A Disector-based Framework for the Automatic Optical Fractionator

Palak Dave[a,*], Dmitry Goldgof[a], Lawrence O. Hall[a], Yaroslav Kolinko[b], Kurtis Allen[a], Saeed Alahmari[c], Peter R. Mouton[a,d]

[a]*Department of Computer Science and Engineering, University of South Florida, Tampa, Florida, 33620, USA*
[b]*Department of Histology & Embryology and Biomedical Center, Faculty of Medicine in Pilsen, Charles University, Pilsen , Czech Republic*
[c]*Department of Computer Science, Najran University, Najran, 66462, KSA*
[d]*SRC Biosciences, Tampa, Florida 33606, USA*

Abstract

Stereology-based methods provide the current state-of-the-art approaches for accurate quantification of numbers and other morphometric parameters of biological objects in stained tissue sections. The advent of artificial intelligence (AI)-based deep learning (DL) offers the possibility of improving throughput by automating the collection of stereology data. We have recently shown that DL can effectively achieve comparable accuracy to manual stereology but with higher repeatability, improved throughput, and less variation due to human factors by quantifying the total number of immunostained cells at their maximal profile of focus in extended depth of field (EDF) images. In the first of two novel contributions in this work, we propose a semi-automatic approach using a handcrafted Adaptive Segmentation Algorithm (ASA) to automatically generate ground truth on EDF images for training our deep learning (DL) models to automatically count cells using unbiased stereology methods. This update increases the amount of training data, thereby improving the accuracy and efficiency of automatic cell counting methods, without a requirement for extra expert time. The second contribution of this work is a Multi-channel Input and Multi-channel Output (MIMO) method using a U-Net deep learning ar-

*Corresponding author
Email address: palakdave@usf.edu (Palak Dave)

chitecture for automatic cell counting in a stack of z-axis images (also known as disector stacks). This DL-based digital automation of the ordinary optical fractionator ensures accurate counts through spatial separation of stained cells in the z-plane, thereby avoiding false negatives from overlapping cells in EDF images without the shortcomings of 3D and recurrent DL models. The contribution overcomes the issue of under-counting errors with EDF images due to overlapping cells in the z-plane (masking). We demonstrate the practical applications of these advances with automatic disector-based estimates of the total number of NeuN-immunostained neurons in a mouse neocortex. In summary, this work provides the first demonstration of automatic estimation of a total cell number in tissue sections using a combination of deep learning and the disector-based optical fractionator method.

Keywords: Unbiased Stereology, Automatic Optical Fractionator, Disector Stacks, Cell Counting, Overlapping Cell Segmentation, U-Net, Microscopy Image Stack

---

1. Introduction and Related Work

Accurate quantification of stained cells in tissue sections is a critical component of basic and clinical research for a wide range of scientific and medical disciplines, including experimental research in psychology, neuroscience, pathology and oncology among others. Using artificial intelligence (AI)-based approaches to analyze stained tissue sections, our group has focused on combining automatic deep learning (DL) and unbiased stereology for quantification of stained cells and other biological structures in tissue sections. For example, we have shown DL with a convolutional neural network (CNN) such as U-Net can estimate the total number of immunostained cells in extended depth of field (EDF) images with accuracy comparable to manual counts but with higher throughput, repeatability and reduced error from human factors [1, 2, 3, 4, 5]. An EDF image refers to a synthetic 2D construct that shows stained cells of interest at their maximal plane of focus through a z-axis stack of digital images (disector

stack) [6, 7, 8, 9]. These focus-stacked EDF images allow for comparisons of AI and non-AI based approaches for automatic segmentation by different methods [1, 2, 3, 10, 11, 4, 12]. Analyses of EDF images from the same disector stacks allow for comparisons of accuracy, repeatability and efficiency of cell counts using different applications of the optical fractionator method ([13, 14, 15]). However, a limitation of cell counts in disector stacks converted to EDF images is undercounting due to the possibility of overlapping cells in the z-plane (masking) as schematically illustrated in Figure 1.

Here we propose a disector-based DL approach that avoids stereological bias associated with counts of 3D objects (cells) based on their 2D profiles, i.e., the Corpuscle problem, as well as undercounting on 2D EDF images due to masking [10]. This approach uses the optical fractionator method, a combination of the optical disector and fractionator sampling scheme [13], facilitated by a DL model for automatic segmentation and counting of cells on a stained tissue section by thin focal plane optical scanning through a systematic-random series of z-axis tissue volumes (disector stacks) that represent a known fraction of the total reference volume. Thus, this contribution includes 1) a disector-based enhancement of our previous EDF-based DL approach [10]; and 2) an automatic extension of the ordinary optical fractionator method [13].

A range of previous approaches have been proposed for segmentation of overlapping cells due to masking in EDF images of various datasets. In overlapping cervical cell segmentation work presented by Lee and Kim [16], cell nuclei detected as a first step act as an indicator of presence of a cell in overlapping cells in EDF images. A similar technique was used to segment cells in genome-wide RNAi screening images [17] where cell nuclei are usually not overlapping and have good contrast due to dark color, which facilitates nuclei detection in the first step. However, that study used images of multichannel fluorescent probes that allow for interference (subtraction) imaging of co-localized proteins in specific cell compartments, which is not possible with general DAB-based immunostaining used in the present work.

A widely used method for overlapping cell segmentation in absence of flu-

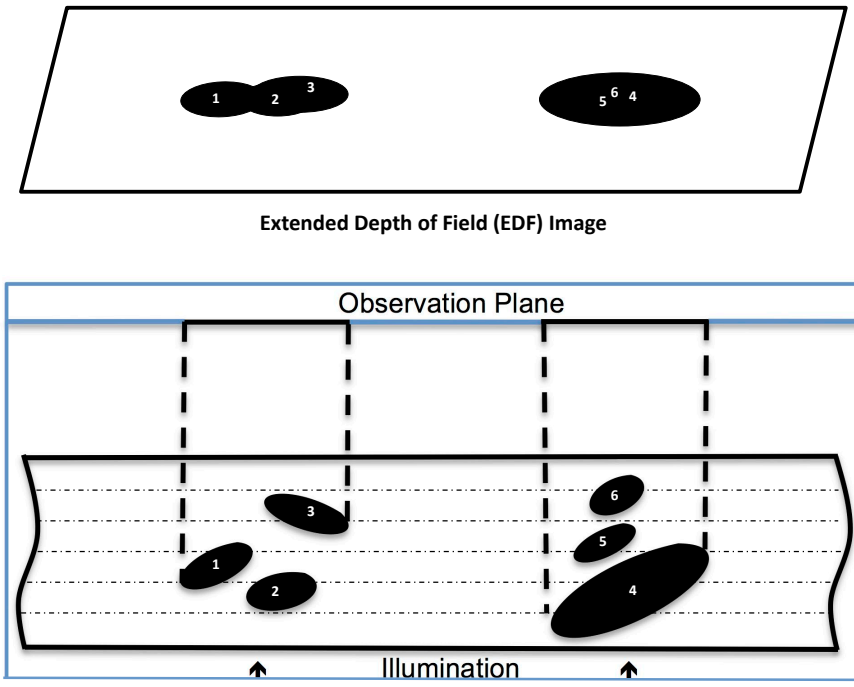

Figure 1: Schematic showing overlapping objects (masking) along a single viewing axis caused by objects clustered together (1,2,3) and a larger object (4) masking the observation of smaller objects (5,6). As shown in the top panel, the clustered and masked objects are not separable in the Extended Depth of Field (EDF) image. Counting along the multiple planes parallel to the observation plane may resolve false negative counts along a single viewing axis. Such multiple parallel planes are represented by horizontal dotted lines (a through e). Using single optical plane scanning along the viewing axis, all cells may be accurately counted at the best plane of focus; or when they first come into focus, i.e. (1,d), (2,d), (3,b), (4,e), (5,c), (6,a). Notably, the number of cells counted in this manner converges on the true or expected value, thereby avoiding under-counts due to cell clustering and masking.

orescent co-localization is the Distance Transform (DT) followed by watershed segmentation [18, 19, 17, 1, 2, 3, 5, 4]. This technique works for touching or overlapping cells with a small amount of overlap (i.e., 'bottle-neck' shaped connections). However, the DT fails to provide extrema/seeds for the watershed transform in case of highly overlapping cells.

