AutoEMage user manual

CHENG Yuanhao
Version 1.0.0
Contact: chengyuanhao@iphy.ac.cn

July 14, 2023

Contents

1 Introduction

AutoEMage is a software that automates file transfer, motion correction, CTF estimation, image display, outlier detection, particle picking and 2D class averaging during cryo-EM data acquisition. Through specific settings, file transfer, data preprocessing and monitoring can be executed simultaneously with data acquisition. In this process, AutoEMage does not require manual intervention or surveillance, which lives up to its name. Once outliers are detected, it will send emails to users so that they can come back in time and have a check. Moreover, it provides users with a graphical user interface, in order that they can not only see real-time preprocessed micrographs, but also some results of data processing such as 2D classes.

2 System Requirements and Installation

AutoEMage is developed using Python3 and Perl on Linux, including a few programs such as MotionCor2, CTFFIND4 and some programs of RELION and IMOD, while its graphical user interface is based on PyQt6. Therefore, these programming languages and cryo-EM programs should be installed before using AutoEMage, which can be downloaded through the following links:

- PyQt6: https://pypi.org/project/PyQt6/ Note that PyQt6 may be successfully installed by inputting "python3 -m pip install PyQt6" on the command line.
- Python3: https://www.python.org/
- Perl: https://www.perl.org/
- MotionCor2: https://hpc.nih.gov/apps/MotionCor2.html
- CTFFIND4: https://grigoriefflab.umassmed.edu/ctffind4
- RELION: https://relion.readthedocs.io/en/latest/Installation.html
- IMOD: https://bio3d.colorado.edu/imod/download.html

Warning: MotionCor2 supports cuda of different versions, but RELION does not support cuda 12.1.
In addition, some libraries such as PyQt6-WebEngine, matplotlib and mrcfile need manual installation. After successful installation, directories of MotionCor2, CTFFIND4 and AutoEMage can be added to environment variables in this way: add the following lines to the .bashrc file:
#add MotionCor2 path
export PATH=MotionCor2 directory:$PATH
#add CTFFIND 4 path
export PATH=CTFFIND directory:$PATH
#add AutoEMage path
export PATH=AutoEMage directory:$PATH

3 Usage

3.1 Parameters Setting

Before first use, users should manually adjust several basic parameters according to the electron microscope they use. As a variety of electron microscopes show up in recent years, this adjustment cannot be achieved automatically by AutoEMage. File names of the gain file, images and directories automatically saved by the electron microscope should be modified, which corresponds to lines 75-145 in the script Auto_mv.pl.

3.2 Graphical User Interface and Functions

3.2.1 Registration and Login

After parameters setting, users can run "autoemage.py" on the command line to open the user graphical interface. Up to now, the login interface is only used for keeping a record of users' emails, in order that

Figure 1: AutoEMage login interface

AutoEMage can send emails and store data for each user respectively. Sending email alerts is one convenient function of AutoEMage whose email address is cryo_EM_2022@163.com. To make use of this function, users had better type in their current email addresses during registration. The login interface is shown in Figure 1.

Following successful login, the main interface shows up (Figure 2). During data acquisition, the left panel will display several physical quantities varying with time and the right panel will exhibit plots of "preprocessed micrograph–power spectrum–drift path–CTF fit".

