Segmenting Large Electron Microscopic Image Volumes

An Introduction to NBCR image analysis and segmentation tools

National Biomedical Computation Resource Summer Training Program @ UC San Diego

Alex Perez
August 1st, 2016
a3perez@ucsd.edu

Hands-On Session 1-1

The Manual and Automated Segmentation of Organelles in 3D EM Data

Goals

By the end of this session, you will be able to:

1. Use IMOD programs to manipulate SBEM image stacks
2. Generate training datasets that are compatible with running CHM jobs through the SLASH portal
3. Run CHM jobs through the SLASH portal and interpret their results

1. Download and Unzip the Dataset

A. Download the ZIP file entitled *Download_Dataset_1-1.zip* from the BBDTC portal

B. Unzip the contents to an easily accessible location:

    i. <u>For Windows</u>: Unzip the file into a new folder in your Cygwin home directory. If you used the default install path of C:\cygwin64, the home directory will be C:\cygwin64\home\<username>.

    ii. <u>For Mac</u>: Unzip the file into a new folder (e.g. /Users/<username>/Download_Dataset_1-1).

    iii. <u>For Linux</u>: Unzip the file into a new folder (e.g. /home/<username>/Download_Dataset_1-1).

    Try the command: *unzip Download_Dataset_1-1.zip –d Download_Dataset_1-1*

D. Open your terminal program. Enter the directory to which you unzipped the dataset by using the *cd* command in the terminal:

    For Windows/Linux (example): *cd /home/<username>/Download_Dataset_1-1*
    For Mac (example): *cd /Users/<username>/Download_Dataset_1-1*

E. List the directory's contents using the *ls* command. You should see one file, *sbem_tiles_2Dbin2.mrc*

Example Output

2. Determine the Properties of the Dataset

A. At the command line, type *header sbem_tiles_2Dbin2.mrc*

  • The program *header* is part of the IMOD software suite*, and reads pertinent details from MRC files. The output should contain the following:

```
Number of columns, rows, sections .....    500    500    10
Map mode ............................    0   (bytes - signed in file)
Start cols, rows, sects, grid x,y,z ...   0    0    0    500    500   10
Pixel spacing (Angstroms).............   77.98    77.98    300.0
Cell angles ..........................   90.000   90.000   90.000
Fast, medium, slow axes ...............   X   Y   Z
Origin on x,y,z ......................   0.000    0.000    0.000
Minimum density ......................   117.00   ( -11.000   in file)
Maximum density ......................   182.00   ( 54.000   in file)
Mean density .........................   152.45   ( 24.446   in file)
tilt angles (original,current) ........  0.0  0.0  0.0  0.0  0.0  0.0
Space group,# extra bytes,idtype,lens .   0    0    0    0
```

Voxel dimensions of the file, in (X Y Z).

0 = 8-bit, unsigned integers
1 = 16-bit, signed integers
2 = 32-bit, floating point
6 = 16-bit, unsigned integers

Voxel sizes, in Angstroms per voxel, in (X Y Z). This dataset has anisotropic voxels, with 7.798 nm/voxel in X and Y, and 30 nm/voxel in Z. Anisotropic voxels are common in SBEM datasets.

* For an exhaustive list of IMOD programs and help documents, visit: http://bio3d.colorado.edu/imod/doc/program_listing.html

