A Review of Nuclei Detection and Segmentation on Microscopy Images using Deep Learning

With Applications to Unbiased Stereology Counting

Saeed S. Alahmari, *Member, IEEE,* Dmitry Goldgof, *Fellow, IEEE,* Lawrence O. Hall, *Fellow, IEEE,* and Peter R. Mouton

***Abstract*—The detection and segmentation of stained cells and nuclei are an essential prerequisite for subsequent quantitative research for many diseases. Recently, deep learning has shown strong performance in many computer vision problems, including solutions for medical image analysis. Furthermore, accurate stereological quantification of microscopic structures in stained tissue sections plays a critical role in understanding human diseases and developing safe and effective treatments. In this paper, we review the most recent deep learning approaches for cell (nuclei) detection and segmentation in cancer and Alzheimer disease with emphasis on deep learning approaches combined with unbiased stereology. Major challenges include accurate and reproducible cell detection and segmentation of microscopic images from stained sections. Finally, we discuss potential improvements and future trends in deep learning applied to cell detection and segmentation.**

***Index Terms*—Nuclei Segmentation, Nuclei Detection, Nuclei, Cell, Microscopy Image Analysis, unbiased stereology, Deep Learning, Convolutional Neural Network, Generative Adversarial Network, Active Learning.**

I. Introduction

THE combination of microscopic imaging, machine learning and computer-aided diagnosis is emerging as a powerful tool for quantifying and classifying cell and tissue abnormalities. Accurate detection and segmentation of individual nuclei and cells could provide rigorous quantitative morphological analysis to guide the diagnostic decisions of clinicians, drive biomedical research and develop novel strategies for the therapeutic management of disease [1][2]. Traditional approaches by pathologists and cytopathologists lack the feasibility for accurate detection and segmentation of microscopic structures with variable staining and tissue structures, conjoined and deformed cells, cell densities, etc. Therefore, a major research concern is the development of an accurate, reproducible and automatic method for detection and segmentation of nuclei and cells [3].

The physical appearance of stained cells and nuclei on microscopic images presents serious challenges for machine learning algorithms. The performance of conventional machine learning methods is limited by their dependence on extracting engineered features that describe the data, i.e., raw microscopic images. Designing and engineering representative features to learn to do cell detection and segmentation of an image is a daunting process, given the complexity of biological structures in microscopy images, including cell variation, complexity of cell texture, and background texture. The recent emergence of deep learning shows relatively superior performance on a wide range of diverse tasks in computer vision [4][5], speech recognition [6][7], natural language processing [8][4], and medical image analysis [9][10]. LeCun et al. [11] define deep learning as "representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level"[11] (page 1).

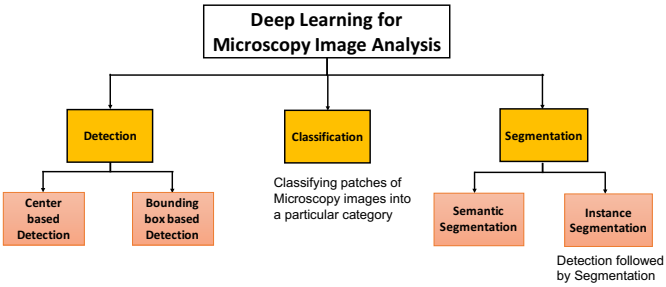


Fig. 1: The main tasks of deep learning applied to microscopy image analysis

In this paper, we review the most recent deep learning approaches for cell detection and cell segmentation which were published from 2016 through the end of 2019. A classification of deep learning applications in microscopy images is shown in Fig. 1. In this paper, we focus only on cell detection and segmentation. We defined two types of cell detection: near-center detection and cell body detection using a bounding-box approach. For cell segmentation, two types are defined: semantic segmentation and instance segmentation. A visual summary of the papers reviewed in this manuscript is shown in Fig. 2. Additionally, we survey recent work in unbiased stereology, provide details of challenges and discuss potential future endeavors. We offer a brief highlight of the basics of deep learning in Appendix A (Supplementary Material); for further information on the topic of deep learning, the interested reader can refer to [4][11]. There have been some reviews that

S. Alahmari is assistant Professor at the Computer Science department of the Najran University, Najran,66462 KSA e-mail: ssalahmari@nu.edu.sa

D. Goldgof and L. Hall are distinguished Professors and researchers at the Computer Science and Engineering department of the University of South Florida, Tampa-FL,33620 USA

P.R. Mouton is Chief Scientific Officer of SRC Biosciences in Tampa, Florida and Courtesy Faculty in the Computer Science and Engineering department of the University of South Florida, Tampa-FL,33620 USA

Manuscript received TBD; revised TBD.

surveyed earlier work in microscopy image analysis, such as [12] [13] [14], which summarized deep learning approaches for detection, classification, segmentation, and registration in microscopy images. To the best of our knowledge, there have been no reviews of deep learning and applicability to cell detection and segmentation for unbiased stereology.

II. Unbiased Stereology

Unbiased stereology is the field of study concerned with the accurate quantitative analysis of objects in three-dimensional (3D) space. The accuracy is achieved through the use of sampling methods and geometric probes specifically designed to avoid all known sources of non-random error (bias) when making estimates of first order stereology parameters [Volume, Surface Area, Length, Number] and their variation [16]. For example, making an unbiased estimate of the total number of objects (e.g., cells, nuclei) requires several steps: define a reference space; sample the reference space in a systematic-random manner; count the number of stained objects using a 3-D virtual probe (disector) using unbiased counting rules; and scale the resultant count to the entire reference space [17] [18]. Thin focal plane sampling using the disector probe avoids the Corpuscle problem introduced by sampling and counting 3D objects based on their appearance on 2D planes (i.e., profiles) through the reference space [19]. Unbiased stereology counting on a stack of images is illustrated in Fig. 3.

With manual stereology to obtain ground truth counts, a trained technician slowly focuses from the top to the bottom of each disector stack while counting each object of interest as it comes into focus. The technician repeats this thin focal plane scanning process for 200 disector stacks spaced in a systematic-random manner through the entire reference space. The six-sided disector probe has three exclusion planes (left, bottom, first z-plane) and three inclusion planes (right, top, last z-plane; 2D disector frame is shown in Fig. 4). It is worth noting that exclusion lines (planes in 3D) extend past the upper left, and lower right corners to avoid objects being counted twice (i.e., edge effects), as depicted in Fig. 4. In the applications of deep learning to unbiased stereology, stacks of images through each disector volume (disector stacks) are captured automatically using standard computer-assisted stereology equipment (i.e., microscope with motorized three-axis stage and digital imaging (Stereologer, SRC Biosciences, Tampa, FL)). Each disector stack consists of multiple z-axis planes with one micron spacing.

The disector frame shown in Fig. 4, determine whether a cell should be counted or not based on the exclusion line (red left and bottom lines). If a cell is touching the exclusion line then it is not counted based on unbiased stereology rules, otherwise, cells within the disector frame or touching the inclusion line (green top and right lines) are counted based on unbiased stereology rules. In-order to automate the process of unbiased stereology counting, each cell has to be identified accurately, before applying unbiased stereology counting rules based on disector frame. Therefore, we reviewed segmentation and detection methods in this paper, and we discussed the applicability of these methods to identify cells for unbiased stereology counting.

III. Nuclei Analysis approaches using Deep Learning

In this section, we summarize deep learning approaches applied to microscopy images for cell detection and cell segmentation. We focus on two aspects in this section: 1) deep neural networks that detect part of a cell, which can be either near center cell detection (i.e., localization) or entire cell body detection using a bounding box based detection approach. 2) deep neural networks for segmentation, which can be semantic segmentation where objects are segmented from the background without identifying each object's mask (pixel-wise classification), or instance segmentation, where each cell in a microscopy image is detected then segmented to provide a mask for each detected cell body.

A. Detection

Detection of a nuclei (cell) using deep learning is an essential step in computer-assisted microscopy image analysis, such as cancer tissue grading. Hence, an accurate cell detector plays a significant role in the diagnosis and study of many diseases. The diversity of shape, crowdedness (overlap), stain, and appearance variabilities of nuclei complicate the process of detection-based analysis. Deep learning has generated much interest versus conventional computer vision algorithms in detecting cells in images of stained histopathology sections. There have been many proposed works of CNN based architectures to localize nuclei (cell) location (e.g., centroids or an enclosed bounding box of nuclei). This section discusses cell-based detection methods in detail. Tables I and II summarize papers reviewed in this section.

1) Nuclei near-center based detection:

*a) **Stack autoencoder based methods**:* Sparse autoencoder (SAE) is an unsupervised neural network autoencoder with sparsity penalty [4], which consists of encoding the input x to latent high level features h, and reconstructing the learned features (decoder) to the output $x^{'}$. In medical image analysis where labeled data is hard to obtain, an autoencoder plays a significant role in learning features that can be used for classification and detection of nuclei.

In [20], a stacked sparse autoencoder (SSAE) was proposed to detect nuclei in whole slide images of breast cancer tissues using a sliding window. SSAE was trained to learn the high-level features, and then a softmax classifier was used as the output layer to determine if a patch 34*34 contains cancer nuclei (positive) or not (negative). If a patch contains nuclei, then the center of the patch is marked as detected. Additionally, SSAE was compared with a shallow, sparse autoencoder as well as traditional methods for nuclei detection such as Expectation-Maximization active contour [21] and Blue Ratio thresholding [22]. The results of SSAE with a softmax layer show superior results of 84.49% (F-measure) and average precision (AveP) of 78.83% on an H&E stained breast histopathology images dataset [20]. A similar approach was applied to Pap smear images of cervical cancer (ISBI14 dataset) to detect and classify image patches extracted using a sliding window approach with different sizes [23]. This approach uses SSAE followed by

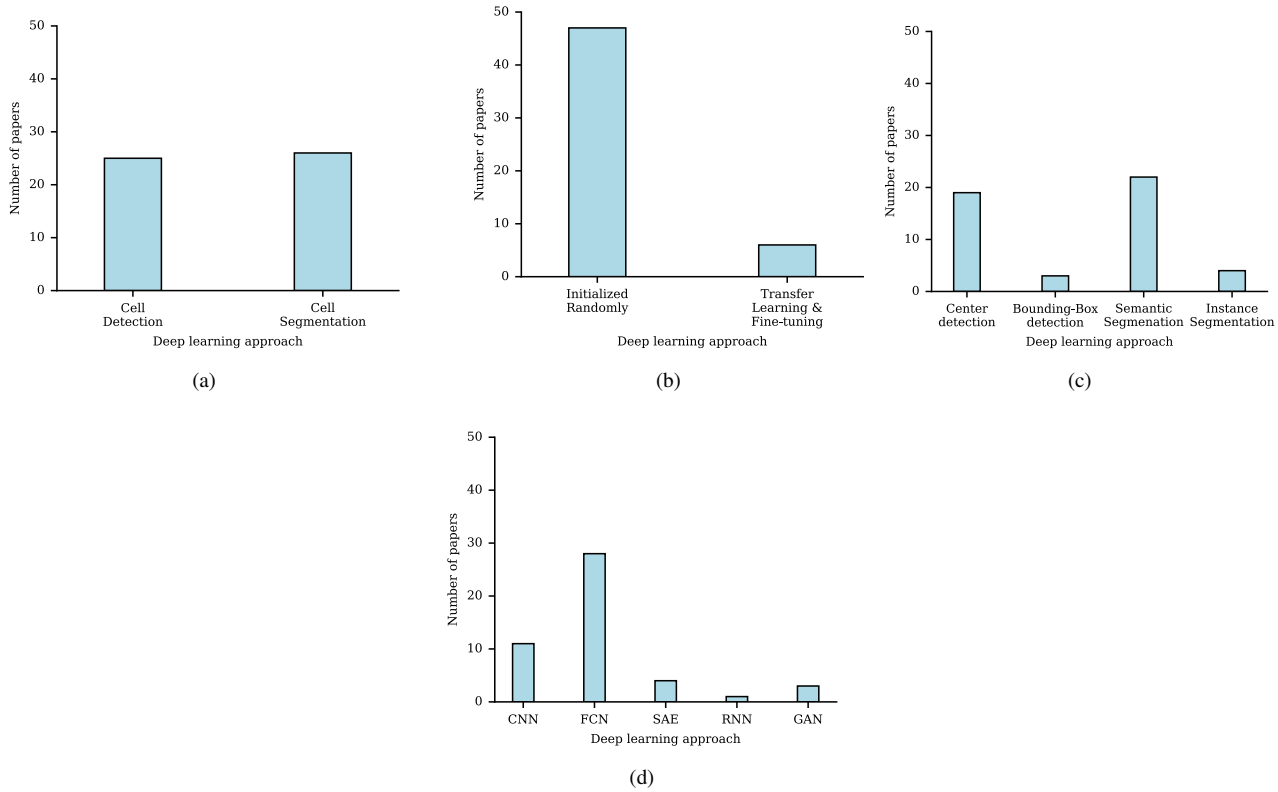


Fig. 2: Summary of papers reviewed in this manuscript: (a) number of papers published recently for cell detection and cell segmentation in microscopy images, (b) number of papers that use transfer learning and tuning versus using random initialization, (c) number of papers that use near-center based cell detection, bounding box based detection, semantic segmentation, and instance segmentation. (d) number of papers reviewed in this manuscript per neural network type: Convolutinoal Neural Network (CNN), Fully Convolutional Neural Network (FCN), Stack Autoencoder (SAE), Recurrent Neural Network (RNN), and Generative Adversarial Networks.

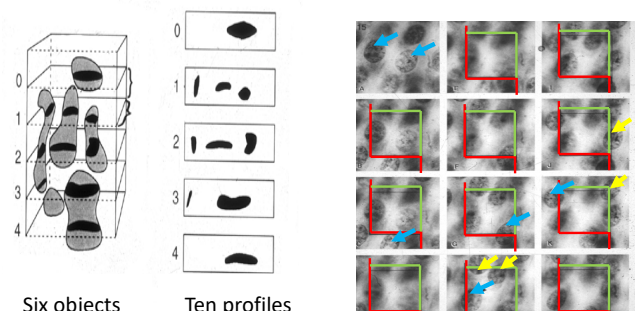
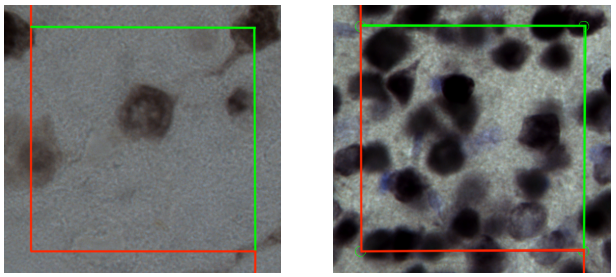


Fig. 3: Illustration of the bias associated with profile counting on tissue sections, where a) shows the true count of cells (total 6 cells) vs. the profile count (total 10). b) Z-axis scanned images (stacks), where true cell counting in tissue stack (volume) is shown in yellow arrows using thin focal-plane scanning in z-axix using the unbiased stereology optical disector probe with exclusion lines (red left and bottom lines), and inclusion (green upper and lower lines). Blue arrows indicate cells that are not counted based on unbiased stereology counting rules. [15]

a softmax layer to detect the presence of nuclei in patches based on features (representations) learned by the autoencoder.

A hybrid deep autoencoder with curvature gaussian was proposed by Song et al. [24] for detecting cells on bone marrow histology images. The curvature gaussian model was used to generate probability maps that were used in addition to the actual bone marrow histology images to train an autoencoder. The result of the autoencoder is also a probability map that is post-processed to obtain the local maxima, which represent the centroids of the detected cell. The results showed superior performance with an F1 score of 0.9483 though with a lower precision compared to the precision performance achieved by SC-CNN [25]. An extended version of the hybrid autoencoder [24] for bone marrow images nuclei detection and segmentation was proposed in [26] to simultaneously detect and classify the bone marrow nucleus. Another autoencoder was added to obtain the class-map. Then the class-map and probability map were combined in the decoding path to obtain classification and detection of a nucleus. The input to the autoencoder is a hematoxylin (H) channel patches of size 29*29 obtained using the deconvolution method [27].

