



HYDRA

Computer Vision for Data Quality Monitoring

Thomas Britton, Torri Jeske

David Lawrence

Kishansingh Rajput

Introducing Hydra

- Hydra aims to be an extensible framework for training and managing A.I. for near real time monitoring

- If you need it to tell a dog from cat I can have hydra do that, without system modification, now

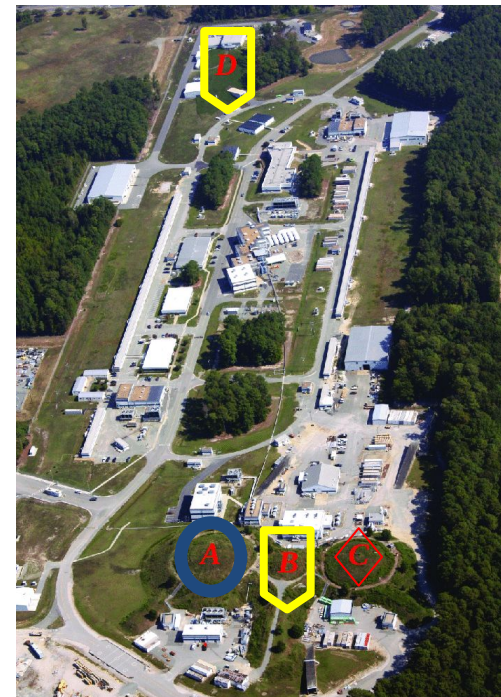
- Most importantly, Hydra allows me to embrace my inner sloth:



Koboldpress.com

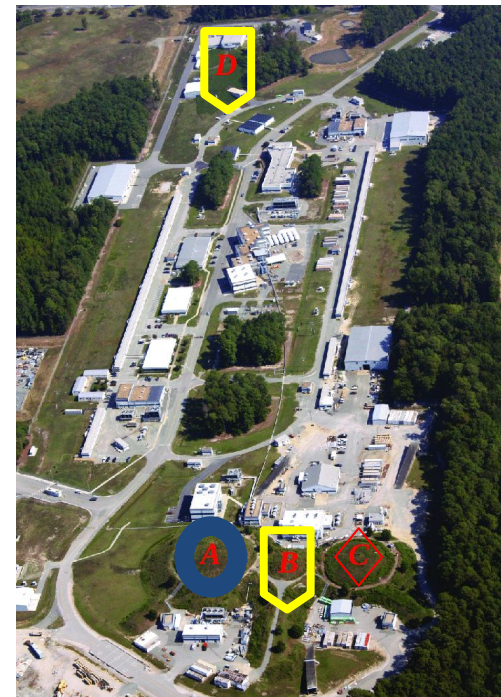
Deployed site wide

- Instance of Hydra deployed in all halls
 - Fully in B and D ◻
 - In A but not being fully utilized
 - Technically in Hall-C ○
 - Working to make its adoption more green-field ◊



