

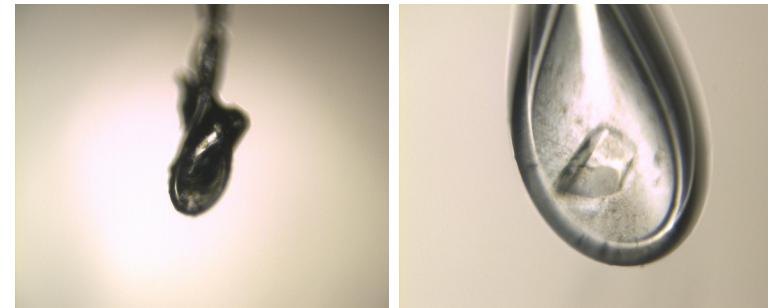
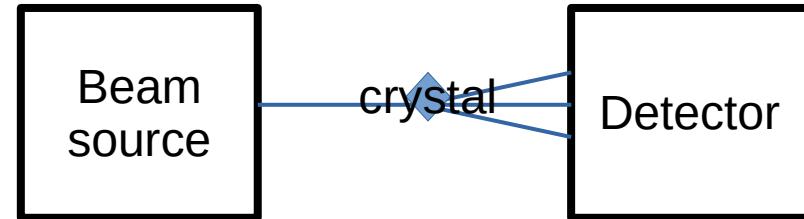


# **Automated ML-based Sample Centering for Macromolecular X-Ray Crystallography with MXAimbot**

Isak Lindhé, October 2021

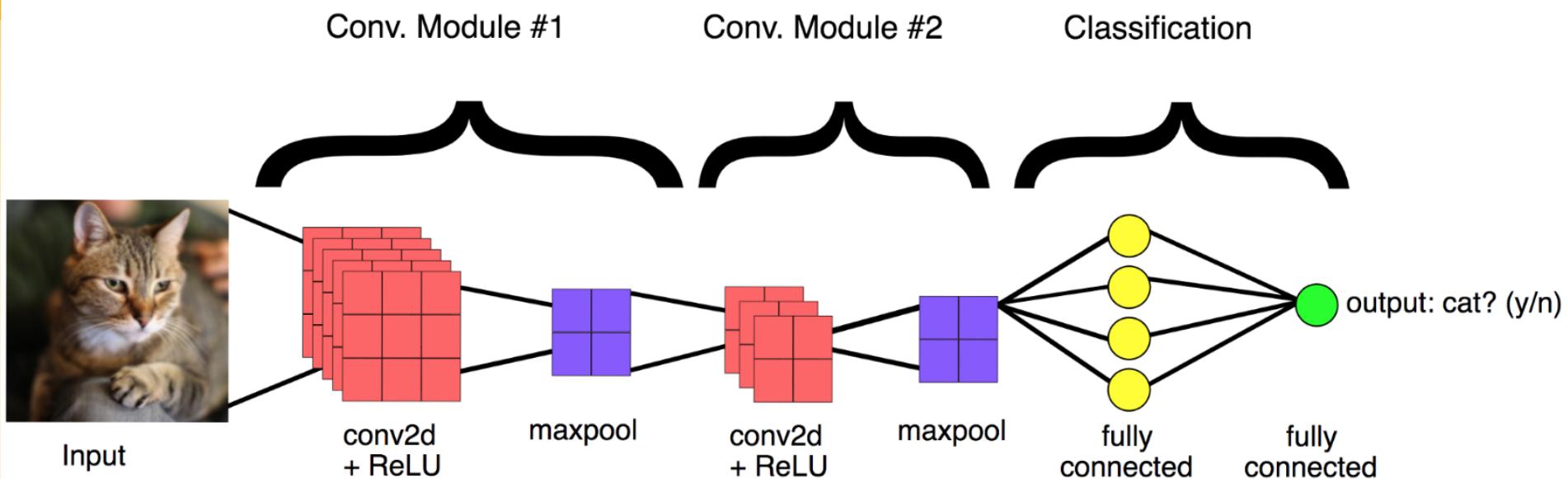
# Problem description

- The crystal must be centered in the beam in order to produce a good diffraction pattern.
- Existing methods for centering:
  - Manual centering
    - User clicks on the sample in a GUI three times to center it
    - **Time consuming**
  - X-ray centering with mesh scan
    - The brute force solution
    - “expose a big area around the loop and take the best result”
    - **Can cause radiation damage to some samples.**
- **MXAimbot: what if we did manual centering but with an artificial neural net instead of a slimy biological one?**



