



Automated Cell Counts on Tissue Sections by Deep Learning and Unbiased Stereology

Saeed S. Alahmari^{a,*}, Dmitry Goldgof^a, Lawrence Hall^a, Hady Ahmady Phoulady^b, Raj H. Patel^c, Peter R. Mouton^{a,d}

^a Department of Computer Science and Engineering, University of South Florida, Tampa, FL, USA

^b Department of Computer Science, University of Southern Maine, Portland, ME, USA

^c Department of Chemistry University of South Florida, Tampa, FL, USA

^d SRC Biosciences, Tampa, FL, USA

ARTICLE INFO

Keywords:

Unbiased stereology
Automatic optical fractionator
Convolution neural network (CNN)
Deep learning
Adaptive Segmentation Algorithm (ASA)

ABSTRACT

In recent decades stereology-based studies have played a significant role in understanding brain aging and developing novel drug discovery strategies for treatment of neurological disease and mental illness. A major obstacle to further progress in a wide range of neuroscience sub-disciplines remains the lack of high-throughput technology for stereology analyses. Though founded on methodologically unbiased principles, commercially available stereology systems still rely on well-trained humans to manually count hundreds of cells within each region of interest (ROI). Even for a simple study with 10 controls and 10 treated animals, cell counts typically require over a month of tedious labor and high costs. Furthermore, these studies are prone to errors and poor reproducibility due to human factors such as subjectivity, variable training, recognition bias, and fatigue. Here we propose a deep neural network-stereology combination to automatically segment and estimate the total number of immunostained neurons on tissue sections. Our three-step approach consists of (1) creating extended-depth-of-field (EDF) images from z-stacks of images (disector stacks); (2) applying an adaptive segmentation algorithm (ASA) to label stained cells in the EDF images (i.e., create masks) for training a convolutional neural network (CNN); and (3) use the trained CNN model to automatically segment and count the total number of cells in test disector stacks using the optical fractionator method. The automated stereology approach shows less than 2% error and over $5 \times$ greater efficiency compared to counts by a trained human, without the subjectivity, tedium, and poor precision associated with conventional stereology.

1. Introduction

Unbiased stereology is a set of theoretical and practical methods that allow for theoretically accurate (unbiased) estimation of stereology parameters for stained cells by carefully avoiding all known sources of methodological bias (West, 2012)(Mouton, 2011). Examples of common stereology parameters include counts of total cell numbers and cell density; region and mean cell volumes; surface area and surface density; and total length and length density (Burke et al., 2009; Mouton, 2011). Though based on theoretically unbiased principles, a current weakness of state-of-the-art stereology systems is their dependence on human data collectors. As a result, current stereology systems are labor intensive, costly and prone to counting errors due to variable user training, subjectivity and fatigue.

We have previously reported an automatic stereology approach

called Adaptive Segmentation Algorithm (ASA) for quantifying the total numbers of NeuN immunostained neurons from extended depth of field (EDF) images (Mouton et al., 2017; Bradley and Bamford, 2004). As shown in Mouton et al. (2017), the ASA approach allows for comparable accuracy as trained human counters. However, a limitation of the ASA is the domain specificity, since, modifications of ASA parameters are needed to account for variations in cell size, staining intensity, and background illumination. With the goal of more robust performance across different staining domains, here we expand on that approach to include an artificial intelligence (AI) technique known as deep learning (LeCun et al., 2015).

Deep neural networks have lately generated considerable interest in the medical imaging field where they have shown significantly better performance over conventional engineered images analysis algorithms (Lee et al., 2017). Although the idea of neural networks has been

* Corresponding author.

E-mail address: saeed3@mail.usf.edu (S.S. Alahmari).

<https://doi.org/10.1016/j.jchemneu.2018.12.010>

Received 30 May 2018; Received in revised form 21 December 2018; Accepted 22 December 2018

Available online 27 December 2018

0891-0618/ © 2018 Elsevier B.V. All rights reserved.

around for a long time, the emerging revolution of deep neural networks was partially due to the development of convolution neural network (CNN); optimization algorithms (Hinton and Salakhutdinov, 2006; Nair and Hinton, 2010; Srivastava et al., 2014; Ioffe and Szegedy, 2015); highly efficient computation resources; and the availability of large datasets (big data). Deep learning refers to the representation learning methods that start from raw data and have multiple hidden layers (neither input nor output) (LeCun et al., 2010). For instance, images are the raw data for a neural network to learn a certain task (i.e., segmentation), whereas hidden layers contribute to features learning process. In the medical field, large sets of labeled data are mostly hard to find, which remains an obstacle to deep learning based applications. Data augmentation such as the rigid and non-rigid transformation of images can be used to address this problem. In this paper, a deep learning architecture for medical image segmentation (Unet) was used to create segmentation masks of EDF images (Ronneberger et al., 2015). In Section 4, we explain Unet in further detail.

In this paper, we propose a pipeline method that includes state-of-the-art deep learning architecture for segmentation and unbiased stereology of histology images with improved accuracy, precision and throughput efficiency as compared to current manual cell counting stereology systems. A novel feature of our approach is that verified segmentation masks from the ASA are used as ground truth for training a deep neural network to make automatic counts of stained cells (neurons) on test images. This innovation effectively reduces the human effort on data labeling by using an unsupervised algorithm ASA to create masks for EDF images. This patent-pending approach [FAST (fully automatic stereology technology)] includes several automatic and semi-automatic option for verification of results by trained experts.

