CNN-based analysis of organoid growth

Shan Zhou
Facebook

shanzhou@stanford.edu

Timothy Daley
Stanford University
Departments of Statistics and Bioengineering

tdaley@stanford.edu

Abstract

High-throughput analysis of imaging data is critical for analysing large data typical of modern biological investigations. Here we investigate a convolutional neural network-based approach to analyse the growth of organoids based off of imaging data.

1. Introduction

Predicting tumor growth rate is a first step in determining treatment options for cancer patients. Fast growing tumors necessarily require more aggressive treatment. It would be beneficial if patients could avoid aggressive treatments when possible.

Alexandra Sockell of the Fordyce and Curtis labs in the Bioengineering and Genetics departments, respectively, of Stanford University has developed a microfluidic device to isolate single cells of a tumor and allow them to grow into organoids within the microwell. Organoids are three dimensional stem cell-like cultures that organize into a "mini-organ" [6], and can be used to study cancer in a more natural environment than traditional cell lines [2]. The objective of her research is to study the mechanisms of tumor growth by subjecting a large number of individual cells to a wide range of treatments and conditions and track their condition by imaging. She has taken 14 days of imaging over 8 conditions. For each day, there are approximate 38,000 well images across all conditions. The number of cells per well is approximately Poisson, with most of the wells not containing any cells, 25% have one cell, and smaller portion have more than 1. We believe that the large number of images should provide a sufficient amount of data and information content to apply deep convolution neural network approach. Our hypothesis is that the state of the cells in the early days should be related to their final state. Therefore our objective to determine whether the early stages of the organoid can predict the growth rate and final state of the organoid.

Previous approaches for high-throughput analysis of organoid imaging data did not look at single-cell microwell level data. Instead they typically relied upon a large number of cells to quantify cell proliferation or death [4], used cell counting assays to calculate growth [7], or used single cell tracking to calculate cell motility [3]. To our knowledge, no deep learning approaches have been proposed to analyse organoid imaging data, despite the large success in deep learning to analysing imaging data across a broad spectrum of applications. We have found successful CNN approaches in related biological analysis such as high content screening [**?**].

The purpose of this project is to provide a proof of principle for CNN-based high-throughput analysis of organoid imaging data.

2. Data

The data is composed of 193×193 greyscale images of 4800 microwells for each of 6 wells, imaged across 14 days. So we have $6 \cdot 4800 = 28,800$ wells with 14 images per well. The images were obtained via computational stitching [5]
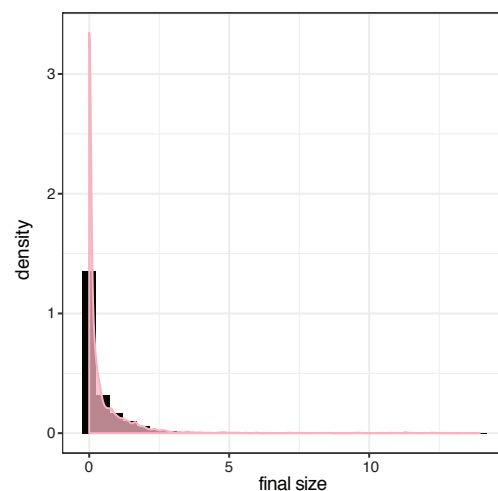

Figure 1. Distribution of normalized final sizes. There is a large peak at zero corresponding to empty wells or cells that died.

of multiple larger images of well. A small number of images appear distorted. We filtered the images by the estimated area of their interior, computed by the hyst2 function in openCV [1], to try to reduce such artifacts. If the estimated area was larger than possible for any of the 14 days, then we removed all images of that well from the data. After filtering we kept 8,763 microwells. Example images are shown in figure (reference needed).

3. Methods

To start we approached the regression problem straightforward. We found difficulties with overfitting using only day 1 on the images from a single well. We therefore decided to first focus on the classification problem, to predict from the early days whether the final day is empty or not, using whether the computed hyst2 area is zero or not as zero and one, respectively. As a proof of principle, we attempted to overtrain a deep convolutional neural network consisting of three convolutional layers, with kernel sizes of 2, 3, and 3, respectively, and channels of 64, 32, 16, respectively. Each convolutional layers was followed by a batch normalization layer, a ReLU layer, and then a max pooling layer with a kernel size of 2. The convolutional layer were followed by a square fully connected layer, followed by a ReLU layer, a fully connected layer with output size of one, and finally a sigmoid for output. We applied this network to all images from a single well, and attempted overfitting on day 1,

References

[1] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[2] J. Drost and H. Clevers. Organoids in cancer research. *Nature Reviews Cancer*, page 1, 2018.

[3] D. T. Geum, B. J. Kim, A. E. Chang, M. S. Hall, and M. Wu. Epidermal growth factor promotes a mesenchymal over an amoeboid motility of MDA-MB-231 cells embedded within a 3D collagen matrix. *The European Physical Journal Plus*, 131(1):8, 2016.

[4] J. Jabs, F. M. Zickgraf, J. Park, S. Wagner, X. Jiang, K. Jechow, K. Kleinheinz, U. H. Toprak, M. A. Schneider, M. Meister, et al. Screening drug effects in patient-derived cancer cells links organoid responses to genome alterations. *Molecular systems biology*, 13(11):955, 2017.

[5] S. Preibisch, S. Saalfeld, and P. Tomancak. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics*, 25(11):1463–1465, 2009.

[6] A. C. Rios and H. Clevers. Imaging organoids: a bright future ahead. *Nature methods*, 15(1):24, 2018.

[7] T. A. Sebrell, B. Sidar, R. Bruns, R. A. Wilkinson, B. Wiedenheft, P. J. Taylor, B. A. Perrino, L. C. Samuelson, J. N. Wilking, and D. Bimczok. Live imaging analysis of human gastric epithelial spheroids reveals spontaneous rupture, rotation and fusion events. *Cell and tissue research*, 371(2):293–307, 2018.