SERNN cluster analysis manual  
© 2022 Kanchanawong Lab, Mechanobiology Institute, National University of Singapore

Contents

[1. Quick overview 2](#_Toc102662957)

[2. Detailed step-by-step guide 3](#_Toc102662958)

# Quick overview

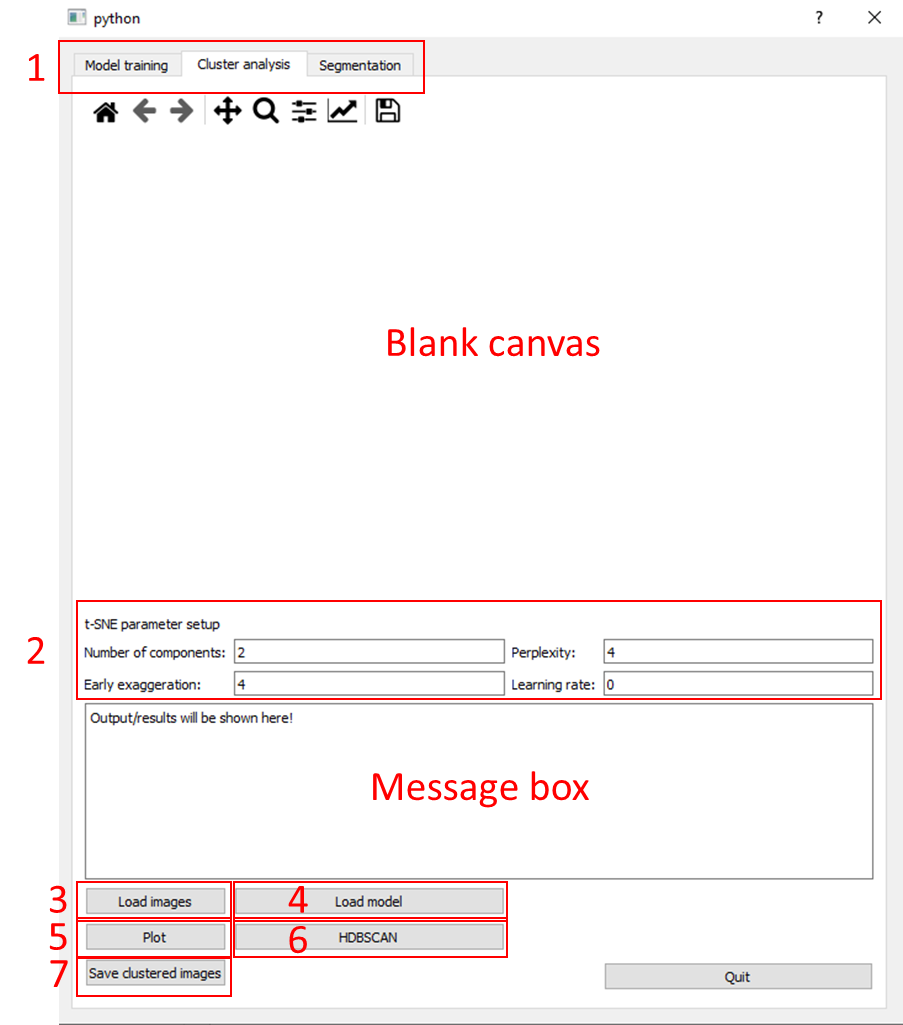


Figure 1

Figure 1 shows the overall layout for the software. To perform cluster analysis, follow the 7 steps as indicated on figure 1:

1. Ensure that the tab is set to “Cluster analysis”.
2. The default t-SNE parameters are as shown in figure 1. These parameters can be changed any time before performing the t-SNE plot (step 5). Please note that the learning rate value will change after plotting the t-SNE plot if it is left at default value of 0.
3. Load the folder containing your images. All the images to be used should be contained within the same folder. Please do not have subfolders in the main folder as they will not be loaded.
4. Load the trained model. You can use your own model or the provided SERNN model, “Trained\_SERNN\_model.h5”.
5. Click on “Plot” to perform the t-SNE plot. This step will take a while depending on the model size and number of images used. The plot will be shown in the blank canvas. All the data points will be red in colour.
6. Click on “HDBSCAN” to perform HDBSCAN clustering. The colour of the data points will change depending on the number of clusters and the clusters that the points belong to.
7. Click on “Save clustered images” to save the images with the cluster label. The folder containing these clustered images will be in the same directory as the Python script.