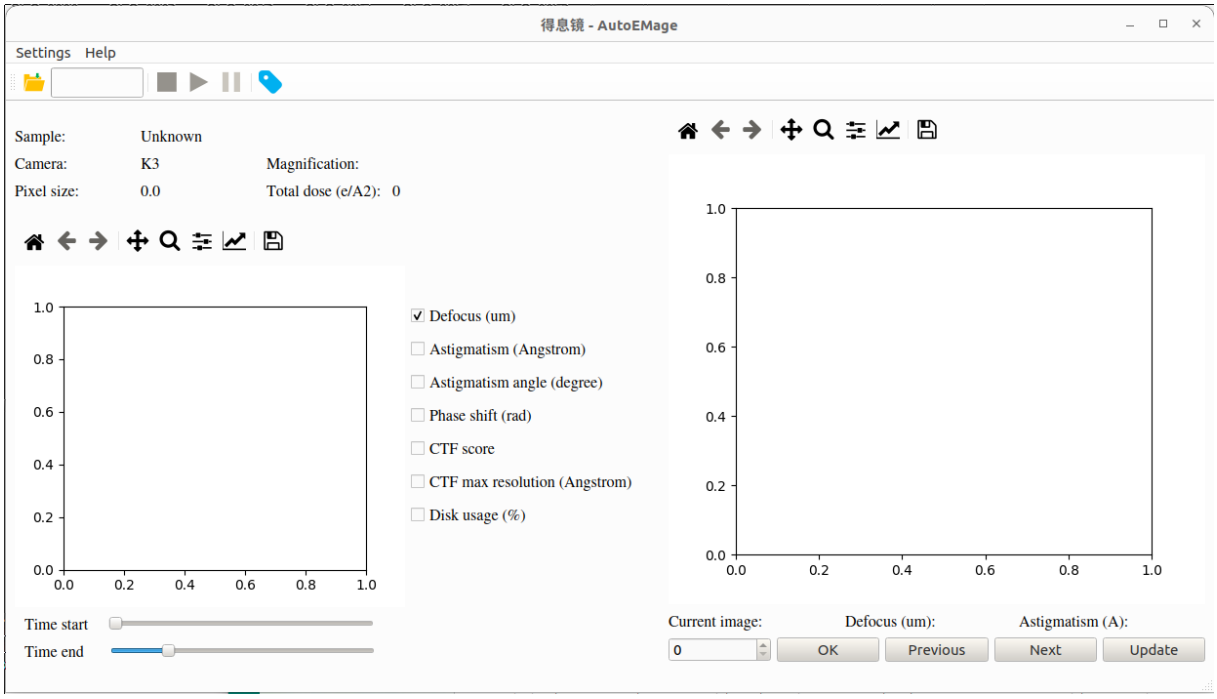

Figure 2: The main GUI of AutoEMage

3.2.2 File Transfer

In the main interface, users can click the "File Transfer" button at the top left corner to open the window to input some necessary experimental values, as shown in Figure 3. The EM directory refers to the directory

where the electron microscope automatically saves images, and the destination directory is the directory of users' mobile disk drive or any other directory to store data. The job name should contain no space. After inputting necessary values, users can start automatic file transfer by clicking "OK", while the progress will update in the progress bar. As the file transfer goes on, gain correction, motion correction and CTF calculation will simultaneously begin and the results will show up once finished, as shown in Figure 4.

