radiasoft

RadiaSoft™ is an industry leader in high-level research & design and scientific consulting for beamline physics and machine learning.

Who are we?

Matthew Kilpatrick
Associate Scientist
kilpatrick@radiasoft.net

RadiaSoft headquarters:
1790 38th St., Suite 306
Boulder, Colorado

Supporting US Industry

RadiaSoft supports high-tech industry in the US, including applications in medicine, agriculture, energy and homeland security:

radiabeam TECHNOLOGIES

PROTOM proton therapy technologies

BARTOSZEK ENGINEERING

NUSANO

REVEAM

MODERN ELECTRON

Supporting research labs around the world

Lawrence Livermore National Laboratory

Argonne NATIONAL LABORATORY

OAK RIDGE National Laboratory

Fermilab

CERN

J-PARC

BROOKHAVEN NATIONAL LABORATORY

Jefferson Lab

Education & Training

RadiaSoft collaborates with universities and training programs:

- U.S. Particle Accelerator School
- Korea University
- UCLA
- Strathclyde University

RadiaSoft personnel volunteer to teach at the
US Particle Accelerator School (USPAS)

supports the SAGE Summer Camp:

https://conf.slac.stanford.edu/sage

Small Business Innovation Research (SBIR) Grants

- Small businesses to engage in Federal Research/Research and Development (R/R&D)
  - Highly competitive programs that encourage domestic
  - Potential for commercialization
  - Collaboration with National Research Institutes, e.g. ORNL
  - Multiple phases of research funding

- Two projects currently funded with ORNL
  - Automated sample alignment within a beamline
  - Automated bragg peak detection with interactive visualization

DOE Award #: DE-SC0021555

Automated Sample Alignment for Neutron Scattering Using Image Segmentation Networks

Dr. Matthew Kilpatrick

J.P. Edelen (P.I.), **M. Henderson**, I. Pogorelov (RadiaSoft)

S. Calder, R. Gregory, C. Hoffmann, B. Krishna, G. Taufer, B. Vacaliuc (ORNL)

kilpatrick@radiasoft.net

Site Visit: Oak Ridge National Labs

October 19-21, 2022

radiasoft

Boulder, Colorado USA | radiasoft.net

U.S. DEPARTMENT OF ENERGY | Office of Science

Introduction

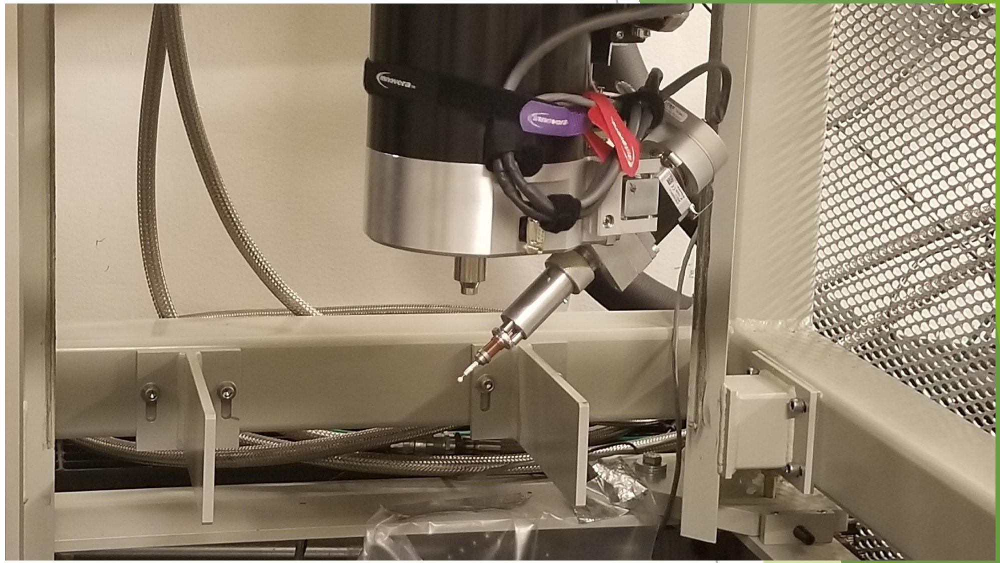
Background & Motivation

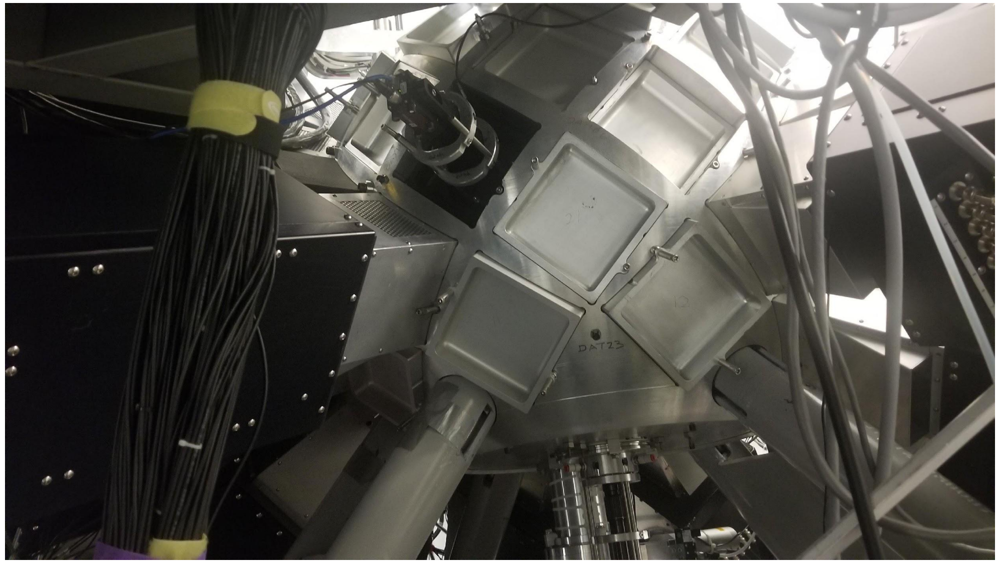
- Sample alignment, tedious but critical
  - Neutron production time is limited
  - Some activities require constant realignment
  - User facilities especially face schedule constraints
- Machine learning (ML) is a key automation tool
  - Current alignment tasks require human image recognition
  - Convolutional neural networks (CNN) for computer vision
- Alignment protocols vary between beamlines
  - Opportunity to employ & test transfer learning

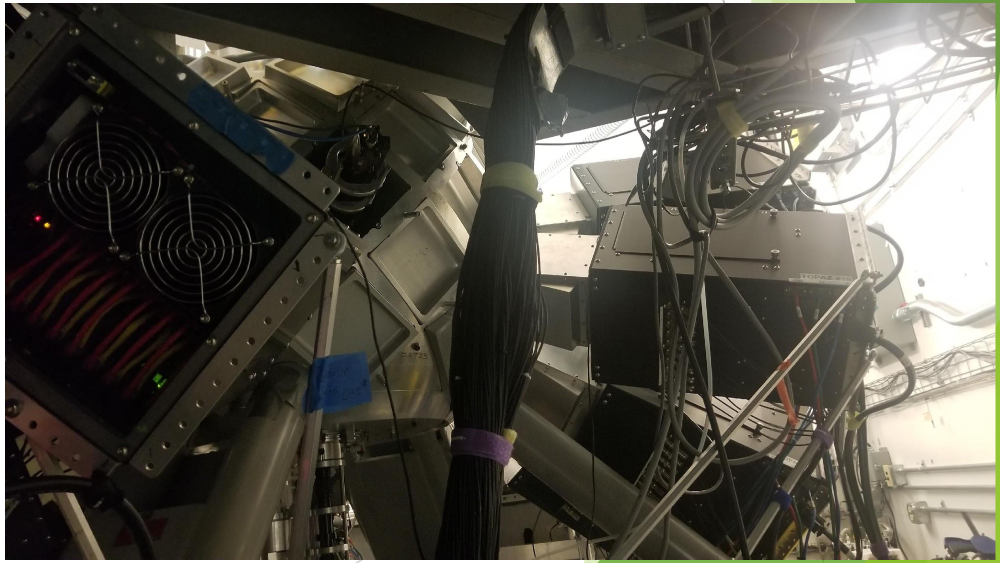
The Beamlines

- TOPAZ
  - Spallation Neutron Source (SNS)
  - Collision chamber with neutron sensors, photon camera, & env. Controls
  - Point-and-click control automation already in place
- HB2A Beamline
  - High-Flux Isotope Reactor (HFIR)
  - Sample container surrounded by neutron sensors & a **neutron** camera
    - Can only retrieve test alignment images during production
  - Requires direct motor controls

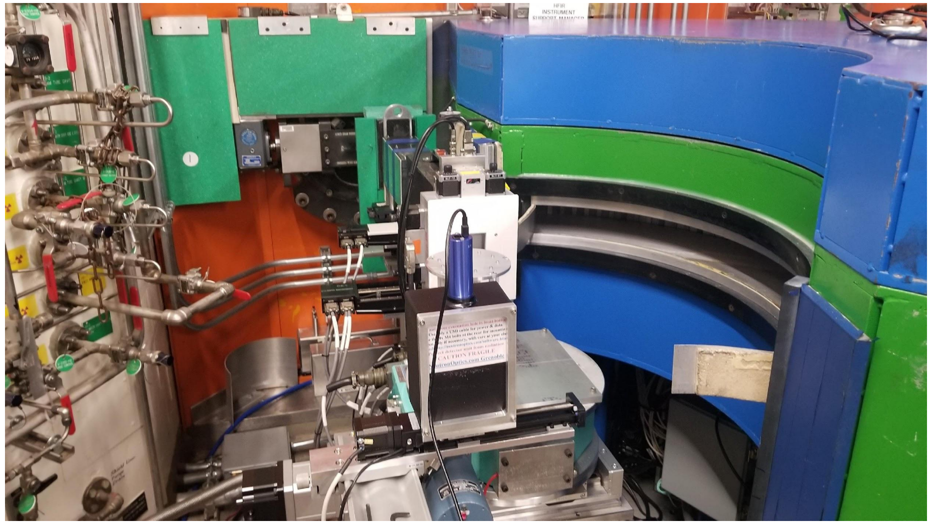
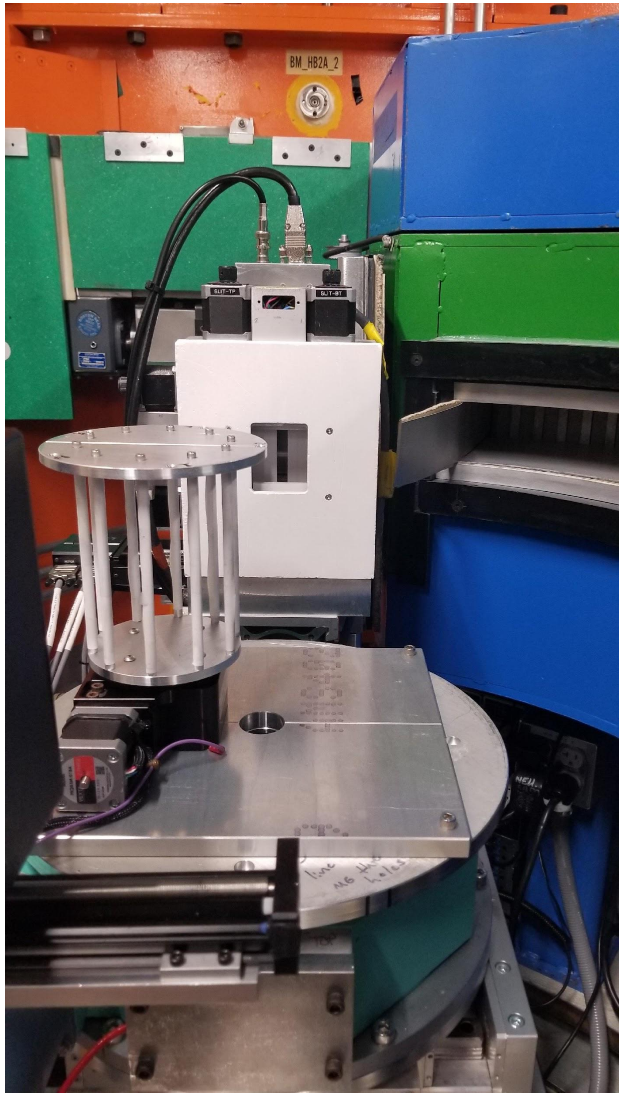
TOPAZ







