CNN-based analysis of organoid growth

Shan Zhou
Facebook
shanzhou@stanford.edu

Timothy Daley
Stanford University
Departments of Statistics and Bioengineering
tdaley@stanford.edu

Alexandra Sockell
Stanford University
Department of Genetics
asockell@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. However, we have found successful convolutional neural network (CNN) approaches in related bio-

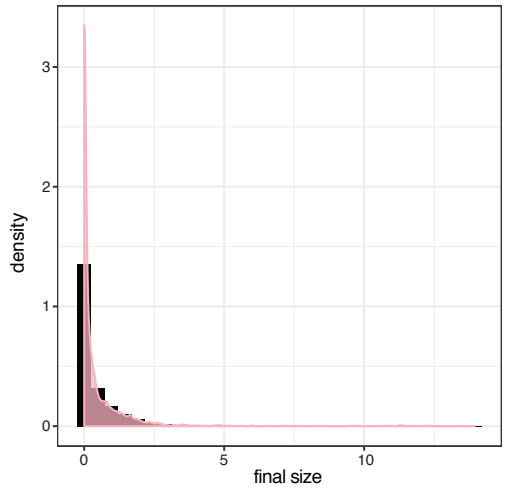

Figure 1. Distribution of normalized final sizes. There is a large peak at zero corresponding to empty wells or cells that died.

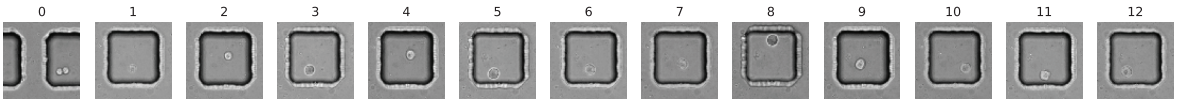

Figure 2. All 14 images from a single microwell.

logical analysis such as high content screening [8], but this is not similar enough for us to compare against.

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 (example shown in figure 2). So we have $6 \cdot 4800 = 28{,}800$ wells with 14 images per well. The images were obtained via computational stitching [5] 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 4.

One difficulty with this project is that we did not obtain all the data at the start. We began with one experiment (4800 microwells), and then obtained data from more experiments as they were processed. We found that some models built on the first experiment did not generalize well, and a major objective is to ensure that our model can is generalizable across biological conditions.

3. Methods

3.1. Proof of principle model

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, day 2, and day 13, using cross-entropy loss. Day 13 was included as a sanity check, since the hyst2 area was computed using the day 13 images. The training loss for days 2 and 13 quickly went to zero, while the loss for day 1 did not go below the loss for random guessing (Figure 3). We therefore excluded days 1 and 0 from further consideration because they do not appear to be informative towards our objective. It may be that day is informative when combined with other days, but not by itself.

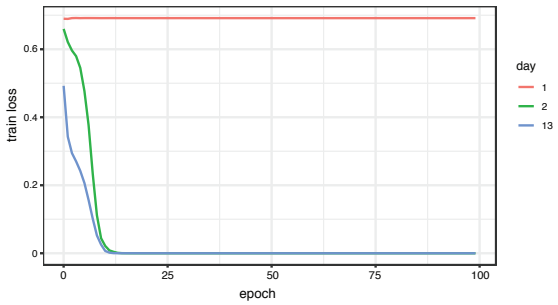

Figure 3. Cross-entropy loss for the training set using days 1, 2, and 13 to predict whether hyst2 area is zero at day 13.

3.2. Deep CNN for multi-day input

After the initial proof of principle we built a classification CNN to predict whether the microwell was empty at the end of the time point (indicating cell death), using hyst2 area as a proxy for whether a cell is empty or not. If the hyst2 area at day 13 is zero then we say that microwell is empty and if the hyst2 area is greater than zero then we say the microwell is non-empty. This CNN had five convolutional layers, all with batch normalization, ReLU, and max pooling following, with channels sizes of 32, 64, 128, 256, and 256 and kernel sizes of 5, 3, 3, 3, 3 (Figure 4). We used cross-entropy as the loss function and the Adam optimizer with learning rate $10^{-}4$. To facilitate generalization of our model we used Dropout regularization with $p = 0.5$. To take into account the temporal nature of our data we used multiple day images as input. We treated each separate image as an input channel (since each image is a single greyscale channel). We initially used days 2 and day 8 images as input, and later used days 2, 5, and 8 as input.