2. Dataset

The use of animals in this study was approved by the USF Institutional Animal Care and Use Committee according to NIH guidelines as described by Mouton et al. (2017). The dataset used in this experiment was sampled in a systematic-random manner from the neocortex of Tg4510 mouse brains. Among the phenotypic changes in these mice are progressive neuron degeneration and activation of neuroglia cells in the neocortex and other brain regions (Mouton et al., 2017; Santacruz et al., 2005; Spires et al., 2006). Neu-N immunostained neurons were segmented and counted in disector volumes according to current stereology principles (Mouton et al., 2017; West et al., 1991), and the total neuron number was calculated using the optical fractionator feature in the Stereologer system (SRC Biosciences, Tampa, FL) (Mouton et al., 2017). Briefly, after manual counts at systematic-random x–y locations through the neocortex of nine mouse brains, disector stacks (10 z-axis images 1 μ m apart) were captured at the same locations using high magnification objective (100 $\times$ oil, 1.4 NA) for brightfield microscopy. Table 1 shows the number of sections and number of z-axis stacks (disector stacks) obtained and converted into an EDF images, where the interval for stacks is 1.0 μ m and the section

Table 1
Datasets mouse ID, number of sections per mouse and total number of stacks per mouse.

Mouse ID	Number of sections	Number of stacks
02	8	113
03	6	121
14	8	90
17	7	91
29	8	135
21	7	102
24	8	103
67	8	104
09	6	107

sampling fraction is 1/6th. A total of 966 EDF images with their corresponding ASA masks were used for automatic counts of Neu-N immunostained neurons as described in the following section.

3. Adaptive Segmentation Algorithm (ASA)

As detailed in Mouton et al. (2017), the ASA consists of multiple steps optimized to segment neocortical NeuN-immunostained neurons. The ASA includes a Gaussian mixture (GMM), morphological operations, Voronoi diagrams, and watershed segmentation. Our ASA starts with EDF images to segment NeuN immunostained cells within an ROI using a GMM, where a GMM uses pixel intensity for the expectation maximization algorithm (EM) to estimate its parameters followed by thresholding and morphological operations to get separate cells. A processed EDF image using opening then closing by reconstruction was used in the watershed foreground and background markers extraction. These foreground and background markers used in applying watershed segmentation were followed by segmentation approximation using a Voronoi diagrams algorithm, and finally a smoothing process to enhance cell boundaries using the Savitzky–Golay filter (Savitzky and Golay, 1964).

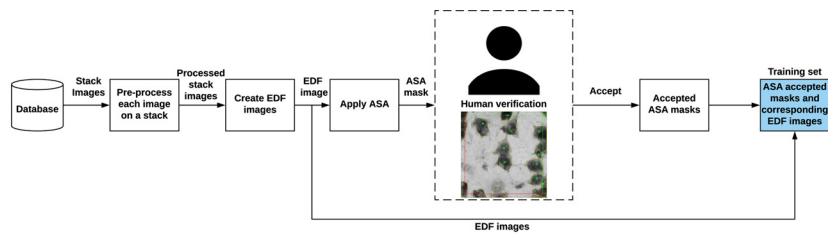
4. Deep learning

Deep learning is a representation learning technique that can learn to perform specific tasks (i.e., detection, classification, and segmentation) from images, sound, text, or numbers (LeCun et al., 2015). This technique does not require handcrafting features, but instead learns discriminant and other powerful features automatically (Litjens et al., 2017). Initially, deep learning was inspired by the biological brain functions such as communication and processing of tasks on biological neurons (Goodfellow et al., 2016). An example of deep learning is the common CNN, a learning algorithm that consists of one or more convolution layers that serves as the main building block of a CNN for learning to recognize features of images. Although a version of CNN has existed since 1980, computation power and data availability had been an obstacle until recently (Rawat and Wang, 2017).

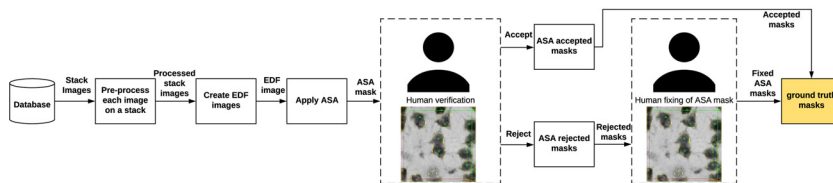
In this paper, we use a CNN based architecture for medical image segmentation known as Unet (Ronneberger et al., 2015). This neural network is a simple, fast, and end-to-end fully convolutional network (FCN) that contains contraction and expansion paths to capture context and learn precise localization. Unet has 19 convolution layers where a max pool layer follows every two convolution layers. Moreover, it has skipping connections between the encoding and decoding path to add precise localization of high resolution to segmented objects (Ronneberger et al., 2015). Each convolution layer in encoding and decoding paths are 2D convolution operations with a filter size of 3×3 followed by rectified linear unit (ReLU) activation function. A max-pooling layer of size 2×2 follows each of the two convolution layers in the encoding path. In contrast, an up-sampling layer of size 2×2 follows each of the two convolution layers in the decoding path. The last layer is a 2D convolution layer of filter size 1×1 followed by a sigmoid activation function.

5. Methods

This section outlines steps in our proposed deep learning-based unbiased stereology approach to segment and count cells in immunostained tissue sections. Our approach starts with data preparation and verification, then training the deep learning model, after that a post-processing step of predicted masks, and finally counting cells step to calculate the total number of cells using the unbiased optical fractionator method.