HB2A


Problem Identification

Computer Vision Problem

- Recognize a sample in an image
  - Define sample location with a 2D mask
- Find the sample center of mass
  - Convert 2D masks into pixel coordinates
- Extend to new beamlines with minimal retraining
  - Design modular interface for machine controls
- Quantify ML prediction quality
  - Provide feedback to users
  - Alert in cases of need for human intervention
    - First need to *identify* these cases

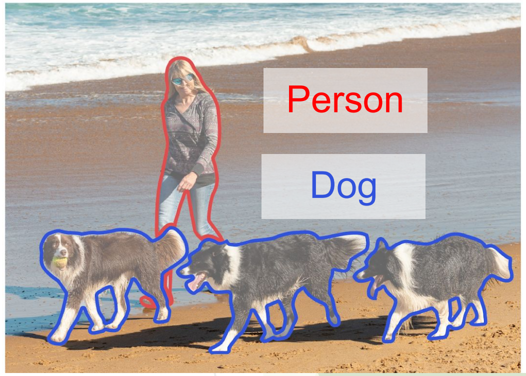
Types of Computer Vision Tasks

Image Classification

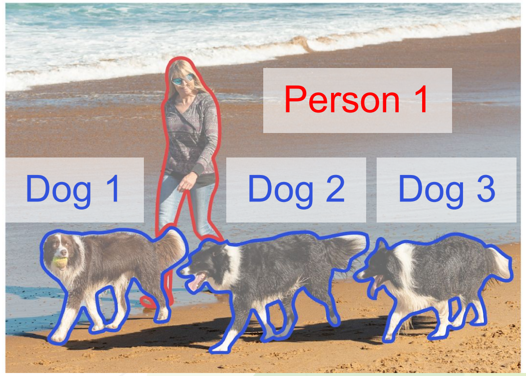
Classification with Localization

Object Detection

nuline



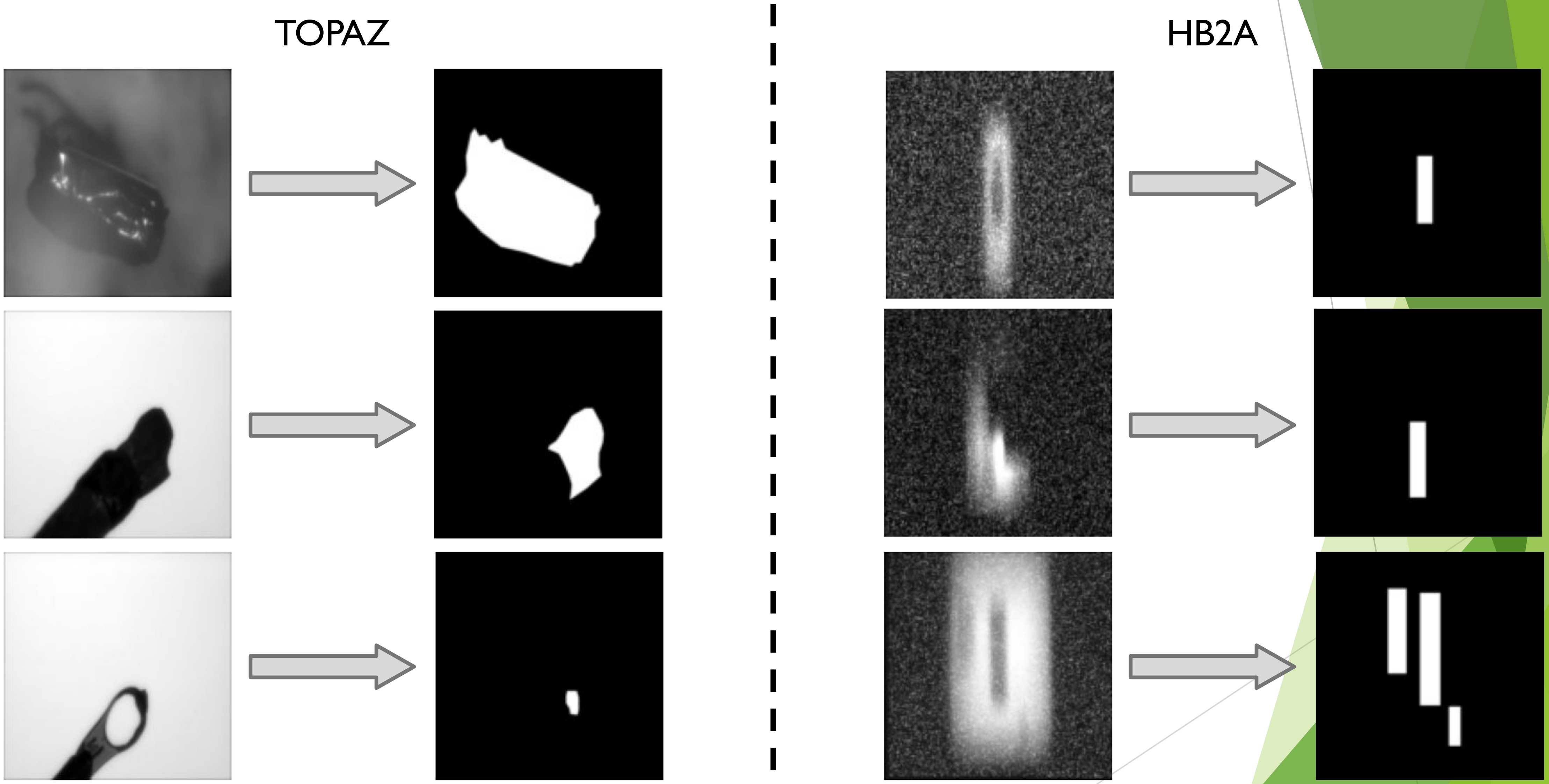
Semantic Segmentation



Instance Segmentation

[Images courtesy of Qualcomm Developer Network]()

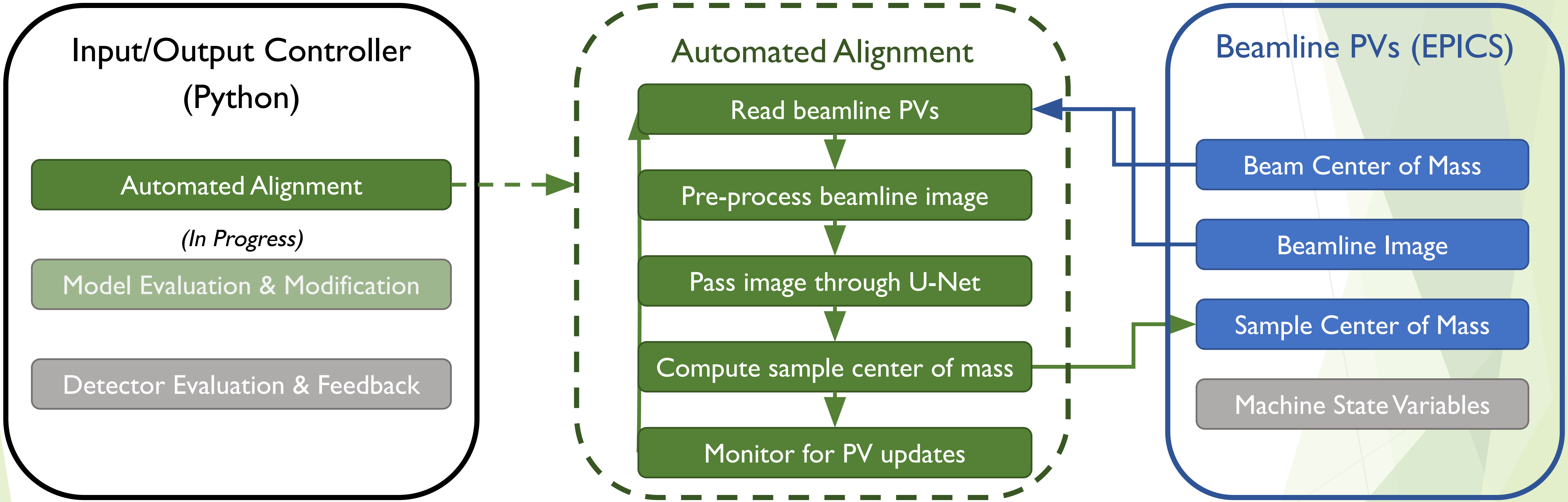
Sample Image Segmentation



Controls Problem

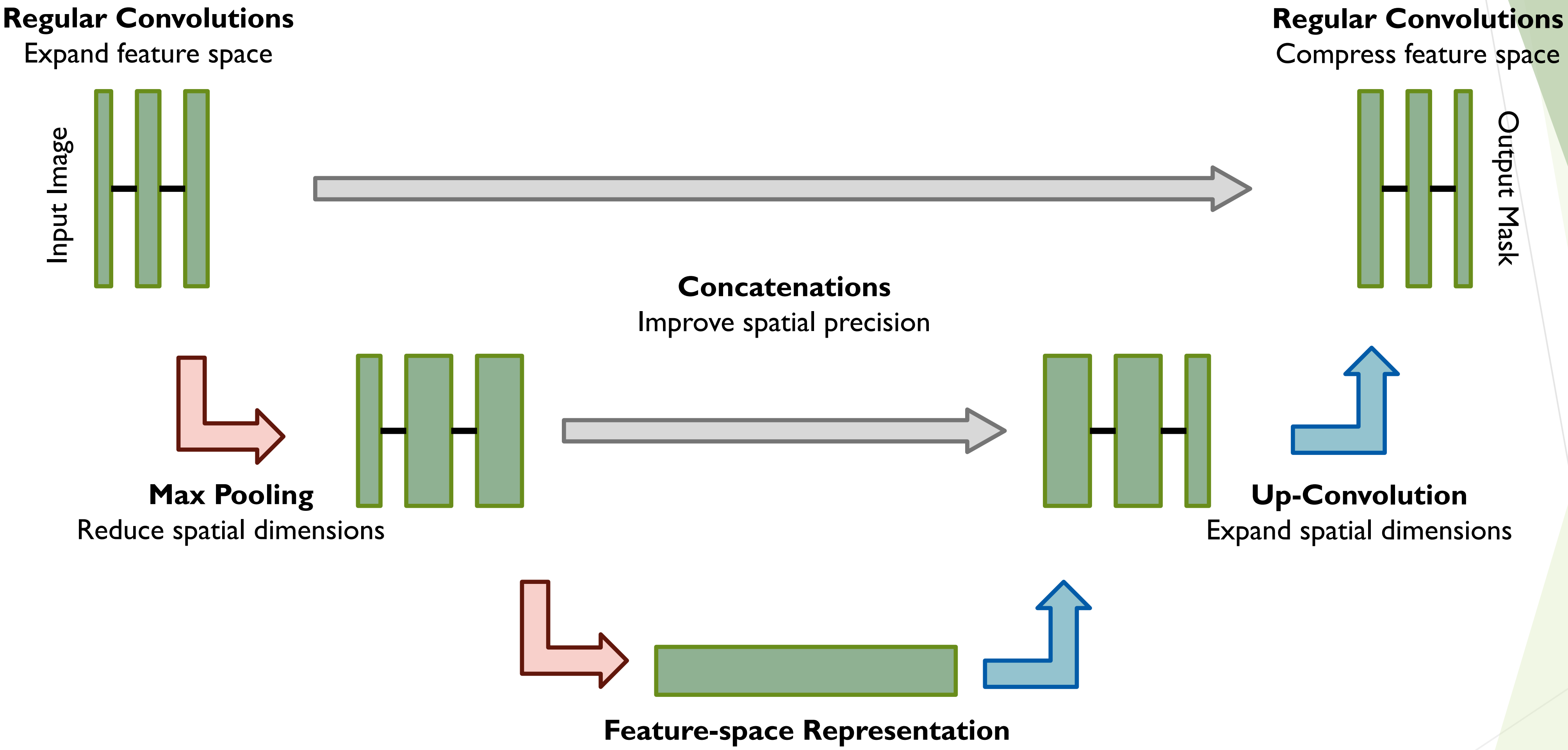
- Take image inputs from beamline
- Pass images through ML model
  - Retrieve predicted mask & center of mass
  - Retrieve uncertainty measurement
- Evaluate quality of predictions/state of controls
- Pass values back to beamline controls
  - Sample/beam center of mass offsets
  - HB2A: motor control information
  - Human intervention needed?
- Retrain networks with feedback from cycle

