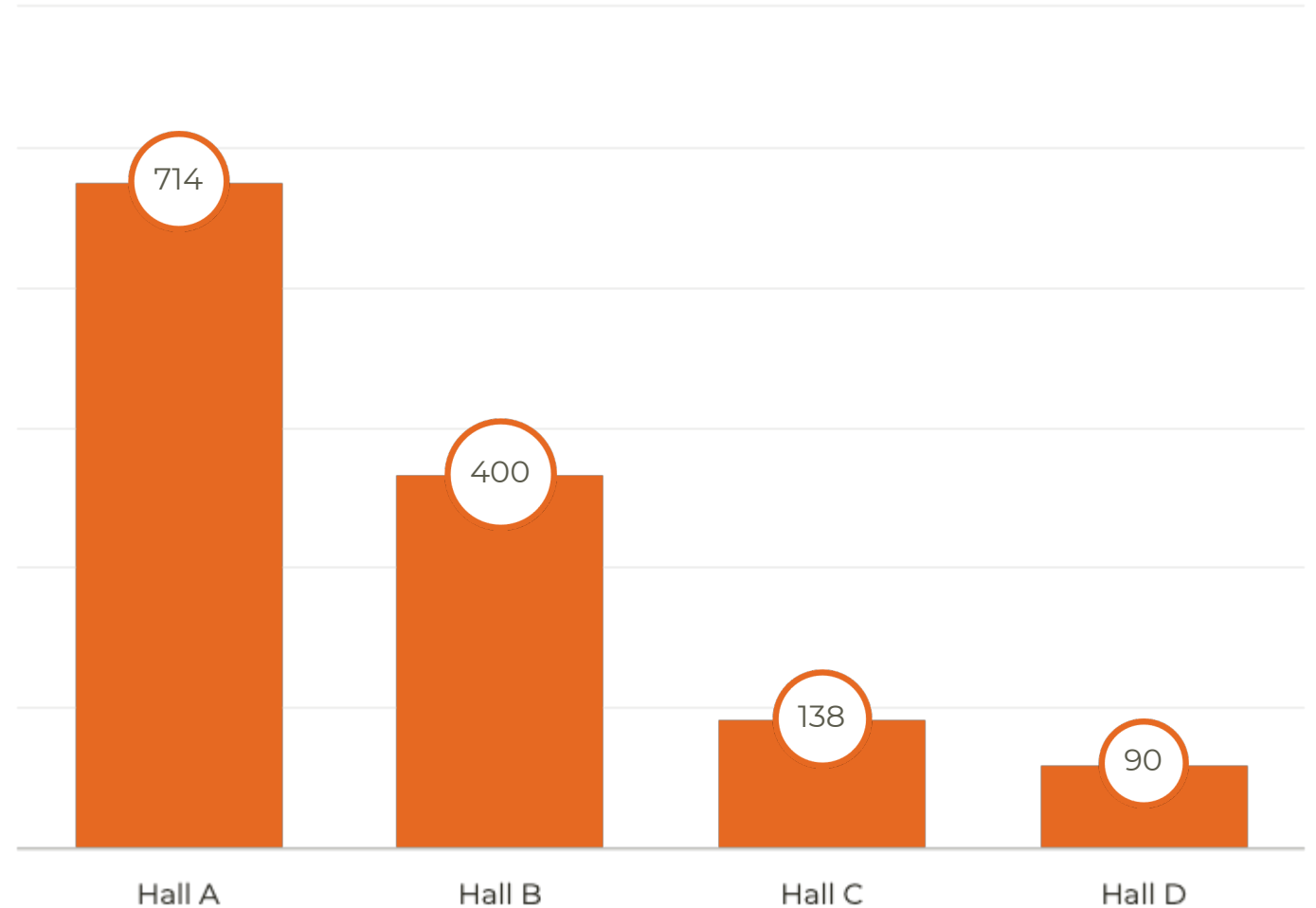
davidl@jlab.org

KISHANSINGH RAJPUT

Online monitoring is **tedious**.

Problem: varying levels of expertise, inconsistent monitoring, too many plots

Problem: Images are good most of the time. Rare "bad" instances are easily missed.
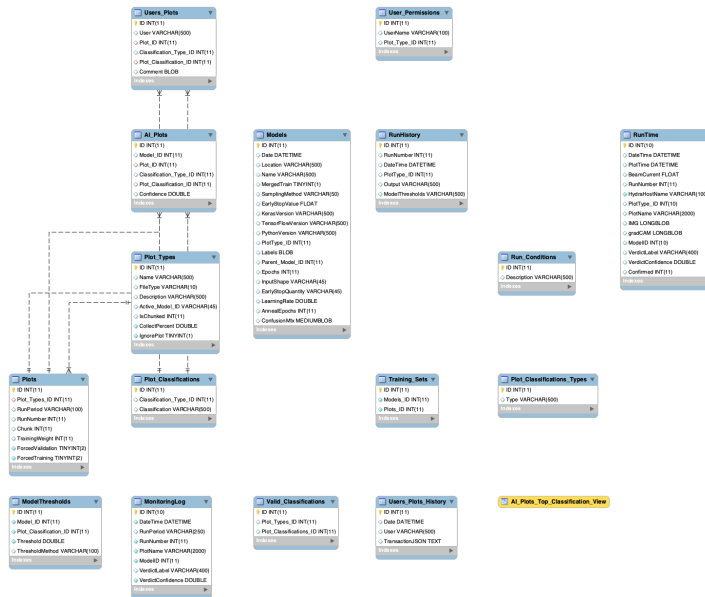
Use computer vision to aid shift crew in all 4 experimental halls.

