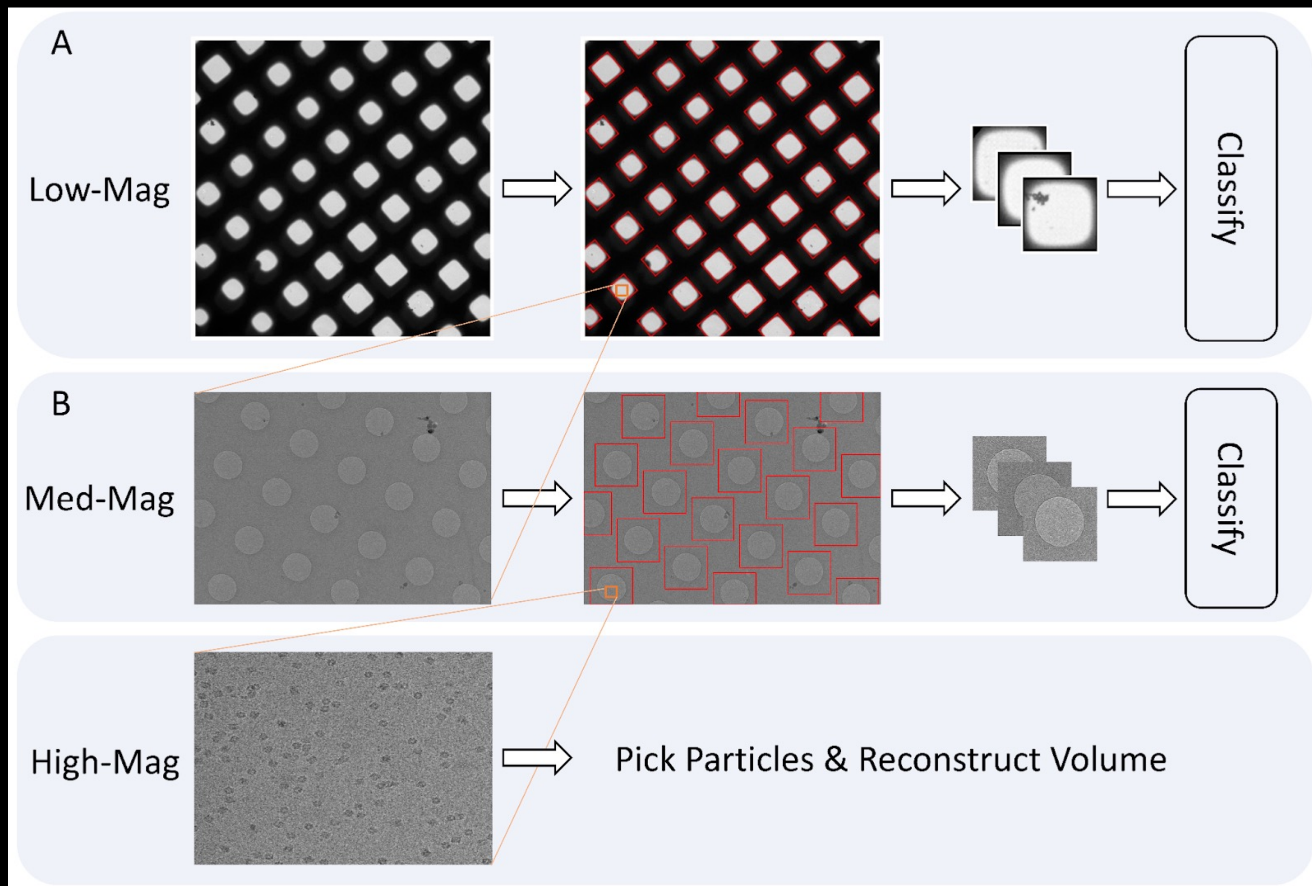
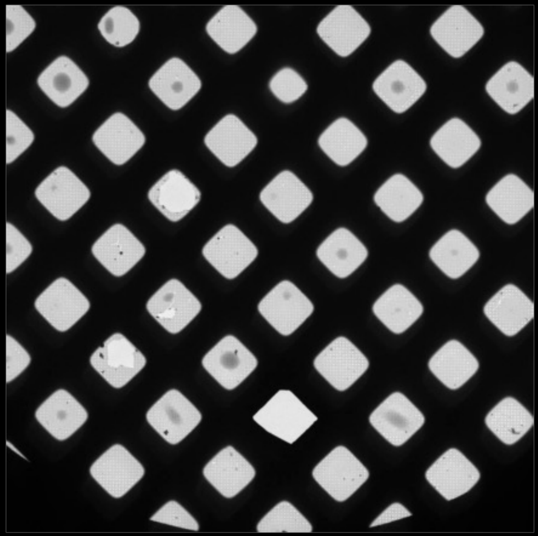
Learning to automate cryo-electron microscopy data collection with Ptolemy

Smart Data Collection Workshop April 2022

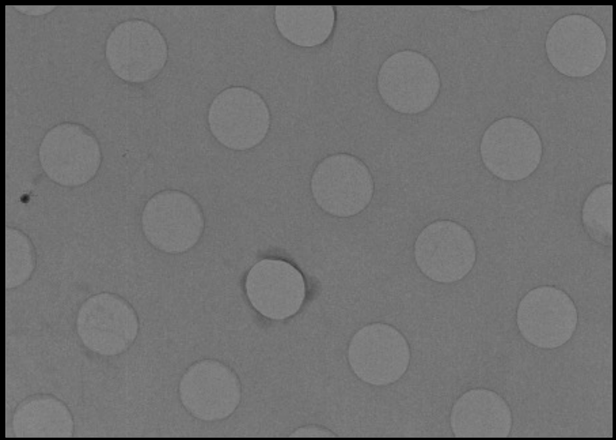

A
Low-Mag
Classify
B
Med-Mag
Classify
High-Mag
Pick Particles & Reconstruct Volume

Datasets

- 76 historical data collection sessions
- 1.3k grid tile images w/ square target coordinates
- 11k targeted squares
- 28k square tile images w/ hole target coordinates
- 410k targeted holes