Controls Problem Block Diagram



Solution Methods

UNet Architecture



Contraction Path →

Extract features, lose spatial information

Expansion Path →

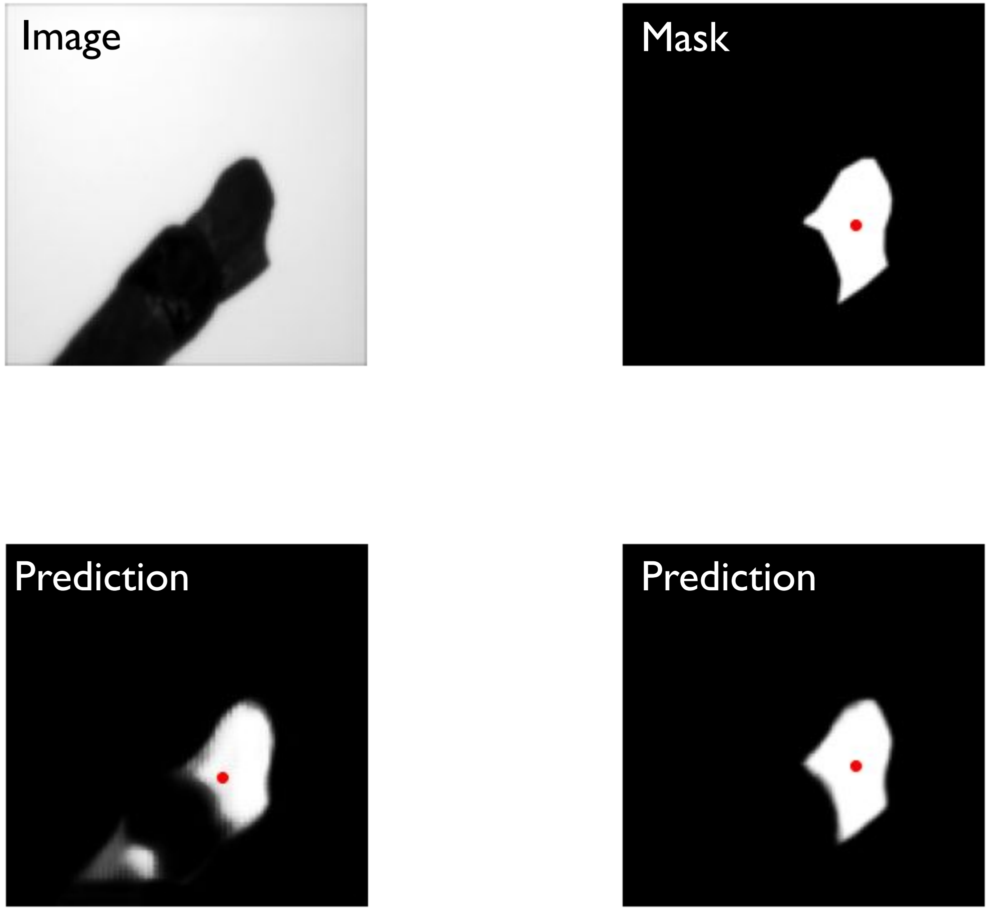
Expand features, regain spatial information

UNet Image Segmentation

- Begin with rough architecture (sub-optimal)
  - Trained on Topaz or HB2A data
- Optimize network parameters on beamline data
  - 2 optimization conditions: Topaz/HB2A data
- Conduct fresh training
  - 2 training conditions: Topaz/HB2A data (regardless of optimization)
- Perform extended OR transfer training
  - 2 training conditions: Topaz/HB2A data (regardless of optimization & initial training)
- Train 20 randomly initialized models for each setup
  - Ensemble learning: provides statistics for predictions/errors

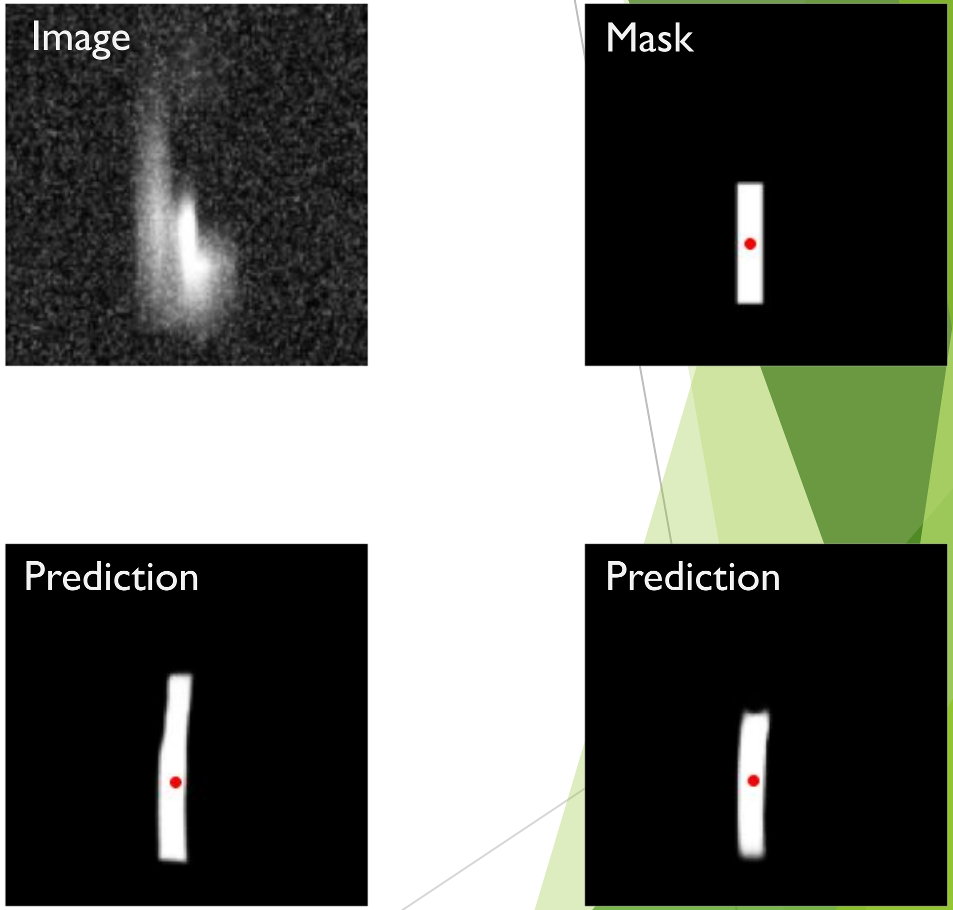
UNet Predictions

TOPAZ



(Suboptimal Model) (Optimized Model)

HB2A



(Suboptimal Model) (Optimized Model)

Transfer Learning

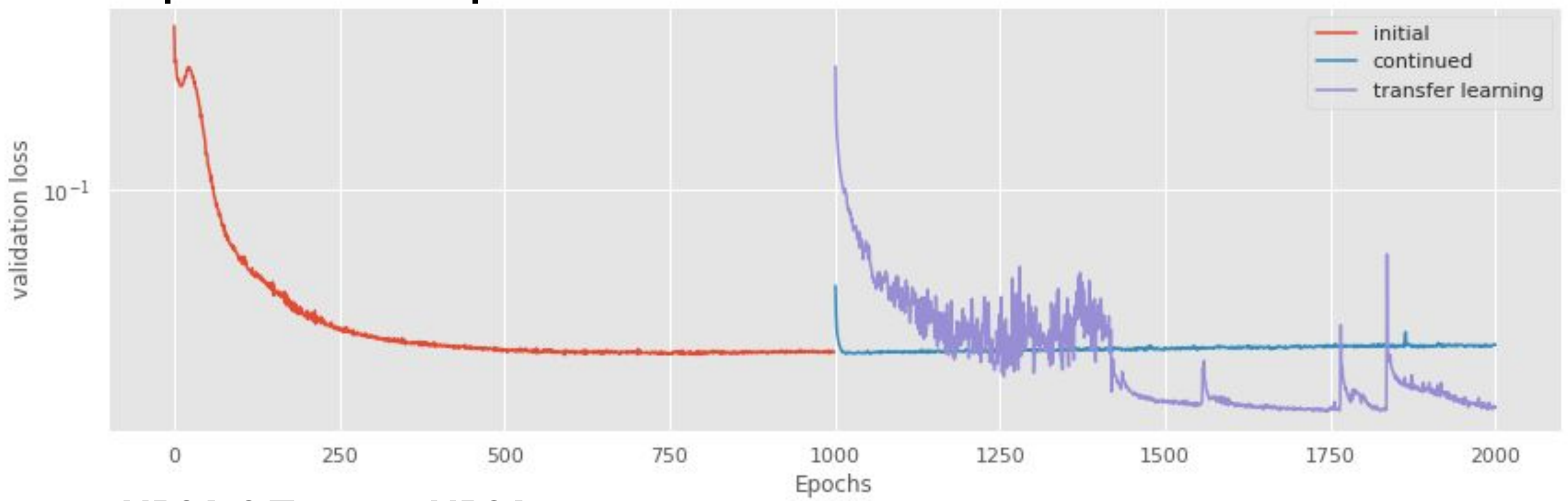
- Retrain existing networks with data from another source
  - Test extensibility of models to new beamlines
  - Improve predictions for original beamline data
- Conduct transfer learning under each unique setup (4 total)
  - TOPAZ to HB2A transfers (2 models, optimized for TOPAZ/HB2A)
  - HB2A to TOPAZ transfers (2 models, optimized for HB2A/TOPAZ)
- Quantify & compare:
  - Errors during learning
  - Uncertainties during testing

Transfer Learning Losses

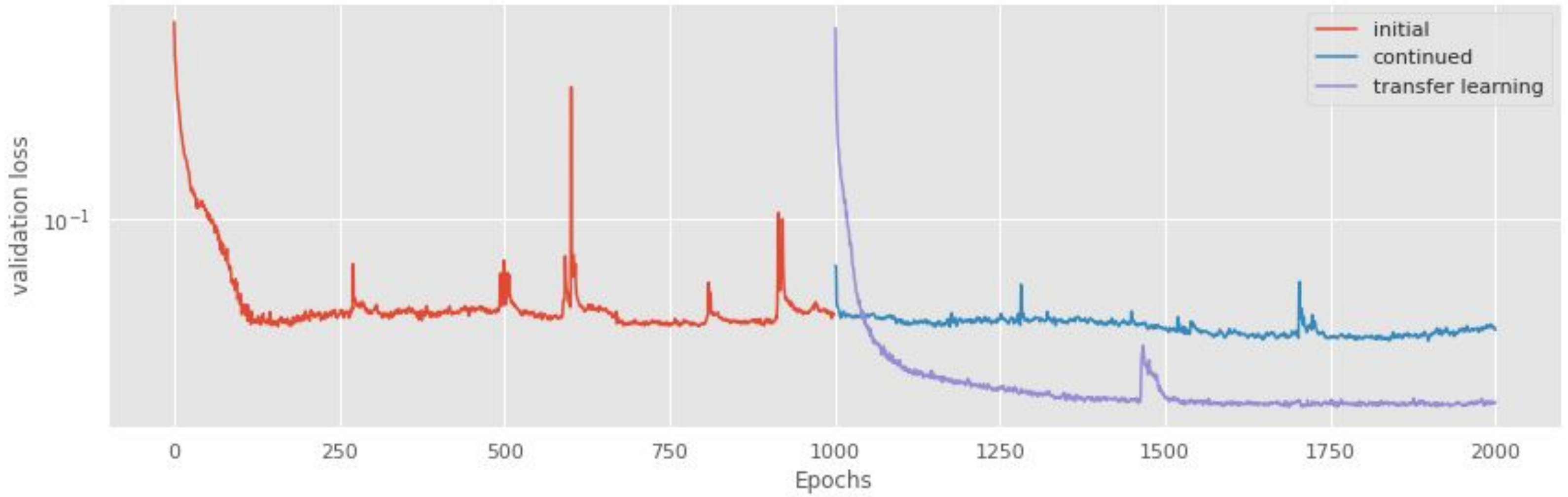
Takeaways

- Initial training
  - Converged to low loss (pixels)
- Extended training
  - Loss pattern generally continues
- Transfer learning
  - Lower loss is achieved