Approximate **number of individual histograms per experiment per run**, monitored by the shift crew for each experimental hall.
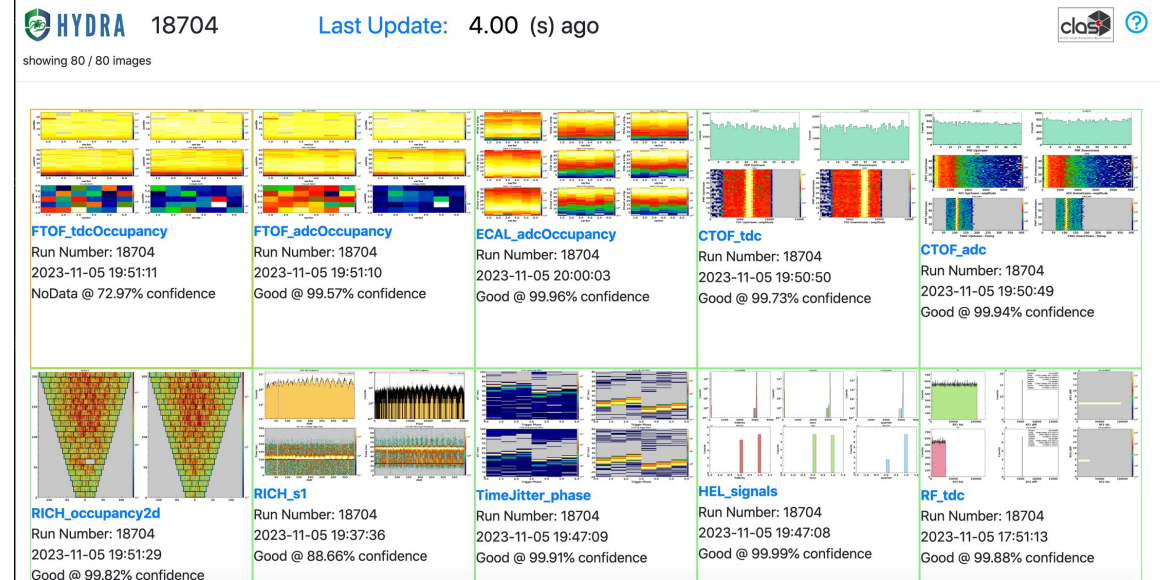
Online Data Quality Monitoring with **HYDRA**

An extensible framework for training, managing, and evaluating AI for real time data quality monitoring.

1/ MySQL back end

Unique plot identifiers, model training parameters, classifications, user permissions, labels, and more are all stored in MySQL database.

2/ Web based front end

Web based front end for labeling, monitoring, and model validation.

EPSCI develops and maintains Hydra in all Halls in collaboration with Users

Online Data Quality Monitoring with **HYDRA**

- ## Correct, frequent classifications

  Train InceptionV3 (default) for each image using a labeled data set

- ## Explainable predictions

  Can the model tell us about its predictions?

  Generate Gradient Weighted Class Activation Maps (GradCAM) to highlight regions of image important for classification

- ## Replacing shift crew duties

  The shift crew should still be following standard monitoring procedures.

  shift crew 🤝 AI

Updates from *last year*

Hydra has undergone a lot of changes!

2022:

Images sent to labeler when shift crew submits to logbook via mon12

Models trained with ~100s of images

Hydra "walked" over input directory

2023:

Images are automatically sent every N electron triggers.

Some models trained with ~1000s of images

GradCAM heat maps available for Bad images online

Hydra now listens for images

Unified front end

Parallel predicts coming *very* soon

80 images every 5 minutes >> GlueX

Hydra for CLAS12

- ## 80 total plot types

  35 images with active model in production

  Remaining either have <u>no bad examples</u> and/or <u>no active model</u>.

  Models with no bad examples for training were pulled from production to avoid the scenario where Hydra cannot classify an image as bad.

- ## Images are "chunked"

  Hydra looks every 10k electron triggers, which is about every 5 minutes

  Events in one chunk are **<u>NOT</u>** in another

- ## Hydra keeps most images for labeling

  This is on purpose - most models are over fit and need a lot more training data.