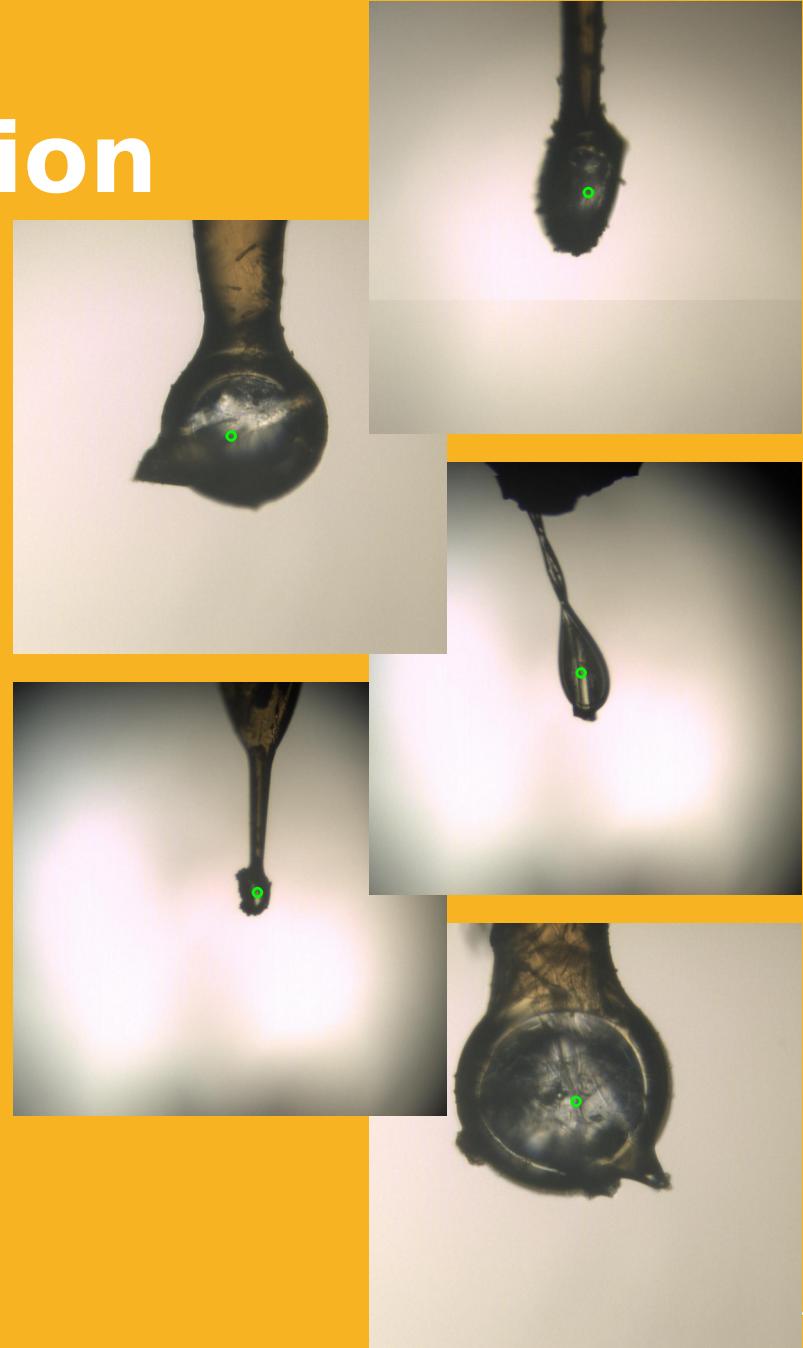
# A very quick introduction to convolutional neural networks

- A very flexible machine learning model which can be trained to approximate the relationship between given sets data.
- In this case: relationship between pictures of crystals and center location of those crystals in the image



