RCNN-SliceNet: A Slice and Cluster Approach for Nuclei Centroid Detection in Three-Dimensional Fluorescence Microscopy Images

Liming Wu* Shuo Han* Alain Chen*
Paul Salama† Kenneth W. Dunn‡ Edward J. Delp*
* Video and Image Processing Laboratory (VIPER), Purdue University, West Lafayette, Indiana
† Department of Electrical and Computer Engineering, Indiana University-Purdue University, Indianapolis, Indiana
‡ Division of Nephrology, School of Medicine, Indiana University, Indianapolis, Indiana

Abstract

Robust and accurate nuclei centroid detection is important for the understanding of biological structures in fluorescence microscopy images. Existing automated nuclei localization methods face three main challenges: (1) Most of object detection methods work only on 2D images and are difficult to extend to 3D volumes; (2) Segmentation-based models can be used on 3D volumes but it is computational expensive for large microscopy volumes and they have difficulty distinguishing different instances of objects; (3) Hand annotated ground truth is limited for 3D microscopy volumes. To address these issues, we present a scalable approach for nuclei centroid detection of 3D microscopy volumes. We describe the RCNN-SliceNet to detect 2D nuclei centroids for each slice of the volume from different directions and 3D agglomerative hierarchical clustering (AHC) is used to estimate the 3D centroids of nuclei in a volume. The model was trained with the synthetic microscopy data generated using Spatially Constrained Cycle-Consistent Adversarial Networks (SpCycleGAN) and tested on different types of real 3D microscopy data. Extensive experimental results demonstrate that our proposed method can accurately count and detect the nuclei centroids in a 3D microscopy volume.

1. Introduction

High quality data generated by optical microscopy can provide useful information for understanding biological tissue structures [1]. Traditional optical microscopes have limitations due to light scattering, leading to images with low resolution and contrast [2]. Two-photon fluorescence microscopy allows deeper tissue imaging with near-infrared light [3]. The strongly focused subpicosecond pulses make it possible to generate high-quality three-dimensional (3D) images in living cells and deeper tissues [4].

Quantification results can be useful for obtaining information for subsequent analysis such as cell tracking, disease diagnosis, and new drug development [5]. Robust and accurate nuclei counting and localization are the first steps in quantifying biological structures. Due to large variations of cell type, size, or microscopy modality, nuclei counting and localization remains a challenging problem especially in a 3D volume where nuclei are crowded and overlap with each other [6].

Segmentation approaches are popular for nuclei counting and localization. This is because segmentation can separate foreground nuclei and background structures. Many segmentation methods use thresholding to determine binary masks of nuclei. In [7] a cell nuclei segmentation method that uses median filtering, Otsu's thresholding [8], and morphological operation to segment nuclei is described. Otsu's thresholding can estimate a threshold by minimizing within-class variance of the foreground and background pixels. Similarly, ImageJ's 3D object counter, known as JACoP [9, 10], uses Otsu's thresholding to segment the 3D image volumes. It then uses 3D connected components to remove small artifacts and estimates the number of objects and centroid coordinates of each object in a 3D volume. The counting and localization accuracy still suffers from over-segmentation caused by artifacts and noise. The approach in [11] was integrated in [12] and proposed a graph-cuts-based binarization that used multi-scale Laplacian-of-Gaussian filtering to extract the image foreground. CellProfiler [13] provides customized "pipelines" by adding different functional modules for image segmentation and quantitative analysis.

Another successful method for segmentation is active contours [14]. It minimizes an energy function with the assumption that the approximate shape of the boundary is known. As an extension, a segmentation method known as Squassh [15, 16] couples image segmentation with image restoration and uses generalized linear models for energy minimization. It enables co-localization and shape analyses

for better quantifying subcelluar structures. Similarly, a 3D region-based active contour method described in [17] incorporated the 3D inhomogeneity correction for solving the inhomogeneous intensity issue. These segmentation techniques cannot easily separate touching objects.

Watershed segmentation has been used for this problem. Traditional watershed [18] uses the regional minimum as the flood point and builds barriers when different water sources meet which results in over-segmentation. This was improved in [19] by using a new marker-controlled watershed segmentation with conditional erosion. Similarly, Volumetric Tissue Exploration and Analysis (VTEA) described in [20] is an image analysis toolbox integrated in ImageJ which uses 2D watershed segmentation to separate different nuclei on each focal plane and merges the 2D segments to 3D segments based on the segments centroid distance. It can separate different nuclei and provide centroid information of each nucleus.

More recently, convolutional neural networks (CNN) has been widely used in microscopy image segmentation. The popular image segmentation framework known as SegNet [21] is an encoder-decoder architecture. It was originally used for scene understanding applications and has been extended to 3D and used in fluorescence microscopy volume segmentation [22, 23]. Similarly, [24] and [25], an encoder-decoder type of fully convolutional neural network was proposed for dense volumetric segmentation. It uses deconvolution and shortcut concatenation to better reconstruct the location and context information, which result in better segmentation results in biomedical research with limited labeled data. An improved 3D U-Net using watershed, known as DeepSynth, was described to segment different nuclei instances [26]. The nuclei centroid detection accuracy of these methods normally depends on the watershed segmentation accuracy.

To distinguish different objects, the region proposal based network, known as regional convolutional neural network (R-CNN), was proposed in [27] for object localization. It uses selective search to generate regions of interest (RoIs) that may contain candidate objects and classify the RoIs into different categories. However, this network is relatively slow for training and inference. This has been improved in [28] by directly extracting the RoIs from the feature map that generated from another deep CNN. More recently, [29] replaced selective search with a new region proposal network (RPN) for learning the RoIs. This significantly reduced the training time. Alternatively, another category of object detectors [30, 31] that is based on one-stage regression network achieved competent results. In [32], the Faster R-CNN was used to detect nuclei in malaria images and in [33], the modified Faster R-CNN was used for pneumonia detection in CT images. These object detection models only work on 2D images and are difficult to extend to 3D.

An approach described in [34] took advantage of 2D object detectors and estimated the 3D bounding boxes from the segmented pointcloud. But the 3D location accuracy is highly dependent on the 2D object detection accuracy. This has been improved in [35] with a 3D object detection network that is based on deep Hough voting model to directly regress the 3D object centroid. Similarly, [36] proposed a SliceNet using the slice-and-fuse strategy for object detection in high-resolution 3D volumes. It detects the object on each slice of a volume from 3 different directions and merges the results to a fused 3D detection.

