MULTILAYER ENCODER-DECODER NETWORK FOR 3D NUCLEAR SEGMENTATION IN SPHEROID MODELS OF HUMAN MAMMARY EPITHELIAL CELL LINES

Mina Khoshdeli, Garrett Winkelmaier, and Bahram Parvin

Biomedical and Electrical Engineering Department, University of Nevada, Reno, NV, USA

ABSTRACT

Nuclear segmentation is an important step in quantitative profiling of colony organization in 3D cell culture models. However, complexities arise from technical variations and biological heterogeneities. We proposed a new 3D segmentation model based on convolutional neural networks for 3D nuclear segmentation, which overcomes the complexities associated with non-uniform staining, aberrations in cellular morphologies, and cells being in different states. The uniqueness of the method originates from (i) volumetric operations to capture all the three-dimensional features, and (ii) the encoder-decoder architecture, which enables segmentation of the spheroid models in one forward pass. The method is validated with four human mammary epithelial cell (HMEC) lines—each with unique genetic makeup. The performance of the proposed method is compared with the previous methods and is shown that the deep learning model has a superior pixel-based segmentation, and an F1-score of 0.95 is reported.

Index Terms— Convolutional neural networks, 3D nuclear segmentation, encoder-decoder architecture, 3D spheroid model, volumetric convolution

1. INTRODUCTION

Cancer is a disease of cellular disorganization, which requires delineation of nuclear regions. The goal of this manuscript is to delineate nuclear regions in spheroid models of human mammary epithelial cell (HMEC) lines. Spheroid models represent the organizations of premalignant and malignant HMEC lines uniquely [1, 2] and in terms of specific genetic aberrations. In other words, while image-based classification of cells is nearly impossible in 2D assays, 3D colony organization tends to be unique with respect to genetic aberrations. The first step for characterizing colony organization is nuclear segmentation in 3D spheroid models that have been imaged with confocal microscopy. The traditional approach for 3D nuclear segmentation involves a cascaded of custom-designed processes for making the solution tractable. However, in this paper, we show that extension of emerging methods in deep learning can be trained to learn complex patterns. The complexities arise from cells being in different states, non-uniformity in staining, nuclear size, overlapping nuclei, and the fact that cancer cells tend to have large nucleoli.

Organization of this paper is as follows. In section 2, previous research is summarized. Section 3 outlines the approach. Section 4 outlines experimental design, which includes validation of the computational pipeline with a diversity of cell lines--each with a specific mutation. Section 5 concludes the paper.

2. LITERATURE REVIEW

Nuclear segmentation is an important step in profiling colony organization, disease progression, and quantifying protein expression. It has been widely studied in 2D (e.g., histopathology) problems [3, 4]. However, for this paper, we summarize the previous research as they relate to 3D samples that are imaged by fluorescence microscopy. The main complexities arise from differentially expressed fluorescent signals (e.g., cells being at different states, staining non-uniformities), cells being at different scales, or cells having different morphologies. At the same time, overlapping cells, as a result of fixation, increased complexities.

Classical techniques are based either on the geometrical morphology of the nuclear shapes, or on features that are derived from intensity (e.g., gradient), or a hybrid combination of shape and intensity derived features. In [5], a segmentation method based on gradient flow tracking (GFT) has been presented, which is a generalization of regularized centroid transform [6] to 3D data. The steps of this method include computation of regularized gradient vector field, tracking gradient flow, and labeling each voxel based on a converged sink position that corresponds to the nucleus centroid. In [7], a hybrid method based on