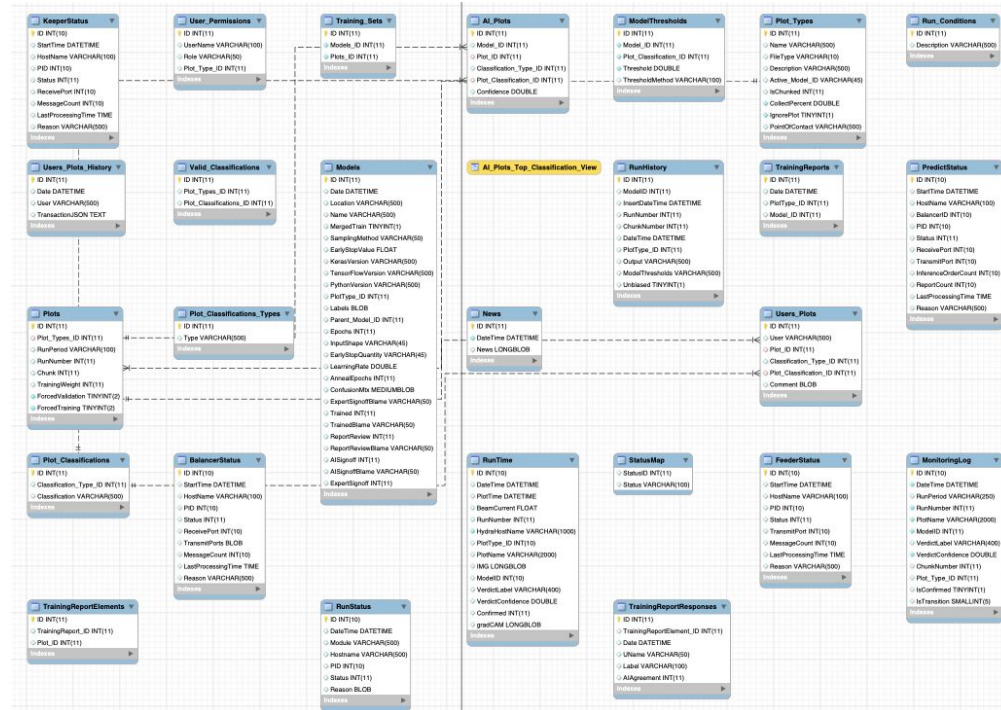
Challenges

- **Differing protocols**
 - Not everyone is set up to provide images, or in the granularity required
- **Differences in scale**
 - GlueX has about a dozen plots.
CLAS has 81
- **Requires some degree of buy-in from disparate groups**



Back End

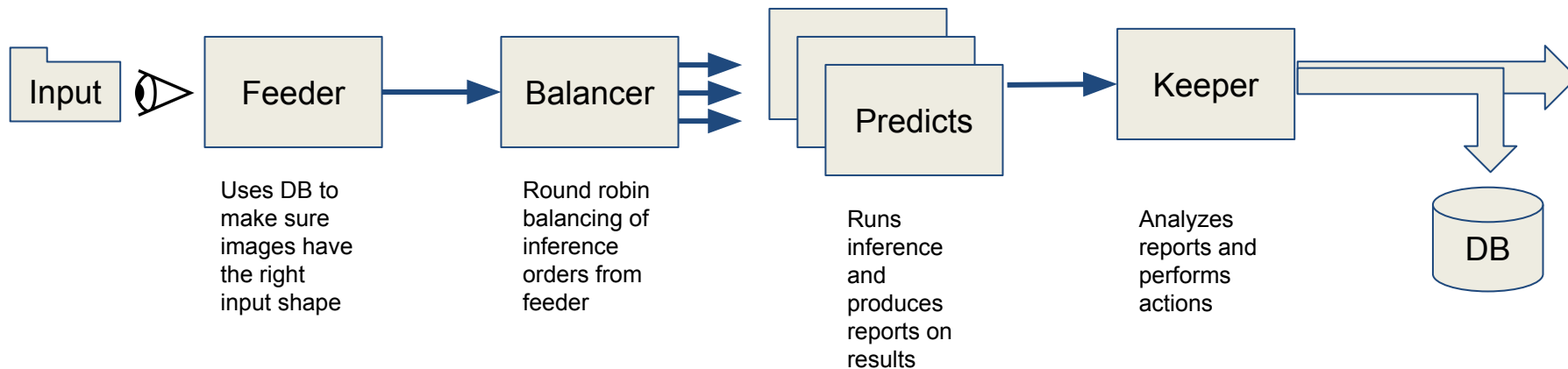
- Hydra is supported by a MySQL database which holds both training data and operational data
- Model configurations are also stored/read from the database
- Saves a snapshot of every inference and training
 - A whole host of data to analyze and utilize



Hydra System

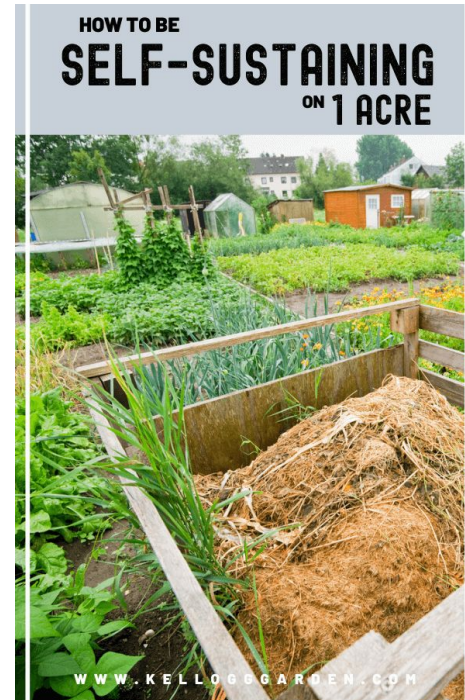
- Parasitic to normal operations
 - Meant to aid not replace!
- Feed images into the input directory and Hydra handles the rest