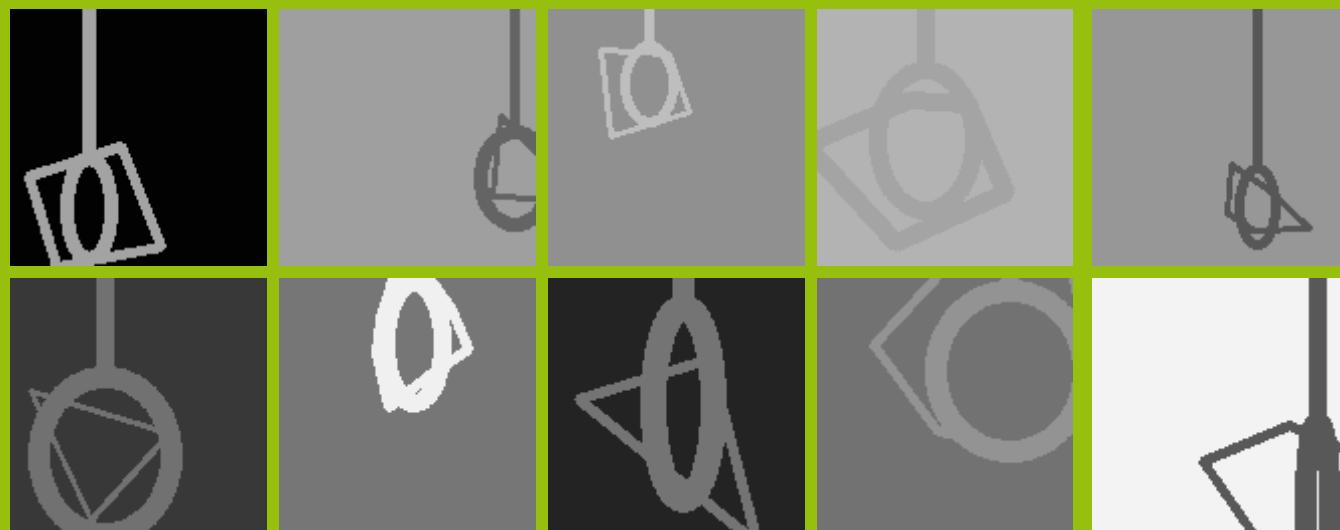
# Dataset Collection

- Crystal images from fragmentation screening experiments at the BioMAX beamline at MAX IV
- Filtered with classical computer vision techniques
  - Keyframe detection
  - Sharpness filter
  - Brightness filter
- Manually annotated with crystal center points
  - Size of manually annotated set: 1517 images
  - 60% for training, 20% for validation, 20% for testing



# Priming the predictor

- Pretraining on synthetic data
- Helps the neural net get the hang of the real data



# Model and Optimizations

- Convolutional Neural Network Regressor
  - Input: Grayscale, batch-normalized Image of crystals (128x128x1)
  - Output: Estimated coordinates of the crystals center point
- Images are randomly flipped to increase variation in each batch (data augmentation)
- Early stopping
  - “Stop training when the model stops improving”
  - Prevents the model from getting fixated on its training data (overfitting)

