EMAN2.91+ Particle Picking Tutorial

Using e2boxer

This tutorial was updated in April, 2021. It should not be used with versions of EMAN2 older than EMAN2.91.

- ➡ Boxes like this one will contain additional information and tips.
- ➡ The main source for EMAN2 documentation is the Wiki at: http://eman2.org . There is also a Google Group for support and discussions: http://groups.google.com/group/eman2 .
- ➡ **GUI Tips:** EMAN2 will work best with a 3-button scroll mouse, though there are alternatives using keyboard modifiers for one button mice or trackpads on Macs.
  - In most EMAN2 windows (2-D images, 3-D volumes, plots, etc.), the middle mouse button will open a control panel for the widget, which is different for each widget type
  - The right mouse button is used for panning in 2-D or 3-D image windows, and can be used to zoom (by shift+dragging), and to reset the zoom (clicking) in plot windows.
  - The scroll-wheel will generally act as a zoom. control-panel for more precise control
  - If you have a one button mouse or trackpad, holding down the alt/option key combined with a mouse click will serve the same role as a middle-click.
- ➡ Text you see in *italics* will generally refer to labels in the GUI, such as buttons to press. Text you see in **bold**, are commands to be typed in. Items like: <param> are parameters you should fill in (without the < >). Items like: [param] are optional parameters (without the [ ]).
- ➡ Check your version: The command e2version.py will tell you exactly what version of EMAN you are using. When reporting bugs or asking questions on the mailing list it is critical to include all of the lines of the output of this program with your question.

Introduction to the Tutorial

The standard single particle analysis tutorial (http://eman2.org/Tutorials) makes use of pre-picked particle locations, and skips the particle picking process. The main reason for this is that doing a good job at particle picking can take a fair bit of time to learn, particularly for beginners, so in our typical workshop setting, there simply isn't enough time to tackle this issue. This short tutorial covers this topic in more depth.

One perspective on particle picking is that it isn't important, and even if you select a bunch of bad particles, you can just sort it out later through either 2-D classification or 3-D quality control measures. While it is absolutely true that it is possible to complete a good single particle reconstruction even with a sloppy job of particle picking, it is also true (we've tried it on several projects) that you will get a better, cleaner-looking structure if you pick your particles carefully. Unfortunately automated or semi-automated methods for identifying and removing bad particles aren't perfect. They will help (sometimes a lot), but they will still leave a fraction of bad particles in the data set, and even if the resolution isn't impacted, the quality of the structure will be.

Tutorial

1. Set up project

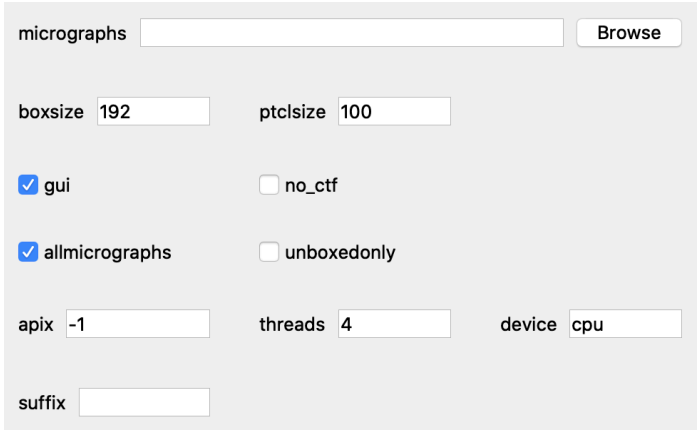
While you may use any data you like while learning the particle picker, the tutorial will be based on the standard single particle reconstruction tutorial data set. If you follow that tutorial, you will need to go through step 3 before you are ready to start this tutorial.

If working with your own data or the tutorial, you must have an EMAN2 project folder. That folder must already contain an *info/* folder and a *micrographs/* folder (containing the imported micrographs). If you have followed the standard tutorial past step 3, you will also have other folders containing the results of the refinement.

- Run **e2projectmanager.py** from the project folder

2. Launch particle picker

- *Particles → Interactive Particle Picking*
- *Launch*



➡ The box size and particle size can be changed within the GUI, but give it an initial guess at least. Note that both sizes are in pixels, not Å or nm.