Directory Structure:

```
Download_Dataset_1-1
        └── sbem_tiles_2Dbin2.mrc
```

```
/home/aperez/Download_Dataset_1-1

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ header sbem_tiles_2Dbin2.mrc

 RO image file on unit   1 : sbem_tiles_2Dbin2.mrc      Size=       2442 K

 Number of columns, rows, sections .....     500     500      10
 Map mode ..........................      0    (bytes - signed in file)
 Start cols, rows, sects, grid x,y,z ..    0     0      0    500    500     10
 Pixel spacing (Angstroms).............   77.98     77.98     300.0
 Cell angles ..........................   90.000   90.000   90.000
 Fast, medium, slow axes ..............    X     Y     Z
 Origin on x,y,z ......................   0.000      0.000      0.000
 Minimum density ......................   117.00    (  -11.000    in file)
 Maximum density ......................   182.00    (   54.000    in file)
 Mean density .........................   152.45    (   24.446    in file)
 tilt angles (original,current) .......   0.0   0.0   0.0   0.0   0.0   0.0
 Space group,# extra bytes,idtype,lens .     0      0      0      0

     9 Titles :
 raw2mrc: Converted to mrc format.                         19-Feb-12  12:22:31
 NEWSTACK: Images copied                                   21-Feb-12  11:37:20
 NEWSTACK: Images copied                                   21-Feb-12  12:19:51
 NEWSTACK: Images copied             , floated to range    22-Feb-12  08:54:59
 NEWSTACK: Images copied                                   24-Feb-12  12:04:59
 NEWSTACK: Images copied, transformed, floated to means    28-Feb-12  22:16:28
 BOXSTARTEND:    10 starts clipped out and averaged         3-Aug-15  13:50:47
 NEWSTACK: Images copied                                    3-Aug-15  13:52:31
 NEWSTACK: Images copied                                    3-Aug-15  13:52:50


aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ |
```

Example Output

---

3. Create New Directories for Outputs

A. Make a new directory for your training data, using the command *mkdir training_data*

B. In the same manner, make two more directories entitled 'test_data' and 'output'.

> Note: If you accidentally create a directory with the wrong name, you can remove it using the *rmdir* command. For example:
>
> *mkdir training_datuhhh*
> *rmdir training_datuhhh*

Directory Structure:

```
Download_Dataset_1-1
    ├── output
    ├── sbem_tiles_2Dbin2.mrc
    ├── test_data
    └── training_data
```
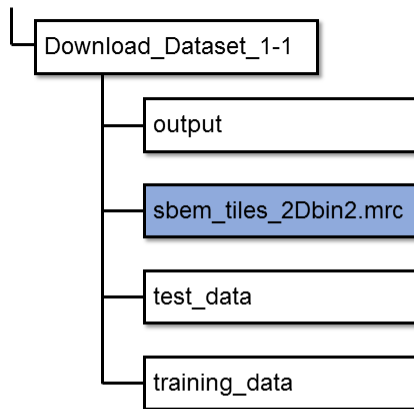
```
/home/aperez/Download_Dataset_1-1

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ mkdir training_data
aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ ls
sbem_tiles_2Dbin2.mrc  training_data
aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ mkdir test_data output
aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ ls
output  sbem_tiles_2Dbin2.mrc  test_data  training_data
aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$
```

Example Output

4. Extract Tiles for Training and Testing

To keep the training time manageable for this course, we will create a training set with two of the ten tiles contained in *sbem_tiles_2Dbin2.mrc*. For reference, training a CHM classifier with a full set of fifty, 500 x 500 pixel tiles can take over 24 hours.

A. Pick two numbers from 1-10 to specify the tiles that will serve as training data. These can be your favorite numbers, randomly chosen, etc.

B. Extract these two tiles to a new MRC stack using the IMOD program *newstack*. For example, if you chose the numbers 5 and 8, the command would be:
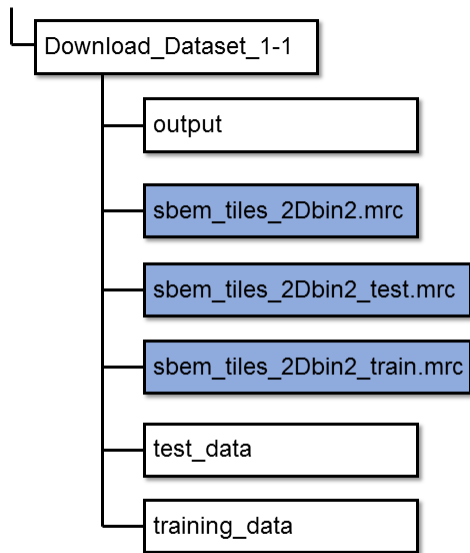
newstack –secs 5,8 –fromone sbem_tiles_2Dbin2.mrc sbem_tiles_2Dbin2_train.mrc