→ Messages passed via OMQ allowing for more distributed deployment



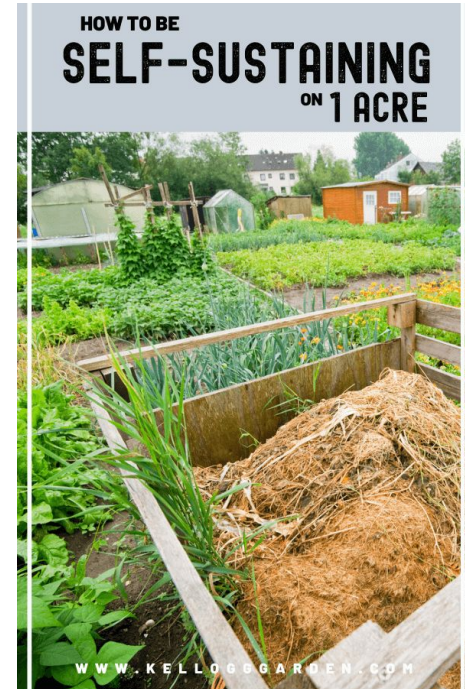
Self-sustaining

- The goal is to provide a system which is more or less self-sustaining
 - Further Education
 - Unbiased, configurable, sampling
 - Unconfirmed, Bad examples
 - How is a model performing?
 - If labeling is up-to-date we can leverage the unbiased selection to compute a trailing accuracy



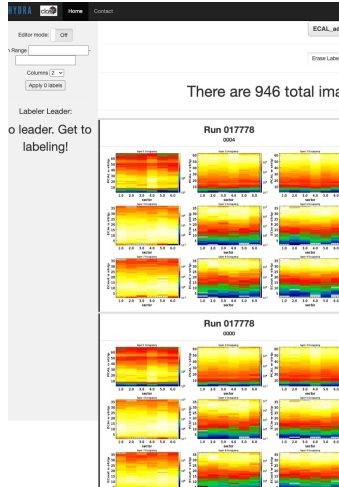
Self-sustaining cont.

- Possible to trigger retrainings based on changes in the running accuracy
 - Other conditions?
- Developing methods to enable better administration of the Hydra system
 - Monitoring Hydra



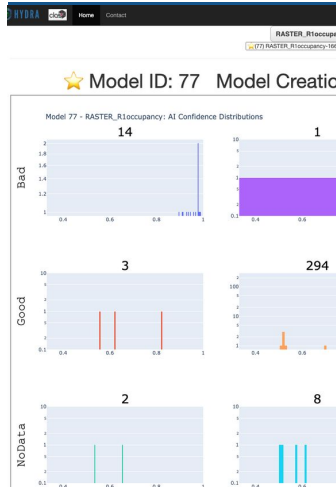
HYDRA: Front End

Web based for user convenience.



