Rutgers Winter Boot Camp
# Cellular Tomogram Annotation

Muyuan Chen
2018-01-12

# Tutorial

# What you will need …

- **Software:**

  - EMAN2.2

  - EMAN2.2 patch

- **Data**

  - AnnotationTutorial

    - tomoseg_init

    - NNET

# EMAN2

- Single particle reconstruction

- Tomography

  - Tilt series alignment & reconstruction

  - **Cellular tomogram annotation**

  - Subtomogram averaging

  - CTF correction

# Hardware and system recommendations

- Linux workstation

- Nvidia GPU with CUDA7.5+ (>10x speed up for training)

- Enough memory for 2 (binned) tomograms (~2GB for 1Kx1Kx256)

- >=3 button mouse

# Hardware and system recommendations

- If you do not have recommended hardware…

    - You can still go through the tutorial on a laptop using CPUs

    - Certain steps can be very slow…

    - All output files for the workshop are available

# Dataset

- PC12 cell

- Binned by 4

  - 864x864x94, 28 Å/pixel

- 4 features to annotate

  - Microtubules

  - Ribosomes

  - Single membranes

  - Double membranes
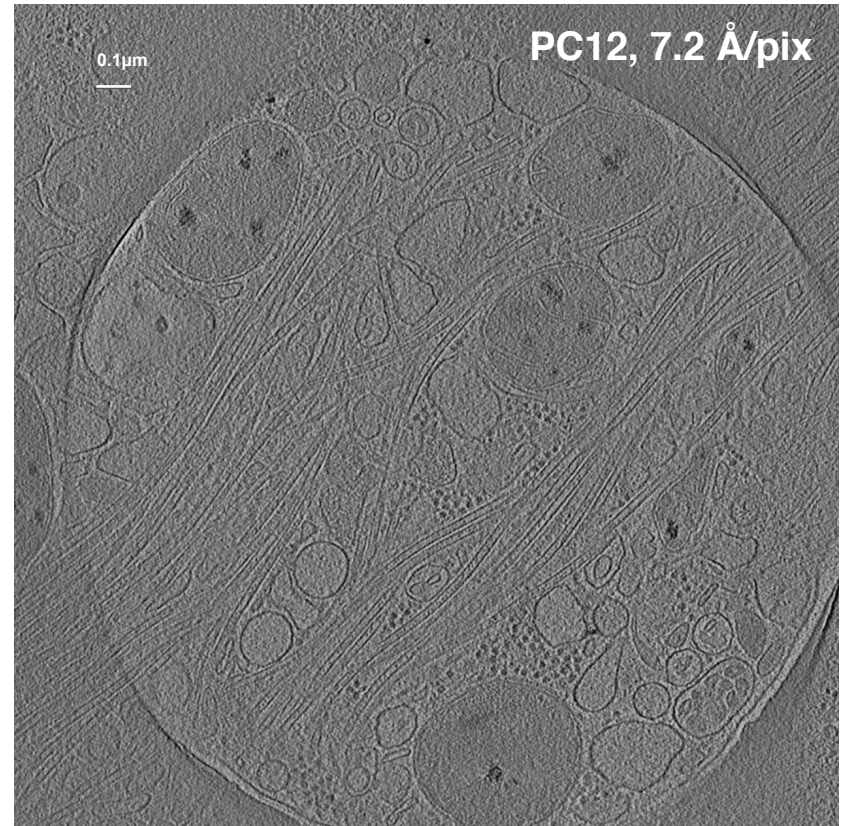
# Setting up working directory

- mkdir tomoseg

- cd tomoseg

- e2projectmanager.py


- **Make sure to run everything <u>within</u> the working directory!**

# Directory structure

./

./info

./rawtomograms

./particles

./particles3d

./spt_xx

…

e2display.py Browser

Path: ./tomoseg_mt.hdf                                    Info

Filter:                                                   Sel All

EMEN2
SSH
Root
Current
Home

| Row | Name | Type | Size | Dim | N | Date |
|---|---|---|---|---|---|---|
| 0 | display.py | Text | 43 k | 900 ln | - | 2017/11/14 16:31:44 |
| 1 | e2boxercache | Folder | | | | 2017/11/14 14:12:17 |
| 2 | info | Folder | | | | 2017/11/15 13:25:24 |
| 3 | nnet_save_dbmemb.hdf | Image Stack | 5 m | 15 x 15 | 1684 | 2017/11/14 15:18:45 |
| 4 | nnet_save_memb.hdf | Image Stack | 5 m | 15 x 15 | 1684 | 2017/11/14 14:56:20 |
| 5 | nnet_save_mt.hdf | Image Stack | 5 m | 15 x 15 | 1684 | 2017/11/14 13:56:39 |
| 6 | nnet_save_ribo.hdf | Image Stack | 5 m | 15 x 15 | 1684 | 2017/11/14 14:30:15 |
| 7 | particles | Folder | | | | 2017/11/14 15:13:57 |
| | 0 rawtomograms-pc12-DI144h-102413t3-bin2_bin2_dbmemb_bad.hdf | Image Stack | 2 m | 64 x 64 | 100 | 2017/11/14 15:13:11 |
| | 1 rawtomograms-pc12-DI144h-102413t3-bin2_bin2_dbmemb_good.hdf | Image Stack | 195 k | 64 x 64 | 8 | 2017/11/14 15:09:59 |
| | 2 rawtomograms-pc12-DI144h-102413t3-bin2_bin2_dbmemb_good_seg.hdf | Image Stack | 195 k | 64 x 64 | 8 | 2017/11/14 15:11:03 |
| | 3 rawtomograms-pc12-DI144h-102413t3-bin2_bin2_dbmemb_good_trainset... | Image Stack | 94 m | 64 x 64 | 4000 | 2017/11/14 15:13:49 |
| | 4 rawtomograms-pc12-DI144h-102413t3-bin2_bin2_memb_bad.hdf | Image Stack | 2 m | 64 x 64 | 100 | 2017/11/14 14:50:58 |
| | 5 rawtomograms-pc12-DI144h-102413t3-bin2_bin2_memb_good.hdf | Image Stack | 195 k | 64 x 64 | 8 | 2017/11/14 14:46:10 |
| | 6 rawtomograms-pc12-DI144h-102413t3-bin2_bin2_memb_good_seg.hdf | Image Stack | 195 k | 64 x 64 | 8 | 2017/11/14 14:47:30 |
| | 7 rawtomograms-pc12-DI144h-102413t3-bin2_bin2_memb_good_trainset.hdf | Image Stack | 94 m | 64 x 64 | 4000 | 2017/11/14 14:51:34 |
| | 8 rawtomograms-pc12-DI144h-102413t3-bin2_bin2_mt_bad.hdf | Image Stack | 2 m | 64 x 64 | 100 | 2017/11/14 13:51:24 |
| | 9 rawtomograms-pc12-DI144h-102413t3-bin2_bin2_mt_good.hdf | Image Stack | 195 k | 64 x 64 | 8 | 2017/11/14 13:47:34 |
| | rawtomograms-pc12-DI144h-102413t3-bin2_bin2_mt_good_seg.hdf | Image Stack | 195 k | 64 x 64 | 8 | 2017/11/14 13:49:35 |
| | rawtomograms-pc12-DI144h-102413t3-bin2_bin2_mt_good_trainset.hdf | Image Stack | 94 m | 64 x 64 | 4000 | 2017/11/14 13:52:08 |
| | rawtomograms-pc12-DI144h-102413t3-bin2_bin2_ribo_bad.hdf | Image Stack | 2 m | 64 x 64 | 100 | 2017/11/14 14:18:14 |
| | rawtomograms-pc12-DI144h-102413t3-bin2_bin2_ribo_good.hdf | Image Stack | 195 k | 64 x 64 | 8 | 2017/11/14 14:12:28 |
| | rawtomograms-pc12-DI144h-102413t3-bin2_bin2_ribo_good_seg.hdf | Image Stack | 195 k | 64 x 64 | 8 | 2017/11/14 14:16:20 |
| | rawtomograms-pc12-DI144h-102413t3-bin2_bin2_ribo_good_trainset.hdf | Image Stack | 94 m | 64 x 64 | 4000 | 2017/11/14 14:18:47 |
| 8 | particles3d | Folder | | | | 2017/11/14 14:06:37 |
| | 0 pc12-DI144t3-102413t3__mt.hdf | Image Stack | 43 m | 48 x 48 x 48 | 97 | 2017/11/14 14:04:21 |
| 9 | rawtomograms | Folder | | | | 2017/11/14 14:40:41 |
| | 0 pc12-DI144h-102413t3-bin2_bin2.hdf | Image | 281 m | 868 x 864 x 94 | 1 | 2017/11/14 13:41:42 |
| 10 | rec | Folder | | | | 2017/11/14 13:41:18 |
| | 0 pc12-DI144h-102413t3-bin2.rec | Image | 563 m | 1736 x 1728 x 188 | 1 | 2013/10/25 17:07:51 |
| 11 | spt_mt | Folder | | | | 2017/11/14 16:01:41 |
| | 0 output.hdf | Image | 457 k | 48 x 48 x 48 | 1 | 2017/11/14 14:54:32 |
| | 1 ref.hdf | Image | 456 k | 48 x 48 x 48 | 1 | 2017/11/14 14:15:07 |
| | 2 tmpout.hdf | Image Stack | 456 k | 48 x 48 x 48 | 6 | 2017/11/14 14:54:32 |
| 12 | tomoseg_dbmemb.hdf | Image | 281 m | 868 x 864 x 94 | 1 | 2017/11/14 15:31:28 |
| 13 | tomoseg_memb.hdf | Image | 281 m | 868 x 864 x 94 | 1 | 2017/11/14 15:05:34 |
| 14 | tomoseg_mt.hdf | Image | 281 m | 868 x 864 x 94 | 1 | 2017/11/14 13:59:49 |
| 15 | tomoseg_ribo.hdf | Image | 281 m | 868 x 864 x 94 | 1 | 2017/11/14 14:36:26 |
| 16 | trainout_nnet_save_dbmemb.hdf | Image Stack | 5 m | 64 x 64 | 300 | 2017/11/14 15:18:42 |
| 17 | trainout_nnet_save_memb.hdf | Image Stack | 5 m | 64 x 64 | 300 | 2017/11/14 14:56:17 |
| 18 | trainout_nnet_save_mt.hdf | Image Stack | 5 m | 64 x 64 | 300 | 2017/11/14 13:56:37 |
| 19 | trainout_nnet_save_ribo.hdf | Image Stack | 5 m | 64 x 64 | 300 | 2017/11/14 14:30:13 |
| 20 | wiener.txt | Plot | 864 | 48 x 2 | - | 2017/11/14 16:20:13 |

