Contents

Overview

This project extends WARP to add several functions that I found useful for downstream processing of frame series and tilt series.

This is a prototype. This prototype has been useful for me to experiment with which new functions can be implemented and how. This means, the code works but is not easy to read and requires extensive refactoring and optimization.

It is made with EPU and TOMO in mind. Some of the functions (Grid screening, Grouping particles based on AFIS) are made to work with EPU.

It is built on top of the latest version of the open-source code of WARP. Therefore, the developments introduced in the so-called “nightly patch” of WARP or the WARP 1.1 beta are not available because they are not open source.

The data processing code of the original WARP has not been modified in any way. This means, the processing of data is exactly the same as in the vanilla WARP.

You will notice that the WARP icon has a new color (a bit closer to Karolinska Institutet’s plum). In this way you won’t get confused with the original WARP.

This project is being developed and tested at the Karolinska Institutet 3D-EM facility.

Features

- Built-in Stacker
- Support for EPU session queues
- Assign particles to different optics groups based on aberration-free image shifts (AFIS)
- Automatic 2D/3D classification in CryoSPARC
- Automatic tilt series alignment and tomogram reconstruction in AreTomo

Requirements

Install original WARP in the system (Download & Install (warpem.com))

For CryoSPARC

A Linux workstation with CryoSPARC installed, with access to the same filesystem where WARP is running. I have tested with CryoSPARC 4 up to version 4.2.1.

For AreTomo

WARP will use SSH to communicate with a Linux workstation that will run AreTomo. This workstation needs access to the same filesystem where WARP is running. For me, installing AreTomo in WSL in the same machine running WARP works great.

Setting up the Linux workstation

Generating and installing an SSH key in the CryoSPARC/AreTomo workstation

WARP needs a private SSH key file to communicate with the CryoSPARC/AreTomo Linux workstation (or WSL).

Due to limitations of the SSH.NET library, only the key format ’ed25519’ is supported in the Ubuntu 22.04 or new Fedora releases.

To generate a key in ed25519 format (Ubuntu 22.04 or newer)

ssh-keygen.exe -t ed25519

Then install the matching public key in the Linux workstation.

- In Linux: `ssh-copy-id -i ~/.ssh/id_ed25519 user@host`
- Or in Windows: `cat ~/.ssh/id_ed25519.pub | ssh user@host "cat >> ~/.ssh/authorized_keys"`

To generate a key in the RSA classic format (Ubuntu 20.04 or older)

If using RSA, the private key file must be in the *classic* format, due to the limitations in the SSH.NET library.

To generate a key in the *classic* format use:

- In Linux: `ssh-keygen -m PEM`
- Or in Windows: To generate a key in the *classic* format in Windows, use PuttyGen from the Putty package (PuTTY: a free SSH and Telnet client (greenend.org.uk))

Then install the matching public key in the Linux workstation.

- In Linux: `ssh-copy-id -i ~/.ssh/id_rsa user@host`
- Or in Windows: `cat ~/.ssh/id_rsa.pub | ssh user@host "cat >> ~/.ssh/authorized_keys"`

In case of problems, observe the advice in the "Troubleshooting" section in this site: What is ssh-copy-id? How ssh-copy-id works?

AreTomo

The commands "`newstack`" (from the IMOD package (IMOD Download (colorado.edu))), "`aretomo`" (Software | UCSF Macromolecular Structure Group), "`ffmpeg`" ("`sudo apt install ffmpeg`" in Ubuntu) and "`nvidia-smi`" should be available to run in the command line. In the case of "`aretomo`", I suggest to create the softlink "`/usr/bin/aretomo`" to whatever AreTomo executable you want to use.

Cryosparc

A regular CryoSPARC installation will work.

A note about WSL

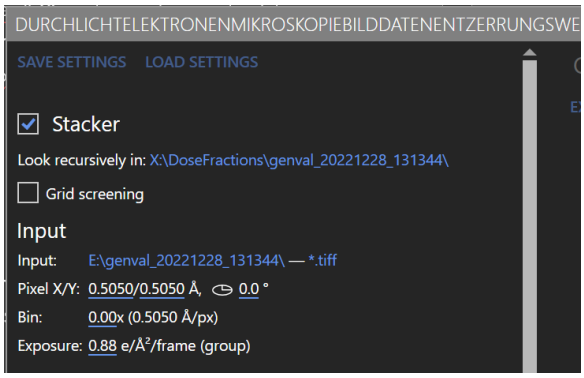
WSL2 automatically mounts the Windows host filesystems on the virtual machine, but it has very poor drive performance. This will slow down AreTomo a lot. To circumvent this issue, mount the Windows host filesystem as a network drive using samba.