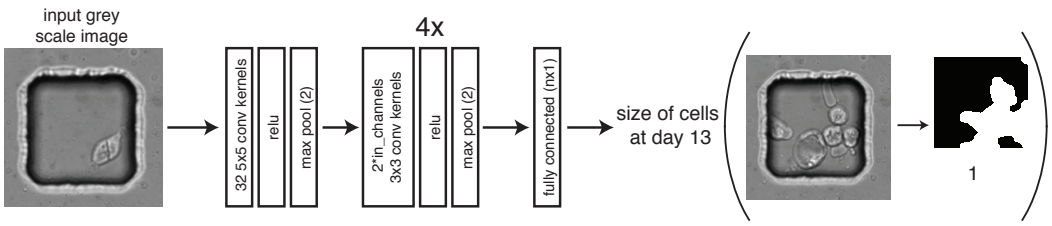

Figure 4. Example workflow of our CNN algorithm. The input is a 193×193 greyscale image of the cell at day one. We pass this through a convolutional neural network to whether the final well is empty, using the hyst2 area as a proxy.

3.3. Pretrained model

We also applied pre-trained models, using transfer learning to train only the final output layer, keeping the other layers constant. We found success with the ResNet 18 model, but not as good as our previously discussed model.

3.4. Data augmentation

Because of the paucity of experiments, we explored data augmentation approaches. Because the primary information is in the center of the image, we hypothesized that random cropping may be a suitable data augmentation strategy. Additionally, the information content of the images should be rotation invariant. Therefore we also explored rotations for data augmentation.

4. Experiments

For classification, first we used one simple 3-conv,1-fully connected layer without any regularization to train 100 images. We obtained overfitting, as the training loss went to 0 and the training accuracy went to 1. However, the validation accuracy remained constant, equal to random guessing. We then added more convolutional layers and more parameters in each of the convolution network, obtaining the model described in section 3.2, and training with more data, 4510 microwells. We obtained 0.19 as the best validation loss , 0.8 as the best training accuracy and 0.6 as the best validation accuracy across all epochs (figure 6). If we run for more epochs, then the training loss goes to zero and the training accuracy goes to one. However, the validation accuracy does not improve. This supports the decision of early stopping. To improve generalization we then added dropout. This improved the validation accuracy slightly, to 0.65. However, the best improvement is when we added more microwells to the training. Then the validation accuracy improved to 0.8. The test accuracy for this model is 0.77 (figure 5). We explored adding data augmentation for this model, but this did not appreciably improve the validation accuracy.

To validate our approach we generated saliency maps for several example images, shown in figure 7. We should note that we did not perform data augmentation on any of the images, so that the shifting and rotation shown in the first and last example images are present in the real data. The saliency maps show a clear indication that the convolutional filters are picking up the signal contained in the microwell. The boundaries of the microwell contain only noise, and even when the images are rotated or shifted, the interior of the microwell is highlighted in the saliency map. This is evidence that our convolutional model is correctly identifying the salient information contained in the images.

4.1. Regression

Now that we were satisfied with the results of the classification problem, we turned back to the regression problem. We took the pretrained ResNet18 for feature generation and

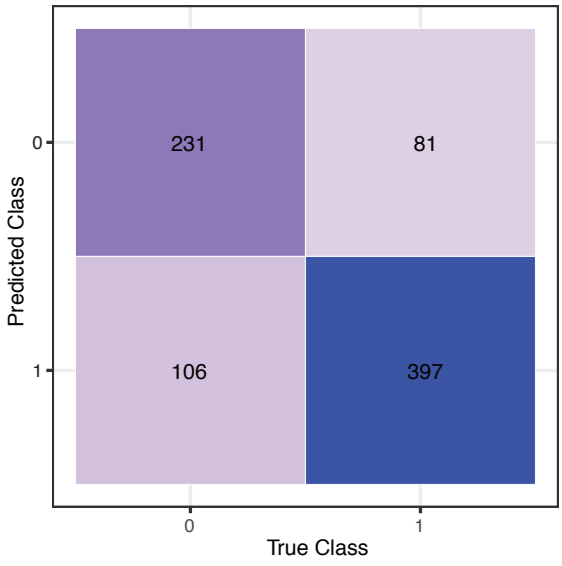

Figure 5. Confusion matrix for the classification problem on the 815 microwell test set.