The two major advantage of these approaches is they do



(a) NeuN single stain (b) NeuN dual stain

(c) Iba1 single stain (d) Iba1 dual stain

Fig. 4: Examples of the disector frame (2D view) on four datasets of microscopy images of stained sections of a mice brain. The green lines of the disector frame are the inclusion lines, whereas the red lines are the exclusion lines.

not require labeling of data and they can effectively learn feature representations effectively. The primary disadvantage is the use of sliding windows to extract overlapping patches for training stacked autoencoders.

b) ***CNN based cell detection methods****:* Convolutional neural networks (CNN) have attracted increasing interest in image recognition, object detection, and segmentation for many reasons, such as shift-invariance, shared weights — which reduces the learning time cost with many parameters — and spatial feature learning. CNN is often referred to as the neural network that has stacks of convolutional layers followed by fully connected layers at the output layer of the neural network.

A spatially constrained convolutional neural network (SC-CNN) was proposed by Sirinukunwattana et al. [25] to detect and classify nuclei on colorectal adenocarcinoma images. The proposed method uses a deconvolution of H&E stain to separate stains [27]. In this approach, hematoxylin stain intensity of the nucleus is used as input to SC-CNN, where the ground truth and the output y are represented as probabilities of a pixel being close to the center of the nucleus in a given image patch as shown in Equation 1. SC-CNN has two convolutional layers, two max-pooling layers, followed by two fully connected layers and a parameter estimation layer to estimate the center of nuclei and the height of the probability map and spatial constraint layer that gives higher probabilities to pixels close to the nuclei center. SC-CNN is designed to detect the centroid of a nuclei and provide a confidence score of the detection. The best result using SC-CNN shows a higher F1-score (0.802) compared to other methods for detection of nuclei [28][29][30][31]. However, this approach uses sliding window-based patches extraction of size 27*27, which is time expensive in large microscopy images, and a single nuclei is the maximum prediction per patch.

$$y = \begin{cases} \frac{1}{1+(\|z_j - z_m^c\|_2^2)/2} & if \ \|z_j - z_m^c\|_2 \ \leq d \\ 0 & otherwise \end{cases} \quad (1)$$

where z_j and z_m^c represent y_j coordinates and the center of the nucleus m, where $d = 4$ is a constant radius.

In high-resolution microscopy images, a sliding window is used to extract overlapping patches and feed them to a neural network; however, this approach is time-consuming and computationally expensive due to the redundancy between adjacent patches. The K-sparse kernel approach reduces computation and time costs by inserting zero entries in the kernel to make them K spatially apart. This approach was applied to subtype cancer tumor cell detection using microscopy images by Wang et al. [32]. This method combines the detection and classification of lung cancer cell histology images from TCGA (The Cancer Genome Atlas) [32]. The neural network is based on LeNet [33] and employed the K-sparse kernel method that separates neighboring items of the kernel with zeros [34]. The lung cancer tiles are of size 512*512. The input to LeNet for training is patches of size 40*40 extracted from the original tiles, whereas testing input image size is 551*551 after padding the original tile images. This method alleviates the burden of the sliding window over high-resolution histology images during testing. This technique achieved superior results compared to previously proposed approaches [35][36]. In [37] another fast deep learning-based detection method which uses K-sparse kernel was proposed (i.e., inserting zero entries in the kernel to make them K spatially apart). Additionally, a pre-fetching technique was proposed to alleviate the bottleneck of reading data from disk. This method was able to detect cells on tile images of the whole slide images very quickly (approximately 1000 cells per second detection rate) using NLST dataset[1]. However, the performance did not improve compared to previously proposed work [35].

Nuclear shape and morphology variations in microscopy images present a challenge to accurate detection and segmentation. A shape prior convolutional neural network (SP-CNN) was proposed by Tofighi et al. [38]. The network aims to use expert based shapes in detection [39] [40]. The network has two parts: learnable layers composed of convolutional layers with ReLU activation, whereas the fixed processing part aims to apply the prior knowledge in calculating the regularization of the SP-CNN. The prior shape was applied as follows: 1) Canny edge detection was applied on the input image, 2) the activation map that resulted from the learnable part of SP-CNN was multiplied element-wise with the Canny edges. The resulting mask is convolved with several prior shape masks S_i, where $i = \{1, 2, ..., n\}$. The results of SP-CNN is the detection

[1]https://biometry.nci.nih.gov/cdas/studies/nlst/

of cells that comply with the shape prior provided during training; this architecture detects the location of the cell, but not the enclosed detection of the whole cell. The authors also proposed a tunable SP-CNN approach in [41], where the prior shape becomes trainable to adapt to the nucleus shape for different images and to avoid the redundancy of shapes created by experts. Additionally, an approach to overcome the variation in shape and crowds of nuclei was proposed in [42]. This probabilistic method uses a neural network and Mixture Density Networks (MDN) [43], where nuclei detection was defined as mapping an input patch to a probability density function (PDF). The Gaussian Mixture Model (GMM) was used for PDF modeling, where a neural network based on an 18 layer Resnet [44] was employed to learn GMM parameters. The experimental results on a colorectal cancer dataset (CRC) [25] of H&E colorectal adenocarcinomas images of size 500*500 obtained at 20x magnification were superior when compared to [25][20][45].

The advantage of these approaches is that they use prior shape and alleviate the problem of a sliding window by using a K-sparse kernel, though they require large labeled sets of training datasets.

Fully Convolutional Neural Networks (FCN) are neural networks that do not contain any fully connected layers, but instead use convolutional layers even at the classification layer at the output of the neural network such as U-Net [9]. FCN often contains deconvolution layers to up-sample the learned latent features. Although, FCN is often used for segmentation, the approaches reviewed in this subsection uses FCN in one of the following ways: 1) replacing dense layers (fully connected layers) with convolutional layer since convolutional layers have less parameters compared to dense layers [46]. This approach mainly classifies pixels in overlapping patches as part of a cell or not. Then, a post-processing step to estimate a centroid was done using moments based on a rough cell shape predicted by the model. 2) transforms weak annotation (clicks near the cell center) into density maps, then train FCN on microscopy images with associated ground truth (density maps). The trained model is used to predict density maps on the test set, which is post-processed to extract the peaks for each cell [45][47].

One challenging aspect of microscopy images is the high-resolution of the images which can be alleviated by using a smaller sliding window over the entire image during testing, and extracting the patches for training a deep learning model; however, this approach is time-consuming. Huang et al. proposed a sparse kernel technique for a deep convolutional neural network [46] to accelerate the detection of cells in high-resolution histopathology images. The proposed convolutional neural network has two convolutional layers, each followed by a max-pooling layer and 1*1 convolutional layer. The training set contains patches of tiled whole slide images, and the test set is also extracted patches from whole slide images. The K-sparse kernel technique [34] was used during testing for convolutional and max-pooling layers by inserting zero value entries between every two adjacent values in a column-wise and row-wise manner to make them K-pixels apart. Additionally, the fully connected layers were replaced by a convolutional layer with a filter size of 1*1. The training of this method was done on annotated patches (if a patch contains a cell, then classify each pixel as positive otherwise classify all pixels of the patch as negative). This method accelerates the testing phase by testing on bigger patches, to reduce the time cost and redundancy of the sliding window approach.

A fully convolutional neural network based on deep residual networks has been used for improving cell detection. For instance, the proposed work in [48] integrates deep residual networks and Hough voting for mitotic cell detection. They have used two branches at the output layer of the network with different dropout rates, followed by Hough voting to predict the radius and the angle of a cell location on histology images. Their work was based on invasive breast carcinoma histology images from the AMID13 challenge dataset [49]. Hough voting and the residual network gave a higher recall of 0.686; while having a lower precision and F1-score compared to other methods [49][50]. A deep voting approach was proposed by Xie et al. [51] for cell localization using voting of offsets from local patches of the image, where each voting provides a confidence score. The weighted votes from all the testing patches were collected to compute the final voting density maps similar to the Parzen-window estimation of size 5×5 and σ was 1. The nucleus position was identified by the local maxima of the density maps. This method requires only a weak annotation (click near the nucleus center). This method had an F1-score of 0.8152 using 44 Ki67-stained Neuroendocrine Tumor (NET) microscopy images. In [45], a deep residual neural network was proposed for cell detection using a regression approach. This method is inspired by [44] [9] to output the probability density prediction of the same size as the input. The residual blocks contain an exponential rectified unit (ELU) convolutional layer, dropout, and scaling layer. The experimental results were applied to four datasets: neuroendocrine tumor (Net) dataset [52], Hela cervical cancer [36], breast cancer dataset, and a bone marrow dataset [53]. The proposed method showed a higher F-1 score compared to [28][45]. The results are shown in Table II.

A CNN based architecture that uses images and corresponding ground truth (density map) as input was proposed by Wang et al. [47]. The ground truth (density map) was generated from weak annotations using Equation 2, where D_c is the Euclidean distance, and α, d_m constrained the shape of the density maps.

$$d(x) = \begin{cases} e^{\alpha(1-)\frac{D_c(x))}{d_m}} - 1 & if D_c(x) < d_m \\ 0 & otherwise \end{cases} \quad (2)$$

The ground truth is represented by density maps generated from dot annotations (i.e., weak annotation), where a peak on a density map corresponds to a cell dot annotation. This approach uses a network inspired by semantic segmentation using a Fully Convolutional Network (FCN) [54] with two modifications: 1) skipping paths were added in a similar manner to that presented in the U-Net architecture [9] and 2)

max-pooling layers were eliminated and replaced with convolutional layers with a stride of 2*2. A similar regression-based approach that uses a density map as ground truth was proposed by Zhu et al. [55] to detect and count cells on histology images using FCN. Their deep learning architecture has three blocks of convolutional layers, max-pooling pairs; followed by up-sampling layers. Experiments with four different datasets were presented: breast cancer stained images dataset, Insect cells image dataset, Vesicales image dataset, and deep convolutional adversarial generated network DCGAN images.

Jacobs et al. investigated transfer learning and tuning of deep learning to detect and classify nuclei on H&E stained histology images [56]. The initial model was trained on a colon histology dataset [25]; then tuning the trained model was done on a prostate histology dataset. The proposed regression-based neural network detects the presence of a cell where the ground truth is a density map with peaks corresponding to cell centers. The results endorse the idea of transfer learning for nuclei detection.

*c) **Evaluation study**:* Although deep learning has had success in microscopy image analysis for detection, different parameter settings are essential to get the best performance. Hofener et al. [57] presented an evaluation study of essential parameter settings of a neural network on multiple datasets for cell detection. The parameter settings evaluated include augmentation, dropout, post-processing, and FCN decoding (i.e., up-sampling or dilation). Additionally, a post-processing method of probability map (Pmap) was presented where a 3*3 median filter was applied followed by Gaussian smoothing ($\alpha = 2$). This post-processing method was showing higher results compared to other post-processing methods.

*d) **Adversarial Neural Network based methods**:* Residual attention generative adversarial network was proposed by Li et al [58] to detect cells on microscopy images. The generator is a combination of U-Net [9] and Residual Attention Network [59], where the U-Net segments the cells on the microscopy images. The attention mechanism goal is to orient the neural network focus to the important region of the image (i.e., nuclei). The discriminator is compromised of eight convolutional layers. Experimental results on a colorectal adenocarcinomas data set of H&E microscopy images [25], showed superior results compared to [25][29], where the F1-score was 0.847.

2) Bounding Box based nuclei detection: Bounding box cell detection is an approach where the deep learning model localizes an object in an image by providing an object bounding box and a confidence score. Deep learning is used to create a regression model for the bounding box offsets $x, y, width, height$, which represents a bounding box upper-left corner x-axis, y-axis location, and a bounding box width and height respectively.

Cell proposal is the procedure of generating proposals that are used later for detection or tracking by choosing the most optimal set of cell proposals based on confidence score and other constraints. In the work of Akram et al. [2] a CNN based method was presented for proposing cells in the Fluo-N2DL-HeLa dataset [60]. Their deep learning network was based on Zeiler and Fergus [61] with some experimental modifications. Additionally, their networks output the confidence score of the bounding box and the bounding box starting point and dimensions. This CNN architecture achieved a high precision of 0.95 and recall of 0.90.

In [62], Fast R-CNN was adopted to detect cells on colon cancer H&E stained histology images (colorectal adenocarcinoma dataset) [25]. A VGG-16 architecture pre-trained on the ImageNet dataset [63] [64] was used for initialization. The region proposal network (RPN) was built on top of VGG-16 to generate bounding box proposals and confidence scores for Fast R-CNN. The results of Fast R-CNN showed an improvement in cell detection compared to SC-CNN [25]. Another approach to detect and segment cells in microscopic images was proposed by [65]; the approach combines a Single Shot Detection neural network (SSD) [66] and ResNet101 as a backbone network [44]. The model first localizes the cells (i.e., cell detection) then cropped versions of activation maps from shallow and deep layers were bi-linearly up-sampled to get the mask of an instance in the input image. This approach is referred to as instance segmentation, as discussed in Section III-B2.

A deep learning approach to detect mitotic cells was proposed by [67]. This multi-stage neural network approach which consists of deep segmentation network, deep verification network, and deep detection network. The deep segmentation network is for generating deep mitosis segmentation when only a weak label is given (only the cell center pixels are annotated). Training is done on the 2012 MITOSIS data set challenge which provides pixel-level ground truth, then testing was done on the 2014 MITOSIS data set challenge which only provide center pixel annotation (weak annotation). The 2012 MITOSIS data set has 50 histopathology images of 40X magnification [68]. The 2014 MITOSIS data set has 1696 histopathology images scanned at 40X magnification [69]. The deep detection network for mitotic cells is based on the Faster R-CNN approach which uses Region Proposal Network (RPN) followed by region-based-classification [70]. The deep verification network was a ResNet50 based classification neural network that detected patches of fixed size from the deep detection network and classified each patch as mitotic cells or not mitotic cells with a probability score. This verification network's goal is to reduce false-positive detection. The results of this approach show state-of-the-art performance on the 2012 MITOSIS data set, where the precision was 0.854, recall was 0.812, and F1-score was 0.832.

3) Shape-fitting based cell detection: Due to the overlap of nuclei in microscopy images, individual nuclei detection is a challenging task. A Bayesian object recognition method was proposed in [71], where an ellipse is fitted to each segmented mask by a CNN. Cytological H&E images of breast cancer

were used for the experiment, where color deconvolution [72] was applied to extract hematoxylin density. The hematoxylin patch images of size 43*43 were the input to a CNN with four convolutional layers with a max-pooling layer after each consecutive convolutional layer. At the output layer of the neural network two fully connected layers were applied to predict the class of each pixel of patch images where classes are nuclei, cytoplasm, nuclei edge, and background. The resultant semantic segmentation was turned into nuclei masks by assigning other classes as background. After that, an ellipse fitting approach was conducted to fit each binary mask with an ellipse. This approach yielded better results, even with clumped cytological nuclei.

B. Segmentation

Segmentation of nuclei in microscopy images plays a significant role in medical diagnosis and computer-aided-diagnoses (CAD) [79]. Additionally, segmentation provides a basis for other quantification of cells in microscopy images such as counting, and shape analysis. Nuclei segmentation plays an important role in cancer grading by utilizing the nucleus shape [80]. It is also used as an indicator of prognosis [81]. Microscopy image cell segmentation is a challenging task due to the variability of cell shapes, orientations, staining, scanners, and acquisition. Therefore, the ordinary image segmentation algorithm's performance is limited. Deep learning showed success in medical image segmentation, including microscopy images cell segmentation, which enables the model to learn discriminative features that allow the model to perform better than ordinary algorithms. A summary of segmentation based methods reviewed in this subsection is shown in Tables V, VI, and VII.