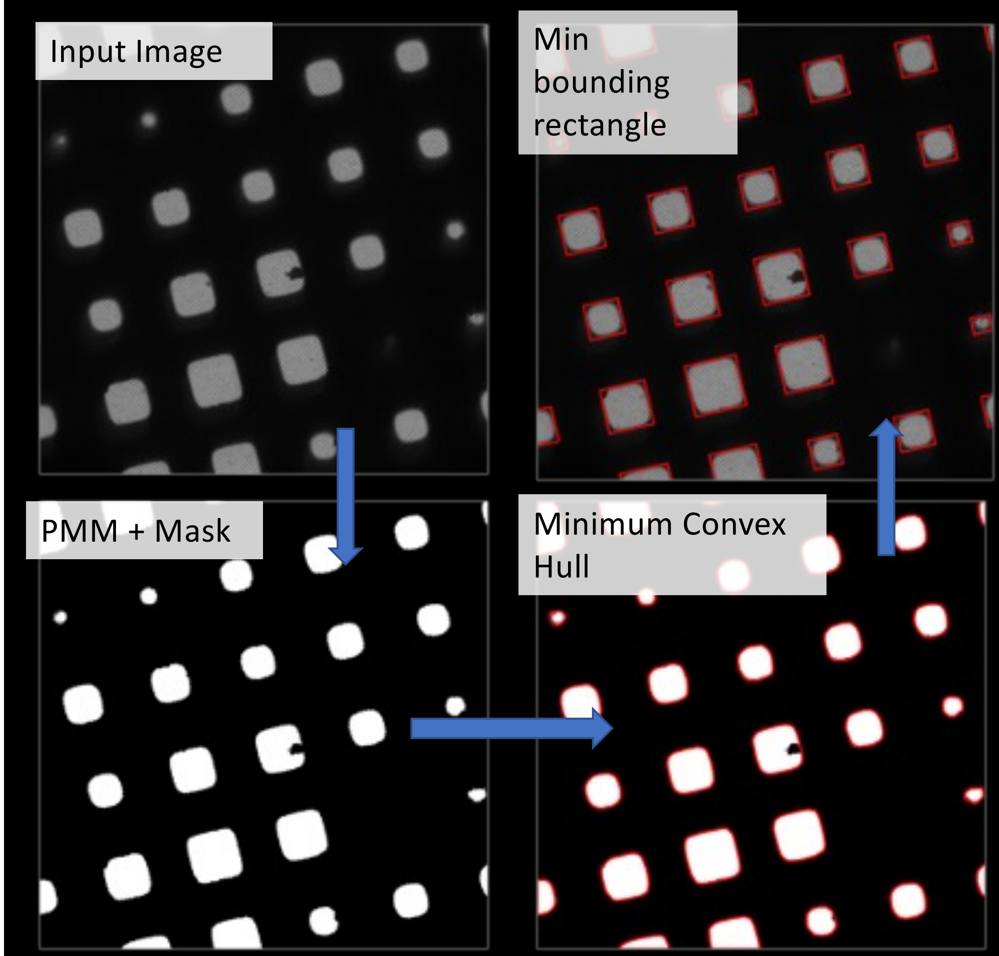


Grid Tile

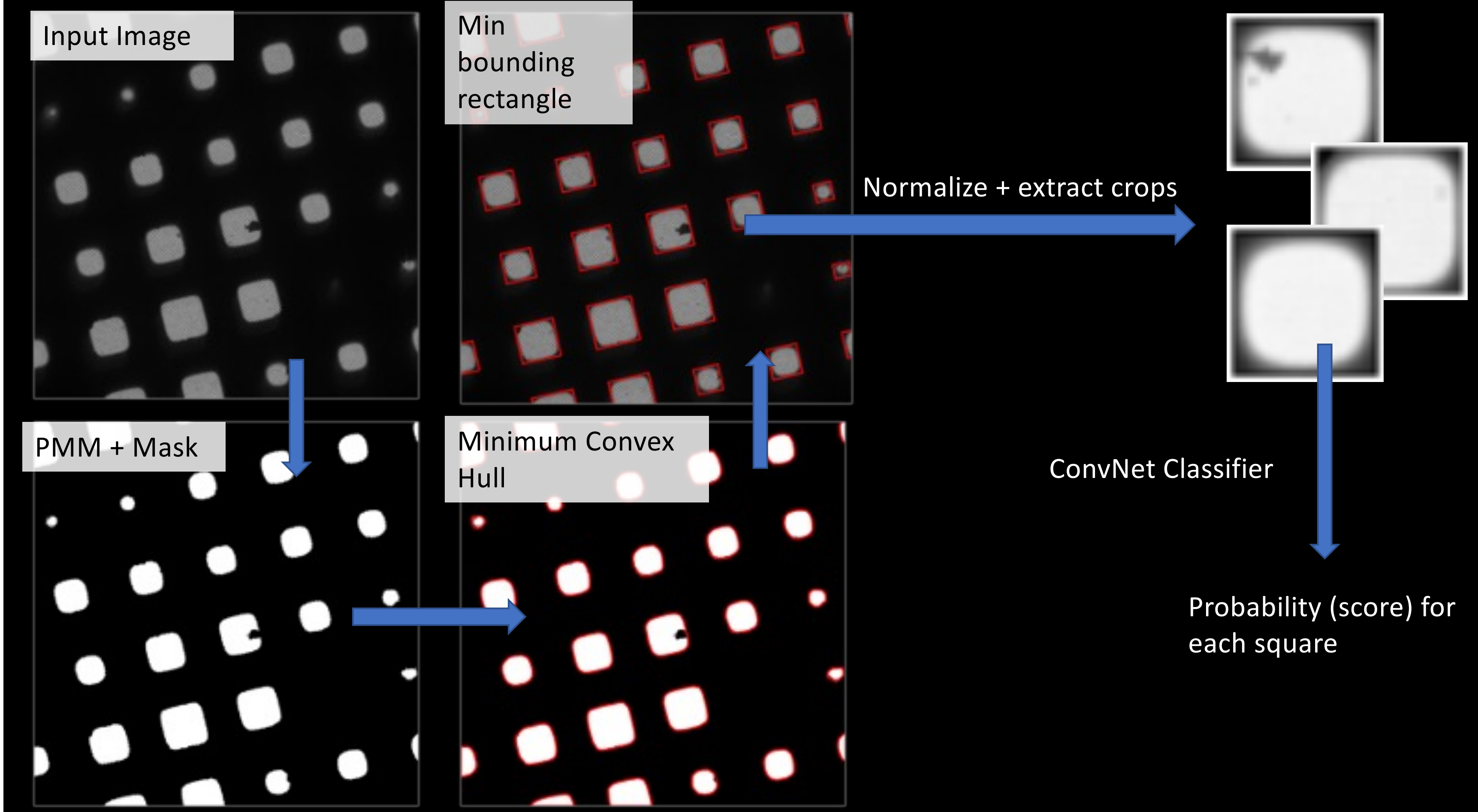


Square Tile

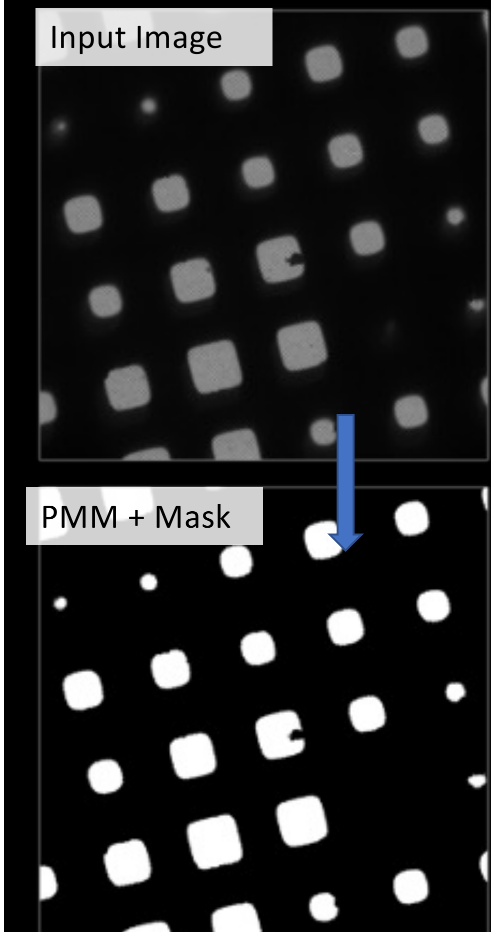
Low2Med: Workflow

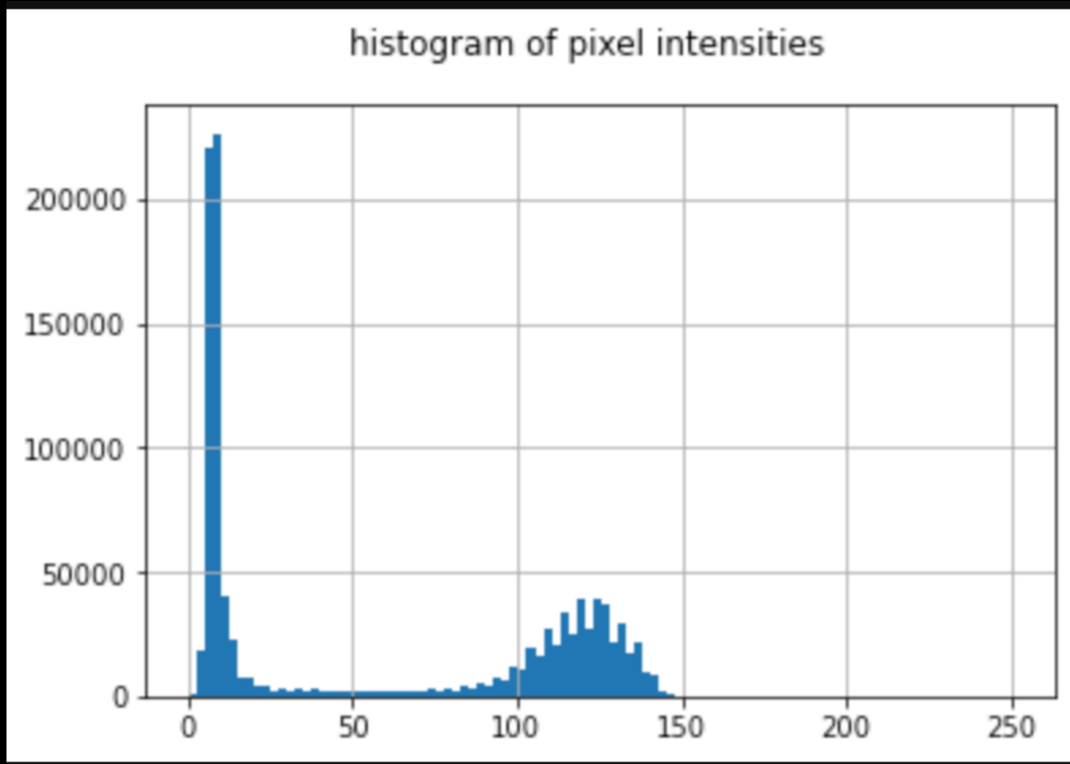


Low2Med: Workflow



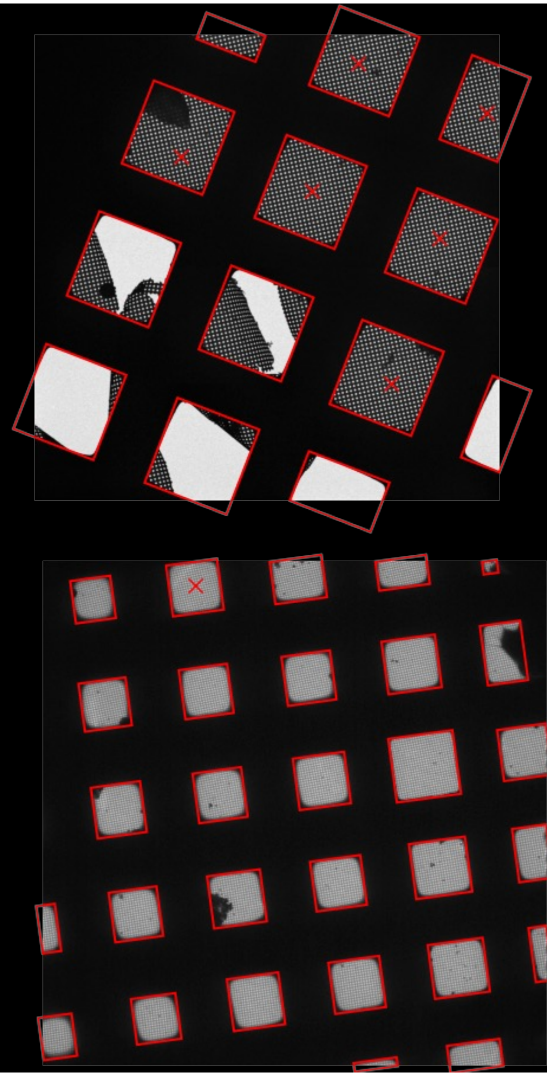
Low2Med: Why Mixture Model Works





Data + Training