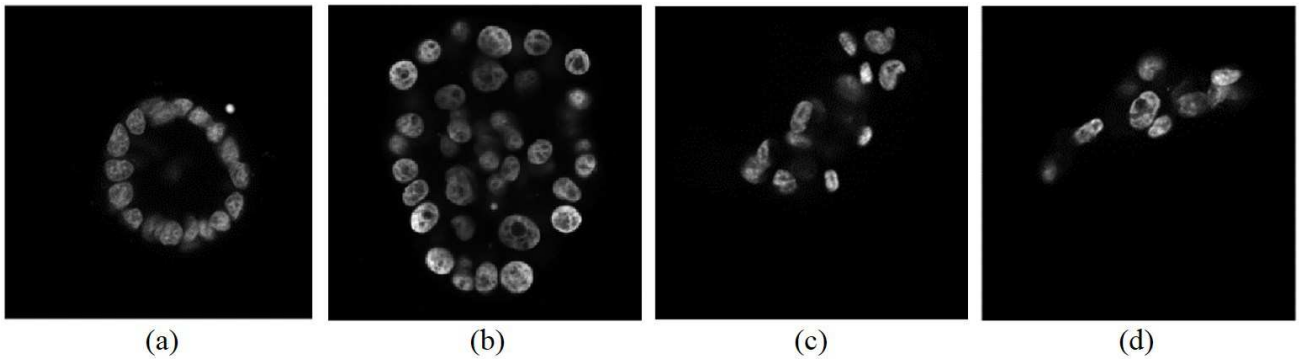


Fig. 1. The middle section of 3D spheroids shows diverse phenotypes (from left to right) MCF10A, MCF7, MDA-MB-468, and MDA-MB-231.

iterative voting and geometric partitioning have been proposed. In [8, 9], a cascade of 3D operations was implemented to homogenize intensity distributions within each nucleus, and overlapping nuclei were partitioned based on surface curvatures. In [10], nuclei segmentation of fluorescence microscopy images has been obtained using inter- and intra-region discriminative information in three steps. First, an initial segmentation has been achieved by dynamically thresholding local regions. Second, the false positive regions have been filtered by computing similarity features for the adjacent nuclei. Third, nuclei localization has been performed by using intra-region contrast information. Although [5, 7] are successful for the nuclei with homogeneous intensity, they will fail for the nuclei with aberrant nuclear morphometry (e.g., elongated nuclei) and with heterogeneous internal structures (e.g., large nucleoli). This is in part due to the facts that these methods were developed for normal and premalignant cells. However, cancer cells can present heterogeneous aberrant structures. In [11], an atlas-based model has been proposed for 3D nuclei segmentation based on the assumption that a prior model exists for the cellular organization. However, due to the organizational heterogeneity of spheroids [12] and mutation properties of each cell line, corresponding to a phenotype, it is rather difficult to find a prior model for the growth of 3D cell culture models.

In the recent years, convolutional neural networks [13] have become popular for addressing a range of problems (e.g., image classification [14-16], image recognition [17], image segmentation [18-21]). The main advantage of the convolutional neural networks, compared to the traditional computer vision methods, is that complex patterns can be learned automatically with sufficient sample size. For example, it is difficult to model a nucleus with a large nucleolus. In fact, since the presence of nucleoli interferes with the task of segmentation, previous researchers have developed methods to filter them out [8]. In [22], a ten-layer convolutional neural network has been trained with a synthetic dataset for segmentation of nuclei in fluorescence microscopy images. The synthetic dataset has been generated based on the assumption that a nucleus can be modeled as an ellipsoid shape. Therefore, each data sample contains a few ellipsoid volumes, which has been blurred by a Gaussian filter. A Poisson noise has been added to the final samples to crease more realistic dataset. Although the network performs well on the synthetic data set, it performs poorly on the real data. The reasons are (i) nuclei have different and complex shapes, not only elliptic, (ii) the intensity of voxels inside a nucleus is not uniform, and (iii) the synthetic data does not include touching nuclei. In [23], the two-dimensional U-Net [20] model has been trained for segmenting 3D microscopy images of plant cell boundaries. In [24] , the two-dimensional SegNet [18] model has been used for segmentation of yeast cell in multi-modal fluorescence microscopy images. In [25], we have shown that 3D convolutional neural network has superior performance for nuclear detection of 3D cell colonies. In this paper, we proposed to use the convolutional neural network for 3D nuclear segmentation. To this aim, we generalize the encoder-decoder architecture for 3D image segmentation by performing volumetric convolution and pooling. Moreover, we have validated the proposed model with manually annotated data of human mammary epithelial cell lines.