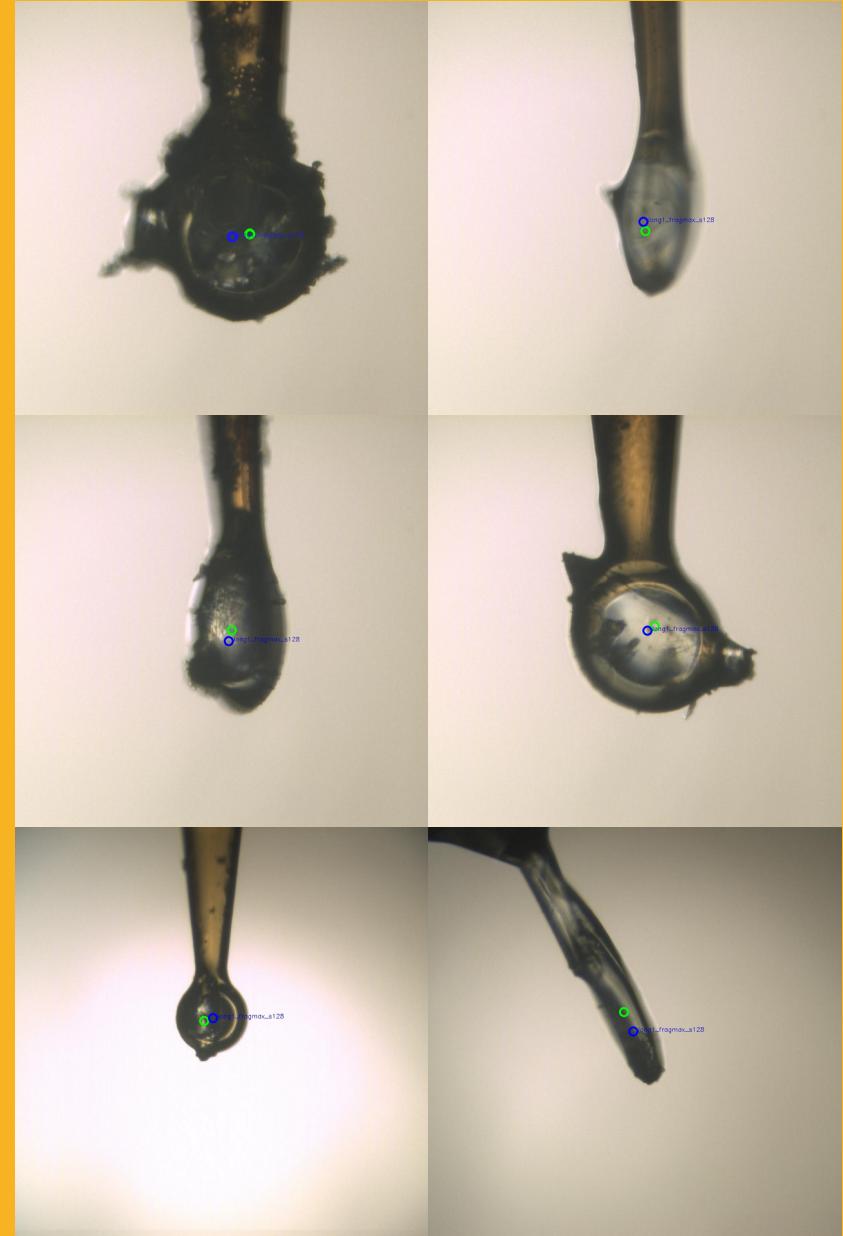
# Results

Current best model has a median distance between prediction and annotation coordinates of **3.3% of the image width.**