- 1.3k total grid tile images
- **98.8% recall** of selected squares
- Extract 41k squares, 30k that user did not select, 11k selected
- Predict user selections using CNN on crops, LogReg/RF on image features



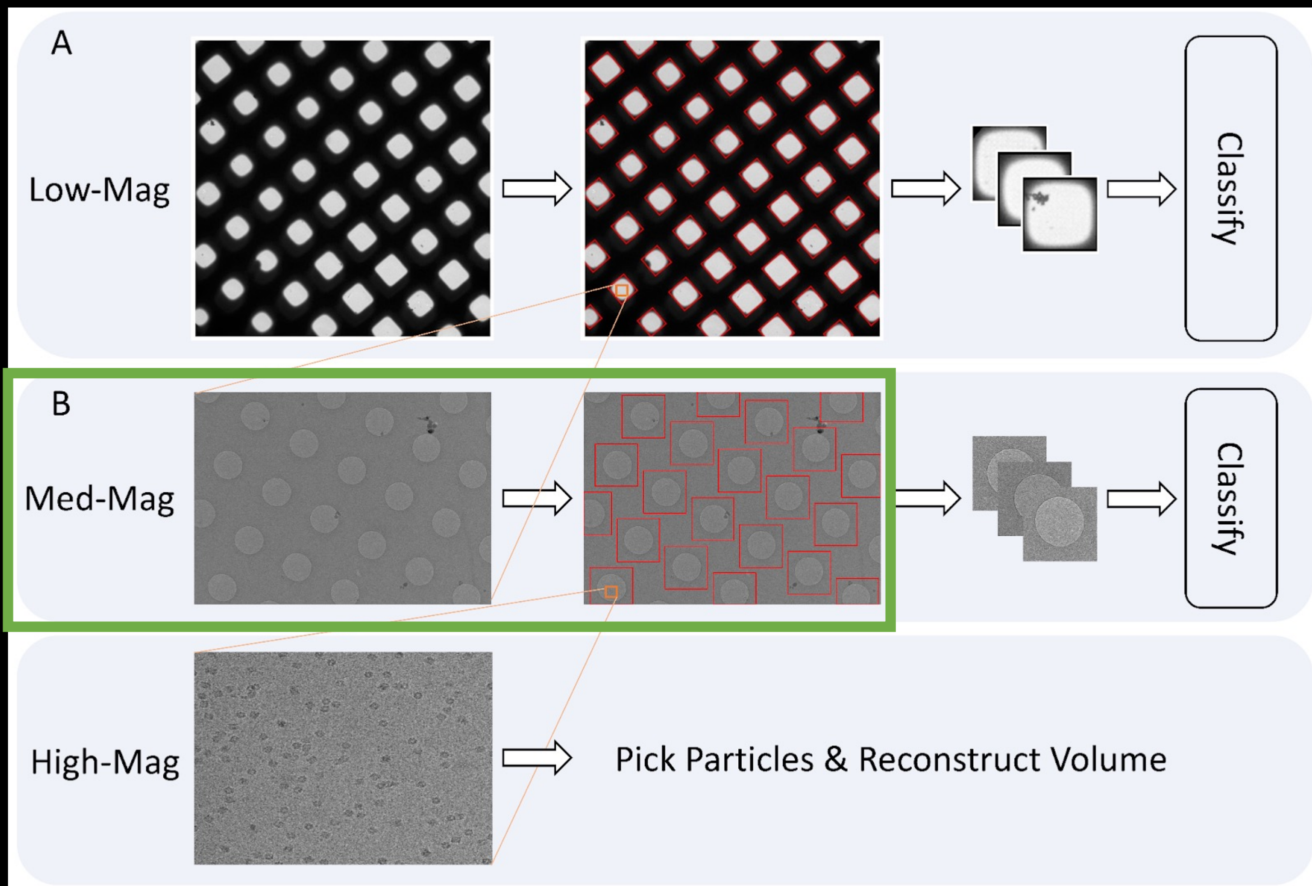
RF and CNN reasonably classify, session generalization is hard