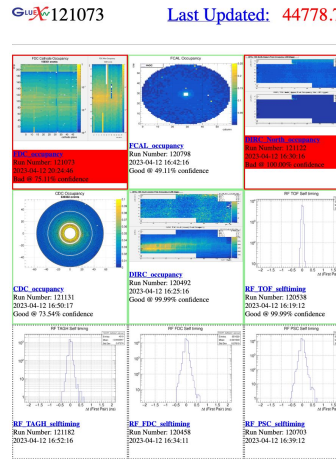
Data Labeler

Efficiently label hundreds (thousands) of images



Library

Contains enhanced confusion matrix, thresholds, active model designations



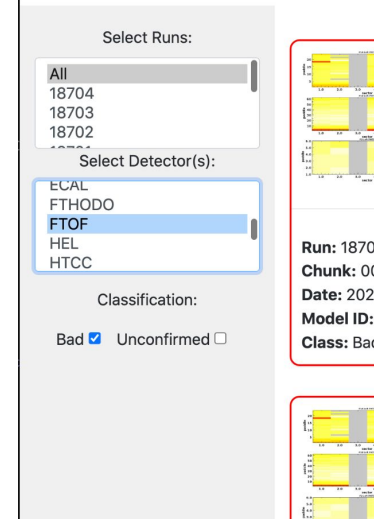
Run

See near real time predictions



Grafana

Dashboard displays all predictions over time



Log

Display concerning plots sorted by detector from previous day

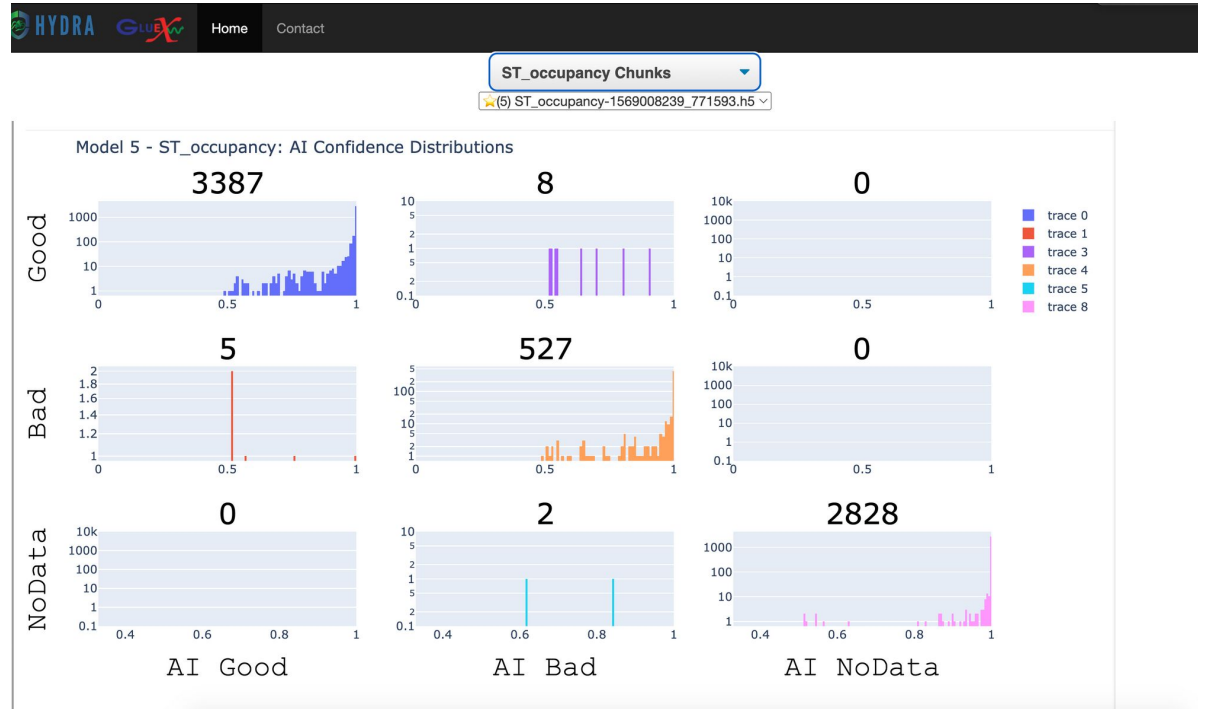
HYDRA: Data Labeler

Efficiently label multiple monitoring plots. Labels and images are automatically uploaded to database.