  Eventually the percentage of images kept will be reduced.

Offline Tools
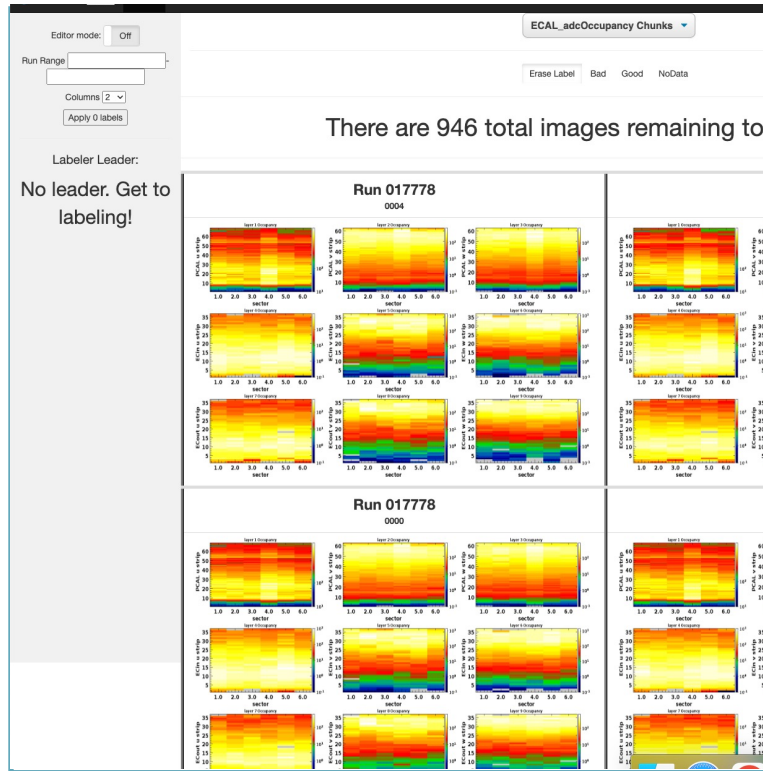
Labeler

Keep labeling images please :)
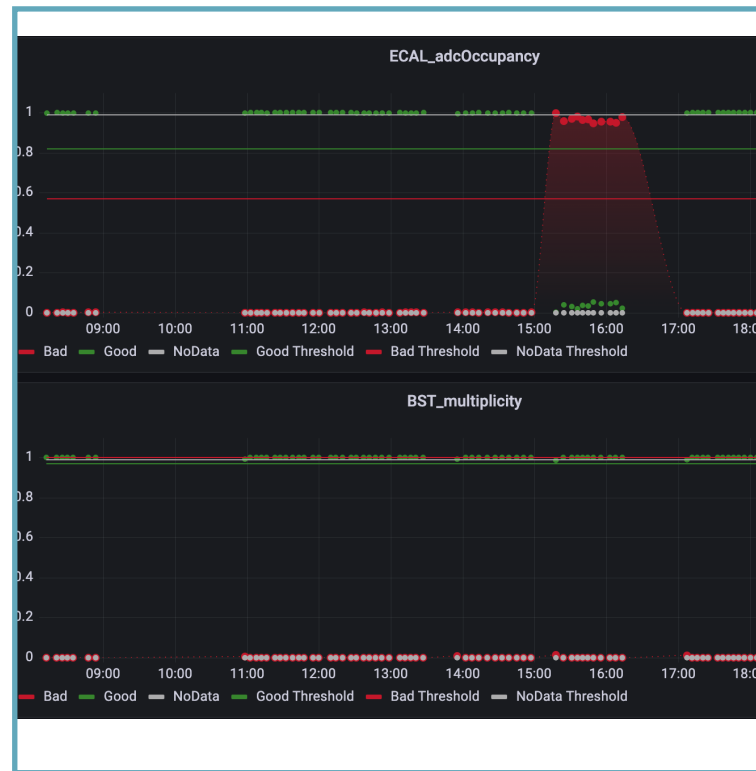
Grafana

See output from model over time. Dots correspond to softmax probability per class.