Show 3D    Show Stack    Show 2D    Chimera    ProjXYZ
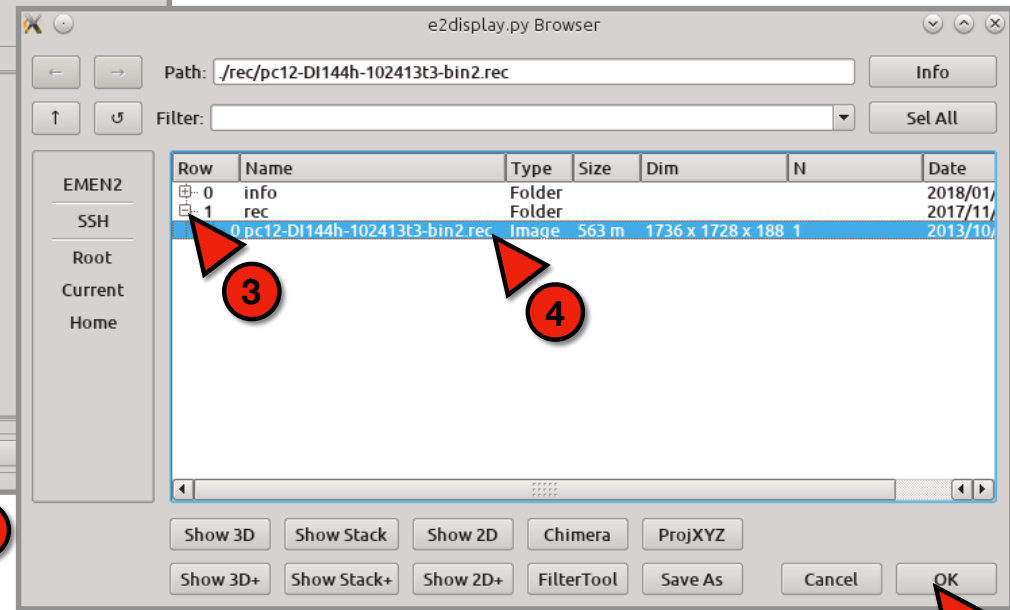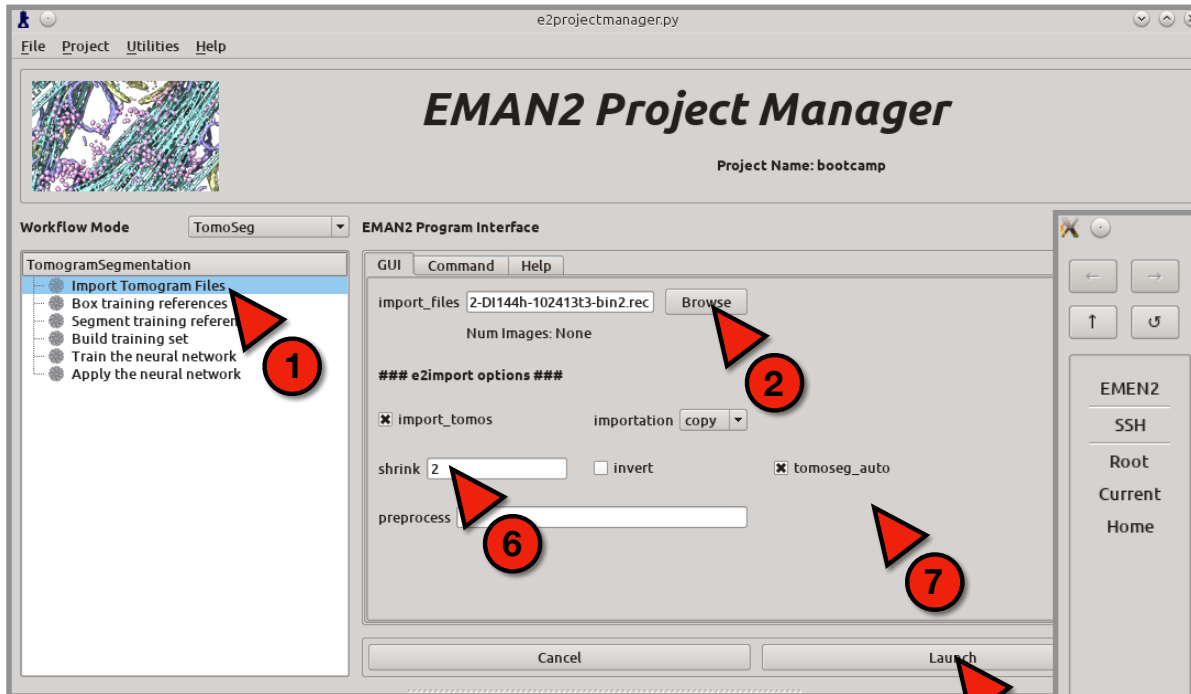
Show 3D+    Show Stack+    Show 2D+    FilterTool    Save As

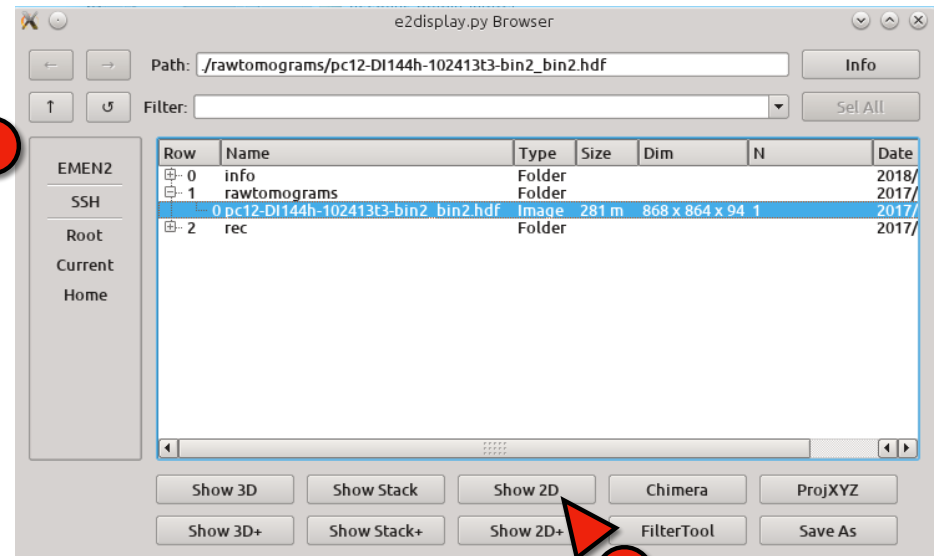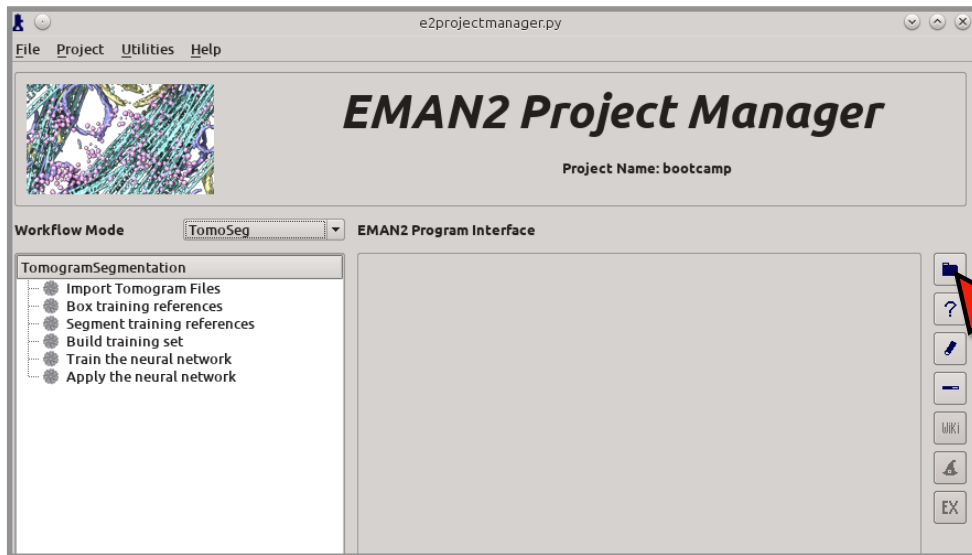# e2projectmanager
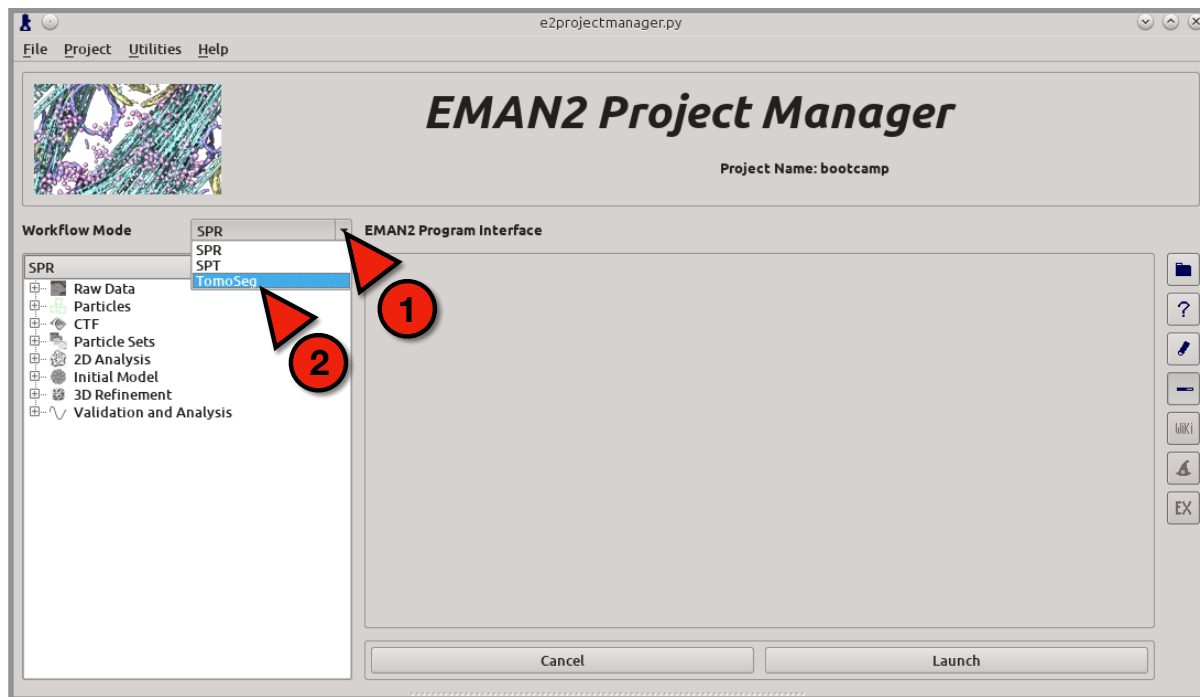
# Import tomogram



