Generative Model for Imputing Imaging Mass Spectrometry from Serial Two-Photon Tomography

Kevin Yang
University of British Columbia
Vancouver, BC V6T 1Z4
`kyang@bccrc.ca`

Abstract

Serial Two-Photon Tomography (STPT) and Imaging Mass Spectrometry (IMC) are two popular imaging techniques in tumour analysis. STPT images describe tumour morphology, whereas IMC images describe protein abundance. Having both data modalities for the same tissue sample is often beneficial in clinical applications, as understanding the tumour landscape from both a morphological and proteomic perspective can inform treatment decisions. However, it is difficult to obtain both data modalities for a single tissue sample. To mitigate this issue, we present a Generative Model that, for the same tissue sample, only requires STPT images to impute corresponding IMC images. 18 aligned STPT and IMC physical sections have been identified and have been used for preliminary training; we are currently reaching out to relevant personnel to obtain more data. A naïve model based on convolutional layers has been trained to predict IMC from STPT images with poor results.

1. Project Progress

1.1. Core Tasks

Core tasks required to complete the project are 1) obtain aligned IMC/STPT images 2) try a simple convolutional model 3) explore more complex models.

18 aligned STPT and IMC physical sections have been identified. Each physical section contains two STPT images corresponding to two imaging depths (30 μm and 38 μm) (Bressan et al., 2021), and a 40-channel IMC image dataset - each channel corresponding to a different protein. The methodology to align raw STPT and IMC images is described in the Methods Section of González-Solares et al. (2021). 18 aligned STPT and IMC images may not be sufficient, so we are currently searching for more aligned images by reaching out to the original authors of the IMAXT project paper (González-Solares et al., 2021).

This reconstruction problem is like the image colorization problem in that we are attempting to predict an image with many channels (IMC) from an image with a fewer number of channels (STPT). In our case, we are not trying to go from 1 value per pixel to 3 values per pixel, as in grayscale to RGB; rather, we are aiming to predict a 40-channel IMC image dataset using only two 4-channel STPT images. A simple model which uses ResNet-18 and a series of convolutional and upsampling layers has been implemented but gave poor results.

Considering this, we intend to explore more complex models. IMC channels may share dependencies, so using an architecture that capitalizes on that is likely to increase accuracy. Also, pre-training the model on a related dataset might give a slight boost in performance.

1.2. Difficulties

The size of each image is restrictive with respect to training time. There is a total of 18 physical sections in the preliminary dataset. Each STPT image is 2.4GB, and there are two STPT images per physical section. On top of that, each IMC image is 668MB, and there are 40 IMC images corresponding to the 40 proteins for each STPT physical section. Unsurprisingly, training on such a large image dataset takes a long time, even with substantial computing power (48 CPUs, 128GB RAM, 1 GPU).

A formidable challenge lies in finding a model that accomplishes the reconstruction task. It is difficult to find a model powerful enough to go from two 4-channel STPT images to an IMC dataset with 40 channels. Perhaps there is simply not enough information in the STPT images alone to accurately reconstruct the entire 40 channel IMC dataset.

2. References

[1] González-Solares, E. A., Dariush, A., González-Fernández, C., Yoldaş, A. K., Sa'd, M. A., Millar, N., Whitmarsh, T., Chornay, N., Falciatori, I., Fatemi, A., Goodwin, D., Kuett, L., Mulvey, C. M., Ribes, M. P., Qosaj, F., Roth, A., Vásquez-García, I., Watson, S. S., Windhager, J., … Walton, N. A. (2021). *The Imaging and Molecular Annotation of Xenografts and Tumours (IMAXT) High Throughput Data and Analysis Infrastructure.* ResearchGate. https://doi.org/10.1101/2021.06.22.448403

[2] Bressan, D., Mulvey, C. M., Qosaj, F., Becker, R., Grimaldi, F., Coffey, S., Vogl, S. L., Kuett, L., Catena, R., Dariush, A., González-Fernández, C., González-Solares, E. A., Sa'd, M. A., Yoldaş, A. K., Whitmarsh, T., Falciatori, I., Watson, S. S., CRUK IMAXT Grand Challenge Team, Joyce, J. A., …Hannon, G. J. (2021). *Exploration and analysis of molecularly annotated, 3D models of breast cancer at single-cell resolution using virtual reality.* bioRxiv. https://doi.org/10.1101/2021.06.28.448342