Table 2. Performance metrics of different ML models on held-out-sessions.

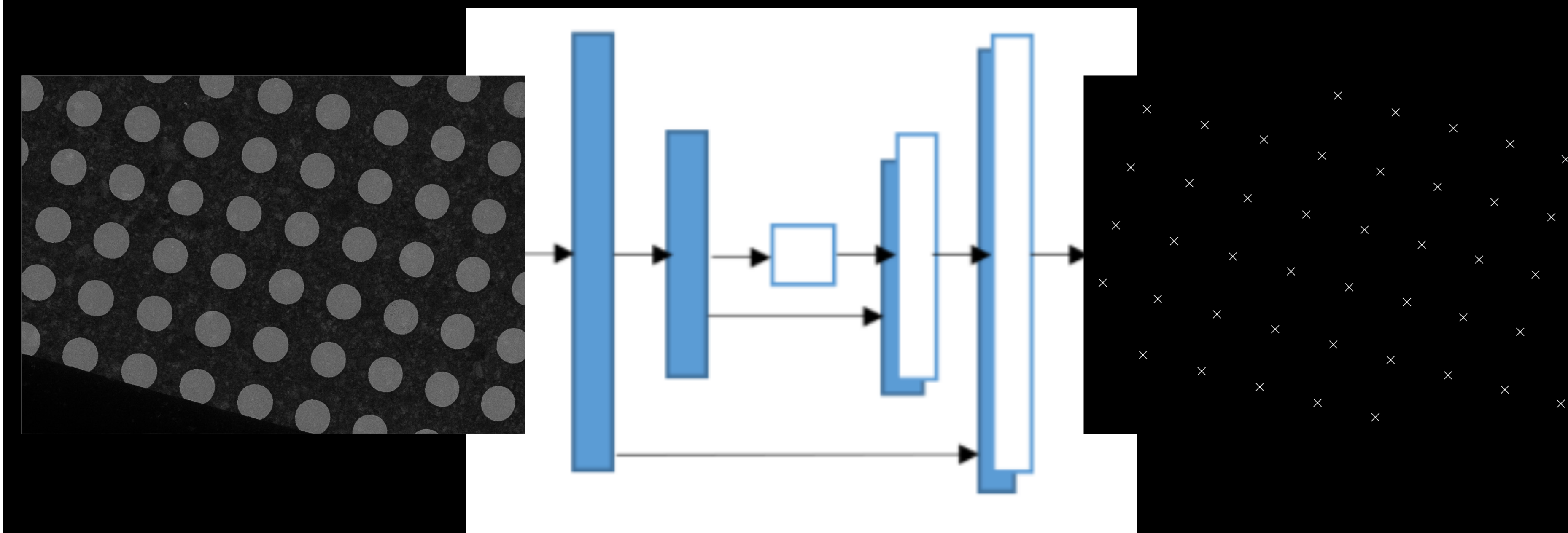
	Session Split		Random Split	
Model	ROC AUC	Avg Precision	ROC AUC	Avg Precision
LogReg	0.539	0.258	0.499	0.259
RF	0.603	**0.344**	**0.867**	**0.734**
CNN	**0.608**	0.331	0.733	0.489

- 400 model squares ≈ 100 operator squares
- Data contains many false negatives
- **Session generalization is hard** – we aren't doing screening
- RF does well
- Eliminates *bad* squares

Example images: darker blue are higher scoring, darker red are lower scoring

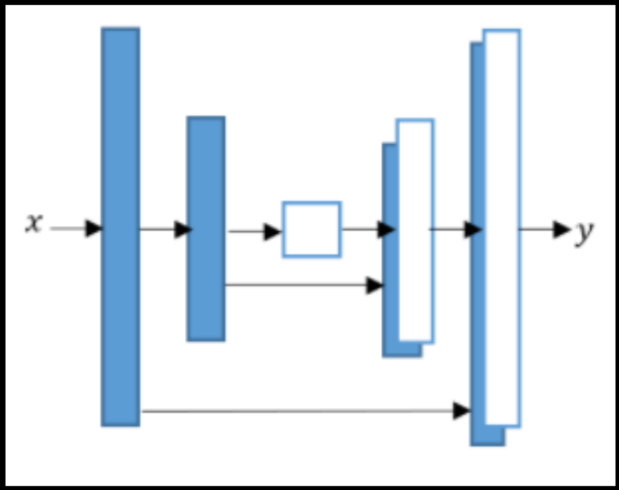

A
Low-Mag
Classify
B
Med-Mag
Classify
High-Mag
Pick Particles & Reconstruct Volume