Abdolhoseini *et al.* [20] proposed an approach to fuse image intensity with the DT followed by watershed segmentation to segment overlapping cells. The design of this fusion approach is based on darker pixel intensity at the center of the cell and lighter intensities near the edges (i.e., the overlapping area). However, this is not always true for overlapping cells, particularly cells with high overlap. Furthermore, this method cannot be used for datasets with non-uniform intensity within a cell, such as the dataset in the present work.

Lin *et al.* [21] proposed a gradient-fused DT followed by watershed to segment fluorescent-labeled cell nuclei in 3D confocal microscope image stacks. The key idea behind the use of gradient information is that there is a higher inter-cell intensity gradient than intra-cell intensity gradient. The challenge with applying this method to focus-stacked 2D (EDF) images of our dataset is that there is little-to-no inter-cell gradient among highly overlapping cells in most instances.

Lugagne *et al.* [22] use the focal signature of individual pixels along the z-axis for cell identification in z-stacks of bright-field microscopy images. Pixels are labeled as one of multiple classes like cell contour, cell interior, halo, etc. A machine learning model [support vector machine (SVM)] is trained with individual pixel focal signature to classify the signal as one of the classes based on the idea that the pixels in each of the classes have distinct intensity profiles along the z-axis. Though versatile, one limitation of this method is that a pixel can belong to only one of the classes, i.e., a pixel belonging to one cell interior cannot belong to another cell interior or cell contour. All the pixels of multiple overlapping cells can be classified as 'cell interior' forming a big blob resulting in under-counting. Also, this method cannot resolve a small cell completely occluded by a larger cell above in the z-axis (masking).

In our case, microscopy image stacks consist of multiple images (n=10) in

each z-stack (disector stack) where each image shows cells at a different focal plane in the z-axis. Different cells can appear in best focus at the same or different focal planes depending on their physical location in the z-stack. Thus, segmenting/counting cells in their singularly best focus plane overcomes masking bias that leads to undercounts in EDF images. Furthermore, cells that highly overlap in the x-y plane can be separated in the z-direction provided 3D context is available to identify the best focus plane for each cell. The 3D cell counting approach proposed here offers numerous advantages over existing DL-based methods, including the need for less training data, which is often a limiting factor in biomedical applications; and lower computational costs in terms of memory requirement and training time, as compared with existing methods with 3D context processing such as 3D convolutional networks and Recurrent Neural Networks (RNN) [23].

The discussion above outlines the need for a new approach for automatic DL-based cell counting in tissue volumes that combines 3D context with the low computational costs and minimal training data requirements. To achieve these goals, we propose a disector-based Multi-channel Input and Multi-channel Output (MIMO) framework that represents a digital realization of the ordinary optical disector method where an expert carries out manual thin focal-plane scanning of z-axis volumes while counting cells of interest at a unique point, e.g., the best plane of focus for every cell in each disector stack. In both approaches, the total number of cells in the reference volume is estimated in an unbiased manner using the fractionator method [15], i.e., the product of the reciprocal of all sampling fractions and the sum of cells counted in all disector volumes.

2. Methods

2.1. ASATP-DL: An update in ASA-DL framework

One of the state-of-the-art methods for automatic profile counting in EDF images of single-immunostained microscopy image stacks is the Adaptive Segmentation Algorithm-Deep Learning (ASA-DL) combination presented by Alah-

mari *et al.* [3]. The same framework has been demonstrated for automatic counts of EDF images by Dave *et al.* [4] after stain-separation of counter-stained tissue sections (NeuN with cresyl violet counterstain). In the ASA-DL framework, the Ground Truth (GT) is divided into two steps as follows. In the first step (count-annotation) an expert identifies and counts (clicks) on the immunostained cells (neurons) of interest to create GT. In the second step (mask-annotation) creates binary image masks (with cells of interest in foreground) for training a deep learning model to automatically segment unseen cells. The mask-annotation, which does not require an expert, is done via manual verification of segmentation mask generated automatically by the handcrafted algorithm ASA [10, 11]. During this verification, a human observer reviews an ASA segmentation mask superimposed on the count-annotation and accepts or rejects the image based on the quality of the segmentation. For example, a false positive (FP) could be the basis for the human observer to reject an image in the ASA segmentation. On completion of this review process, the accepted images are used for training the DL model.

Here we propose a novel update to this verification process in which an ASA-segmented blob (cell) can be identified as an FP automatically if there is no matching annotation (click) in the count-annotation. In this case, FPs are automatically erased from the ASA segmentation masks prior to presentation of the image to the human observer for verification. This automatic FP removal from ASA mask requires no expert time and allows more accepted data for training the DL model. The framework with the update is referred as 'ASATP-DL' (where only True Positives (TP) of the ASA segmentation (ASA-TP) are verified). Figure 2 depicts the process for generating the mask-annotation for the training set in ASA-DL and ASATP-DL. An example of ASA segmentation before and after automatic FP removal is shown in Figure 3. The green crosses ($\times$) indicate count-annotation, the contours represent segmentation by ASA while green and red contours are TPs and FPs, respectively.