(a) Training set preparation steps



(b) Ground truth preparation steps

5.1. Image preparation and mask verification

Three-dimensional stacks of Z-axis images (disector stacks) containing NeuN-immunostained neurons were captured from 6 to 8 mouse brain sections sampled in a systematic-random manner through the neocortex of nine mouse brains as shown in Table 1. This data collection was done using a commercial stereology system (Stereologer, SRC Biosciences, Tampa, FL) consisting of Leica DM2500 bright-field microscope equipped with a motorized X–Y–Z stage, and high-resolution digital camera. Data preparation and verification involved two main steps: training set preparation and ground truth preparation as shown in Fig. 1.

Training set preparation starts with preprocessing stacks of images for converting images to grayscale followed by creating an extended-depth-of-field (EDF) image (Fig. 1(a)). The EDF algorithm based on discrete wavelet transformation that convert each disector stack into a synthetic image in which all stained cells appear at their maximum plane of focus (Bradley and Bamford, 2004). These in-focus cells are “dropped” rather than projected onto a 2D plane. Stereology bias is avoided by ensuring each cell in the disector stack has the same probability ($=1$) of being counted inside a known volume (disector volume). Bias from edge effects (Gundersen, 1977) is avoided through the use of a unique plane (i.e., the most in-focus plane) to determine whether the cell hits an exclusion plane. The ASA is applied to each EDF image for segmentation, resulting in a binary mask of cells. We utilized masks created by ASA rather than using manual annotation directly since manual annotation has only counting marks and does not provide an outline of the cell (i.e., mask) as shown in Fig. 2(a). After creating the ASA mask, a human verification step is performed to verify the agreement between the manual annotation and the ASA mask. If the ASA mask matches the manual annotation except for the exclusion lines (i.e., left and lower disector frame lines applied in a subsequent step), then the human verifier accepts ASA mask. The accepted ASA masks and their corresponding EDF images provided the training set. The steps for training set preparation are illustrated in Fig. 1(a).

Ground truth preparation is similar to training set preparation in the preprocessing step by creating EDF images and creating ASA masks (Fig. 1(a)). In a subsequent verification step an expert verifies the agreement of ASA mask and manual annotation. If ASA masks match manual annotation except for the exclusion lines, then the ASA mask is accepted. If the ASA mask does not match manual annotation, then the ASA mask

Fig. 1. (a) Training set preparation, where every stack image is preprocessed for conversion to grayscale, then EDF image creates an in-focus image from a stack of images, then ASA creates a binary mask. Human verification is applied to accept correct ASA masks. (b) Ground truth preparation, where every image in a stack is preprocessed for conversion to grayscale, then an in-focus EDF image is created from a stack of pre-processed images, then ASA creates a binary mask. Human verification is applied to accept or reject ASA masks based on manual annotation. If an ASA mask is rejected, then a human fixes the mask manually and the result is concatenated with ASA accepted masks to form a ground truth for the whole dataset.

is rejected. It should be noted that rejection does not eliminate those rejected masks from our study; instead, the expert fixes the rejected masks manually to create ground truth for test images. Expert-based manual fixing of rejected masks creates a segmentation of cells missing on the ASA masks (false negatives) and eliminates cells that were wrongly segmented by the ASA (false positives). The fixed masks and accepted ASA masks together provide the ground truth for the whole dataset. Finally, manual correction of rejected ASA masks allows for calculation of the dice coefficient similarity metric. The schematic for preparation of ground truth is shown in Fig. 1(b).

To minimize artifacts caused by rotation or elasticity, data augmentation was applied to EDF images prior to cropping. EDF images were reduced by cropping 20 pixels beyond the disector lines for training and testing. Fig. 2 shows inclusion/exclusion lines of the disector frame on the manual annotation (Fig. 2(a)); an EDF image after cropping 20 pixels around the disector frame (Fig. 2(a)); predicted mask before post-processing (Fig. 2(c)); and cells contours after post-processing overlaid on top of the manual annotation image (Fig. 2(d)).

5.2. Training and testing models

For training the deep learning model, EDF images and their corresponding accepted ASA masks (i.e., the training set) were augmented to increase the number of instances for the training set for a better performance via a more general deep learning model. Data augmentations used on this data are rotation augmentation, elastic deformation (Simard et al., 2003), or combinations thereof, rotational augmentation of 45° and 15° , or combined augmentation of elastic images using a rotation of 45° and 15° . Augmented images and their corresponding masks of size 400×400 were used to train the Unet architecture with the Adam optimizer algorithm implementation of Keras with TensorFlow backend (Chollet et al., 2015; Kinga and Adam, 2015; Abadi et al., 2015). The Adam optimizer learning rate was set to $1e^{-4}$, while exponential decay rates for the moment estimates hyperparameters β_1 and β_2 were set to 0.9 and 0.999 respectively (Kinga and Adam, 2015). Training the deep learning model is the process of using the training set for learning discriminant and powerful features from the input images using the deep learning (i.e., Unet) which yields a trained model that can segment cells on an unseen test set. For validation purposes, we have used leave-one-out cross-validation on the mouse level, which means training was performed on images of eight mice and testing on