The screenshot displays the HYDRA Data Labeler web interface. At the top, there is a navigation bar with the HYDRA logo, a 'cloud' icon, and links for 'Home' and 'Contact'. Below the navigation bar, the interface is divided into several sections. On the left, there is a control panel with the following elements: 'Editor mode: Off', a 'Run Range' input field, a 'Columns' dropdown menu set to '4', and an 'Apply 0 labels' button. Below this, the 'Labeler Leader' section indicates 'No leader. Get to labeling!'. The main area of the interface shows a grid of 8 monitoring plots, arranged in two rows of four. Each plot is titled 'Run 017778' or 'Run 017777' with a sub-identifier (0004, 0002, 0000, 0006). The plots display various data visualizations, including time-series plots and bar charts. At the top right of the main area, there is a dropdown menu set to 'LTCC_tdc Chunks' and buttons for 'Erase Label', 'Bad', 'Good', and 'NoData'. A large message in the center of the main area states 'There are 811 total images remaining to be labeled'.

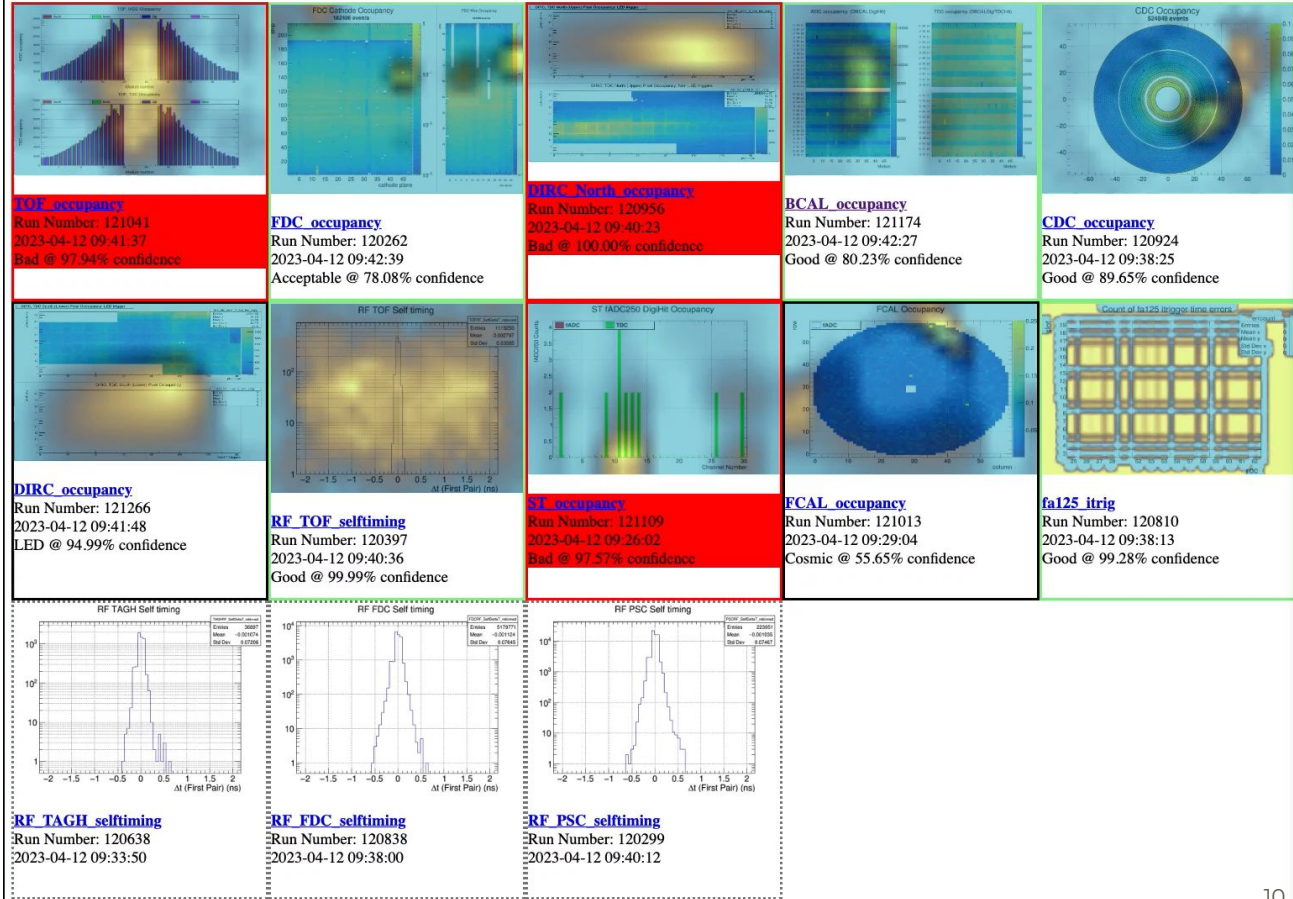
HYDRA: Library

Visualize model performance, thresholds, active models, etc.



