[I only saw two emails in the comments; was there supposed to be a third recipient as well?]

This generally sounds like an interesting project.

However, trying to segment images based on "raw pixels" (or downsampled images) is probably not going to work very well. Instead, I might recommend that you use a Bag of Words representation for this task (something that we are going to talk about in class in a couple of weeks).

The way this works is by first extracting sparse features from all the training images and then using k-means over that space to compute a "vocabulary" (basically the location of the means). Once you have a "vocabulary" you can turn images into histograms, where the number of items in each bin is "how many times that "mean" (of the k-means) appeared in this image. The histograms of each image can be compared to one another.

This is a very common pipeline in image classification and something I think you should consider replacing the "middle" of your project. Once you have k-means implemented, and likely discover that it does not work very well on your data, you can

I would prefer you do the Bag of Words pipeline over playing around with hyperparameters (which doesn't make for a pretty "fun" final project). This tutorial seems to have a reasonably good over view of how to implement this using tools like sklearn:

https://kushalvyas.github.io/BOV.html

You should feel free to use this tutorial as your "comparison" component if you'd like, but for your main implementation, I generally expect you to try to implement much of this yourself (except the feature detection & description). It sounds like most of this you were already planning to implement; the real difference is that you'll be using these features, rather than a downsampled image.

Let me know over Piazza if you have questions.

Best,

Prof. Stein