Installation

No installation is required. Download the "Release" folder. Directly double click on WARP.exe to run.

Usage

Stacker



EPU places dose fraction movies into a folder hierarchy. On can use Stacker to move the movies into one directory, where WARP is working. I have used part of the Stacker2 code (GitHub - dtegunov/stacker2: Automatically compress and move EPU's MRC movies) to implement this function.

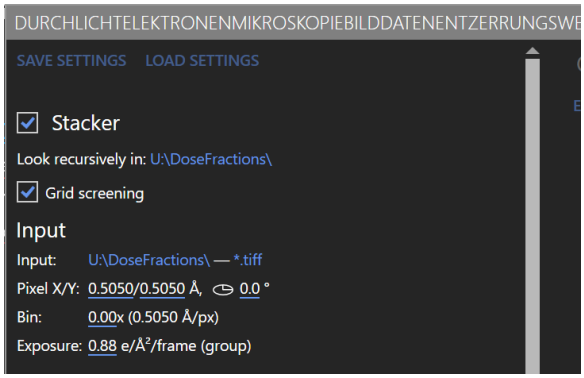
“Look recursively in:” = The directory where to look recursively for movies. The extension indicated the Input section will be used (*e.g.* *.tiff)

The movies will be moved into the “Input” folder in the Input section.

In case the “Look recursively in:” folder is the same or is contained in the target “Input” folder, WARP will automatically create a subdirectory named “WARP” inside the “Input” folder. The movies will be stacked into Input\WARP and do all the processing there to avoid conflicts.

Stacker will also move *.mdoc files.

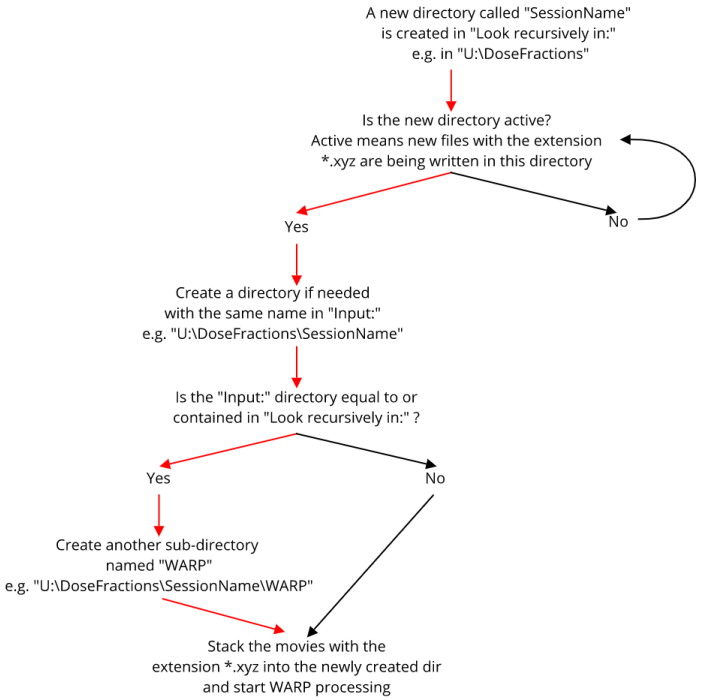
Grid screening



In EPU, one can program a “Session queue” to, for example, collect datasets from different grids automatically. If the “Grid screening box” is ticked, WARP follows the “Session queue” and switches to every new session that EPU starts. WARP will switch to a new session even if there are still unprocessed items.

This is the logic how it multi-grid works: EPU creates a new directory for a new session. WARP will detect when a new directory is created inside the “Look recursively in:” folder. Then WARP will check whether this directory is an active EPU session. This means, WARP will ask: “Are there new files with the extension *.xyz being created in this directory?” If yes, WARP will move to that directory.

