SegEM: Semi automated image analysis toolkit for Connectomics

==

This toolkit provides all code required for the analysis of large-scale volume EM data for Connectomics.

(Berning, Boergens, Helmstaedter)

There are two main use cases:

1. De-novo training of CNNs for the analysis of a novel EM dataset.

2. Connectomic analysis of an already volume-segmented data set based on skeleton reconstructions.

Prerequisites:

+ The Matlab jobmanager has to be operational. If no compute cluster with a Matlab jobmanager is available, the jobmanager 'local' will usually work by default.

+ Matlab-mex compiler has to be configured. This can be done via 'mex -setup' in command line or the Matlab command prompt. This is required to compile some of the routines for your computer.

Usage

Set the Matlab working directory to the root folder of this code repository.

Execute initalSettings.m and provide the requested information in the GUI (alternatively, these variables can be set via command line):

+ Data Directory = Parent directory containing the raw EM data in the "Knossos" format (www.knossostool.org), located in directory "datasets/[datasetName]/color/1/" (use case 1); this directory should also contain the volume segmentation and skeleton reconstructions (use case 2) stored in directories "dataset/[datasetName]/20141007T094904/local/" and "supplement/extracted/", respectively. An example dataset (retina dataset, ek563) is available via www.segEM.io.

+ Name of jobmanager = Name of operational Matlab job manager ('local' in case no cluster is available)

+ Output Directory = Directory to store data and figures generated by SegEM scripts

+ Code version = choose **SegEM** for de-novo analyses (use case 1); choose legacy version for analysis of retina dataset ek563 (use case 2)

Press "OK". Then, the relevant scripts will be opened in the Matlab editor.

The scripts are most conveniently executed using the cell mode in Matlab (pressing Ctrl+Enter executes the current script section; see "cell mode" help in Matlab).

Use case 1: start with cnnStart.m, then proceed to cnnParameterSelection.m, mainSeg.m, wholeDatsetFwdPass.m and skeletonsToContacts.m; these scripts will be opened in the matlab editor automatically (s. above)

Use case 2: start with skeletonsToContacts.m. When working with the retina dataset ek563, use legacy version skeletonsToContacts_legacy.m, instead.

Brief description of most relevant **SegEM** code files

+ Step 1 - cnnStart.m: load training data, train convolutional neuronal networks (parallel network training on multiple GPUs)

+ Step 2 - cnnParameterSelection.m: Automated hyper-parameter selection and variation of learning rates for each CNN layer

+ Step 3 - mainSeg.m: steps from classification result by CNN to segmentation (grid parameter search for watershed segmentation) including skeleton based split-merger metrics

+ (Optional) - visSeg.m: visualize a given segmentation, e.g. object collected along skeletons in dense skeletonization, movies of segmentation and similar

+ Step 4 - wholeDatasetFwdPass.m: Apply trained CNN classifier (trained in step 1 or 2) and watershed segmentation steps (with parameters optimized in step 3) to whole dataset

+ Step 5 - skeletonsToContacts.m: Volume reconstruction and contact detection based on skeleton reconstructions. This requires volume segmentation (either via step 4 or via the segmentations provided at www.segEM.io)

To further analyze the retina dataset ek563, the following SegEM legacy code is provided:

+ main_legacy.m: load training data, train a convolutional neuronal network, parallel network training

+ mainSeg_legacy.m: steps from classification result by CNN to segmentation (grid parameter search for watershed segmentation) including skeleton based split-merger metrics

+ visSeg_legacy.m: visualize a given segmentation, e.g. object chains, segMovies, errorMovies = detected merger errors etc.

+ skeletonsToContacts_legacy.m: Use segmentation of whole dataset (provided) to use skeleton reconstructions for volume reconstructions or contact detection between cell pairs