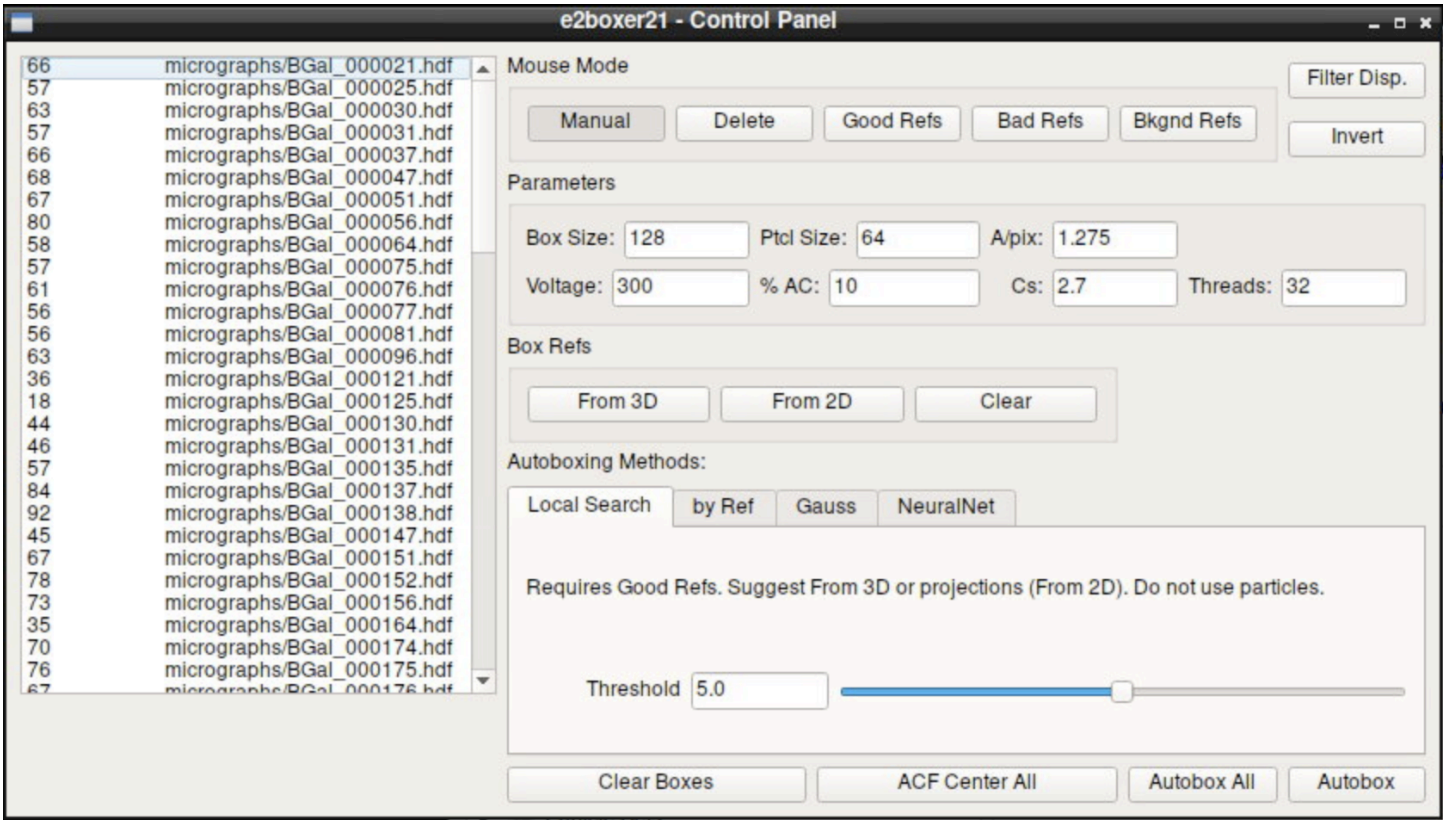
➡ apix of -1 will cause it to use the apix value from the project.

➡ *device = cpu* is for the TensorFlow based particle picker. If you are running on a Linux box with CUDA configured/installed for EMAN2, then *device = gpu* instead. On Macs, CUDA is no longer supported, and you must use *cpu*.

➡ threads is used only when autopicking from multiple images. Normally this should be the number of physical cores on your computer, not the number of 'hyperthreads'.

3. e2boxer graphical interface

After pressing Launch, you should see three windows appear. One will be empty, one should contain the image of your first micrograph, and the third will be the control-panel for the program:



On the left you should see the list of all of the micrographs in your project. The number 0 next to each indicates how many particles have currently been selected from that micrograph.

- ➡ If you are using a project folder where you have already completed the standard tutorial beyond step 3, you will see non-zero numbers, as particles have already been picked and imported into the project. While you can click the *clear boxes* button in the control panel to erase these boxes for each micrograph individually. It may be easier to start out with a fresh project folder and only go through step 3.
- ➡ The micrograph display is a standard single-image display window. Middle-clicking will open a control panel, the mouse-wheel should zoom, etc. If you are unfamiliar with this, you should read the Quick GUI Intro tutorial, or watch the similar YouTube video.