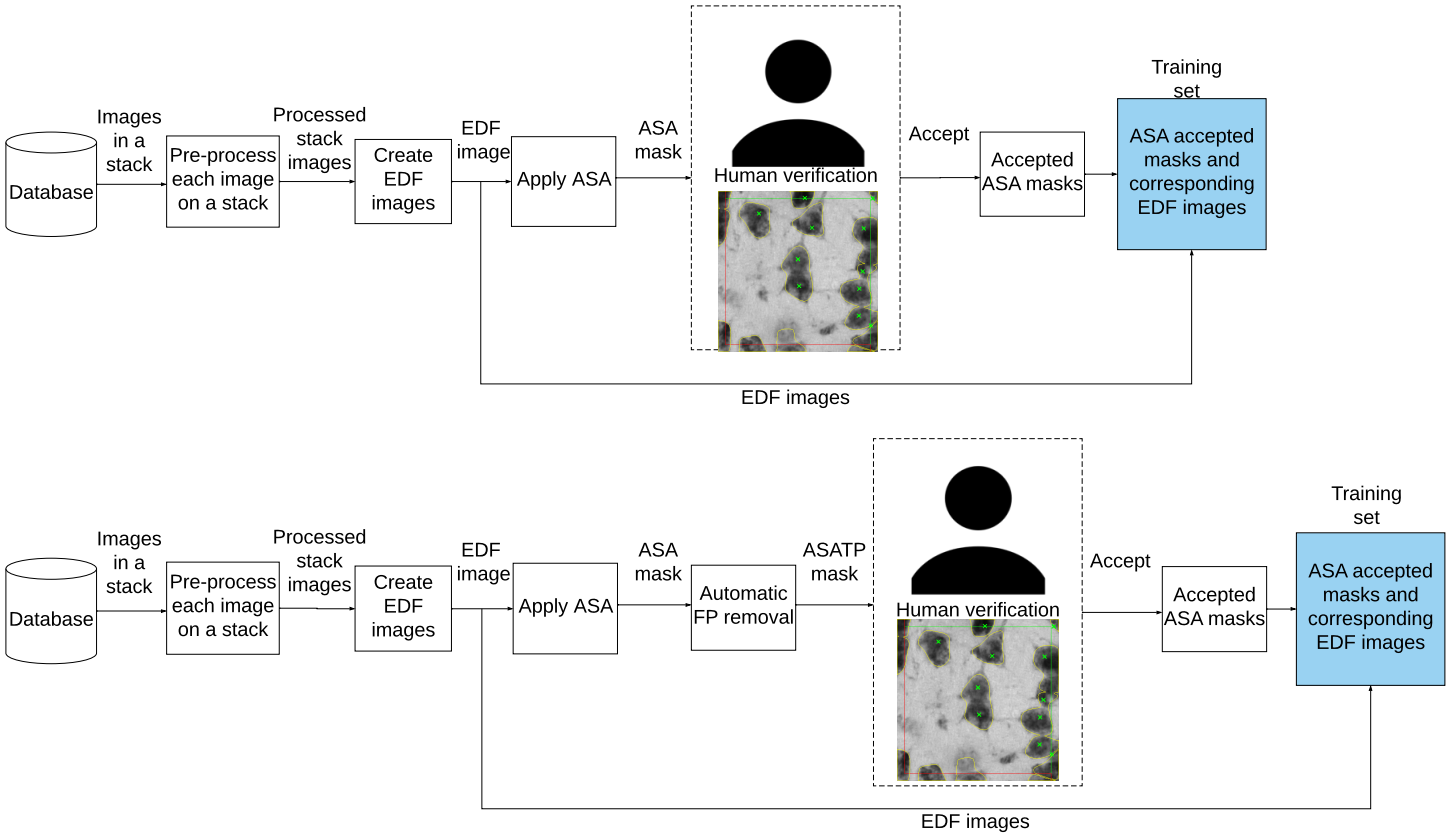

Figure 2: Top: the process of generating the mask-annotation in ASA-DL framework. Bottom: the same process in the ASATP-DL framework where the FPs are automatically removed from the ASA segmentation mask before manual verification.

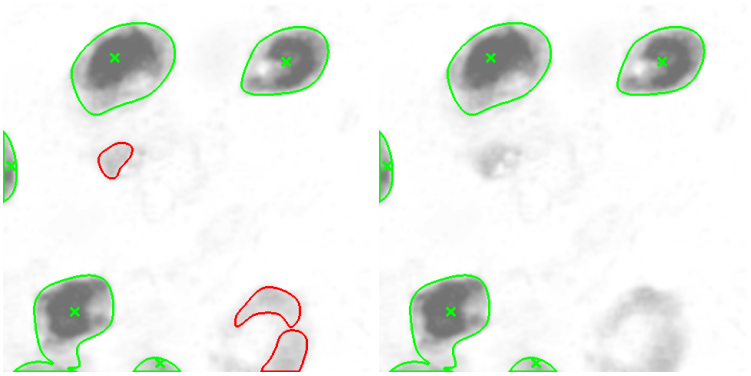

Figure 3: Left: ASA segmentation mask verified (and rejected) in the ASA-DL approach. Right: ASA segmentation mask verified (and accepted) after FP removal in the proposed ASATP-DL approach. The green cross marks indicate count-annotation and the contours represent segmentation by ASA. Green contours are TPs and red contours are FPs.

2.2. Disector-based MIMO Framework

An overview of the proposed disector-based MIMO framework for cell counts is shown in Figure 4. In the first step (far left), the color image stacks are converted to grayscale using stain separation [4]. Next, the grayscale image stack is fed as input to a trained MIMO U-Net model for automatic segmentation/ counting of cells. The prediction map stack is then post-processed to avoid over-segmentation of a cell and remove some of the FPs. Finally, Gundersen's unbiased counting rules are applied based on inclusion and exclusion planes in 3D [24]. Each of these steps are explained in detail in the following subsections.

2.2.1. Disector-based MIMO U-Net

A U-Net [25] deep learning architecture was used in the present work based on its previous performance segmenting biomedical images.

As discussed in the Section 1, we propose to use a 2D U-Net model [25] with MIMO to exploit the 3D context in an image stack for identifying the best focus plane for each cell and to avoid the high computational cost and training data requirement for known 3D methods, e.g., RNNs. To segment a

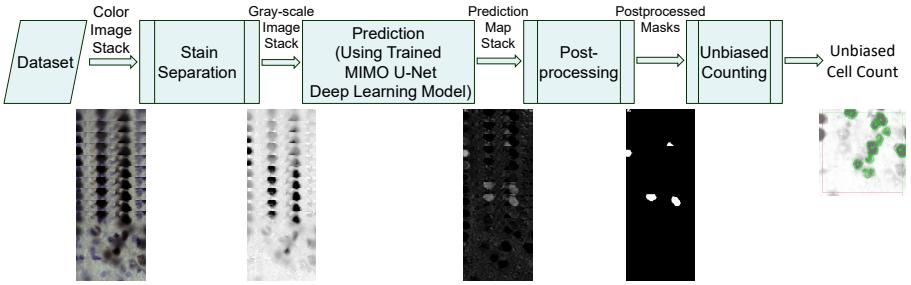

Figure 4: An overview of the proposed disector-based MIMO framework

cell in the channel (optical plane) with its best (sharpest) focus, the number of input and output channels are the same as the number of planes (z-axis planes) in a stack (here 10 channels). Each image plane in the z-stack is treated as an input channel and each output channel is foreground of the corresponding image plane in the stack where foreground for an input channel contains the cells in best focus. In other words, a cell that appears in several z-axis planes is segmented only once at a unique point, i.e., its sharpest focal plane.

2.2.2. Color to Gray Conversion

In order to use each image plane of a stack as an input channel, the plane has to be a single channel (gray-scale) image. Since visible light microscopy images are typically three channel [R (red), G (green), B (blue)] color images, it is necessary to transform the color images to gray scale. The conventional BT.601-7 (studio encoding parameters for digital television) Equation 1 [26] can be used for color to gray conversion of microscopy images of cells in tissue sections immunostained with a single color, e.g., neurons immunostained with a nuclear (NeuN) with colorization by the diaminobenzidine (DAB) reaction forming a brown product. In Equation 1, R, G, and B are the intensity values of the red, green, and blue channels respectively, and Y is the gray scale intensity value. A common option in histological staining is to combine a primary immunostain or histochemical stain with a counterstain to enhance definition of cellular

structures. Examples of common dual stain combinations are hematoxylin and eosin (H&E) and a DAB-based immunostain, e.g., NeuN, with a Nissl stain such as cresyl violet (DAB-CV). For the present study, NeuN-immunopositive nuclei appear brown while Nissl substances stained blue violet with the CV counter-stain are not of interest. Importantly, Equation 1 would convert both the brown and blue objects in the dual stain images to gray, which is not desirable. Stain separation is done on each image to remove the CV counter-stain and thereby generate a single channel grayscale image corresponding to the primary (brown) stain. This approach uses a digital stain separation method proposed by Dave *et al.* [4] with two modifications as described below. It should be noted that a constant (fixed) stain color basis is required across the images of a stack to maintain the focus profile of a cell as expected after stain separation. That is, the focus profile of a cell in a stack is expected to monotonically increase until the best focus plane then monotonically decrease thereafter.

$$Y = 0.2989 * R + 0.5870 * G + 0.1140 * B \tag{1}$$

The sparse solution for stain separation used by Dave *et al.* [4] reduces the background noise. However, stain separation with such a sparse solution can result in 'holes' (zero valued pixels) within cells in some of the planes in an image stack. Such 'holes' are not a problem while working with EDF images since they are present only in some of the planes and are filled with the contents from the other planes of the stack by the EDF algorithm. On the other hand, the high sparsity can be a problem while using each plane as an input channel, especially because the aim is to segment each cell in its best focus plane. Hence, the sparsity coefficient is made zero (no sparsity) for the disector-based approach presented here.