**Always import tomograms before processing!**
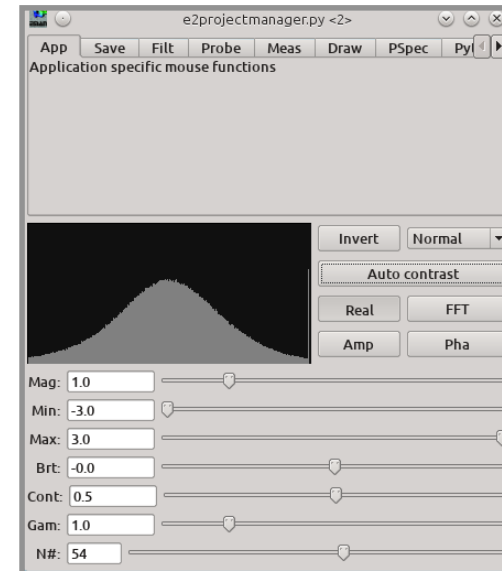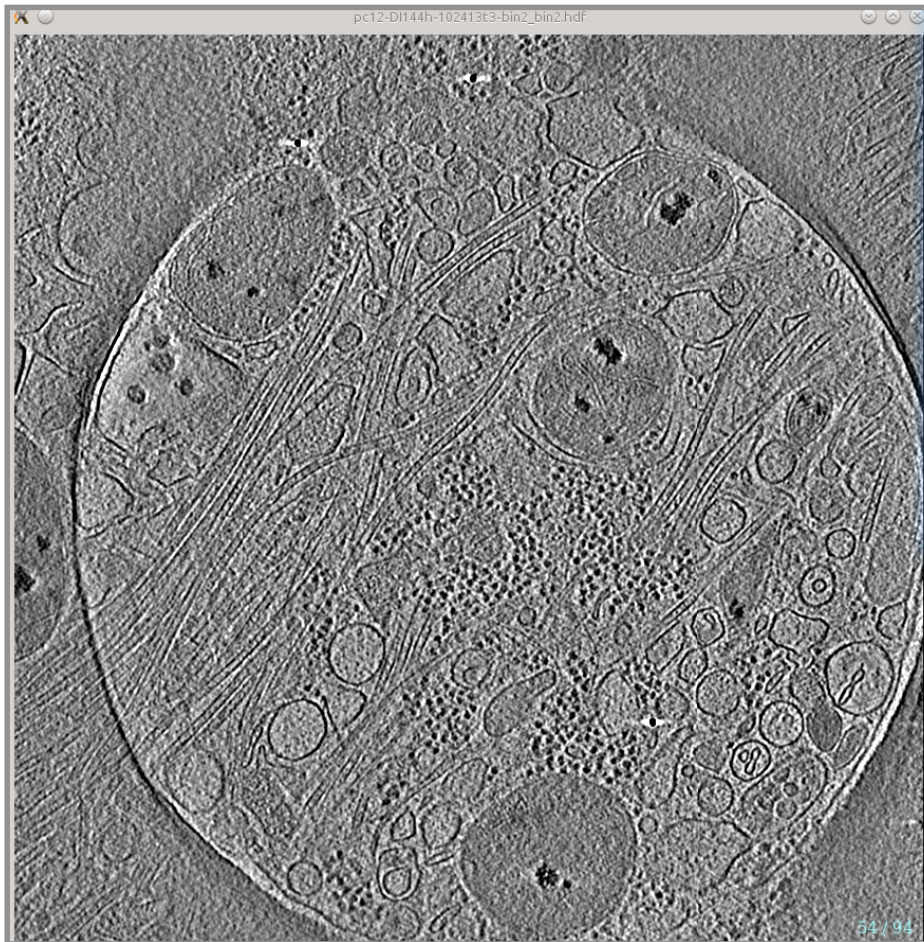
If you are using a Mac: you may need to click somewhere else (i.e. another box or command tab) before click *Launch* …

This will generate **rawtomograms/xxx_bin2.hdf**

# Visualize tomogram

# Visualize tomogram



⬆⬇ to change planes

Control panel:
Mouse middle / Option + mouse left
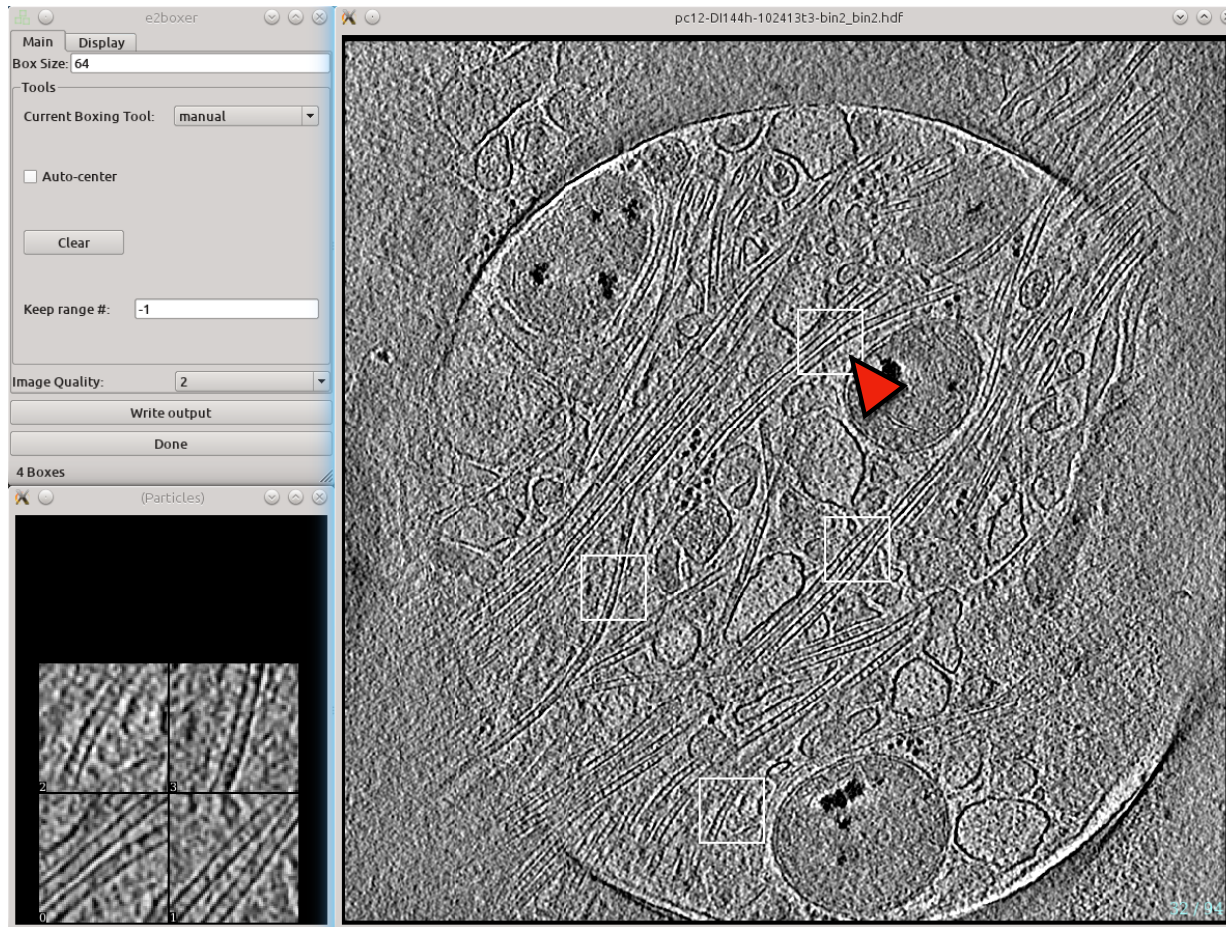
# Box training references

# Box training references



⬆⬇ to change planes

Mouse scroll to zoom in/out
Mouse left to select particles
Shift + Mouse left to remove selections