Topaz & HB2A to Topaz

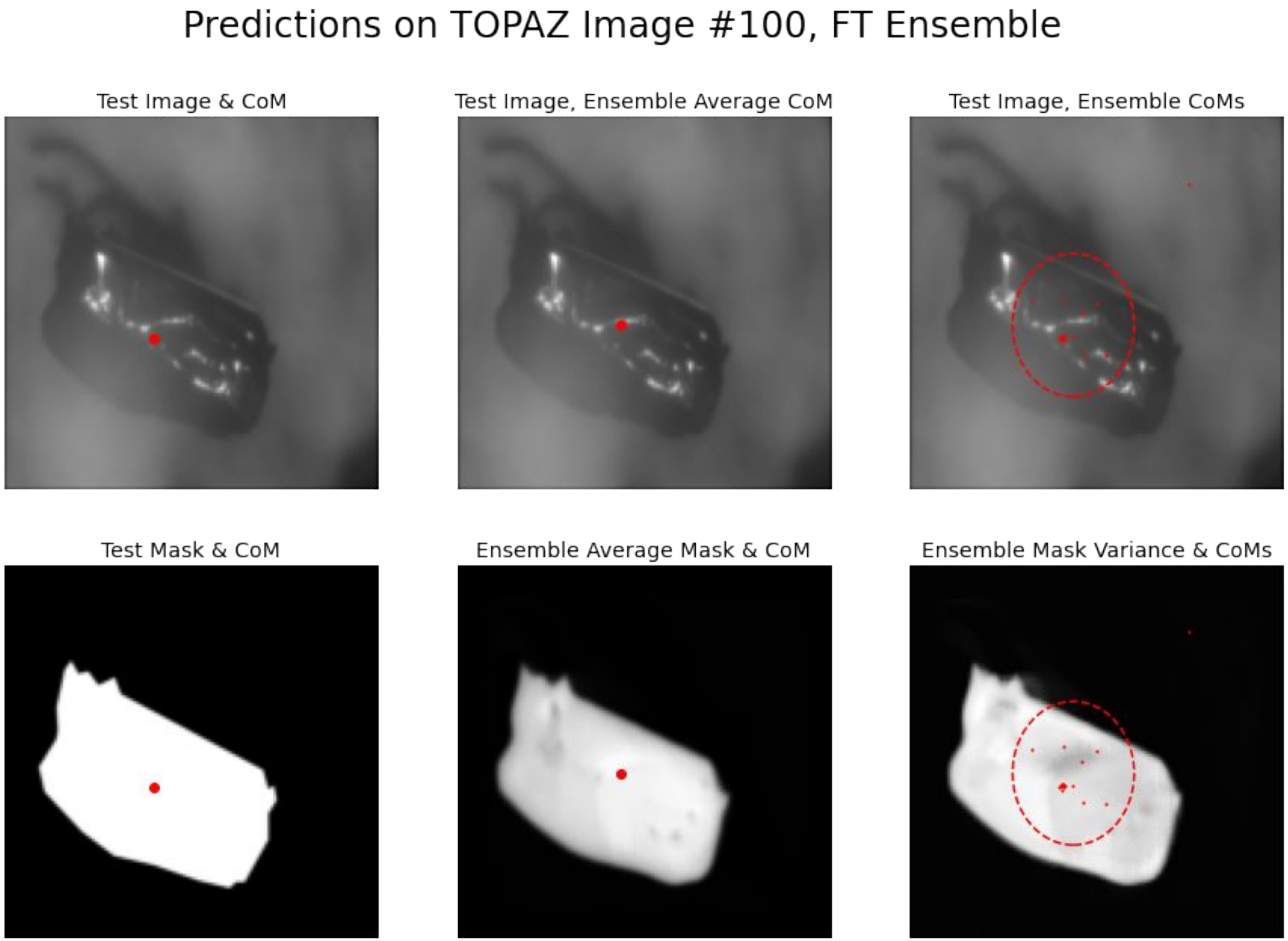


HB2A & Topaz to HB2A



Uncertainty Quantification

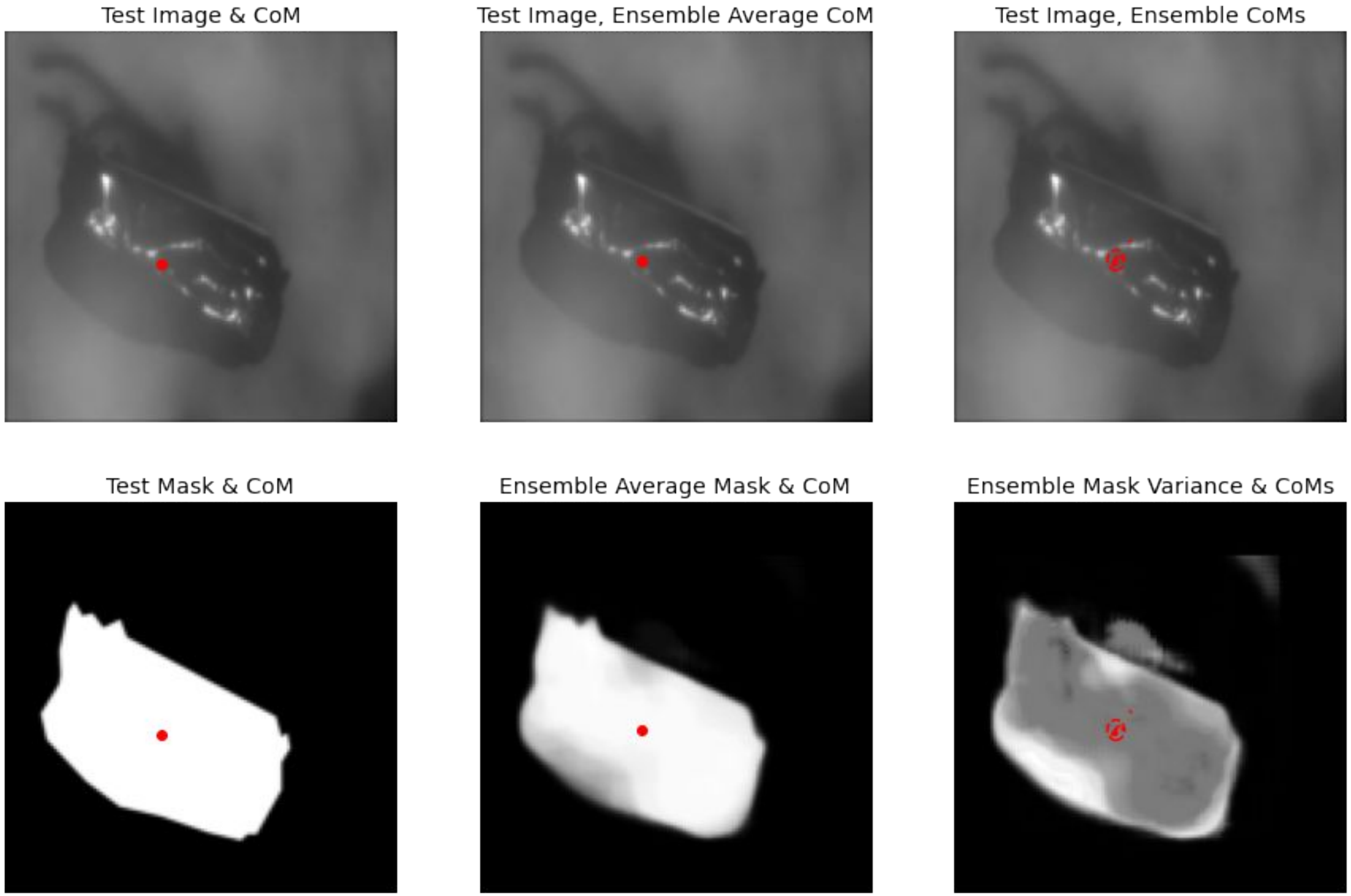
- During supervised training
  - Real error compared to human-defined masks
- During testing & operations
  - Ground-truth data ("correct" masks) not available
  - Employ statistics from ensemble predictions
    - Variance (disagreements) between many trained models
- Compare ensemble vs. ground-truth uncertainties
  - Ideally, both normally/χ^2 distributed (error/variance)
  - Do squared errors & ensemble variances share similar distributions?



Takeaways

- Sub-optimal model
- Suitable CoM prediction
- Large ensemble spread
  - High uncertainty
- Mask variance
  - High over entire sample

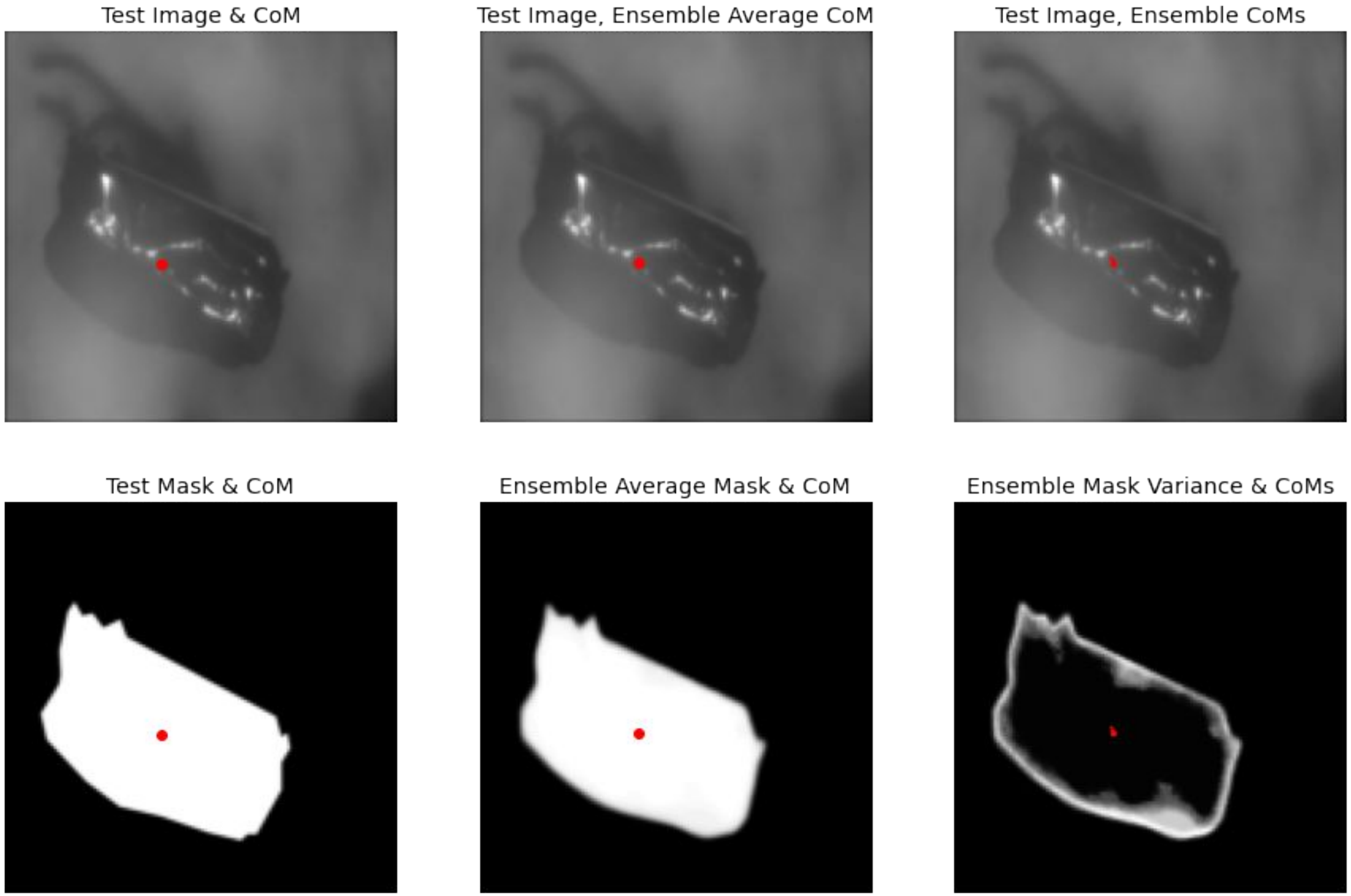
Predictions on TOPAZ Image #100, TT Ensemble



