02-261 – Quantitative Cell and Molecular Biology Lab

Programming Lab: High Content Screening

In microscopy experiments, we cannot simply rely on the appearance of images to make strong conclusions. We require computational methods to quantify the observed phenomena. In some situations, a single measurement is not sufficient to capture the variability of an observation. In this case, researchers may use machine learning methods to identify experimental outcomes. To do so, one must calculate features which are numerical descriptors of the images. Then a model can be learned to predict the labels of images based on the values of these features. In this Lab, you will design and program feature calculation methods and assess the accuracy of models learned using those features.

You have been provided with the images you generated last week. Refer to the naming scheme from the previous lab for labels.

Your task is to design a set of at least five types of image features allowing you to distinguish between different classes using a RandomForest classifier.

Field-based Classification Tasks:

1. Learn a model to predict the treatment for a field in your section.
2. Learn a model to predict which group generated the field image in your section.
3. Learn a model to predict the day on which the field image was generated.

Note: You may use the same five features for each task. RandomForest has the capability to select features for you thereby ignoring features which are irrelevant to the classification task.

(12 unit only) Object-Level Classification Task:

Modify your feature calculation to take as input segmented objects. Segment the fields into objects around nuclei and learn a classifier to predict the treatment for objects from images in your section.

Programming Assignment: For each classification problem, you will submit the following:

1. Feature Calculation Description – Describe the process for the calculation of features including images or figures where helpful. Explain why you chose these features.
2. Show the feature importance for each feature. Comment on importances relative to your expectations. Why do you think some features were more important than others?
3. Generate confusion matrices for each classification task.
4. Conclusions and Error discussion: Describe your classification results and their implications on this experiment. For a few misclassifications, look at fields which were misclassified and qualitatively describe differences between those images and fields which were correctly classified. Describe whether these issues could be addressed with better imaging/laboratory technique or if you could address these issues computationally through better feature calculations.
5. Commented code.

Submission: You will submit your assignment in two parts. The first part should be a PDF. The second submission will be a python file named uniquely for you.

Potentially Useful information:

Hoechst 33342

Matplotlib is a library designed for making nice plots.

Numpy is a library designed for math with a special emphasis on arrays.

Scipy is a library built around numpy for specialized scientific tasks.

Skimage is a library for image processing tasks.

You may or may not need to use any of these sources of information, but they may prove useful.

Loading Images with Scipy

Contrast Stretching with skimage

Image processing with Scipy

Otsu Thresholding with Scipy

Watershed Segmentation with Scipy

Gallery of matplotlib plots with Examples

More segmentation

Calculating size of objects

Edge detection

Feature Calculation