1) Semantic segmentation: Semantic segmentation refers to classifying each pixel p of an image x to a particular class c where $c \in \{c_1, c_2, ..., c_n\}$ for n classes [82][83]. In the following subsections, we discuss neural networks contributions to address microscopy image cell segmentation challenges.

a) ***Superpixel and resolution reduction based methods:*** High resolution microscopy images require a large memory and significant computation resources. Therefore, preprocessing steps such as patch extraction are mostly applied to reduce input image size and thus reduce the memory and computation time, one such technique is the use of superpixels [84]. A Deep learning model with superpixels was proposed in [85] to learn localized features efficiently and to reduce input image sizes. The approach involves extracting superpixel information using the Simple Linear Iterative Clustering algorithm (SLIC) [84]. Images were converted to the HSV color domain where the V color component was used as an input to deep learning. A modified version of LeNet [33] was used to train and test disjoint subsets of cervical cancer images. Tareef el at. [86] proposed a similar method to segment cells based on extracted superpixel patches segmentation by CNN followed by dynamic shape modeling. The superpixel patches were created using SLIC [84], then classification was done for each patch, where a patch could belong to either nuceli, background, or cytoplasm. After that, Voronoi segmentation and dynamic shape level set were performed to segment the nucleus and cytoplasm using the Overlapping Cervical Cytology Image Segmentation challenge dataset from ISBI2014. However, these two approaches used small patches of size 50*50 and 16*16 respectively from superpixel images with a single cell per patch which is time consuming. Janowczyk et al. proposed a resolution adaptive deep hierarchical (RADHicaL) learning approach for nuclei segmentation [87]. The authors trained M AlexNets [88] each with a different resolution. The resolution of the input images was reduced with factors $f \in \{1, 0.5, 0.25, 0.1\}$, where factor $f = 1$ represents the original image. The final output is the maximum of probabilities starting with the lowest resolution, and only uses a higher resolution if the objects within the patch need further accurate segmentation. This approach was compared to the same neural network that uses the original image (40x magnification), where the RADHicaL approach showed slightly lower F-score compared to its counterpart. However, the computation time was reduced by about 85%. Therefore, there is a trade-off between efficiency and speed when training deep neural networks for whole-slide pathological images.

b) ***Methods for cell, stain, and shape variance:*** The variability of images in intensity, size, shape, and orientation of cells in fluorescence microscopy images pose a severe challenge to learning algorithms including deep learning. In [89], a method was proposed to deal with images and cells variability by adding extra connections to bypass pooling layers. This method is known as MIMO-Net (Multi-input Multi-Output neural network) which uses an encoder-decoder approach similar to U-Net [9]. However, they generate auxiliary branches to make multiple outputs that are combined for the final output. The inputs to the neural network are two-channel images corresponding to two main components of the images Ecad (membranes marker) and DAPI (nuclear marker). The results were superior to previously proposed methods such as [54][90][9]. Another approach to deal with the variation of cell shapes and appearance was proposed in [91]. This deep learning method uses a sliding window approach to extract patches for training a neural network, where shape selection from a dictionary is performed during training, and outputs probability maps. The neural network consists of two convolutional layers each followed by a max-pooling layer. At the output layer of the neural network, two fully connected layers were used where the last layer is used to make predictions by applying the softmax activation. Training images are YUV color space image patches with a fixed size. For testing, a sliding window approach was utilized to extract patches of specific sizes. For accurate segmentation of overlapping cells, the authors proposed an iterative region merging algorithm that works as follows: 1) calculating distance map [92] using the probability maps, 2) H-minima transform [93] was applied to the inverse of the distance map to get the initial markers, 3) an iterative process to expand the markers based on the distance map until the iteration before two markers

TABLE I: Summary of the datasets, results, and cell detection methods, where precision, recall, F1-score, and accuracy are denoted by P,R,F1, and Acc respectively.

References	Datasets and source code	Task	Results	Summary
[47]	**Datasets:** 1) Bone Marrow dataset (BM) [53]: contains 11 stained images of size 1200*1200 from 8 healthy subjects 2) Histopathology images of breast cancer [73], which consists of twenty stained images of size 100*100 **Source code:** not available	Detection	1) Bone Marrow dataset (BM): P = 0.9138 R = 0.8985 F1 = 0.90602 Breast cancer dataset: F1= 0.8810	Regression based detection of cell centroids. The annotation of the datasets are dot based annotation of cell centroids
[74]	**Dataset:** Ten micrographs of gastric cancer tissue scanned at 40x prisma dataset **Source code:** not available	Detection	Training set: Acc 96.88% Test set results not provided in the author manuscript	Classification based detection and counting of Tumor-infiltrating lymphocytes (TILs) using a simple CNN. Uses sliding window approach during testing
[55]	**Dataset:** 1) Camlyon 17: Breast cancer pathology H&E stained images 2) Insect cell dataset 3) Vesicles image dataset 4) DCGAN generated images **Source code:** not available	Detection	1) Breast Cancer dataset: R = 0.900 P = 0.909 F1 = 0.904 2) Insect cell dataset: R = 0.902 P = 0.944 F1 = 0.923 3) Vesicles image dataset: R = 0.917 P = 0.902 F1 = 0.909 4) DCGAN generated images: R = 0.826 P = 0.910 F1 = 0.866	FCN applied to the datasets, followed by post-processing The detection uses the maxima of intensity map to detect a certain cell
[57]	**Dataset:** 1) H&E colorectal adenocarcinoma: Total 100 images of resolution 500*500 associated with dot annotation [25] 2) Ki-67 stained breast tumor tissue section images of resolution 450*450 [75] **Source code:** not available	Detection	F1 = 0.827	An evaluation study of different parameter tuning and settings of FCN, such as augmentation, dropout, post-processing, upsampling method. Detection goal is the centroid of the cells
[46]	**Dataset:** Histopathology images from National Lung Screening Trial (NLST) dataset. Total tiles are 215 of resolution 512*512 [76] **Source code:** not available	Detection	F1 = 0.786	K-sparse kernel based acceleration of testing phase of detection on WSI. The method uses patches centroid on nuclei to train CNN. The detected cell centroid is calculated based on raw image moment postprocessing
[24]	**Dataset:** Bone Marrow Dataset **Source code:** not available	Detection	P = 0.9273 R = 0.9702 F1 = 0.9483	Hybrid Autoencoder and Curvature Gaussian model proposed to detect cells intensity map. A post-processing to extract local minima of cell was used
[56]	**Dataset:** 1) prostate H&E stained images. Magnification = 20x, Total : 400 images, resolution 250*250 pixels with weak annotation for training 2) Colon H&E stained images dataset. Magnification =20x **Source code:** not available	Detection	P = 0.846 R = 0.882 F1 = 0.864	Transfer learning and fine-tuning of a pre-trained regression based CNN on prostate H&E stained images into another dataset of Colon H&E dataset. Detection goal is the centroid of the nuclei
[42]	**Dataset:** Colorectal cancer histology dataset of H&E images. Magnification = 20x. Total 100 images. Resolution 500*500 pixels [25] **Source code:** not available	Detection	P = 0.788 R = 0.882 F1 = 0.832	ResNet [44] deep learning network with Mixture density network (MDN) to learn from different cell Gaussian distribution
[71]	**Dataset:** Breast Cancer microscopy images. 25 malignant and 25 benign Total images: train = 20 images, test = 20 images. Resolution = 500*500 pixels **Source code:** not available	Detection	Acc = 91.33%	Deconvolution method was applied to H&E stained images to extract the hematoxylin stain (represents nuclei) [72]. Then Extracted patches of size 43*43 were used to train a CNN. To create a rough localization of nuclei. Then, Ellipse fitting was applied to detect and separate touching nucleus structures
[23]	**Dataset:** Cervical Cancer dataset (ISBI2014): Total images: 16 pap smear of resolution 1024*1024 pixels **Source code:** not available	Detection	Acc = 87.6%	Sparse Stack autoencoder was used to train and learn latent features, then a sliding window was used to test on a testing set followed by softmax layer
[48]	**Dataset:** Invasive breast carcinoma histology images from AMID13 challenge dataset [49] **Source code:** not available	Detection	P = 0.547 R = 0.686 F1 = 0.609	Deep residual network with two top branches followed by deep voting based on hough transform and bilinear interpolation
[2]	**Dataset:** Fluo-N2DL-HeLa dataset [60], contains 92 frames from 2 time lapse sequences of fluorescent HeLa cells images **Source code:** available here	Detection	P = 0.963 R = 0.996	Cell proposal network to detect and localize, cells by providing a bounding box of each cell
[25]	**Dataset:** Total of 100 H&E stained histology images of colorectal adenocarcinomas. Resolution 500*500 pixels extracted from 10 WSI of 9 patients. Magnification = 20x **Source code:** available here	Detection	P = 0.758 R = 0.827 F1 = 0.791	Deconvolution of H&E stain to separate stains [27] was applied, then regression based detection of center of nuclei using CNN
[26]	**Dataset:** Bone marrow hematopoietic stem cells dataset **Source code:** not available	Detection	P = 0.9129 R = 0.9641 F1 = 0.9378	Stained separated H&E images based on [27] used in two Autoencoders rufor class map and erythroid and myeloid cells detection

TABLE II: Summary of the datasets, results, and cell detection methods, where precision, recall, F1-score, and average precision are denoted by P,R,F1, and AP respectively. (continuation of Table I)

References	Datasets and source code	Task	Results	Summary
[45]	**Dataset:** 1) Neuroendocrine tumor (Net) dataset [52]: Total 59 cropped Ki-67 bright field stained images, resolution 400*400*3, magnification 20x. 2) Hela cervical cancer [36]: Total 22 phase contrast microscopy images. 3) Breast cancer dataset from National Cancer Institute, 2013.): total 70 images. 4) Bone marrow dataset [53]: total 11 H&E stained bright-field microscopy images. Resolution = 1200*1200 pixel from 8 patients **Source code:** not available	Detection	Net dataset: P = 0.94 R = 0.92 F1 = 0.93 Hela dataset: P = 0.98 R = 0.98 F1 = 0.98 Breast Cancer dataset: P = 0.89 R = 0.91 F1 = 0.90 Bone marrow dataset: P = 0.86 R = 0.94 F1 = 0.90	Unet based ResNet for cell detection. The detected part of the cells are centroids
[37]	**Dataset:** Here is the dataset of extracted tiles from WSI. Resolution: 512*512 pixels **Source code:** available here	Detection	P = 0.83 R = 0.84 F1 = 0.83	CNN with K-sparse kernel to speed the detection
[62]	**Dataset:** Colorectal adenocarcinomas dataset of 100 H&E stained microscopy images of size 500*500 [25]. **Source code:** not available	Detection	F1 = 0.714	VGG16 faster R-CNN to detect cells location
[77]	**Dataset:** H&E stained microscopy images containing seven different tissue samples: Breast, Bladder, Colon, Liver, Kidney, Prostate, and Stomach **Source code:** available here	Detection & Segmentation	P = 0.799 R = 0.955 F1 = 0.860	A pipeline to segment nuclei using Unet and post-processing using erosion. The detection is based on centroids of the segmented nuclei
[32]	**Dataset:** Lung cancer cells histology images from TCGA (The Cancer Genome Atlas). Total : 300 512 *512 lung cancer histology images. 270 for training and 30 for testing **Source code:** not available	Detection	P = 0.8029 R = 0.8683 F1 = 0.8215	Neural network architecture based on LeNet [33] and uses K-sparse kernel to reduce sliding windows cost during testing
[65]	**Dataset:** CNS stem cells population dataset [78] **Source code:** not available	Detection & Segmentation	AP = 0.8739	Detect and Segment cells using a combined Single Shot Detection neural network [66] and ResNet101 as a backbone network approach [44]
[38]*	**Dataset:** Colorectal adenocarcinomas dataset of 100 H&E stained microscopy images of size 500*500 [25]. **Source code:** available here	Detection	P = 0.803 R = 0.843 F1 = 0.823	Detection of the location of the cells by including a prior shapes created by experts for neural network training.
[67]	**Dataset:** 1) 2012 ICPR MITOSIS dataset of 50 hisopathology images scanned at 40X magnification, and associate pixel level is provided (masks) [68]. 2) 2014 ICPR MTOSIS dataset of 1696 micrscopy images scanned at 40X magnification. The annotation is a single click on cell centroid [69]. **Source code:** available here	Detection	P = 0.854 R = 0.812 F1 = 0.832	A multi-stage neural network approach which consists of deep segmentation network, deep verification network, and deep detection network. The deep segmentation network is for generating segmentation of mitosis region when only a weak label is given (only the cell center pixels are annotated), where training is done on 2012 MITOSIS data set challenge which provide pixel-level ground truth, then testing was done on 2014 MITOSIS data set challenge which only provide center pixel annotation (weak annotation).
[58]	**Dataset:** Colorectal adenocarcinomas dataset of 100 H&E stained microscopy images of size 500*500 [25]. **Source code:** not available	Detection	P = 0.859 R = 0.836 F1 = 0.847	The generator is a combination of U-Net [9] and Residual Attention Network [59], where the U-Net segment the cells on the microscopy images.
[29]	**Dataset:** Total of 537 H&E stained histopathological images digitized at 40x magnification, where size of each image is 2200x2200. Images was randomly split for training, validation, and testing. **Source code:** not available	Detection	P = 0.888 R = 0.828 F1 = 0.844	Stack Sparse autoencoder trained to learn distinguishable features of nuclei, where the images are patches extracted using sliding window. Then a classifier was applied to identify each patch as nuclear or non-nuclear.

merge, and 4) a morphology smoothing operation was applied to the segmentation mask. The dice coefficients on a brain tumor, neuroendocrine tumor (Net) dataset, and a breast cancer datasets are 0.85, 0.92, and 0.80 respectively.

Variation of stain and stain absorption presents a challenge for accurate segmentation. Therefore, preprocessing steps such as stain normalization are often applied to microscopy images. A stain-aware multi-scale network (Sams-Net) was proposed to segment cells on H&E stained images by accounting for the variability of stains inside the cell, and the clumping of cells challenges [94]. A stain normalization method [95] was applied to neural network input images to avoid data preparation artifacts and account for stain variations. Sams-Net uses a predefined weights map that penalizes the loss function (pixel-wise cross-entropy) to achieve better cell segmentation. The predefined weights were calculated based on 1) giving more chance for the nuclei to be segmented correctly when class imbalance occurred, and 2) providing more opportunities for the boundary to separate touching cells, and to a nuclei which absorbed less stain. SamsNet is based on FCN [54] and uses residual blocks, where three residual blocks are on the downsampling and two are on the upsampling. The input images are in three sizes to prevent the loss of spatial information during downsampling: first residual block takes

H&E image of size 100*100*3; second residual block takes H&E image of size 50*50*3, and the third residual block takes H&E image of size 25*25*3. The segmentation result using Sams-Net showed slightly higher results compared to U-Net [9] using the CPM dataset [96].