8 good references used

# Box training references
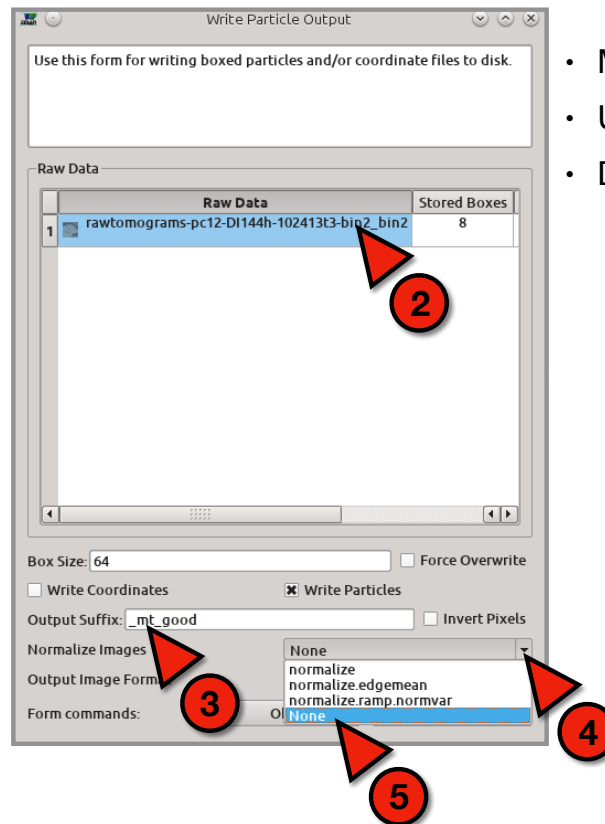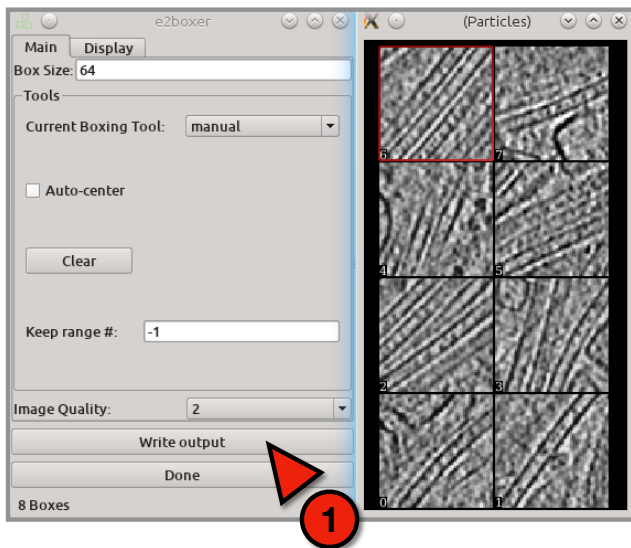


- Make sure **Box Size** is 64
- Use different **Output Suffix** for different features
- Do not normalize images

This will generate **particles/xxx_mt_good.hdf**

# Segment training references

# Segment training references
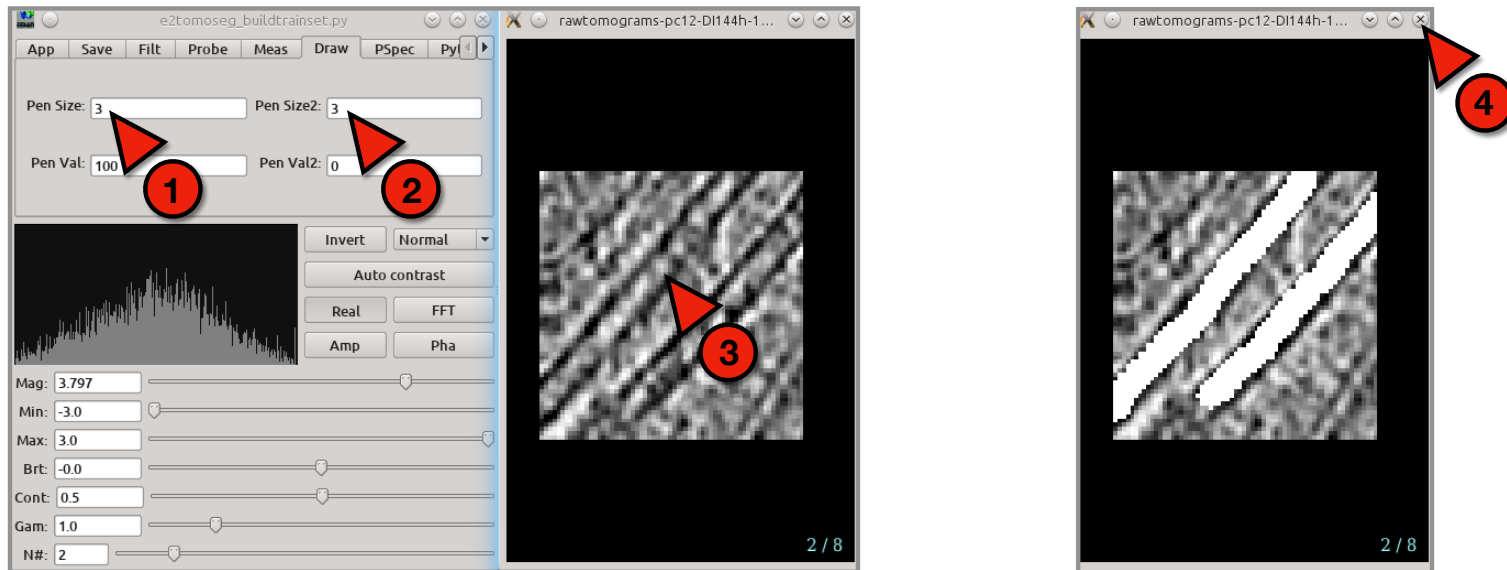


↑↓ to switch between images

This will generate **particles/xxx_mt_good_seg.hdf**

# Box negative references
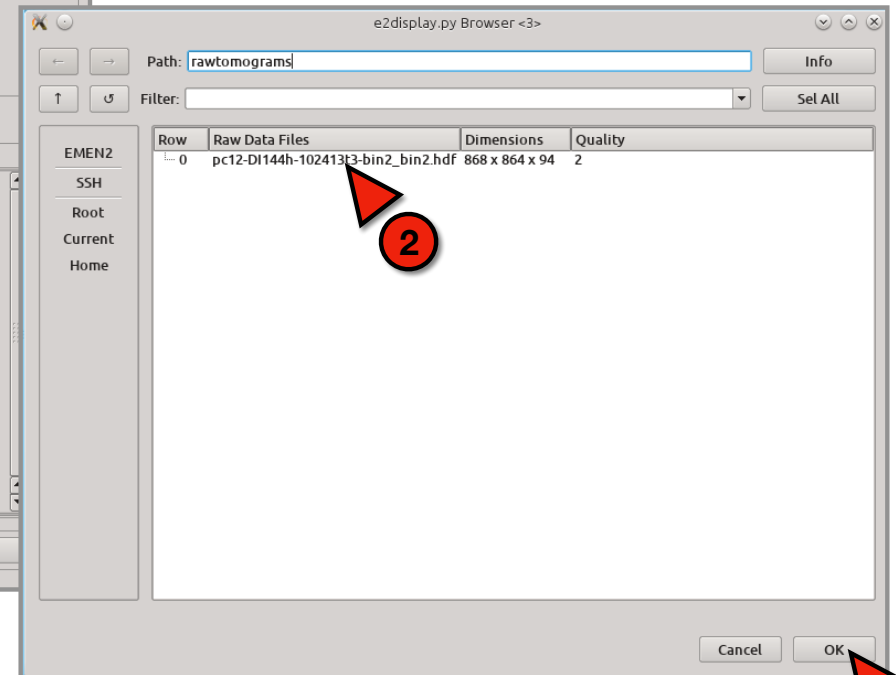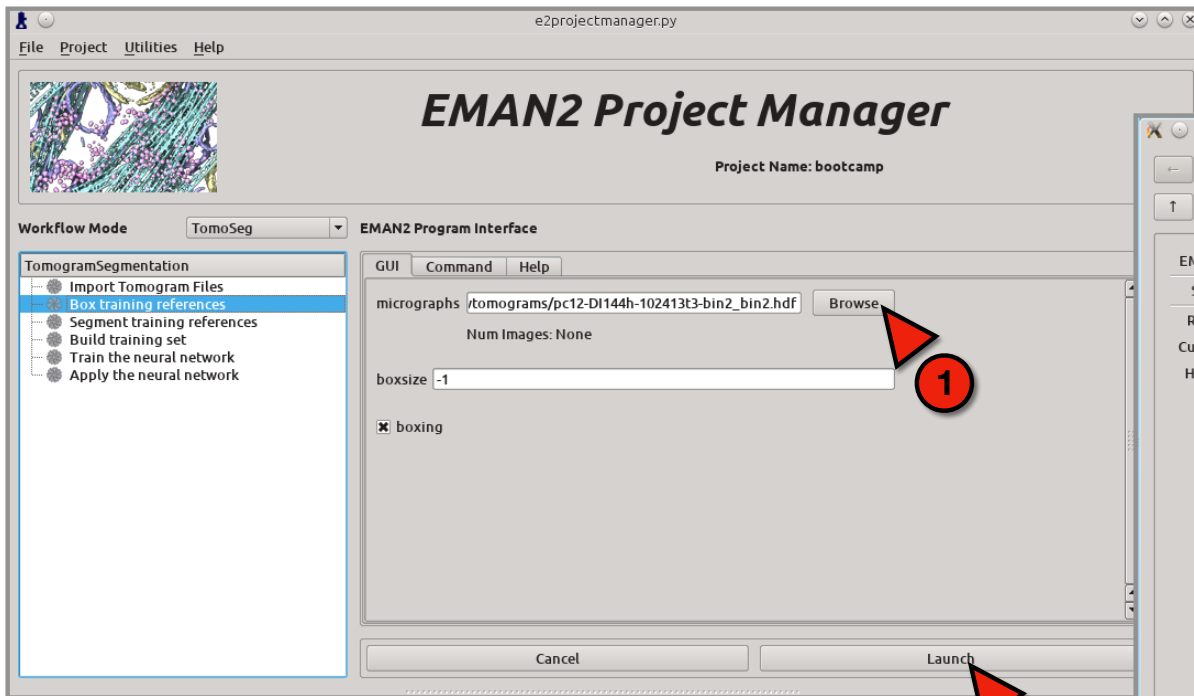
# Box negative references
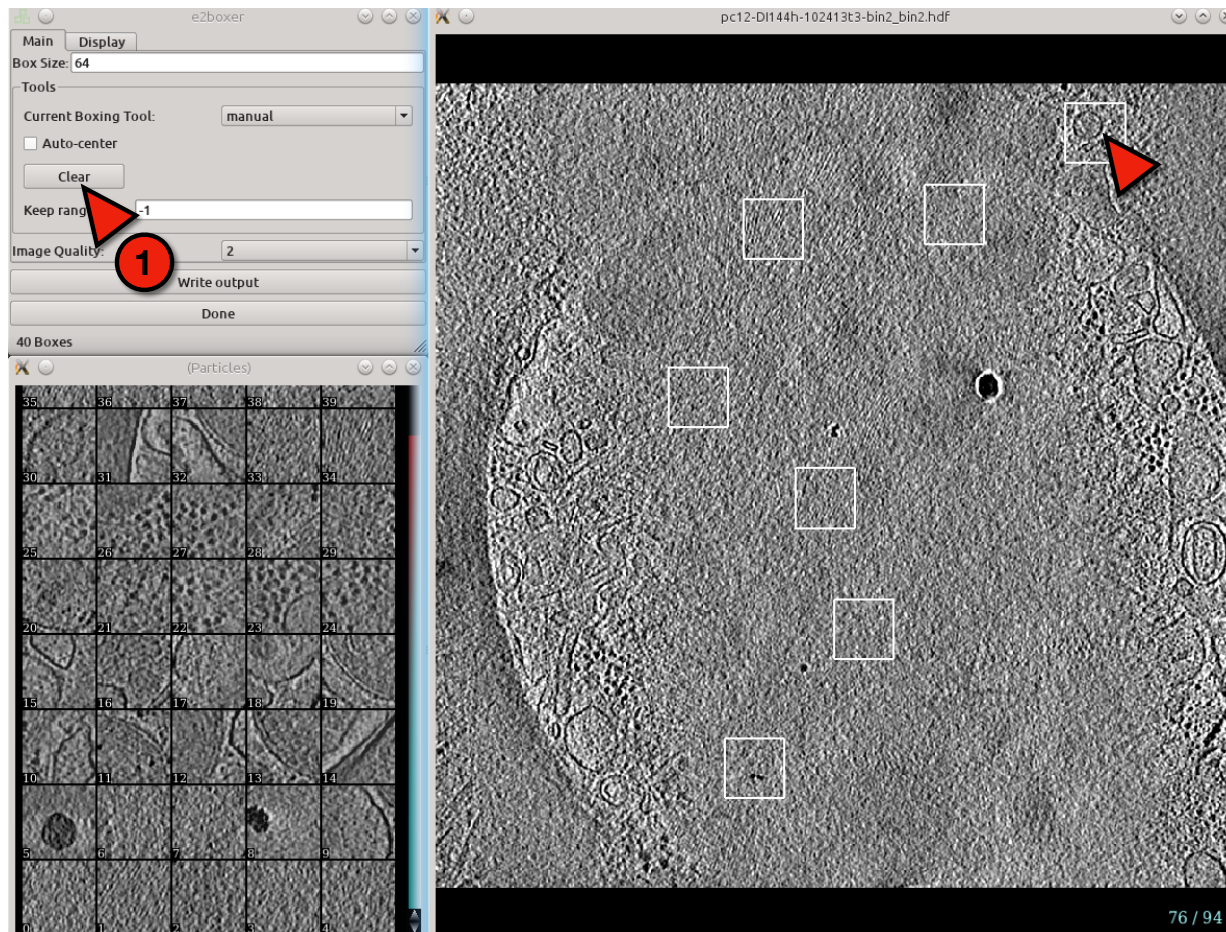