As mentioned earlier, the CV counterstain improves the contrast definition of the objects of interest stained with the brown color of the primary immunostains. Hence, if both stains are present at a pixel, it is imperative to combine both stains while transforming the color image to gray scale. In the first step, the

stain density maps of the primary stain (brown) and counterstain (blue) are obtained through stain separation. Second, at all the pixels where primary stain density is above a threshold (here, 5% of the maximum possible value), the stain density values of the two stains are summed to a single grayscale value, effectively removing the counter-stain while preserving the intensity perceived by the human for the object of interest.

2.2.3. Prediction Maps Post-processing

The threshold value for the prediction confidence maps is first selected based on the smallest absolute error rate on the validation data. Next, the blobs smaller than a minimum area threshold are discarded for noise removal. A minimum area threshold of 500 pixels is used for the blobs inside the prediction map image and 200 pixels is used for the blobs which are touching any of the edge of the prediction map image considering the fact that it can correspond to a cell with partial cell body outside the field of view (disector box).

As discussed earlier, there are ten output channels (same as number of input channels) and each cell is expected to be segmented in the corresponding best focus plane. It is interesting to note that one plane before or after the best focus slice can have a prediction confidence value lower than the best focus slice and remain above the prediction confidence threshold. Based on the fact that the planes in the dataset used in this study are 1 µm apart and the neurons are 3D volumetric structures, with a spread in the z-axis as well (generally larger than 2 µm spread in the z-axis), there is a high probability that any two overlapping blobs in any two consecutive planes belong to a single cell. Hence, such overlapping blobs are combined and assigned to the plane with the bigger blob among the two participating blobs.

Since the blobs at this stage might have holes, any holes in all the blobs are filled as a cell body is expected to be a solid. Finally, any blob with circularity less than a circularity threshold (here, 30%) are filtered out since NeuN-stained nuclei are roughly spherical concave structures. In post-processing we used circularity-based filtering to help discard some of the FPs. To assess the extent

of systematic error (bias) related to the 3D shape of the objects, we computed the results with and without the circularity-based filtering. The average error rates (Table 3) are 4.82% and 5.93% with and without the shape assumption, respectively, for a difference of 1.11%. Thus, circularity-based filtering based on an assumption about the object shape only slightly affects the results and is not an absolute requirement for the proposed method.

3. Collection of Image Datasets

The image dataset for these experiments (Section 4) consisted of images from tissue sections processed for NeuN immunostain and cresyl violet counterstain (NeuN-CV) from five mice, which represent a subset of the NeuN counter-stained dataset from our prior published work [4].

Animals for this study were aged 6–8 months Tg4510 male mice (n=3) and male non-tg littermate controls (n=2) sampled at random from the colony at the Byrd Alzheimer's Disease Center at the University of South Florida in Tampa, FL. Under deeply anesthesia on an isothermal pad, mice were perfused with 25 ml of cold sterile buffered saline, their brains removed and one hemisphere immersion fixed for 24 hours in fresh 4% phosphate buffered paraformaldehyde then transferred to Dulbecco's phosphate buffered saline, cryoprotected in 10, 20 and 30% sucrose and stored at $4°C$. 50 µm frozen sections were collected with a sliding microtome, with every 6^{th} section sampled in a systematic-random manner to obtain 8-12 sections through neocortex of each brain. The number of coronal sections analyzed varied due to difference in the length of the brain along the rostro-caudal axis. As shown by Gundersen *et al.* in [27], this range of sections ($\sim 8-12$) is sufficient to capture the between-section variation for any size or shape of reference space. Average thickness of the samples after post-processing was ∼18.5 µm.

Staining protocols were conventional immunostaining with primary and secondary antibodies and DAB-based colorization followed by standard counterstaining in a 2% cresyl violet (CV) solution, as we have previously detailed [28].

On the day of immunostaining, sections were moved into a multi-sample staining tray for blocking of endogenous peroxidase 10% methanol, 3% H202 in PBS; 30 min. Tissue samples were permeabilized with 0.2% lysine, 1% Triton X-100 in PBS solution and incubated overnight in anti-NeuN primary antibody (Millipore). After washing in PBS, sections were incubated in biotinylated secondary antibody (Vector Laboratories, Burlingame, CA), washed after 2 hours and incubated with Vectastain® Elite® ABC kit (Vector Laboratories, Burlingame, CA) for enzyme conjugation. In the final step, sections were colorized in 0.05% diaminobenzidine in 0.03% H202 with nickel intensification, mounted onto slides, dehydrated, and cover slipped.

Disector stacks were collected in a known systematic-random fraction of the reference space similar to manual analysis of disector volumes with the ordinary optical fractionator method for unbiased estimates of the total number of stained cells in an anatomically defined reference space [13, 29, 30, 31]. The work was done with assistance from a computerized stereology system (Stereologer®, SRC Biosciences, Tampa, Florida, USA) consisting of an Olympus microscope equipped with automatic XY stepping motors and Z-axis focus motor for automatic collection of z-axis image stacks (disector stacks). Starting at a random X-Y location in the upper left quadrant of the reference space (neocortex) in the first section, images were collected in a systematic-random sampling manner at intervals of 1200 µm × 1200 µm, and repeating on 8–12 sections sampled in a systematic-random manner through the entire mouse neocortex. Each disector stack consists of a set of ten images (1 µm apart) in the z-axis captured using 100x oil lens (NA=1.4) with a guard volume of at least 2 µm. This level of sampling stringency achieved a low sampling error as evidenced by coefficient of error ~ 0.10 (CE $\sim 10\%$) as shown by [27]. The time required for unsupervised (automatic) collection of images for each case (46 to 66 disector stacks) was $\sim$ 30 to 45 minutes. The dataset used is summarized in Table 1.

The 'Ground Truth (GT)' for a mouse brain in this work is the sum of the cells counted in all disector stacks sampled through an anatomically defined reference space (mouse neocortex) in a systematic-random manner as we have

previously reported [1, 2, 3, 4, 5, 10, 11, 12].

Similar to the ASATP-DL approach, the GT is divided into two levels in the present work. The first is to identify and count the true number of cells (NeuN neurons) in the sampled disector stacks, also referred to as 'count-annotation' where expert knowledge was used to distinguish between neurons and various glia cells. Similar to an optical disector, an expert (YK) used thin focal plane scanning of a disector stack to count NeuN neurons only once at the best focus plane. NeuN-immunostained cells in disector stacks were counted (clicked) using Gundersen's unbiased counting rules to avoid edge effects [24]. As shown in Figure 5, cells were included that fell within the disector volume or touched one of the top, upper or right-side inclusion planes in 3D. Cells were excluded that fell outside the disector volume or touched either the bottom, lower or left-side exclusion planes.

A method of Köhler illumination [32] was used with a $\frac{1}{3}$ open condenser diaphragm for data collection. The stacks which did not conform to this rule and as a result had unclear boundaries/edges of cells were considered as low quality on the recommendations of a domain expert (YK). About 22% of the total stacks were discarded due to poor quality. Number of stacks discarded per case is listed in the data summary in Table 1. For the mice with a significantly higher number of good quality stacks, a number of stacks were discarded in a systematic-random manner in order to balance the dataset, i.e., every 3^{rd} stack with a random start at stack 1, 2, or 3 was discarded from the 98 stacks from case PI3-20 to obtain 66 stacks. Also, the annotation on the selected data (subset) was done by one expert (YK) during this work to achieve uniform GT across all stacks.

