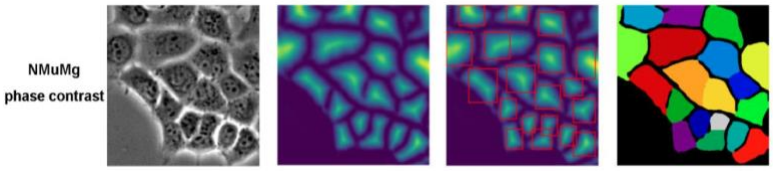
For this semester's project, I am exploring the application of deep learning (convolutional neural networks, *CNN* for short) models and computer vision (*CV*) methods for cancer cell tracking. Cell tracking is a crucial step in our lab's projects which are to discover the environmental conditions that cause tumor cells' change of behavior, most importantly tumor cell invasion. To accomplish this task, we need to relate the cell shape and motility behavior with the specific environmental condition because this specific cell behavior is a marker for tumor cell invasion. Currently, cell tracking is done manually which is time consuming and prone to human error through constantly repeating the task of distinguishing cells, adjusting parameters for software, etc. Thus, automation of tracking cancer cells is of great importance to the project as it will reduce total time spent on this crucial, labor-intensive step and increase accuracy of our data.

The project is split into two related goals: first and most importantly, I will train two CNNs and perform the CV technique of the watershed algorithm to track cells. This process is outlined by a paper from one of our collaborators on the project: https://arxiv.org/abs/1803.10829. The end product of this goal is to have the process functioning for our lab's data as shown below (source: link above).



Once the first goal is accomplished such that our lab members can track cells with ease and efficiency, I will research other methods to the tracking problem and apply them to our lab's data to fine-tune and improve upon the cell tracking to our needs if time permits. Hence, the second goal is to compare/contrast the pros and cons of different approaches to the cell tracking problem. There are many as highlighted in https://www.nature.com/articles/nmeth.4473.

For both of these goals, I am doing background research on neural networks and computer vision methods, as there are several different approaches to cell tracking that have their advantages depending on the specific conditions of the datasets. I'm currently past the background research stage for the most part. Now I will be training multiple neural networks which involves creating suitable datasets from the specifications of the network and using them as inputs to improve the CNN's results; this is the general process of deep learning.

Specifically, I will be working with the Python language along with Jupyter Notebook, Keras, and TensorFlow which are important technologies used by data scientists for machine learning projects. Along with the standard deep learning algorithms, I'll be working with algorithms such as the well - known *Adam* method: https://arxiv.org/abs/1412.6980 for optimization.

Additionally, I will be using source control (GitHub), which is a crucial tool used by software engineers, to maintain the project. For best case scenario, I'd like to make both training the CNNs and imaging as simple as possible for the lab members, which means having them input an image and get the final output without working directly with the background neural network code. Lastly, if there is time left, I will continue altering and optimizing the cell tracking program, taking inspiration from the many other cell tracking methods as shown in goal #2, to further improve our lab's data analysis process if needed.

For the final presentation, I will discuss (1) the results of applying the process from goal #1 to our data and (2) the current eco-system of approaches to the cell-tracking challenge, their pros/cons, and their possible benefits to our data as discussed in goal #2. I will also do progress reports and other presentations where I may, for example, discuss the ideas behind the deep learning process to the lab members.