Project outline

1. Project overview

I am working in biophysics :)
We will have video and tracked positional data on cells during wound healing and a process called morphogenesis.

2. Data description

We will use video and tracked positional data of cells during wound healing and morphogenesis. The dataset will allow us to analyze both spatial and temporal dynamics of individual cells and their protein clusters.

Expected features/variables:

1) Protein cluster locations and intensity
2) Cell-wall data
3) Staining, ie. protein type
4) Epithelial features (size, shape of the animal as a whole)
5) Frame number/time point
6) Imaging conditions (e.g., magnification, fluorescence channel)

Expected dataset size:
Several thousand to tens of thousands of cell observations across multiple time-lapse experiments.

Potential issues:
Data quality may vary due to imaging noise, focus drift, or cell overlap, which could complicate segmentation and tracking. Availability of well-annotated ground-truth labels is limited, making supervised training challenging.

3. Machine learning approach

We expect to estimate spatial and biochemical features of the cells in the image data. For example segmenting different protein clusters that are moving in and out of focus, which would be hard using classical computing.

We will most likely use some pre-trained models, and maybe, after hand-labeling some data, fine-tune it using supervised ML.

4. Implementation plan

I am pretty confident in PyTorch from my Master thesis.

Evaluating the performance would be a tradeoff between time saved (if we were hand-labeling everything) and performance. I think the validation from our colleagues from biology would be part of our accuracy metric.

5. Potential use of foundation models

Hard to see how, in this very hard science it is not more of a liability at the moment. If you allow me to be meta for a moment, I could use LLM's to solve assignments for courses. This answer could have come from a foundation model (it did not, but it might have saved me time, ie. helped my project)

6. Challenges and considerations

Main challenges:

1) Not having a ground-truth.
2) Explaining/justifying its use for biologists
3) Being bored by necessary hand-labeling of data

Ethical:

As always, using black-box methods comes with explainability concerns. We will have to be certain we can vouch for any data processing and conclusions!

7. Expected outcomes and impact

I hope to achieve precise segmentation and tracking of cells and dynamic protein clusters during wound healing and morphogenesis, reducing the need for labor-intensive manual annotation. This will enable quantitative, high-throughput analysis of spatial and biochemical cellular features that are otherwise difficult to measure.

These results could advance biophysics and developmental biology by providing reproducible, data-driven insights into cellular dynamics and morphogenetic processes, potentially revealing phenomena that were previously unobservable.

8. Next steps

1) I have already started
2) Mainly talking to our collaborators on teh format/quality of the data
3) Trying to circumvent the garbage in / garbage out problem
4) Python, Torch, ssh to a server with NVIDIA RTX PRO™ 6000