Pipeline utilization guide

1. Pipeline description and scope

The pipeline help utilization of pre-trained deep-learning model Cellpose for segmentation of 3D images of ovaries (or round objects) without extensive annotation and training with specialized data. Pre-processing steps (*see 3.a and 3.b*) are developed for 3D image stacks to improve visualization and the segmentation process, especially for heterogeneous fluorescent staining. Post-processing steps (*see 3.d*) are developed for correction of segmentation results and resolution of Cellpose limitation for segmentation of a large variety of objects size (*i.e* 20um to 1200um in diameter for medaka ovaries). It works with either cytoplasmic fluorescent signal or contour staining of follicles.

2. Requirements and installation

Details can be found in "Installation" file for :

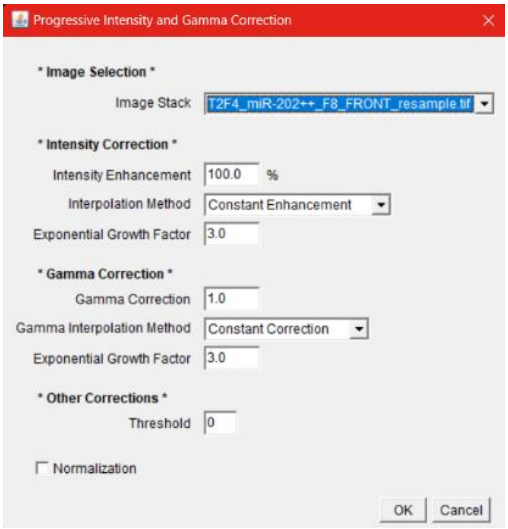
- ❑ Cellpose installation
- ❑ CLIP Fiji macro tool installation
- ❑ Installation of Fiji plugins used by the pipeline or needed for running CLIP Fiji macro tool.

3. Application guide

a. Manually prepare stacks depending on image acquisition (Optional)

Depending on image acquisition modalities, few steps may be needed before others automatic steps of the pipeline.

- **Intensity enhancement** : helps homogenization of fluorescent signal through depth of 3D stacks.



This step we uses the plugin from Murtin et *al*, (Murtin 2016): ***Progressive_Intensity_and_Gamma_Correction***. It increases or decreases the intensity and/or gamma contrast of each section progressively depending on the depth in the 3D volume stack.

Further explanations for setting parameters can be found here (**See p.124 and 164-165 for examples**) :

https://theses.hal.science/tel-01715589/file/these.pdf

- **Contrast enhancement :** helps homogenization and recovering of fluorescent signal.

For this step, we used the plugin ***Enhance Local Contrast (CLAHE)***. It is available in our CLIP macro tool : *...> CLIP_Image menu > Step_by_step > Contrast*

Contrasts can be enhanced on individual 3D stacks if images look different from each other in a batch. Otherwise, it can be applied on batch during the automatic process available in our CLIP macro tool (*see 3.b*) :

...> CLIP_Image menu > CONTRAST EDGE N2V

Or

...> CLIP_Image menu > CONTRAST N2V EDGE

Further explanations for setting parameters can be found here : https://imagej.net/plugins/clahe

- **Downscaling** : Reduce size of 3D stack and pixel size of objects on image.

Large 3D stacks may need extensive time for processing and high usage of computer RAM or GPU. Downscaling is recommended for images >2-3GB, but may depend on hardware specifications (RAM and GPU availability). Downscaling can also help adjusting the size of objects to segment (in pixels) for further segmentation step with Cellpose (*see 3.b*).

It can be applied on batch during the automatic process available in our CLIP macro tool (*see 3.b*) :

...> CLIP_Image menu > Step_by_step > Downscaling

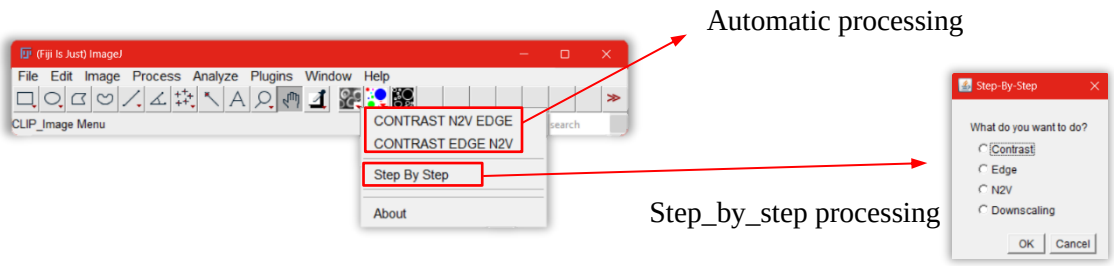
- **3D registration** : Reconstruct full 3D stack from two image acquisitions