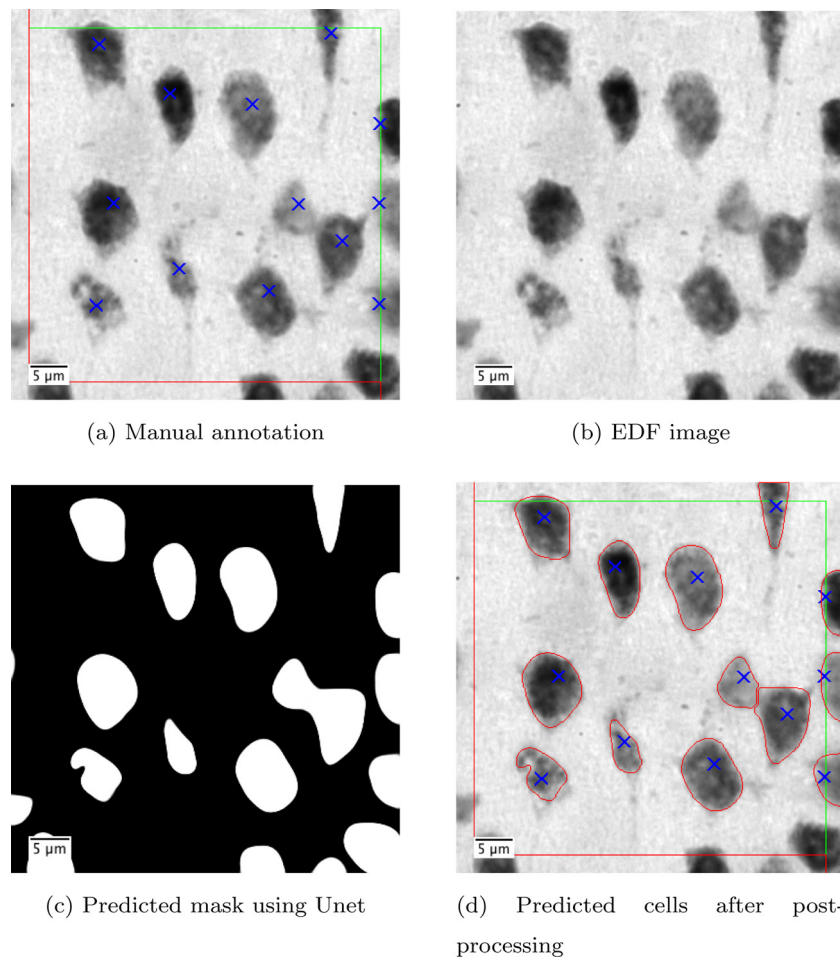


Fig. 2. Examples from our dataset where (a) is the manual annotation (counted neurons are denoted by blue marks), (b) is the EDF image, (c) predicted mask (before post-processing), (d) predicted cells contours (after post-processing) overlaid on top of manual annotation image. The contours in (d) shows the counted cells by the proposed method, where all cells marked in (a) were counted correctly and there is no overcount.

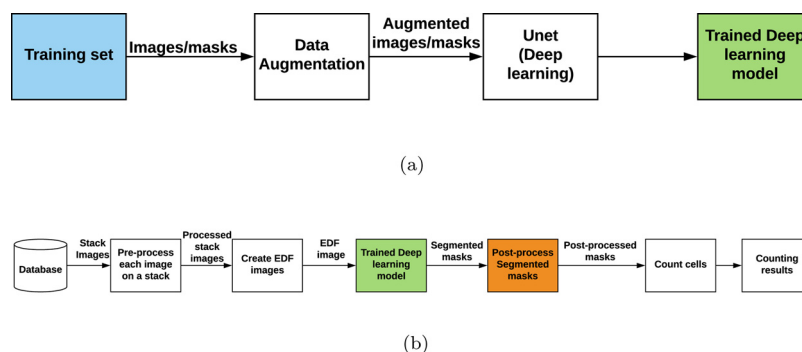


Fig. 3. (a) Training deep learning (i.e., Unet). (b) Testing the trained model followed by post-processing and counting cells.

images of the ninth mouse alternatively. This approach ensures that training set and testing set images do not overlap and avoids pre-selection of a separate test set that could be biased. Thus, a trained model is the result of training a deep learning model that can be used for testing a separate mouse. The training stage is illustrated in Fig. 3(a).

For testing a deep learning model, stacks of images were converted to grayscale, followed by creating an in-focus image using the EDF algorithm. The EDF images of the test mouse were provided to the trained model to predict segmentation masks. Testing the deep learning model is illustrated in Fig. 3(b).

5.3. Post-processing and counting

Each deep learning model created segmentation masks of the test set that were subsequently post-processed for three goals: to remove small amounts of segmented stain artifacts on an image background; to remove cells touching the exclusion line (i.e., left and lower disector box lines); and to separate touching cells. Removing the small amount of noise was done by thresholding the predicted mask based on the area in pixels. For instance, using an area of 250 pixels (i.e., $3.86 \mu\text{m}^2$) as a threshold removes small segmented stain artifacts on the background of EDF images. This threshold was selected by surveying our dataset to find the small stain artifacts to be excluded. Touching cells were

separated using the watershed algorithm, and a minimum cell size was imposed to avoid over-segmentation (Meyer, 1994). It should be noted that there was no maximum range size of cells to be counted. Therefore, the proposed method can segment and count hypertrophy and atrophy cells as long as the size of an atrophy cell is above the minimum area threshold (250 pixels). After post-processing, a counting step was applied using the unbiased optical fractionator method. This approach uses the disector principle (Sterio, 1984) to avoid overcounts by (1) identifying cells when their top-most point moves from out-of-focus to in-focus within the disector stack; and (2) adjusting the final count for edge effects by applying Gundersens exclusion/inclusion planes in XYZ Gundersen (1977). After imposing unbiased counting rules in the post-processing step and obtaining total count of cells (ΣQ^-), Eq. (1) (Optical Fractionator formula) was applied to estimate the total number of cells in the neocortex, where (ΣQ^-) is the summation of counted cells based on the unbiased stereology approach in a given sample (i.e., summation across all sections of a sample to reflect the number of counted cells in a known fraction of the reference space) using manual counting; ASA or deep learning, f_1 is the reciprocal of the section sampling fraction; f_2 is the reciprocal of the area sampling fraction; and f_3 is the reciprocal of the thickness sampling fraction (West et al., 1991).