Next, binary image masks (with cells as foreground in corresponding best focal plane) were generated to train the deep learning model ('mask-annotation') to segment NeuN neurons. After the count-annotation, no further expert knowledge is required to draw masks (boundaries) around identified neurons, allowing mask-annotation by trained student researchers to minimize the requirement for expert time. As for count-annotation, a boundary around a cell was drawn in its

Mouse ID	# Stacks Used	Ground Truth Count	# Stacks Discarded
PI3-18	65	312	3
PI3-19	46	274	57
PI3-20	66	394	32*
PI3-21	62	580	35
PI3-22	60	357	2

Table 1: Summary of neuron (NeuN) counterstain dataset used and the number of stacks discarded per case due to poor illumination setting during data collection. * Good quality stacks discarded to balance the dataset.

best focal plane. A check is performed on every stack to match the cell count in both annotations to ensure that no human error is introduced while going from count-annotation to mask-annotation. The required tools for these annotations were developed in-house during this work.

4. Experimental Results

The number of accepted images per mouse after manual verification of ASA and ASA-TP masks is listed in Table 2. The number of images in Table 2 represents the number of stacks since each stack is converted into an EDF image. It should be noted that the small number of accepted images after ASA mask verification is because ASA was originally designed for a single immunostained dataset. The number of accepted images are significantly larger, and hence provide a larger training set for subsequent training of a DL model, by automatic FP removal discussed in Section 2.1. Furthermore, the number of accepted images is sufficient to train the DL model to obtain an average error rate of 6.05% (Table 3) using the ASATP-DL approach.

A five-fold experiment using leave-one-mouse-out for testing strategy was performed for both EDF-based ASATP-DL and the disector-based MIMO approach. In each fold, data from one mouse was left out as a test-set for the trained model performance evaluation on unseen data (i.e., not seen by the

Mouse ID	# Images	# ASA Accepted	# ASA-TP Accepted
PI3-18	65	12	32
PI3-19	46	5	26
PI3-20	66	2	29
PI3-21	62	2	17
PI3-22	60	6	24

Table 2: Number of images accepted per mouse after ASA vs ASA-TP verification.

model during training) since the model is expected to perform well on the training data. From each of the other four mice, 80% and 20% of the data contributed to the train set and validation set, respectively. This regimen is repeated five times leaving a different mouse data out for testing each time. In effect, the approach is validated on each mouse to ensure the fairness of the performance evaluation.

As for the ASA-DL framework, elastic and rotation augmentations were performed on the train-set to obtain a large and diverse dataset for training a deep learning model. Training the deep learning model is an automatic process that primarily requires unsupervised computer time, i.e., no expert time is required once the training is initiated. The actual training time depends on multiple factors such as amount of training data, hardware capacity, training parameters etc. In the present work, approximately 4 hrs and 12 hrs allowed for one-fold training in the ASATP-DL and MIMO approaches in that order. Furthermore, the training time is a one-time investment per stain and cell type in that less than 15 minutes is needed to count cells in a complete reference space ($\sim$ 60 to 90 disector stacks) using a trained DL model. The predicted cell counts by both EDF-based ASATP-DL and disector-based MIMO approaches are reported in Table 3.

The motivation behind this work was to develop an automatic disector-based method that avoids under-counting in previous approaches due to over-

Test Mouse ID	#Stacks	Ground Truth Count (G)	ASATP-DL Count (A)	ASATP-DL Error Rate(%) ($\frac{\lvert G-A\rvert*100}{G}$)	MIMO Count (L)	MIMO Error Rate(%) ($\frac{\lvert G-L\rvert*100}{G}$)
PI3-18	65	312	295	5.45	319	2.24
PI3-19	46	274	290	5.84	275	0.36
PI3-20	66	394	413	4.82	426	8.12
PI3-21	62	580	558	3.79	596	2.76
PI3-22	60	357	394	10.36	319	10.64
Average Error Rate (%)				6.05		**4.82**

Table 3: Automatic counting results using EDF-based ASATP-DL and disector-based MIMO approach. Where, the fifth and last columns refer to % error; and, the last row refers to the mean % error for all mice analyzed.

lapping/occluded cells in the z axis of EDF images. The under-counting error is expected to be relatively high in the stacks with higher cell density. Also, while improving on the under-counting issue in high cell density stacks, the new method should perform at least as well as the existing EDF-based method on stacks with low cell density. To test this hypothesis, stacks were divided into low and high cell density groups (based on manual stereology counts) then the performance of the disector-based MIMO approach was assessed on high and low cell density groups independently. The low cell density group consists of stacks with ≤ 10 cells per stack and high cell density group with >10 cells per stack. Notably, the stacks are divided into two groups only for analysis purposes after stereology counts at the case level where each case contains areas of both low- and high-density stacks. No extra experiments were performed on the two groups independently. The cell count result comparison for both groups is provided in Tables 4 and 5. Under-counting and over-counting error is indicated by negative and positive error rates, respectively.

As indicated in Table 4, the proposed method has significantly lower absolute error on the low cell density stacks. Table 5 shows that the proposed MIMO method has significantly lower under-counting as compared to the EDF-based ASATP-DL approach on the high cell density stacks as expected. Further-

Test Mouse ID	#Stacks	Ground Truth Count (G)	ASATP-DL Count (A)	ASATP-DL Error Rate(%) ($\frac{(A-G)*100}{G}$)	MIMO Count (L)	MIMO Error Rate(%) ($\frac{(L-G)*100}{G}$)
PI3-18	57	202	208	2.97	205	1.49
PI3-19	37	152	187	23.03	159	4.61
PI3-20	51	179	227	26.82	193	7.82
PI3-21	35	225	238	5.78	232	3.11
PI3-22	50	202	260	28.71	180	-10.89
Average Absolute Error Rate (%)				17.46		**5.58**

Table 4: Automatic counting results using EDF-based ASATP-DL and proposed disector-based MIMO method for low cell density stacks with <=10 cells per stack where under-counting and over-counting error are indicated by negative and positive error rate, respectively; and, the last row refers to the mean % absolute error for all mice analyzed.

Test Mouse ID	#Stacks	Ground Truth Count (G)	ASATP-DL Count (A)	ASATP-DL Error Rate(%) ($\frac{(A-G)*100}{G}$)	MIMO Count (L)	MIMO Error Rate(%) ($\frac{(L-G)*100}{G}$)
PI3-18	8	110	87	-20.91	114	3.64
PI3-19	9	122	103	-15.57	116	-4.92
PI3-20	15	215	186	-13.49	233	8.37
PI3-21	27	355	320	-9.86	364	2.54
PI3-22	10	155	134	-13.55	139	-10.32
Average Absolute Error Rate (%)				14.68		**5.96**

Table 5: Automatic counting results using EDF-based ASATP-DL and proposed disector-based MIMO method for high cell density stacks with >10 cells per stack where under-counting and over-counting error are indicated by negative and positive error rate, respectively; and, the last row refers to the mean % absolute error for all mice analyzed.