Library

Confusion matrix about each trained model, adjust thresholds, etc

HYDRA: Data Labeler

Efficiently label multiple monitoring plots. Labels and images are automatically uploaded to database.
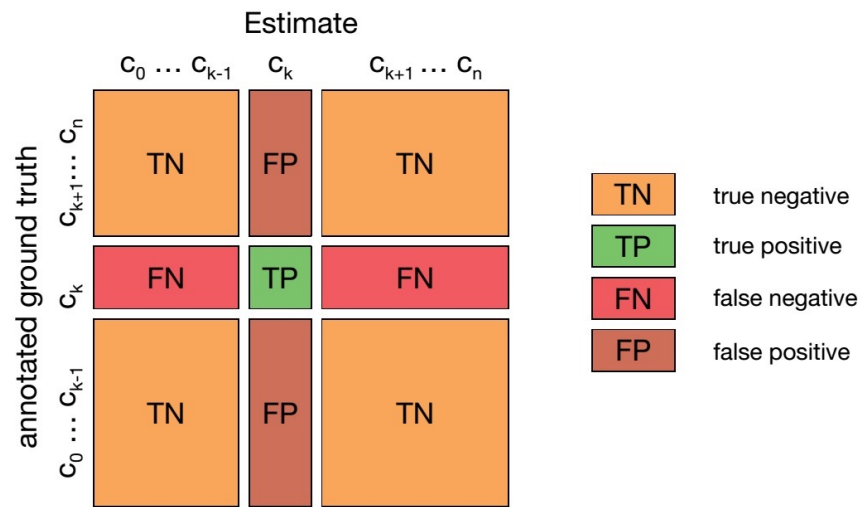
Most plot types have ~1000s of images to label.

Please label ☺

HYDRA: Library

Visualize model performance, thresholds, active models, etc.

Threshold: based on max f1 score per class by default.