Large 3D tissues, as medaka adult ovaries, may need two steps of image acquisition for complete imaging in depth (*i.e* acquisition from recto side, then from verso side of the tissue). Registration of corresponding recto/verso images is performed using the plugin ***Fijiyama*** from Fernandez et Moisy (Fernandez et Moisy 2021). Information and tutorials can be found here : https://imagej.net/plugins/fijiyama#tutorial-1--two-images-registration.

b. Automatic image pre-processing (Fiji CLIP_Image menu)

This part of the CLIP macro is for pre-processing images before segmentation process with Cellpose. Segmentation accuracy is sensitive to image noise and signal heterogeneity between objects. This enhancement process is thus designed to improve segmentation accuracy of suboptimal 3D images with low SNR or poor signal homogeneity.

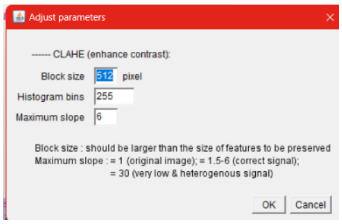
❍ CLIP_Image menu :



- ❑ Automatic image pre-processing : consists in different enhancement steps performed all-in-one. Two options are available with automatic steps following each other in different order (*Contrast, N2V, Edge* OR *Contrast, Edge, N2V*). A *Downscaling* step is also performed at the end of each option.

- ❑ Step_by_step image pre-processing : Enables performing each pre-processing steps independently. Parameters can be tested, adjusted and validated for each step. It is very useful, and recommended, to begin analysis of new datasets with this Step_by_step function.

❍ Pre-processing steps :



- **Contrast :** Enhancement of local contrasts helps fluorescent signal recovering on images. *Maximum slope* defines amplitude of histogram equalization. Further explanations for setting parameters can be found here : https://imagej.net/plugins/clahe. Other parameters are defined as follows : mask = "*None*"; fast = true; process_as_composite = false;

- **Edge :** Uses morphological gradient (difference of a morphological dilatation and erosion with the same structuring element) to enhance edges of the image. External gradient is subtracted from image, resulting in darker external borders of objects (Edge_border). Internal gradient can also be added on image, resulting in brighter signal of features (Edge_border&cyto).

- **N2V :** Performs image denoising from neural network prediction. A trained bioimage.io model is needed to perform this step. Training can be performed directly on image data. More information can be found here https://imagej.net/plugins/n2v#usage .

- **Downscaling :** While a first downscaling of images may be needed to accelerate image processing or avoid memory issues, another downscaling is performed at the end of pre-processing step. Downscaled image can thus be used for a second Cellpose segmentation that would be less accurate for small follicles but allows segmentation of largest follicles. **Use the same rescaling factor for each dimension.**

c. Cell segmentation with Cellpose

Cellpose was launched from Anaconda prompt (command line). Authors (Stringer et *al*, (Stringer et al. 2021)*)* provided detailed information and examples for utilization of Cellpose from command line or a notebook : See documentation https://cellpose.readthedocs.io/en/latest/command.html, https://github.com/MouseLand/cellpose

In our examples, we used :