The next few pages will show an example of a cluster analysis.

# Detailed step-by-step guide

1. First, check that you have the following files in the same directory as shown in figure 2 below. The “test\_images” folder contains the images to be clustered. The “RUN” script is the file to be run for cluster analysis. The “Trained\_SERNN\_model.h5” file is the file containing the trained model that will be used to extract features from the test images. Please note that you may use your own test images and model as long as they are in the same directory.

Text

Description automatically generated

Figure 2

1. Next, run the “RUN” script and the following window should pop up (figure 3 left). Click on the “Cluster analysis” tab at the top and the layout should change (figure 3 right).

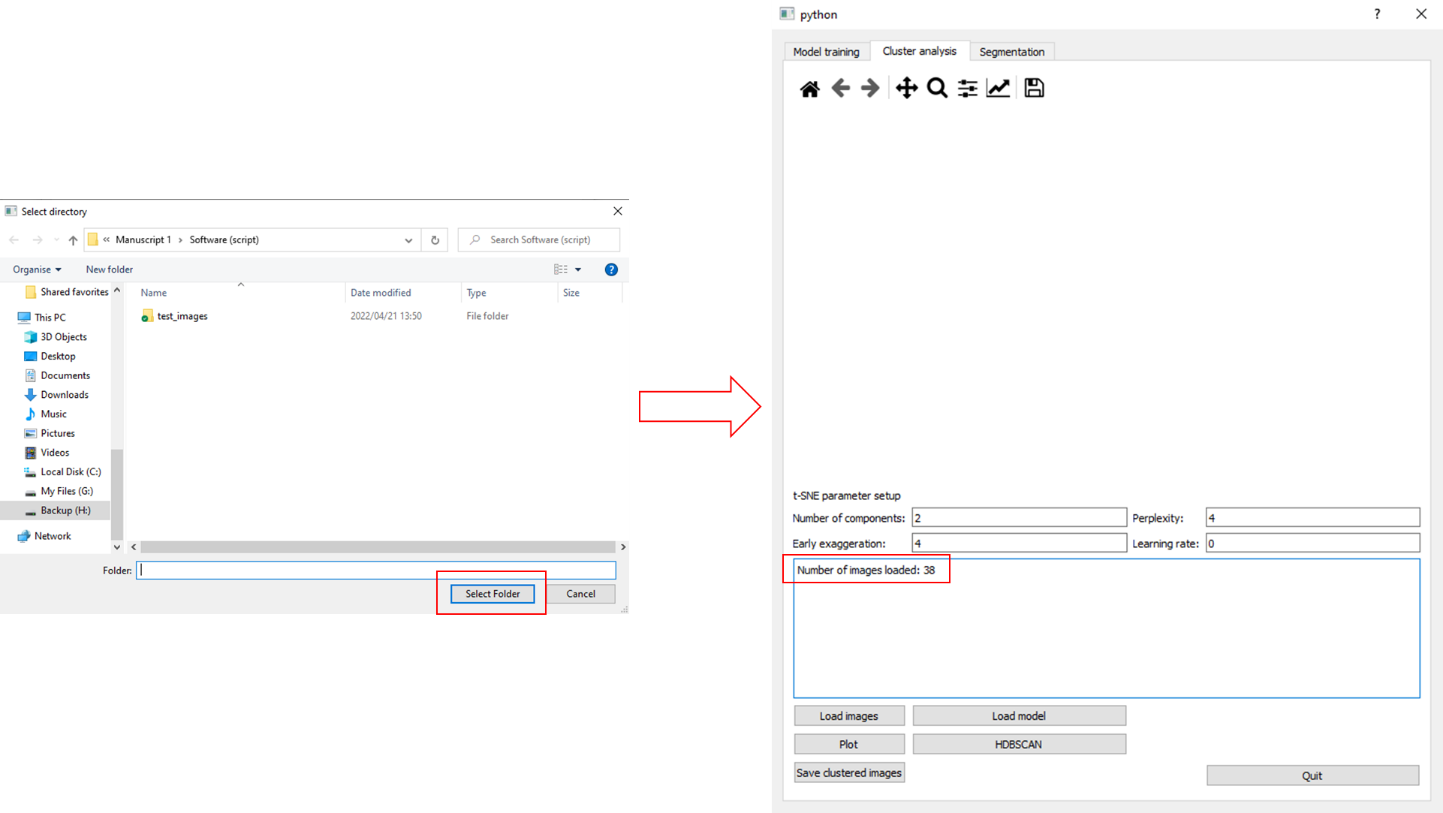


Figure 3

1. Press on the “Load images” button and select the folder containing the test images before clicking on “Select Folder”. Once that is done, there will be message in the message box stating how many images have been loaded (Figure 4).