The image below illustrates the how the multi-grid logic works. For the example settings above, the decision path is shown with the red lines.



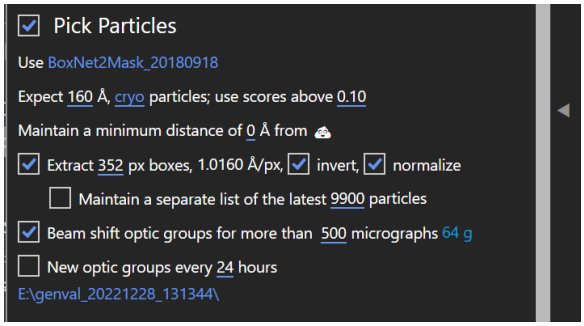
If no working directory is found when you click “Start processing”, the button will change to “Waiting”. You can still click on “Waiting” to stop WARP.

Optics groups based on AFIS

One can group particles into “Optic groups” based on the x and y beamshift of each micrograph with EPU. This only works with regularly spaced holes, such as QuantiFoil or UltraAuFoil grids.

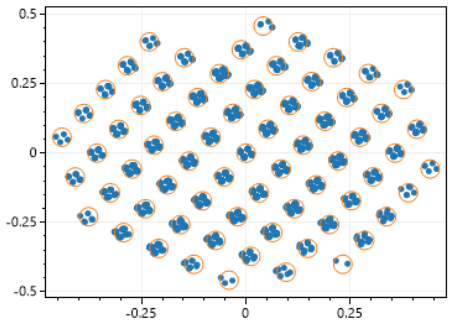
This option is available in the “Pick Particles” section as

- “Beam shift optic groups for more than xxx micrographs”
- “New optic groups every xx hours”



The beam shifts are written by EPU into XML files. WARP will look for the XML files in the directory indicated below the “New optic groups every x hours” checkbox (*e.g.* E:\genval_20221228_131344). These files are copied to a new folder in the WARP working directory called “microscopeXMLfiles”.

WARP will make one optics group for each foil hole. You can click in the number of optic groups (e.g. 64 g) to bring up an image such as this one and inspect the grouping.



Each blue dot is a micrograph, and the orange circles represent the optic groups. The x and y axis are the beam shifts extracted from the XML files.

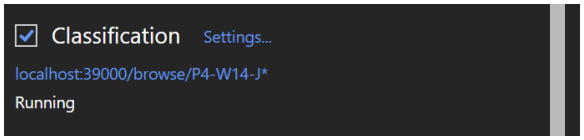
If "Grid screening" is on

The path where to look for XML files must go one directory up. For example, "E:\genval_20221228_131344" should become "E:\". Warp will add the SessionName in the background.

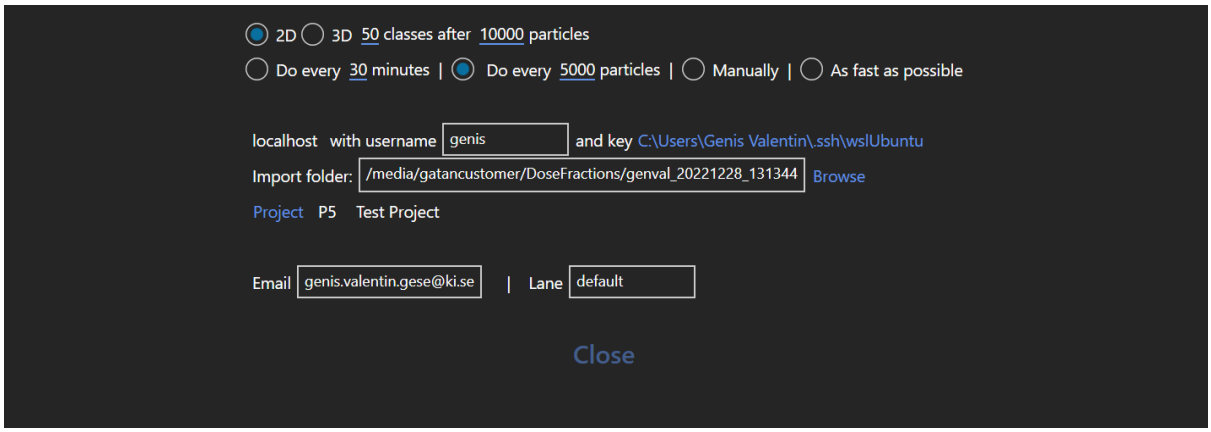
Automatic classification in CryoSPARC

WARP can periodically queue classification jobs in CryoSPARC.