3. METHOD

The proposed nuclear segmentation framework consists of a deep convolutional neural network for region-based

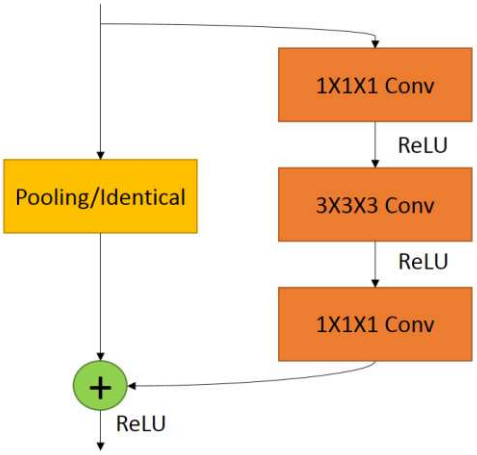


Fig. 2. The Bottleneck module consists of two parallel branches of pooling and convolutional. The output of these two branches are then added element-by-element.

segmentation of 3D images, and a post-processing stage for separating overlapping of nuclei.

3.1. Proposed convolutional neural network for 3D image segmentation

A convolutional neural network is constructed by stacking layers of convolution and pooling in an alternating fashion. In a typical 2D CNN, convolution and pooling operations are performed spatially to compute feature maps from the local neighborhood, which originated from the previous layer. We suggest that for the problem of 3D segmentation, it is desirable to capture the local neighborhood information in the third dimension as opposed a series of 2D convolutions for each slice [26], i.e., volumetric convolution and pooling are preferred. However, volumetric convolution and pooling, for construction of the feature maps, increase the computation significantly. We address this issue in two ways. First, all 3D convolutions are 3-by-3-by-3, which has a low computational cost. Second, a 1-by-1-by-1 convolution is applied to each feature map layer, which significantly reduces the dimensionality of the data.

More recently, the concept of volumetric operations, in deep learning, have been applied for the task of classification and segmentation [27, 28], where classification is typically through layers of convolution and pooling and segmentation is through the encoder-decoder architecture. The encoder-decoder architecture was first introduced in SegNet [1].

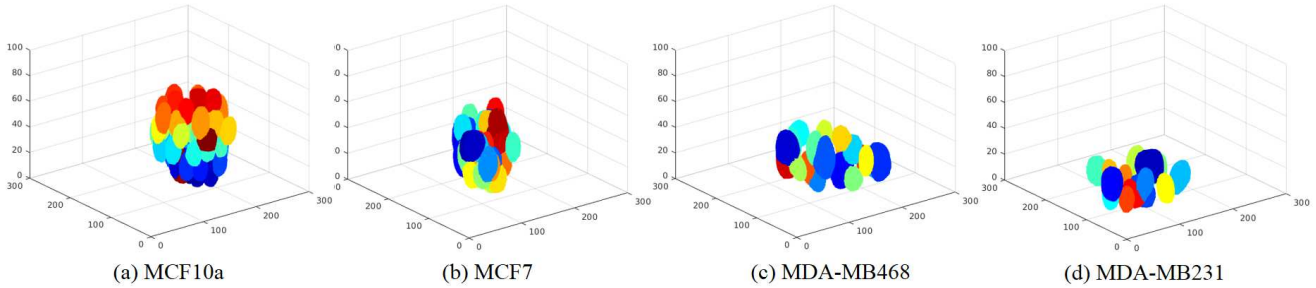
Table 1: Architecture of the Encoder and Decoder modules for training with patch size of 120-by-120-by-48 pixels.