*c) **Methods to address touching and overlapping cells problem**:* Touching and overlapping cells in dense regions in microscopy images are challenging for segmentation algorithms. To address this issue, Kumar et al. proposed a technique for nuclear segmentation in histopathology images of various human tissues and organs [97]. This technique predicts ternary probability maps instead of binary probability maps. The ternary probability maps correspond to the background, nucleus, and boundary of a nucleus. This method ensures that overlapping and touching cells are segmented correctly instead of segmented jointly or relying on postprocessing scheme such as watershed to separate touching nuclei. The convolutional neural network proposed is simple yet powerful. It consists of three convolutional and pooling layer pairs, followed by two fully connected layers and the output layer. However, this approach uses small patches for training and testing which is time expensive. Contours of cells are essential to overcome the problem of touching cells; therefore, a deep contour-aware network (DCAN) approach was proposed by Chen et al. for the detection and segmentation of cells and their contours [98]. DCAN is an end-to-end fully convolutional network that has one downsampling path for learning deterministic features followed by two branches of upsampling for nuclei and contour segmentation individually. This network is trained as a single neural network where weights on the downsampling path were initialized with a DeepLab model [99] trained on PASCAL VOC 2012 dataset [100]. In contrast, upsampling paths weights were initialized randomly from a Gaussian distribution. The results of this network show superior results on the 2015 MICCAI gland segmentation challenge dataset. Another Boundary-Enhanced Segmentation network (BESNet) was proposed by Oda et al. to improve the segmentation of nuclei in ganglion H&E stained histopathology images [101]. BESNet is similar to U-Net [9]; however, it has two decoding paths, one for the nuclei segmentation, and the other for boundary segmentation. Additionally, a boundary-enhanced loss function was proposed to penalize when the output differs from the ground truth. The results of BESNet showed superior performance over the U-Net approach [9] in dice coefficient. A method for nuclei segmentation on H&E stained histopathology images using regression was proposed by Naylor et al [102]. This regression based method generated a distance map to overcome the issue of touching and overlapping nuclei in a clumped area of cells in histopathology images. This method provides an alternative approach to predicting the object (nuclei) and its contour [97][103]. The regression based deep neural networks used to create the distance map were a pre-trained FCN on ImageNet[54][64], U-Net [9], and pre-trained Mask R-CNN with ResNet 101 backbone [44] on COCO dataset [104], where pre-trained networks were fine-tuned using two datasets [97][105]. The regression-based distance map was post-processed using morphology dynamics [106] to obtain the right nuclei.

Preprocessing of microscopy images and postprocessing of the deep learning models results (i.e., probability maps) are often applied to get an accurate segmentation and to overcome the overlapping and touching cells. Pan el at. [107] proposed a segmentation method of nuclei in H&E stained breast cancer histopathology images dataset [2] using sparse reconstruction and deep learning. This approach consists of three steps: 1) applying sparse reconstruction to remove the background and emphasize the nuclei in the stained images which work as follows: i) converting RGB images to grayscale, then ii) applying a smoothing filter called anisotropic diffusion filter (ADF), after that, iii) applying the K-SVD algorithm on gray level images, followed by iv) Batch-OMP for orthogonal matching pursuit to get a denoised image as input for a CNN 2) CNN training and testing, 3) post-processing operations using morphology operations such as opening operations. Another approach is a FCN that has encoder-decoder paths similar to U-Net [9] as proposed in [108]. The method segments a nuclei and boundary simultaneously. The inputs to the model are random patches of H&E stain normalized images using sparse non-negative matrix factorization (SNMF) [109]. The authors tried deconvolution of H&E, where the hematoxylin channel is used as input to deep neural network and concluded that stain normalization is better than stain deconvolution to train deep neural networks. In the testing phase, overlapped patches are used to test and train the model, and the results of testing on patches are assembled. The architecture outputs probabilities of background, boundary, and nuclei. A postprocessing step was applied to the prediction to get optimal results.

Due to the overlapping of nuclei in histopathology images, a deep learning method that uses two U-Net architectures to segment the cells and create a regression-based distance map was proposed by Mahbod et al. [110] to enhance the separation of the cells in segmented masks. The method uses stain normalized H&E images using [111] to train a U-Net for segmentation. Another U-Net was trained to create the distance map. The distance map was smoothed using a Gaussian, then maxima per nuclei were extracted to seed the watershed algorithm. Then a postprocessing step was applied to fill holes and remove small objects. A method for segmenting cells in H&E stained tissue was proposed in [77]. This method uses overlapping patches to train and test a U-Net deep learning architecture [9]. To maximize the separation of touching cells, a loss function that combines edges and cell body loss was used during training of the model. The segmented cells were smoothed and eroded to separate touching cells; then, a geometric centroid was calculated to count the number of cells in an image.

In [112], a deep learning architecture similar to U-Net [9] was proposed to segment cells and cytoplasm on fluorescence

[2] http://medicine.yale.edu

images. The input images are preprocessed using a top-hat filtering kernel to suppress background illumination followed by down-sampling the image to be 10x magnification. After that, a patch of size 176*176 with 16 pixels overlap was extracted to train a deep learning model. The results from the deep learning model were a probability map that was used for postprocessing. The postprocessing of the probability maps involves applying a Laplacian of Gaussian (LoG) blob detector [113] multiple times to improve blobs on a probability map. Then, Otsu thresholding was applied to extract a nuclei binary mask [114]. To address the touching cells issue, an inverse distance transform was applied, then an H-minima transform was applied at a certain prefixed level h [115]. The value of h was used as a seed for the watershed algorithm to separate touching cells. The results of inference on test set patches of the same image are stitched together to form a single image.

d) ***Cascaded deep learning approach****:* A cascaded deep learning approach was proposed by Wang et al. to segment nuclei in histology images of brain tissues [116]. This method trained two CNNs, where the probability maps from the first neural network were fed to the second neural network in order to learn high-level contextual information from the probability map and information from the original image patches to improve segmentation quality. The results of this method show superior results over a non-cascaded CNN.

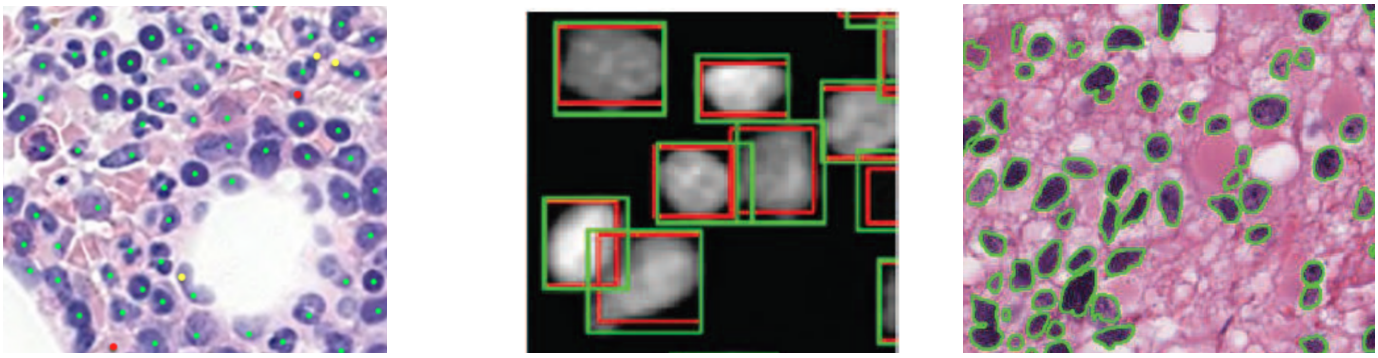
e) ***Evaluation studies****:* A study of spatial information effect using three deep learning architectures was conducted by Hatipoglu [117]. This study used a CNN, a deep belief network (DBN), and SAE to segment cellular and extracellular objects in histopathology images. For instance, patches of histopathology images were extracted using windowing methods of different sizes. The findings of this study show that a CNN has the dominant performance with high spatial information (i.e., the largest patch size 13*13) using the breast cancer dataset [118] and kidney renal clear cell carcinoma dataset [119]. In [105], an experimental study of nuclei segmentation in histopathological images was done using three different neural networks: a shallow neural network called PangNet [120], a fully convolutional neural network for semantic segmentation (FCN) [54], and DeconvNet [121]. Additionally, an ensemble of FCN and DeconvNet was created for this approach [121]. The posterior probability maps were post-processed based on morphological dynamics [106]. The result of training the aforementioned neural networks on a breast cancer dataset [105] shows relatively higher performance for the ensemble and FCN compared to the other two neural networks. In a study by Baltissen et al., a comparison between multiple supervised and unsupervised methods for Glioblastoma cell segmentation in microscopy images was performed [122]. These methods include global thresholding; local thresholding; fast marching level set method [123]; regions competition [124]; K-means clustering; random forests; and deep learning. The deep learning architecture was based on U-Net [9] with atrous spatial pyramid pooling (ASPP) [99]. The best performance was obtained using deep learning which outperformed other supervised and unsupervised methods by a great margin.

f) ***Transfer learning****:* A transfer learning approach was proposed to handle the difference of channels between the datasets the model was trained on, and the dataset a model will be tuned [125]. This method used the U-Net architecture [9] with residual connections [44] and ASPP [99]. This network was trained on a DAPI single channel Glioblastoma dataset and transferred to a stained (four channels) Glioblastoma dataset for fine-tuning. The transfer learning approach proposed was of two types: 1) copying the transferred weights of a single channel to all channels of the target dataset, and 2) weights from single channel images trained model transferred to only one channel of the transfer model, whereas, the remaining channels were initialized using MSRA initialization [126]. The second transfer learning approach showed superior results over prior approaches.

g) ***Recurrent Neural Network based segmentation****:* Her2Net: a deep learning architecture for segmenting and classifying cells and cell membranes on human epidermal growth factor receptor-2 (HER2) stained images of breast cancer was proposed in [127]. Her2Net is an encoder-decoder architecture that involves 16 convolutional layers, two max-pooling layers, two spatial pyramid pooling layers, and a TLSTM (trapezoidal long-short-term memory). TLSTM is basically a combination of four LSTM's to prevent cellular structures from distortion on a pixel basis. This neural network achieved higher results compared to SegNet [128], Bayesian SegNet [129], and U-Net [9] using the Her2 dataset [130].

h) ***Adversarial Neural Network based approach****:* An approach to segment cells using an adversarial neural network was proposed in [131]. This approach was inspired by generative adversarial networks (GAN) [132], however it does not generate images from a random noise vector, but instead it estimates the segmentation of cells in an image. The proposed method consists of estimator and discriminator networks, where the estimator learns to generate a cell segmentation, while the discriminator learns to discriminate between the generated mask and the manual ground truth mask (manual mask). The estimator has five convolutional layers, where each layer is followed by batch normalization and activation, the output is a probability map of three classes (background, foreground, and cell boundary). To enable discriminator to distinguish between the probability map and the manual ground truth, a Rib Cage blocks was designed. Rib Cage blocks has three inputs: estimator output, manual ground truth, and combination of both. The performance of this approach was demonstrated on H1299 data set [133], where the best Jaccard measure was 0.806.

An approach to train a neural network with millions of parameters with less annotated data was proposed by Majurski et al. [134]. The approach uses generative adversarial networks (GAN) to learn weight parameters [132], then the weights from the discriminator are transferred to the encoder part of U-Net architecture [9]. The U-Net architecture was further



(a) Near Cell Center Detection [24] (b) Bounding Box detection [2] (c) Segmentation [94]

Fig. 5: Examples of a) Near-cell center detection based approach [24], where the green dots represents the true position detection, yellow dots represents the false positive detection, and the red dots represents the false negatives. b) Bounding box based cell detection approach [2] where the green bounding boxes are the ground truth and the red bounding boxes are detected by deep learning. c) Segmentation of cells approach [94], where the contours of the masks are overlaid on top of the microscopy image

trained with small manually annotated microscopy images to segment contours of cells. The authors also, tried transfer learning using a model trained on the annotate Common Objects in Context (COCO) data set [104]. Furthermore, the authors also tried augmentations of the small manually annotated microscopy images data set for training U-Net. The best performing approach was transfer learning from GAN (i.e., transfer the weights from the discriminator to U-Net encoder with further training and augmentation), where the contour dice coefficient was 0.637 using absorbance microscopy images of human iRPE cells [135].

2) Instance segmentation: Instance segmentation refers to the process of identifying each object, which is achieved by localizing the objects followed by segmenting each localized object (i.e., detection followed by segmentation) [83].

*a) **Bounding Box localization based instance segmentation**:* A detection based segmentation of neural cells using deep learning was proposed in [136]. This method uses a VGG-16 based Single Shot Detector (SSD) to find the offset of instances on an image. Then, an upsampling approach similar to [9] was utilized to preserve spatial information and to get masks of a size similar to the input image size. This approach outperformed multi-task network cascades (MNC) [137], and fully convolutional instance aware semantic segmentation (FCIS) [138]. A cell proposal based-segmentation method was introduced in [2], where two neural networks were trained. The first neural network proposed K cells, each with a score S. This neural network was based on [2]. Then the second neural network used ROI-pooling to extract a fixed-sized features map of size 25*25 from the first neural network. The extracted feature maps were concatenated with other coarse feature maps from lower convolutional layers to localize cells better. The output of the second neural network (segmentation neural network) was 25*25 probability maps, which were resized using bicubic interpolation to the actual size of the bounding box provided by the first neural network. A significant advantage of using the first neural network is that it uses a Whole Slide Image (WSI) rather than using a sliding window. This method was tested on three datasets: PhC-HeLa [36], Fluo-N2DL-HeLa [60], and Hist-BM [53] and compared to other cell detection methods. The results showed the superior performance of the two-stage cell proposal and segmentation deep learning compared to MSER [36][97].

A detection and segmentation approach was proposed in [65]; the approach combines a Single Shot Detection neural network (SSD) [66] and ResNet101 as a backbone network [44]. The SSD localizes the cell on the image, then cropped versions of activation maps from shallow and deep layers were bi-linearly up-sampled and utilized to get the mask of a cropped instance of the input image.

A composite deep learning method followed by post-processing using the local fully connected conditional random field (LFCCRF) was proposed by Liu et al. for automatic segmentation of cervical nuclei [139]. This method uses Mask R-CNN deep learning [140] to provide detection-based instance segmentation. The segmentation masks were refined using LFCCRF because Mask R-CNN uses more semantic features of low resolution, which cause some boundary loss on the predicted mask. Therefore, the refinement of the nuclei boundary is needed for accurate segmentation. This method used the Herlev dataset for experimentation [141].

*b) **Centroid based instance segmentation**:* While deep learning has shown tremendous improvement in medical imaging research with superior results over its counterpart (classical image processing techniques), annotation of data is tedious, error-prone, and time-consuming, which imposes an obstacle for the further success of deep learning. A method that requires the minimum annotation (i.e., centroid marks) was proposed by Thierbach et al. to segment nuclei in histopathology images [142]. This method is based on Mixed-scale Dense (MS-D)

architecture [143] that enables learning effectively with different cell sizes. Light sheet microscopy fluorescence images were used to train a MS-D network where the annotated cells were convolved with a spherical kernel of radius = 3. The predicted centroids on the test set were used as initialization for a Multi-object Geometric Deformable Model (MGDM) to localize cell centers and segment them correctly [144]. This approach showed a higher precision (0.895), recall (0.942), and F1-score (0.918) compared to [145].

IV. Evaluation Metrics and State-of-the-arts methods

A summarized description of the evaluation metrics used in the reviewed papers is shown in Table III. In Table IV, we provide a summary of the methods that used the same training-testing split of a dataset, and we note the state-of-the-art methods. Although, others methods shared the same dataset, training-testing splits were done randomly [47] [45] [146] [2] [42] [62] [105], therefore, it was difficult to conclude which is the true state-of-the-art method for some datasets.

V. Deep Learning application to Unbiased stereology

The current gold standard method for quantifying cell number in tissue sections is unbiased stereology, where a trained user focuses through a stack of images (z-axis) of a section and performs counting based on the disector principle [155]. A study by Groot et al. [156] found that counting neurons by a trained bio-scientist is 33% lower than the actual neuron count due to human recognition limitation with respect to small changes in object quantities [157]. Therefore, an automatic unbiased stereology approach that solves the human limitation in cell countings such as counting subjectivity, recognition bias and time-consumption will offer great advantages over the current unbiased stereology practices.

Deep learning provides an objective approach to classify, segment, and detect cells in microscopy images, where training the deep learning model requires a large number of microscopy images to learn discriminative features. In unbiased stereology, stacks of microscopy images (3D) are obtained by focusing through the section (z-axis). A cell is counted when it appears in-focus and complies with the unabised stereology rules. For more about unbiased stereology, refer to Section II.