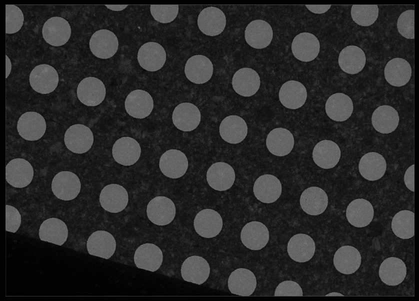
Med2High: Localization w/ U-Net

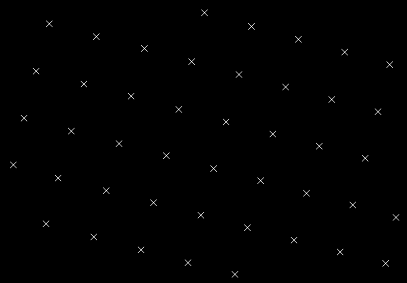


Localization Details

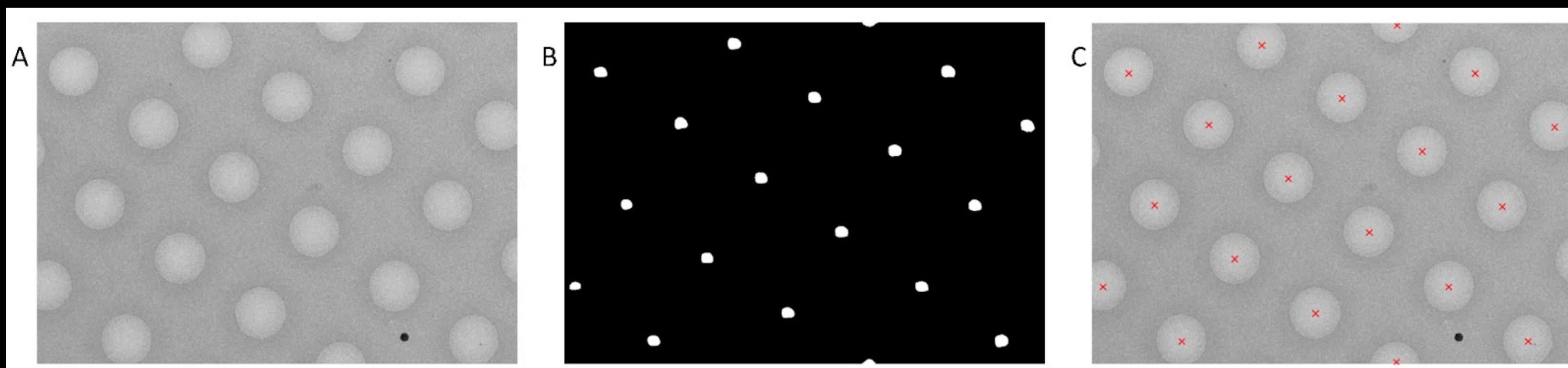
- Data: 28k carbon and gold holey-grid medium-mag images
- Predict operator selection locations from med-mag image using U-Net
- Gaussian smoothing of output + learning of smoothing sigma
    - To address uncertainty in the location where the operator selected

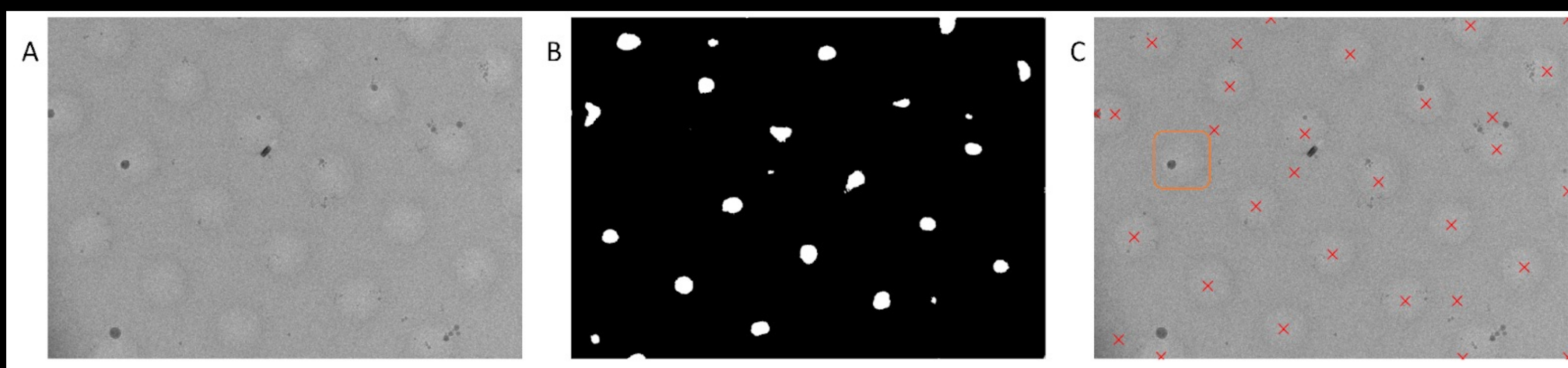






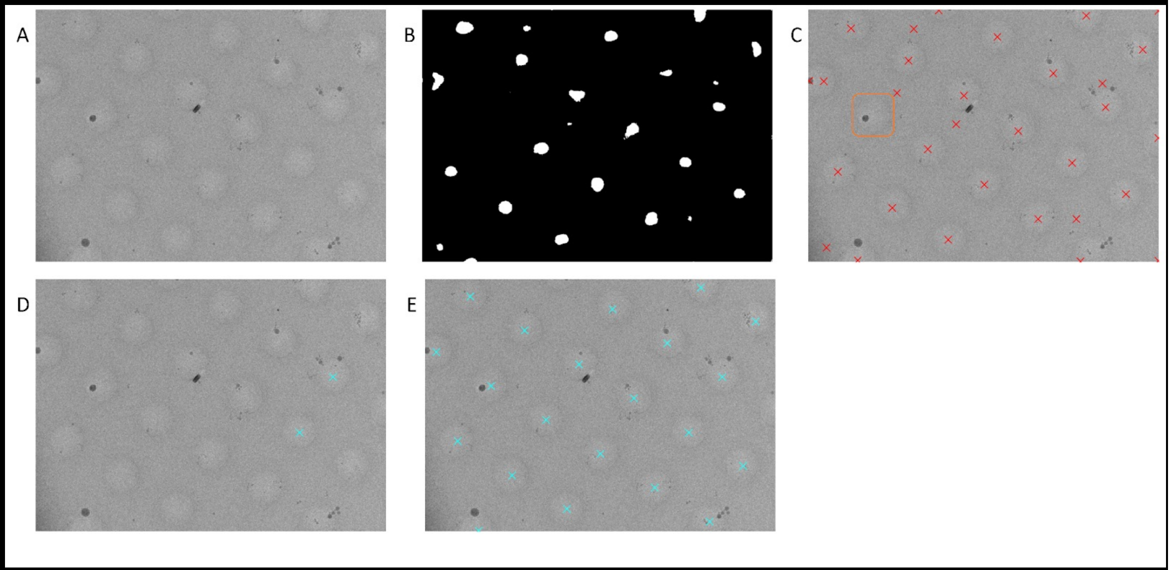
Sometimes U-Net is not enough





The solution: Lattice Fitting

- We know holes lie on square lattice
- Post process w/ lattice-fitting
- Find anchor points for lattice, where lattice points have smallest error from U-Net output



Localization succeeds, lattice fitting improves recall

Table 3. Performance metrics of different methods on held-out sessions for hole localization from medium-mag images. Reported metrics are aggregated by session and averaged.