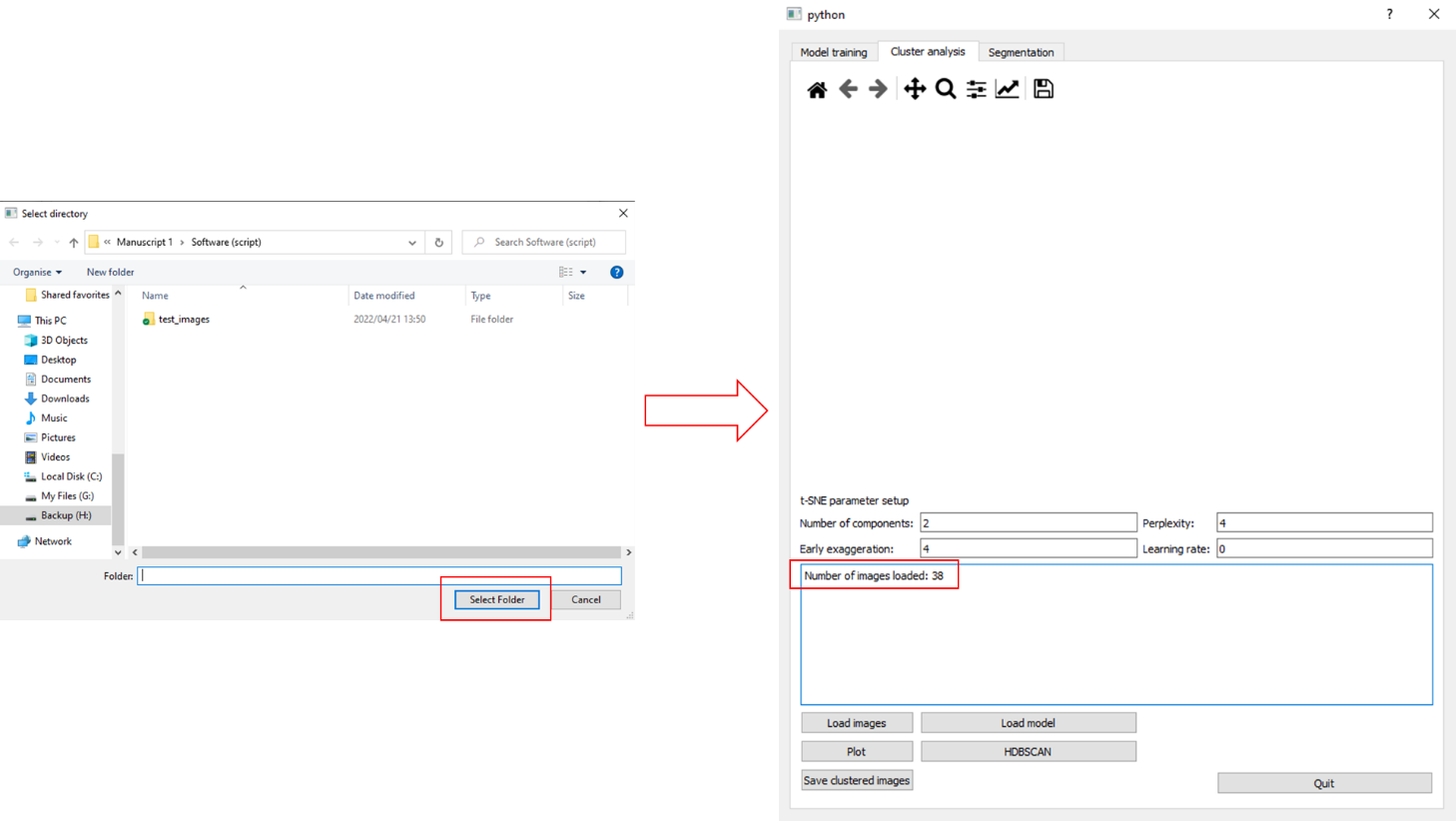


Figure 4

1. Next, press the “Load model” button, select the model to be used, and click on “Open”. The model architecture will be displayed in the message box (figure 5). Please note that this step can take a while.