The classification results are accessible by clicking on the link shown (e.g. "localhost:39000/browse/P4-W14-J*"), which will open a web browser window.



The classification task needs to be configured by clicking on Settings, that will bring up the following menu.



2D vs 3D classification

One can select between 2D or 3D (heterogeneous refinement). Classification will start only after the required number of particles has been extracted. Heterogeneous refinement requires some starting 3D volumes. WARP will look for the latest heterogeneous refinement or ab-initio reconstruction job within the project (indicated below) that matches the required number of classes, and take the starting 3D volumes from there. If no job can be found, WARP will run an ab-initio reconstruction with the number of particles indicated in the first row (*e.g.* 10000 particles).

Import folder and project selection

One needs to indicate CryoSPARC where to find the WARP processing files (*e.g.* the goodparticles star file or average/*mrc micrographs). For this, you must fill in the corresponding path in the CryoSPARC master in the "Import folder" box. Optional: If you have a SSH username and key, you can click on "Browse" to navigate the master filesystem and find the import folder. The SSH username and key can be left blank if you manually type the "Import folder".

One must also select a working Project. Click on "Project" to browse and select the CryoSPARC master project list.

WARP will create a new Workspace inside the "Project" with the name of the EPU session. If the Workspace already exists, WARP will use it.

Email and lane

Type the CryoSPARC user email and the CryoSPARC lane where to queue the classification. If the Email or Lane are incorrect, the classification will fail silently.

Setting up the global.settings

The `global.settings` file is in the WARP binary directory and contains some general configuration options. Here, `global.settings` has gained two new options:

```
<Param Name="ClassificationUrl" Value="localhost" />
<Param Name="CryosparcLicense" Value=" " />
```

"`ClassificationUrl`" is the hostname of the CryoSPARC master (such as "`localhost`" or "`master.lab.ki.se`").

The "`CryosparcLicense`" in the master node is required since CryoSPARC 4 to read the CryoSPARC database and to submit jobs.

If "Grid screening" is on

The path to the "Import folder" must go one directory up. For example, "/media/gatancustomer/DoseFractions/genval_20221228_131344" should become "/media/gatancustomer/DoseFractions". Warp will add the SessionName in the background.

Automatic tilt series alignment and reconstruction in AreTomo

The information to process a tilt series is written by TOMO as an .mdoc file.

The processing of an mdoc file involves the following steps:

- 2D processing of the individual tilt images by WARP
- Generate an un-aligned stack with "`newstack`" in IMOD
- Align the stack and generate a tomographic reconstruction in "`aretomo`"
- Write a corrected `*.mdoc` file and `*.tomostar` file (see below *Re-Importing aligned tilt series into WARP*), which can be used to re-import aligned tilt-series into WARP

The AreTomo processing results are located in a folder named "aretomo" inside the WARP processing directory.

The automatic processing of tilt series is configured in a new tab.

Set up the SSH connection

Fill in the IP of the AreTomo workstation (e.g. “preprocess1” or “localhost” or “aretomo.lab.ki.se”), the SSH username (e.g. “genval”) and select the private SSH key by clicking on the blue label after the Username box.

“MdocFilesDirectoryLinux”: Type in the path to the WARP processing directory or click on “Browse” to bring up a directory picker in the AreTomo Linux workstation. This functionality is designed for ThermoFisher TOMO, which writes *.mdoc files in the same directory as the DoseFractions.

Auto-run AreTomo

The Auto-run aretomo box can be ticked only if the “IP”, “Username” and SSH private key work, and the “MdocFilesDirectoryLinux” exists. Otherwise, it will remain grayed out.

When WARP is processing, “Auto-run AreTomo” will automatically queue newly collected tilt series to a AreTomo. Only one tilt series can be processed per GPU at a time.

