Project Pitch Review
Software Engineering for Scientists
by Jacob Stanley

Overall this seems like a great project. Your objective is well-defined and you have a clear idea of the type of processing you'll need to do for the data. Good work!

Scientific overview:

1. Is any part of the scientific background unclear or confusing? If so, what additional information would be helpful?
   I think you did a great job framing the background of the project---explaining the motivation, giving the right amount of biological context, and outlining the experimental design. There was one thing I was confused by though: my understanding is that the principle goal is to verify that the Riboglow images can be used to recapitulate the granule locations identified by the protein marker (a proof of concept, of sorts), however, it wasn't clear to me what RNA marker you were going to be using to do so. Do you expect that all RNA would be represented in all granules or are you planning to Riboglow label many different RNA's, to improve the chances that the Riboglow will be localized to all granules?
2. What part of the project do you find interesting?
   I think stress granules are an interesting area of research and I've wanted to know how they form. The Riboglow tag seems like a great way to study them! Also, I thought your metaphor of the black dog with the lighted collar was great. ☺

Architecture:

1. What components of the groups' proposed architecture do you think is a good design?
   Your software workflow is very clear and all the steps make sense. Good work!
2. What, if any, are some limitations of the current proposed architecture that you see?
   Generally, I think the problem you're tackling can be well addressed computationally. The 9-step architecture that you outline toward the end of your talk hits all the major points.
3. What components of the architecture do you think might be missing?
   I don't know how the cell segmentation function works, but presumably it has some tunable parameters which may impact how the granules are detected. Similarly, I imagine the granule detection function is going to have tunable parameters that will impact its performance. Therefore, it's possible you will want to iterate over steps 3 through 6 in order to optimize the mask overlapping, with respect to the segmentation and granule detection process. So, this would be an internal loop in your workflow, which would require some sort of quantitative overlap metric like I discuss below (#2).

Technical Implementation:

1. Do the proposed data types seem suitable for the proposed software design? If not, what could the group improve?

I think using numpy arrays for the image masks is an excellent choice. Numpy arrays (like matlab matricies) are designed to use matrix operations. So, comparing masks is as simple as multiplying together the two. This will be much more efficient than looping through the objects and comparing them element by element.

2. Do you anticipate any computational bottlenecks not described by the group?
   I think a potential bottleneck would be the evaluation of imperfect overlap of masks. While it's a good idea to save them so they can be visually inspected, this is not a high-throughput process. My suspicion is that you may have quite a few poor overlapping masks (just the messy nature of data), so you may want to consider coming up with some kind of quantitative metric to evaluate the overlap in an automated fashion, and if the metric is below some conservative threshold, you flag this particular instance in the metadata.
3. Does the delineation of the code development between developers make sense or do you anticipate any code conflicts when merging the code? Does an alternative division of labor seem more suitable?
   There was no explicit delineation of the code between developers, but you've clearly elucidated the different modules of the project, so I imagine you'll divide those between you two.