Model	Precision	Recall	F1
Yolov3[6]	0.395	0.669	0.459
U-Net	0.703	0.984	0.815
U-Net + Lattice Fitting	0.549	**0.993**	0.702
U-Net + Lattice Fitting + Probability Threshold	**0.802**	0.891	**0.837**

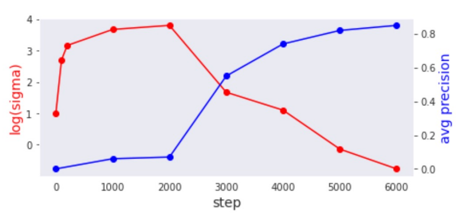
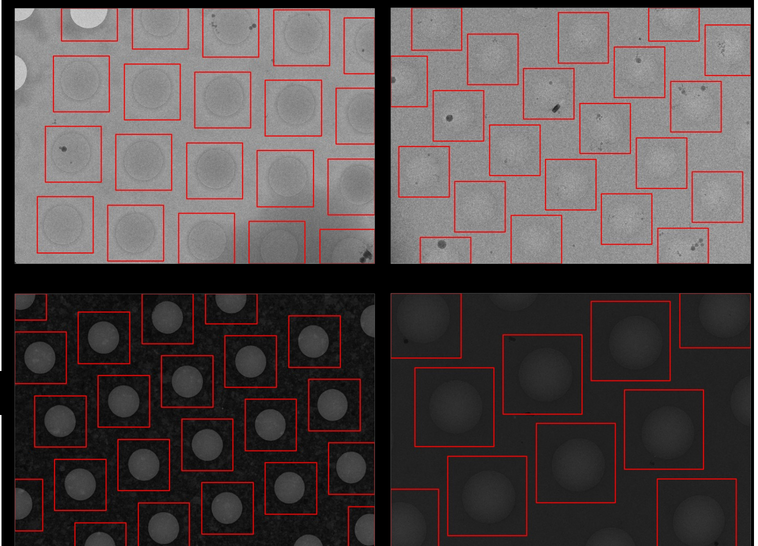
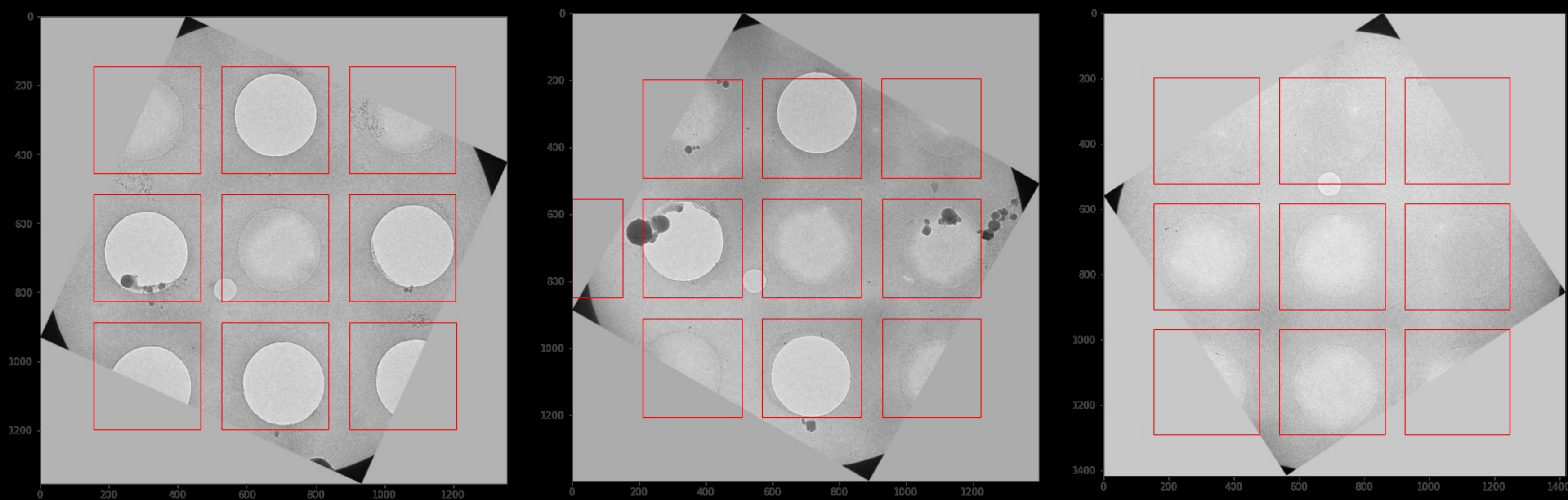


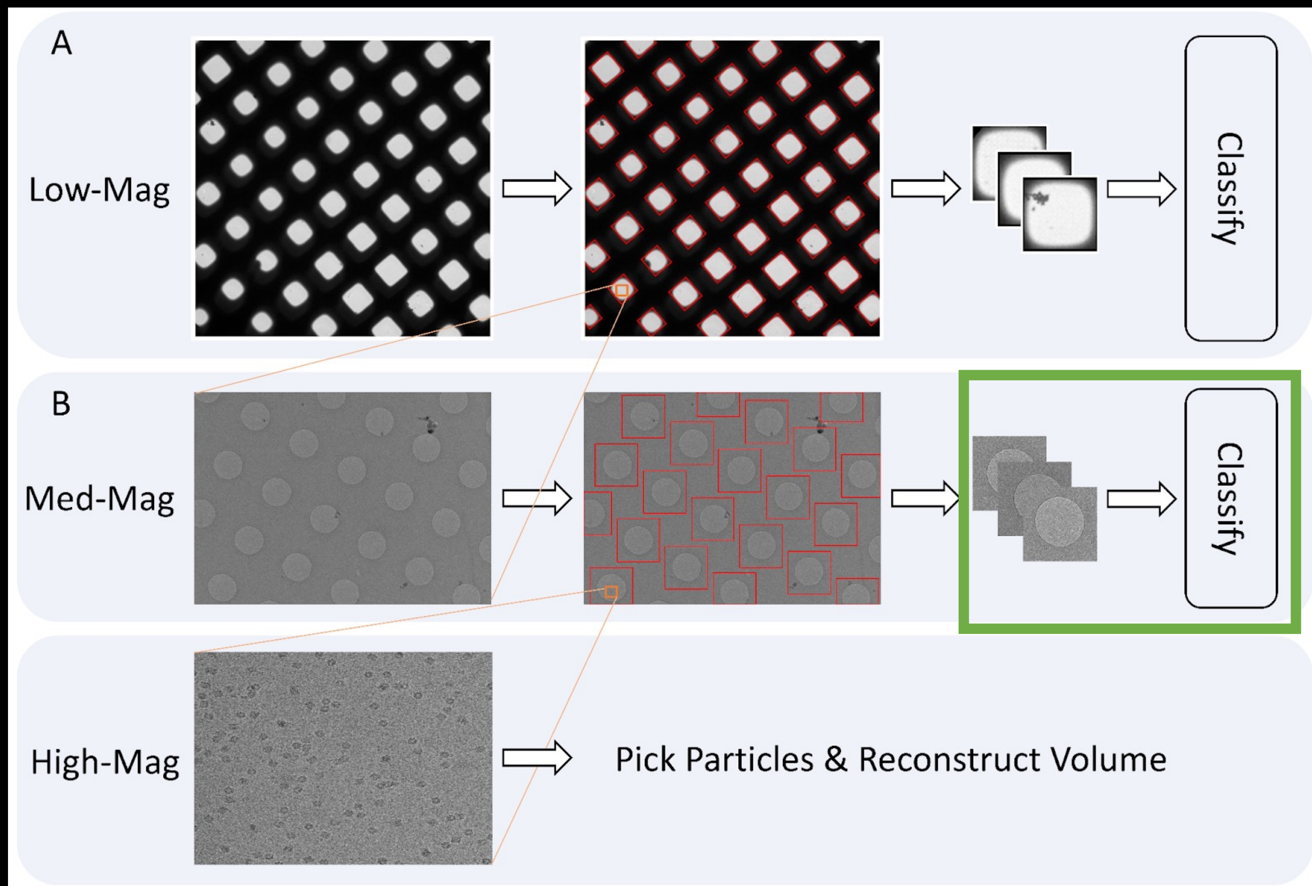
Figure 10. Sigma parameter versus model training progress. We plot the gaussian smoothing sigma parameter against average precision on validation set during training of U-Net.



