Microscopy cell counting and detection with fully convolutional regression networks

DeepGs: Alec Tarashansky, Ehsan Dadgar-Kiani, Yuan Xue, Matthew Kim

Background: With the emergence of new technology, biology is largely becoming a quantifiable discipline. Obscure qualitative descriptions of biological phenomena are now being replaced with precise numerical data. Given the increase in throughput and capacity for measurement, the bottleneck to biological discovery depends heavily on efficient data analysis to match the rate of its acquisition. In the field of cell biology, quantification of cells in microscopic images was traditionally done by manual counting. The authors are motivated to develop an automatic cell detection and counting method in order to improve analysis throughput and reduce technical variances. They trained a modified convolutional neural network (CNN) on a synthetically generated dataset to count cells from experimental microscopic images. They also demonstrated that cell detection can be derived from their density estimation approach without the need for prior data labeling and object segmentation. The advantage of this approach is that it avoids the problem of obtaining a large dataset through manual annotation of the objects, which would defeat the purpose of an automatic method. They presented convincing results that the neural networks trained on standard synthetic data can generalise for different kinds of microscopy images gathered from biological experiments; however, they also discuss how the performance of their model can be further improved by fine-tuning parameters with annotated real data.

A related CNN approach to detect cells in microscopic images has been developed independently using a regression framework[1]. This alternative approach trains a CNN based on weakly annotated images, where a cross is manually placed approximately near the center of each cell. Given the input image patch, this algorithm will generate structured outputs called proximity patches, which exhibit higher values for pixels near cell centers. Although this regression-based approach produced robust detection of cells, it still required manual annotation of the training dataset. In contrast, the authors of our critique paper focused on building density estimation-based models where cell counting and detection in any specific region of the image are predicted simultaneously from the density map.

Data and Model Summary: The authors used synthetic images of clumped cells to train two Fully Convolutional Regression Networks (FCRN) to be able to map an input training image to its corresponding density map. The synthetic images are annotated with red crosses near the cell centres. Each dot annotation corresponds to a Gaussian function with peak value at the dot, indicating the approximate center of the cell. The resulting density maps consist of these Gaussians superimposed at the position of each dot. Integrating the density maps, calculating local maxima, and counting the local maxima within an arbitrary region of the image results in the number of cells within that region.

To regress the density map from its corresponding cell image, the authors design two FCRNs with slightly different architectures, FCRN-A and FCRN-B. The FCRNs use very small convolution kernels of size 3x3 or 5x5 pixels. FCRN-A, using 3x3 kernels, has three convolution - ReLU

(rectified linear units) pooling layers. In order to reconstruct a density map that is the same size as the input image, the max-pooling layers are followed by upsampling - convolution - ReLU pooling layers that undo the spatial reduction. Upsampling is done via bilinear interpolation and the convolution kernel sizes are hyperparameters that can be learned during end-to-end training. FCRN-B uses 5x5 kernels and only uses max-pooling after every two convolutional layers to avoid too much loss of spatial information.

To train the networks, the authors break up the images into random patches of 100x100 pixels that are normalized to zero mean and unit variance. The cost function is defined as the mean square error between a randomly sampled patch and the ground-truth annotation. Stochastic gradient descent with momentum updates the weights with a learning rate that is gradually decreased by a factor of 10. No dropout is used. In order to account for the fact that most of the pixels in the patches will be zeros (since the cells are very small), the Gaussian-annotated ground truths are increased by two orders of magnitude to force the network to fit the Gaussians instead of the background noise. Following pre-training with the patches, the parameters are fine-tuned with whole images.

Results/Summary: Both FCRN architectures see better performance when using using more training images (increasing from 8 to 32 images). Mean absolute error for the proposed FCRN-A performs about as well as the method proposed in *Fiaschi et al.* in which the method incorporates regression forests and structured labels to learn to count. Among the previous papers that used methods for counting on synthetic data-sets from other papers via neural networks and computer vision, the regression forest model of *Fiaschi et al.* had performed the best with the lowest mean absolute error and standard deviations.