To understand the meaning of these arguments and the *newstack* syntax, view the newstack man page in the terminal by typing *man newstack*

C. Create a second stack containing the test data. These are the tiles that will be classified using the CHM model trained from the tiles in *sbem_tiles_2Dbin2_train.mrc*. We will use all eight leftover tiles as test data. In this example, the command would be:

newstack –secs 1,2,3,4,6,7,9,10 –fromone sbem_tiles_2Dbin2.mrc sbem_tiles_2Dbin2_test.mrc

Directory Structure:

- Download_Dataset_1-1
  - output
  - sbem_tiles_2Dbin2.mrc
  - sbem_tiles_2Dbin2_test.mrc
  - sbem_tiles_2Dbin2_train.mrc
  - test_data
  - training_data

```
/home/aperez/Download_Dataset_1-1

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ newstack -secs 5,8 -fromone sbem_tiles_2Dbin2.mrc sbem_tiles_2Dbin2_train.mrc

 RO image file on unit   1 : sbem_tiles_2Dbin2.mrc      Size=      2442 K

 Number of columns, rows, sections .....      500      500       10
 Map mode .........................        0   (bytes - signed in file)
 Start cols, rows, sects, grid x,y,z ...    0     0     0    500    500    10
 Pixel spacing (Angstroms)..............   77.98     77.98     300.0
 Cell angles ...........................   90.000   90.000   90.000
 Fast, medium, slow axes ...............    X    Y    Z
 Origin on x,y,z .......................   0.000     0.000     0.000
 Minimum density .......................  117.00   (  -11.000    in file)
 Maximum density .......................  182.00   (   54.000    in file)
 Mean density ..........................  152.45   (   24.446    in file)
 tilt angles (original,current) ........  0.0  0.0   0.0  0.0   0.0  0.0
 Space group,# extra bytes,idtype,lens .    0         0         0         0

     9 Titles :
raw2mrc: Converted to mrc format.                           19-Feb-12  12:22:31
NEWSTACK: Images copied                                     21-Feb-12  11:37:20
NEWSTACK: Images copied                                     21-Feb-12  12:19:51
NEWSTACK: Images copied                , floated to range   22-Feb-12  08:54:59
NEWSTACK: Images copied                                     24-Feb-12  12:04:59
NEWSTACK: Images copied, transformed, floated to means      28-Feb-12  22:16:28
BOXSTARTEND:    10 starts clipped out and averaged          3-Aug-15  13:50:47
NEWSTACK: Images copied                                     3-Aug-15  13:52:31
NEWSTACK: Images copied                                     3-Aug-15  13:52:50

 NEW image file on unit   2 : sbem_tiles_2Dbin2_train.mrc
 section   input min&max       output min&max   &   mean
    0     122.00    178.00    122.00    178.00    154.11
    1     119.00    180.00    119.00    180.00    150.71

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ |
```

Example Output

---

5. Create IMOD Model File for Training Labels

A. Open the training data stack in 3dmod: *3dmod sbem_tiles_2Dbin2_train.mrc*

B. In this course, we will be generating training data for mitochondria. Using the techniques described in the preceding lecture, manually segment all instances of mitochondria in the two tiles of your training data stack.