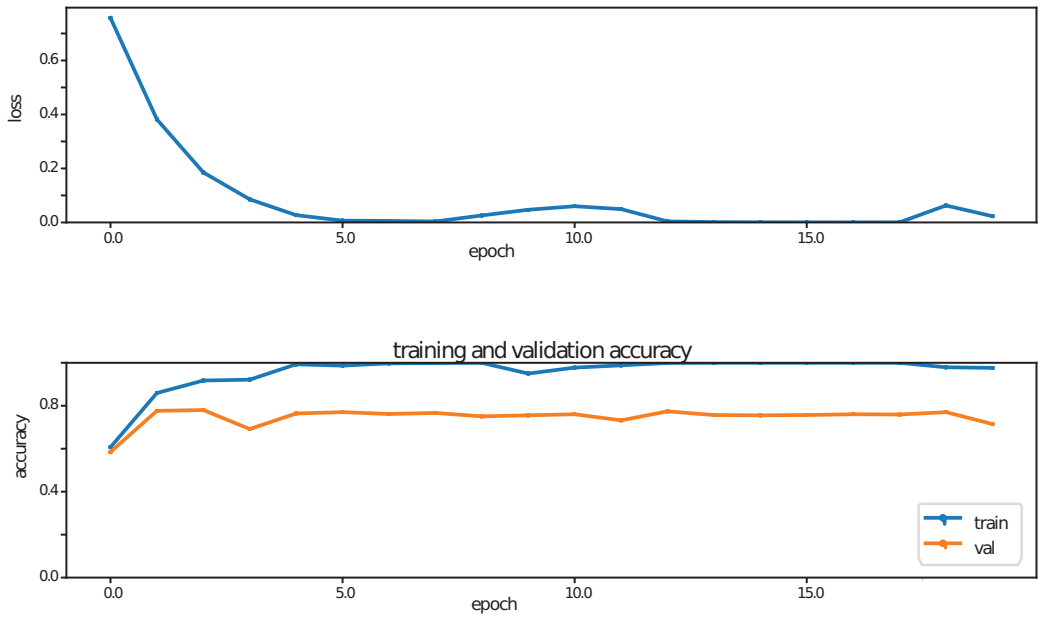

Figure 6. The training cross-entropy loss (top) and training and validation accuracy (bottom) for the classification problem.

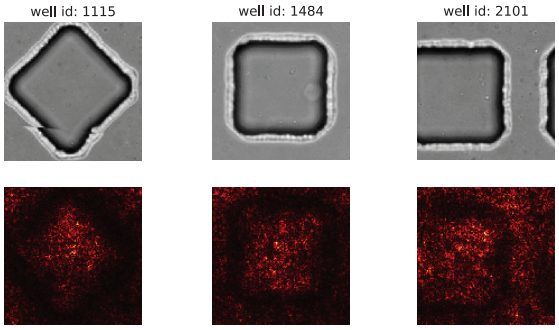

Figure 7. Example images and their corresponding saliency maps.

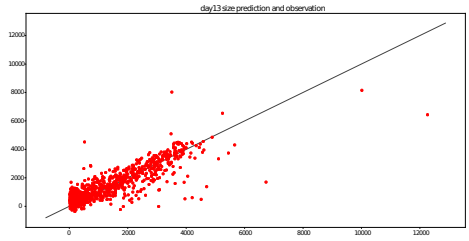

Figure 8. The observed hyst2 area size (x-axis) and the predicted hyst2 area size (y-axis) on the validation set. The correlation is equal to 0.85.

used the Adam Optimizer to train the final layer, with mean square error as the loss function. We looked at only the images that have a final hyst2 area greater than zero. As discussed in section 3.1, the overabundance of zeros prevented us from directly solving this problem. We obtained approximately seven thousand microwells for a training set. We were able to obtain a validation (913 microwells) mean square error of 0.23 after 4 epochs with early stopping (figure 8).

5. Discussion

This work represents a proof of principle for a CNN approach to microwell image analysis. This analysis is complicated by the fact that cells are small in the initial days of the experiment, and this is exactly when the results are most critical for guidance on medical treatment. We suggest several approaches to improve our results:

- more data across more biological conditions will help to improve the generalization of the model;
- the label we used is inherently noisy, therefore using a GAN to predict the final day image may be a better approach (though this would have been a futile effort with the amount of data that we had);
- combining the classification and regression problem, possibly by using something similar to a zero-inflated model likelihood as the loss function.

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.

[8] J. Simm, G. Klambauer, A. Arany, M. Steijaert, J. K. Wegner, E. Gustin, V. Chupakhin, Y. T. Chong, J. Vialard, P. Buijnsters, et al. Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell chemical biology*, 25(5):611–618, 2018.

6. Contributions and acknowledgements

A.S. developed the microwell and imaging system and obtained the imaging data. S.Z. and T.D. developed the methods, performed the analysis, and wrote the paper with input from A.S.