Another challenge of deep learning-based methods is the lack of ground truth labels. Manually labeling ground truth masks or bounding boxes is labor-intensive and time-consuming even for an experienced person especially for 3D volumetric data. One way to address this problem is using data augmentation methods to create some "new" training samples, which include some linear and non-linear transformations such as horizontal flipping, random cropping, color space transformations, or elastic deformations [37, 38]. These traditional data augmentation methods are not appropriate if the available training samples are limited. An alternative way to address this problem is to generate synthetic data. In [39], synthetic objects are rendered with random background, lighting, and texture are used for training and testing Faster R-CNN. More recently, [23, 40, 41] describe an approach for generating synthetic ellipsoidal and non-ellipsoidal nuclei using Generative Adversarial Networks (GANs) in 3D microscopy volumes.

In this paper, we present a slice-and-cluster strategy inspired by [34, 36] that combines the Spatially Constrained CycleGAN [40], RCNN-SliceNet and 3D agglomerative hierarchical clustering (AHC) to detect the centroid of nuclei in a 3D microscopy volume without the need of large amounts of ground truth labels. We tested our method on four different real 3D microscopy data and the evaluation results show our method outperforms the rest of the baseline methods that widely used in biomedical research.

This paper provides an approach for detecting nuclei centroids in large 3D microscopy volumes. The main contribution includes: (1) We propose the slice-and-cluster strategy for extending RCNN-SliceNet's detection results from 2D slices to 3D volumes; (2) We introduce a nuclei centroid estimation method using 3D agglomerative hierarchical clustering to estimate the true 3D nuclei centroids by making use of the RCNN-SliceNet inference results; (3) We provide extensive experimental results to validate the effectiveness of our designed approach. We demonstrate that our proposed method achieves best results both visually and numerically compared to other baseline models and image analysis toolkits. We also show our method is robust to generalize to other microscopy data without training.

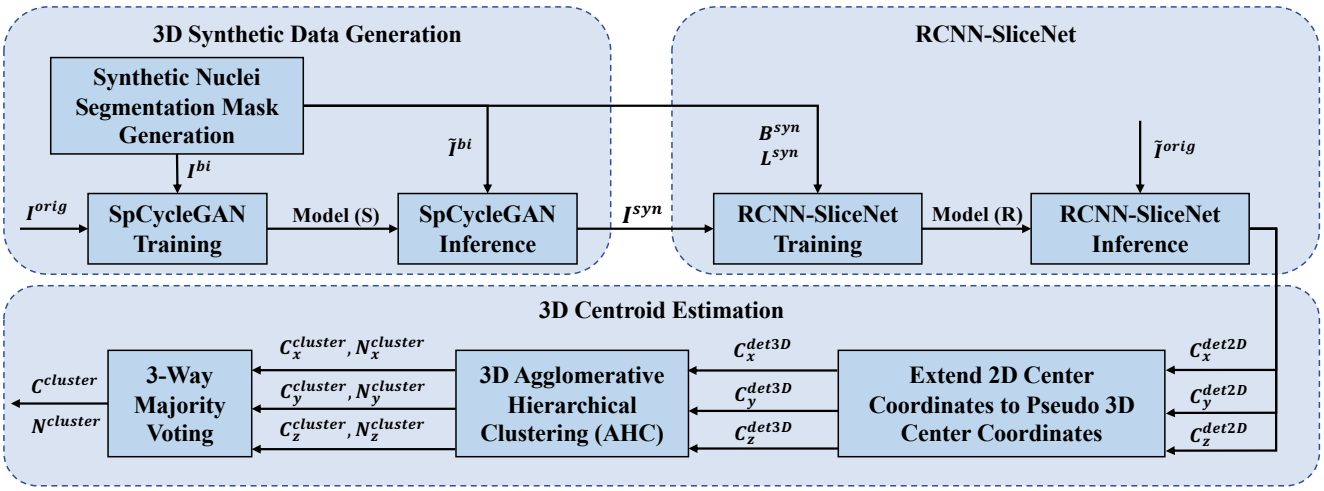


Figure 1. The block diagram of the proposed method

2. Proposed Method

As shown in Figure 1, the block diagram of the proposed system consists of three major parts, 3D synthetic data generation, RCNN-SliceNet, and 3D centroid estimation.

In this paper, I denotes a 3D image volume of size $X \times Y \times Z$. I_z denotes all slices on the z-direction, and I_{z_p} is denoted as the p^{th} slice, which is of size $X \times Y$, along the z-direction in a volume, where $p \in \{1, ..., Z\}$. Similarly, I_{x_p} and I_{y_p} denote the p^{th} focal plane image along the x-direction and along the y-direction. We then denote I^{orig} as the original microscopy volume of size $X \times Y \times Z$. Similarly, I^{bi} denotes the synthetic binary volume and I^{syn} denotes the synthetic microscopy volume. For example, $I^{orig}_{z_{25}}$ is the 25^{th} focal plane of the original microscopy volume along the z-direction. I^{bi} and $\tilde{I}^{bi}$ are different binary volumes use for training and inference of SpCycleGAN [40]. Similarly, I^{orig} and $\tilde{I}^{orig}$ are different microscopy volumes.

B^{syn} is a N by 4 matrix that represents the 2D coordinates of the bounding box for the nuclei on each slices along a direction. L^{syn} is the label of each bounding box, in our case, all $L^{syn} = 1$ which indicates the nucleus. Finally, as shown in Figure 1, the trained RCNN-SliceNet is used to inference on the $x-$, $y-$ and $z-$ directions of a volume. Suppose RCNN-SliceNet is used on the z-direction then we denote C_z^{det2D} as the detected centroid coordinates of nuclei in all 2D slices along the z-direction of a volume, and C_z^{det3D} is the pseudo 3D detected centroid coordinates estimated from C_z^{det2D}.

$C_z^{cluster}$ is the 3D coordinates of cluster centroids and $N_z^{cluster}$ is the number of clusters. Similarly, if the RCNN-SliceNet is used on the x-direction and the y-direction, $C_x^{det2D}, C_x^{det3D}, C_x^{cluster}, N_x^{cluster}$, and $C_y^{det2D}, C_y^{det3D}, C_y^{cluster}, N_y^{cluster}$ will be generated.

2.1. 3D Synthetic Data Generation

Synthetic data generation consists of synthetic nuclei segmentation mask generation and synthetic microscopy volume generation (Figure 1).