Test Mouse ID	ASATP-DL				MIMO			
	Accuracy $(\frac{TP}{TP+FP+FN})$	Precision $(P=\frac{TP}{TP+FP})$	Recall $(R=\frac{TP}{TP+FN})$	F1-Score $(f1=\frac{2*P*R}{P+R})$	Accuracy $(\frac{TP}{TP+FP+FN})$	Precision $(P=\frac{TP}{TP+FP})$	Recall $(R=\frac{TP}{TP+FN})$	F1-Score $(f1=\frac{2*P*R}{P+R})$
PI3-18	0.61	0.78	0.74	0.76	0.69	0.81	0.83	0.82
PI3-19	0.68	0.79	0.83	0.81	0.68	0.81	0.81	0.81
PI3-20	0.56	0.70	0.73	0.72	0.66	0.76	0.82	0.79
PI3-21	0.65	0.80	0.77	0.79	0.65	0.78	0.80	0.79
PI3-22	0.58	0.70	0.77	0.74	0.63	0.82	0.73	0.77
Average value	0.62	0.75	0.77	0.76	0.66	0.79	0.80	**0.80**

Table 6: Comparison of accuracy, precision, recall, and F1-score of EDF-based ASATP-DL vs proposed MIMO method. The higher average F1-score of the proposed method represents better overall performance.

Test Mouse ID	ASATP-DL				MIMO			
	Accuracy $(\frac{TP}{TP+FP+FN})$	Precision $(P=\frac{TP}{TP+FP})$	Recall $(R=\frac{TP}{TP+FN})$	F1-Score $(f1=\frac{2*P*R}{P+R})$	Accuracy $(\frac{TP}{TP+FP+FN})$	Precision $(P=\frac{TP}{TP+FP})$	Recall $(R=\frac{TP}{TP+FN})$	F1-Score $(f1=\frac{2*P*R}{P+R})$
PI3-18	0.58	0.72	0.74	0.73	0.67	0.80	0.81	0.80
PI3-19	0.70	0.74	0.91	0.82	0.67	0.79	0.82	0.80
PI3-20	0.49	0.59	0.74	0.66	0.60	0.73	0.78	0.75
PI3-21	0.65	0.77	0.81	0.79	0.63	0.76	0.79	0.77
PI3-22	0.57	0.64	0.83	0.72	0.61	0.80	0.71	0.75
Average value	0.60	0.69	0.81	0.74	0.64	0.77	0.78	**0.78**

Table 7: Comparison of accuracy, precision, recall, and F1-score of EDF-based ASATP-DL vs proposed MIMO method for low cell density stacks with <=10 cells per stack. The higher average F1-score of the proposed method represents better overall performance.

Test Mouse ID	ASATP-DL				MIMO			
	Accuracy $(\frac{TP}{TP+FP+FN})$	Precision $(P=\frac{TP}{TP+FP})$	Recall $(R=\frac{TP}{TP+FN})$	F1-Score $(f1=\frac{2*P*R}{P+R})$	Accuracy $(\frac{TP}{TP+FP+FN})$	Precision $(P=\frac{TP}{TP+FP})$	Recall $(R=\frac{TP}{TP+FN})$	F1-Score $(f1=\frac{2*P*R}{P+R})$
PI3-18	0.70	0.93	0.74	0.82	0.74	0.83	0.86	0.85
PI3-19	0.65	0.86	0.73	0.79	0.70	0.84	0.80	0.82
PI3-20	0.64	0.84	0.73	0.78	0.70	0.79	0.86	0.83
PI3-21	0.65	0.83	0.75	0.79	0.66	0.78	0.80	0.79
PI3-22	0.61	0.81	0.70	0.75	0.65	0.83	0.75	0.79
Average value	0.65	0.86	0.73	0.79	0.69	0.82	0.82	**0.82**

Table 8: Comparison of accuracy, precision, recall, and F1-score of EDF-based ASATP-DL vs proposed MIMO method for high cell density stacks with >10 cells per stack. The higher average F1-score of the proposed method represents better overall performance.

more, the EDF-based approach has high over-counting error on the low-density stacks since some glia cells are incorrectly counted as neurons by the EDF-based method. Interestingly, in the EDF-based method the over-counting (positive) error on low-density stacks cancels the under-counting error on high-density stacks when considering all stacks of a mouse together to compute the % error rate for the mouse (as reported in Table 3). As a result, the EDF-based method for overall % error appears comparable to that of the proposed method, though the proposed method performs with higher accuracy (lower error rate) on both density groups individually. The same observation can be derived from the results reported in Tables 6 through 8. The higher F1-score for both density groups individually as well as on the mouse level indicates better overall performance of the proposed method.

A few comparison examples between the EDF-based ASATP-DL approach and proposed MIMO method are shown in Figure 5. The top row shows one of the image planes of a stack. Middle and bottom rows demonstrate results by ASATP-DL and proposed MIMO methods in order. All cross marks (green and red) indicate Ground Truth while the red cross marks show FNs (missed cells). The contours represent segmentation by ASATP-DL and MIMO methods with green and red contours showing TPs and FPs, respectively. In the result (bottom row), all cells segmented in respective best focus planes of a stack are demonstrated on an EDF image.

5. Discussion and Conclusion

There was an average of about 3-5% inter-rater variability in 'count-annotation' observed for counts of the same cells in same reference space of the same cases by multiple raters in our group ([12, 33]). This sets a lower bound on the error rate for the automatic methods compared to manual stereology counts. As can be seen in Table 3, the average error rate for the ASATP-DL is already close to the lower bound, meaning there is not much room for average error rate improvement by a new method. However, the low average error in the ASATP-DL

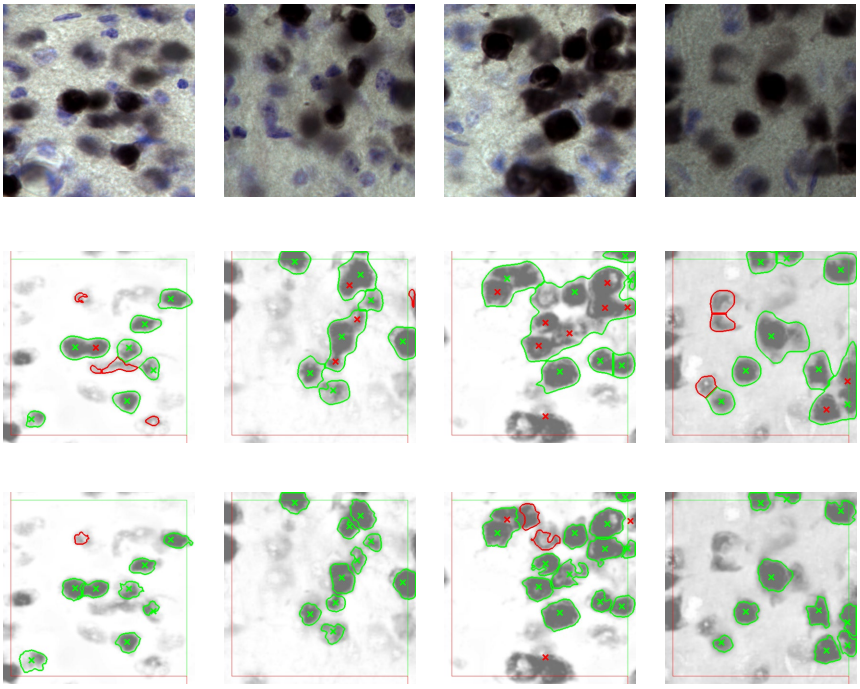

Figure 5: Visualization of results. The top row shows one of the image planes of the stack. Middle and bottom rows demonstrate results by ASATP-DL and proposed MIMO methods in order. All cross marks (green and red) indicate Ground Truth while the red cross marks show FNs (missed cells). The contours represent segmentation by the automatic methods. The green and red contours are TPs and FPs, respectively. The unbiased counting frame includes left and lower exclusion lines (red) and upper and right inclusion lines (green) (lines convert to planes in 3D).

method is due to the nuance of error cancellation between the low and high cell density stacks (as discussed in Section 4). Thus, the ASATP-DL method achieves the desired count by counting incorrect objects to compensate for the under-counting of the highly overlapping cells. In contrast, as shown in Tables 3 through 5 the proposed method reaches the desired count by primarily counting the correct objects in both low- and high-density stacks.