Takeaways

- Optimized model
- Excellent CoM prediction
- Small ensemble spread
  - Low uncertainty
- Mask variance
  - Lower over sample
  - Highest near edges
  - Background artefacts

Predictions on TOPAZ Image #100, HHT Ensemble



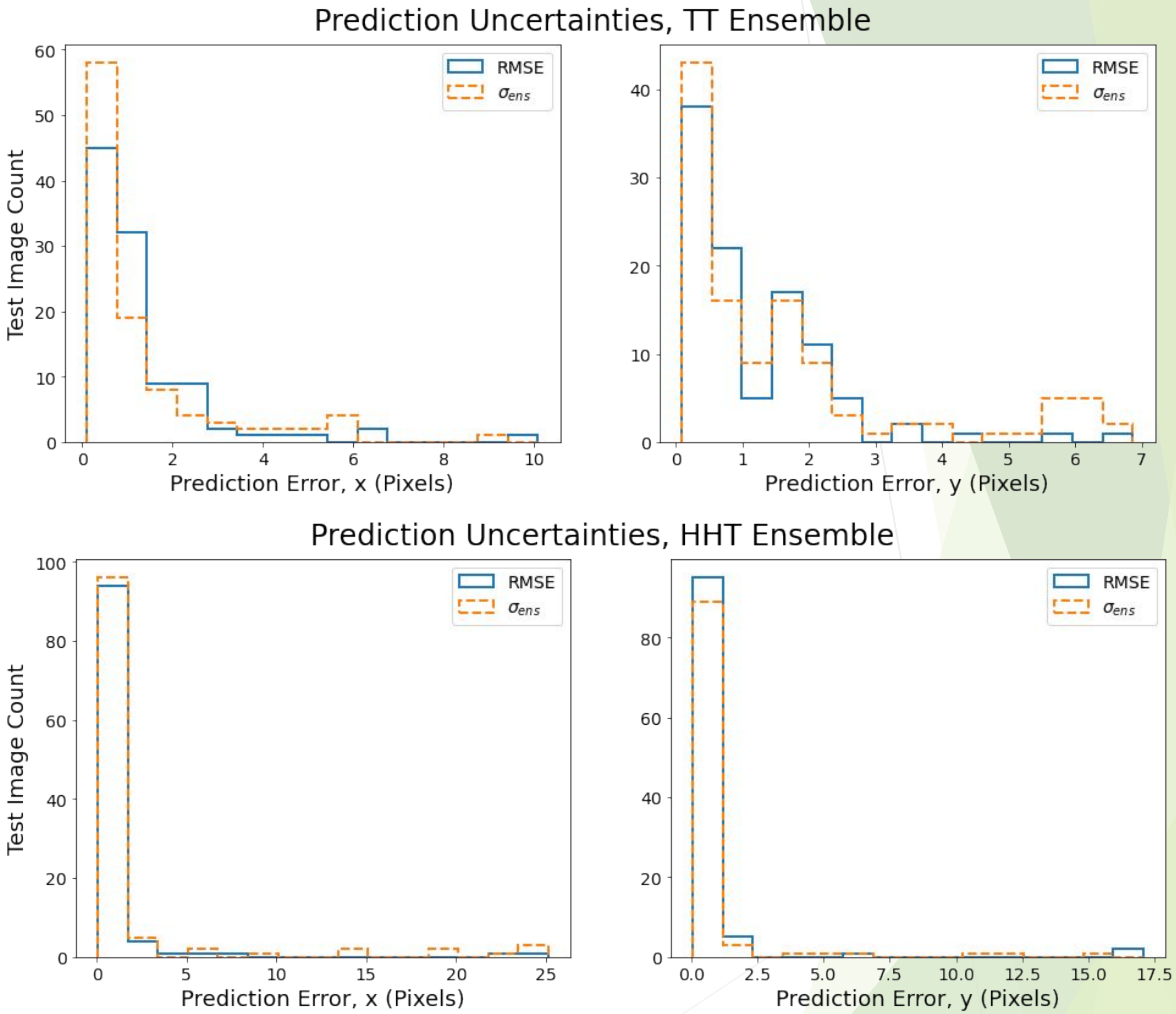
Takeaways

- Transferred model
- Excellent CoM prediction
- Negligible ensemble spread
  - Very low uncertainty
- Mask variance
  - Very low over sample
  - Restricted to edges

Error Distributions

Takeaways

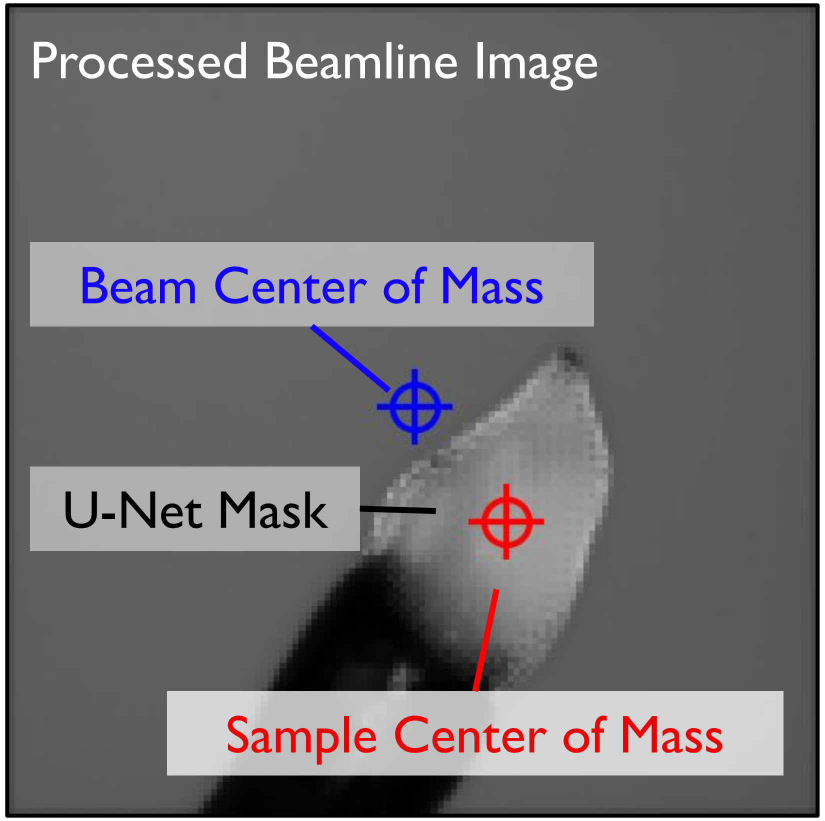
- True & ensemble distributions match!
  - Regardless of overall uncertainty
- Average uncertainties are "low"
  - ~1% of image dimensions
- Edge cases are identifiable as outliers
  - e.g., dark images, no sample, blurred

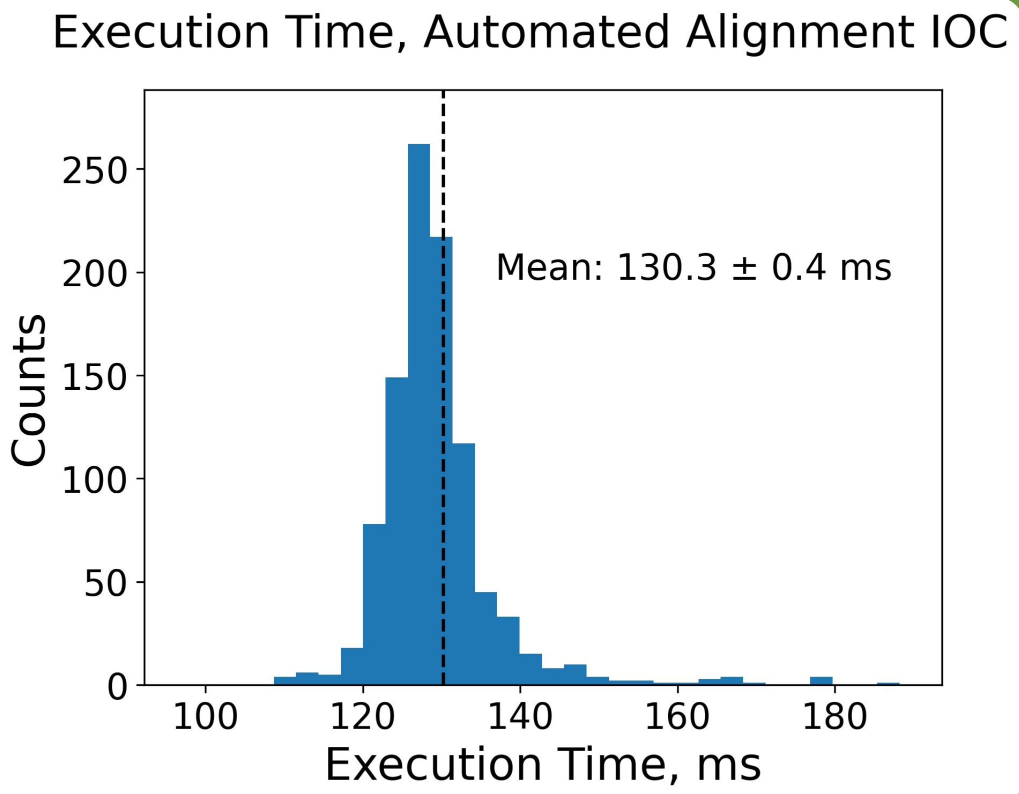


Controls Interface

- Create a mock beamline program
  - Supplies historical beamline data as EPICS variables (PVs)
- Build interface using EPICS libraries in Python
  - Establishes full set of PVs for access (set by user, highly modular)
    - Reads beamline PVs (image & beam CoM, etc.)
    - Sets output PVs (sample CoM, motor positions, etc.)
  - Loads a stored UNet model (efficient HDF5 files)
- Run Python interface during mock beamline operation
  - Read simulated beamline inputs to controller
  - Controller executes ML predictions on inputs
  - Controller passes outputs back to beamline PVs (& timing/plots for testing)

Control Interface Output





Ongoing Work

- Merge interface with motor controls
  - TOPAZ: integrate with existing software (ORNL)
  - HB2A: implement direct motor controls functions
- Integrate detector information
  - TOPAZ: improve overall intensity, possibly automate calibration
  - HB2A: identify correct sample, choose correct detector, etc.
    - Possible application for *instance* segmentation
- Conduct live testing of controls interface
  - Normal operations (both)
  - TOPAZ: temperature ramping sequence, calibration