WARP displays a list of GPUs (*e.g.* GPU 0 and GPU 1 available in the AreTomo Linux workstation). When the SSH connection is correctly set up, WARP will automatically query “`nvidia-smi`” in the AreTomo Linux workstation and update list the available GPUs.

AreTomo settings panel

Most of the AreTomo settings can be changed. The settings for AreTomo reconstruction are described in the AreTomo manual. However, there is some things to bear in mind:

“Dose per tilt”: AreTomo needs to know the accumulated exposure for each tilt image. If the “Dose per tilt” is set to zero, AreTomo will read the accumulated dose for each tilt image in the mdoc file. If it is set to a value greater than zero, AreTomo re-calculates the dose accumulated on each tilt image from this value.

“Sample pre-tilt”: This is useful when one is working with a pre-tilted specimen such a lamella, so that the tilt series collection has a stage pre-tilt (*i.e.* collection starts at an angle different than zero (e.g. 10 degrees)). ThermoFisher TOMO mdoc files do not take the stage pre-tilt into account, but you can correct for the stage pre-tilt here. If your stage pre-tilt is 10 degrees, type -10 degrees in the “sample pre-tilt” field.

“Tilt correction”: AreTomo can estimate a correction angle in case the specimen is not lying completely flat (see AreTomo manual). If the tilt correction first number is “-1”, then this function is completely disabled. If it is “0”, this means that tilt correction will be used for the alignment. If it is “1”, then it will be used for the alignment and the reconstruction. In the case of “1”, WARP will take the tilt correction into account when writing a corrected *mdoc file and *tomostar file. This correction will be applied in addition to the “Sample pre-tilt” correction.

“Include filtered out images”: Whether to use all tilt images listed in the mdoc file or leave out the ones excluded by the WARP filters.

“Copy to all”: It is possible to define different AreTomo settings for each tilt series (see below). Copy to all will overwrite any changes in the settings that may have been made in a tilt series.

“Clear all”: Delete the AreTomo processing results for all tilt series in the list. If processing is still running, kill the job and delete intermediate results.

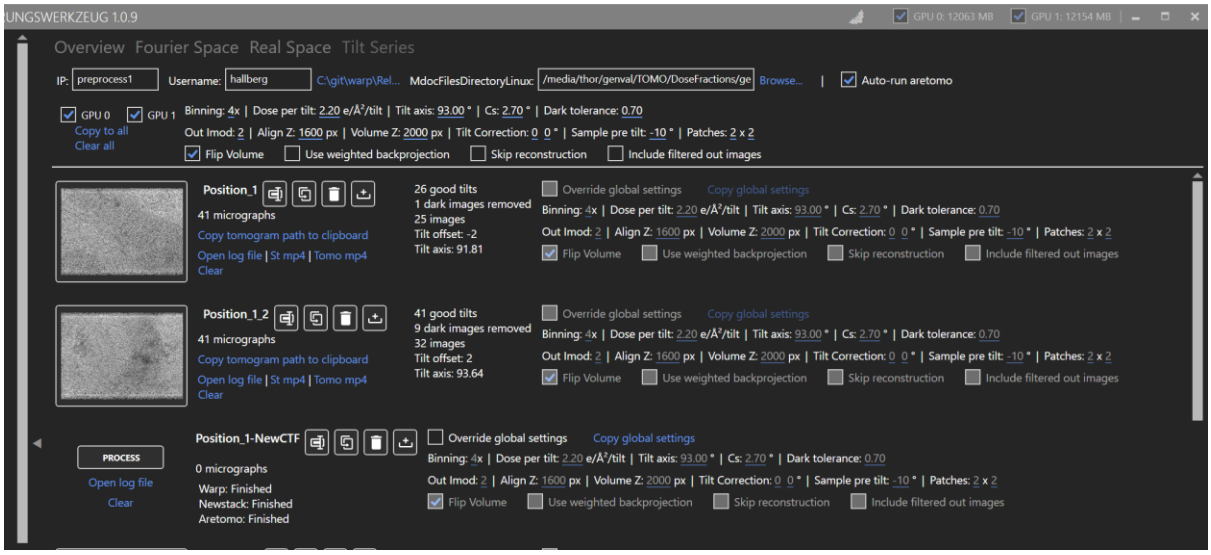
Processing a tilt series in AreTomo

WARP will watch the “Input” for newly created mdoc files. Each mdoc file will be imported as a new tilt series.