$$\text{Total}_{\text{NeuN}} = (\sum Q^-) * f_1 * f_2 * f_3 \quad (1)$$

6. Experiments and results

To fairly evaluate a trained deep learning model, mouse images cannot be in common between training and testing. Therefore, training was performed in eight mice and testing was done on the ninth mouse. This procedure was performed alternatively nine times using cross-validation on the mouse level where a different mouse is left out for testing each time. Eq. (2) was used for calculating the error rate for a single mouse,

$$\text{Error_rate} = \frac{|y_{\text{true}} - y_{\text{pred}}|}{y_{\text{true}}} * 100 \quad (2)$$

where y_{true} is the total number of cells in ground truth (manual annotation), and y_{pred} is the total number of cells in predicted masks. The Dice coefficient that measures spatial overlap of a predicted mask and ground truth mask (Zou et al., 2004) was calculated by Eq. (3), where A is the ground truth mask and B is the predicted mask. This section discusses the experiment when training on EDF images and their corresponding ASA masks which match the manual annotation (i.e., accepted images by a human) as shown in Fig. 3(a). Overall error rate and Dice coefficient results are the average over the results of nine different models. The total number of accepted ASA masks was 167 masks, whereas, the total number of ASA rejected (then fixed) masks was 799 masks. Table 2 shows results when training on original images with no augmentation, in addition to five different augmentation approaches. When no augmentation was applied, the average error rate (i.e., the average error rate over nine mice, when testing on each mouse

individually) was 4.39% with Dice coefficient 0.786. Another experiment tested the effects of elastic augmentation alone on the performance of the CNN model as described in Simard et al. (2003). The results showed improved CNN performance after elastic augmentation with an average error rate 2.44% and Dice coefficient 0.796. Training using rotation (45°) augmented images showed an average error rate 3.44% and a Dice coefficient of 0.801. The lowest average error rate was 1.85%, with Dice coefficient 0.804 when training with rotation (15°) augmentation of elastic augmented images [Table 2, Eq. (3)].

$$\text{Dice} = \frac{2 * |A \cap B|}{|A| + |B|} \quad (3)$$

A comparison between manual annotation, ASA, and deep learning cell counts per mouse is shown in Fig. 4(a), whereas Fig. 4(b) shows a comparison between ASA error rate and deep learning error rate across all mice on our dataset. The average error rate over nine mice for ASA and the deep learning model (11.9% versus less than 1.85%, respectively) are shown in Fig. 4(c). The error rate of the deep learning method proposed here was statistically significant (lower) at $p < 0.05$ using a two-tailed t -test, where the t -value is 5.29, and the p -value is 0.000073 compared to the ASA error rate. Thus, the proposed method was able to precisely segment and count neurons compared to ASA segmentation. Fig. 5 shows three examples of ASA versus our proposed segmentation method using the Unet deep learning architecture, where predicted masks contours are overlaid on top of manual annotation images.

7. Discussion

We have previously proposed an automatic stereology segmentation-based counting method (i.e., ASA) (Mouton et al., 2017) that achieved reasonable results; however, it requires manipulating specific parameters to get the best performance. In our proposed deep learning method, a significant improvement in terms of enhanced segmentation and substantial reduction in error rate was achieved. Both ASA and the Deep Learning involve data acquisition using an X–Y–Z motorized stage to collect disector stacks of images at a systematic randomly sampled (SRS) locations in routine Neu-N immunostained tissue. Here we propose a novel approach using EDF images of the disector stacks and ASA masks matching the manual annotation to train deep learning models. After training and testing the deep learning model, the post-processing step was applied where unbiased counting rules were imposed and separation of touching cells was done. Finally, the optical fractionator was used to estimate the total number of cells in an anatomically defined reference space (mouse neocortex).

The current study found that deep learning achieved more accurate counting results than ASA counts [Table 2]. The lowest average error rate was about 1.85% using rotation 15° of elastic augmented images as compared to the ASA error rate of 11.94%, difference in error rates of 85%. Compared to manual stereology, both the deep learning method proposed here for automatic stereology and our previous ASA approach (Mouton et al., 2017) avoid the subjectivity, recognition bias, variable training, and fatigue of manual data collectors. Reproducibility was improved by the use of an objective method (i.e., deep neural network) to segment and count target cells.

Lack of a large number of properly annotated images remain an obstacle to many researchers, especially in the medical field, because manual image annotation is labor-intensive, tedious, and error-prone work. Another issue is the lack of expert labor to annotate a large number of images for building a more robust model. Therefore, utilizing a pre-existing algorithm (i.e., ASA) to generate masks for EDF images overcomes the significant challenges of creating masks manually. As shown here, user intervention was not eliminated but reduced, since ASA masks need a verification step to accept masks that match manual annotation.

We investigated a number of different data augmentations methods

Table 2

Results of our experiments when testing on each mouse independently (i.e., leave a mouse out for each mouse in our dataset).