Machine Learning for Automated Analysis of 3D Neutron Scattering Data

Dr. Matthew Kilpatrick

D. Bruhwiler (PI), K. Bruhwiler, E. Carlin,

R. Gregory, C. Hoffman, J. Kohl, Z. Morgan, A. Savici

kilpatrick@radiasoft.net

October 21, 2022

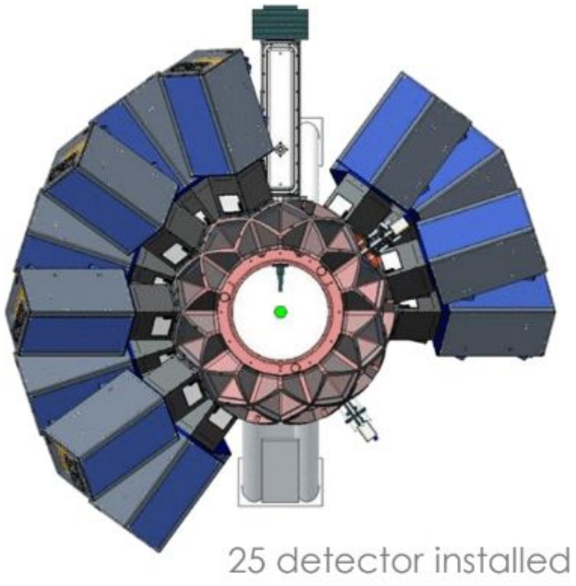
Oak Ridge National Laboratory Seminar

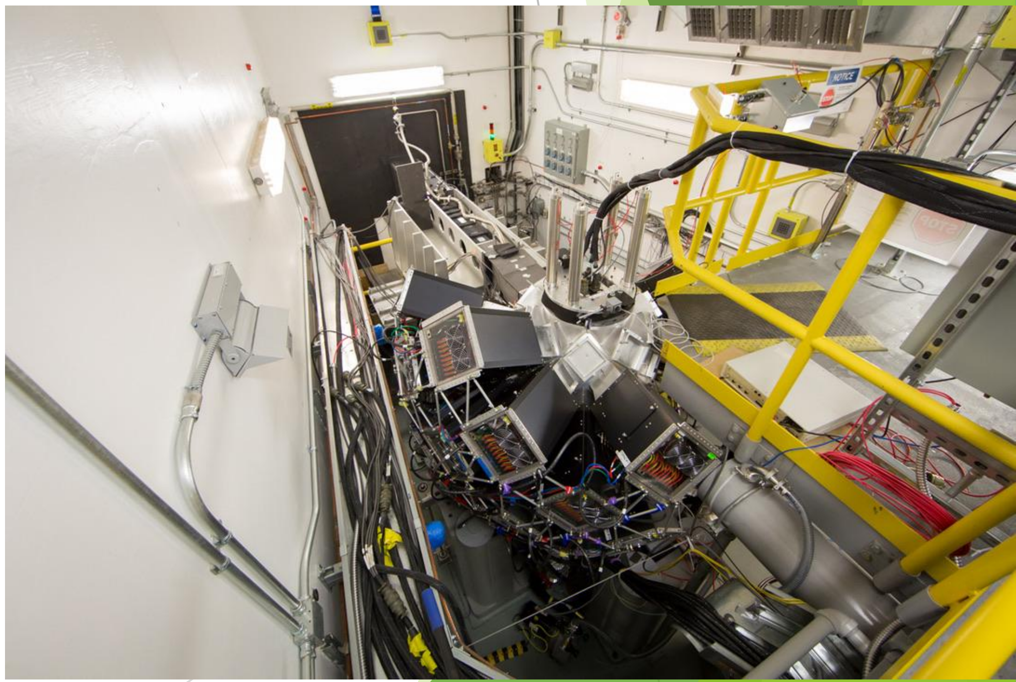
radiasoft

Boulder, Colorado USA | radiasoft.net

TOPAZ Single-crystal Diffractometer

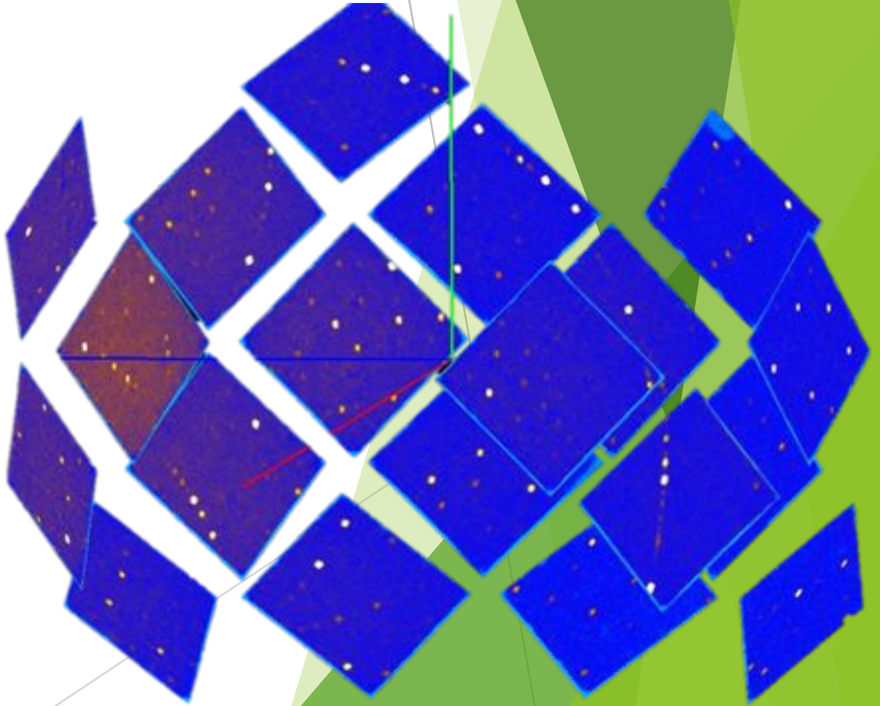
- Located at ORNL and receives neutrons from the Spallation Neutron Source
- Samples can be measured with high precision for volumetric sampling in reciprocal space (momentum measurements)
  - Samples are rotated to measure all aspects of the lattice
  - Temperature control from 5 K - 450 K
  - Broad Q coverage





Current Analysis Methods

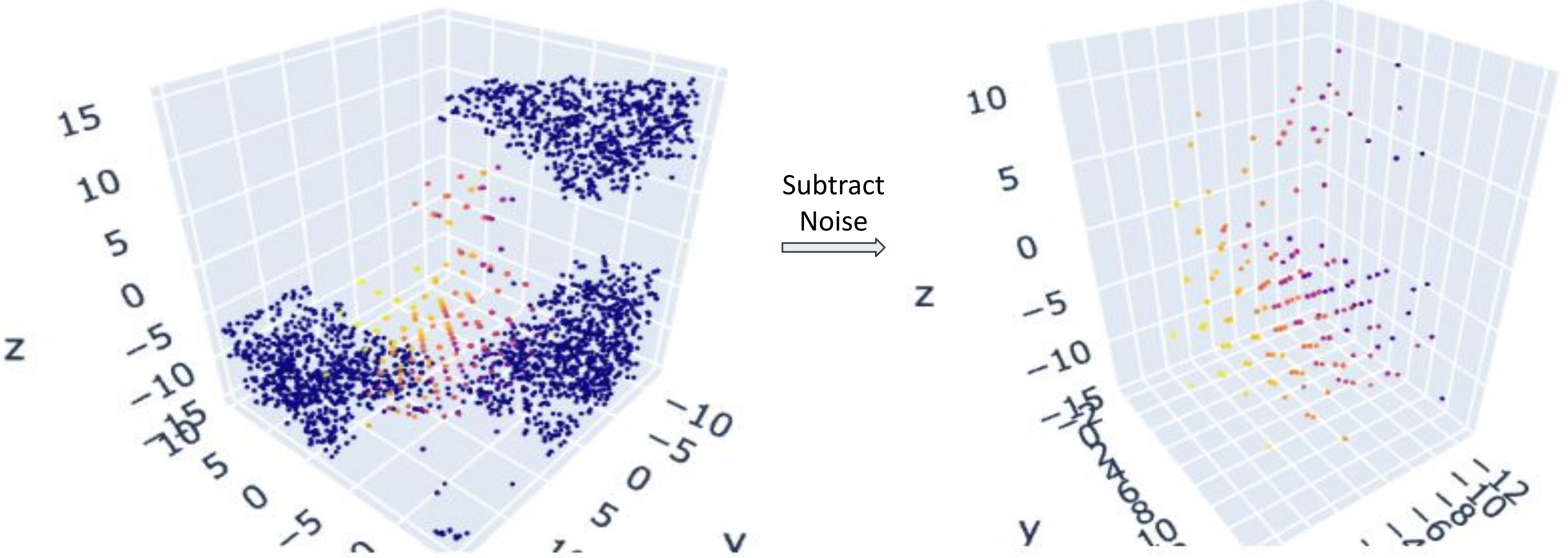
- Mantid
  - Open source community developed application
  - Algorithms use raw data processed within 2D slices of the 3D data
- Difficulties
  - Large datasets can be up to 100 Gb in size and current tools limit interactivity
  - Displaying data requires a high level of user interaction
  - 2D slices may miss key features of the diffraction data
  - May be slow to run due to optimization shortfalls
- How can we improve this?
  - Machine learning
    - Reduces user interaction
    - Automated Bragg peak identification
- Can we improve the interactivity of the data visualization?
  - Collaboration with NVIDIA Index to allow for real time visualization
  - Interactive 3D methods to analyze measurements



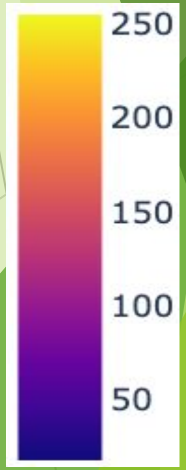
https://www.mantidproject.org/

Analysis of Scattering Data for a Single Silicon Crystal

- **Preprocessing the data**
  - Measurements of the particles in momentum space
    - Looking for pockets of constructive interference
  - Measurements projected onto a 3D mesh
- **DBSCAN is a density based clustering algorithm**
  - Can identify oblong clusters and attribute data to noise if it doesn't match anything
  - Sparse background can be removed easily



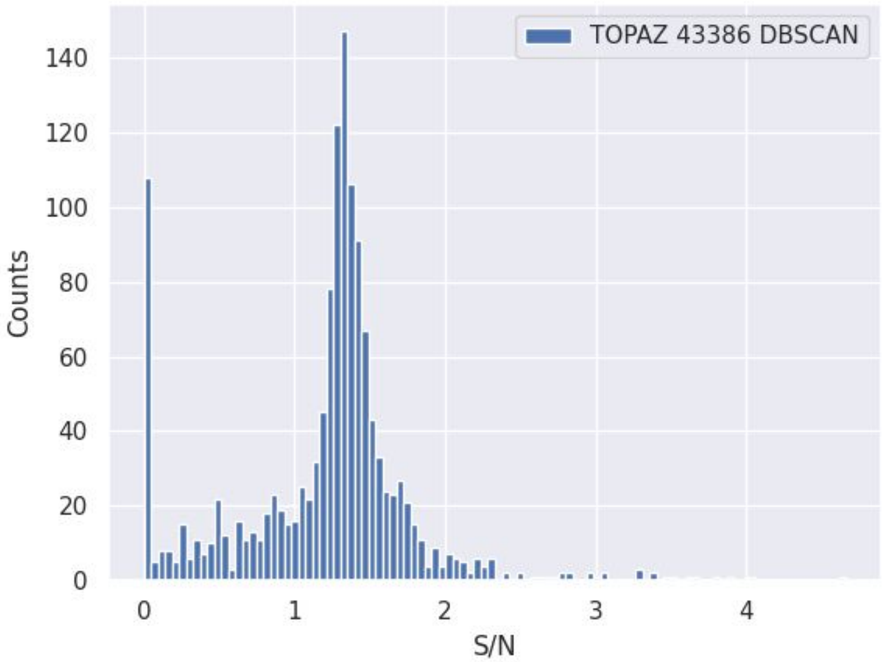
Silicon
Cluster ID

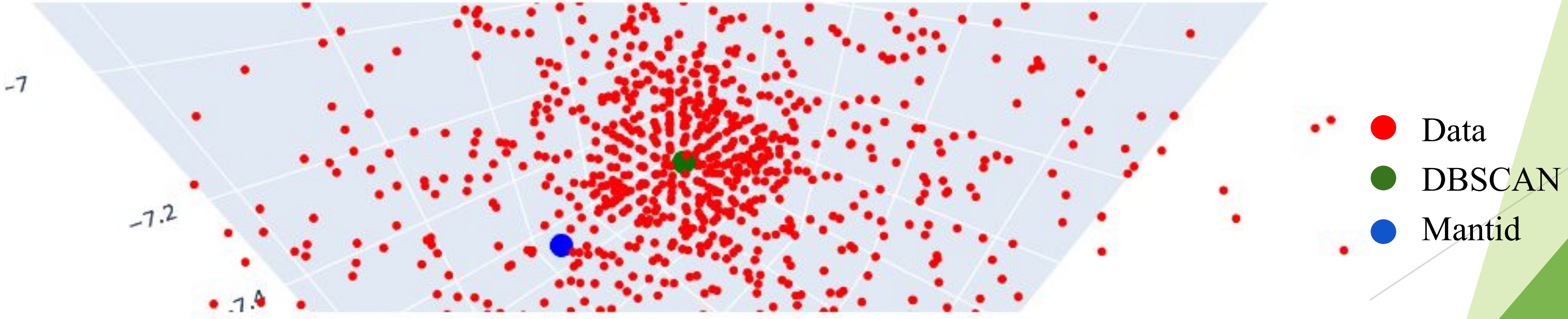


How to determine a good Bragg peak?