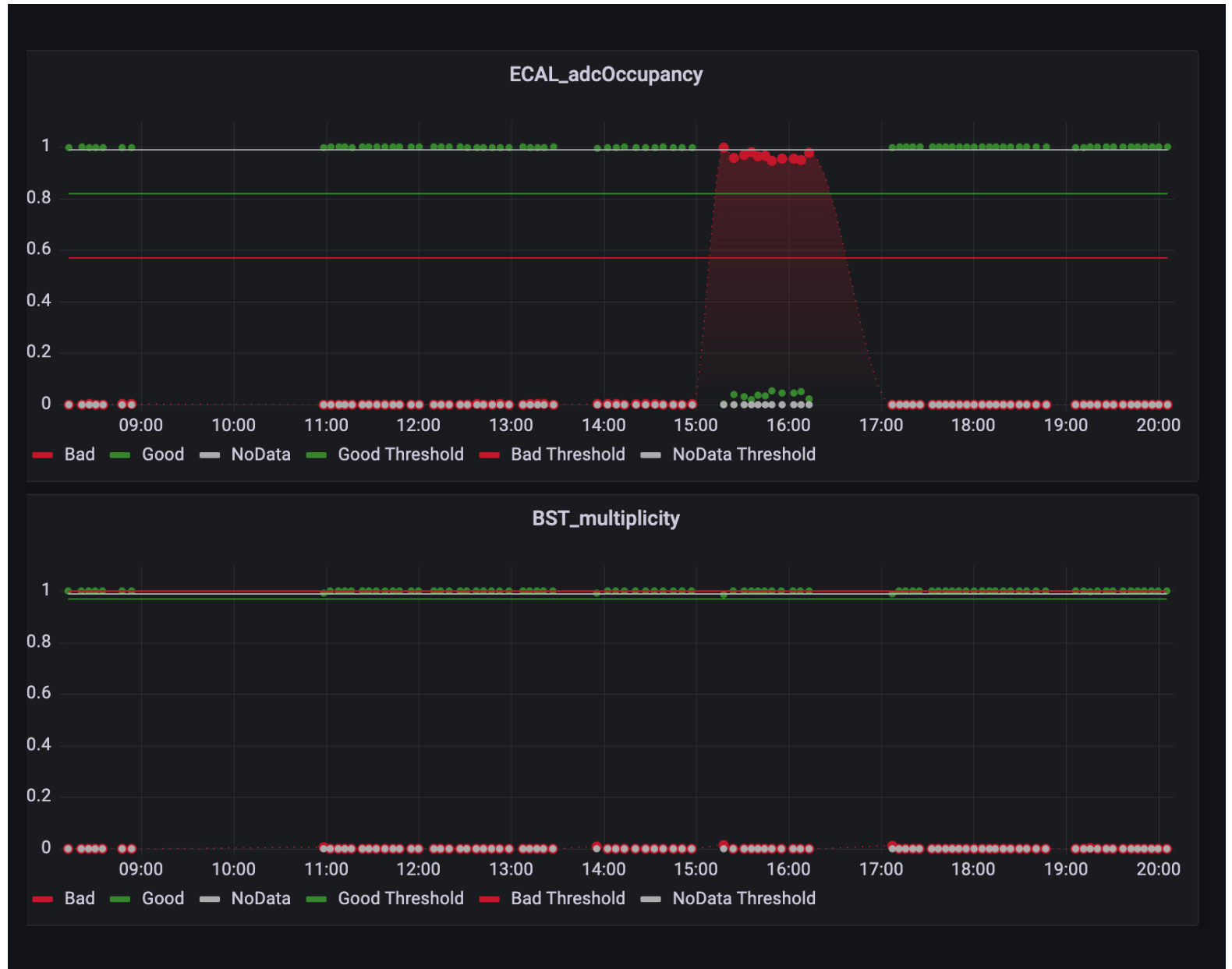# HYDRA: Grafana Dashboard

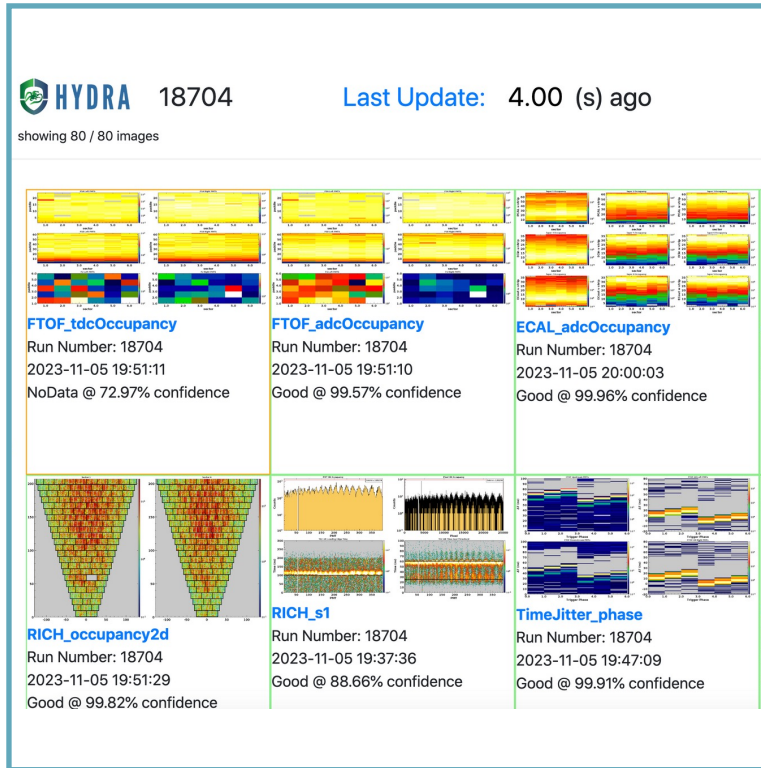Look at any prediction over time.