Enhanced confusion matrix

Each cell contains AI confidence distribution and total counts



HYDRA: Run

Watch predictions in real time from anywhere

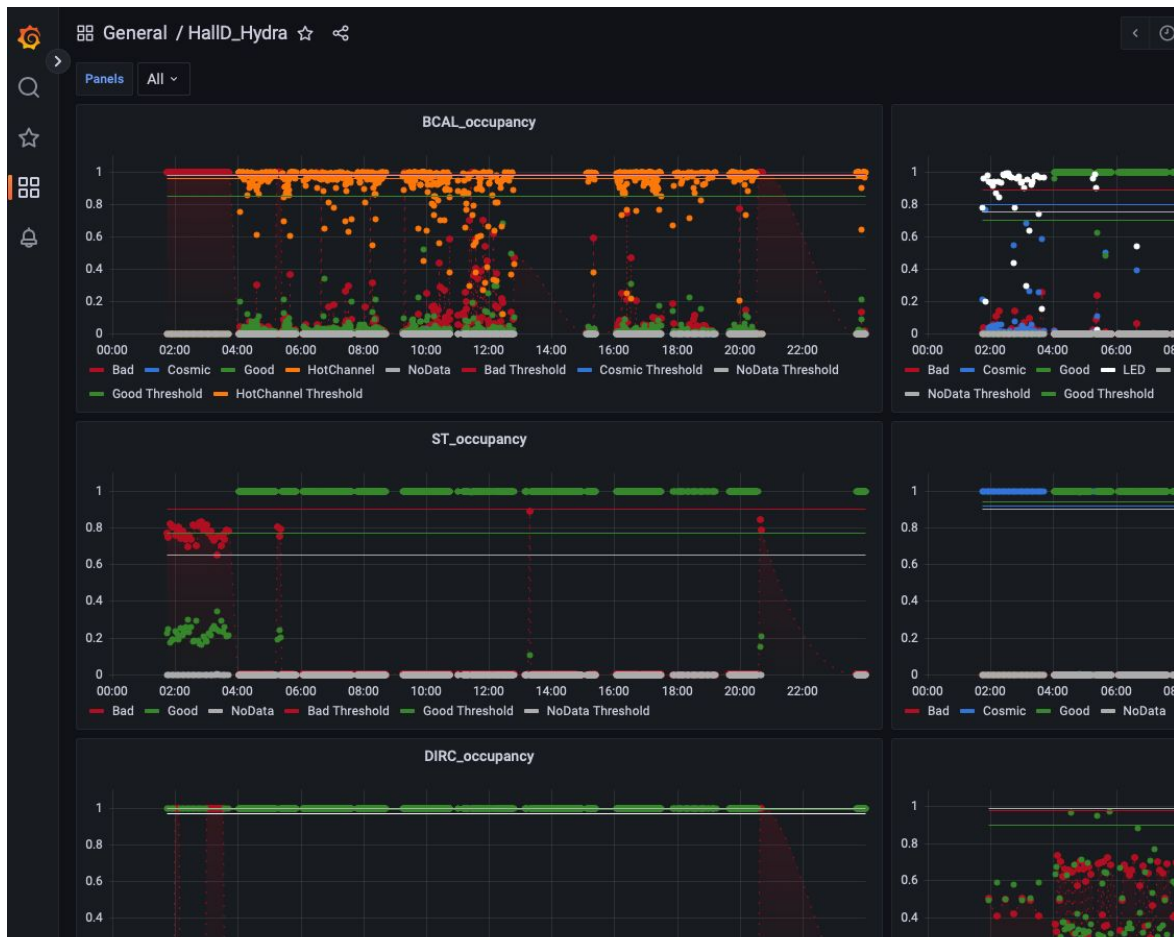
GradCAM visualizations are optional

HYDRA:

Grafana

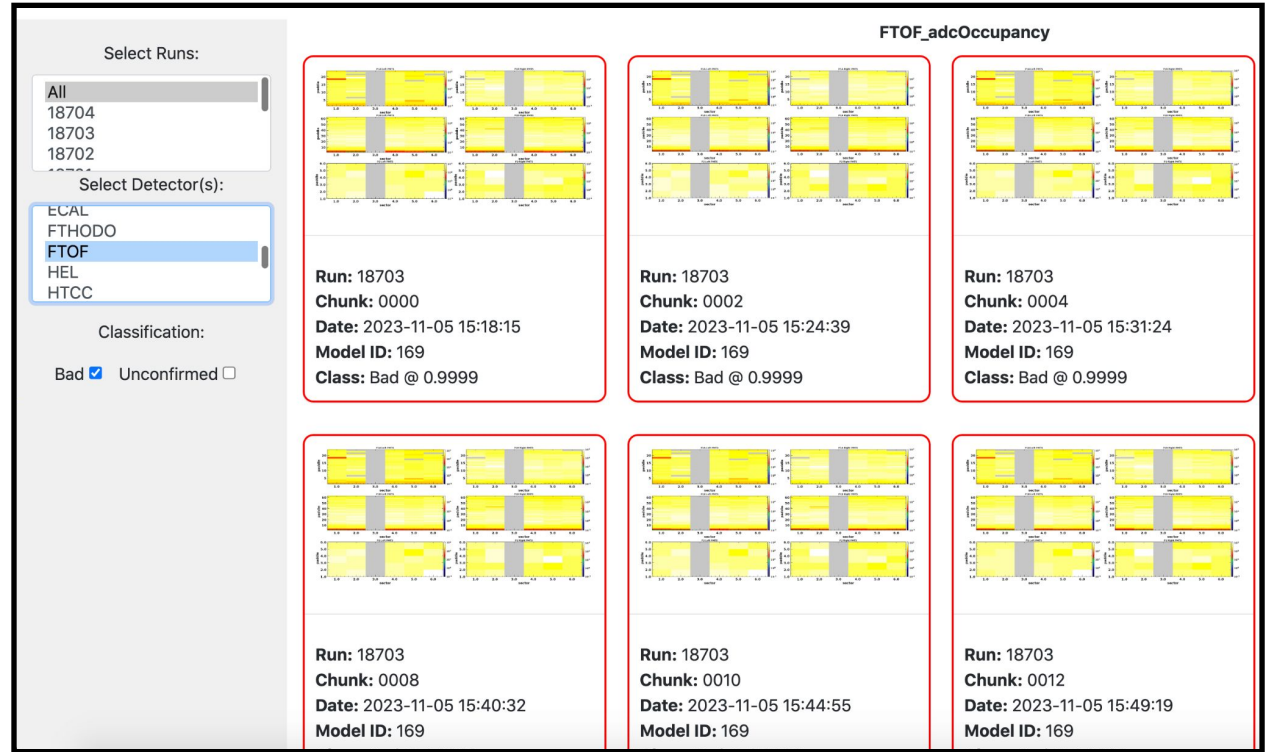
Dashboard

Look at any prediction over time.



HYDRA: Log

See 'Bad' or 'Unconfirmed' images from previous runs, separated by detector. Easily identify when a particular problem started



Other Future Hydra developments

- Siamese model(s)
 - Two step processing
 - First a generic anomaly approach
 - Second is a diagnostic step in the face of anomalies
- Masking
 - Ability to ignore parts of an image in inference
 - known/accepted problems aren't considered
- Kubernetesization
 - Use Kubernetes for easier maintenance/deployment
- AI human interface improvements
 - Enable non AI experts to maintain/manage the hydra system

Chang Woo Lee



Hydra alarmed on the FA125 itrigr errors. Indeed, during run 101072 after about 8.8 M events, the error occurred. We immediately ended the run and restarted R 101073. The error had cleared.

Kudos to Thomas, Naomi and Hydra!

p.s. We plan to change the SD board within the next hour.

HydraRun also saw the FDC problem, which I probably would have missed inspecting it by eye.

Hydra alarm on fa125_itrigr during Run 101076

Questions?