The nuclei are assumed to have ellipsoidal shape based on the observation of the original microscopy volume. Candidate nuclei I^{can} in different size, location and orientation are generated in a 3D volume I^{bi}. The affine transformation described in [23] is used to generate nuclei in different orientation. The size of I^{can} is determined by the length of semi-axes $\mathbf{a} = (a_x, a_y, a_z)$, which are appropriately chosen based on the nuclei size in original microscopy volume. The k^{th} nucleus candidate $I^{can,k}$ with intensity k is generated by Equation 1. In order to differentiate nuclei instances, each candidate nucleus $I^{can,k}$ has a unique gray-level intensity k. The maximum overlap between two candidate nuclei is set to be at a threshold of T_{ov} voxels.

$$I^{can,k} = \begin{cases} k, & \text{if } \frac{\tilde{x}^2}{a_x} + \frac{\tilde{y}^2}{a_y} + \frac{\tilde{z}^2}{a_z} < 1 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Candidate nucleus $I^{can,k}$ is added to I^{label}, which is initialized as a $128 \times 128 \times 128$ volume with intensity 0. $B^{syn,k}$ and L^{syn} are the ground truth bounding boxes coordinates and the category of the k^{th} nucleus.

The spatially constrained CycleGAN (SpCycleGAN) from [40] is used for synthetic microscopy volume generation. SpCycleGAN is an extension of CycleGAN [42] with an additional generative network H and a spatial constraint loss $\mathcal{L}_{spatial}$ that can generate synthetic microscopy images while maintaining the nuclei location and contour from the synthetic binary volume.

SpCycleGAN consists of 5 networks G, F, H, D_1, and D_2. G maps I^{bi} to I^{orig}, and D_2 attempts to distinguish between the generated images $G(I^{bi})$ and original images

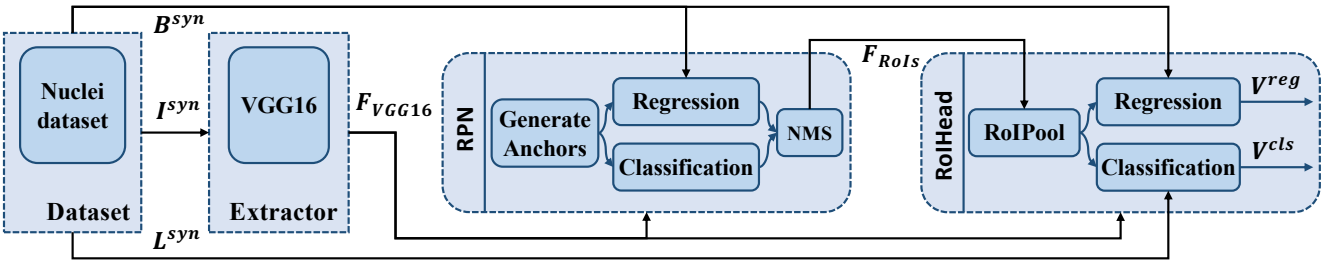


Figure 2. Block diagram of the RCNN-SliceNet

I^{orig}. Similarly, F maps I^{orig} to I^{bi}, and D_1 attempts to discriminate the binary images I^{bi} and translated images $F(I^{orig})$. $G(I^{bi})$ is the translated microscopy like volume and $F(I^{orig})$ is the translated binary like volume. To keep the spatial information consistent, the network H with the same architecture of G is introduced to measure the spatial constraint loss $\mathcal{L}_{spatial}$ between $F(G(I^{bi}))$ and I^{bi}. The loss function of SpCycleGAN is described in Equation 2

$$\begin{aligned}\mathcal{L}(G,F,H,D_2,D_1) &= \mathcal{L}_{GAN}(G,D_2,I^{bi},I^{orig}) \\ &+ \mathcal{L}_{GAN}(F,D_1,I^{orig},I^{bi}) \\ &+ \lambda_1\mathcal{L}_{cycle}(G,F,I^{orig},I^{bi}) \\ &+ \lambda_2\mathcal{L}_{spatial}(G,H,I^{orig},I^{bi}) \quad (2)\end{aligned}$$

where $\mathcal{L}_{GAN}$ and $\mathcal{L}_{cycle}$ are the adversarial and cycle losses [42], and $\mathcal{L}_{spatial}$ is the spatial consistency loss [40].

2.2. RCNN-SliceNet

As shown in Figure 2, the block diagram of the proposed RCNN-SliceNet nuclei localization system can be divided into three parts: training data preparation, region proposal network, and RoIHead network.

The RCNN-SliceNet was modified based on the use of Faster R-CNN model for pneumonia detection on CT images [29, 33]. The 13 convolution layers from a pre-trained VGG16 on ImageNet were used to obtain the features from the input image. The feature map F_{VGG} is of size (C,H,W), where C is the number of channels and H and W represent the height and width of the feature map. F_{VGG} and ground truth bounding boxes B^{syn} are sent into a region proposal network (RPN). The RPN will provide the regions of interest (RoIs) F_{RoIs} from F_{VGG} indicating where might be an object. During training RPN, a sliding window moves on F_{VGG}, generating 9 fixed size anchor boxes a_{1-9} at each sliding position in three scales (s_1^2, s_2^2, s_3^2) and three aspect ratios (r_1, r_2, r_3). s, r are the area, and the ratio of width to height of the anchor a_i. The training loss of RPN $\mathcal{L}_{RPN}$ consists of the bounding box classification loss $\mathcal{L}_{cls}$ and the bounding box regression loss $\mathcal{L}_{reg}$.

$$\mathcal{L}_{RPN} = \mathcal{L}_{cls} + \lambda\mathcal{L}_{reg} \quad (3)$$

where λ is a constant controlling the balance of $\mathcal{L}_{cls}$ and $\mathcal{L}_{reg}$ [29]. $\mathcal{L}_{cls}$ is measured by the negative log softmax function over $K = 2$ categories. Instead of directly regressing the bounding box coordinates, B^{syn} was parameterized for better convergence. The smooth $L1$ loss function defined in [28] was used to measure the bounding box regression loss $\mathcal{L}_{reg}$. Non-maximum Suppression (NMS) was used to remove duplicate boxes.

As shown in Figure 2, the RoIs F_{RoIs} along with B^{syn} are sent to RoIHead network to further estimate the location and category of the nuclei. The RoIPool layer [43, 28] will convert all F_{RoIs} in different dimension to the fixed size feature maps F_{pooled} and F_{pooled} is mapped to two output vectors: softmax probability vector $V_{cls} \in \mathbb{R}^2$ and bounding box coordinates vectors $V_{reg} \in \mathbb{R}^{2\times4}$. The loss function of RoIHead $\mathcal{L}_{RoIHead}$ is the same as $\mathcal{L}_{RPN}$ described in Equation 3.

The RCNN-SliceNet was trained on 2D slices of synthetic microscopy images I^{syn} on the z-direction and tested on original microscopy images I^{orig} along the x-direction, the y-direction, and the z-direction, respectively. The 2D detected centroid coordinates are denoted as C_x^{det2D}, C_y^{det2D}, and C_z^{det2D}.