Y axis: "Confidence" – output from model inference

Points per inference correspond to number of labels

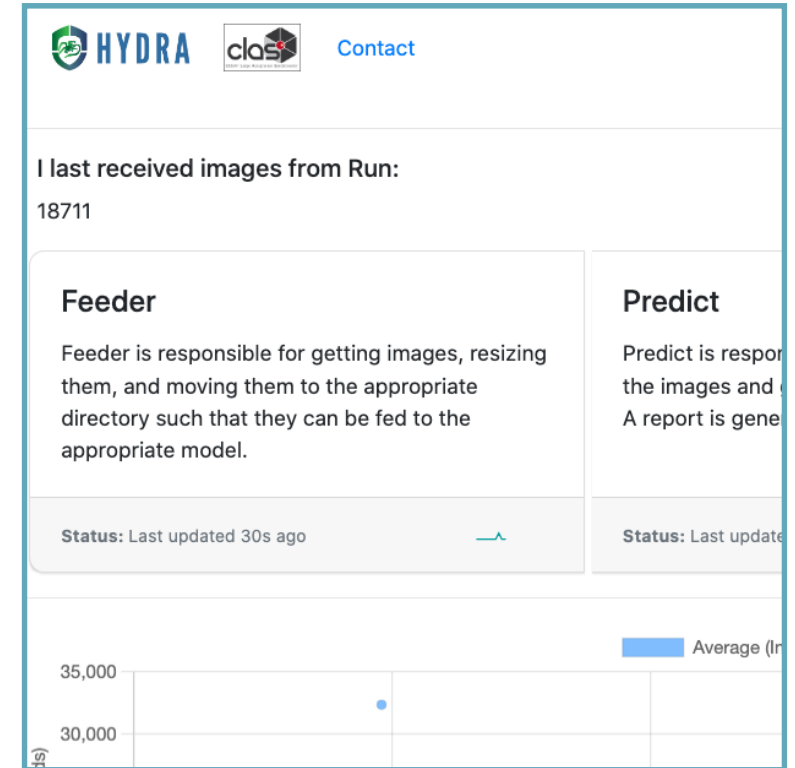Lines correspond to threshold values

Online Tools

Run

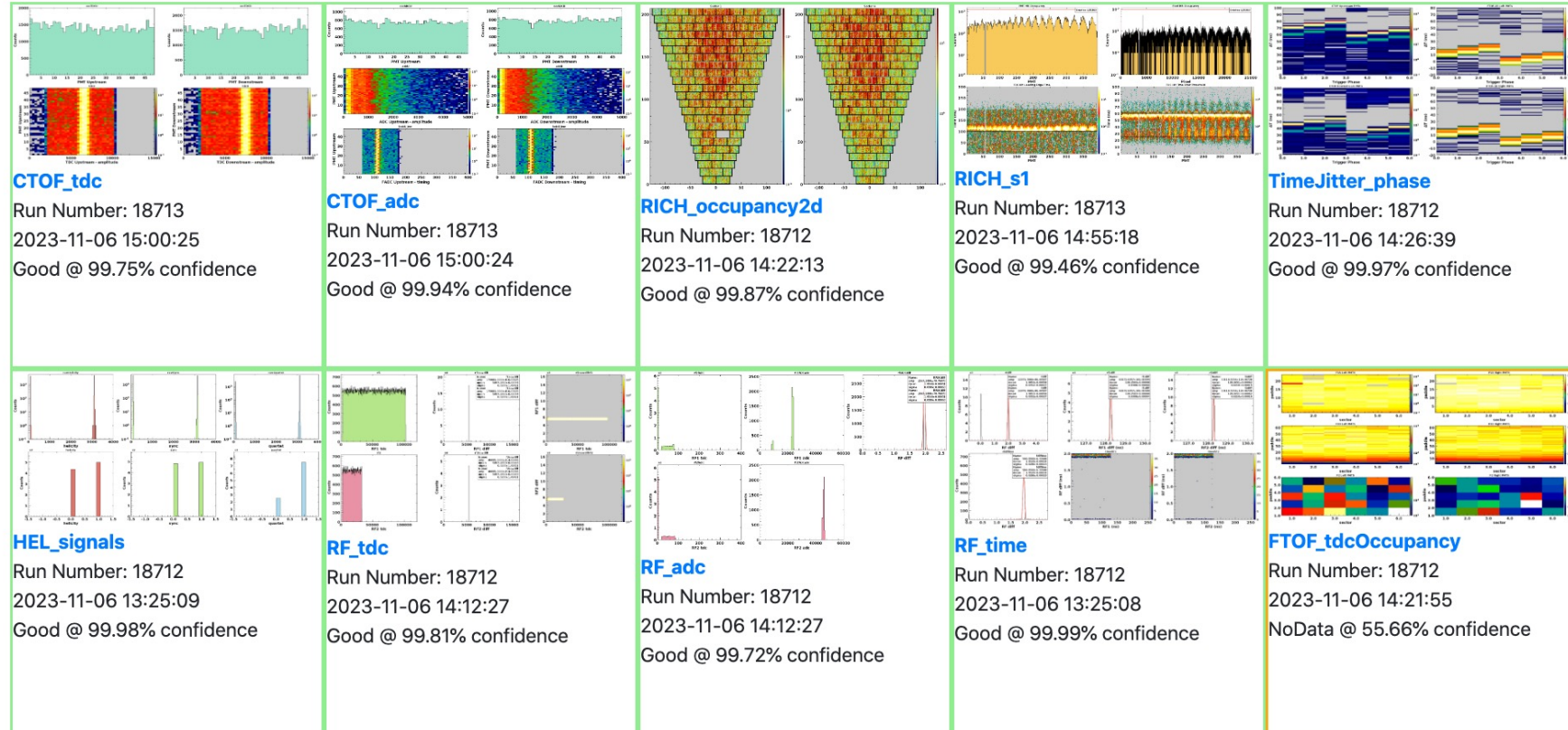Real time classifications. **Should be up in counting house!**

Log

Bad or potentially concerning plots for the previous 24 hours

Status

See if back end processes are running, time it takes for an image to go through Hydra, etc

HYDRA: Run

Watch predictions in real time from anywhere

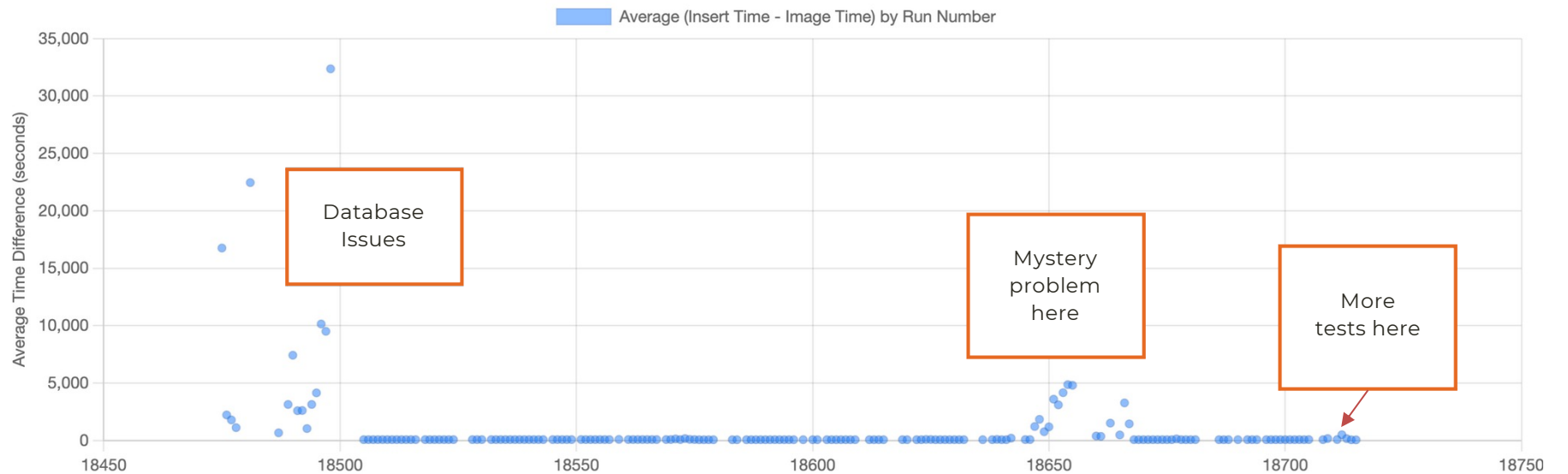GradCAM visualizations are available for Bad images

