

Assignment 3 Challenge Question

Due: Friday, November 15, 2024 at 1:00 PM

Problem by Laura Miron and Adrian Sanborn
Data by Adrian Sanborn and Cynthia Hao

In this challenge question, we have provided microscopy images of fixed human bone cancer cells (U-2 OS) whose nuclei have been stained with the Hoechst fluorescent dye. Your task is to segment the nuclei and compute statistics about their characteristics. The purpose of this problem is to become familiar with the programmatic methods for analyzing image files. We recommend using a conda environment to keep track of the packages you use – see software handout for details.

Files and Setup

Within this downloaded directory, we have provided you with four `.tif` files, called `nucleus_10x_img*.tif`. To view the images, use [ImageJ](#) or any other program for scientific images. We provide a function in the starter code called `load_image` to view a TIF file and load it as a numpy array.

Functions provided in the [scipy](#), [scikit-image](#), [Pillow](#), or [Open-CV](#) packages may be useful (but not required). If you choose to use these functions, you can install them by typing in your conda environment:

```
$ pip install scipy
$ pip install scikit-image
$ pip install Pillow
$ pip install opencv-python
```

Note: You may have to use `pip3` instead of `pip`, depending on your conda environment.

Tasks

1. Segment the nuclei in the given images, and display the segmentation as a binary mask. You may use any method you like, including any of these steps:
 - a. **Preprocessing:** smooth, sharpen, or other operations.
 - b. Compute an appropriate threshold value and convert the image to a binary mask. See also **Otsu's method** or **adaptive thresholding** from [scikit-image](#).
 - c. **Fill holes**, e.g. using [scipy.ndimage.morphology.binary_fill_holes](#).
 - d. Separate objects using a [watershed algorithm](#).
 - e. **Label objects**, e.g. using [scipy.ndimage.label](#)

2. Compute the cross-sectional area of each nucleus and plot a histogram of the areas. Each pixel is roughly 0.4 micrometers across.
3. Compute the total Hoechst intensity of each nucleus — this gives a measure of DNA content — and plot a histogram of the intensities.
4. Create a scatter plot of the areas versus the intensities.

Submission

Append the following to your Assignment 3 pdf submission on gradescope:

1. A binary segmented image of each of the four input files, e.g. nuclei white, background black.
2. Your histograms of areas and intensities.
3. Your scatter plot of area versus intensity.
4. A complete explanation of your analysis pipeline. Illustrate each step of your pipeline with example images. Why did you design it this way? **Note:** You are not required to submit your code, but feel free to include screenshots of your code in addition to the example images.
5. An explanation of other approaches you tried. Why did they perform more poorly? Illustrate with example images.
6. A brief discussion interpreting any trends you observe in the nucleus area and intensity data. Are there aspects of the pipeline that could still be improved, and how would this impact these results?