- Clear existing boxes first
- Box ~10 times more negative references than positive ones
- Cover background and non-interest features

100 negative references used

# Box training references
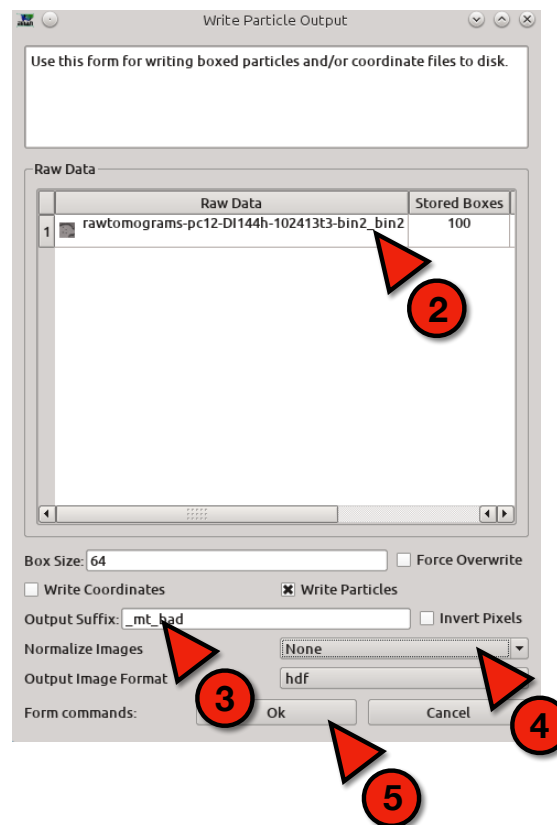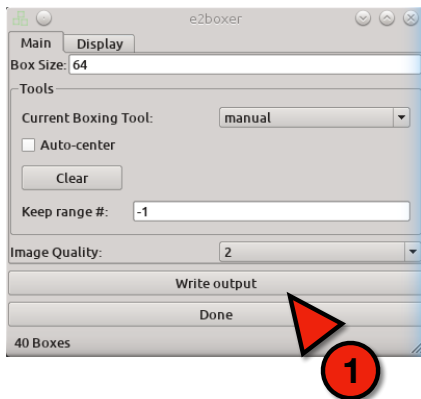


- Use a different **Output Suffix**
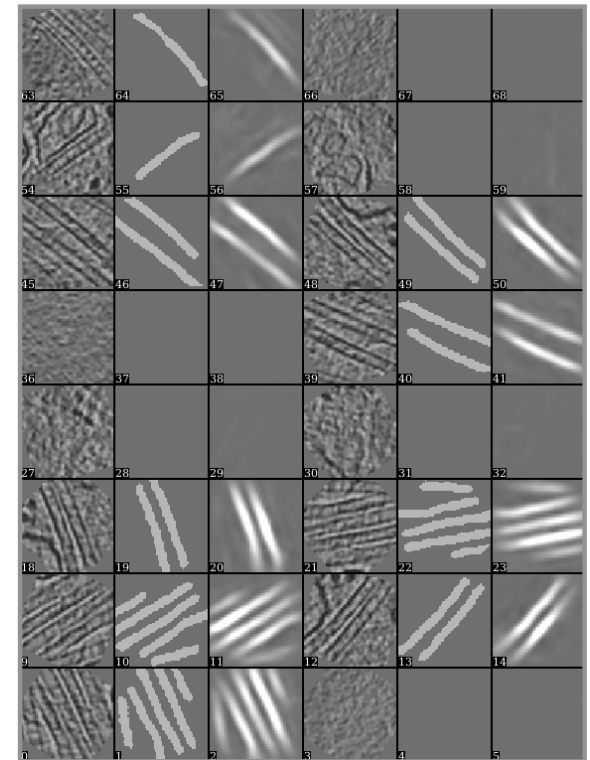
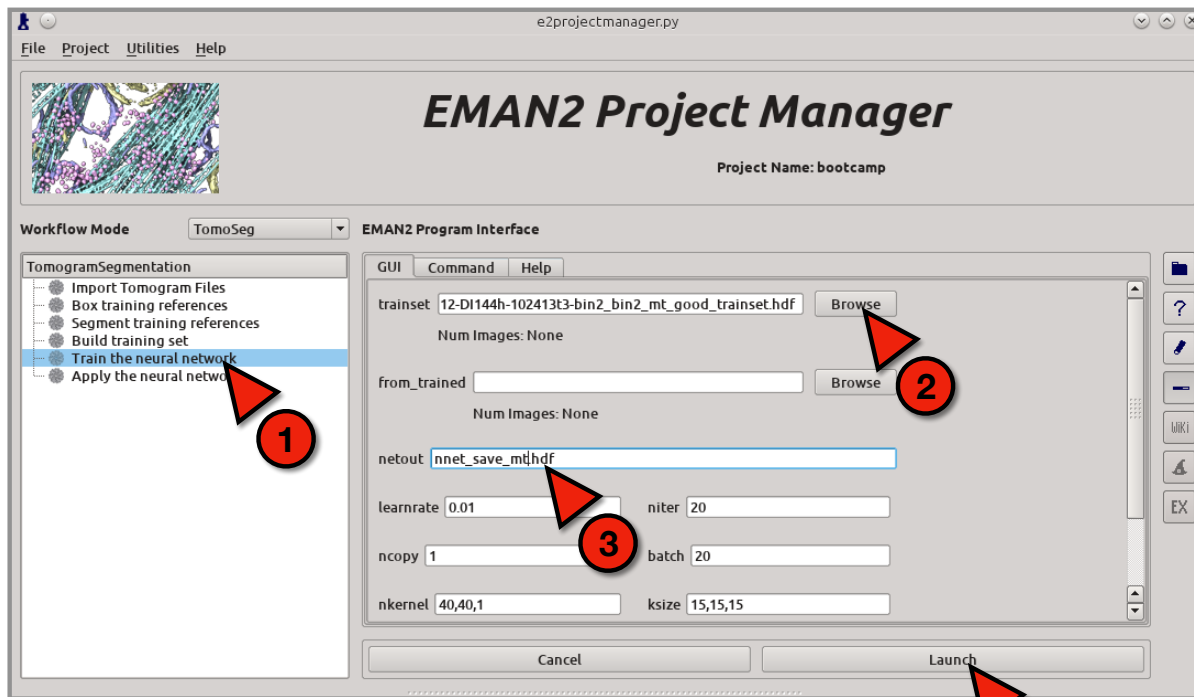This will generate **particles/xxx_mt_bad.hdf**