Semantic segmentation of objects using deep learning architectures such as U-Net [9] classifies each object in an image to a particular class y, where $y \in \{y_1, y_2, ..., y_n\}$. For instance in binary classification, y can be either foreground (i.e., part of the cell) or background. Alahmari et al. proposed an automatic algorithm that uses a U-Net semantic segmentation deep learning architecture to perform the segmentation of cells on microscopy images. The deep learning input images are inspired by the Adaptive

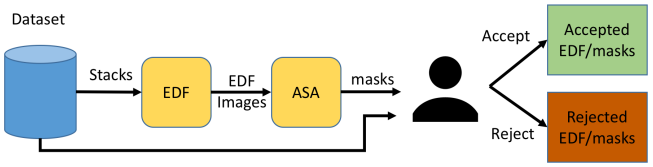


Fig. 6: Data preparation process and ground truth generation using ASA

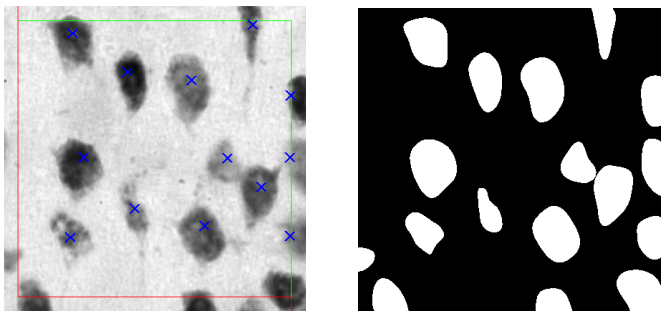
Segmentation Algorithm (ASA) [158], that makes an Extended Depth of Field image (EDF) from 3D stacks [159]. The ground truth of each EDF image was initially created by the ASA [158] followed by a human checking to verify the correctness of the ASA generated masks. Fig. 6 shows the process of data and ground truth preparation. After training and testing the deep learning model, a post-processing technique was applied to perform unbiased stereology counting rules in microscopy images of the mice brain neocortex [160]. Although this approach alleviated a considerable burden of manually labeling (i.e., creating pixel-wise masks) for cells on EDF images by using an unsupervised learning algorithm (i.e., ASA) to generate masks followed by human verification, the performance of the deep learning model trained using the accepted images from initial masks generation step is limited due to the simplicity of the accepted images/masks generated by ASA. Therefore, leveraging the remaining unlabeled images (i.e., rejected by a user in the verification step) is critical for increasing the number of training examples for a better performing deep learning model.

To leverage the unlabeled data, Alahmari et al. proposed an algorithm called Iterative Deep Learning (IDL) [161]. The IDL extends a previously proposed algorithm [160], where a trained model f on a training dataset S is used to create masks for an unlabeled pool of microscopy images U. After that, a human verifies all masks generated by a deep learning model $f(U)$. The human verification basically makes a binary decision either accept or reject, where the accept decision is made in the case that a mask of each cell in an EDF image corresponds to an annotated cell in the manually annotated images (i.e., images with clicks for counted cells as shown in Fig. 7), otherwise, reject. The accepted images and masks are added to train set S for the next deep learning model training. This approach shows a performance boost compared to the baseline [160]; however, this approach is time-consuming because a user verifies all masks generated by a deep learning model for all unlabeled pool images U, especially if U is large. IDL is described in Algorithm 1.

An Active Deep Learning (ADL) approach was proposed by Alahmari et al. to reduce the number of masks verified by a human during the IDL process [162]. This ADL approach uses snapshot saving to save multiple models while training a single deep learning model. The snapshot models capture variation in the deep learning training where each model is saved at a different time during gradient convergence [163].

TABLE III: A summarized description and equations of the evaluation metrics that were used in the reviewed papers

Evaluation Metric	Formula	Description
Precision (P)	$P = \frac{TP}{TP+FP}$	Where true positive and false positive are represented by TP and FP respectively.
Recall (R)	$R = \frac{TP}{TP+FN}$	Where true positive and false negatives are represented by TP and FN respectively.
F1-score (F1)	$F1 = \frac{2\times P\times R}{P+R}$	Where Precision and Recall are denoted by P and R respectively
Accuracy (Acc)	$Acc = \frac{TP+TN}{TP+FP+TN+FN}$	Where true positive, true negative, false positive and false negative are denoted by TP, TN, FP, and FN respectively
Average Precision (AP)	$AP = \frac{1}{11}\sum_{r\in\{0,0.1,....,1\}} P_{interp}(r)$	The AP summarizes the shape of the interpolated precision/recall curve, and is defined as the mean precision at a set of eleven equally spaced recall levels [100]
Dice Coefficient (DC)	$DC = \frac{2*\lvert A \bigcap B\rvert}{\lvert A\rvert+\lvert B\rvert}$	Is a measurement of semantic segmentation overlap with the ground truth, where A is the ground truth and B is the predicted mask.
(Intersection over Union (IoU) Also known as Jaccard index (J)	$IoU = \frac{\lvert A \bigcap B\rvert}{\lvert A \bigcup B\rvert}$	Is a measurement of detection and segmentation overlap with the ground truth, where A is the ground truth and b is the predicted mask or bounding box.
Similarity Metric (SM) [112]	$SM(R,T) = k(\frac{1}{N}\sum_{i=1}^{N}\max\frac{2\lvert r_i \bigcap t_j\rvert}{\lvert r_i\rvert+\lvert t_j\rvert}) +(1-k)(\frac{\lvert 2\lvert P_T^R\rvert\rvert}{N+M})$	Where k is the weight factor. The first term compute average maximum overlap between each label in the ground truth $r_i \in R$, while the second term compute the ratio of true positive labels to all labels.



(a) Manual annotation (b) mask

Fig. 7: An example of a) manual annotation (counted neurons have blue X's), b) cell mask.

The ADL approach creates masks for images in unlabeled pool U using the snapshot models, followed by a confidence measure based on majority voting as shown in Algorithm 3. The ADL approach reduces the time of human verification by $\sim 25\%$ compared to IDL, because a human looks at images with high confidence rather than looking at all the images. ADL is described in Algorithm 2.

ALGORITHM 1 : Iterative Deep Learning

Input: 1) Initial training dataset S^0 which consists of EDF images and corresponding masks S^0_{EDF}, S^0_{mask}. 2) Unlabeled pool of EDF images U^0.
Output: An extended training set S^t, training model M^t on the updated training set S^t.

$S^t \leftarrow S^0$, where S^0 represents initial training set generated using ASA, M^0 is the baseline model, $t = 0$ is the baseline.

for $t = 1, 2, 3, ..$ **do**
/*Test on U^t using M^{t-1} and verify by user */
1) $R^t \leftarrow M^{t-1}(U^{t-1})$
2) Human verify ALL masks $\in R^t$
/* Update S^t and U^t and train M^t
1) $S^t \leftarrow S^{t-1} \bigcup R^t_{accepted}$
2) $U^t \leftarrow U^{t-1} \setminus R^t_{accepted}$
/* Train a new model */
3) $M^t \leftarrow M^t(S^t)$
end for

VI. Discussion

Deep Learning performance in microscopy image cell segmentation and detection based unbiased stereology quantification is significantly better than the traditional machine learning and image processing techniques [160]. However, there are some prerequisites for a more accurate unbiased stereology counting method, such as post-processing and obtaining large labeled datasets. Deep learning segmentation and detection of cells for unbiased stereology quantification in microscopy images requires a post-processing step for several reasons: 1) to apply the unbiased stereology counting rules by removing cells that intersect with exclusion lines of the disector as shown in Fig. 4. The removed cells are not counted based on unbiased stereology counting rules see Section II, 2) to separate

TABLE IV: Summary of methods that share same dataset, and the training-testing split was the same as specified by the dataset publisher. Results for the state-of-the-art methods are in bold.

Dataset	Methods	State-of-the-art
H&E colorectal adenocarcinoma [25]. Total of 100 images of resolution 500*500 associated with dot annotation. The images were extracted from 10 WSI of 9 patients. The images were digitized at 20X magnification.	1) [25] CNN regression based approach to detect the center of nuclei in H&E microscopy images. Training and testing were done using on two 50-50 split of the data. 2) [38] Prior shape incorporated to detect cell location using neural network. 50-50 split of training using [25] code. 3) [58] Generator Adversarial Network (GAN) based approach, where the generator is a combination of U-Net and Residual Attention Network for segmenting cells on microscopy images. 50-50 split of dataset for training and testing using evaluation code provided in [25] .	F1-score for: 1) [25] is 0.802 2) [38] is 0.823 3) [58] is **0.847**
H&E stained tissue images from TCGA of different tissues such as liver, breast, kidney, prostate, colon, bladder, and stomach [97].	1) [97] CNN based approach with ternary probability maps corresponding to the background, nucleus, and boundary of nucleus. Split provided by dataset provider for same organ and different organ training and testing. 2) [102] Regression based approach of distance map based segmentation to overcome the issue of touching nucleus. Split the dataset two training validation and testing, the testing set is the same as the test set for [97]. 3) [108] U-Net based approach to segment cells in images using sparse non-negative matrix factorization (SNMF). Using same data splitting provided by [97], where training was done on 16 images, and testing was done on 14 images (8 same tissue images, and 6 different tissue images). Same tissue images are from Breast, Liver, Kidney, Prostate. Different tissue images are from Bladder, colon, stomach.	F1-score for: 1) [97] is 0.826 2) [102] is 0.786 3) [108] is **0.854**

ALGORITHM 2 : Active Deep Learning

Input: 1) Initial training dataset S^0 which consists of EDF images and corresponding masks S^0_{EDF}, S^0_{mask}. 2) Unlabeled pool of EDF images U^0.
Output: An extended training set S^t, training model M^t on updated training set S^t.

$S^t \leftarrow S^0$, where S^0 represents initial training set generated using ASA, M^0 is the baseline model, $t = 0$ is the baseline.

for $t = 1, 2, 3, ..$ **do**
/*Test on U^{t-1} using M_s^{t-1} where $s \in \{10, 15, 20, ..., 100\}$ and calculate confidence based on ensemble followed by verification by a user */
1) $R^t \leftarrow Query(\{M_{snapshot}\}, U^{t-1}, K)$
2) Human verify masks $\in R^t$
/* Update S^t and U^t and train M^t
1) $S^t \leftarrow S^{t-1} \bigcup R^t_{accepted}$
2) $U^t \leftarrow U^{t-1} \setminus R^t_{accepted}$
/* Train a new model */
3) $M^t \leftarrow M^t(S^t)$
end for

ALGORITHM 3 : Query

Input: 1) Set of Snapshot models $M_{snapshot} = \{M^{t-1}_{10}, M^{t-1}_{15}, ..., M^{t-1}_{100}\}$, 2) Unlabeled pool of EDF images U, and 3) Confidence score threshold K
Output: A set of most images/masks F with high confidence score $\geq K$.

$E_{sum} = 0$ /* Zeros entries matrix */
for $M \in M_{snapshot}$ **do**
/*Test on U^{t-1} using M, and calculate confidence based on ensemble*/
1) $R_s \leftarrow M(U^{t-1})$ /*test on unlabeled set*/
2) $\forall$p(i,j)$\in$R$_s$, do: if $(p(i, j) > 0.5) = 1$, else $p(i, j) = 0$,
3) $E_{sum} \leftarrow \sum(E_{sum}, R_s)$ /* pixel-wise sum*/

end for
i) $E_{ensemble} \leftarrow E_{sum}/T$, where T = total number of $M_{snapshot}$
ii) $f \leftarrow \sum_m^n E_{ensemble}/(n + 1)$, where n = number of non zeros pixels of $E_{ensemble}$, m, n are the dimensions of $E_{ensemble}$
iii) $F \leftarrow R_{high}$ where $R_{high} \subset R_s$ of images with $f \geq K$ /* get masks where of high confidence above or equal K */

touching cells or remove small noise in the background [160][161][162].

TABLE V: Summary of the datasets, results, and cell segmentation methods, where precision, recall, F1-score, average precision, dice coefficient and Intersection over Union are denoted by P,R,F1,AP,DC,and IoU respectively.

References	Datasets and source code	Task	Results	Summary
[85]	**Dataset:** 133 digitized histology images **Source code:** not available	Segmentation	P = 0.9762 R = 0.9843 F1 = 0.9802	Extracted patch of nuclei based on Superpixel clustering, to train CNN based on LeNet to detect cells in image patches
[97]	**Dataset:** Here is the H&E stained tissue images from TCGA, from different tissues such as liver, breast, kidney prostate, colon,bladder, and stomach. **Source code:** available here	Segmentation	DC = 0.7623 F1 = 0.8267	The CNN proposed has ternary probability maps corresponding to the background, nucleus, and boundary of nucleus. Small patches of size 51*51 pixels were used in training/testing the proposed CNN
[89]	**Dataset:** A multi-channel fluorescence images acquired images of tissue from mouse pancreata [147] **Source code:** not available	Segmentation	DC = 0.824 F1 = 0.718 Object Dice (OD) = 0.741 Pixel Accuracy =0.835 Object Hausdorff (OH) = 27.5	training multi-input and multi-output with different input resolution for better cell localization, and adding extra convolutional layers to bypass pooling layer to preserve features
[105]	**Dataset:** Here is the Histopathology H&E stained images of size 512* 512 pixels **Source code:** not available	Segmentation	PangNet [120] P = 0.814 R = 0.655 F1 = 0.675 DeconvNet [121] P = 0.864 R = 0.773 F1 = 0.805 FCN [54] P = 0.823 R = 0.752 F1 = 0.763 Ensemble P = 0.741 R = 0.900 F1 = 0.802	An experimental study of three neural networks. Additionally, an ensemble of the FCN and DeconvNet was created to improve the results
[127]	**Dataset:** Total of 158 Whole Slide Images (WSI) of breast cancer tissue, 79 stained with H&E, and another 79 stained using HER2 monoclonal antibody. Resolution of each WSI is 100,000*80,000 pixels of magnification 40x. Extracted images of size 2048*2048 pixels by cropping **Source code** not available	Segmentation	P = 0.9664 R = 0.9679 F1 = 0.9671	CNN to segment cells and cell membrane in stained microscopy images. The CNN uses two blocks of Trapezoidal Short-term-Memory (TLSTM)
[136]	**Dataset:** Neural cell images from time-lapse micrscopy video, total images is 386 of size 640*512 pixels **Source code:** not available	Detection based Segmentation	AP = 0.857 IoU = 0.812	Detection based segmentation using Single Shot Detector (SSD) with VGG16 backbone. Then segmentation masks obtained using upsampling similar to [9]
[122]	**Dataset:** Total of 50 fluorescence microscopy tissue images of glioblastoma cells **Source code:** not available	Segmentation	DC = 0.911 IoU = 0.843	A comparison between supervised and unsupervised machine learning approaches' performance in segmenting cells. The approaches compared are: Global thresholding, local thresholding, fast marching level set method [123], region competition [124], random forest based segmentation using Weka [148] and Ilastik [149], and Deep learning with and without atrous spatial pyramid pooling (ASPP) [9] [150]
[94]	**Dataset:** Nuclei segmentation challenge MICCAI 2017 [96] **Source code:** not available	Segmentation	DC = 0.855	A residual based FCN that uses multiple scales input images. A custom loss function was used to account for stain variation
[102]	**Dataset:** Multiple tissues stained H&E microscopy images from two datasets [97] [105] **Source code:** available here	Regression based segmentation	F1 = 0.7793	A regression of distance map based segmentation to overcome the issue of touching nucleus using pre-trained models. Post-processing was applied to the final segmentation

Accurate segmentation and detection of cells are critical since the removal of cells touching the exclusion line is based on deep learning segmentation or detection performance. For instance, a cell that is touching the exclusion line of the disector may not get segmented entirely (i.e., cell body segmented partially); thus, the cell may not be removed by applying the unbiased stereology counting step. Similarly, if the bounding box detection of objects does not enclose the entire cell, the cell may not be removed prior to applying unbiased stereology cell counting. Another issue with detecting a cell using a bounding box detection approach is that the cell is completely inside the bounding box, but the bounding box is loose. Therefore, the cell will be removed before applying the unbiased stereology since the bounding box is touching the exclusion lines of the disector, although, the cell body is not touching the exclusion line.