Encoder Module	Input/Output dimension	Filter dimensions
Input	120×120×48×1	
Conv.	120×120×48×16	3×3×3
Max-Pooling	60×60×24×16	2×2×2
Bottleneck 1	30×30×12×64	3×3×3
Bottleneck 2	30×30×12×64	3×3×3
Bottleneck 3	30×30×12×64	3×3×3
Bottleneck 4	30×30×12×64	3×3×3
Bottleneck 5	15×15×6×128	3×3×3
Bottleneck 6	15×15×6×128	3×3×3

Decoder Module	Input/Output dimension	Filter dimensions
Bottleneck 7	30×30×12×64	3×3×3
Bottleneck 8	30×30×12×64	3×3×3
Bottleneck 9	60×60×24×16	3×3×3
Full conv.	120×120×48×C	2×2×2

The encoder-decoder network is a cascade of two convolutional neural networks (e.g., an encoder and a decoder), where the encoder module performs feature extraction from the input sample, and the decoder module is used to up-sample the output of the encoder. The existing models are either symmetric (have the same architecture for encoder and decoder, e.g., SegNet[18], UNet[20]), or asymmetric (have shallower decoder than the encoder, e.g., ENet [29]). Most of the successful 2D models are constructed from a stack of series blocks named bottleneck, where each bottleneck is a parallel combination of a convolution layer and a pooling layer. Similar to 2D, the proposed 3D model, is constructed from stacking of bottlenecks. The architecture of the bottleneck is shown in **Fig. 2.** The pooling branch can perform both up-sampling and down-sampling. The convolutional branch includes three convolutional layers: (i) two 1-by-1-by-1 convolutions at the beginning and end of a branch, and (ii) a primary convolutional layer with the filter size of 3-by-3-by-3. After each convolution layer, there are Batch Normalization and ReLU as the activation function. The outputs of these two branches then merge through element-wise addition.

After comparing several configurations, the architecture of **Table 1** demonstrated the best performance for a specific patch-size, but we are continuing with our evaluation. The proposed



(a) MCF10a (b) MCF7 (c) MDA-MB468 (d) MDA-MB231

Fig. 3. Visualization of nuclear segmentation and colony organization for (a) premalignant MCF10A versus malignant cell lines of (b) MCF7, (c) MDA-MB-468, and (d) MDA-MB-231.

architecture is an asymmetric one (i.e., the decoder module is smaller than the encoder.) This helps to reduce the computational cost, which is more important for 3D volumetric analysis.

3.2. Post-processing for separation of touching nuclei

Separation of the clump of nuclei after segmentation is a necessary step for delineating nuclei. Therefore, a final post-processing step is added. There are various techniques to separate a clump of nuclei, which include: marker-based watershed and curvature based partitioning [8]. Each of these methods has their advantages and disadvantages. However, the current post-processing step is limited to the marker-based watershed method because of its computational simplicity and open source availability.

4. EXPERIMENTAL SETUPS AND RESULTS

This section outlines the experimental design, validation of the trained model, and the performance results. An important component of the experimental design is that segmentation needs to be validated with multiple cell lines, where each line has a unique genetic aberration. Experimental design includes 4 cell lines: (a) MCF7, which is progesterone positive, estrogen positive, and ERBB2 negative; (b) MDA-MB-468, which is ERBB2 negative with EGFR amplification; (c) MDA-MB-231, which is progesterone negative, estrogen negative, and ERBB2 negative; and (d) MCF10A, which is non-transformed and recapitulate lumen formation. Subsequently, colony formations were harvested in days 2, 5, 7, and 12. Each cell line contributes uniquely to 3D colony organization for validating higher order bioinformatics analysis.

The dataset consists of 68 image stacks (e.g., colonies), which were imaged with confocal microscopy and were also used in [8]. Each sample is imaged at 0.25-micron resolution in the XY directions and 1-micron resolution in the Z direction. The size of each sample is approximately 256-by-256-by-80 pixels. Each sample data has been divided into four non-overlapping patches of size 120-by-120-by-48. Data augmentation (e.g., rotation, flipping) has been applied to increase the number of data. The samples were divided equally between training and testing datasets (50% -50%).