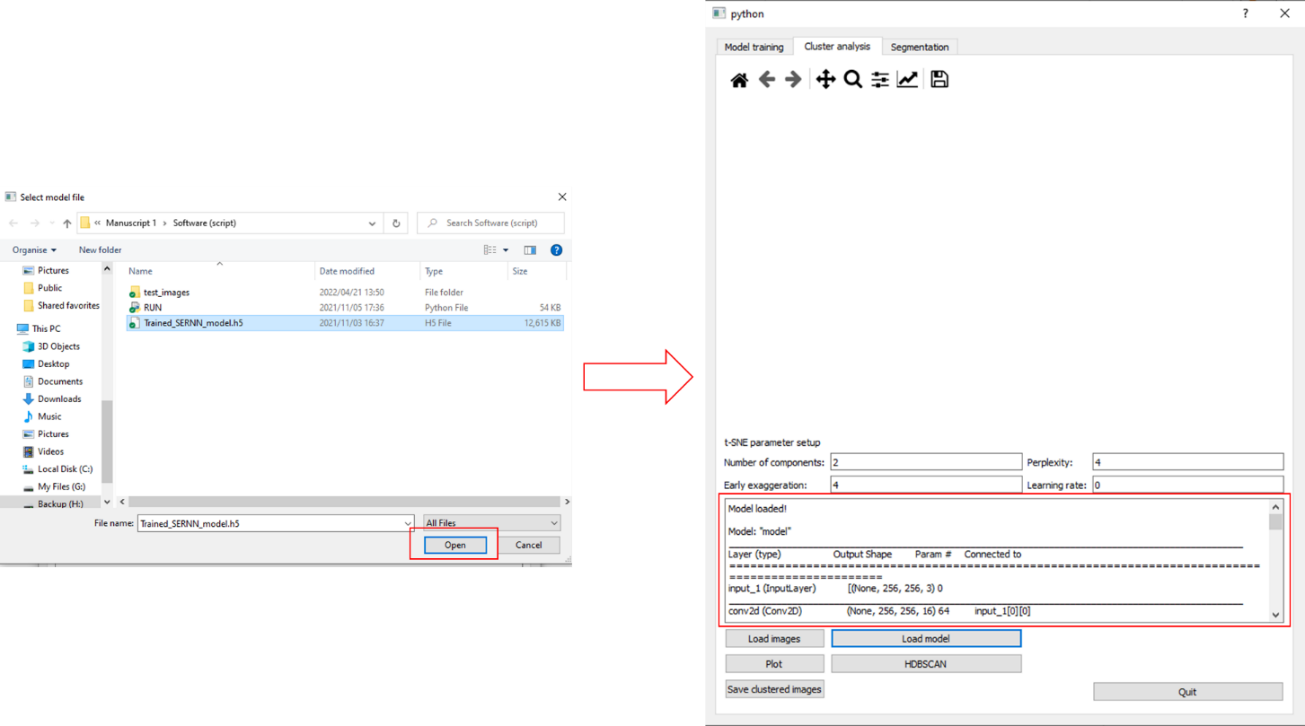


Figure 5

1. You are now ready to perform the t-SNE plot. You may change the t-SNE parameter setup values if you wish to. If the learning rate value is left at the default value of “0”, a value of will be used. For this example, the learning rate will not be changed and will be left at the default value of “0”. As 38 images have been used for the example, the learning rate value should be changed to 3.167. To perform the t-SNE plot, click on the “Plot” button. Please note that this step can take a while.

Chart, scatter chart

Description automatically generated

Figure 6

1. The t-SNE plot will be shown in the blank canvas as shown in figure 6. As expected, a value of 3.167 is shown for the learning rate. You can save the t-SNE plot