As an additional reference, you can view the file:
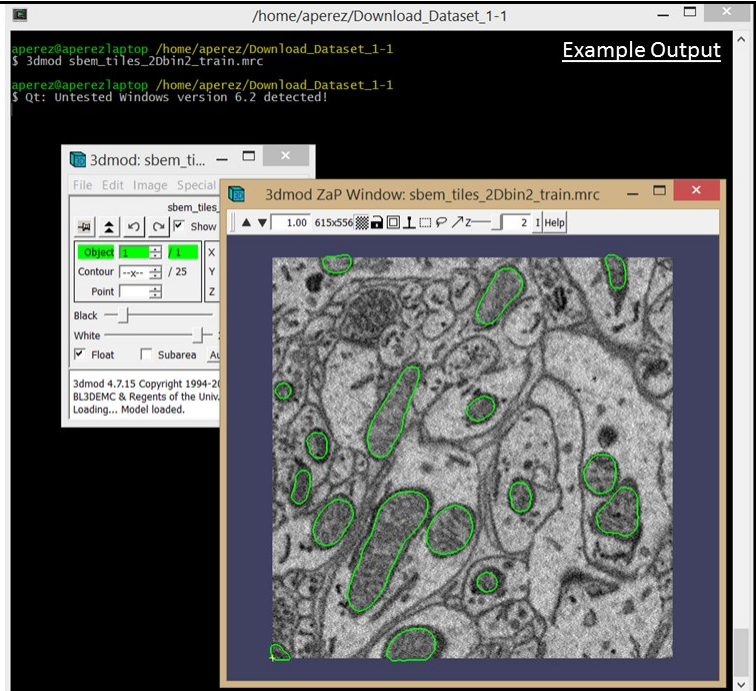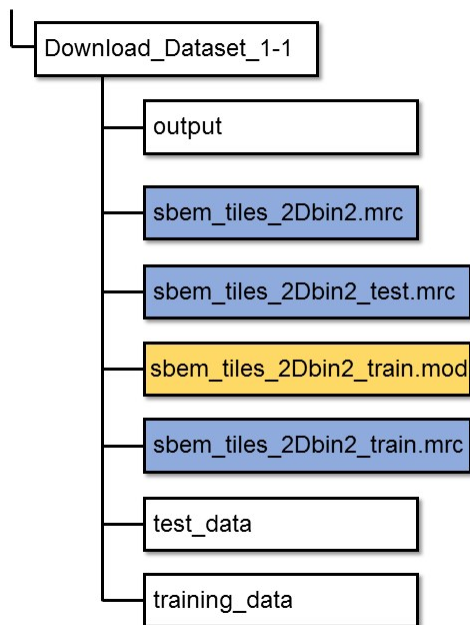
Generating_CHM_training_data_with_IMOD.pdf

Or visit the following URLs:

http://bio3d.colorado.edu/imod/doc/3dmodguide.html – Detailed information about manual segmentation and model file structure in IMOD.

https://www.youtube.com/watch?v=BsNSVLIQ-cE – Useful video illustrating the use of IMOD's Drawing Tools

C. Save the model file as *sbem_tiles_2Dbin2_train.mod*

Directory Structure:

```
┌─ Download_Dataset_1-1
   ├─ output
   ├─ sbem_tiles_2Dbin2.mrc
   ├─ sbem_tiles_2Dbin2_test.mrc
   ├─ sbem_tiles_2Dbin2_train.mod
   ├─ sbem_tiles_2Dbin2_train.mrc
   ├─ test_data
   └─ training_data
```


6. Generate a Training Label Stack

The goal of this step is to generate a new MRC stack with the same dimensions as the training images we previously segmented. However, in this stack, all pixels inside of the traced contours will have values of one, and all pixels outside of the traced contours will have values of zero. In this way, a binary label stack will be created that will serve to tell the CHM training algorithm where mitochondria are. We will use the IMOD program *imodmop* to generate this stack.