Training of the network is achieved by the Adam optimization algorithm [30]. The learning rate is set at 1e-5 initially, and then it is decreased by a factor of 10 in every100 iterations. Dropout regularization is added after each convolution layer to avoid overfitting. L_2 weight norm regularization is also used with the coefficient of 5e-4. All convolutional layers are initialized with a zero-mean Gaussian distribution with a variance of 0.01.

Performance of the proposed model is compared with the previous methods, which include curvature-based partitioning (CBP) [8], gradient flow tracking (GFT) [5], and Watershed (Table 2). The results are evaluated based on pixel- and object- level precision, recall, and F1-score defined as below.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall}$$

In the object-level evaluation, a nucleus considered to be a true positive (TP), if its intersection over union (IoU) with a ground truth object is more than 50%. The IoU of object A and B is defined as below.

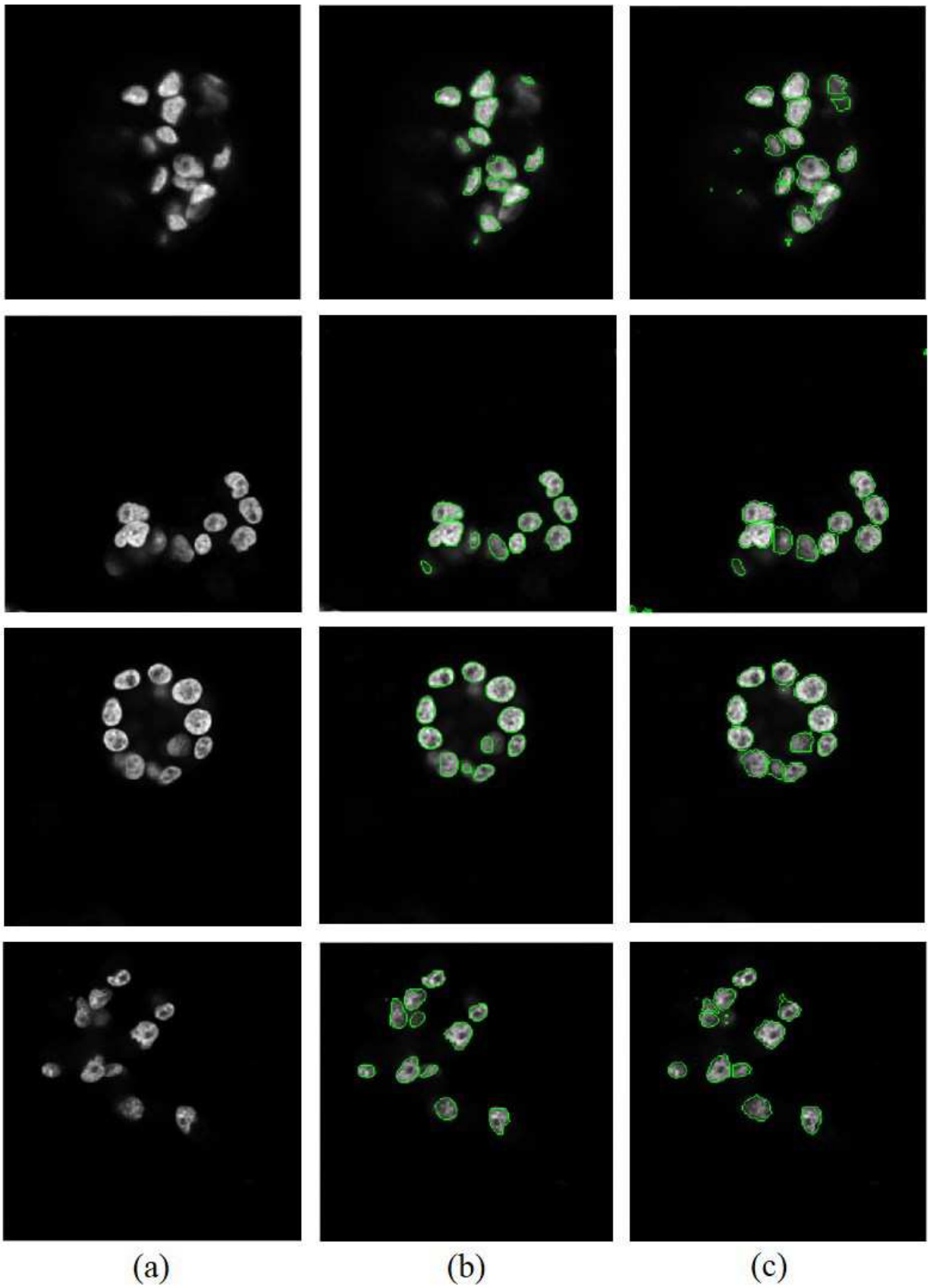


Fig. 4. Nuclei are segmented and overlaid on a single section of 3D colonies. (a) original image, (b) ground truth, (c) segmentation results.

Table 2. The proposed encoder-decoder architecture has a superior pixel-level (e.g., accuracy of a pixel being part of a cell) performance as compared to previous methods.

Method	Object-based precision	Object-based recall	Object-based F1-score	Pixel-based precision	Pixel-based recall	Pixel-based F1-score
Proposed	0.89	0.85	0.87	0.94	**0.93**	**0.95**
CBP	**0.97**	**0.95**	**0.96**	0.86	0.84	0.85
Watershed	0.75	0.67	0.71	**0.95**	0.44	0.60
GFT	0.88	0.87	0.88	**0.95**	0.82	0.88

$$IoU(A,B) = \frac{A \cap B}{A \cup B}$$

The pixel-based accuracy of the proposed method is about 7% superior to the other approaches. However, the object-based accuracy of the curvature-based partitioning [8] is higher than our method. There are a few possible explanations for this observation: (i) it is the effect of down-sampling in the encoder module, which causes loss of nuclei boundaries, (ii) convolutional neural networks cannot capture higher-level information such as the shape of the objects. Therefore, the proposed model cannot distinguish touching nuclei. Examples of nuclear segmentations for each of the cell lines are shown in **Fig. 3** and **Fig. 4.**