- Calculate a Signal-to-Noise for each matched peak
  - A Bragg peak should have a large Signal-to-Noise ratio
  - Background events result in a minimal ratio
  - Tunable cut to allow for optimized data cleaning

$$\frac{S}{N} = \frac{I_{\delta S}}{\sqrt{I_{\delta B}^{2} + N_S}}$$





Comparison Cluster Identification with Mantid

- **Peak locations**
  - Fast Fourier Transform (FFT) to calculate UB matrix (orientation matrix) and predict Miller indices
- **Limitations of Mantid peak finding**
  - Needs to be told how many peaks to look for
  - Use the number of peaks that is found by DBSCAN
- **What do we compare?**
  - UB matrix is not unique
    - Due to the symmetry of the Bragg peaks you can rotate the matrix and still predict the peaks in the proper location
  - Miller indices are not unique
    - Due to the rotation of the UB matrix different hkl indices can map to the same peak
    - Confirm values are 'close' to an integer value
  - Compare peak locations
    - Clean with Signal-to-noise

$$\underbrace{Q_s}_{\text{Measured Peaks}} = 2\pi \cdot \underbrace{UB}_{\text{Calculated with FFT from Mantid}} \cdot \underbrace{\begin{pmatrix} h \\ k \\ l \end{pmatrix}}_{\text{Predicted Miller Indices}}$$

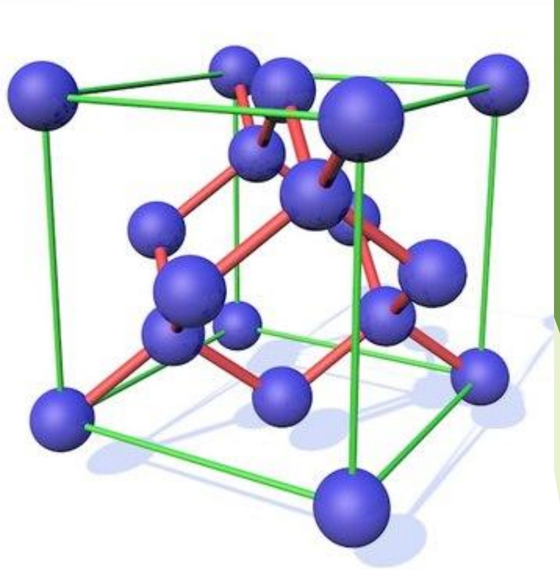
https://docs.mantidproject.org/v4.0.0/algorithms/FindUBUsingFFT-v1.html
https://docs.mantidproject.org/v4.0.0/concepts/Lattice.html#lattice

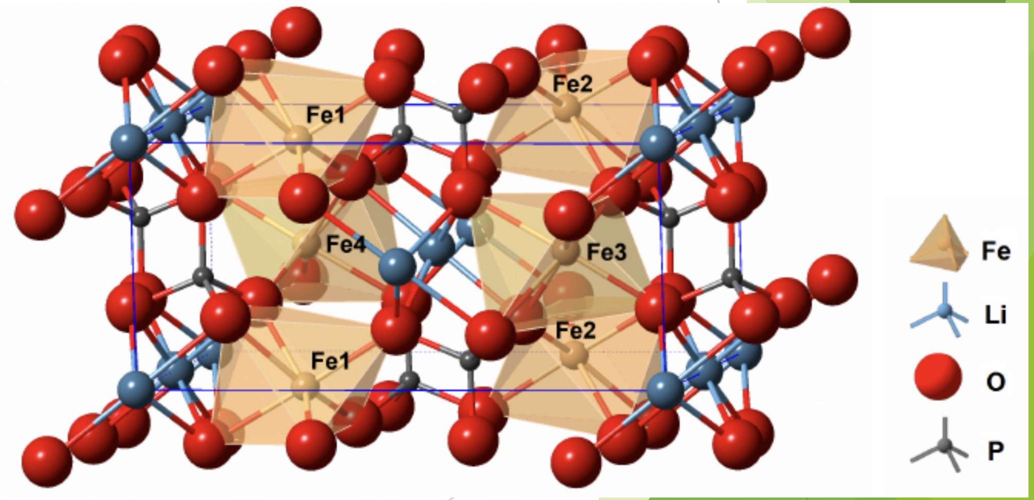
TOPAZ Data

- **Silicon**
  - Clean dataset for testing methods
  - print("Hello World!")
  - FCC unit cell

- **LiFeO (LiFePO4)**
  - More noisy background
  - Rectangular unit cell

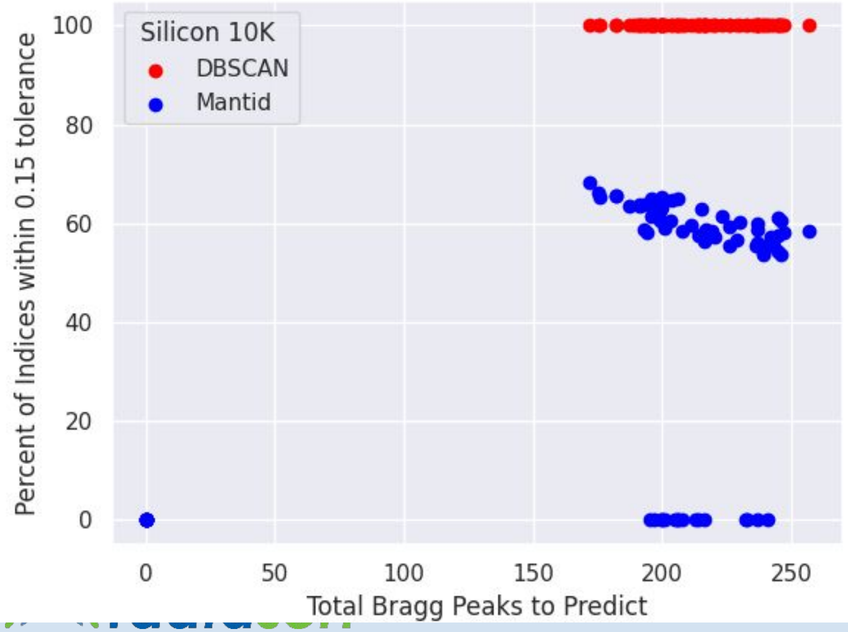
- **Benzil**
  - Much more difficult
  - Includes a diffuse background structure
  - Still being processed…

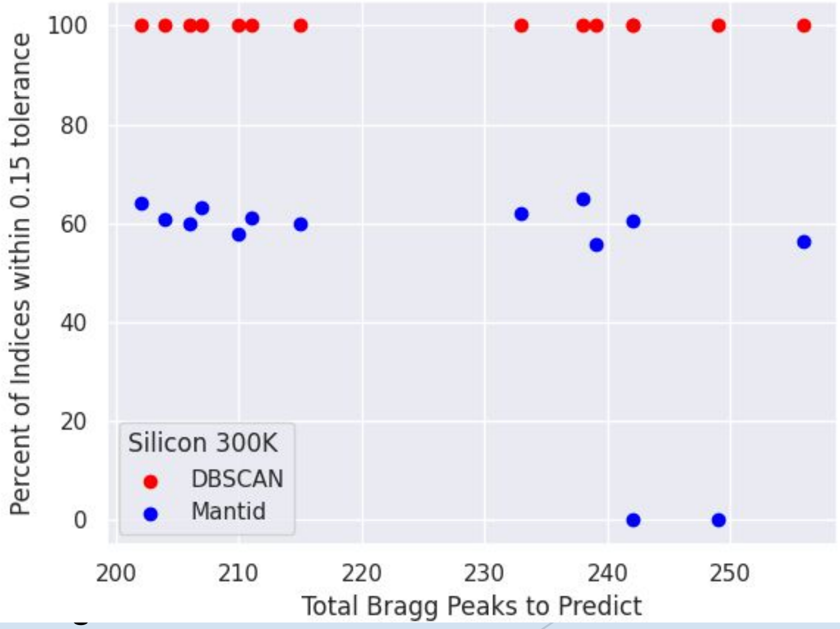




Silicon 10 K and 300 K

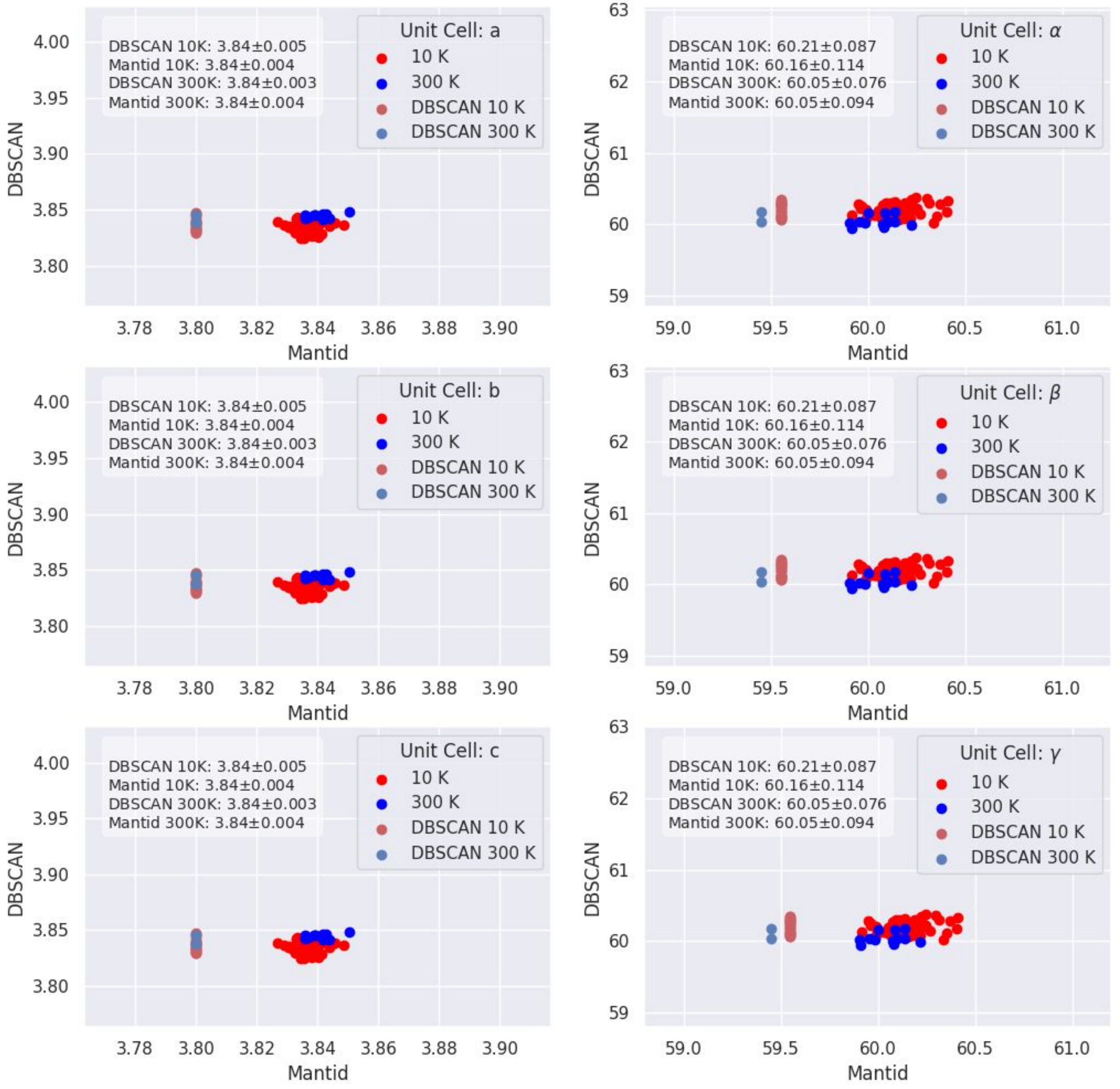
- Predict UB matrix with identified Bragg peaks
  - Silicon 10K
    - DBSCAN finds 100% of Bragg peaks within tolerance
    - Mantid finds 55-65% of Bragg peaks within tolerance
    - 24 3D silicon datasets could not find a UB matrix with Mantid peaks
  - Silicon 300K
    - DBSCAN finds 100% of Bragg peaks within tolerance
    - Mantid finds 55-65% of Bragg peaks within tolerance
    - 2 3D silicon datasets could not find a UB matrix with Mantid peaks