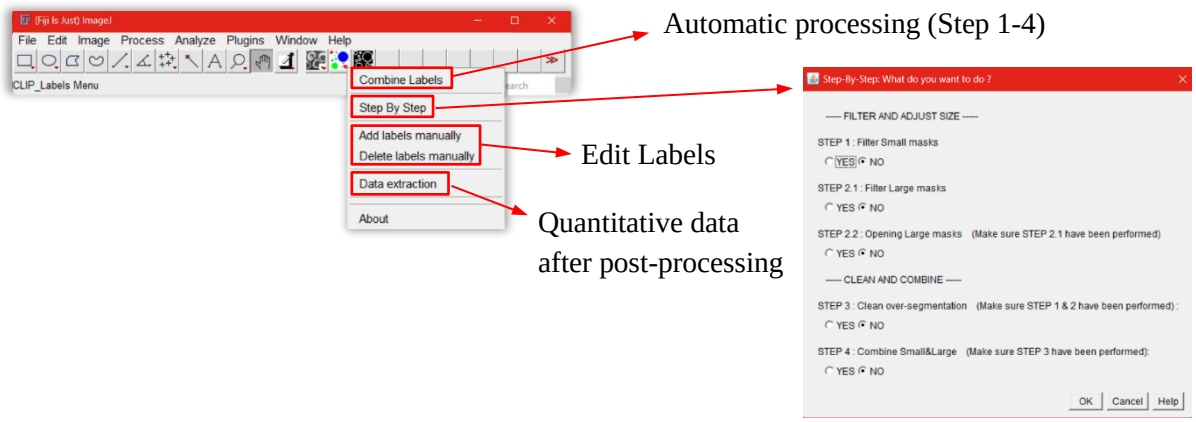
```
(base) L:\Deep_learning_dependencies\cellpose-master>
(base) L:\Deep_learning_dependencies\cellpose-master> conda activate cellpose
(cellpose)  L:\Deep_learning_dependencies\cellpose-master> python  -m  cellpose  --dir
D:\Analyse_3D\For_cellpose  --pretrained_model  cyto  --diameter  30  --resample  --do_3D  --
flow_threshold 0.6 --cellprob_threshold -2 --chan 0 --use_gpu --batch_size 2 --save_tif --
no_npy
```

--*Diameter* was changed to 60 when needed to segment largest follicles missing in previous segmentation runs.
--*Batch size* was set according to our GPU memory limitation to avoid crash at the end of segmentation process.
--*Cellprob_threshold* was determined with few tests on Cellpose GUI in order to avoid ill-shaped segmented labels but keeping max probability of finding cells.
Anisotropy was set directly in python file "models.py" as --anisotropy argument wasn't available in precedent versions of Cellpose (now available in Cellpose v.2).

d. Automatic Label filtering and Combine labels (Fiji CLIP_Labels menu)

This part of the CLIP macro is designed for post-processing of segmented images. Cellpose segmentation results may conserve small false positives labels or missing large labels when image presents a great variety of objects size. Post-processing here aims at filtering false positives and combining results (where labels are missing) with other segmentation results where large labels are well segmented (*i.e* when image downscaling is done before Cellpose).

❍ **CLIP_Labels menu :**



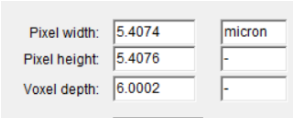
- ❑ Automatic image post-processing (Step 1-4) : consists in different steps performed all-in-one. At least two segmentation results of a given image sample are required for CombineLabels, as it is designed for combination of largest labels from one image with the other labels on an image lacking proper segmentation of large follicles.

- ❑ Step_by_step image pre-processing (Step 1-4) : Enables performing each automatic post-processing steps independently. Parameters can be tested, adjusted and validated for each step. It is very useful, and recommended, to begin analysis of new datasets with this Step_by_step function. Steps are further detailed in the following section.

- ❑ Edit labels : Enables manual label editing after combination step to correct some remaining segmentation errors (Add_manually/Delete_manually).

- ❑ Data extraction : Creates a file containing labels measure results (number, size of follicles).

WARNING : Make sure image resolution is calibrated before proceeding. Calibration is normally done during pre-processing steps.



❍ **Post-processing steps :**

- **STEP 1 : Filter small masks :**

Size and shape filtering process of "Small_Masks" image (*i.e* segmentation results obtained with Cellpose from original resolution images (_cp_masks)). Size filter refers to a volume calculated for each label in the image. An equivalent diameter for each volume can be calculated, so that the required input parameter is a diameter. Shape refers to a measure of sphericity, as oocytes are mostly round.

Required parameters and how to choose settings :

- *Diameter :* Indicated diameter will determine the lower 3D equivalent volume limit for labels to keep in the analysis. Choice of diameter limit will depend on image resolution and size of follicles. Cellpose segmentation will produce different rates of precision for smaller objects (< 30 or < 10 px). **Find a value that eliminates poor segmentation labels for smallest objects.**

- *Sphericity :* Labels that present a sphericity lower than the indicated value will be deleted. Small labels, representing small follicles, are expected to be mostly round, so sphericity is often superior to 0.4, while false positive often have irregular shapes, thus being mostly eliminated. However, pushing sphericity limit very close to 1 may lead to loss of true positives. **Focus on finding the best value to delete maximum false positive while avoiding loss of true positives.**

Example :

1. For Medaka Adult ovary sample with image resolution of 5.41*5.41*6.0002 µm.
   - *Diameter =50um.*
   - *Sphericity = 0.45 or 0.5*

30 pixel diameter parameter used for Cellpose segmentation matches follicles of ~160 um. A limitation at 50 um diameter will correspond to ~10 pixels diameters for Cellpose, which seems close to its limits for good segmentation results. In those image resolution conditions, Cellpose results have correct accuracy for follicles between 10 to 100 pixels diameter (50um to 600um).