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel 3D convolutional neural network design based on the encoder-decoder architecture for nuclear segmentation in 3D organoid models. The advantages of the proposed model are that: (i) its encoder-decoder architecture generates the segmentation results in one forward pass through the network; and (ii) its pixel-based accuracy is higher than previous methods. Future works include: (a) integrating engineered feature with the deep learning model; (b) constraining the proposed model with the nuclear shapes; and (c) extending the model with high-level features to complete perceptual surface boundaries for delineating adjacent nuclei.

6. REFERENCES

[1] P. A. Kenny, G. Y. Lee, C. A. Myers, R. M. Neve, J. R. Semeiks, P. T. Spellman, K. Lorenz, E. H. Lee, M. H. Barcellos-Hoff, O. W. Petersen, J. W. Gray, and M. J. Bissell, "The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression," *Molecular Oncology,* vol. 1, pp. 84-96, 2007.

[2] J. Han, H. Chang, O. Giricz, G. Y. Lee, F. L. Baehner, J. W. Gray, M. J. Bissell, P. A. Kenny, and B. Parvin, "Molecular predictors of 3D morphogenesis by breast cancer cell lines in 3D culture," *PLoS Comput Biol,* vol. 6, no. 2, pp. e1000684, Feb, 2010.

[3] F. Xing, and L. Yang, "Robust Nucleus/Cell Detection and Segmentation in Digital Pathology and Microscopy Images: A Comprehensive Review," *IEEE Reviews in Biomedical Engineering,* vol. 9, pp. 234-263, 2016.

[4] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential," *IEEE Reviews in Biomedical Engineering,* vol. 7, pp. 97-114, 2014.

[5] G. Li, T. Liu, A. Tarokh, J. Nie, L. Guo, A. Mara, S. Holley, and S. Wong, "3D cell nuclei segmentation based on gradient flow tracking," *BMC Cell Biology,* vol. 8, no. 40, 2007.