A.   Create the binary label stack using the following command:

imodmop –mask 1 sbem_tiles_2Dbin2_train.mod sbem_tiles_2Dbin2_train.mrc sbem_tiles_2Dbin2_labels.mrc

To understand the meaning of these arguments and the *imodmop* syntax, view the imodmop man page in the terminal by typing *man imodmop*

B.   Visualize the label stack using 3dmod:

3dmod sbem_tiles_2Dbin2_labels.mrc

C.   In the 3dmod window, select Image → Pixel View. Check the 'File Value' box. Enter Model mode. Right-click (or left-click, depending on OS) on some of the white blobs in the ZaP window, and confirm that their pixels are one-valued.
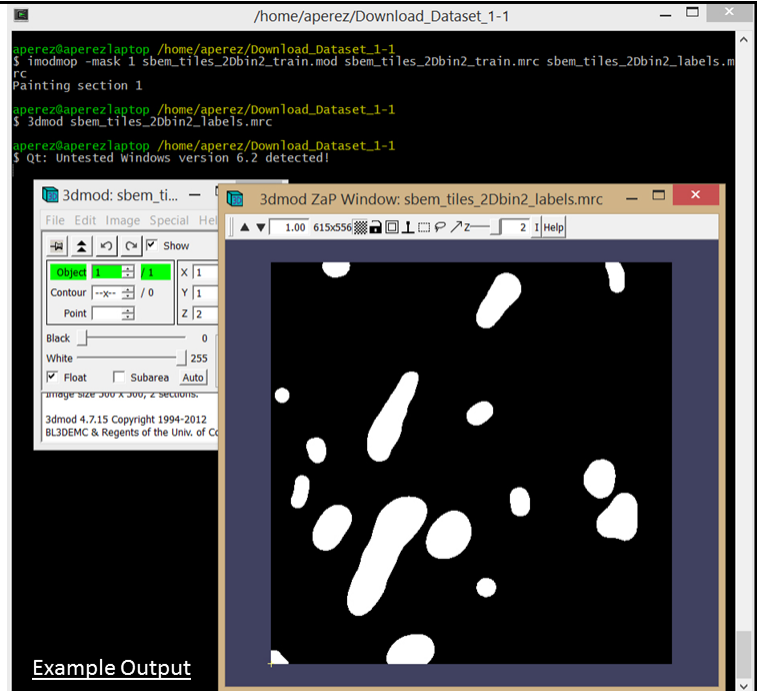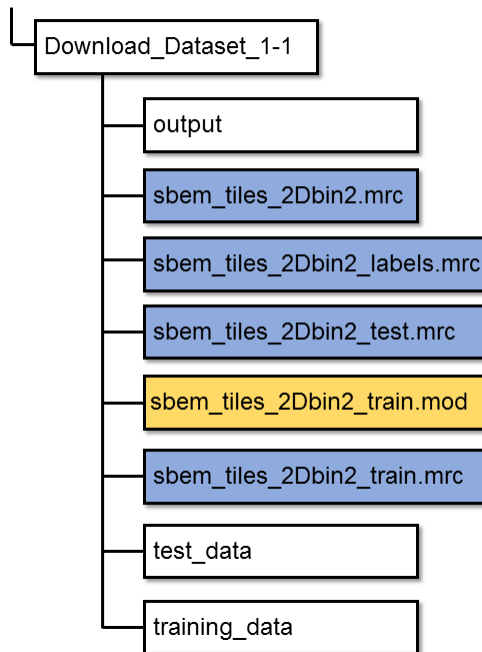
7. Generate Training Image and Label PNGs

The CHM algorithm requires inputs in the form of PNG files. Thus, before we can continue to job submission with the SLASH portal, we must first convert from MRC to PNG.

A. First, we need to make sub-directories under the *training_data* folder for images and labels:

mkdir training_data/images
mkdir training_data/labels

B. Next, convert the MRC stack of raw training images to individually numbered PNGs using the IMOD program *mrc2tif*. The '-p' argument forces conversion to PNG, rather than TIF:

mrc2tif –p sbem_tiles_2Dbin2_train.mrc training_data/images/image

C. The SLASH portal requires training images to have simple, sequentially numbered filenames (i.e. 01.png, 02.png). We can perform this renaming using the following commands:

cd training_data/images
N=1; for file in *.png; do new=$(printf "%02d.png" $N); mv $file $new; let N=N+1; done

D. Repeat steps B-C for the binary label stack. The output PNGs should go to training_data/labels. See the next slide for the exact commands necessary to do this.