If “Auto-run AreTomo” is selected, the tilt series will be queued for alignment and reconstruction by AreTomo.

In the following screenshot shows the list of tilt series. The two first entries in the list, which are labelled “Position_1” and “Position_1_2” are already aligned and reconstructed by AreTomo. The series labelled “Position_1_NewCTF” has not been processed yet.

Newly imported tilt series are named after the corresponding mdoc file.



Each tilt series has the following actions:

	Rename. This changes the label of the tilt series, but the original mdoc file is not renamed.		
	Duplicate. Creates a copy of the mdoc file, which is then imported as a new tilt series. This is useful to compare different AreTomo settings for the same tilt images.		
	Delete. Deletes a tilt series and **also the original mdoc file**.		
	Add to “Auto-run AreTomo” queue		Remove from “Auto-run AreTomo” queue

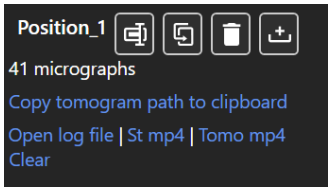
Tilt series after AreTomo processing

Position_1 in the image above is an example of a processed tilt series.

First panel

The thumbnail image displays the XY projection of the reconstructed tomogram. Clicking on the thumbnail image opens the XY and XZ projections in a larger image viewer. The XY and XZ projections are automatically created by AreTomo after a tomogram is reconstructed.

Second panel



"41 micrographs" is the number of tilt images in the *mdoc file.

"Copy tomogram path to clipboard" is useful to inspect the tomogram in another program, for example ChimeraX.

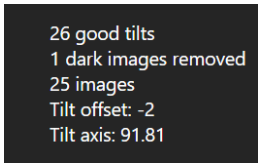
"Open log file" shows you the Newstack and the AreTomo log files.

"St mp4" will play a movie of the aligned tilt images to visually judge the quality of the image alignments.

"Tomo mp4" will play a movie going through the tomogram volume to visually judge the quality of the tomographic reconstruction.

"Clear" deletes the processing results, so that you can re-process this tilt series.

Third panel



"X good tilts": How many tilt images went into AreTomo. This equals the total number of images in the series minus the filtered out (blue items) after 2D processing in WARP.

"X dark images removed": Number of images discarded by AreTomo. Set "DarkTol" to zero for keeping all images.

"X images": How many images went into the reconstruction. These are also the images that will go in the corrected *mdoc and *tomostar files (see below *Re-Importing aligned tilt series into WARP*).

Tilt offset and Tilt axis. See the AreTomo manual for more info.

Fourth panel

This panel shows the settings that were used in this particular AreTomo run. Since this series is already processed, this panel is grayed out.

Tilt series before and during AreTomo processing

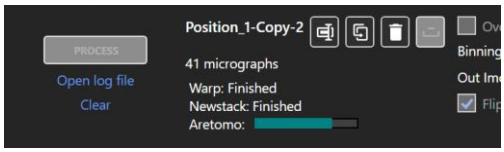
First panel

The "Processing" button will launch AreTomo immediately. The job will always go to GPU 0. This will skip the queue, even if there is something already running in GPU 0.

"Open log file" shows you the Newstack and the AreTomo log files.

“Clear” kills the processing immediately and deletes intermediate results.

Second panel



It displays the processing status of this tilt series.

Jobs can be “Waiting”, “Starting”, “Finished” or “Failed”. If any of the steps has failed, you can open the log file to identify the issue, then “clear” the tilt series and finally queue or process it again.

While the AreTomo step is running, a progress bar is shown.

Third panel

Here you can override the AreTomo settings set above for this particular tilt series.

Important: If the “Override global settings” is unticked when the processing starts, the settings in this panel will be overwritten with the global settings set on the top part of the page. This way avoids confusion as the panel always shows the settings that were used for the processing.