The *Filter Disp* button in the upper right corner is a toggle, which will apply a filter to the visual display of the whole micrograph to make particles easier to see and contamination easier to identify. This is just a filter on the visual display. When particles are extracted from the micrograph or particles are automatically located, this button has no effect. Similarly *Invert* can be used to invert the contrast of the display, which helps some users better visualize particles.

The *Mouse Mode* section of the control panel determines what the left mouse button will do in the micrograph window. In the default *Manual* mode:

- clicking will select a new particle
- clicking and dragging on an existing particle will let you reposition it
- holding down shift and clicking will delete an existing particle
- Manually boxed particles will appear as a blue box with a circle inside it. The size of the circle is the specified *Ptcl Size*. The selected particle will also appear in the *Particles* window which started empty when e2boxer was launched. You may also delete particles from that window by holding down shift and clicking.

The *Delete* mode is similar to holding shift in *Manual* mode.

Good Refs, *Bad Refs* and *Bkgnd Refs* allow you to select specific types of reference particles for use by the autopicking algorithms (explained below). When any of these three modes is used, a new window for that mode will appear, and any clicked-upon particles will be added to that window. The reference particles picked in this way are not specific to one micrograph. If you select a different micrograph, the particles in these windows will remain the same, and *selected reference boxes are not marked in the original micrograph*.

To delete a reference particle, hold down shift and click on the particle in the corresponding reference window (not on the micrograph). Again, no visual indicators of reference boxes will appear on the micrograph, because the reference boxes may come from many different micrographs.

The *Parameters* section should be fairly self-explanatory. The CTF parameters aren't used at present. This is handled in a separate process. So, the only 2 meaningful parameters are A/pix and threads.

The Box Refs section provides a mechanism for generating *Good Refs*, either from a 3-D volume (it will generate a set of projections in all different orientations) or from a set of 2-D particle images or class-averages. The *Clear* button allows you to clear all current reference particles.

Finally, in the bottom section of the window we have the controls for the various autoboxing algorithms. This version of EMAN2 has 4 different autopicking algorithms: *Local Search, by Ref, Gauss* and *NeuralNet*. These algorithms are completely separate from one another, and if you perform automatic picking, whichever of these three is currently highlighted will be the algorithm which is used. This list may be expanded in future versions of EMAN2. We have had some discussions about including external methods as options here, but have not yet completed this.

Each of the different automatic picking algorithms has a brief description in the box explaining the requirements for using it. Each algorithm has different requirements for the type and quantity of reference particles required. NeuralNet is the most accurate picker, by far, but is also the most time consuming to configure and use. The NeuralNet picker has its own tutorial available on eman2.org.

Here is a brief summary of the 4 pickers, but it is likely that you will need to experiment with them to really understand the advantages and disadvantages of each:

- By Ref - This is a classic reference based particle picker. To use it, you need to have high quality "good references" in all possible 3-D orientations. The algorithm will do in-plane rotation of the references and cross-correlate to look for peaks. It is recommended that you use projections of a 3-D map (low resolution) as references.
- Local Search - This algorithm is similar to By Ref, but this algorithm works very differently. It heavily downsamples the particles and references, and actually performs a 2-D alignment of each putative particle to each reference to identify the best particles in the image. In theory this should produce fewer false positives. Reference requirements are similar to By Ref, though a smaller number of references may be fine.
- Gauss - This is a simple and fast reference-free picker, which provides a simple solution for easy particle picking cases, where the particles have good contrast and are monodisperse. This may work well for things like viruses or ribosomes. It is very fast and since it requires no references, it's easy to try. It likely won't work well for most projects. It was ported from the old boxer program by a volunteer (Vadim Kotov).
- NeuralNet - This is the most accurate boxer by far, both in terms of false positives and false negatives, when trained properly. It has its own tutorial (linked above). It is based on a pair of neural networks, one to discriminate between putative particles and the background, and a second to discriminate between real particles and contamination or other high contrast non-particles, which is why it has two thresholds.

Note that once you have completed selecting particles, you do not need to save your box locations to files manually. The box locations are continuously stored in the appropriate .json files in the info/ folder in your project as you use the program. There is no need to generate '.box' files for all of your micrographs for EMAN. The reason we provide that capability is for use with other software packages requiring particle locations. When you finish locating particles, you can simply exit the program, and then use the "generate output" tool in e2projectmanager to actually extract the windowed particles from the micrographs and produce particle stack files.