# Build training set
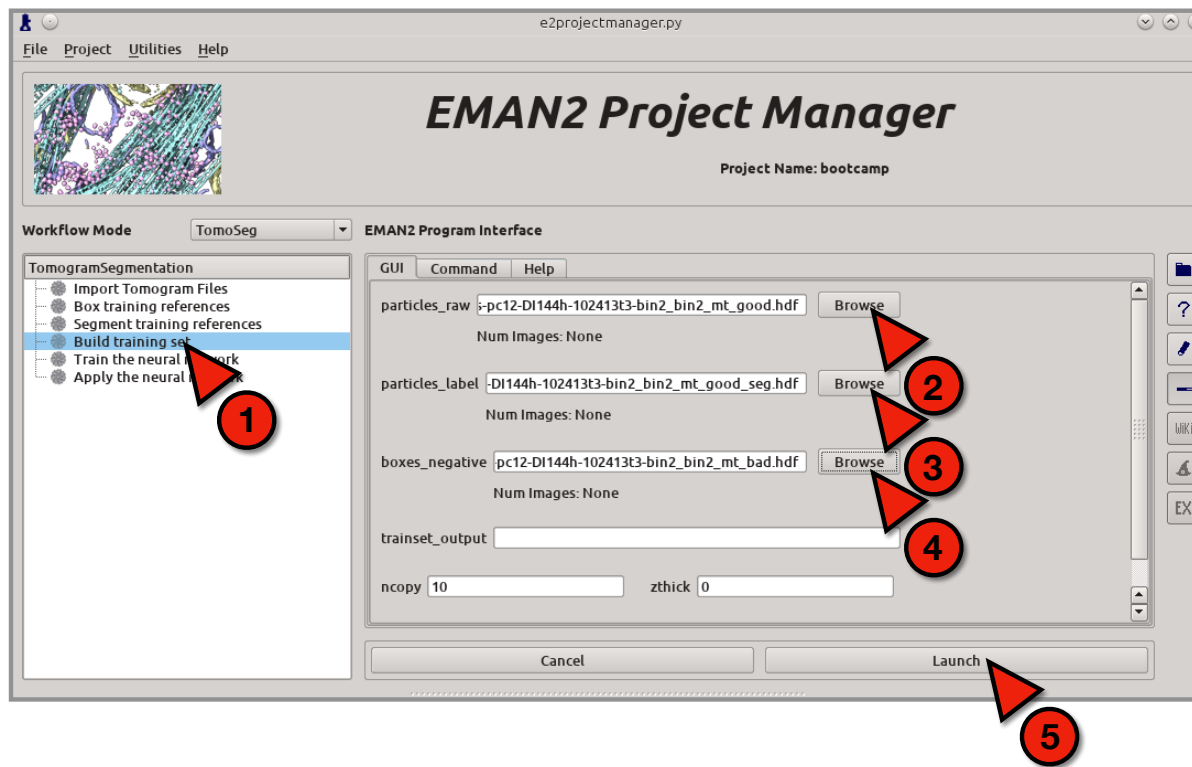


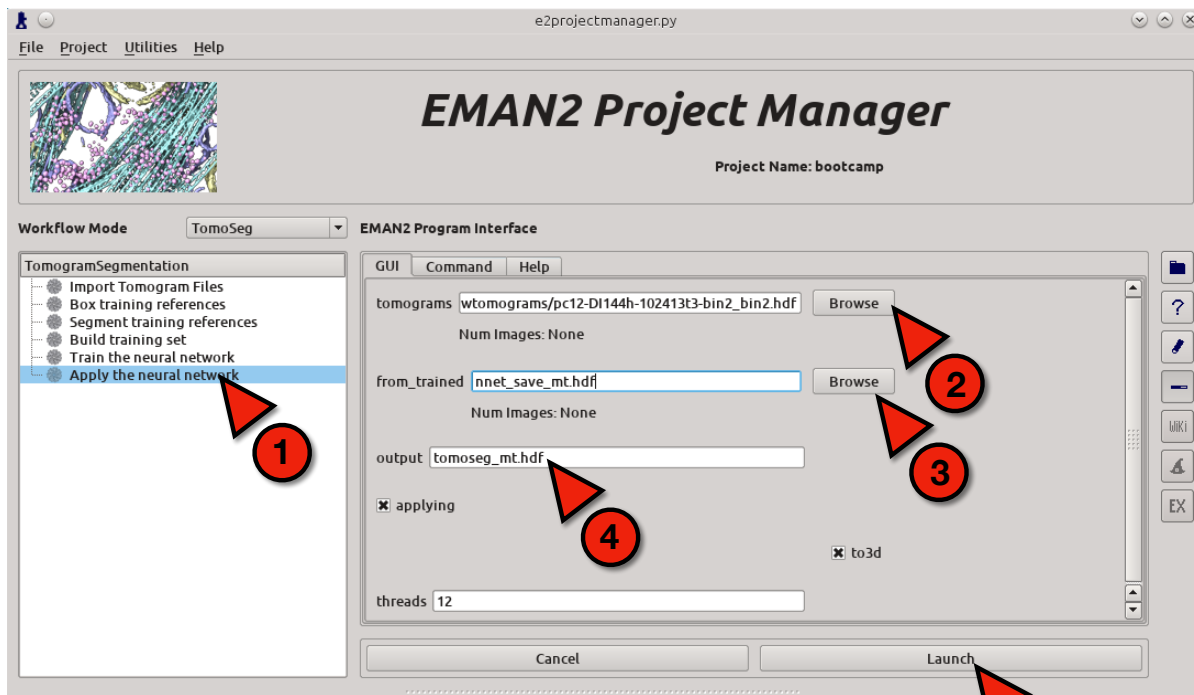This will generate **particles/xxx_mt_good_trainset.hdf**

# Train neural network

trainout_nnet_save_mt.hdf



This will generate **nnet_save_mt.hdf** and **trainout_nnet_save_mt.hdf**

# Apply to tomogram



This will generate **tomoseg_mt.hdf**

- **from_trained** should be the same as **netout** in **Train neural network** section (NOT the trainout_xx file!)
- Set **threads** to the number of cores on your computer
- Make sure there is enough memory to open two tomograms

**This may take ~30min on a laptop… It is recommended to copy the pre-generated output files to your working folder.**

**OR**

**Clip and test on a small region using:**
e2proc3d.py  rawtomograms/xxx_bin2.hdf rawtomograms/xxx_bin2_clip.hdf --clip 480,480,24
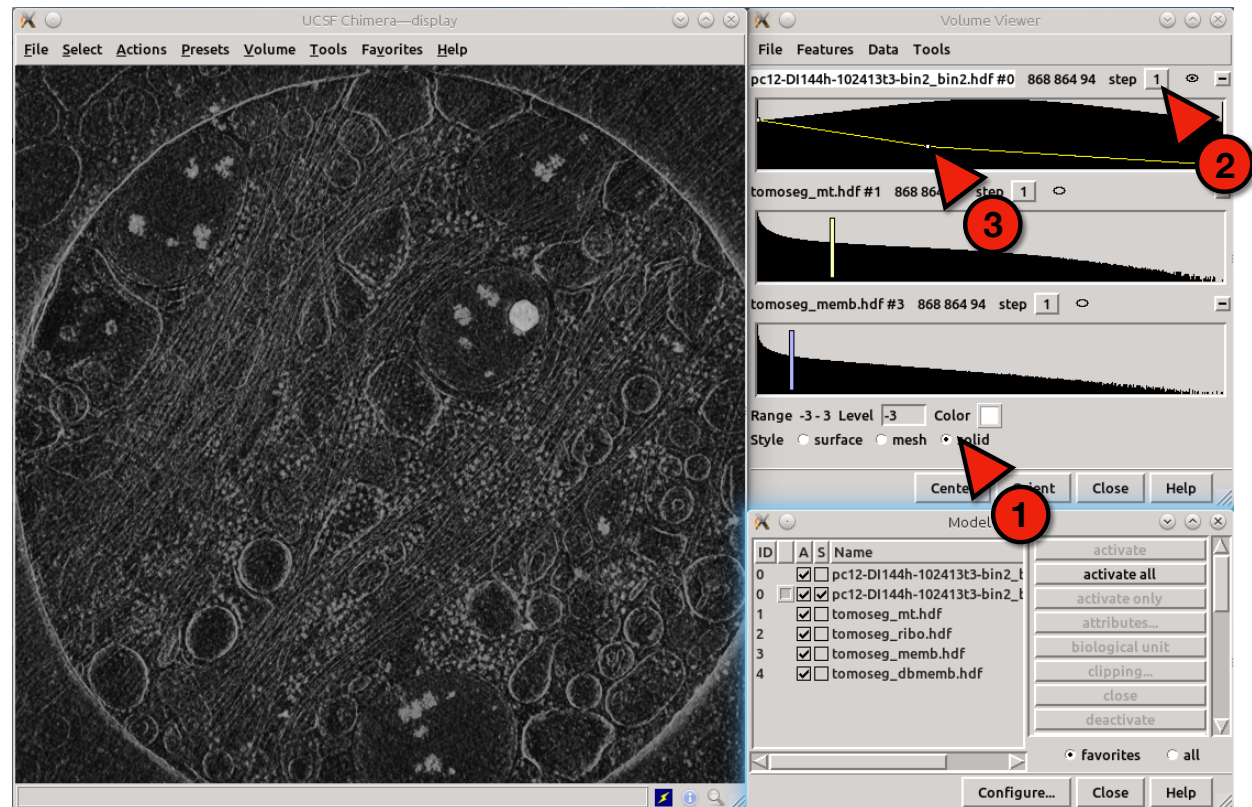
**This will take ~2min…**
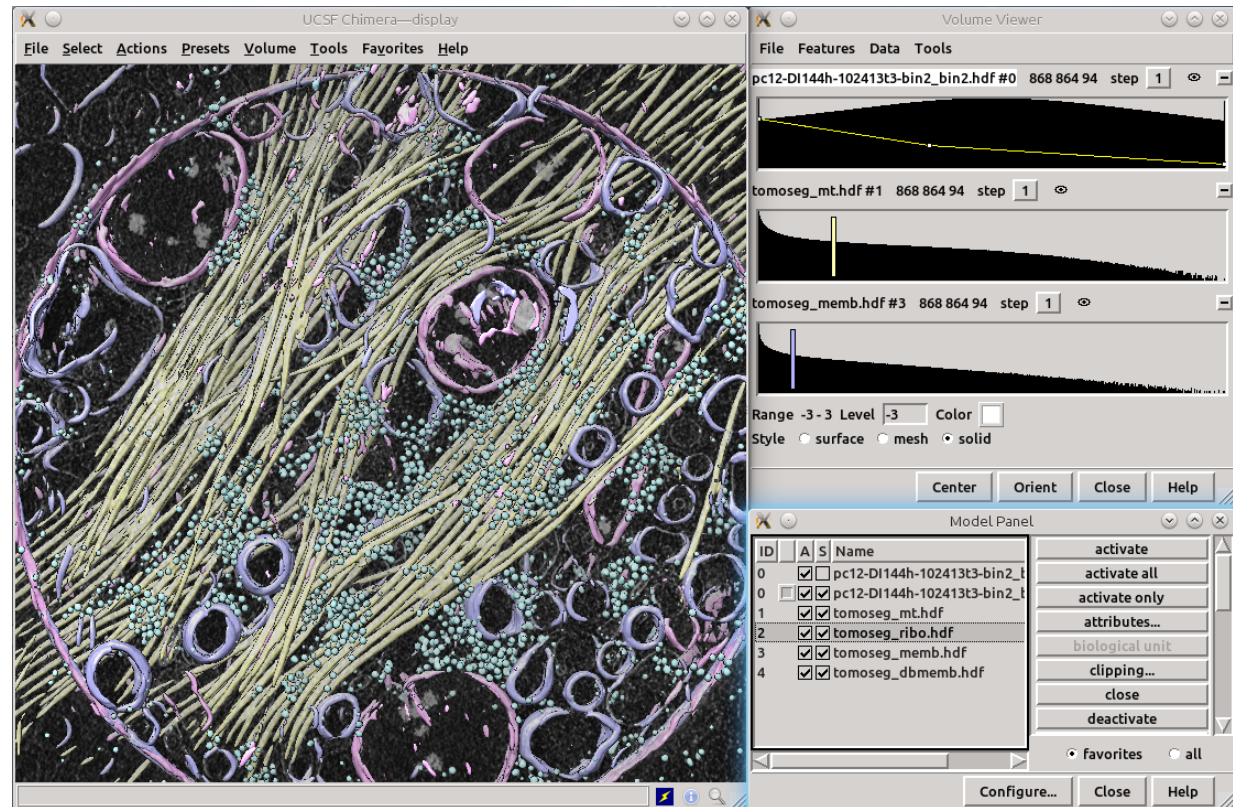
# Repeat previous steps for each feature of interest