2.3. 3D Centroid Estimation

Suppose the RCNN-SliceNet is used on the z-direction. The detected nuclei centroids C_z^{det2D} from model **R** are 2D coordinates. For any nucleus in I^{orig}, multiple detected centroids will be generated in C_z^{det2D} due to the appearance of the nucleus in multiple focal planes. Thus, as shown in Figure 1, the first step of 3D centroid estimation is to extend 2D detected centroids coordinates C_z^{det2D} to pseudo 3D detected centroid coordinates C_z^{det3D}. This step is achieved by adding the slice number as coordinates. The overview of 3D AHC clustering is shown in Figure 3. For example, if the RCNN-SliceNet inference is along the z-direction of I^{orig} and there are 7 nuclei were detected on 15^{th} focal plane. The detected 2D centroid coordinates are C_{z15}^{det2D}, and the detected pseudo-3D centroid coordinates C_{z15}^{det3D} is generated by adding 15 as the z coordinates. Similarly, if the RCNN-SliceNet inference is along the x-direction of I^{orig}, the detected 2D centroid coordinates on the p^{th} fo-

cal plane is extended to the pseudo 3D centroid coordinates $C_{x_p}^{det3D}$ by adding p as the x coordinates. To estimate the

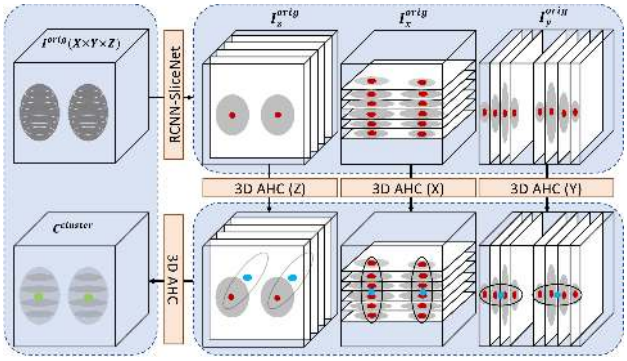


Figure 3. Overview of the RCNN-SliceNet detection and 3D AHC clustering for nuclei centroid estimation

3D nuclei centroid, the agglomerative hierarchical clustering (AHC) [44] with average linkage criterion is used to obtain structural information from C_z^{det3D}. AHC tries to create nested clusters by greedily merging a pair of clusters that have a similar property. The average linkage criterion forces AHC to minimize the average distances between all cluster pairs. The Lance-Williams dissimilarity [45] L in Equation 4 is used to estimate the dissimilarity between the newly merged cluster $i \cup j$ and an external cluster e. AHC will initially treat each point as a cluster and iteratively estimate all inter-points dissimilarities and form clusters from two closest points. Eventually all points will be merged as one cluster.

$$L(i \cup j, e) = \alpha_i L(i, e) + \alpha_j L(j, e) + \beta L(i, j) + \gamma |L(i, e) - L(j, e)| \quad (4)$$

where α_i, α_j, β, γ define the agglomerative criterion, and $|\cdot|$ represents the absolute value.

The number of clusters $N_z^{cluster}$ is estimated using the Silhouette Coefficient [46]. For a given number of clusters $k, k \in [1, n]$, the mean intracluster distance $a(i)$ and the average of the mean intercluster distance $b(i)$ for each point $i \in C_z^{det3D}$ is estimated using Equation 5.

$$a(i) = \frac{1}{n_c - 1} \sum_{i,j \in C_c, i \neq j} d(i, j)$$
$$b(i) = \frac{1}{k - 1} \sum_{q, q \neq c} \left(\frac{1}{n_q} \sum_{i \in C_c, j \in C_q} d(i, j) \right) \quad (5)$$

$a(i)$ measures the distance between i and all other samples in its cluster. c denotes the cluster where i is in, n_c is the number of points in cluster c, C_c are all points in cluster c, and $d(i, j)$ is the Euclidean distance between i and j. $b(i)$ measures the distance between i and other clusters q.

$$SC_k = \frac{1}{n} \sum_{i \in C_z^{det3D}} \frac{b(i) - a(i)}{\max(a(i), b(i))}$$
$$N_z^{cluster} = \underset{k}{\operatorname{argmax}}\, SC_k, k = 1, ..., n \quad (6)$$

The Silhouette Coefficient SC_k when cluster number is k is given by Equation 6. The optimal number of clusters $N_z^{cluster}$ is the number of clusters that corresponds to the highest Silhouette Coefficient, and the cluster centroid $C_z^{cluster}$ are the centroid coordinates of all clusters. Similarly, the $C_x^{cluster}$, $N_x^{cluster}$, and $C_y^{cluster}$, $N_y^{cluster}$ are obtained by repeating the entire process on the x-direction and the y-direction of the volume.

In the final step, 3-way majority voting was used to better estimate the cluster centroid. Three-way majority voting is achieved by using another 3D AHC that cluster the points of $C_x^{cluster}$, $C_y^{cluster}$, and $C_z^{cluster}$. Ultimately, we use $C^{cluster}$ as the centroid coordinates of the nuclei and use $N^{cluster}$ as the number of nuclei in a 3D volume.

3. Experimental Results

3.1. Datasets

We used four different types of microscopy data for evaluation, denoted as Data-I, Data-II, Data-III and Data-IV. The Data-I and Data-II are fluorescent-labeled (Hoechst 33342 stain) nuclei collected using two-photon fluorescent microscopy of rat kidneys. Data-I consists of one volume in size of $X \times Y \times Z = 128 \times 128 \times 64$ pixels, Data-II consists of 16 volumes in size of $128 \times 128 \times 32$ pixels, Data-III consists of 21 volumes in size of $128 \times 128 \times 19$ pixels, and Data-IV consists of 90 volumes in dimension of $807 \times 565 \times 129$. Data-I, Data-II and Data-III were provided by Malgorzata Kamocka and Kenneth W. Dunn of Indiana University and were collected at the Indiana Center for Biological Microscopy [47]. Data-IV is BBBC024vl [48] from the Broad Bioimage Benchmark Collection.