as PNG file by clicking on the floppy disk icon that is located right under the “Segmentation” tab. To perform HDBSCAN clustering on the dimensionally reduced points, click on the “HDBSCAN” button and the colour of the data points will change (Figure 7).

Chart, scatter chart

Description automatically generated

Figure 7

1. To save the images, click on the “Save clustered images” button.

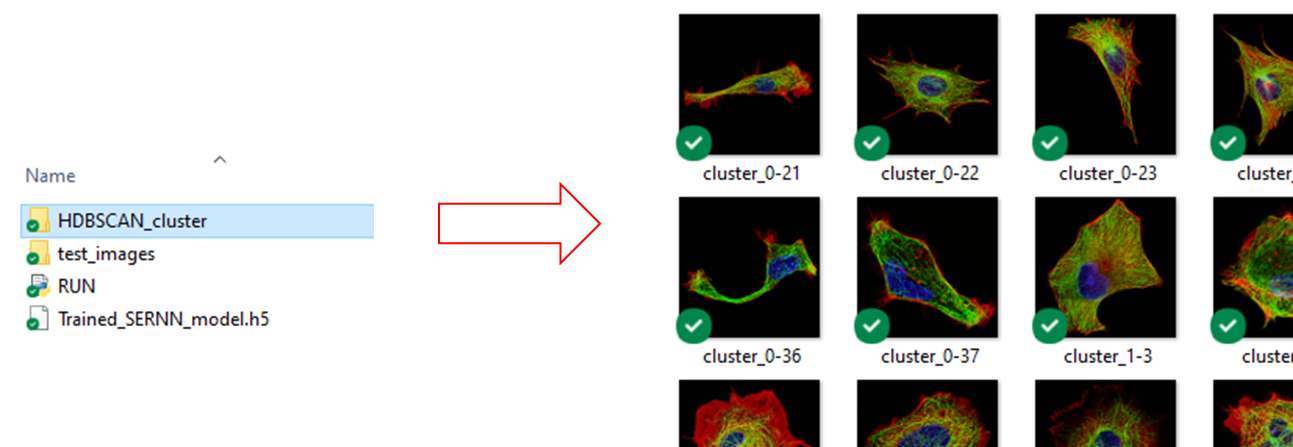


Figure 8

The clustered images will be saved in the same directory as the other files (as shown in figure 8, named “HDBSCAN\_cluster”). The name of each image follows the following structure:

“cluster\_{cluster number}-{image index}”

The order of the image index will follow the same order as the one found in the folder containing the test images.