Learning to automate cryo-electron microscopy data collection with Ptolemy

Smart Data Collection Workshop April 2022
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
Low2Med: Workflow

Input Image

PMM + Mask

Min bounding rectangle

Minimum Convex Hull
Low2Med: Workflow

- Input Image
- PMM + Mask
- Minimum Convex Hull
- Normalize + extract crops
- ConvNet Classifier

Probability (score) for each square
Low2Med: Why Mixture Model Works

Input Image

PMM + Mask

histogram of pixel intensities
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

• 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

---

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Session Split</th>
<th>Random Split</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ROC AUC</td>
<td>Avg Precision</td>
</tr>
<tr>
<td>LogReg</td>
<td>0.539</td>
<td>0.258</td>
</tr>
<tr>
<td>RF</td>
<td>0.603</td>
<td><strong>0.344</strong></td>
</tr>
<tr>
<td>CNN</td>
<td><strong>0.608</strong></td>
<td>0.331</td>
</tr>
</tbody>
</table>

---

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yolov36</td>
<td>0.395</td>
<td>0.669</td>
<td>0.459</td>
</tr>
<tr>
<td>U-Net</td>
<td>0.703</td>
<td>0.984</td>
<td>0.815</td>
</tr>
<tr>
<td>U-Net + Lattice Fitting</td>
<td>0.549</td>
<td>0.993</td>
<td>0.702</td>
</tr>
<tr>
<td>U-Net + Lattice Fitting + Probability Threshold</td>
<td><strong>0.802</strong></td>
<td>0.891</td>
<td><strong>0.837</strong></td>
</tr>
</tbody>
</table>

**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
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>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>ROC AUC</th>
<th>Avg Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN (padding)</td>
<td>0.748</td>
<td>0.742</td>
<td>0.808</td>
</tr>
<tr>
<td>CNN (avg pool)</td>
<td>0.758</td>
<td>0.796</td>
<td><strong>0.878</strong></td>
</tr>
</tbody>
</table>
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 squares/holes per session
- Holes: find holes with low ctf resolution
- Squares: find squares with many good holes
- Assumption: square model > hole 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 ångström/pixel
Superresolution (unbinned) Images: 12.1 Ångström/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