Directory Structure:
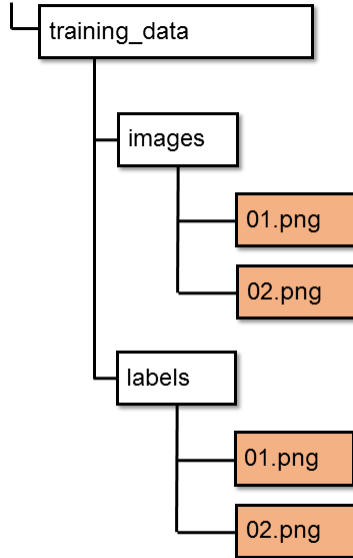
```
aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ mkdir training_data/images

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ mkdir training_data/labels

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ mrc2tif -p sbem_tiles_2Dbin2_train.mrc training_data/images/image
Qt: Untested Windows version 6.2 detected!
Writing PNG images. ..

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ mrc2tif -p sbem_tiles_2Dbin2_labels.mrc training_data/labels/label
Qt: Untested Windows version 6.2 detected!
Writing PNG images. ..

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ cd training_data/images

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1/training_data/images
$ ls
image.000.png   image.001.png

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1/training_data/images
$ N=1; for file in *.png; do new=$(printf "%02d.png" $N); mv $file $new; let N=N+1; done

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1/training_data/images
$ ls
01.png   02.png

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1/training_data/images
$ cd ../labels

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1/training_data/labels
$ N=1; for file in *.png; do new=$(printf "%02d.png" $N); mv $file $new; let N=N+1; done

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1/training_data/labels
$ ls
01.png   02.png

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1/training_data/labels
$ cd ../..

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ |
```
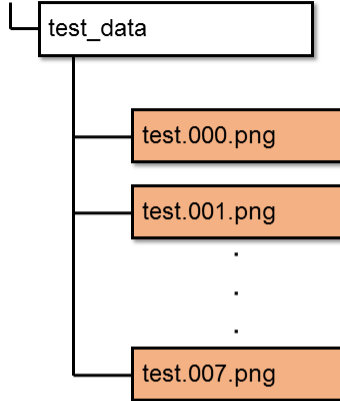
8. Generate Test Image PNGs

A. Finally, we need to create PNG files of the test images:

mrc2tif –p sbem_tiles_2Dbin2_test.mrc test_data/test

Unlike the training PNGs, the test PNGs do not have any restrictions on filename conventions, so we can leave the filenames as they are.