A sphericity of 0.45 or 0.5 was used to eliminate obvious segmentation errors. Note that it also eliminates largest labels that are not well segmented because of out-of range size (due to image resolution).

2. Same parameters were used for Trout ovary sample (same image resolution)

- **<u>STEP 2 : Filter large masks</u> :**

Size and shape filtering process of "Large_Masks" image (*i.e* segmentation results obtained with Cellpose at lower resolution (RES_cp_masks)). Size filter refers to a volume calculated for each label in the image. An equivalent diameter for each volume can be calculated, so that the required input parameter is a diameter. Shape refers to a measure of sphericity, as oocytes are mostly round.

<u>Required parameters and how to choose settings :</u>

- *Diameter :* Indicated diameter will determine the lower 3D equivalent volume limit for labels to keep in the analysis. Choice of diameter limit will depend on image resolution and labels wanted for combination. **Find a value that enables keeping only largest labels that are missing or poorly segmented at STEP 1.**

- *Sphericity :* Labels that present a sphericity lower than the indicated value will be deleted. Large labels segmented at lower image resolution may present less spherical shapes than small labels. **Focus on finding the best value to delete maximum false positives or poorly segmented labels while avoiding loss of true positives.**

- *Opening :* Morphological operation consisting in a succession of an erosion with a dilatation. It smooths out labels shape but also slightly reduce their size with each iteration. Number of iterations to apply depends on the minimal size kept for large labels to be combined. It should be used carefully, as it highly modifies the spherical shape of labels when highly iterated. **Large labels overlap with neighbours (from STEP 1 image) have to be minimized.**

<u>Examples:</u>

1. For Medaka Adult ovary sample with image resolution of 10.82*10.82*12.0004 µm (RES_cp_masks) & total range of follicles 20um-1000um.
   - *Diameter = 600um*
   - *Sphericity =0.25*
   - *Opening = 6*

600um diameter matches here a diameter of ~55 pixel, and is the accuracy limit for STEP 1 image. In this image resolution condition, good accuracy may be obtained up to 100 pixels = 1100um diameter.

Sphericity of 0.25 will help eliminate obvious false positives (epithelium or other segmentation artifacts) but avoid loss of large follicles that are not perfectly segmented (shape fitting) but can be improved by opening operation and thus kept in the analysis.

Opening of 6 applied to the labels >600um helps smoothing and size reduction in preparation for combination. Large labels overlap with neighbours is reduced but shape is well preserved.

2. For Trout ovary sample with image resolution of 10.82*10.82*12.0004 µm (RES_cp_masks) & total range of follicles 20um-600um.
   - *Diameter = 350um*
   - *Sphericity =0.25*
   - *Opening = 3*

Trout sample (in this example) has a smaller range of follicle size (depending on sampling). Image resolution is similar to medaka sample, but with lower segmentation quality. Medium follicles (350-600um) are segmented more accurately with downscaled image, so they are selected at this step by applying *Diameter 350.* As follicles are not very large, *Opening* is not strong, set to 3.

- **STEP 3 : Clean over-segmentation before combination :**

The main problematic resolved by this step is over-segmentation present at presumptive locations of large follicles. This happen on an image that has not been downscaled and containing follicles with diameters out of range for Cellpose segmentation. The strategy here is to delete those over-segmented labels before addition of the correctly segmented large labels obtained from STEP 2.

How it works

Large labels are used for detection of regions where over-segmentation needs to be deleted. Incorrect labels within or touching those regions are then deleted. Thus, to avoid deleting at the same time some correct labels in the neighbourhood of large labels, different erosions are applied depending on large labels size so that they overlap with maximum over-segmentation labels but not neighbourhood. For that, label image from STEP 2 is split in three 3D stacks containing each a defined range of label size.