FCRN-A performs slightly better than than FCRN-B for the synthetic dataset. The authors hypothesize that the size of the receptive field is more important than being able to provide detailed information over the receptive field. So a relative increase in loss of spatial information is countered for FCRN-A by larger receptive fields. Essentially, the difficulty is in the regression of large cell clumps and in order to accommodate larger clumps, a larger receptive field is required. Furthermore, the results show that fine-tuning increases the performance of FCRN-A by reducing error by 1%.

When the authors tested FCRN on real microscopy data, FCRN-B undercounted ground-truth RPE cells by a marginal 0.85% and Precursor T-Cell LBL cells by 1.97%. These results were closer to the ground-truth than that of FCRN-A and the authors hypothesize that unlike synthetic data, real data should contain smaller cell clumps, giving FCRN-B the edge in a small receptive field.

Weaknesses: The model undoubtedly obtained very high accuracies on both the training and test data sets. However there are still several weaknesses we observed in both the methods and resultant analysis. First of all, although one of the main premises behind this paper is the ability to train a network on a synthetic dataset and have it generalize to real microscopy images, it would have been fruitful to at least attempt training on real biological data. Training primarily on a synthetic dataset

seems like it would limit the predictive power of the model because it would be missing out on some of the subtleties of real data. The authors claim that this was infeasible since real biological image data was not large enough to train such an FCRN architecture, but a standard technique for bootstrapping images is to apply random deformations, using both affine and nonlinear transformations to increase the dataset size [2]. The paper made no mention of such an attempt to do bootstrapping on real biological data, although they do use simple data augmentation techniques on the synthetic dataset. This could have potentially addressed one of the main weaknesses of the model, identifying cells amongst clumps, for it could have allowed the convolutional layers to pick up on more subtle features on the input to help differentiate between cells.

Additionally, each cell in the synthetic dataset has homogeneous fluorescence intensity, whereas real biological cells may have relatively more membrane-bound fluorophores which would result in the periphery being more intense than the cell center. The authors did not characterize how cell counting performance changes when training on more realistic synthetic intensity maps. In fact, they do not even adequately describe what types of stained images (i.e. DAPI staining, antibody staining, etc.) they are interested in counting cells for. Also, a fairly significant fraction of imaged cells will invariably be undergoing some cell transition state such as mitosis. Perhaps the authors should have mentioned how cell-counting performs with morphologically dynamic cells to address the fact that not all cells will have the same shape as the synthetic training data.

Furthermore, the explanation surrounding what information has been encoded in the different convolutional layers left much to be desired. The authors briefly mention that it appears that for clumps of cells, *Conv3* captures edges around the cell, while *Conv4* captures concavity information. However, the authors failed to provide an adequate description of how these features (edges, concavity, etc.) specifically inform the cell identification process.

Future Steps: An interesting extension of this paper's work would be to support cell counting and detection in three dimensional volumes. An example of this would be neuronal cell counting from tissue clearing data such as CLARITY[3] or iDISCO[4]. Although an architecture that handles this added spatial dimension is inherently more complicated to train, it may address the biggest limitation of the density map prediction task described in this paper, which is that it occasionally fails to differentiate between cells that are clumped relatively close together. An added spatial dimension may help resolve overlapping cells that would be misidentified due to a poor 2D viewing angle. To our knowledge this technique has not been adapted to three dimensional cell counting.

References:

[1] Beyond Classification: Structured Regression for Robust Cell Detection Using Convolutional Neural Network. Yuanpu Xie, Fuyong Xing, Xiangfei Kong, Hai Su, Lin Yang (2015). Conference paper. DOI: 10.1007/978-3-319-24574-4_43.

[2] Chollet, Francois. "Building Powerful Image Classification Models Using Very Little Data." Blog post. *The Keras Blog*. N.p., 05 June 2016. Web. 27 Nov. 2016.

[3] Chung, K. et al. Structural and molecular interrogation of intact biological systems. Nature 497, 332–337 (2013).

[4] Renier et al., 2014N. Renier, Z. Wu, D.J. Simon, J. Yang, P. Ariel, M. Tessier-Lavigne. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imagingCell, 159 (2014), pp. 896–910