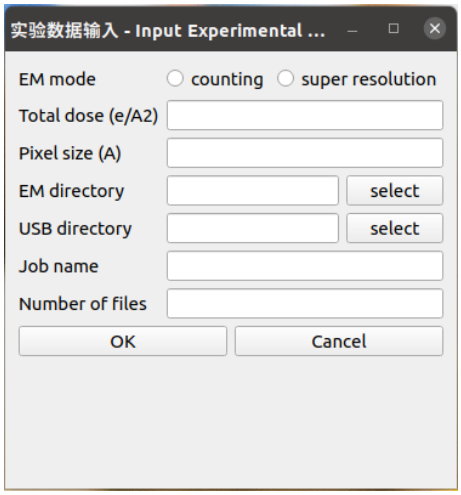

Figure 3: The input UI of experimental parameters

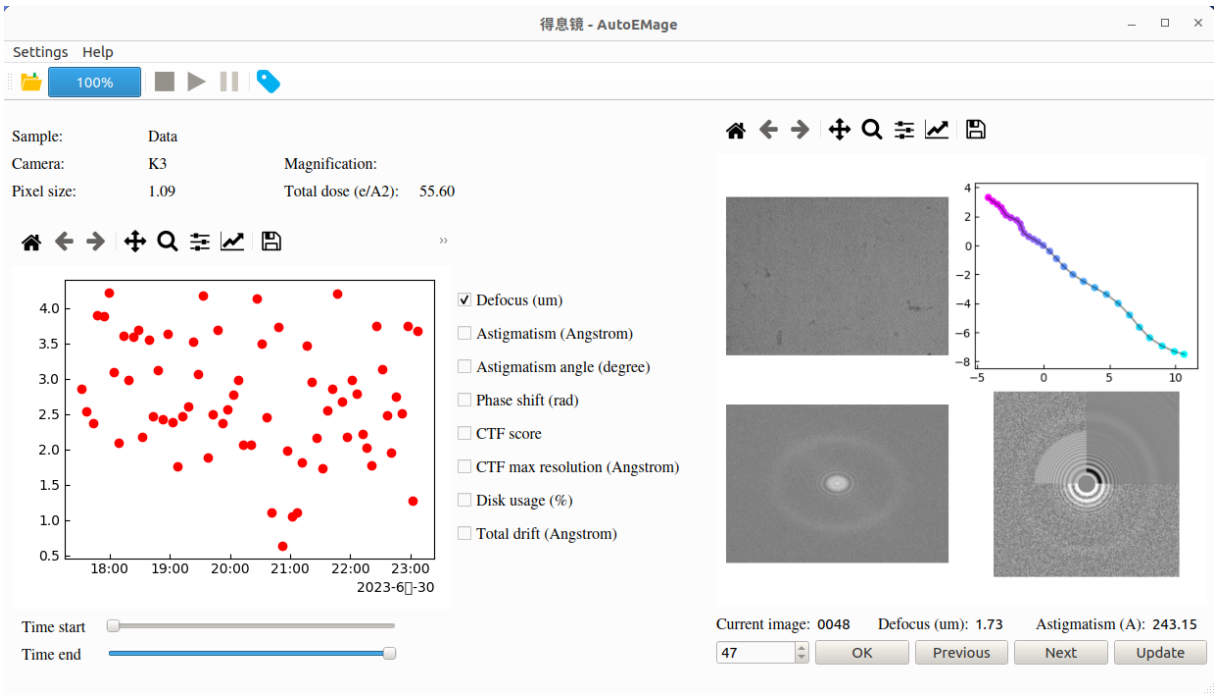

Figure 4: The main GUI of AutoEMage during data acquisition

3.2.3 Variables Display

The left panel displays a few physical variables changing with time during data acquisition. Time sliders below the graph are used for displaying variables in a specific time slot. As thousands of images are usually collected for a project, thousands of data points will be produced. If all of them are plotted on a single graph, the graph

will be too crowded to exhibit detailed variations. The first time slider controls the start of the time slot, and the second slider determines the end of the time slot.

3.2.4 Autopicking and 2D class averaging

To use the function of autopicking and 2D class averaging, users just need to input the smallest diameter and the largest diameter of target particles, as shown in Figure 5. Autopicking and 2D class averaging is based on programs of RELION, such as relion_autopick, relion_preprocess and relion_refine. If users are familiar with RELION, then they can modify the corresponding scripts according to their samples to make the best of these programs. They are lines 192-242 in the script autoemage_threads.py.

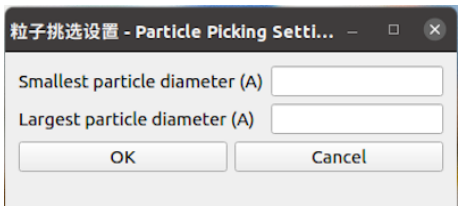

Figure 5: The input UI of particle picking

3.2.5 Outliers Detection and Email Reminder

Data collection usually stores thousands of images, so outliers are inevitable. When outliers appear frequently within a short amount of time, for instance 10 out of 50 images, AutoEMage will send an email to users so that they can come back to check the EM settings as well as their samples, as shown in Figure 6.

Up to now, AutoEMage can detect two kinds of outliers: images with large drift and images with low CTF scores, as shown in Figure 7. These images are unlikely to contribute to the high-resolution reconstruction.

When the task is finished or the hard disk drive is nearly full, AutoEMage will also send a reminding email to users.

3.3 Incomplete Functions

AutoEMage is an open source software for academic users, so its biggest advantage is the freedom that users have to modify and add any function they need. As this software was used by different groups in their own labs, they have left their special scripts that have witnessed the development of this software as well as the ever growing requirements of a practical software in the field of cryo-EM. Here I point out these special codes for users' reference. For example, lines 242-340 in the script Auto_ctf_find.pl and the script Stigma_check_hxj_v1.pl. These are also reflected in "EM settings" located in the toolbar of the main interface.