Augmentation task	Average error rate (%)	Standard deviation	Average Dice coefficient
No augmentation	4.39	2.69	0.786
Elastic	2.44	1.50	0.796
Rotation 45°	3.44	4.03	0.801
Rotation 15°	3.89	2.61	0.802
Rotation 45° of elastic	3.68	2.72	0.804
Rotation 15° of elastic	1.85	2.07	0.804

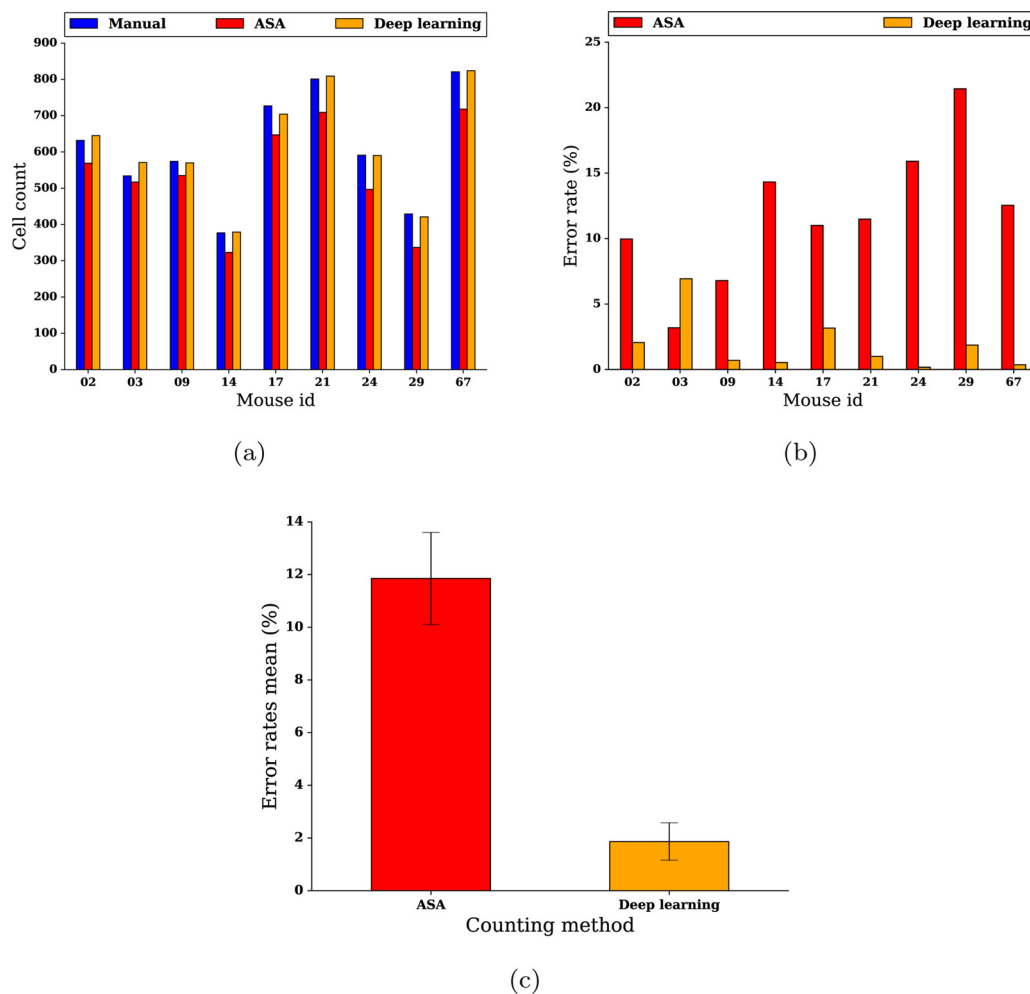


Fig. 4. (a) Comparison between manual, ASA, and deep learning cell count results per mouse, (b) a comparison between ASA and deep learning error rate per mouse, and (c) shows error rate mean and standard error of the mean of ASA and deep learning.

to increase the number of training images, thereby facilitating learning of the segmentation by the CNN model. The small number of training images was a result of (1) removing images with data acquisition issues such as microscopy lighting level and tissue fold artifacts which represent about 1.5% images of the total number of images on our dataset; and (2) ASA masks based rejection of images/masks that do not match manual annotation.

The semi-automatic and fully automatic unbiased stereology cell counting can be achieved on brain cells (NeuN-immunostained neurons) from EDF images created from disector stacks. The EDF images have the potential for large cells to block counts of smaller cells located distal to the observation plane (masking). Thicker sections are only an issue when they involve more densely packed cells, but not for thicker sections without high cell packing (e.g., Golgi stained sections). As for human cell counters using manual stereology, highly dense cell packing would introduce segmentation error for our automatic approach. Therefore, the effects of masking on cell counts are expected to vary as a direct function of the cell packing density. Fortunately, high-density cell-packed regions are relatively rare in histology images. The present work shows that any systematic error from masking, for neuron counts in the mouse neocortex, is negligible for all practical purposes, as evidenced by the demonstration of less than 2% error using a CNN model trained from EDF images of accepted ASA masks [Table 3]. Our ongoing work is assessing techniques to minimize and eliminate potential optical artifacts from masking and other sources relative to gold standard counts [Table 3], including both semi-automatic and fully automatic approaches [Table 3].

A predesign diagram for the stereology system that includes outlining an ROI, capturing stacks of images, auto-generating ASA and deep learning segmentation, cell count, and exporting results is shown in Fig. 6.

8. Conclusions

We propose a novel AI-based method that uses deep learning and unbiased stereology counting criteria for automatically segmenting and estimating the total number of immunostained cells in defined ROIs. The lowest error rate of 1.85% was achieved using a deep learning model trained on elastic and rotated EDF images compared to an error rate of 11.94% for ASA alone. Another novel contribution in this work is the demonstration that image augmentation can be a critical tool for improving CNN performance after the automatic generation of ground truth using the ASA. Thus, automatic generation of ground truth followed by a verification process with automatic segmentation algorithms such as ASA can be used to generate labels (i.e., segmentation masks), thereby reducing the considerable time and effort for manual data annotation to create masks.

