Part of this code is from : https://github.com/BPHO-Salk/PSSR

Results obtained with this code : https://github.com/thibaut1998e/centriole_internship

## *General overview of the code :*

train.py : train a neural network with fastai and save a model  (most of this code is from PSSR)

inference.py : test the model and save the results

transformations.py : various transformations to  apply on images (crop, padd, convolution, resizing …)

apply_transformation.py : contains a method which applies a list of transformations to all the images in a folder and save the outputs.

paths_definitions.py : paths constants

main.py : apply degradation model, train a neural network and test it

plot_graphs.py : plot several curves

topaz_commands : methods which execute topaz command lines

split_source_data_train_valid.py

files to clean data : invert_channels.py , delete_list.py, delete2D_images.py

convert_lif.py, convertles .lif in. tiff

dossier utils (mostly from PSSR) : definition of several architectures, metrics and losses (feature loss, l1 loss)

The first step is to get  .tiff images of centrioles using .lif files. To do that use the python file *convert_lif.py* by specifying at the top of it the source and destination folders. Set also the variable *number_of_channels*. The destination folder is then split in 2 separate folders deconv and raw each of them are split in 2 channels c1 and c2.

Then you can clean the data by removing unwanted data, to do this use the python file *delete_images* by specifying the folder and the path of the txt file containing names of undesired images. You can also delete 2D images, which are often undesired with the code *delete2Dimages*. For some images channels are inverted by mistake, use similarly the code *invert_channels* to fix it.

The next step is to split the data into 2 subfolders train and valid, use the code *split_source_data_train_valid*. You can specify the proportion of validation.

Before strarting using the code, modify the constant path variables in the file *paths_definition* if needed.

Then create the conda virtual environment with the file *env.yml* with the command line :

*conda create --file env.yml -n virtual_env*.

Another environment is needed to use codes which execute topaz commands lines, to install topaz and create the environment follow their guide at : https://github.com/tbepler/topaz.

Then you are ready to use the code ! 😊

A method which is useful for many purposes in the code is the method *transforms()* in the python file *apply_transformation.py*. It takes as input a source folder *folder_in*, a destination folder *folder_out* and a list of transformations. A transformation is a function which takes as input a 2D or 3D array (representing an image) and returns a 2-3D array or a list of 2-3D array. The method applies to all images in *folder_in* the list of transformation and stores the outputs in *folder_out*.

Source images, got from the code *convert_lif,* are 3D images but we want 2D images to train our model. Moreover, to train a model with fastai , it is required for all images to have the same shape, and the shape is slightly different from one image to another. We then need to take one (or several) section(s) for each image and crop them. To do that we call *transforms()* with the method *cross_section*(), in the file *apply_transformation*.

Then to add spots to HR images use the method *add_spots()*. It is the only transformation which takes times so its better to do it only once.

Once it's done you can run the code *main.py*. It does 3 things : apply the degradation model to HR images by convolving, resizing and normalizing. Then train and save a model with fastai by calling the code *train.py*. Finnaly it tests the model with the code i*nference.py*. LR images are generated from the folder HR_spots, using also the method *transforms()*, called by *conv_resize_norm()*. You can choose training and testing parameters in this file. Results are saved in the folder *test_results*.

## **Topaz :**

Methods which executes topaz commands are located in the file *topaz_commands.py*

Topaz data are located in the folder *data_centriole_detection,* which contains 2 folders train and valid and hand-labelled positions of centriole in 2 .txt files, *valid_labels.txt* and *train_labels.txt*.

The method *train_topaz_model* uses these data to train a model which detects position of centrioles by executing a topaz command line. It first rescales if needed the images so that the centriole size in pixel matches the receptieve field of their neural network, then trains a model. You can visualize training curves with the method *plot_metrics* in *plot_loss.py*.

Then to predict the postition of centers and associated scores, you can either use a model that you have trained or use the model *centriole_detection_epoch100.sav*, already trained on this data set. To do that call the method *compute_center_particles()* in the same file. Results are saved in a txt.

After that, you can use this txt file to crop images on these predicted centers. There are 2 methods for that one for 2D images and one for 3D images.

For 2D images call the method *crop_topaz()* in *apply_transformations.py*. This method first calls the method *get_center_dict_from_txt()* which returns a dictionnary from the txt file computed by topaz. Keys are names of images and values a list of *Center* object, a center is defined by 2 attributes its score and its position (a couple). Then images are cropped with this dictionnary.

For 3D images call *crop_threeD_topaz()*. It uses the transformation *crop_threeD_with_center_dict(),* which has several options.

This transformation detects the position in 3D of centrioles given positions in 2D for all the slices. You need to specify a radius : center closer than radius, in two different slices or in the same slice, are considered to represent the same particule.

If the option *average* = True is given the center in 3D is computed as the average of all centers (closer than raidus) in 2D. This is a weighted average by the scores. So if you set a threshold lower than 0 in *get_center_dict_from_txt()* make sure that you have normalized the scores (with option *normalize* in *get_center_dict_from_txt())*.

You can save graphs of intensity through slices with the option *save_figure*. In these figures we can visualize 2 things : slices of local minima of intensity correspond to end slices of centrioles. Besides local maxima of intensity are slices where the centriole is the most visible. We can use this fact to cropp images in a smarter way.

First by setting the option cut = True, patches are cropped between 2 local minima of intensity.

Secondly with average = False, the 3D center is defined as the 2D center of the slice of highest intensity.