TABLE VI: Summary of the datasets, results, and cell segmentation methods, where precision, recall, F1-score, accuracy, dice coefficient and Intersection over Union are denoted by P,R,F1,Acc,DC,and IoU respectively. (continuation of Table V)

References	Datasets and Source code	Task	Results	Summary
[139]	**Dataset:** Herlev dataset [141] **Source code:** not available	Segmentation	P = 0.96 R = 0.96 F1 = 0.95	Segmentation of cells in Pap smear dataset using Mask-RCNN followed by post-processing using LFCCRF
[87]	**Dataset:** Total of 141 H&E histology images from 137 patients scanned at magnification of 40x with a resolution of 2000*2000 pixels **Source code:** available here	Segmentation	F1 = 0.8218	Adaptive hierarchical approach to segment cells in large histology images using patches of size 32*32.The approach uses M AlexNet with different M resolutions, and higher resolution is used only if further refined segmentation is needed
[101]	**Dataset:** Total of 224 of Intestine H&E images of resolutions 1636*1088 pixels **Source code:** not available	Segmentation	DC = 0.74 P = 0.818 R = 0.723	A Unet like architecture with two decoder paths: one for cell body and the other decoder for boundary of cells to enhance segmentation
[107]	**Dataset:** Total of 58 H&E histopathology images of breast cancer from Yale, David Rimm's Laboratory **Source code:** not available	Segmentation	Acc = 92.45% P = 0.8241 R = 0.8604 F1 = 0.8393	Applying sparse reconstruction method on input of H&E stain images before training neural network
[86]	**Dataset:** ISBI2014 dataset **Source code:** not available	Segmentation	P = 0.994 R = 0.911	Superpixels pre-processing of input images followed by training a CNN. The post-processing of the masks includes the dynamic shape level and Voronoi Segmentation
[142]	**Dataset:** Temporal lobe cortex microscopy images from human brain of resolution 2560*2160 pixels **Source code:** not available	Segmentation	P = 0.895 R = 0.942 F1 = 0.918	Input images with weak annotation (centroids of cells) were used to train Mixed-scale Dense (MS-D) architecture [143] The predicted centroids in the test set were used as a seed for Multi-object Geometric Deformable Model (MGDM) [144] to segment cells correctly
[116]	**Dataset:** Mouse brain microscopy images **Source code:** not available	Segmentation	DC = 0.767	A cascaded approach of two neural networks, where the probability map from the first is used to train the second neural network to learn high-level contextual information
[125]	**Dataset:** Total of 50 tissue microscopy DAPI stained images of glioblastoma cells acquired at 63x **Source code:** not available	Segmentation	IoU = 0.7981	This method uses U-Net architecture [9] with residual connections [44] and ASPP [99], and fine-tune the neural network on different datasets with different number of image channels by copying the weights for each channel
[108]	**Dataset:** 1) H&E-stained image dataset from 7 organ (MOD) [97], total images is 30 of 1000*1000 pixel resolution 2) Breast cancer histopathology image dataset (BCD) [151], total images is 39 of 1000*1000 pixel resolution 3) Breast cancer H&E stained images dataset (BNS) [105], total 33 images, with resolution of 512*512 pixels **Source code:** available here	Segmentation	1) MOD dataset: P = 0.813 R = 0.914 F1 = 0.854 DC = 0.812 2) BCD dataset P = 0.942 R = 0.915 F1 = 0.923 DC = 0.862 3)BNS dataset P = 0.920 R = 0.7835 F1 = 0.84 DC = 0.83	A neural network similar to Unet to segment cells in normalized stain H&E images using sparse non-negative matrix factorization (SNMF) [109]. The neural network uses patches from original images during training, and overlapping patches during testing
[110]	**Dataset:** H&E dataset from 7 different organs [111] **Source code:** not available	Segmentation	DC = 0.7932 F1 = 0.8188	A two neural network approach, where the first neural network outputs the probability map of cell segmentation, and the second neural network creates a regression output for the distance map, which used to seed watershed for refined segmentation
[152]	**Dataset :**1) FIB-SEM dataset [153] 2) ATUM-SEM dataset [154] **Source code:** available here	Segmentation	ATUM-SEM : P = 0.911 R = 0.934 F1 = 0.922 FIB-SEM : P = 0.882 R = 0.938 F1 = 0.909	3D residual network to segment Mitochondria

Thus, unbiased stereology cell counting on a microscopy image requires: 1) accurate cell body segmentation, and 2) accurate cell body detection. Some of the other methods, such as those that detect or segment the cell center only [20] [23] [24] [26] [25] [32] [37] [38] [43] [46] [48] [45] [47] [55] [56] are not suitable for unbiased stereology because entire cell body segmentation or detection is required for applying unbiased stereology rules.

In Fig. 8, examples are given to illustrate unbiased stereology rules and the effect of inaccurate segmentation or

TABLE VII: Summary of the datasets, results, and cell segmentation methods, where F1-score, average similarity metric, jaccard measure, and dice coefficient are denoted by F1, Avg SM J,and DC respectively (continuation of Table V and VI)

References	Datasets and source code	Task	Results	Summary
[112]	**Dataset:** Fluorescent microscope images **Source code:** not available	Segmentation	Avg SM = 0.86	FCN similar to U-Net for training and testing followed by post-processing of feature maps
[98]	**Dataset:** Gland segmentation challenge dataset Nuclei segmentation challenge dataset **Source code:** not available	Segmentation	DC = 0.876	Deep Contour Aware Neural network (DCAN) of one encoder initialized using DeepLab model [99], and two decoders (one for nuclei body and the other for nuclei contour) initialized randomly
[117]	**Dataset:** 1) Breast cancer dataset [118], 2) Kidney renal clear cell carcinoma dataset [119]. **Source code:** not available	Segmentation	1) Breast Cancer dataset CNN: F1 = 0.8967 SAE: F1 = 0.8874 DBN: F1 = 0.8785 2) Kidney renal dataset CNN: F1 = 0.9156 SAE: F1 = 0.9123 DBN: F1 = 0.8793 Note: results shown here are for highest test samples with window size = (13*13)	An empirical study of the effect of spatial information (different window sizes patches extracted from original image) on the results of CNN, SAE, and DBN.
[146]	**Dataset:** Microscopy images from three datasets: PhC-HeLa [36], Fluo-N2DL-HeLa [60], and Hist-BM [53] **Source code:** available here	Proposal based segmentation	Fluo-N2DL-HeLa: DC = 0.874 PhC-HeLa: DC = 0.818 Hist-BM: DC =0.823	Two neural networks: the first proposes K cells and the second neural network uses these proposed cells to get the segmentation masks
[131]	**Dataset:** Microscopy images from H1299 dataset:consisting of 72 frame of size 512*640 pixels [133] **Source code:** available here	Segmentation	J = 0.806	Unlike generative adversarial networks, this approach does not generate images from random noise vector, but instead it estimate segmentation of cells on an image The proposed method consists of estimator and discriminator networks, where the estimator learns to generate cell segmentation, while the discriminator learns to discriminate between the generated mask and the manual ground truth mask (manual mask).
[134]	**Dataset:** Microscopy images of human iRPE cells. Images patches of size 256*256 was extracted, where total was 1000 images. [135] **Source code:** not available	Segmentation	DC = 0.637	The approach train generative adversarial networks (GAN) to learn weight parameters, then the weights from the discriminator is transferred to the encoder part of U-Net architecture. The U-Net architecture was further trained with small manually annotated microscopy images to segment contours of cells. Moreover, transfer learning using model trained on annotate Common Objects in Context (COCO) data set was done. [104]. Furthermore, the authors tried also augmentations of the small manually annotated microscopy images data set for training U-Net.

detection. The disector box is shown on top of the images, where green lines are the inclusion lines (cell touching inclusion lines is counted), whereas, the red lines are the exclusion lines (cell touching the exclusion lines is not counted). In Fig. 8a, a disector box is shown with multiple objects (i.e., cells), one of the cells is touching the exclusion line, and thus is excluded from unbiased stereology counting. However, to exclude this cell, an accurate segmentation or detection has to show that the cell is touching the exclusion line. For instance, Fig. 8b and Fig. 8c show the accurate detection and segmentation (respectively) of the cell touching the exclusion line. On the other hand, Fig. 8d and Fig 8e show inaccurate segmentation and detection of the cell touching the exclusion line. This cell will be counted (false-positive), although the cell body is touching the exclusion line. Another example of inaccurate detection is shown in Fig. 8f, where the cell in the lower left will not be counted due to the intersection of the bounding box with the disector exclusion line, although the cell itself is not touching the exclusion line (false-negative).

Microscopy image analysis must take into consideration the variation in scanners, staining, sampling, and tissue texture. Recent work has been trying to address these issues by designing neural networks to overcome stain and multi-tissue variation [94]. These methods account for stain variations between cells and within the cell body by applying stain normalization [95] and designing a weighted loss function. The stain normalization methods such as [109] [111] [95] reduce the variations of stains caused by stain absorption variations, scanners, and light setting differences. Another technique to overcome the stain variation and complications

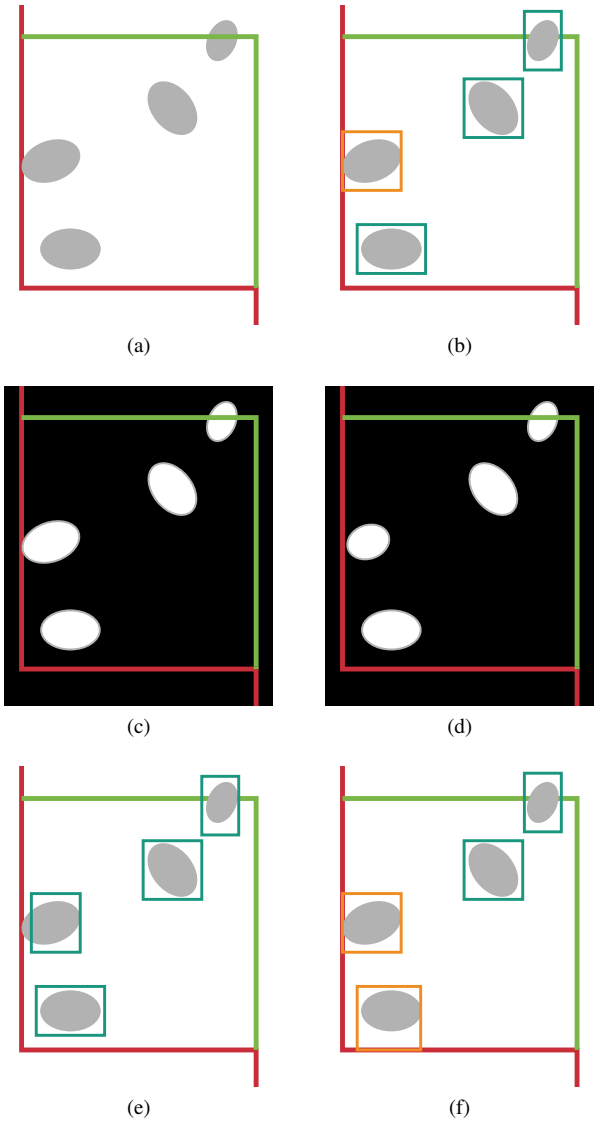


Fig. 8: Examples of inaccurate segmentation and detection impact on unbiased stereology cell counting. a) image with three cells, one of the cells is touching the exclusion line (red line), b) illustration of accurate detection c) illustrates accurate segmentation, d) an example of inaccurate segmentation (because mask on the left is not touching exclusion line) , e) and f) are examples of inaccurate detection. Orange bounding box indicates that the cell will not be counted since it touches exclusion line.

is to remove the counter-stain while maintaining the principle stain which is called stain deconvolution [27][72]. However, some research found that using the stained normalized microscopy images is better than color deconvolution since color deconvolution could remove some critical information of cell body and tissue structure in microscopy images [108]. Additionally, two stains may get absorbed together by a cell body, which makes stain decomposition (i.e., deconvolution) difficult.

In some diseases such as cancer and Alzheimer's, where the cells are continually changing, some cells appear small in size or have deformed shape due to inflammation. Therefore, the size of cell and shape variation due to the cell mutation is a challenging task in microscopy image analysis and unbiased stereology in particular. Therefore, studying microscopy images requires collecting and labeling a large number of images to enable supervised learning algorithms such as deep learning, to learn discriminative features that generalize to more subjects and perform accurately in cell segmentation and detection.

Supervised learning algorithms learn to map input x to the output y (i.e., $y = f(x)$), therefore, the supervised learning algorithm f needs to learn features from a substantial training set X with associated labels Y. Training set X is currently large and sufficient to train a good learning model due to the availability of sophisticated scanners and imaging tools. However, obtaining labels Y remains difficult because manual labeling is time-consuming and labor-intensive. Additionally, microscopy images require an expert to produce the label rather than relying on a crowd-sourcing technique [119] to annotate images. Therefore, leveraging unlabeled data using active learning, transfer learning, unsupervised learning algorithms, and self-supervised learning approaches is a crucial step for quickly obtaining less expensive labels to train a deep learning model. For instance, [160] applied an unsupervised learning algorithm to get initial segmentation masks followed by a human verification process to accept or reject rather than manually creating pixel-wise labels (masks). In [162][164][165], active learning approaches were applied to leverage unlabeled data by generating labels for images in unlabeled pool U. Followed by decision making by a human (i.e, either accept/reject or relabel uncertain instances) to add images to the training set. This approach increases and diversifies training set instances to train a better performing deep learned model.

Whole slide microscopy images (WSI) contain a large number of objects; therefore, crowding, touching, and overlapping of objects (i.e., cells) in microscopy images are obstacles to more accurate segmentation and detection. Several studies have proposed to segment each cell body and its contour using deep learning to obtain a separate mask for each cell accurately [101][97][108]. However, this approach is very applicable when cells are touching; but, when partial overlap of cells occurs, these approaches fail. Another method that uses shape information to segment cells of specific shapes could segment partially overlapped cells [38]. However, this approach is also limited since cells could appear in any shape, especially in the case of Alzheimer and cancer tissues, where cells are continuously deforming.

Despite the high correlation of augmented images to the original image, image argumentation such as random cropping, rotation, flipping, and elastic deformation could help balance the datasets for better classification of two or more classes. However, the class imbalance within the same image is im-

portant for better pixel-level segmentation or object detection. For instance, microscopy images are large resolution images; therefore, cropping or patch extraction is usually done to reduce deep learning input image resolution and thus to reduce the computation of deep learning models. Hence, the chance of having microscopy image patches with different cell quantities is high, which creates an imbalance of object and background (i.e., two-class pixel-level classification). In [166], a focal loss was proposed to penalize the loss for a well-classified class during the training of a one-stage object detector such as SSD. Therefore, accounting for cell imbalance when extracting patches is important in training segmentation or detection model on microscopy images.

VII. Conclusion and Future Trends

Current practices to handle high-resolution microscopy images are to extract overlapping or non-overlapping patches that are used individually as input to the deep learning architecture. After training, results from patches are aggregated together to obtain the final output. Although this approach reduces the overwhelming computation time significantly, it does not take the spatial information of the WSI into account. For instance, the relations between the nucleus and the surrounding cytoplasm and tissue structure carry useful information that is relevant and important in many cancer gradings of a tissue sample. Hence, a method that integrates neighboring patches for unified feature learning using deep learning could improve the performance and efficiency of current deep learning models trained on patches of WSI. One potential approach is to input more related patches at a particular point in the neural network architecture. Thus, a fusion of the learned features from one patch with deep features of the other patch for further training could improve the performance of a deep learning model.