Required parameters and how to choose setting

--*Split large masks : Diam1, Diam2 :* Defined diameters will identify the size range of follicles contained in each stack. Stack 1 contains labels measuring less than *Diam1*. Stack 2 contains labels measuring between *Diam1* and *Diam2.* Stack 3 contains labels measuring more than *Diam2.*

Diam1 will often be the same value than Diameter used in STEP 2 for filtration, thus splitting image in only two stacks of large labels. Another opposite possibility is to apply only one erosion factor by using *Diameter (STEP 2)* value for *Diam2*. **If labels selected from STEP 2 have a great size range (ex 500-1000), it may be preferred to split in at least two stacks.**

--*Erode Large masks: small, medium, large (erosion factor) :* Erosion factor will be applied to labels of different range of size (contained in 3 or less stacks), preventing loss of neighbourhood by avoiding overlap.

The larger labels are, the higher transformation is needed (opening + erosion). High opening at STEP 2 will need less erosion and lower opening will need higher erosion factor. **Prefer using a common medium opening (4-8) at STEP 2 and then adapt size of largest labels through erosion factors.**

Examples

1. For Medaka Adult ovary sample with Large masks starting at *Diameter 600um (step 2), Opening 6* and a total range of follicles 20um-1000um.
   - *Diam1 = 600*
   - *Diam2 = 750*
   - *Small erosion factor =1 (not used)*
   - *Medium erosion factor =4*
   - *Large erosion factor =5*

2. For Trout ovary sample with Large masks starting at *Diameter 350um (step 2), Opening 3* and a total range of follicles 20um-600um.
   - *Diam1 = 400*
   - *Diam2 = 500*
   - *Small erosion factor =2*
   - *Medium erosion factor =3*
   - *Large erosion factor =4*

- **STEP 4 : Combine**

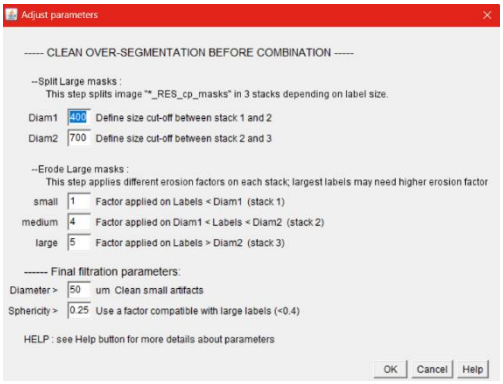
Use results from STEP 1 and STEP 3 and combine them together. No parameter needed. Adjust STEP 2.2 and STEP 3 if results are not satisfying. Label filtration is performed again at the end to clean small artifacts potentially created during combination. Choose parameters as in STEP 1.

- **STEP 5 : Edit labels / Extract data :** Enable manual editing of labels after combination process, then extracting quantitative data from final results.

- Add manually : Strategy is similar to automatic combination, but instead of selecting a full range of large label size, labels can be selected individually and manually.

✓ Image “To_select” should be a Cellpose result (_cp_masks, _RES_cp_masks).
✓ Image “Combined” should be the result of automatic combination (STEP 1 to 4).
✓ Image “Image” should be the pre-processed image used for best resolution segmentation.
✓ Opening will be applied on “To_select” image.

For label selection, “Combined” and “Image” will be merged into “Composite” to help detection of missing labels. Multi-point tool is used to select missing labels on the image “To_select_open”. Synchronize Z-stacks with “synchronize windows” to facilitate stack navigation (select “Composite” and “To_select_open” stacks). Multiple click on the same label is allowed.



After label selection, labels erosion is applied depending on label size range. Strategy is the same as in CombineLabels. Choose diameter and erosion factor in accord with labels that are selected. Estimations of selected labels size can be done by looking at image “*name*_Labels_selection”.

For final filtration, choose parameters for cleaning small label artifacts that may have been created (Diameter), but use sphericity factor compatible with Large labels (<0.4).

Example :

If labels selected are medium (300-400um) and very large (800), use a small Opening factor. It will prevent loss of shape of medium labels. Then to avoid overlap before combination, apply a small erosion factor to medium labels, but higher erosion factor to very large labels.

- Delete manually : Enable manual correction of labels poorly segmented that were not deleted by Filtering process (size and shape).

For label selection, “Combined” and “Image” will be merged into “Composite” to help detection of missing labels. Label edition (MorpholibJ) is used to select labels to delete on the image “Combine”. Click on labels to delete and click on “Remove selected”. When selection is finished, click on Done, then OK on dialog.

- Extract data

Extracts quantitative data from images. Name of images should end by “cb_label_filtered” (*i.e* after CombineLabels or Add_manually/Delete_manually. File extracts Sample name, image resolution, total number of oocytes counted (labels) and each diameter in um.