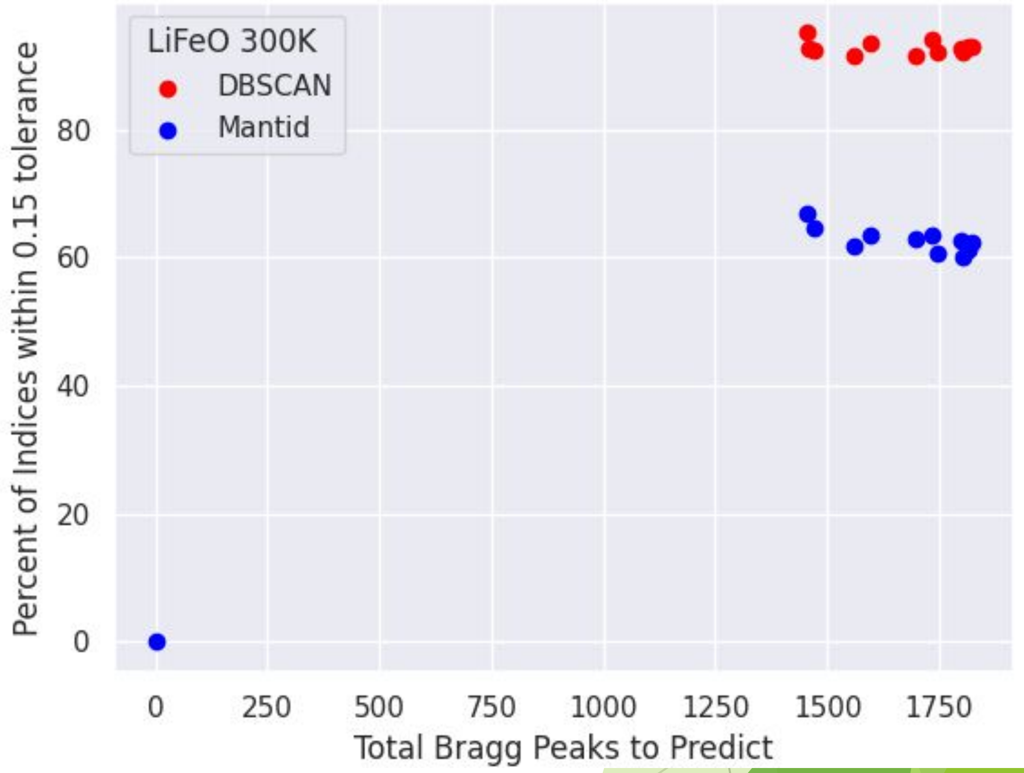
Silicon 10 K and 300 K Unit cells parameters

- There is agreement!
  - Mantid and DBSCAN methods show the same lattice parameters for Silicon
  - Standard deviation is comparable!
- DBSCAN datasets
  - Automated peak finding allows for similar results to Mantid only methods
  - Files with only DBSCAN results are shifted from the normal results
  - All lattice parameters agree with the theoretical value of Silicon



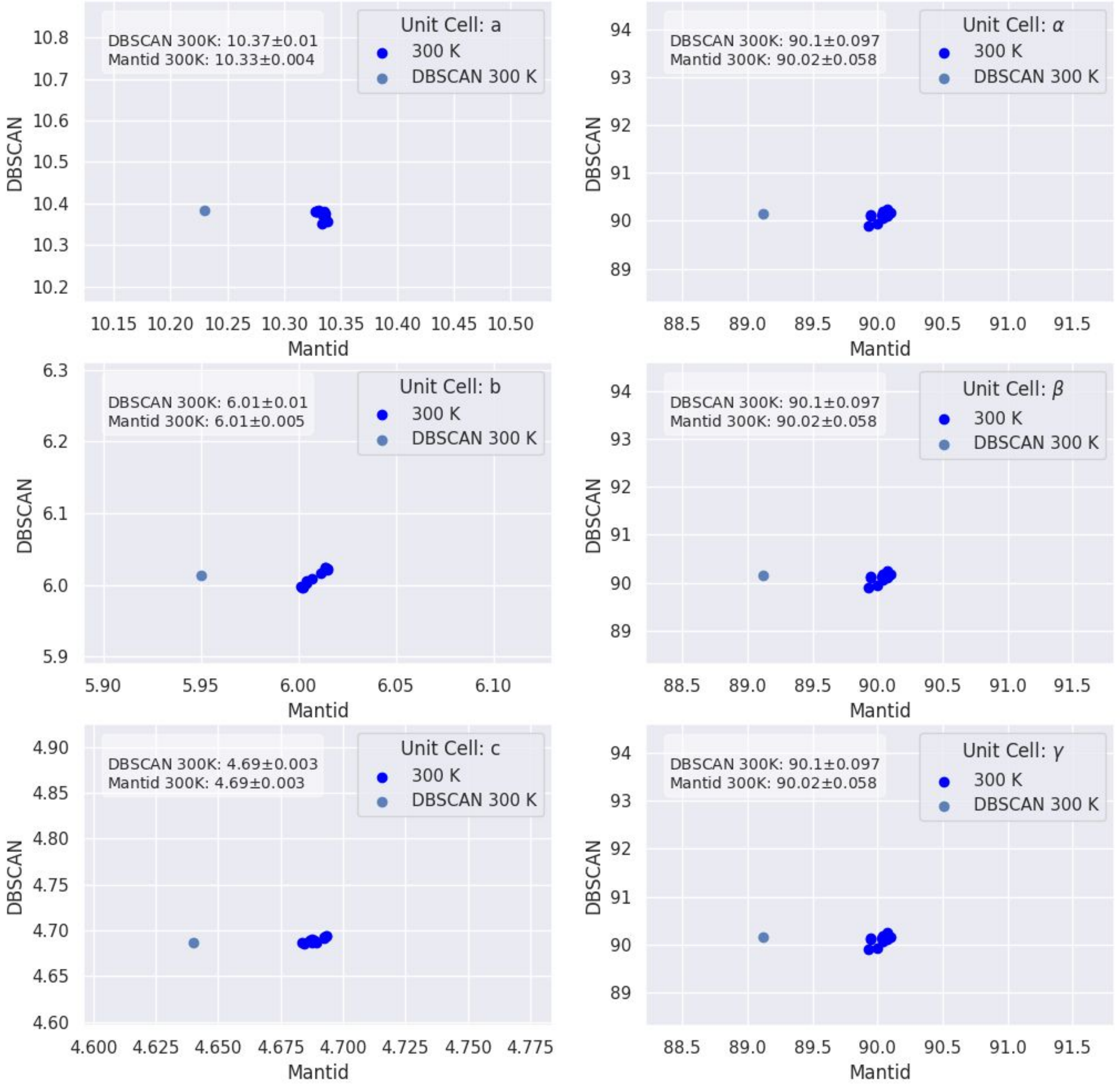
LiFeO at 300 K

- Predict UB matrix with identified Bragg peaks
  - DBSCAN finds ~85% of Bragg peaks within tolerance
  - Mantid finds ~65% of Bragg peaks within tolerance
  - Bottom left point is a point that Mantid couldn't find any Bragg peaks
    - DBSCAN found 1460 Bragg peaks



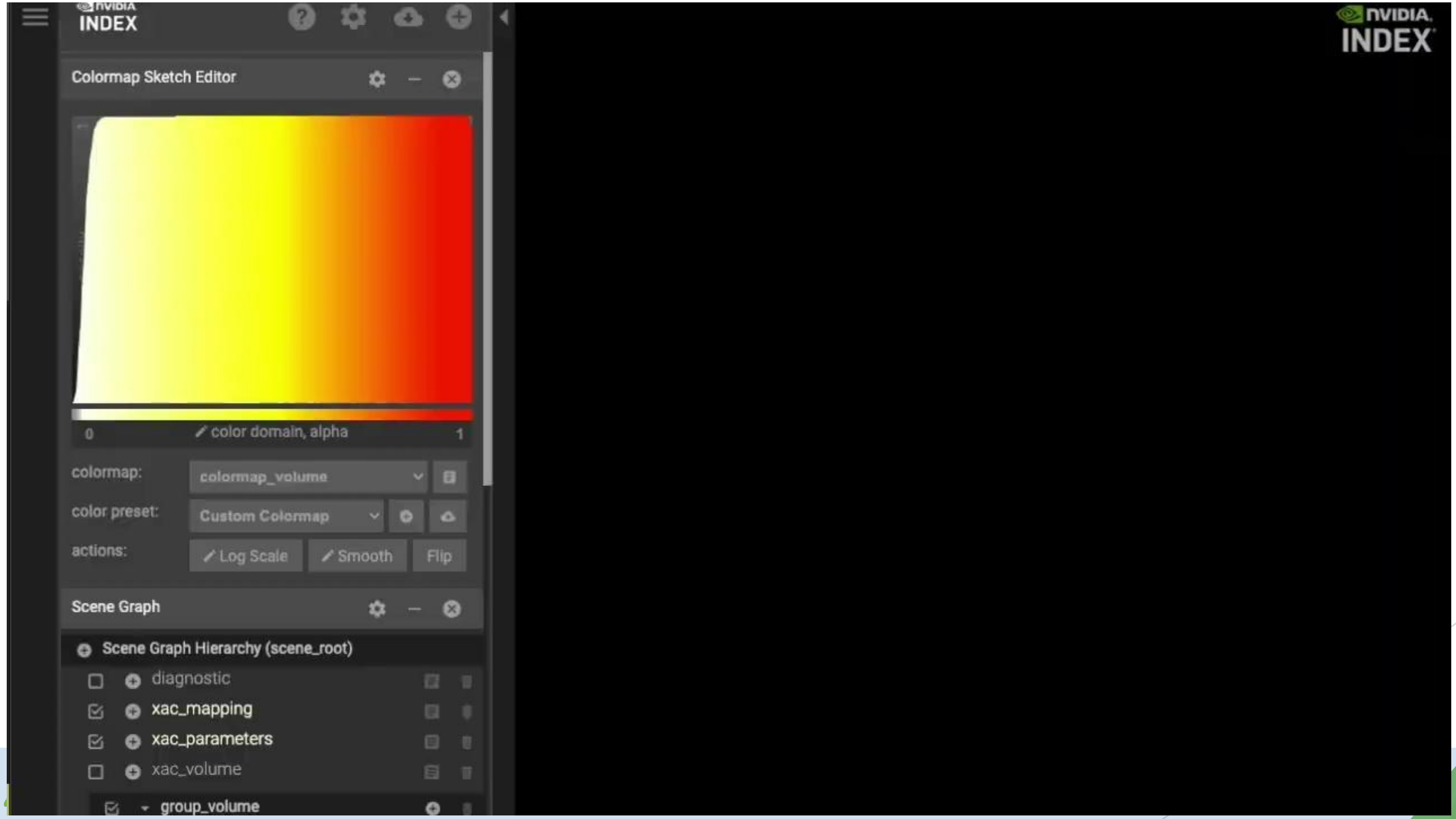
LiFeO at 300 K

- **Predict UB matrix with identified Bragg peaks**
  - DBSCAN finds ~85% of Bragg peaks within tolerance
  - Mantid finds ~65% of Bragg peaks within tolerance
  - Bottom left point is a point that Mantid couldn't find any Bragg peaks
    - DBSCAN found 1460 Bragg peaks
- **Lattice Parameters**
  - Theoretical: a = 10.334 Å, b = 6.008 Å, c = 4.693 Å
  - Both methods show great agreement!



ADARA Live Streaming into NVIDIA's IndeX

- Software developer: Evan Carlin
  - Adapted the ADARA live streaming
  - Live reading of data to allow for interactive real-time vizualization



Summary

- Advantages of DBSCAN:
  - Density Based clustering algorithm can identify Bragg peaks without user interaction
  - Robust methods
  - Allows for easy implementation into NVIDIA's IndeX
  - Noise Reduction
  - Real time unit cell calculation!
- Future work
  - Integration of identified peak in NVIDIA's Index visualization for interactive analysis
  - Full 3D streaming of data to see data in real time
  - Interactive projection of Miller indices
  - Robust methods for confidence level

THANK YOU!

Acknowledgements

- Our ORNL partners
  - Ray Gregory & Gary Taufer (EPICS handling of camera data & SPICE support)
- NVIDIA
  - Marc Nienhaus, Alexander Kuhn, Dragos Tatulea, Jörg Mensmann, Steffen Roemer,
- RadiaSoft
  - David Bruhwiler, Evan Carlin, Jon Edelen, Morgan Henderson
- DOE Office of Science, Basic Energy Sciences
  - SBIR and STTR Program, award # DE-SC0021555, DE-SC0021551