During the synthetic binary volume generation, 54 synthetic binary volumes ($I^{bi01} - I^{bi54}$) for Data-I, Data-II and Data-III are generated, respectively. The semi-axes **a** are randomly chosen in the range of 4 and 8 for synthetic Data-I, 10 and 14 for synthetic Data-II, 6 and 10 for synthetic Data-III. The corresponding numbers of candidate nuclei N^{syn} in a volume are set to 400, 40, and 50, and the allowed overlapping threshold T_{ov} are set to 5, 10, and 10, respectively. $I^{bi001} - I^{bi004}$ along with the corresponding original microscopy data I^{orig} are used to train the SpCycleGAN. The rest of synthetic binary volumes $I^{bi005} - I^{bi054}$ are inferenced on the trained SpCycleGAN to generate synthetic microscopy data $I^{syn005} - I^{syn054}$. All synthetic binary volumes I^{bi} and synthetic microscopy volumes I^{syn} are in size $X \times Y \times Z = 128 \times 128 \times 128$.

Both the SpCycleGAN and RCNN-SliceNet were implemented using PyTorch. The SpCycleGAN was trained using Adam optimizer with a constant learning rate 0.0002 for the first 100 epochs and linearly decayed to 0 in the next 100 epochs. The generators G, F, and H use 9 blocks of ResNet. The loss parameters are set to $\lambda_1 = \lambda_2 = 10$. As shown in Figure 4, the SpCycleGAN generates synthetic microscopy images that look like the original microscopy images. It can maintain the size and the shape of the nuclei from the binary mask. The RCNN-SliceNet

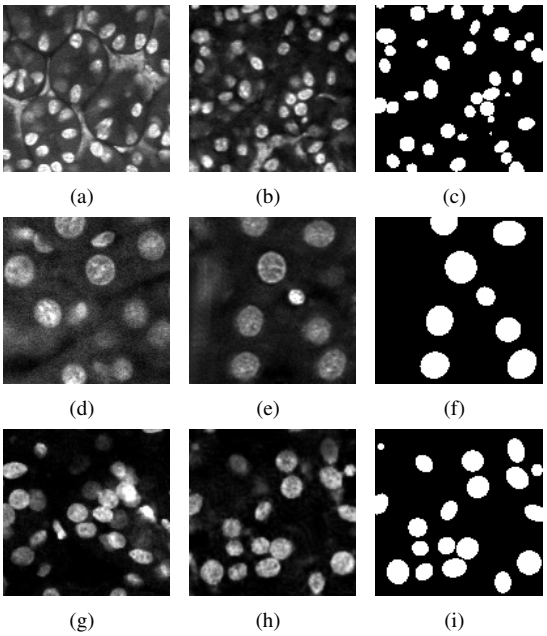


Figure 4. Synthetic microscopy images (second column) generated based on the original microscopy images (first column) and the synthetic ground truth masks (third column) using SpCycleGAN. First row is Data-I, second row is Data-II, and third row is Data-III

was trained on synthetic volumes and tested on original microscopy volumes. The 3D ground truth masks for the original microscopy volumes are manually annotated using ITK-SNAP [49].

For Data-I evaluation, all models are trained on synthetic Data-I and tested on original Data-I. Similarly, for Data-II evaluation, all models are trained on synthetic Data-II and tested on original Data-II. For Data-III, we use transfer learning that all models are pre-trained on synthetic Data-III and cross validated on original Data-III. During cross validation, the 21 volumes of original Data-III are randomly divided into 3 equal sets. We then recursively update the pre-trained model on one set and test on other two sets. For Data-IV evaluation, we test the generalizability of these models from one type of microscopy data to another. Thus, we use the model that trained on synthetic Data-III and inference it on Data-IV.

During RCNN-SliceNet training, all slices along the z-direction of 30 synthetic volumes $I^{syn005} - I^{syn054}$ are

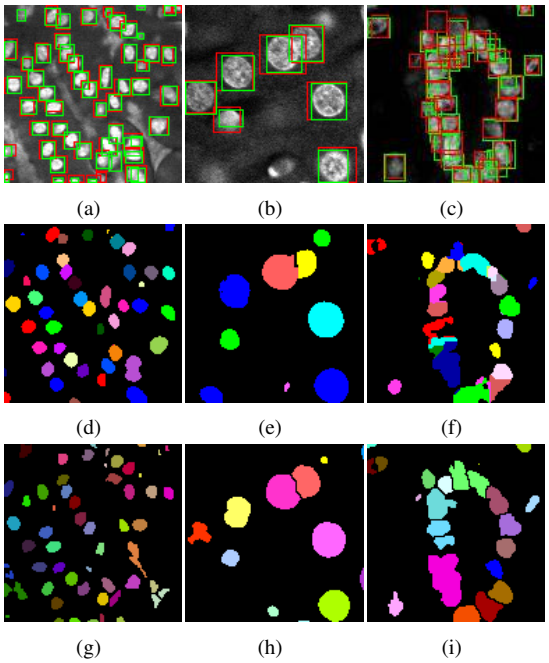


Figure 5. Examples of RCNN-SliceNet detection (first row), 3D U-Net segmentation (second row), and VTEA segmentation (third row) on original Data-I (first column), Data-II (second column) and Data-III (third column). The ground truth bounding box is in green and the detected bounding box is in red

used for training and validating RCNN-SliceNet. The validation set is randomly chosen 20% from the entire training set. Scales and aspect ratios of the three anchors are set to $(64^2, 128^2, 256^2)$, and $(0.5, 1, 2)$ based on the nuclei size. The RCNN-SliceNet model was trained for 30 epochs with the Stochastic Gradient Descent (SGD) optimizer and an initial learning rate of 0.001 that decays 50% every 5 epochs. The training images are normalized using Caffe normalization method and randomly flipped horizontally and vertically while training. During testing, the 3-way inference was used to run RCNN-SliceNet along the x-direction, the y-direction, and the z-direction of a volume. Note that for Data-III and Data-IV, due to the small number of slices on the z-direction, we only used RCNN-SliceNet on the z-direction.

3.2. Evaluation Metrics

The Mean Absolute Percentage Error (MAPE) is used to estimate the counting accuracy.

