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 data, Location of raw EM data in "Knossos" format (www.knossostool.org) is located in the datasets/[datasetName] alongside a segmentation that can be used for skeleton based whole dataset volume reconstruction. An example dataset (retina dataset, ek563) is available via segEM.io. This folder should also contain different data depending on use case: Volume segmented training data in matlab files in onlineMaterial/extracted/stackKLEE for visualization of training data and/or onlineMaterial/extracted/targetKLEE for targets CNN is trained on (both provided on segem.io for retina dataset e_k0563 and cortex dataset 2012-09-28_ex145_07x2). In the subfolder /supplement/extracted/ all supplementary data supplied with the paper and on segem.io (or analog data for novel dataset) should be put.

+ 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 scipts 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 (->cnnParameterSelection.m->mainSeg.m->wholeDatsetFwdPass.m->skeletonsToContacts.m); will all be opened in this order after setting directories in GUI

Use case 2: start with skeletonsToContacts_legacy.m (other scripts only provided for purpose of documentation, will require some changes, e.g. changing folders manually, to run on local computer)

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

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

+ cnnParameterSelection.m: Automated hyperparameter selection and variation of learning rates for each CNN layer

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

+ wholeDatasetFwdPass.m: Apply trained CNN classifier (trained in step 1) and watershed segmentation steps (with parameters optimized in step 3)

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

To futher analyze the retina dataset ek563, the following segEM legacy code is provided:

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