Directory Structure:

```
test_data
    ├── test.000.png
    ├── test.001.png
    .
    .
    .
    └── test.007.png
```

```
/home/aperez/Download_Dataset_1-1
aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ mrc2tif -p sbem_tiles_2Dbin2_test.mrc test_data/test
Qt: Untested Windows version 6.2 detected!
Writing PNG images. ........

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ ls test_data/
test.000.png   test.002.png   test.004.png   test.006.png
test.001.png   test.003.png   test.005.png   test.007.png

aperez@aperezlaptop /home/aperez/Download_Dataset_1-1
$ |
```
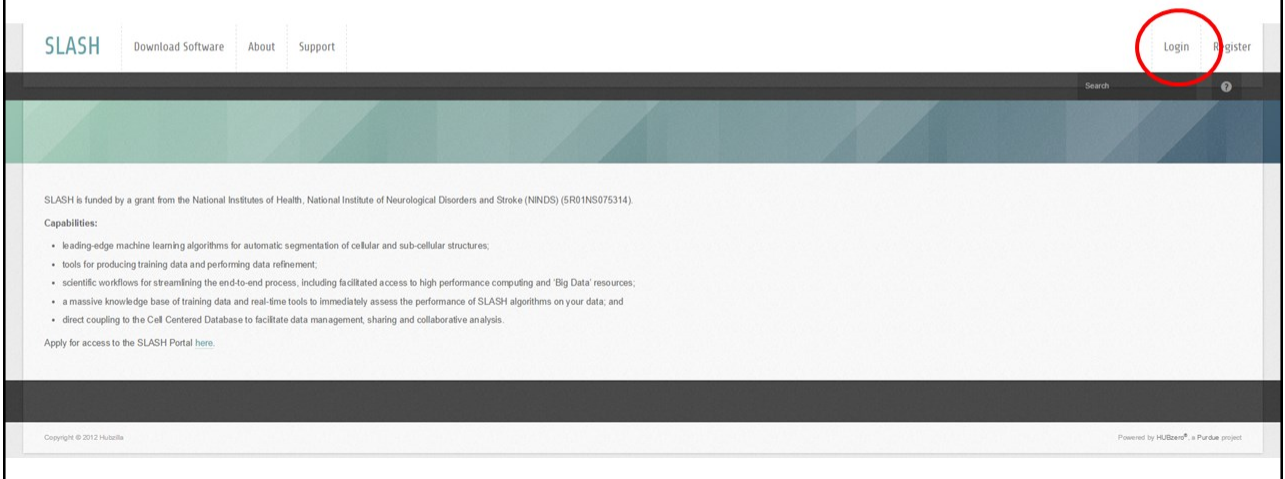Example Output

9a. Run a CHM Train Job through the SQUIRE Portal

A.   Open your browser, and navigate to http://cellsegmentation.org

B.   Click the 'Login' button, and enter your assigned username and password when prompted

SLASH   Download Software   About   Support   Login   Register

Search

SLASH is funded by a grant from the National Institutes of Health, National Institute of Neurological Disorders and Stroke (NINDS) (5R01NS075314).

Capabilities:

- leading-edge machine learning algorithms for automatic segmentation of cellular and sub-cellular structures;
- tools for producing training data and performing data refinement;
- scientific workflows for streamlining the end-to-end process, including facilitated access to high performance computing and 'Big Data' resources;
- a massive knowledge base of training data and real-time tools to immediately assess the performance of SLASH algorithms on your data; and
- direct coupling to the Cell Centered Database to facilitate data management, sharing and collaborative analysis.

Apply for access to the SLASH Portal here.

Copyright © 2012 Hubzilla   Powered by HUBzero®, a Purdue project

9b. Run a CHM Train Job through the SQUIRE Portal

C.   Expand the 'Workflow Service' menu and select 'Workflows'.

D.   Under the 'Automated Segmentation' category, select 'CHM Train'

E.   Choose a job name. Something like '<username> train' would be appropriate

F.   Under 'Training Data', select the pre-loaded 'nbcr class training data' workspace file. Click Apply.

G.   Update your email address if it is not correct. Job notification emails will be sent here.

H.   Click 'Submit Job'.

After pressing Submit, you will be automatically re-directed to your 'My Jobs' page. You should see your submitted CHM Train job at the top, with its status set to 'In Queue'. Within a few minutes, the status should change to 'Running' and you will receive a notification email when the job starts.

The pre-loaded training data has two, 500x500 tiles similar to the data you segmented. Training with a dataset of such a size may take up to a couple of hours.

10. Run a CHM Test Job through the SQUIRE Portal

A. Go back to the 'Workflows' listing under the 'Workflow Service' tab.

B. Under 'Automated Segmentation', select the 'CHM' workflow.

C. Choose a job name. Something like 'nbcr01 test' would be appropriate.

D. Under 'Trained Model', select the workspace file corresponding to the model you just submitted the train job for. Even though this job hasn't finished, you can queue up the test job so that it will run when the training job is complete.

E. Under 'Input Image', select the 'nbcr class testing data', which is a set of test images that have been pre-loaded for this session.

F. Update your email address if necessary, and press 'Submit Job'.