Data labeling is expensive and time-consuming due to the vast availability of data nowadays. Leveraging unlabeled data is of great interest in many supervised learning-based applications. Many applications can rely on crowd-sourcing techniques [119] to get labels such as (annotating animal images), where expert in the field knowledge is not needed to obtain optimal labels. In the medical field, annotation (labeling) requires expert knowledge, and can not be done by crowd-sourcing based approaches. Therefore, active learning and semi-supervised learning approaches [162][161] to obtain pseudo labels for instances from unlabeled data to train more deep learning models in an iterative approach are critical and useful for more accurate deep learned models. Recently, self-supervised learning approaches that leverage unlabeled data have shown promise to reduce the expert effort in data labeling and to improve learning model performance. For instance, [167] uses worker (U-Net) and supervisor (ResNet) models trained together to learn optimal object segmentation, where the worker model learns pixel-level classification, and a supervisor model learns regression of dice coefficient. Then, unlabeled data is used to generate labels where the supervisor guides the learning process. We believe that active learning and self-supervised learning approaches have a great potential for success in microscopy image analysis.

In microscopy image analysis where quantification of the number of cells is critical, accurate segmentation and detection of cells plays a significant role in understanding the behavior of diseases. However, microscopy images present many challenges such as cell overlapping, variations in stains, stain absorption, sampling of tissues, microscopy cameras, and light settings during microscopy image acquisition. Although there have been works that tackle some of these issues as discussed in Section VI, the generalizability of such methods to large datasets of different organs and stains is unknown because these methods are stain and tissue-specific. Therefore, we believe a more comprehensive evaluation of the applicability of prior cell detection and segmentation to different stains, tissue samples using large datasets will present an excellent contribution to the field of microscopy image analysis.

Sabour et al. introduced a new type of neural network called Capsule Networks [168]. Capsule Networks replace max-pooling layers with convolutional layers with strides and a dynamic routing algorithm [168]. Additionally, each piece of information at a single neuron of the neural network is stored as a capsule vector rather than a scalar in CNN based networks. Henceforth, more information is stored, such as orientation, and the magnitude of the spatial features. The dynamic routing algorithm [168] routes capsule vectors based on their agreement to the next layer of the capsule network. This technique enables the preservation of information and maintains information on the relationship of part-to-whole. Since the relationship between cell, surrounding objects, and cytoplasm carries information. Capsule Networks for microscopy image analysis will enable learning robust features that take the relationship of nuclei and the surrounding tissue parts into account. SegCaps [10] introduced deconvolution modules for Capsule Networks in a similar approach to U-Net [9] to perform segmentation on the LUNA16 dataset. We believe that Capsule Networks applied to microscopy image analysis for cell segmentation, detection, classification, and unbiased stereology will have huge success and will be the future trend.

In unbiased stereology, bounding-box localization of cell-based counting is an approach that could alleviate the burden of counting overlapping cells and may require less post-processing compared to segmentation based unbiased stereology cell counting. Therefore, we believe localization-based unbiased stereology could gain more interest in the future. On the other hand, unbiased stereology counting is based on stacks of images (3D); therefore, using individual frames from stacks to train a deep learning model for detection and segmentation of cells could improve the performance of unbiased stereology cell counting compared to a single in-focus image from each stack [158][162].

This review paper provides a summary of deep learning-

based cell detection and segmentation approaches in the literature. Furthermore, the article highlights the recent applications of deep learning to unbiased stereology. Also, we offer rules for applying deep learning for cell segmentation and detection-based unbiased stereology cell counting. Limitations and challenges of deep learning in microscopy image analysis were presented. We also sketched future work and trends that we believe will gain more interest.

ACKNOWLEDGMENT

Grant support: National Science Foundation STTR grants 1746511 (Phase 1) and 1926990 (Phase 2).

REFERENCES

[1] M. N. Wernick, Y. Yang, J. G. Brankov, G. Yourganov, and S. C. Strother, "Machine learning in medical imaging," *IEEE signal processing magazine*, vol. 27, no. 4, pp. 25–38, 2010.

[2] S. U. Akram, J. Kannala, L. Eklund, and J. Heikkilä, "Cell proposal network for microscopy image analysis," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3199–3203.

[3] V. Grau, A. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield, "Improved watershed transform for medical image segmentation using prior information," *IEEE transactions on medical imaging*, vol. 23, no. 4, pp. 447–458, 2004.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[6] A. Izbassarova, A. Duisembay, and A. P. James, "Speech recognition application using deep learning neural network," in *Deep Learning Classifiers with Memristive Networks*. Springer, 2020, pp. 69–79.

[7] Y.-H. Tu, I. Tashev, S. Zarar, and C.-H. Lee, "A hybrid approach to combining conventional and deep learning techniques for single-channel speech enhancement and recognition," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2531–2535.

[8] L. Deng, D. Yu *et al.*, "Deep learning: methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[10] R. LaLonde and U. Bagci, "Capsules for object segmentation," *arXiv preprint arXiv:1804.04241*, 2018.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[12] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *Ieee Access*, vol. 6, pp. 9375–9389, 2017.

[13] F. Xing, Y. Xie, H. Su, F. Liu, and L. Yang, "Deep learning in microscopy image analysis: A survey," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 10, pp. 4550–4568, 2017.

[14] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[15] P. R. Mouton, *Principles and practices of unbiased stereology: an introduction for bioscientists*. JHUP, 2002.

[16] ——, *Unbiased stereology: a concise guide*. JHU Press, 2011.

[17] D. Sterio, "The unbiased estimation of number and sizes of arbitrary particles using the disector," *Journal of microscopy*, vol. 134, no. 2, pp. 127–136, 1984.

[18] H.-J. G. Gundersen, "Stereology of arbitrary particles* a review of unbiased number and size estimators and the presentation of some new ones, in memory of william r. thompson," *Journal of microscopy*, vol. 143, no. 1, pp. 3–45, 1986.

[19] S. D. Wicksell, "The corpuscle problem: a mathematical study of a biometric problem," *Biometrika*, pp. 84–99, 1925.

[20] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 119–130, 2016.

[21] H. Fatakdawala, J. Xu, A. Basavanhally, G. Bhanot, S. Ganesan, M. Feldman, J. E. Tomaszewski, and A. Madabhushi, "Expectation–maximization-driven geodesic active contour with overlap resolution (emagacor): Application to lymphocyte segmentation on breast cancer histopathology," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1676–1689, 2010.

[22] H. Chang, J. Han, A. Borowsky, L. Loss, J. W. Gray, P. T. Spellman, and B. Parvin, "Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association," *IEEE transactions on medical imaging*, vol. 32, no. 4, pp. 670–682, 2012.

[23] R. Mufidah, I. Wasito, N. Hanifah, M. Faturrahman, and F. D. Ghaisani, "Automatic nucleus detection of pap smear images using stacked sparse autoencoder (ssae)," in *Proceedings of the International Conference on Algorithms, Computing and Systems*. ACM, 2017, pp. 9–13.

[24] T.-H. Song, V. Sanchez, H. EIDaly, and N. M. Rajpoot, "Hybrid deep autoencoder with curvature gaussian for detection of various types of cells in bone marrow trephine biopsy images," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 1040–1043.

[25] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[26] T.-H. Song, V. Sanchez, H. EIDaly, and N. Rajpoot, "Simultaneous cell detection and classification with an asymmetric deep autoencoder in bone marrow histology images," in *Annual Conference on Medical Image Understanding and Analysis*. Springer, 2017, pp. 829–838.

[27] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1729–1738, 2014.

[28] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang, "Beyond classification: Structured regression for robust cell detection using convolutional neural network," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2015, pp. 358–365.

[29] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 119–130, 2015.

[30] M. Kuse, Y.-F. Wang, V. Kalasannavar, M. Khan, and N. Rajpoot, "Local isotropic phase symmetry measure for detection of beta cells and lymphocytes," *Journal of pathology informatics*, vol. 2, 2011.

[31] Y. Yuan, H. Failmezger, O. M. Rueda, H. R. Ali, S. Gräf, S.-F. Chin, R. F. Schwarz, C. Curtis, M. J. Dunning, H. Bardwell *et al.*, "Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling," *Science translational medicine*, vol. 4, no. 157, pp. 157ra143–157ra143, 2012.

[32] S. Wang, J. Yao, Z. Xu, and J. Huang, "Subtype cell detection with an accelerated deep convolution neural network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 640–648.

[33] 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.

[34] H. Li, R. Zhao, and X. Wang, "Highly efficient forward and backward propagation of convolutional neural networks for pixelwise classification," *arXiv preprint arXiv:1412.4526*, 2014.

[35] H. Pan, Z. Xu, and J. Huang, "An effective approach for robust lung cancer cell detection," in *International Workshop on Patch-based Techniques in Medical Imaging*. Springer, 2015, pp. 87–94.

[36] C. Arteta, V. Lempitsky, J. A. Noble, and A. Zisserman, "Learning to detect cells using non-overlapping extremal regions," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2012, pp. 348–356.

[37] Z. Xu and J. Huang, "Detecting 10,000 cells in one second," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 676–684.

[38] M. Tofighi, T. Guo, J. K. Vanamala, and V. Monga, "Deep networks with shape priors for nucleus detection," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 719–723.

[39] S. Ali and A. Madabhushi, "An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery," *IEEE transactions on medical imaging*, vol. 31, no. 7, pp. 1448–1460, 2012.

[40] M. E. Leventon, W. E. L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *5th IEEE EMBS International Summer School on Biomedical Imaging, 2002.* IEEE, 2002, pp. 8–pp.

[41] M. Tofighi, T. Guo, J. K. Vanamala, and V. Monga, "Prior information guided regularized deep learning for cell nucleus detection," *IEEE transactions on medical imaging*, 2019.

[42] N. A. Koohababni, M. Jahanifar, A. Gooya, and N. Rajpoot, "Nuclei detection using mixture density networks," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2018, pp. 241–248.

[43] C. M. Bishop, "Mixture density networks," Citeseer, Tech. Rep., 1994.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[45] Y. Xie, F. Xing, X. Shi, X. Kong, H. Su, and L. Yang, "Efficient and robust cell detection: A structured regression approach," *Medical image analysis*, vol. 44, pp. 245–254, 2018.

[46] J. Huang and Z. Xu, "Cell detection with deep learning accelerated by sparse kernel," in *Deep Learning and Convolutional Neural Networks for Medical Image Computing*. Springer, 2017, pp. 137–157.

[47] D. Wang, K. Wu, C. Gu, and X. Guan, "Time efficient cell detection in histopathology images using convolutional regression networks," in *Control Conference (CCC), 2017 36th Chinese*. IEEE, 2017, pp. 10 962–10 965.

[48] T. Wollmann and K. Rohr, "Deep residual hough voting for mitotic cell detection in histopathology images," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 341–344.

[49] M. Veta, P. J. Van Diest, S. M. Willems, H. Wang, A. Madabhushi, A. Cruz-Roa, F. Gonzalez, A. B. Larsen, J. S. Vestergaard, A. B. Dahl *et al.*, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Medical image analysis*, vol. 20, no. 1, pp. 237–248, 2015.

[50] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2013, pp. 411–418.

[51] Y. Xie, X. Kong, F. Xing, F. Liu, H. Su, and L. Yang, "Deep voting: A robust approach toward nucleus localization in microscopy images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 374–382.

[52] D. Dhall, R. Mertens, C. Bresee, R. Parakh, H. L. Wang, M. Li, G. Dhall, S. D. Colquhoun, D. Ines, F. Chung *et al.*, "Ki-67 proliferative index predicts progression-free survival of patients with well-differentiated ileal neuroendocrine tumors," *Human pathology*, vol. 43, no. 4, pp. 489–495, 2012.

[53] P. Kainz, M. Urschler, S. Schulter, P. Wohlhart, and V. Lepetit, "You should use regression to detect cells," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 276–283.

[54] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[55] R. Zhu, D. Sui, H. Qin, and A. Hao, "An extended type cell detection and counting method based on fcn," in *Bioinformatics and Bioengineering (BIBE), 2017 IEEE 17th International Conference on*. IEEE, 2017, pp. 51–56.

[56] J. G. Jacobs, G. J. Brostow, A. Freeman, D. C. Alexander, and E. Panagiotaki, "Detecting and classifying nuclei on a budget," in *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. Springer, 2017, pp. 77–86.

[57] H. Höfener, A. Homeyer, N. Weiss, J. Molin, C. F. Lundström, and H. K. Hahn, "Deep learning nuclei detection: A simple approach can deliver state-of-the-art results," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 43–52, 2018.

[58] J. Li, W. Shao, Z. Li, W. Li, and D. Zhang, "Residual attention generative adversarial networks for nuclei detection on routine colon cancer histology images," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2019, pp. 142–150.

[59] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.

[60] M. Maška, V. Ulman, D. Svoboda, P. Matula, P. Matula, C. Ederra, A. Urbiola, T. España, S. Venkatesan, D. M. Balak *et al.*, "A benchmark for comparison of cell tracking algorithms," *Bioinformatics*, vol. 30, no. 11, pp. 1609–1617, 2014.

[61] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.

[62] X. Zhang, G. Chen, K. Saruta, and Y. Terata, "An end-to-end cells detection approach for colon cancer histology images," in *Tenth International Conference on Digital Image Processing (ICDIP 2018)*, vol. 10806. International Society for Optics and Photonics, 2018, p. 108065D.

[63] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[64] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[65] J. Yi, P. Wu, M. Jiang, D. J. Hoeppner, and D. N. Metaxas, "Instance segmentation of neural cells," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

[66] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[67] C. Li, X. Wang, W. Liu, and L. J. Latecki, "Deepmitosis: Mitosis detection via deep detection, verification and segmentation networks," *Medical image analysis*, vol. 45, pp. 121–133, 2018.

[68] L. Roux, D. Racoceanu, N. Loménie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. Le Naour, and M. N. Gurcan, "Mitosis detection in breast cancer histological images an icpr 2012 contest," *Journal of pathology informatics*, vol. 4, 2013.

[69] "Mitos-atypia-14-dataset," 2014. [Online]. Available: https://mitos-atypia-14.grand-challenge.org/Dataset/

[70] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv preprint arXiv:1506.01497*, 2015.

[71] M. Kowal, M. Żejmo, and J. Korbicz, "Nuclei detection in cytological images using convolutional neural network and ellipse fitting algorithm," in *International Conference on Artificial Intelligence and Soft Computing*. Springer, 2018, pp. 157–167.

[72] A. C. Ruifrok, D. A. Johnston *et al.*, "Quantification of histochemical staining by color deconvolution," *Analytical and quantitative cytology and histology*, vol. 23, no. 4, pp. 291–299, 2001.

[73] M. N. Gurcan, A. Madabhushi, and N. Rajpoot, "Pattern recognition in histopathological images: An icpr 2010 contest," in *Recognizing Patterns in Signals, Speech, Images and Videos*. Springer, 2010, pp. 226–234.

[74] E. Garcia, R. Hermoza, C. B. Castanon, L. Cano, M. Castillo, and C. Castanñeda, "Automatic lymphocyte detection on gastric cancer ihc images using deep learning," in *Computer-Based Medical Systems (CBMS), 2017 IEEE 30th International Symposium on*. IEEE, 2017, pp. 200–204.

[75] J. Molin, A. Bodén, D. Treanor, M. Fjeld, and C. Lundström, "Scale stain: Multi-resolution feature enhancement in pathology visualization," *arXiv preprint arXiv:1610.04141*, 2016.

[76] N. L. S. T. R. Team, "The national lung screening trial: overview and study design," *Radiology*, vol. 258, no. 1, pp. 243–253, 2011.

[77] K. Chen, N. Zhang, L. Powers, and J. Roveda, "Cell nuclei detection and segmentation for computational pathology using deep learning," in *Proceedings of the Modeling and Simulation in Medicine Symposium*. Society for Computer Simulation International, 2019, p. 12.