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{N_i^{cluster} - N_i^{gt}}{N_i^{gt}} \right| \qquad (7)$$

where N is the number of volumes, $N_i^{cluster}$ represents the estimated number of nuclei in i^{th} volume, and N_i^{gt} is the ground truth number of nuclei for i^{th} volume. The centroid-based nuclei detection accuracy is evaluated using

Table 1. Evaluation of nuclei counting and centroid detection on original Data-I, Data-II and Data-III. The counting accuracy is evaluated using the mean absolute percentage error (MAPE%). The centroid-based accuracy is evaluated using average precision (AP%) and mean Average precision (mAP%)

	Microscopy Data-I				Microscopy Data-II				Microscopy Data-III			
	MAPE	AP_4	AP_8	mAP	MAPE	AP_6	AP_{12}	mAP	MAPE	AP_6	AP_{10}	mAP
ImageJ[10]	43.66	10.86	21.69	16.48	16.67	41.88	57.23	50.22	40.25	35.43	41.32	38.68
3D Watershed	14.08	55.25	70.66	65.75	44.34	58.40	65.04	63.38	43.02	35.91	41.83	39.54
Squassh[16]	76.06	9.97	14.38	12.67	17.63	44.05	55.13	50.62	37.91	33.76	42.38	38.57
CellProfiler[13]	5.28	41.39	56.95	51.66	28.40	62.46	66.70	65.55	32.74	44.00	58.02	51.73
VTEA[20]	13.73	30.20	42.68	37.84	14.06	60.65	63.14	61.77	11.00	63.02	69.72	66.87
V-Net[24]	17.61	73.78	80.70	78.97	27.16	49.87	56.41	54.26	11.12	61.13	69.92	66.87
3D U-Net[25]	20.77	73.60	78.12	76.95	15.57	57.74	63.74	62.16	12.77	60.58	70.68	66.50
DeepSynth[26]	8.80	74.24	81.98	80.00	14.49	64.06	69.45	68.12	17.10	61.26	70.24	66.86
Proposed	**1.41**	**86.41**	**87.86**	**87.52**	**9.76**	**75.63**	**76.42**	**76.11**	**8.54**	**73.91**	**80.97**	**78.50**

Table 2. Evaluation results of microscopy Data-IV using pre-trained models on synthetic microscopy Data-III

	Microscopy Data-IV			
	MAPE	AP_{15}	AP_{25}	mAP
V-Net[24]	1.50	95.41	95.67	95.55
3D U-Net[25]	2.00	94.30	94.40	94.34
DeepSynth[26]	2.78	91.31	91.94	91.61
Proposed	**1.11**	**98.71**	**98.80**	**98.79**

Average Precision (AP) and mean Average Precision (mAP) inspired by [50]. If the Euclidean distance between an estimated centroid and a ground truth centroid is less than T_{dist}, then we say the estimated centroid and ground truth centroid are matched. We use greedy matching which means each ground truth centroid always matches with its nearest detected centroid. We then estimate the True Positives (number of matched estimated centroids and ground truth centroids), False Positives (number of estimated centroids that have no associated ground truth centroids matched), False Negatives (number of ground truth centroids that have no associated estimated centroids matched), and the total number of nuclei in a volume, respectively. Also, we define AP_t as the average precision when T_{dist} is set to t where $t \in \{4, 5, ..., 8\}$, $\{6, 7, ..., 12\}$, $\{6, 7, ..., 10\}$, and $\{15, 16, ..., 25\}$ for Data-I, Data-II, Data-III, and Data-IV. The AP is estimated by the area under the precision/recall curve [50] and the mAP is estimated using Equation 8. For every detected centroid, we assume their confidences are all 1.

$$\mathrm{mAP} = \frac{1}{|T_{dist}|} \sum_{t \in T_{dist}} \mathrm{AP}(t) \tag{8}$$

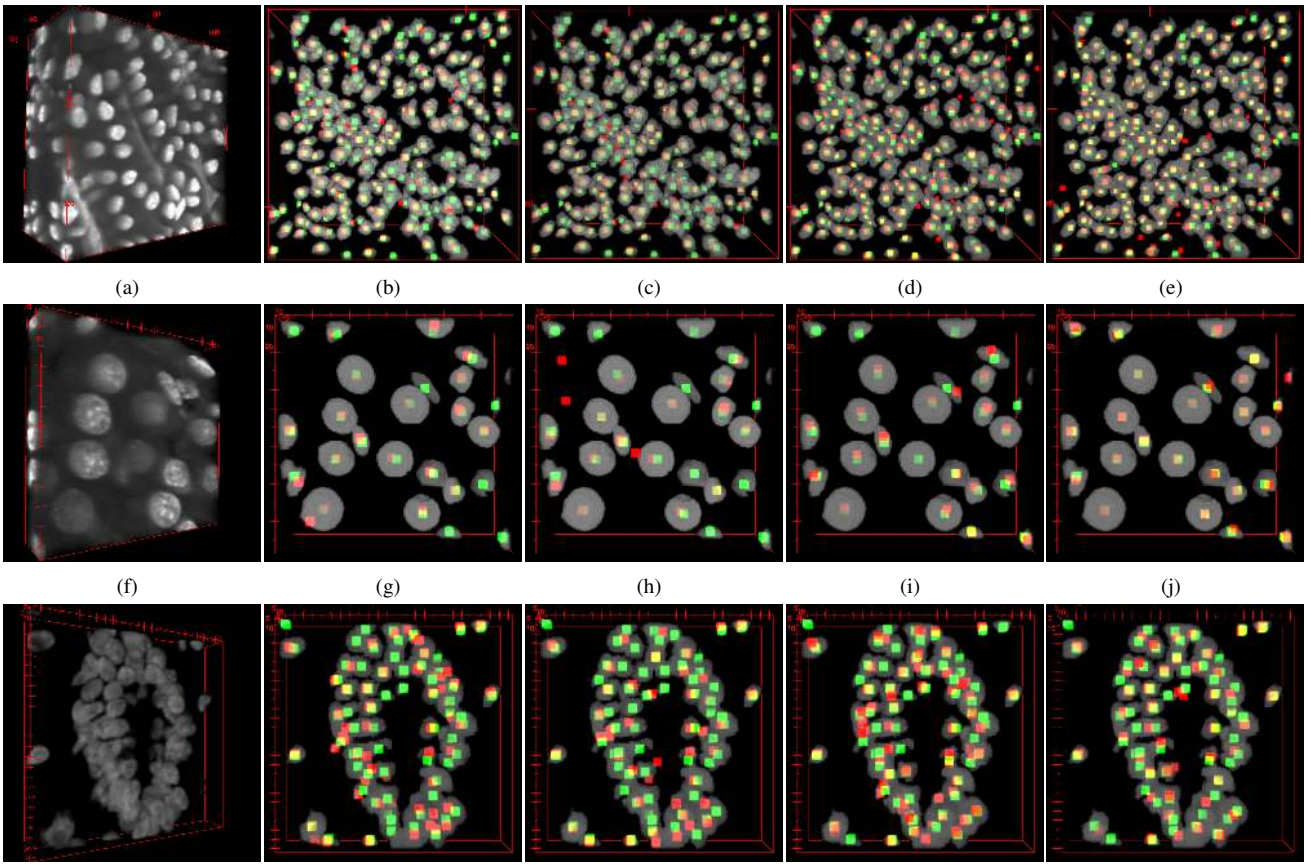
where $\mathrm{AP}(t)$ is the Average Precision given a distance threshold $T_{dist} = t$. The evaluation accuracy is shown in Table 1.