# Visualize results in UCSF Chimera

- Open **rawtomograms/xxx_bin2.hdf** in Chimera
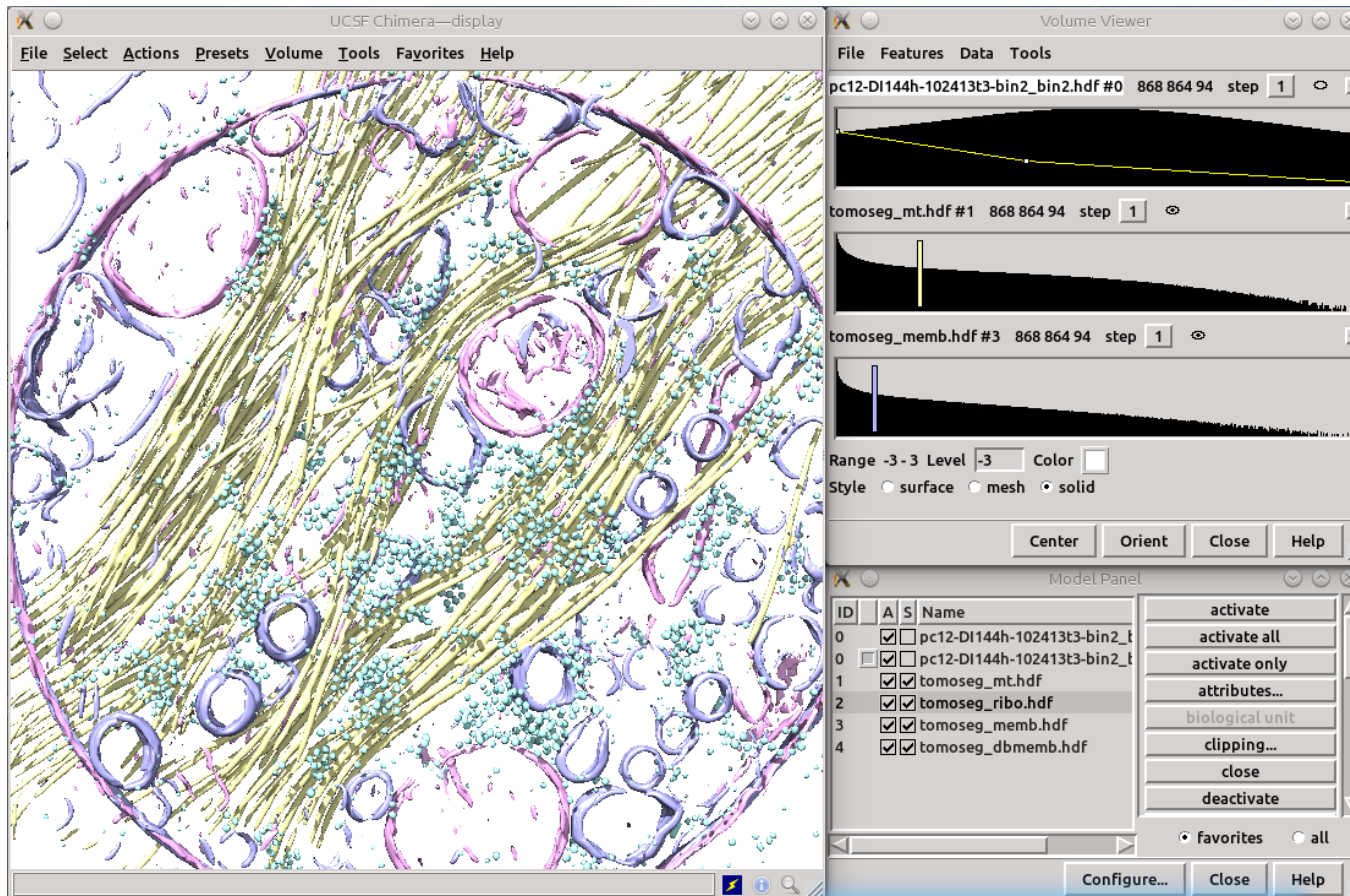- Set up volume rendering
- Set step to 1
- Invert contrast

# Visualize results in UCSF Chimera

- Open **tomoseg_xx.hdf** in Chimera
- Set isosurface threshold (1 should be fine)
- Set step to 1