Although these studies focused on counting immunostained brain cells (neurons) on tissue sections, the approach can be applied to stereology studies of stained cells from any tissue. This study opens a new frontier for deep learning in combination with unbiased stereology research that could accelerate a wide range of life science research and drug discoveries. Our future work includes segmentation based unbiased counts of cells using more precise region-based neural networks

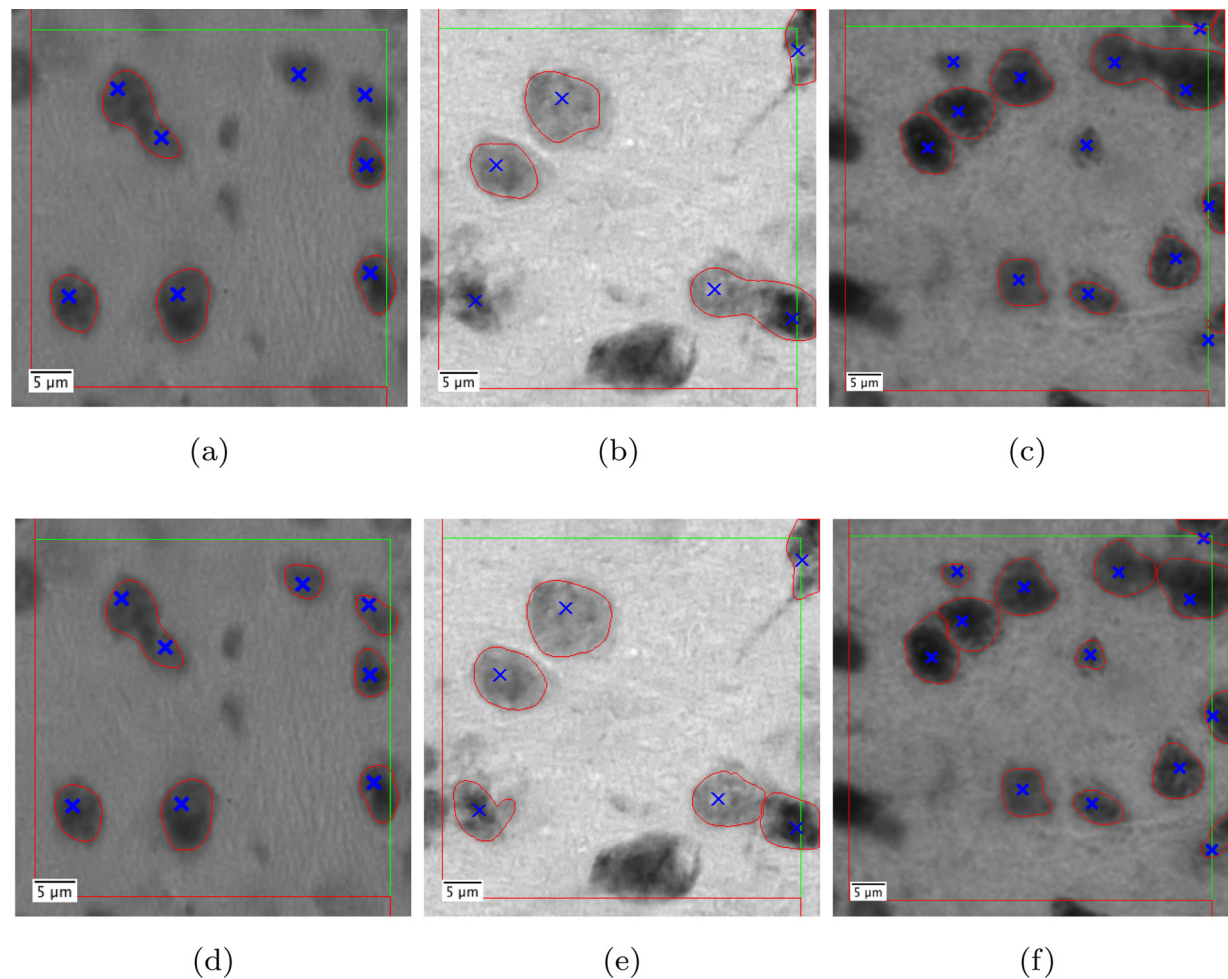


Fig. 5. Comparison between ASA (a–c) and deep learning (d–f) cell segmentation, where cell mask contours are overlaid on top of manual annotation image, and a line of 5 μm is shown for scale.

such as Mask-RCNN (He et al., 2017), and expanding this approach to other stereology parameters, including length and local size (mean cell volume) estimators, and a range of semi-automatic and fully automatic stereology approaches.

Conflict of interest

The authors declare that the research was conducted in the absence of any real or perceived conflict of interest. For their work on this project, co-author Peter R. Mouton, Ph.D., Chief Scientific Officer of SRC Biosciences (Tampa, FL, USA), principal investigator and owner of Intellectual property for the Stereologer system, and co-investigator Hady Ahmady Phoulady, Ph.D., received grant support from competitive Small Business Innovative Research (SBIR) grants and Small

Business Technology Transfer Research (STTR) grants from the National Institutes of Health (NIH) and the National Science Foundation (NSF), respectively. All other co-authors are owners of intellectual property for this work and received grant support from sub-awards on these grant to the University of South Florida (Tampa, FL, USA).

Acknowledgments

The authors would like to thank Dr. Marcia Gordon of Michigan State University (Grand Rapids, MI) for the generous donation of the stained tissue sections for these studies. The work was supported by National Science Foundation grant (#1746511), (#1513126), and a grant from the Florida High Technology Corridor grant program.

Table 3
Options for manual, semi-automatic and fully automatic counts by optical fractionator to estimate total cell number. A single asterisk (*) denotes the gold standard methods, and double asterisks (**) denote the future and our ongoing work not reported in this paper.

