

Deep learning for analysing immune cell interactions

Leonore Papaloizos - 2264897v

December 20, 2019

1 Proposal

1.1 Motivation

The initiation of an immune response in our immune system depends on the interaction strength between different types of immune cells. Interactions between immune cells can be enhanced or inhibited by the application of drugs. Applying deep learning to the analysis of microscope images of t-cells and dendritic cells would allow us to explore cell interaction under different drug conditions.

1.2 Aims

This project will run microscope images of immune cells through different algorithms in order to evaluate whether different levels of interactions can be uncovered in this way. A convolutional autoencoder will be used for dimensionality reduction in order to be associated with visualisation techniques such as t-sne and UMAP. Image segmentation techniques will be explored for calculating direct overlap between t-cells and dendritic cells. These two techniques could also be combined for further research. The effectiveness of the algorithms for uncovering the impact of drugs on immune cells can be evaluated by comparing the results with truth labels provided by biological researchers.

2 Progress

- Getting to grips with the framework of choice: the project will be implemented in Python, using Keras for model development.
- Background research conducted on autoencoders and convolutional neural networks for image processing.
- Background research conducted on techniques for image segmentation.
- Wrote code to process large cell images into usable dataset with Numpy compression files.
- Developed a convolutional autoencoder model that outputs a lower-dimension reconstructed image.
- Used the encoded bottleneck images from the autoencoder model to evaluate whether clusters of cells are found using t-sne and UMAP.
- Explored different image processing methods to evaluate impact on autoencoder and clustering performance.

- Wrote code to visualise t-sne and UMAP output in a plot, as well as algorithm steps in a GIF.
- Used K-Means as an image segmentation techniques to get masks from images of t-cells and dendritic cells and calculated overlap.
- Used masks obtained with K-Means as a processing technique by masking out redundant data from the images.

3 Problems and risks

3.1 Problems

The following issues were encountered in the project so far.

- The provided dataset is very large and cannot be used all at once. Using a chunk too large runs out of memory both locally and on Google Colab. Subsets of it have to be used for computations.
- Compressing the dataset has been a struggle as my machine runs out of memory when trying to, but it is also a very large dataset to transfer to another machine.
- The autoencoder was not outputting a good enough reconstruction or learning anything. This was because of a lack of appropriate activation function.
- The autoencoder was only outputting black images, even if it was learning something. This was because of issues with pass-in arrays and unclear documentation in Numpy.
- The combination of autoencoder and t-sne/UMAP has not found successful clusters around different drug conditions.
- The processing method which performs the best in terms of autoencoder reconstruction and t-sne clustering also has wild loss scores in the millions.

3.2 Risks

- Many different autoencoder structures and processing methods can be combined to be explored, which is time-consuming. **Mitigation:** will conduct research to narrow down to one processing method and two parameters to tune to evaluate by the start of the semester.
- There are many techniques I could explore to analyse images to find if they more or less overlap. **Mitigation:** will narrow it down to one other technique on top of K-Means by the start of the semester.
- The algorithms currently being developed could be biased towards the subset of the data being used. **Mitigation:** pick three subsets of the data to run the developed algorithms against.
- Unsure how to evaluate success of the project. **Mitigation:** will liaise with biological researcher providing dataset to gather expectations as well as expected metrics for given images.

4 Plan

4.1 Semester 2

- Week 1-2: Complete autoencoder tuning and choice of image processing. **Deliverable:** code for an autoencoder model and processing method with explanations for the choices made.
- Week 1-3: Liaise with biological researcher to explore current tools being used to uncover overlaps between immune cells. Collate results obtained by their research for future evaluation. **Deliverable:** detailed evaluation plan, report on current research methods for evaluating interaction, and precise Excel tables of research metrics.
- Week 3-4: Finalise the image segmentation technique for calculating overlap and obtaining masks. **Deliverable:** code for the image segmentation technique, with an explanation of how it works and why it was chosen.
- Week 5: Investigate using the masks obtained for a UNet like autoencoder structure. **Deliverable:** code for the UNet structure, working with the cell dataset.
- Week 6-7: Construct three separate datasets (DMSO, one CK, mixed), and analyse them with the algorithms developed. Explore other labels associated with the dataset (compound, compound concentration). **Deliverable:** compressed dataset files, plots visualising the results and performance of the models.
- Week 8: Compare results obtained through own algorithms with results obtained through biological research. **Deliverable:** quantitative measures of comparison with ground truth values accompanied with visual plots.
- Week 9: Final clean up of code and packaging up. **Deliverable:** polished research code, model code available in usable tools.
- Week 8-10: Dissertation write up. **Deliverable:** first draft submitted to supervisor two weeks before final deadline.

5 Ethics and data

This project does not involve human subjects or data. No approval required.