Dashboard

Shows status indicators for Hydra

Preliminary benchmark: Average time (by run number) for each image to flow through Hydra

I last received images from Run:

18715

Feeder

Feeder is responsible for getting images, resizing them, and moving them to the appropriate directory such that they can be fed to the appropriate model.

Status: Last updated 30s ago

Predict

Predict is responsible for running inference on the images and generating gradCAM heatmaps. A report is generated for each inference.

Status: Last updated 30s ago

Keeper

Keeper is responsible for analyzing the report written by predict. It sends images to the labeler based on the designated collect percentage.

Status: Last updated 30s ago

Average (Insert Time - Image Time) by Run Number

Database Issues

Mystery problem here
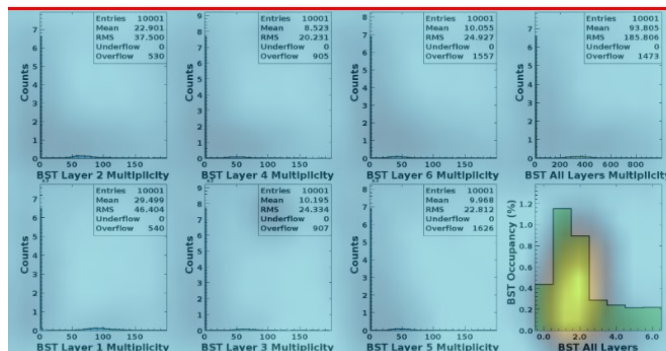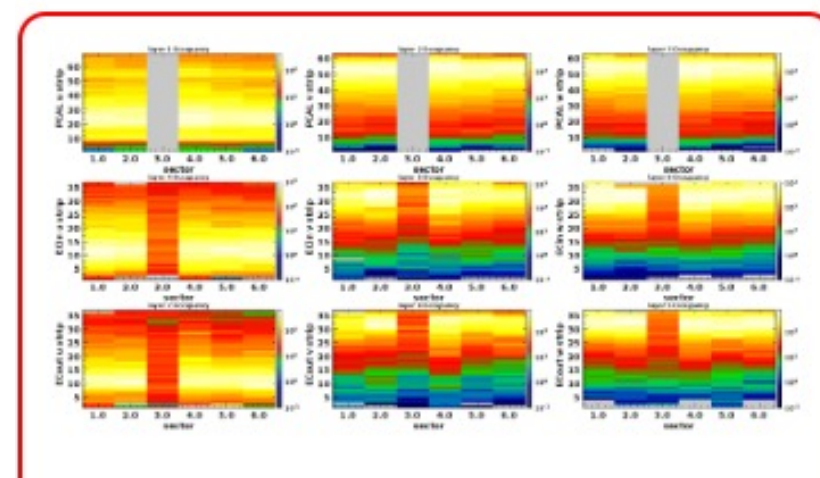
More tests here