Method	Counting approach	Type
1) Manual stereology*	User clicks while thin focal-plane optical scanning in z-axis	Fully manual
2) Disector Stacks*	User clicks while scrolling through each disector stack	Semi-automatic
3) EDF counts	User clicks on EDF images created from disector stacks	Semi-automatic
4) ASA – EDF images (Mouton et al., 2017)	ASA counts cells on EDF images	Fully automatic
5) ASA – EDF with edits	User edits ASA counts on EDF images	Semi-automatic
6) CNN – EDF images	Automatic counts from CNN model applied to EDF images	Fully automatic
7) Iterative deep learning**	CNN model retuning after user edits of initial CNN results	Semi-automatic
8) 3D deep learning**	CNN model trained on each z-plane of disector stacks	Fully automatic

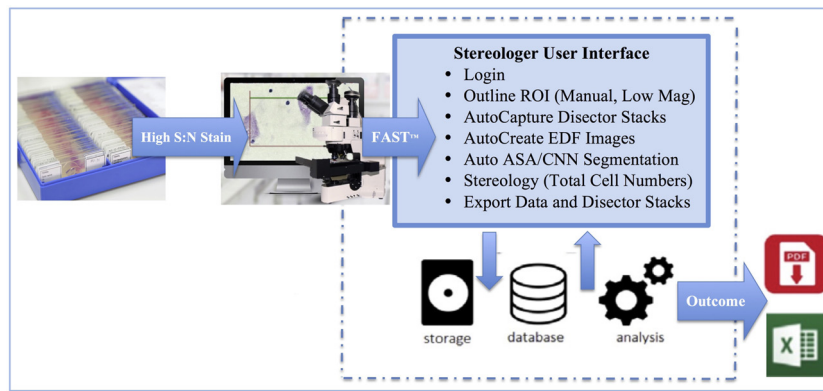


Fig. 6. Stereology system design that include outline ROI, auto capture stacks, auto count using deep learning, and export results.

References

- West, M.J., 2012. Basic Stereology for Biologists and Neuroscientists. Cold Spring Harbor Laboratory Press.
- Mouton, P.R., 2011. Unbiased Stereology: A Concise Guide. JHU Press.
- Burke, M., Zangenehpour, S., Mouton, P.R., Ptito, M., 2009. Knowing what counts: unbiased stereology in the non-human primate brain. *J. Vis. Exp.* 27.
- Mouton, P.R., Phoulady, H.A., Goldgof, D., Hall, L.O., Gordon, M., Morgan, D., 2017. Unbiased estimation of cell number using the automatic optical fractionator. *J. Chem. Neuroanat.* 80, A1–A8.
- Bradley, A.P., Bamford, P.C., 2004. A one-pass extended depth of field algorithm based on the over-complete discrete wavelet transform. *Image and Vision Computing'04 New Zealand (IVCNZ'04)* 279–284.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436.
- Lee, J.-G., Jun, S., Cho, Y.-W., Lee, H., Kim, G.B., Seo, J.B., Kim, N., 2017. Deep learning in medical imaging: general overview. *Korean J. Radiol.* 18 (4), 570–584.
- Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. *Science* 313 (5786), 504–507.
- Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* 807–814.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15 (1), 1929–1958.
- Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning* 448–456.
- LeCun, Y., Kavukcuoglu, K., Farabet, C., 2010. Proceedings of 2010 IEEE International Symposium on Convolutional networks and applications in vision. In: *Circuits and Systems (ISCAS)*. IEEE. pp. 253–256.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer. pp. 234–241.
- Santacruz, K., Lewis, J., Spires, T., Paulson, J., Kotilinek, L., Ingelsson, M., Guimaraes, A., Deture, M., Ramsden, M., McGowan, E., et al., 2005. Tau suppression in a neurodegenerative mouse model improves memory function. *Science* 309 (5733), 476–481.
- Spires, T.L., Orne, J.D., SantaCruz, K., Pitstick, R., Carlson, G.A., Ashe, K.H., Hyman, B.T., 2006. Region-specific dissociation of neuronal loss and neurofibrillary pathology in a mouse model of tauopathy. *Am. J. Pathol.* 168 (5), 1598–1607.
- West, M., Slomianka, L., Gundersen, H.J.G., 1991. Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. *Anat. Rec.* 231 (4), 482–497.
- Savitzky, A., Golay, M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36 (8), 1627–1639.
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88.
- Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. *Deep Learning*, vol. 1 MIT Press, Cambridge.
- Rawat, W., Wang, Z., 2017. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput.* 29 (9), 2352–2449.
- Gundersen, H.J.G., 1977. Notes on the estimation of the numerical density of arbitrary profiles: the edge effect. *J. Microsc.* 111 (2), 219–223.
- Simard, P.Y., Steinkraus, D., Platt, J.C., et al., 2003. Best practices for convolutional neural networks applied to visual document analysis. *ICDAR*, vol. 3 958–962.
- Chollet, F., et al., 2015. Keras. <https://github.com/fchollet/keras>.
- Kinga, D., Adam, J.B., 2015. A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*.
- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org. URL: <https://www.tensorflow.org/>.
- Meyer, F., 1994. Topographic distance and watershed lines. *Signal Process.* 38 (1), 113–125.
- Sterio, D., 1984. The unbiased estimation of number and sizes of arbitrary particles using the disector. *J. Microsc.* 134 (2), 127–136.
- Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells, W.M., Jolesz, F.A., Kikinis, R., 2004. Statistical validation of image segmentation quality based on a spatial overlap index: scientific reports. *Acad. Radiol.* 11 (2), 178–189.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE. pp. 2980–2988.