Localization generalizes to external images



Recall: 0.95, Precision: 0.69



A
Low-Mag
Classify
B
Med-Mag
Classify
High-Mag
Pick Particles & Reconstruct Volume

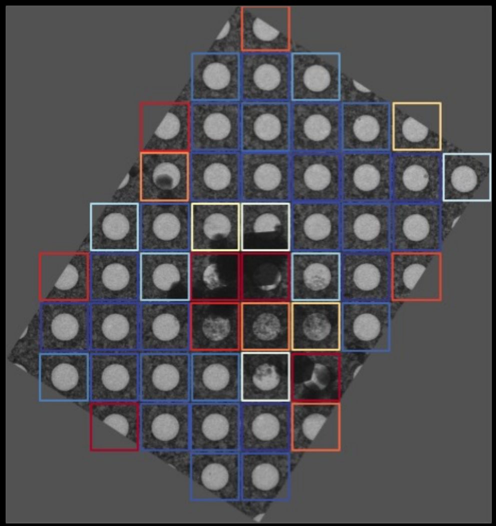
Med2High: Classification

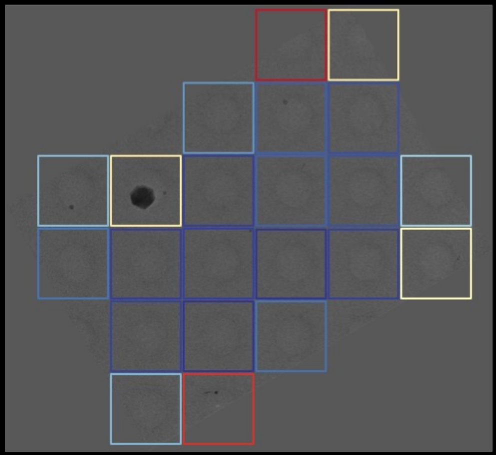
- 571k hole crops extracted from med-mag images
  - 410k targeted
- Large variation in image sizes, because large variation in hole sizes
- CNNs trained on crops
  - Padding vs avg pool

Models learn to classify, average pooling helps

Table 4. Performance of hole classification CNNs on hold-out sessions.

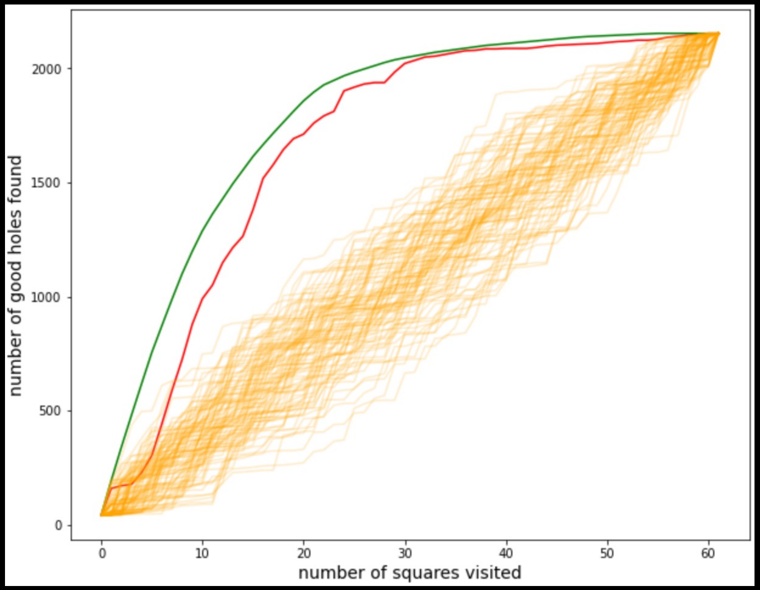
Model	Accuracy	ROC AUC	Avg Precision
CNN (padding)	0.748	0.742	0.808
CNN (avg pool)	0.758	0.796	**0.878**





Ongoing work: Active Learning

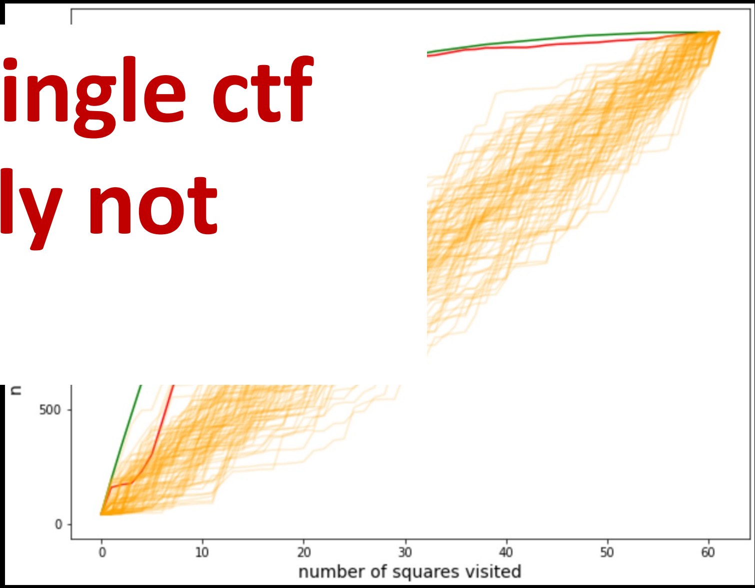
- Goal: learn characteristics of good and bad squares/holes per session
- Holes: find holes with low ctf resolution (angstroms)
- Squares: find squares with many good holes
- Assumption: square model > hole model
- Use Gaussian Process, square image features