# Visualize results in UCSF Chimera

# Mask density map with annotation

Make contrast inverted tomogram

      **e2proc3d.py rawtomograms/xxx_bin2_clip.hdf rawtomograms/xxx_bin2_clip_inv.hdf —mult -1 —process normalize**

Mask tomogram with annotations

 **e2proc3d.py tomoseg_mt.hdf density_mt.hdf —process threshold.binary:value=1 —multfile rawtomograms/xxx_bin2_clip_inv.hdf**
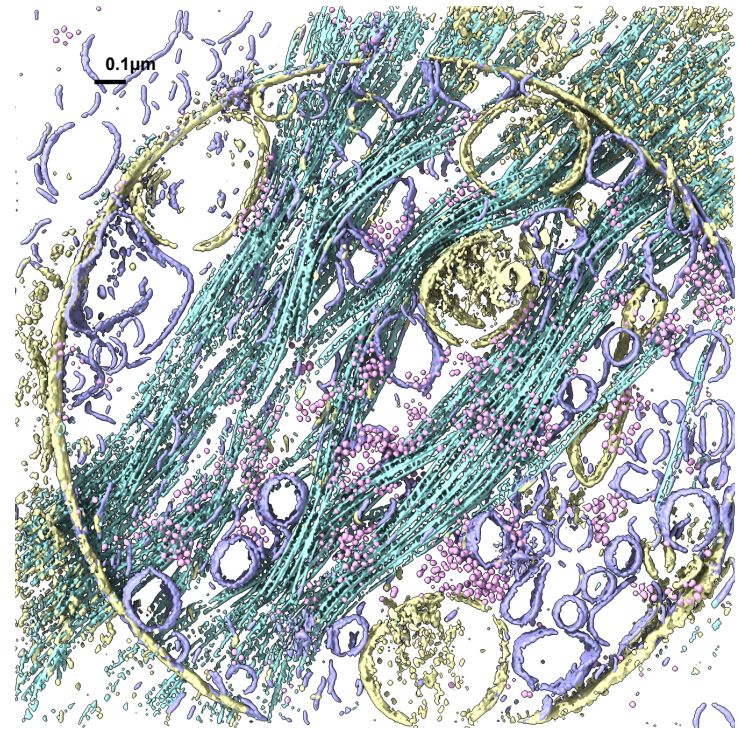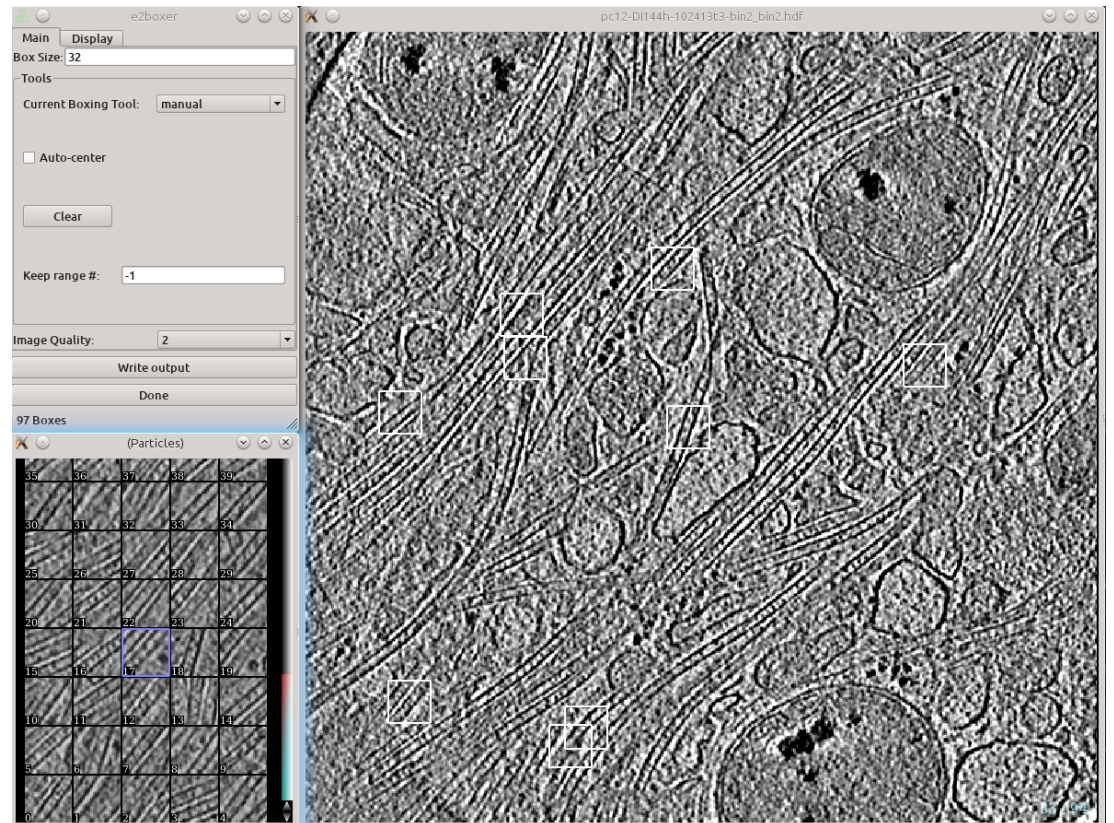
# Next steps

- Annotation on other tomograms

- CTF correction

- Subtomogram averaging

- Statistical analysis

# Subtomogram averaging from annotation

Add EMAN2/examples to your path before proceeding…

# Extract subtomograms from annotation

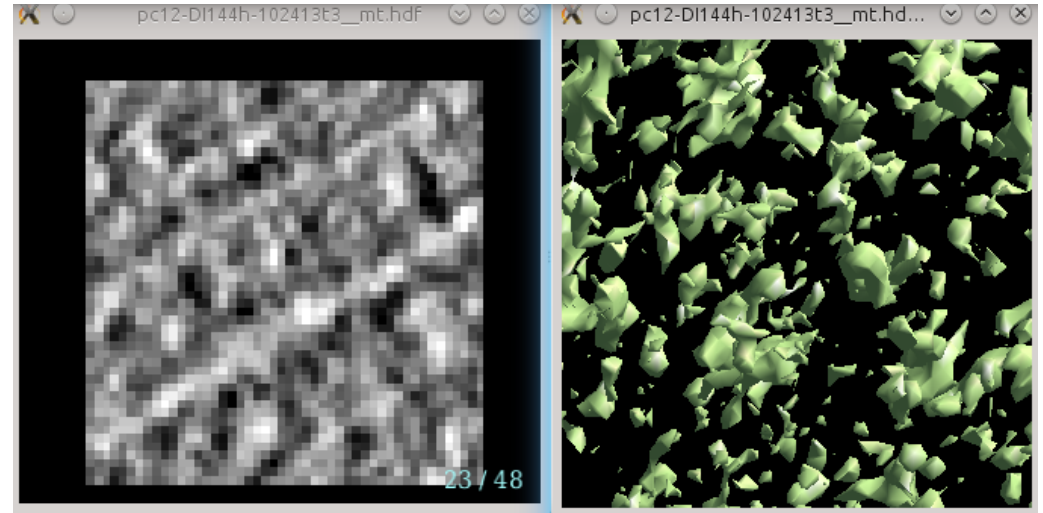**extractptclfromseg.py tomoseg_mt.hdf rawtomograms/xxx_bin2.hdf --random 100 --edge 32 --thresh 1.5**

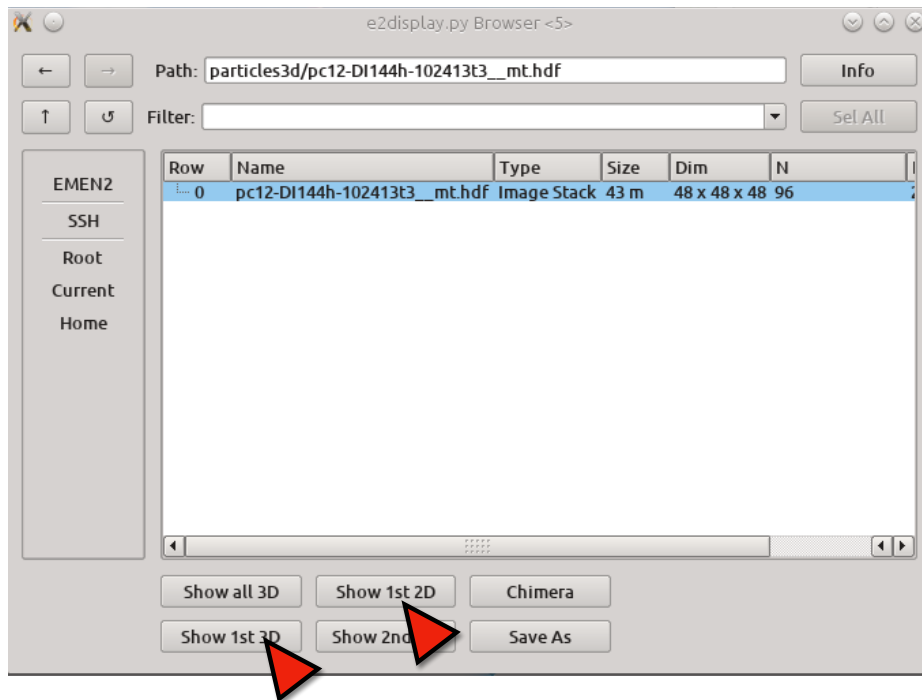

Visualize boxes using **Box training references** or
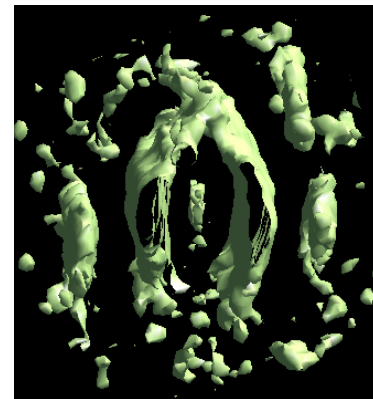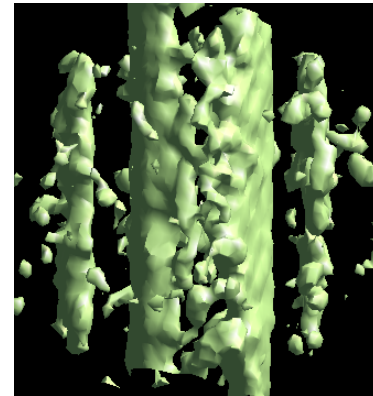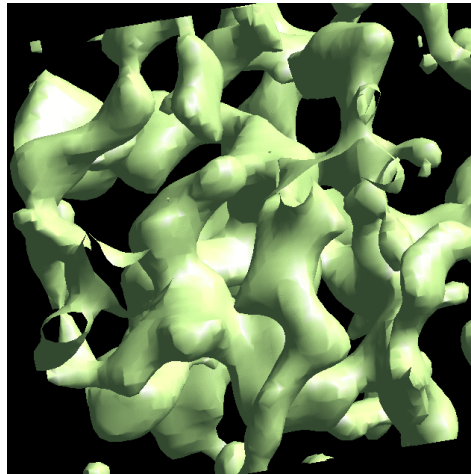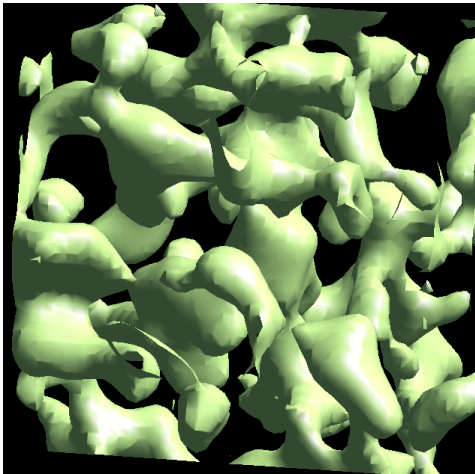**e2spt_boxer.py**

# Generate particles

**extractptclfromseg.py rec/xxx.rec rawtomograms/xxx_bin2.hdf --genptcls particles3d/xxx__mt.hdf --boxsz 24**



Take a look at the first particle and make sure it is correct

# Initial model for subtomogram averaging

**e2spt_sgd.py particles3d/xxx__mt.hdf —path spt_mt/ —niter 5**

# Subtomogram averaging

- e2spt_align.py

- e2spt_average.py

- Results depend on data quality…