HYDRA: Log
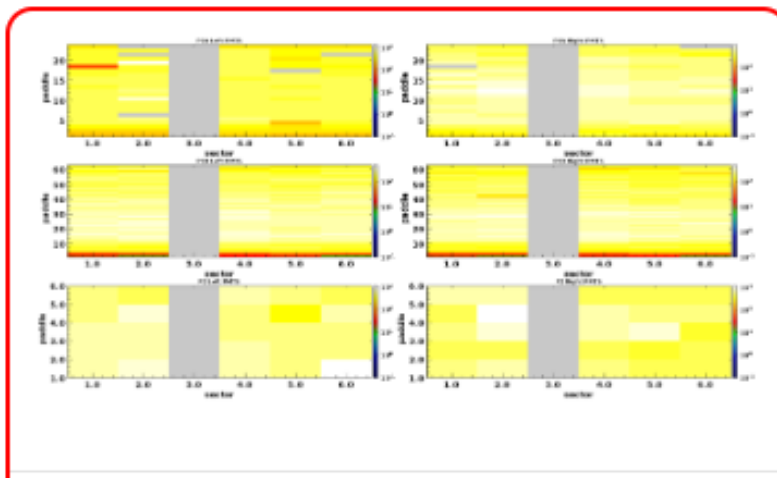
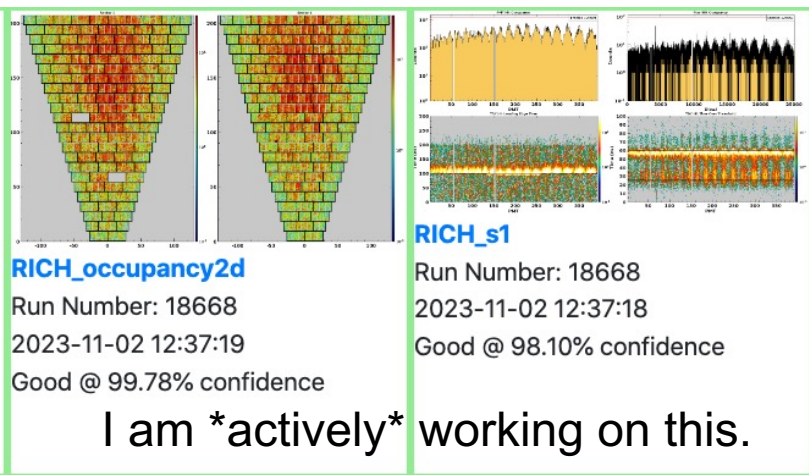See 'Bad' or 'Unconfirmed' images from previous runs, separated by detector.

Easily identify when a particular problem started

Performance so far
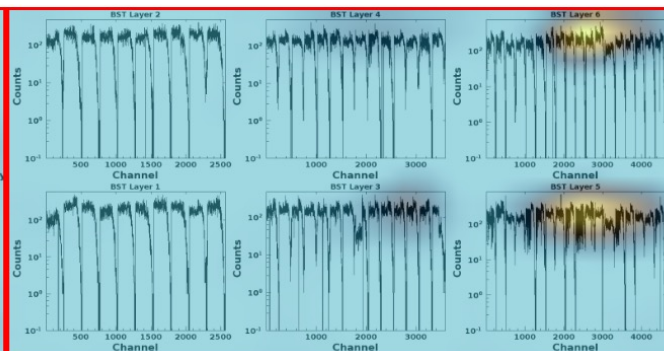
Issues Hydra has caught AND missed

RICH_occupancy2d
Run Number: 18668
2023-11-02 12:37:19
Good @ 99.78% confidence

RICH_s1
Run Number: 18668
2023-11-02 12:37:18
Good @ 98.10% confidence

I am *actively* working on this.
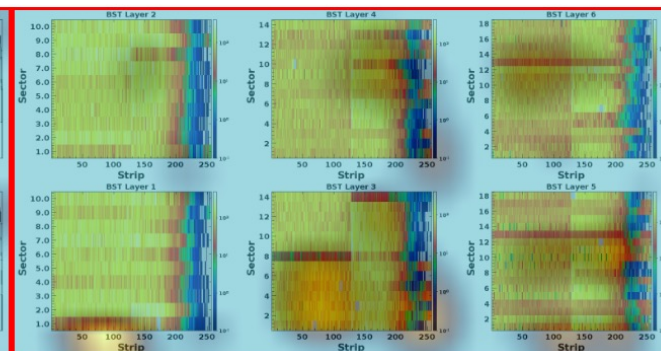
BST_multiplicity
Run Number: 18356

BST_hits1d
Run Number: 18356

BST_hits2d
Run Number: 18356

To Do

Visual Indications of Hydra health

Give feedback for common errors

Tutorials

I'm a new graduate student/shift taker, how do I use Hydra?

Playground

Have a place to train a model on labeled data set

Test different models, training methods

In Progress

Retraining!

As we get more labels, I retrain models.

Parallel Hydra

Multiple "copies" of back end processes running at the same time

Conclusions + Acknowledgements

- Hydra is under active development!

- Hydra is healthy overall!
  Parallel Hydra will be deployed next maintenance day!
  Models still need considerable training. Please label.

- Thanks to Nathan Baltzell, Dan Carman, and labelers for being a part of getting Hydra implemented.

- We would love your feedback!

all web pages can be found at clasweb.jlab.org/hydra/

- How can I delegate labeling responsibilities?

  Just send Thomas (tbritton@jlab.org) or myself (roark@jlab.org) an email with the appropriate JLab username.

- Can we train our own model?

  Sure! InceptionV3 is the default. We can experiment with different models.

- Why didn't Hydra catch this problem?

  1. Has it seen similar problems before?
  2. Did the model have a "bad" label available?
  3. Did the confidence for the classification drop?

- Can we contribute to Hydra?

  Absolutely.

FAQs

Current Labeling Stats