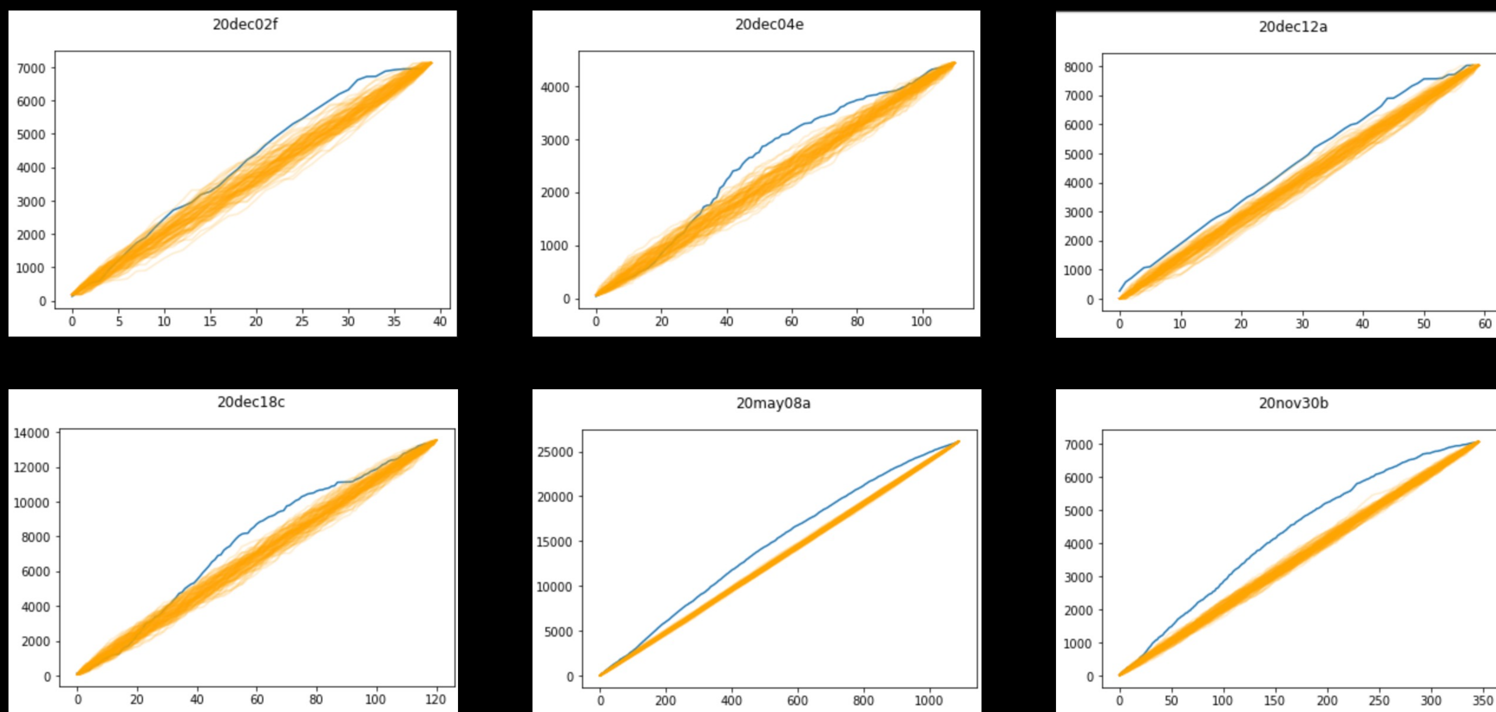
Ongoing work: Active Learning

- Goal: learn characteristics of good and bad sc
- Holes: finc resolution
- Squares: f good hole
- Assumptic model
- Use Gaussian Process, square image features

But apparently single ctf metric is probably not enough!



Active learning generalizes to real sessions



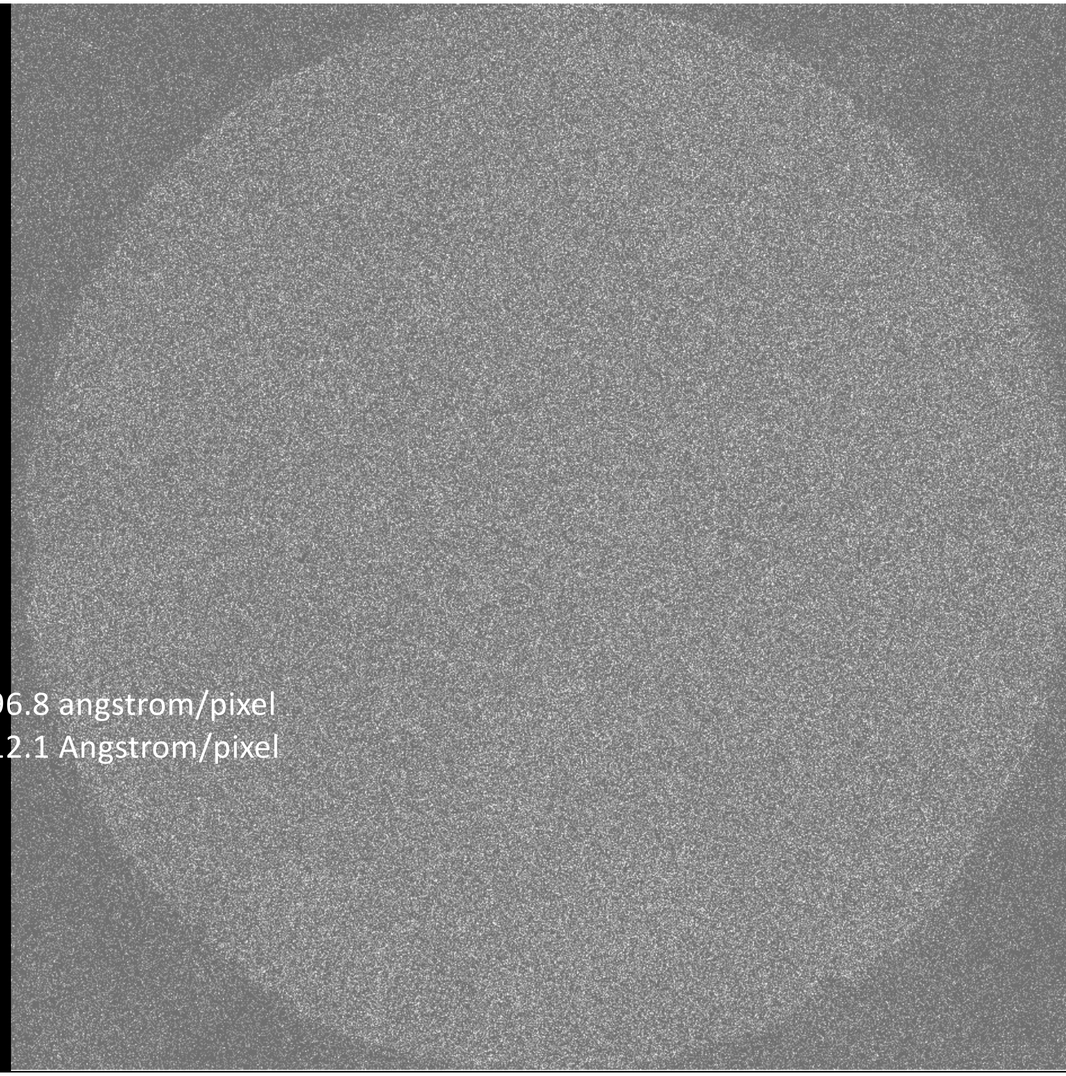
Future development & questions

- Data upload server
- Persistent model
- Modularity for non SPA use-cases
- Integration w/ collection software
- Revisit hole and square classification after Active Learning
- Better metrics/labels for active learning and beyond
- Can we detect hole xy locations directly from grid tile images?

Future development & questions

- Data upload server
- Persistent model
- Modularity for non SPA use-cases
- Integration w/ collection software
- Revisit hole and square classification after Active Learning
- Better metric for active learning
- A long tail of edge cases
  - Lacy, Chameleon, Dealing with bad grids, live processing integration, superresolution integration, better labels, different meshes, *where* in the square/hole to collect from beyond just the center/the tiling

Superresolution classification



Current Med Mag Images: 96.8 angstrom/pixel
Superresolution (unbinned) Images: 12.1 Angstrom/pixel

Thanks!

Operators: Hui Wei, Anjelique Sawh, Eugene Chua, Huihui Kuang, Joshua Mendez, Kashyap Maruthi


Anchi Cheng


Alex Noble


Tristan Bepler


Bridget Carragher


Clint Potter

NEW YORK STRUCTURAL BIOLOGY CENTER

SIMONS MACHINE LEARNING CENTER

SIMONS ELECTRON MICROSCOPY CENTER