In the deep learning field, one of the high-value resources is time required by the domain expert (well-trained data collector). The proposed disector-based MIMO method solves the under-counting issue in the current state-of-the-art EDF-based approach with no extra expert time requirement. That is, the same count-annotation time by an expert is required for both the EDF-based and proposed disector-based MIMO methods.

We compared the time requirement for analysis of one case by the EDF-based approach and the proposed disector-based MIMO method. Both methods were run on the same computer with Intel i7-4790 CPU and 16 GB RAM. The DL model was generated on Nvidia 1080Ti GPU with 11 GB frame buffer. Using a trained DL model, the EDF-based approach required $\sim$ 13 minutes to count NeuN-immunopositive profiles compared to $\sim$ 12 minutes for the proposed disector-based MIMO method to count NeuN-immunopositive cells, from the same disector stacks for one case. The EDF-based method takes slightly higher time to compute an EDF image from each of the stack. Notably, a well-trained technician requires $\sim$ 2 to 2.5 hours of closely supervised time to count NeuN-immunopositive cells to the same level of sampling stringency with the ordinary (i.e., purely manual) optical fractionator. In contrast, the proposed MIMO method requires $\sim$ 15 minutes of supervised time to outline reference spaces at low power; and $\sim$ 42 minutes of unsupervised time to automatically collect disector stacks ($\sim$ 30 minutes) and collect data for total cell counts ($\sim$ 12 minutes). Thus, throughput efficiency in terms of supervised time is $\sim 5\times$ higher for the proposed MIMO method ($\sim$ 0.25 hour versus $\sim$ 2 to 2.5 hours). Whereas collecting accurate data by manual stereology requires constant attention, subjective decision-making, low magnification outlining of the

reference space boundaries combined with high magnification counting (clicking) by a highly trained technician with substantial experience, the DL-based MIMO method requires minimal expertise to outline the boundaries of the reference space on tissue sections at low magnification.

As in all deep learning applications, a relatively large amount of training data is required for a more accurate and generic model for different image variations. However, in many biomedical applications including the present work data for training deep learning models is a limiting factor. We anticipate that more training data will further improve the performance on the least accurate cases (e.g., PI3-20 & 22), which will further reduce the standard deviation of the error rate.

Explainability of AI ('black-box') is becoming increasingly important for medical applications. Like most deep learning algorithms, the proposed work relies on statistical model-free learning. A machine explanation for a decision made by such systems might not fit with human experts' explanation as discussed by Holzinger and Muller in [34]. A better fit can be obtained by incorporating human-in-the-loop with affective computing to receive feedback about if the machine explanation is understandable for the human expert. A case study on a microscopy image analysis with a similar approach is presented by Holzinger and Muller in [34].

The present work shows proof-of-concept for the disector-based MIMO framework by quantifying total numbers of NeuN neurons in the mouse neocortex, which had been quantified previously using manual stereology methods [35, 36]. The focus of this work is to show that the number of NeuN-stained cells counted by our automatic stereology (MIMO method) is comparable to counts of NeuN neurons by a trained human technician using manual stereology on the same disector stacks. We used modern stereology to make estimates of total neuron number based on unbiased sampling, also known as systematic sampling with a random start, in conjunction with unbiased cell counting (the optical disector principle). As shown by many studies, a systematic-random sample of 8-12 sections as used in the present study provide sufficient statistical power for unbiased

(i.e., accurate) estimates, without the need to count all cells [13, 15, 24, 27, 28].

The proposed method can be used for automatic stereology counts of any discrete objects (cells) labelled with a high signal: noise stain. As discussed in Section 2.2.3, prior knowledge about the object shape only slightly benefits performance ($< 2\%$ impact on the error rate) and it is not an absolute requirement of the proposed MIMO method.

The proposed method can be employed to count heterogeneous (different size or shape) cell populations provided that a representative dataset is available for training the deep learning model. The minimum area threshold used in the postprocessing can be tuned based on minimum cell size in the training data. Furthermore, the proposed method can be directly adapted to image stacks with different cell types by using an independent deep learning model for each cell type. The models can use the same image stacks but learn to count only one type of cell. Another approach to handle different cell types might include using each cell type as an output category (e.g. neuron or glia) and the best focus plane as the output plane. Effectively each output z-plane can have same number of output categories as the number of types of cells.

One weakness of the proposed framework is the requirement of counting cells of interest at a unique point, e.g., the best plane of focus for every cell. In manual stereology, a human expert can count a cell when it first appears (early detector), at the maximal focal plane, or at the last focal plane of visibility (late detector). It is important to select one of the three unique points and use that point across the dataset to effectively teach the process to a deep learning model. Different cells counted at different points can confuse the model. The optimal focal plane was selected as the unique point in the proposed work because the size of the cells of interest with respect to the tissue thickness is such that most of the cells are visible through out the stack (no unique start and end point). Ambiguity can arise when a cell is in a similar level of good (best) focus in two or more focal planes. Such cases are handled through post-processing to merge overlapping detections in consecutive planes to allow for only one detection per cell.

To conclude, we provide two enhancements to the current state-of-the-art applications of deep learning to unbiased stereology counts of immunostained cells in tissue sections. The first is an update in the form of a semi-automatic approach for GT preparation in 3D stacks of microscopy images (disector stacks) for generating more training data without a requirement for additional expert time. Second, we present a novel disector-based MIMO framework that represents an automatic version of the ordinary optical fractionator where all the planes of a 3D disector stack are analyzed as opposed to a single focus-stacked image (EDF image) per stack. The MIMO approach avoids the costly computations of 3D deep learning-based methods by using the 3D context of cells in disector stacks; and prevents stereological bias in the previous EDF-based method due to counting profiles rather than cells and under-counting overlapping/occluded cells. Taken together, these improvements support the view that AI-based automatic deep learning methods can accelerate the efficiency of unbiased stereology cell counts without a loss of accuracy or precision as compared to conventional manual stereology.

6. Future Work

As future work, we plan to apply the proposed method to different datasets with different staining protocols (i.e., single immunostain) to ensure the generalizability of the proposed method. We are also expanding the present approach from DAB-stained tissues to immunofluorescent staining of biological structures.

Acknowledgments

We thank Dr. Hady Ahmady Phoulady for his previous work in this project. Also, we would like to thank Raj Patel for his help with data acquisition.

Funding. This work was supported by National Science Foundation Grants (# 1513126, # 1746511, # 1926990) and a Florida High Tech Corridor Grant (# 20-10) to SRC Biosciences and the University of South Florida. Yaroslav Kolinko

was supported by the Charles University Research Fund (Progres Q39) and the Project No. CZ.02.1.01/0.0/0.0/16_019/0000787 "Fighting Infectious Diseases," awarded by the Ministry of Education, Youth and Sports of the Czech Republic, the Alzheimer Foundation Czech Republic and the Avast Foundation.

Declaration of Interests

The authors declare that this work is presented in the absence of any real or perceived conflict of interest. Authors PRM, LOH, and DG hold a U.S. patent on "Automatic stereology for determining tissue characteristics" (U.S. Patent 11,004,199 B2; awarded May 11, 2021) and related patents (US 2017/0234278 A1; awarded August 17, 2017), which may be related to products or product developments used for performing automatic stereology as described in this and several co-authored papers referenced in this work. Author PRM is owner of Intellectual Property for the Stereologer system used for performance of this work.