3.3. Discussion

Our proposed method was compared with 3D marker-controlled watershed, which was inspired by [19], V-Net [24], 3D U-Net [25], DeepSynth [26], and other commonly used segmentation methods. For VTEA [20], ImageJ's JACoP [10], Squassh [16], and CellProfiler [13], we used contrast enhancement, background subtraction, and gaussian blur to pre-process the image. For V-Net and 3D U-Net, we optimize the results by using marker-controlled 3D watershed as the post-processing to separate touching nuclei, the markers are obtained by using conditional erosion [19]. Finally, 3D connected components test was used to estimate the total number of nuclei and the nuclei centroids.

For our proposed method, we observed some outliers from the output of RCNN-SliceNet that may affect the evaluation accuracy. We use an extra 3D AHC for majority voting to remove outliers. As shown in Figure 5, RCNN based method can easily distinguish different nuclei and can still capture small partially included objects. The segmentation-based methods V-Net, 3D U-Net and DeepSynth achieve good segmentation accuracy but post-processing such as watershed is required to separate different objects. As shown in Table 1, our proposed method outperforms all of the baseline methods. Table 2 shows that our method achieved best generalization results on original Data-IV using the pre-trained model on synthetic Data-III. Figure 6 shows that our method made fewer mistakes when nuclei are overlapping together and can detect the small nuclei even they are on the border of the volume.

Advantages By making use of RCNN, our system can easily distinguish touching nuclei without any post-processing steps such as watershed or morphological operation. Our method can better capture small nuclei and partially included nuclei on the border (see Figure 6 (j)). The slice-and-cluster strategy is robust handling the errors from RCNN-SliceNet's misdetections on some slices. The 3D AHC will treat RCNN-SliceNet's misdetection as outliers and 3-way majority voting will further align the centroid location and remove outliers. Our methods achieved better results than segmentation-based methods on ellipsoidal nuclei data.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

Figure 6. (a), (f), (k) are example testing volumes from original microscopy Data-I, Data-II, and Data-III. The nuclei centroid estimation results for 3D U-Net (second column), V-Net (third column), DeepSynth (fourth column), and proposed method (fifth column). The red cubes are the estimated centroids, the green cubes are ground truth centroids, and the yellow is the overlay of green cubes and red cubes. The gray spheroid is the ground truth mask of the nuclei. The volumes are visualized using the ImageJ's 3D Viewer

Limitations Our method is under the assumption that nuclei are ellipsoidal shape. We observed the 3D AHC made more mistakes on clustering centroids for non-ellipsoidal nuclei volumes. Also, the 3D AHC requires a rough estimation of the number of nuclei in a volume to accelerate the inference time. For example, between 10 and 100. Our method is subject to over-detection problem and the error is mainly from false positive detection due to the imbalanced training samples.

4. Conclusions

In this paper, we presented a nuclei counting and localization method using synthetic microscopy volume generated from SpCycleGAN. We proposed a scalable framework RCNN-SliceNet using the slice-and-cluster strategy to detect each instance of nucleus on each focal plane. Our proposed system extends RCNN capability from 2D images to 3D volumes by using 3D AHC to estimate the centroid information of each nucleus. The experiments show that the proposed system can successfully distinguish in focus and out of focus nuclei as well as the background structures. Our method can accurately count the number of nuclei and estimate the nuclei centroids in a 3D microscopy volume and outperforms the rest of the methods. In the future, we plan to extend SpCycleGAN so it can generate different types of synthetic nuclei in one volume for training the RCNN-SliceNet, and we will improve and redesign our method to count and localize different types of nuclei in a single 3D volume.

5. Acknowledgments

This work was partially supported by a George M. O'Brien Award from the National Institutes of Health under grant NIH/NIDDK P30 DK079312 and the endowment of the Charles William Harrison Distinguished Professorship at Purdue University. The authors have no conflicts of interest. Address all correspondence to Edward J. Delp, ace@ecn.purdue.edu

References

[1] K. Dunn, R. Sandoval, K. Kelly, P. Dagher, G. Tanner, S. Atkinson, R. Bacallao, and B. Molitoris, "Functional studies of the kidney of living animals using multicolor two-photon microscopy," *American Journal of Physiology-Cell Physiology*, vol. 283, no. 3, pp. C905–C916, September 2002. 1

[2] M. J. Booth, "Adaptive optics in microscopy," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1861, pp. 2829–2843, September 2007. 1

[3] R. K. P. Benninger and D. W. Piston, "Two-photon excitation microscopy for the study of living cells and tissues," *Current Protocols in Cell Biology*, vol. 59, no. 1, pp. 4.11.1–4.11.24, June 2013. 1

[4] W. Denk, J. H. Strickler, and W. W. Webb, "Two-photon laser scanning fluorescence microscopy," *Science*, vol. 248, no. 4951, pp. 73–76, April 1990. 1

[5] Y. Xie, X. Kong, F. Xing, F. Liu, H. Su, and L. Yang, "Deep voting: A robust approach toward nucleus localization in microscopy images," *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, pp. 374–382, November 2015, Munich, Germany. 1

[6] T. Hayakawa, V. B. S. Prasath, H. Kawanaka, B. J. Aronow, and S. Tsuruoka, "Computational nuclei segmentation methods in digital pathology: a survey," *Archives of Computational Methods in Engineering*, pp. 1–13, October 2019. 1

[7] K. Y. Win and S. Choomchuay, "Automated segmentation of cell nuclei in cytology pleural fluid images using otsu thresholding," *Proceedings of the International Conference on Digital Arts, Media and Technology*, pp. 14–18, March 2017, Chiang Mai, Thailand. 1

[8] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, January 1979. 1

[9] C. T. Rueden, J. Schindelin, M. C. Hiner, B. E. DeZonia, A. E. Walter, E. T. Arena, and K. W. Eliceiri, "Imagej2: Imagej for the next generation of scientific image data," *BMC bioinformatics*, vol. 18, no. 1, pp. 529–554, November 2017. 1

[10] S. Bolte and F. P. Cordelières, "A guided tour into subcellular colocalization analysis in light microscopy," *Journal of Microscopy*, vol. 224, no. 3, pp. 213–232, December 2006. 1, 7

[11] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 841–852, April 2010. 1

[12] C. S. Bjornsson, G. Lin, Y. Al-Kofahi, A. Narayanaswamy, K. L. Smith, W. Shain, and B. Roysam, "Associative image analysis: a method for automated quantification of 3D multi-parameter images of brain tissue," *Journal of Neuroscience Methods*, vol. 170, no. 1, pp. 165–178, May 2008. 1

