Automated systematic evaluation of cryo-EM specimens with SmartScope

Jonathan Bouvette
Ph.D, visiting fellow
National Institute of Environmental Health Sciences

April 6th 2022,
NYSBC
CryoEM workflow

1. Protein purification
2. Grid preparation
3. Screening
4. Dataset collection
5. Data processing
6. Structure
CryoEM workflow

Sample optimization

- Multiple cycles are required to obtain a good sample
- Most projects require preparing and screening >100 grids
- Each grids take >30 min to screen
Grid Screening

Goal of a screening session

<table>
<thead>
<tr>
<th>Learn as much as possible about the specimen</th>
<th>Thorough sampling</th>
<th>Good grid?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freezing conditions</td>
<td>Different ice thickness</td>
<td>Where are the best areas?</td>
</tr>
<tr>
<td>Sample quality</td>
<td>Find what is good</td>
<td>Enough for a dataset?</td>
</tr>
<tr>
<td></td>
<td>And what is bad</td>
<td>Improvements?</td>
</tr>
</tbody>
</table>

Diagnose and plan the next optimization cycle
Ease the optimization process
Maximize dataset quality
CryoEM workflow
Weekly on the NIEHS Arctica

- 80-100 grids screened:
  - 30 hours of active screening
  - 10 hours of grid preparation

- ~4-7 grid collected:
  - 20 hours of active setup
  - 80 hours of collection
Grid Screening is repetitive

Record atlas

62 x

Save Image
Choose Area
Move stage

210 x

Save Image
Choose square
Move stage
Eucentric

2300 x

Save image
Center on hole
Autofocus

36000 x

Save image
Star over
Manual grid screening – Cutting corners to speed up

- Incomplete metadata
- Suboptimal images
- Hard to navigate the results
- Subjective sampling

Save Image
Choose square
Move stage
Eucentric

Record atlas

Save Image
Choose Area
Move stage

210 x

2300 x

36000 x

Center on hole
Autofocus

Save image
Star over
Goals

• Automate screening
• Provide good sampling
• Complete data
• Intuitive interface
SmartScope – Automated workflow overview

Legend

<table>
<thead>
<tr>
<th>ROIs</th>
<th>Status</th>
<th>BIS type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Queued</td>
<td>BIS center</td>
</tr>
<tr>
<td>Bad</td>
<td>Completed</td>
<td>BIS target</td>
</tr>
<tr>
<td>Cracked</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SmartScope – Layered modular approach to area selection

<table>
<thead>
<tr>
<th>Finders</th>
<th>Classifiers</th>
<th>Selectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object detection</td>
<td>Named labels</td>
<td>Clustering</td>
</tr>
<tr>
<td></td>
<td>Finite number of categories</td>
<td>Tunable number of categories</td>
</tr>
</tbody>
</table>
SmartScope – Layered approach to area selection

- ROIs Detection and classification
- Cluster by area size
- Filtered clusters
- Selection from different clusters

- Good □ Bad □ Cracked □ Partial
- Smallest □ □ Largest
- Smallest □ □ Largest
- Good □ Bad □ Cracked □ Partial □ Queued
SmartScope – Layered approach to area selection

TOIs Detection  Cluster by intensity  Group for BIS  Selection from different clusters

- Target
- Darkest Brightest
- Darkest Brightest Queued
SmartScope – Layered modular approach to area selection

Finders
- Object detection
  - Can also act as a classifier
- RCNN square finder/classifier
- YOLO hole finder
- Binary square finder
- FFT hole finder
- Regular pattern

Classifiers
- Named labels
  - Finite number of categories
- Flow-based square classifier
  - RCNN square classifier

Selectors
- Clustering
  - Tunable number of categories
- Area size clustering
- Signal intensity clustering

Create custom workflows
Add new methods as plugins
Web Interface

- Real-time tracking
- Microscope interaction
Web Interface

- Real-time tracking
- Microscope interaction
Web Interface

- Real-time tracking
- Microscope interaction
- Preprocessing
Web Interface

- Real-time tracking
- Microscope interaction
- Preprocessing
- Annotation
Supervised Automatic screening
Giving the users some freedom

• Change Label
• Modify selection
• Annotation
• Changing parameters

• Micrograph curation (still under work)
Automatic screening
Leveraging early metadata

Faster R-CNN architecture

Identify and classify

Training set:

~ 1500 labeled squares
Hole Finder

- YOLO-based architecture
- AI hole finder is being trained to find holes on multiple grid types.
- Currently 10,000 holes in the training set.
- Precision of 98%, 89% recall
  - Mean-average precision 87%
Screening statistics

- **BIS**
- **No BIS**
- **BIS, n=38**
- **No BIS, n=942**
- **Median, n=980**

**Talos Arctica K2 detector**

- Squares sampled:
  - 1
  - 7
Automatic data collection
Quick setup and high-resolution capabilities
Conclusions

• Automated screening procedure
  – Square finder and classifier
  – Hole finder
  – Clustering methods

• Interactive interface
  – Ability to choose and modify area selection
  – Easy result access and complete bookkeeping

• Data persistence and organization

• Fast data collection setup

• Overnight screening sessions
CryoEM workflow
Weekly at the NIEHS Arctica

• >120 80-100 grids screened:
  – 30 hours of active screening
  – Lightly supervised automatic screening
  – 10 hours of grid preparation

• ~4-7 grid collected:
  – <10 20-hours of active setup
  – >90 80 hours of collection
Short term goals – More Flexibility with modular protocols

Protocol recipe

<table>
<thead>
<tr>
<th>Magnification level</th>
<th>Acquisition method</th>
<th>Finder (1)</th>
<th>Classifier (0 or more)</th>
<th>Selectors (1 or more)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High magnification</td>
<td>Acquisition method</td>
<td>Frames</td>
<td>Preprocessing</td>
<td></td>
</tr>
</tbody>
</table>

- Allow easy addition of Finders, Classifiers, Selectors as external plugins.
- Add acquisition methods to the microscope interface also as plugins.
- Create protocols by mixing existing methods.

Ease the integration of new workflows
Sample variety: virions, filaments, cells
Tomography
Sample-specific navigation roadmap

1. Sample specific state selection
2. User annotation to drive the selection on-the-fly
3. Using preprocessing information as feedback to drive the selection
4. Train AI models to drive the selection and “learn” about the samples
Acknowledgements

Bartesaghi Lab:
Dr. Alberto Bartesaghi
Wendy Qinwen Huang

Copeland Lab:
Dr. William Copeland
Dr. Amanda Riccio

UNC CryoEM:
Dr. Joshua Strauss
Jared Peck

Funding:
NIEHS/NIH

Elizabeth Viverette
Dr. Venkata Dandey
Dr. Kedar Sharma
Kevin John U. Butay
Dr. Kanda Borgognoni
Dr. Amanda Riccio
Dr. Mario J. Borgnia