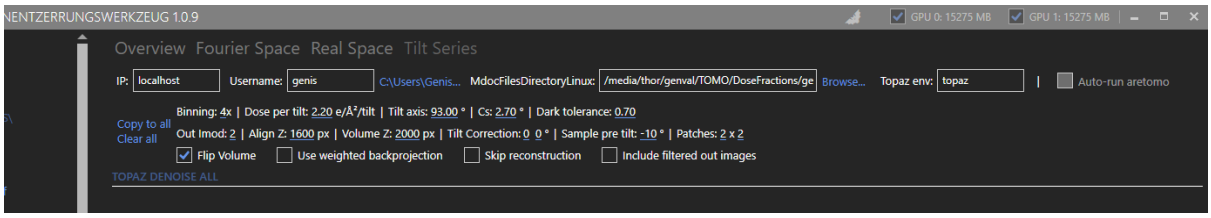
Running Topaz Denoise3D with default settings

If you are happy with the Aretomo results, I have added a convenience button to automatically denoise the tomograms with Topaz.

This will automatically run “`topaz denoise3d`” with the default settings for all reconstructed tomograms.

For this, Topaz needs to be installed in the same workstation where you run Aretomo, in it’s own separate Conda environment.

Just write the name of the topaz Conda environment in the box “Topaz env”, and click “TOPAZ DENOISE ALL”.



Re-Importing aligned tilt series into WARP

You can find the alignment and reconstruction files generated by AreTomo in the “aretomo” subfolder found in the WARP working directory.

When AreTomo finishes the processing of a particular tilt series, WARP will immediately write new *.mdoc and *.tomostar files.

The *.mdoc files are found in the “aretomo” subfolder, while *.tomostar files are located in the WARP working directory.

The *tomostar files can be directly imported into WARP for 3D CTF estimation. There’s no need to run the routine “Import tilt series from IMOD” anymore. The *tomostar files work with WARP 1.0.9 and also the WARP 1.1 beta (https://groups.google.com/g/warp-em/c/dAsmMnlULJA/m/CzlgVkJMCAAJ).

My system has two GPUs, so what I usually do is having two instances of WARP running at the same time, one running on each GPU. As the first instance, I have this custom implementation for 2D processing and running the alignment and reconstruction in AreTomo. In the second instance, I set the input file extension to *.tomostar and do the 3D CTF processing on the fly. I use the WARP 1.1 beta as the second instance.

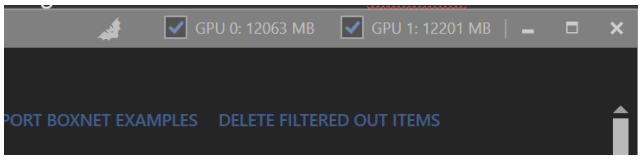
Reconstructing full tomograms and tomogram denoising still must be done manually a posteriori.

About corrected *.mdoc and *.tomostar files

These files:

- Apply the sample pre-tilt correction
- Apply the AreTomo tilt correction, if the first number in “Tilt correction” is set to “1”
- Re-calculate the electron dose, if the “Dose per tilt” is set to a number greater than zero
- The pixel size will be overwritten with the pixel size set in the WARP input options
- Exclude the tilt imges filtered out by WARP, if the “Include filtered out images” is not selected
- Exclude the dark images removed by AreTomo

Delete filtered out items



If you want to save precious disk space, this is the button. This will automatically delete all the WARP processing results (averaged micrographs, particle stacks, etc) for the filtered out items. The original dose fractions are moved into a subfolder named “Trash” in the WARP working directory. So you still have a chance to recover the data =)! You will have to manually delete the “Trash” folder if you really don’t want these data.

WARP also keeps a list of files that are filtered out (and potentially *deletable*) in a file called “filteredoutitems.txt” located in the WARP working directory.

Important: after you delete filtered out items, you might still have a few “blue items”. This is because of the astigmatism filter, which depends on the sigma of the entire dataset, becomes narrower after you remove some outliers.

About previous.settings

The original WARP keeps a file called “previous.settings”. This file is now called “custom_previous.settings”, as it contains a few extra data fields. You can still import “previous.settings” from the original WARP format manually.

Bugs, issues and requests

I will very much appreciate your feedback. Please use the GitHub tools for this.

Contact

Genis Valentin Gese | Postdoctoral researcher
Martin Hällberg group
Cell and Molecular Biology Department | Karolinska Institutet
171 65 Stockholm | Solnavägen 9
genis.valentin.gese@ki.se | ki.se