[6] Y. Qing, and B. Parvin, "Harmonic cut and regularized centroid transform for localization of subcellular structures," *IEEE Transactions on Biomedical Engineering,* vol. 50, no. 4, pp. 469-475, 2003.

[7] J. Han, H. Chang, Q. Yang, G. Fontenay, T. Groesser, M. H. Barcellos-Hoff, and B. Parvin, "Multiscale iterative voting for differential analysis of stress response for 2D and 3D cell culture models," *Journal of Microscopy,* vol. 241, no. 3, pp. 315-326, 2011.

[8] C. Bilgin, S. Kim, E. Leung, H. Chang, and B. Parvin, "Integrated profiling of three dimensional cell culture models and 3D microscopy," *Bioinformatics*, 2013.

[9] C. Bilgin, G. Fontenay, Q. Cheng, H. Chang, J. Han, and B. Parvin, "BioSig3D: High content screening of three-Dimensional cell culture models," *PLoS One,* vol. 11, no. 3, pp. e0148379, 2016.

[10] Y. Song, W. Cai, D. D. Feng, and M. Chen, "Cell nuclei segmentation in fluorescence microscopy images using inter- and intra-region discriminative information." pp. 6087-6090.

[11] L. Qu, F. Long, X. Liu, S. Kim, E. Myers, and H. Peng, "Simultaneous recognition and segmentation

of cells: application in C.elegans," *Bioinformatics,* vol. 27, no. 20, pp. 2895-2902, 2011.

[12] Q. Cheng, C. C. Bilgin, G. Fonteney, H. Chang, M. Henderson, J. Han, and B. Parvin, "Stiffness of the microenvironment upregulates ERBB2 expression in 3D cultures of MCF10A within the range of mammographic density," *Sci Rep,* vol. 6, pp. 28987, 2016.

[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *NIPS*, 2012.

[15] K. Simonyan, and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ImageNet Challenge,* vol. 6, 2014.

[16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *CoRR,* vol. abs/1512.00567, /, 2015.

[17] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks," *CoRR,* vol. abs/1312.6229, /, 2013.

[18] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *CoRR,* vol. abs/1511.00561, /, 2015.

[19] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *arXiv*, 2015.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv*, 2015.

[21] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation," *CoRR,* vol. abs/1606.02147, /, 2016.

[22] D. J. Ho, C. Fu, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei Segmentation of Fluorescence Microscopy Images Using Three Dimensional Convolutional Neural Networks." pp. 834-842.

[23] P. Saponaro, W. Treible, A. Kolagunda, T. Chaya, J. Caplan, C. Kambhamettu, and R. Wisser, "DeepXScope: Segmenting Microscopy Images with a Deep Neural Network." pp. 843-850.

[24] A. S. Aydin, A. Dubey, D. Dovrat, A. Aharoni, and R. Shilkrot, "CNN Based Yeast Cell Segmentation in Multi-modal Fluorescent Microscopy Data." pp. 753-759.

[25] M. Khoshdeli, and B. Parvin, "Feature-Based Representation Improves Color Decomposition and Nuclear Detection Using a Convolutional Neural Network," *IEEE Transactions on Biomedical Engineering,* vol. 65, no. 3, pp. 625-634, 2018.

[26] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, K. Mori, I. Sakuma, Y. Sato, C. Barillot and N. Navab, eds., pp. 246-253: Springer Berlin Heidelberg, 2013.

[27] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal and W. Wells, eds., pp. 424-432, Cham: Springer International Publishing, 2016.

[28] R. Socher, B. Huval, B. Bhat, C. D. Manning, and A. Y. Ng, "Convolutional-recursive deep learning for 3D object classification," in Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 2012, pp. 656-664.

[29] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation," *CoRR,* vol. 1606.02147, 2016.

[30] D. P. Kingma, and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR,* vol. abs/1412.6980, /, 2014.