[78] R. Ravin, D. J. Hoeppner, D. M. Munno, L. Carmel, J. Sullivan, D. L. Levitt, J. L. Miller, C. Athaide, D. M. Panchision, and R. D. McKay, "Potency and fate specification in cns stem cell populations in vitro," *Cell stem cell*, vol. 3, no. 6, pp. 670–680, 2008.

[79] E. Meijering, "Cell segmentation: 50 years down the road [life sciences]," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 140–145, 2012.

[80] C. Genestie, B. Zafrani, B. Asselain, A. Fourquet, S. Rozan, P. Validire, A. Vincent-Salomon, and X. Sastre-Garau, "Comparison of the prognostic value of scarff-bloom-richardson and nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems." *Anticancer research*, vol. 18, no. 1B, pp. 571–576, 1998.

[81] A. Basavanhally, M. Feldman, N. Shih, C. Mies, J. Tomaszewski, S. Ganesan, and A. Madabhushi, "Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: Comparison to oncotype dx," *Journal of pathology informatics*, vol. 2, 2011.

[82] M. Thoma, "A survey of semantic segmentation," *arXiv preprint arXiv:1602.06541*, 2016.

[83] X. Liu, Z. Deng, and Y. Yang, "Recent progress in semantic image segmentation," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1089–1106, 2019.

[84] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk *et al.*, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[85] S. Sornapudi, R. J. Stanley, W. V. Stoecker, H. Almubarak, R. Long, S. Antani, G. Thoma, R. Zuna, and S. R. Frazier, "Deep learning nuclei detection in digitized histology images by superpixels," *Journal of pathology informatics*, vol. 9, 2018.

[86] A. Tareef, Y. Song, H. Huang, Y. Wang, D. Feng, M. Chen, and W. Cai, "Optimizing the cervix cytological examination based on deep learning and dynamic shape modeling," *Neurocomputing*, vol. 248, pp. 28–40, 2017.

[87] A. Janowczyk, S. Doyle, H. Gilmore, and A. Madabhushi, "A resolution adaptive deep hierarchical (radhical) learning scheme applied to nuclear segmentation of digital pathology images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 3, pp. 270–276, 2018.

[88] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[89] S. E. A. Raza, L. Cheung, D. Epstein, S. Pelengaris, M. Khan, and N. M. Rajpoot, "Mimo-net: A multi-input multi-output convolutional neural network for cell segmentation in fluorescence microscopy images," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 337–340.

[90] H. Chen, X. Qi, L. Yu, and P.-A. Heng, "Dcan: deep contour-aware networks for accurate gland segmentation," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2016, pp. 2487–2496.

[91] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 550–566, 2016.

[92] C. R. Maurer, R. Qi, and V. Raghavan, "A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 265–270, 2003.

[93] P. Soille, *Morphological image analysis: principles and applications*. Springer Science & Business Media, 2013.

[94] S. Graham and N. M. Rajpoot, "Sams-net: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 590–594.

[95] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001.

[96] "Computational precision medicine nuclei segmentation challenge website." [Online]. Available: http://miccai.cloudapp.net/competitions/57

[97] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.

[98] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, "Dcan: Deep contour-aware networks for object instance segmentation from histology images," *Medical image analysis*, vol. 36, pp. 135–146, 2017.

[99] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[100] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.

[101] H. Oda, H. R. Roth, K. Chiba, J. Sokolić, T. Kitasaka, M. Oda, A. Hinoki, H. Uchida, J. A. Schnabel, and K. Mori, "Besnet: Boundary-enhanced segmentation of cells in histopathological images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 228–236.

[102] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE Transactions on Medical Imaging*, 2018.

[103] D. A. Van Valen, T. Kudo, K. M. Lane, D. N. Macklin, N. T. Quach, M. M. DeFelice, I. Maayan, Y. Tanouchi, E. A. Ashley, and M. W. Covert, "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments," *PLoS computational biology*, vol. 12, no. 11, p. e1005177, 2016.

[104] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[105] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Nuclei segmentation in histopathology images using deep neural networks," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 933–936.

[106] M. Grimaud, "New measure of contrast: the dynamics," in *Image Algebra and Morphological Image Processing III*, vol. 1769. International Society for Optics and Photonics, 1992, pp. 292–306.

[107] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, L. Zhao, and Y. Fan, "Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks," *Neurocomputing*, vol. 229, pp. 88–99, 2017.

[108] Y. Cui, G. Zhang, Z. Liu, Z. Xiong, and J. Hu, "A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images," *Medical & biological engineering & computing*, vol. 57, no. 9, pp. 2027–2043, 2019.

[109] A. Vahadane, T. Peng, S. Albarqouni, M. Baust, K. Steiger, A. M. Schlitter, A. Sethi, I. Esposito, and N. Navab, "Structure-preserved color normalization for histological images," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 1012–1015.

[110] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, Ö. Smedby, and C. Wang, "A two-stage u-net algorithm for segmentation of nuclei in h&e-stained tissues," in *European Congress on Digital Pathology*. Springer, 2019, pp. 75–82.

[111] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2009, pp. 1107–1110.

[112] Y. Al-Kofahi, A. Zaltsman, R. Graves, W. Marshall, and M. Rusu, "A deep learning-based algorithm for 2-d cell segmentation in microscopy images," *BMC bioinformatics*, vol. 19, no. 1, p. 365, 2018.

[113] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention," *International Journal of Computer Vision*, vol. 11, no. 3, pp. 283–318, 1993.

[114] P.-S. Liao, T.-S. Chen, P.-C. Chung *et al.*, "A fast algorithm for multilevel thresholding," *J. Inf. Sci. Eng.*, vol. 17, no. 5, pp. 713–727, 2001.

[115] C. Jung and C. Kim, "Segmenting clustered nuclei using h-minima transform-based marker extraction and contour parameterization," *IEEE transactions on biomedical engineering*, vol. 57, no. 10, pp. 2600–2604, 2010.

[116] Q. Wang, S. Wang, X. Zhu, T. Liu, Z. Humphrey, V. Ghukasyan, M. Conway, E. Scott, G. Fragola, K. Bradford *et al.*, "Accurate and high throughput cell segmentation method for mouse brain nuclei using cascaded convolutional neural network," in *International Workshop on Patch-based Techniques in Medical Imaging*. Springer, 2017, pp. 55–62.

[117] N. Hatipoglu and G. Bilgin, "Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships," *Medical & biological engineering & computing*, vol. 55, no. 10, pp. 1829–1848, 2017.

[118] E. D. Gelasca, B. Obara, D. Fedorov, K. Kvilekval, and B. Manjunath, "A biosegmentation benchmark for evaluation of bioimage analysis methods," *BMC bioinformatics*, vol. 10, no. 1, p. 368, 2009.

[119] H. Irshad, L. Montaser-Kouhsari, G. Waltz, O. Bucur, J. Nowak, F. Dong, N. W. Knoblauch, and A. H. Beck, "Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd," in *Pacific symposium on biocomputing Co-chairs*. World Scientific, 2014, pp. 294–305.

[120] B. Pang, Y. Zhang, Q. Chen, Z. Gao, Q. Peng, and X. You, "Cell nucleus segmentation in color histopathological imagery using convolutional networks," in *Pattern Recognition (CCPR), 2010 Chinese Conference on*. IEEE, 2010, pp. 1–5.

[121] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1520–1528.

[122] D. Baltissen, T. Wollmann, M. Gunkel, I. Chung, H. Erfle, K. Rippe, and K. Rohr, "Comparison of segmentation methods for tissue microscopy images of glioblastoma cells," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 396–399.

[123] J. A. Sethian, "A fast marching level set method for monotonically advancing fronts," *Proceedings of the National Academy of Sciences*, vol. 93, no. 4, pp. 1591–1595, 1996.

[124] J. Cardinale, G. Paul, and I. F. Sbalzarini, "Discrete region competition for unknown numbers of connected regions," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3531–3545, 2012.

[125] T. Wollmann, J. Ivanova, M. Gunkel, I. Chung, H. Erfle, K. Rippe, and K. Rohr, "Multi-channel deep transfer learning for nuclei segmentation in glioblastoma cell tissue images," in *Bildverarbeitung für die Medizin 2018*. Springer, 2018, pp. 316–321.

[126] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[127] M. Saha and C. Chakraborty, "Her2net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2189–2200, 2018.

[128] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.

[129] A. Kendall, V. Badrinarayanan, and R. Cipolla, "Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding," *arXiv preprint arXiv:1511.02680*, 2015.

[130] "Her2 scoring contest." [Online]. Available: http://www2.warwick.ac.uk/fac/sci/dcs/research/tia/her2contest/

[131] A. Arbelle and T. R. Raviv, "Microscopy cell segmentation via adversarial neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 645–648.

[132] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[133] A. A. Cohen, N. Geva-Zatorsky, E. Eden, M. Frenkel-Morgenstern, I. Issaeva, A. Sigal, R. Milo, C. Cohen-Saidon, Y. Liron, Z. Kam *et al.*, "Dynamic proteomics of individual cancer cells in response to a drug," *science*, vol. 322, no. 5907, pp. 1511–1516, 2008.

[134] M. Majurski, P. Manescu, S. Padi, N. Schaub, N. Hotaling, C. Simon Jr, and P. Bajcsy, "Cell image segmentation using generative adversarial networks, transfer learning, and augmentations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[135] "irpe data set." [Online]. Available: https://isg.nist.gov/deepzoomweb/data/RPEimplants

[136] J. Yi, P. Wu, D. J. Hoeppner, and D. Metaxas, "Pixel-wise neural cell instance segmentation," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018, pp. 373–377.

[137] J. Dai, K. He, and J. Sun, "Instance-aware semantic segmentation via multi-task network cascades," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3150–3158.

[138] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, "Fully convolutional instance-aware semantic segmentation," *arXiv preprint arXiv:1611.07709*, 2016.

[139] Y. Liu, P. Zhang, Q. Song, A. Li, P. Zhang, and Z. Gui, "Automatic segmentation of cervical nuclei based on deep learning and a conditional random field," *IEEE Access*, vol. 6, pp. 53 709–53 721, 2018.

[140] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2980–2988.

[141] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," *Nature inspired Smart Information Systems (NiSIS 2005)*, pp. 1–9, 2005.

[142] K. Thierbach, P.-L. Bazin, W. de Back, F. Gavriilidis, E. Kirilina, C. Jäger, M. Morawski, S. Geyer, N. Weiskopf, and N. Scherf, "Combining deep learning and active contours opens the way to robust, automated analysis of 3d brain cytoarchitectonics," 2018.

[143] D. M. Pelt and J. A. Sethian, "A mixed-scale dense convolutional neural network for image analysis," *Proceedings of the National Academy of Sciences*, vol. 115, no. 2, pp. 254–259, 2018.

[144] J. A. Bogovic, J. L. Prince, and P.-L. Bazin, "A multiple object geometric deformable model for image segmentation," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 145–157, 2013.

[145] O. Hilsenbeck, M. Schwarzfischer, D. Loeffler, S. Dimopoulos, S. Hastreiter, C. Marr, F. J. Theis, and T. Schroeder, "faster: a user-friendly tool for ultrafast and robust cell segmentation in large-scale microscopy," *Bioinformatics*, vol. 33, no. 13, pp. 2020–2028, 2017.

[146] S. U. Akram, J. Kannala, L. Eklund, and J. Heikkilä, "Cell segmentation proposal network for microscopy image analysis," in *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 21–29.

[147] S. Pelengaris, S. Abouna, L. Cheung, V. Ifandi, S. Zervou, and M. Khan, "Brief inactivation of c-myc is not sufficient for sustained regression of c-myc-induced tumours of pancreatic islets and skin epidermis," *BMC biology*, vol. 2, no. 1, p. 26, 2004.

[148] I. Arganda-Carreras, V. Kaynig, C. Rueden, K. W. Eliceiri, J. Schindelin, A. Cardona, and H. Sebastian Seung, "Trainable weka segmentation: a machine learning tool for microscopy pixel classification," *Bioinformatics*, vol. 33, no. 15, pp. 2424–2426, 2017.

[149] C. Sommer, C. Straehle, U. Koethe, and F. A. Hamprecht, "Ilastik: Interactive learning and segmentation toolkit," in *2011 IEEE international symposium on biomedical imaging: From nano to macro*. IEEE, 2011, pp. 230–233.

[150] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[151] M. Veta, P. J. Van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. Pluim, "Automatic nuclei segmentation in h&e stained breast cancer histopathology images," *PloS one*, vol. 8, no. 7, p. e70221, 2013.

[152] C. Xiao, X. Chen, W. Li, L. Li, L. Wang, Q. Xie, and H. Han, "Automatic mitochondria segmentation for em data using a 3d supervised convolutional network," *Frontiers in Neuroanatomy*, vol. 12, p. 92, 2018.

[153] A. Lucchi, Y. Li, and P. Fua, "Learning for structured prediction using approximate subgradient descent with working sets," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1987–1994.

[154] Y. Yang, D.-q. Liu, W. Huang, J. Deng, Y. Sun, Y. Zuo, and M.-m. Poo, "Selective synaptic remodeling of amygdalocortical connections associated with fear memory," *Nature neuroscience*, vol. 19, no. 10, p. 1348, 2016.

[155] M. West, L. Slomianka, and H. J. G. Gundersen, "Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator," *The Anatomical Record*, vol. 231, no. 4, pp. 482–497, 1991.

[156] D. M. De Groot, S. Hartgring, L. Van de Horst, M. Moerkens, M. Otto, M. H. Bos-Kuijpers, W. S. Kaufmann, J. H. Lammers, J. P. OCallaghan, I. D. Waalkens-Berendsen *et al.*, "2d and 3d assessment of neuropathology in rat brain after prenatal exposure to methylazoxymethanol, a model for developmental neurotoxicty," *Reproductive toxicology*, vol. 20, no. 3, pp. 417–432, 2005.

[157] D. L. Brown, "Bias in image analysis and its solution: unbiased stereology," *Journal of toxicologic pathology*, vol. 30, no. 3, pp. 183–191, 2017.

[158] P. R. Mouton, H. A. Phoulady, D. Goldgof, L. O. Hall, M. Gordon, and D. Morgan, "Unbiased estimation of cell number using the automatic optical fractionator," *Journal of chemical neuroanatomy*, vol. 80, pp. A1–A8, 2017.

[159] A. G. Valdecasas, D. Marshall, J. M. Becerra, and J. Terrero, "On the extended depth of focus algorithms for bright field microscopy," *Micron*, vol. 32, no. 6, pp. 559–569, 2001.

[160] S. S. Alahmari, D. Goldgof, L. Hall, H. A. Phoulady, R. H. Patel, and P. R. Mouton, "Automated cell counts on tissue sections by deep learning and unbiased stereology," *Journal of chemical neuroanatomy*, vol. 96, pp. 94–101, 2019.

[161] S. Alahmari, D. Goldgof, L. Hall, P. Dave, H. A. Phoulady, and P. Mouton, "Iterative deep learning based unbiased stereology with human-in-the-loop," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 665–670.

[162] S. S. Alahmari, D. Goldgof, L. O. Hall, and P. R. Mouton, "Automatic cell counting using active deep learning and unbiased stereology," in

2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, 2019, pp. 1708–1713.

[163] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, and K. Q. Weinberger, "Snapshot ensembles: Train 1, get m for free," *arXiv preprint arXiv:1704.00109*, 2017.

[164] Y. Gal, R. Islam, and Z. Ghahramani, "Deep bayesian active learning with image data," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 1183–1192.

[165] Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang, "Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7340–7351.

[166] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[167] Z. Zeng, Y. Xulei, Y. Qiyun, Y. Meng, and Z. Le, "Sese-net: Self-supervised deep learning for segmentation," *Pattern Recognition Letters*, vol. 128, pp. 23–29, 2019.

[168] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Advances in neural information processing systems*, 2017, pp. 3856–3866.