- X distance: 1.6%
- Y distance: 2.6%

green circle is annotation

blue circle is prediction



# Detecting outliers

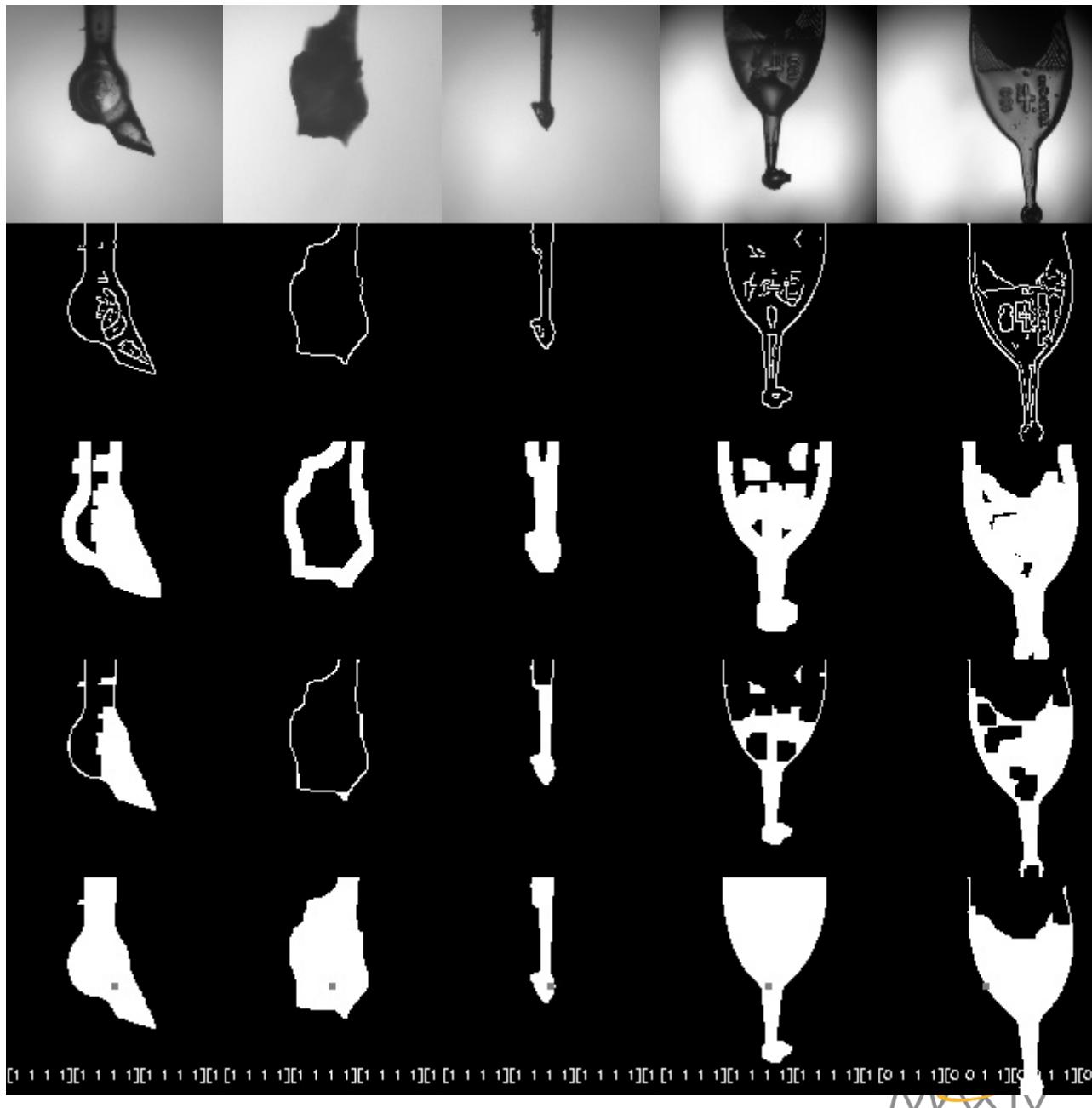
## Raw

# Canny edge detection

# Dilate

## Erode

# Flood fill background



# Next steps

- Full Usability test
  - A beamline scientist uses the workflow in the beamline as if collecting real data without my intervention.
  - Are the centering results acceptable often enough to let it run fully automatically and unsupervised?
  - If not, can it be used to assist the user in manual centering

Applications Places System

MxCuBE-3 Proposal: MX20180479 - Mozilla Firefox

MxCuBE-3 Proposal: MX20180479 - Mozilla Firefox

mxcube.maxiv.lu.se/datacollection

MxCuBE 3

Sample Overview Data collection Sample Changer System log

Beamline Actions ▾

Energy: 12.7000 keV  
Wavelength: 0.9763 Å

Resolution: 3.284 Å  
Detector: 500.000 mm

Transmission: 100.000 %  
Flux: 0 ph/s

Cryo: 0 K

Sample Changer MOVING  
Safety Shutter OPEN  
Ring Current 248.33

Phase Control:

Centring

Beam size: 10

Omega: 360

Kappa: 0

Phi: 0

Sample alignment:

Snapshot Draw grid 3-click Centring Focus Zoom

Diffractometer phase changed to Centring

Run Queue Next Sample (Thau - t2) Settings

Sample: Thau - t2

Queued Samples (0)

mxcube.maxiv.lu.se/samplegrid um

MxCuBE-3 Proposal: M... vis2

The screenshot shows the MxCuBE-3 software interface for a protein crystallography experiment. At the top, the status bar indicates the date as 'Tue Mar 23, 23:19'. The main window title is 'MxCuBE-3 Proposal: MX20180479 - Mozilla Firefox' and the URL is 'mxcube.maxiv.lu.se/datacollection'. The interface is divided into several sections: 'Beamline Actions' (dropdown), 'Phase Control' (beam size, Omega, Kappa, Phi sliders), 'Sample alignment' (button controls, motors dropdown), and a central 'Diffractometer phase changed to Centring' message with a preview image. Top navigation tabs include 'Sample Overview', 'Data collection', 'Sample Changer', and 'System log'. Key parameters displayed are Energy (12.7000 keV), Wavelength (0.9763 Å), Resolution (3.284 Å), Detector (500.000 mm), Transmission (100.000 %), Flux (0 ph/s), and Cryo (0 K). Control buttons for 'Sample Changer' (MOVING), 'Safety Shutter' (OPEN), and 'Ring Current' (248.33) are also present. A 'Run Queue' button and 'Next Sample (Thau - t2)' are visible on the right. The bottom status bar shows the URL 'mxcube.maxiv.lu.se/samplegrid um' and the title 'MxCuBE-3 Proposal: M... vis2'.

# Thanks for listening!

**And thanks to everyone at BioMAX!**

**And thanks to the authors of the paper**

**DeepCentering: fully automated  
crystal centering using deep learning  
for macromolecular crystallography**

**( DOI: 10.1107/S160057751900434X )**

**If you are interested in MXAimbot,  
email me at [isak.lindhe@maxiv.lu.se](mailto:isak.lindhe@maxiv.lu.se)**