References

[1] S. Alahmari, D. Goldgof, L. Hall, P. Dave, H. A. Phoulady, 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.

[2] S. S. Alahmari, D. Goldgof, L. O. Hall, 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.

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

[4] P. Dave, S. Alahmari, D. Goldgof, L. O. Hall, H. Morera, P. R. Mouton, An adaptive digital stain separation method for deep learning-based automatic cell profile counts, Journal of Neuroscience Methods 354 (2021) 109102.

[5] P. Dave, D. Goldgof, L. O. Hall, S. Alahmari, P. R. Mouton, Novel stain separation method for automatic stereology of immunostained tissue sections, Innovation in Aging 3 (Supplement_1).

[6] M. S. Sigdel, M. Sigdel, S. Dinç, I. Dinc, M. L. Pusey, R. S. Aygün, Focusall: Focal stacking of microscopic images using modified harris corner response measure, IEEE/ACM Transactions on Computational Biology and Bioinformatics 13 (2) (2015) 326–340.

[7] M. L. Pusey, R. S. Aygün, Focal stacking for crystallization microscopy, in: Data Analytics for Protein Crystallization, Springer, 2017, pp. 151–176.

[8] A. P. Bradley, P. C. Bamford, A one-pass extended depth of field algorithm based on the over-complete discrete wavelet transform.

[9] B. Forster, D. Van De Ville, J. Berent, D. Sage, M. Unser, Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images, Microscopy research and technique 65 (1-2) (2004) 33–42.

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

[11] H. A. Phoulady, D. Goldgof, L. O. Hall, P. R. Mouton, Automatic ground truth for deep learning stereology of immunostained neurons and microglia in mouse neocortex, Journal of chemical neuroanatomy 98 (2019) 1–7.

[12] R. Patel, S. Alahmari, D. Goldgof, H. Phoulady, P. Dave, L. Hall, P. Mouton, Stereological analysis of neurodegeneration and neuroinflammation in tg4510 mice using manual and automatic stereology, Society for Neurosciences 558.

[13] M. West, L. Slomianka, 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 231 (4) (1991) 482–497.

[14] H. Braendgaard, H. Gundersen, The impact of recent stereological advances on quantitative studies of the nervous system, Journal of neuroscience methods 18 (1-2) (1986) 39–78.

[15] B. Pakkenberg, H. Gundersen, Total number of neurons and glial cells in human brain nuclei estimated by the disector and the fractionator, Journal of microscopy 150 (1) (1988) 1–20.

[16] H. Lee, J. Kim, Segmentation of overlapping cervical cells in microscopic images with superpixel partitioning and cell-wise contour refinement, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2016, pp. 63–69.

[17] P. Yan, X. Zhou, M. Shah, S. T. C. Wong, Automatic segmentation of high-throughput rnai fluorescent cellular images, IEEE Transactions on Information Technology in Biomedicine 12 (1) (2008) 109–117.

[18] N. Harder, M. Bodnar, R. Eils, D. L. Spector, K. Rohr, 3d segmentation and quantification of mouse embryonic stem cells in fluorescence microscopy images, in: 2011 IEEE international symposium on biomedical imaging: From nano to macro, IEEE, 2011, pp. 216–219.

[19] U. Adiga, B. L. Bell, L. Ponomareva, D. Taylor, R. Saldanha, S. Nelson, T. J. Lamkin, Mapping infected cell phenotype, IEEE Transactions on Biomedical Engineering 59 (8) (2012) 2362–2371.

[20] M. Abdolhoseini, M. G. Kluge, F. R. Walker, S. J. Johnson, Segmentation of heavily clustered nuclei from histopathological images, Scientific reports 9 (1) (2019) 1–13.

[21] G. Lin, U. Adiga, K. Olson, J. F. Guzowski, C. A. Barnes, B. Roysam, A hybrid 3d watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks, Cytometry Part A: the journal of the International Society for Analytical Cytology 56 (1) (2003) 23–36.

[22] J.-B. Lugagne, S. Jain, P. Ivanovitch, Z. Meriem, c. Vulin, C. Fracassi, G. Batt, P. Hersen, Identification of individual cells from z-stacks of bright-field microscopy images, Scientific Reports 8. doi:10.1038/s41598-018-29647-5.

[23] J. Chen, L. Yang, Y. Zhang, M. Alber, D. Z. Chen, Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation, in: D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems 29, Curran Associates, Inc., 2016, pp. 3036–3044.
URL http://papers.nips.cc/paper/6448-combining-fully-convolutional-and-recurrent-neural-networks-for-3d-biomedical-image-segmentation.pdf

[24] H. J. G. Gundersen, Notes on the estimation of the numerical density of arbitrary profiles: the edge effect, Journal of microscopy 111 (2) (1977) 219–223.

[25] O. Ronneberger, P. Fischer, 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.

[26] ITU, Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios (2011).
URL https://www.itu.int/rec/R-REC-BT.601/

[27] H. Gundersen, E. Jensen, K. Kiêu, J. Nielsen, The efficiency of systematic

sampling in stereology—reconsidered, Journal of microscopy 193 (3) (1999) 199–211.

[28] P. R. Mouton, Unbiased stereology: a concise guide, JHU Press, 2011.

[29] K. F. Manaye, J. S. Allard, S. Kalifa, A. C. Drew, G. Xu, D. K. Ingram, R. de Cabo, P. R. Mouton, 17α-estradiol attenuates neuron loss in ovariectomized dtg aβpp/ps1 mice, Journal of Alzheimer's Disease 23 (4) (2011) 629–639.

[30] K. F. Manaye, P. R. Mouton, G. Xu, A. Drew, D.-L. Lei, Y. Sharma, G. W. Rebeck, S. Turner, Age-related loss of noradrenergic neurons in the brains of triple transgenic mice, Age 35 (1) (2013) 139–147.

[31] P. Mouton, B. Kelley-Bell, D. Tweedie, E. Spangler, E. Perez, O. Carlson, R. Short, R. Decabo, J. Chang, D. Ingram, et al., The effects of age and lipopolysaccharide (lps)-mediated peripheral inflammation on numbers of central catecholaminergic neurons, Neurobiology of aging 33 (2) (2012) 423–e27.

[32] A. Köhler, New method of illumination for photomicrographical purposes, Journal of the Royal Microscopical Society 14 (1894) 261–262.

[33] P. Delgado, K. Sanchez, A. Anderson, R. Patel, S. Alahmari, D. Goldgof, L. Hall, P. Mouton, Comparison of manual, semi-automatic and fully automatic counts of immunostained neurons in mouse brains, Soc. For Neurosciences, November 8-11, 2021, *in press*.

[34] A. Holzinger, H. Müller, Toward human–ai interfaces to support explainability and causability in medical ai, Computer 54 (10) (2021) 78–86. doi:10.1109/MC.2021.3092610.

[35] D. J. Bonthius, R. McKim, L. Koele, H. Harb, B. Karacay, J. Mahoney, N. J. Pantazis, Use of frozen sections to determine neuronal number in the murine hippocampus and neocortex using the optical disector and optical fractionator, Brain Research Protocols 14 (1) (2004) 45–57.

[36] V. M. Golub, J. Brewer, X. Wu, R. Kuruba, J. Short, M. Manchi, M. Swonke, I. Younus, D. S. Reddy, Neurostereology protocol for unbiased quantification of neuronal injury and neurodegeneration, Frontiers in aging neuroscience 7 (2015) 196.