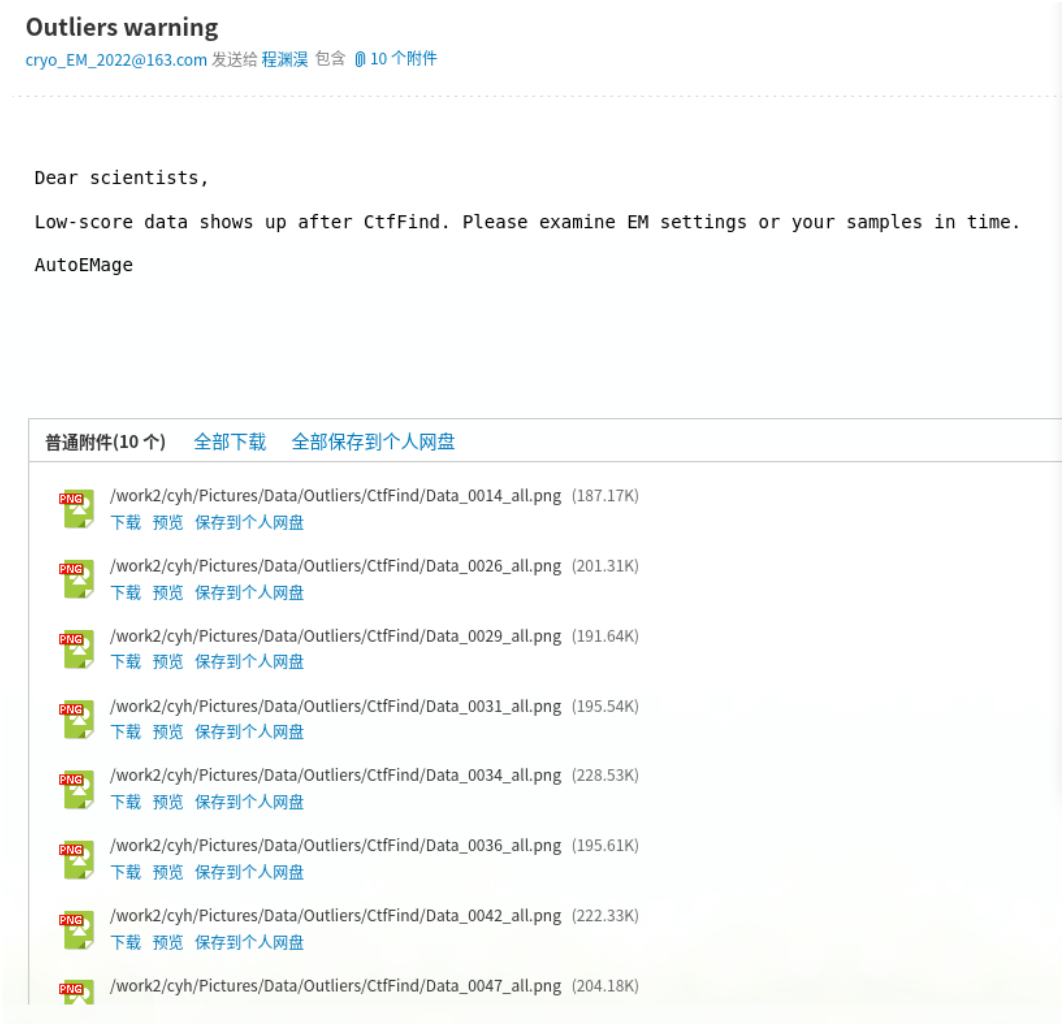

Figure 6: Outliers email

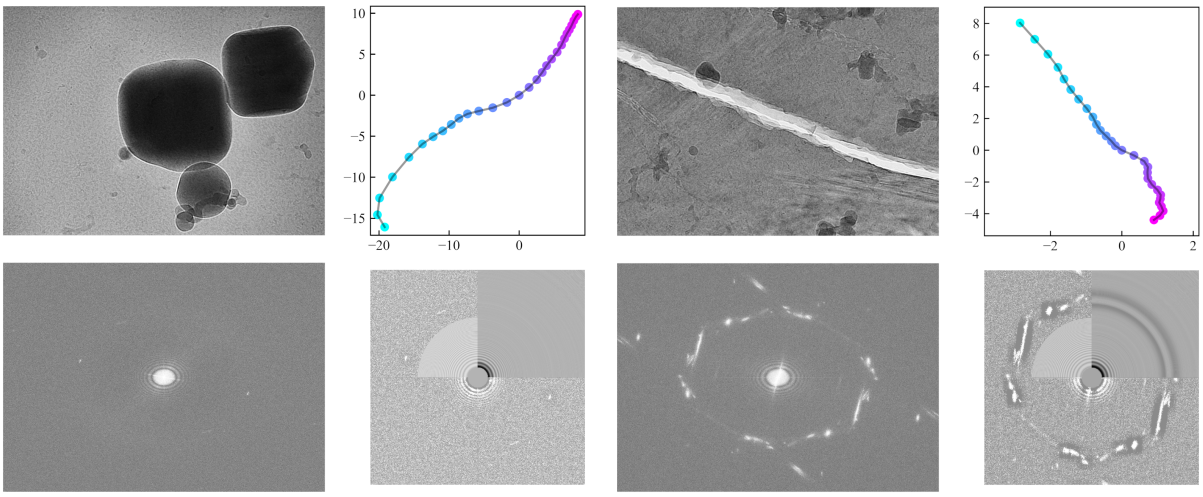

Figure 7: Two kinds of outliers. On the left is an image with large drift. On the right is an image with a low CTF score.