[13] C. McQuin, A. Goodman, V. Chernyshev, L. Kamentsky, B. A. Cimini, K. W. Karhohs, M. Doan, L. Ding, S. M. Rafelski, D. Thirstrup, W. Wiegraebe, S. Singh, T. Becker, J. C. Caicedo, and A. E. Carpenter, "Cellprofiler 3.0: Next-generation image processing for biology," *PLoS biology*, vol. 16, no. 7, pp. e2 005 970–1–17, July 2018. 1, 7

[14] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, January 1988. 1

[15] G. Paul, J. Cardinale, and I. F. Sbalzarini, "Coupling image restoration and segmentation: a generalized linear model/bregman perspective," *International Journal of Computer Vision*, vol. 104, no. 1, pp. 69–93, March 2013. 1

[16] A. Rizk, G. Paul, P. Incardona, M. Bugarski, M. Mansouri, A. Niemann, U. Ziegler, P. Berger, and I. F. Sbalzarini, "Segmentation and quantification of subcellular structures in fluorescence microscopy images using squassh," *Nature Protocols*, vol. 9, no. 3, pp. 586–596, March 2014. 1, 7

[17] S. Lee, P. Salama, K. W. Dunn, and E. J. Delp, "Segmentation of fluorescence microscopy images using three dimensional active contours with inhomogeneity correction," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 709–713, April 2017, Melbourne, Australia. 1

[18] P. Soille and L. M. Vincent, "Determining watersheds in digital pictures via flooding simulations," *Visual Communications and Image Processing'90: Fifth in a Series*, vol. 1360, pp. 240–250, September 1990. 2

[19] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 11, pp. 2405–2414, November 2006. 2, 7

[20] S. Winfree, S. Khan, R. Micanovic, M. T. Eadon, K. J. Kelly, T. A. Sutton, C. L. Phillips, K. W. Dunn, and T. M. El-Achkar, "Quantitative three-dimensional tissue cytometry to study kidney tissue and resident immune cells," *Journal of the American Society of Nephrology*, vol. 28, no. 7, pp. 2108–2118, July 2017. 2, 7

[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, December 2017. 2

[22] C. Fu, D. J. Ho, S. Han, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei segmentation of fluorescence microscopy images using convolutional neural networks," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 704–708, April 2017, Melbourne, Australia. 2

[23] 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," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 834–842, July 2017, Honolulu, HI. 2, 3

[24] F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *Proceedings of the International Conference on 3D Vision*, pp. 565–571, October 2016. 2, 7

[25] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3D u-net: Learning dense volumetric segmentation from sparse annotation," *Medical Image Computing and Computer-Assisted Intervention*, vol. 9901, pp. 424–432, October 2016. 2, 7

[26] K. W. Dunn, C. Fu, D. J. Ho, S. Lee, S. Han, P. Salama, and E. J. Delp, "Deepsynth: Three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data," *Scientific Reports*, vol. 9, no. 1, pp. 18 295–18 309, December 2019. 2, 7

[27] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, June 2014, Columbus, OH. 2

[28] R. Girshick, "Fast r-cnn," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, December 2015, Santiago, Chile. 2, 4

[29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, June 2017. 2, 4

[30] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," *Proceedings of the European Conference on Computer Vision*, vol. 9905, pp. 21–37, October 2016, Amsterdam, The Netherlands. 2

[31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, June 2016, Las Vegas, NV. 2

[32] J. Hung and A. Carpenter, "Applying faster r-cnn for object detection on malaria images," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 56–61, July 2017, Honolulu, HI. 2

[33] L. Wu, "Biomedical image segmentation and object detection using deep convolutional neural networks," M.S. dissertation, Purdue University, Hammond, IN, May 2019. 2, 4

[34] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum pointnets for 3D object detection from RGB-D data," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 918–927, June 2018, Salt Lake City, UT. 2

[35] C. R. Qi, O. Litany, K. He, and L. Guibas, "Deep hough voting for 3D object detection in point clouds," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9276–9285, November 2019, Seoul, Korea (South). 2

[36] A. Yang, F. Pan, V. Saragadam, D. Dao, Z. Hui, J. R. Chang, and A. C. Sankaranarayanan, "Slicenets - a scalable approach for object detection in 3d ct scans," *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pp. 335–344, January 2021, Waikoloa, Hawaii. 2

[37] D. M. Montserrat, Q. Lin, J. Allebach, and E. J. Delp, "Training object detection and recognition cnn models using data augmentation," *Electronic Imaging*, vol. 2017, no. 10, pp. 27–36, January 2017. 2

[38] E. Castro, J. S. Cardoso, and J. C. Pereira, "Elastic deformations for data augmentation in breast cancer mass detection," *Proceedings of the IEEE EMBS International Conference on Biomedical & Health Informatics*, pp. 230–234, March 2018, Las Vegas, NV. 2

[39] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield, "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 969–977, June 2018, Salt Lake City, UT. 2

[40] C. Fu, S. Lee, D. J. Ho, S. Han, P. Salama, K. W. Dunn, and E. J. Delp, "Three dimensional fluorescence microscopy image synthesis and segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2302–2310, June 2018, Salt Lake City, UT. 2, 3, 4

[41] A. Chen, L. Wu, S. Han, P. Salama, K. W. Dunn, and E. J. Delp, "Three dimensional synthetic non-ellipsoidal nuclei volume generation using bezier curves," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, April 2021, Nice, France. 2

[42] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2242–2251, October 2017, Venice, Italy. 3, 4

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, January 2015. 4

[44] F. Murtagh, "A survey of recent advances in hierarchical clustering algorithms," *The Computer Journal*, vol. 26, no. 4, pp. 354–359, November 1983. 5

[45] F. Murtagh and P. Legendre, "Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion?" *Journal of Classification*, vol. 31, no. 3, pp. 274–295, October 2014. 5

[46] H. Zhou and J. Gao, "Automatic method for determining cluster number based on silhouette coefficient," *Advanced Materials Research*, vol. 951, pp. 227–230, May 2014. 5

[47] "Indiana Center for Biological Microscopy." [Online]. Available: http://web.medicine.iupui.edu/icbm/ 5

[48] D. Svoboda, M. Kozubek, and S. Stejskal, "Generation of digital phantoms of cell nuclei and simulation of image formation in 3d image cytometry," *Cytometry Part A: The Journal of the International Society for Advancement of Cytometry*, vol. 75, no. 6, pp. 494–509, June 2009. 5

[49] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, July 2006. 6

[50] M. Everingham, S. M. A. Eslami, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, January 2015. 6, 7