Iterative Deep Learning Based Unbiased Stereology With Human-in-the-Loop

Saeed S. Alahmari
Department of Computer Science and Engineering, University of South Florida, Tampa, FL, USA
Email: saeed3@mail.usf.edu

Dmitry Goldgof
Department of Computer Science and Engineering University of South Florida Tampa, FL, USA
Email: goldgof@mail.usf.edu

Lawrence O. Hall
Department of Computer Science and Engineering, University of South Florida, Tampa, FL, USA
Email: lohall@mail.usf.edu

Palak Dave
Department of Computer Science and Engineering, University of South Florida, Tampa, FL, USA
Email: palakdave@mail.usf.edu

Hady Ahmady Phoulady
Department of Computer Science, University of Southern Maine, Portland, ME, USA
Email: hady.ahmadyphoulady@maine.com

Peter R. Mouton
SRC Bioscience
Tampa, FL USA
peter@disector.com

***Abstract*—Lack of enough labeled data is a major problem in building machine learning based models when the manual annotation (labeling) is error-prone, expensive, and tedious. In this paper, we introduce an iterative deep learning based method to improve segmentation and counting of cells based on unbiased stereology on regions of interest of extended depth of field (EDF) images. This method uses an existing machine learning algorithm called adaptive segmentation algorithm (ASA) to generate masks (verified by a user) for extended depth of field (EDF) images to train deep learning models. Then an iterative deep learning approach is used to feed newly predicted and accepted deep learning masks/images (verified by a user) to the training set of the deep learning model. The error rate in unbiased stereology count of cells on an unseen test set reduced from about 5 % to 1.58 % after 5 iterations of the deep learning based active learning stereology process.**

***Index Terms*—Unbiased Stereology, Active learning, Deep learning.**

I. Introduction

Unbiased stereology is a set of theoretical and practical methods for making accurate counts of stained cells by carefully avoiding all known sources of methodological bias [1] [2]. Examples of common stereology parameters include counts of total cell number and cell density; region and mean cell volumes; surface area and surface density; and total length and length density [3] [1]. However, current computer-assisted stereology systems available to bioscientists and medical scientists are based on a technology developed more than two decades ago. Therefore, a simple study requires tedious counting of hundreds of cells per sample by a well-trained technician [4]. In other words, a simple count of immunostained cells in a defined region of interest (ROI) requires about 2-3 **hours** for a well-trained technician to achieve a reliable result. Though based on theoretically unbiased principles, this approach is prone to data errors and low reproducibility due to user subjectivity, variable expertise, and fatigue. The Adaptive Segmentation Algorithm (ASA) [2] makes stereology counts of total numbers of brain cells (Neu-N immunostained neurons) by automatic segmentation and cell counting on Extended Depth of Field (EDF) images [5] [6]. In Section IV we present ASA details.

A critical aspect of building successful machine learning models is the availability of labeled data. However, labeled data is hard to obtain because the process is time-consuming, labor-intensive, and tedious. Additionally, data labeling in a medical field is mostly restricted to experts in the field and generally cannot be prepared by a crowdsourcing approach for reasons such as the quality of annotation and subject privacy. To overcome stereology images labeling difficulties (i.e., creating pixel-wise labels), we propose an iterative deep learning method to generate segmentation masks of cells on stained NeuN tissue images; then a human-in-the-loop approach was taken to verify each predicted mask and feed correct images-masks pairs to the training set.

Deep neural networks have generated considerable interest in the medical imaging field because they have shown performance advantages over conventional engineered image analysis algorithms. Although the idea of neural-networks has been around for a long time, the recent deep neural networks revolution is partly due to the development of convolution neural network (CNN), optimization algorithms [7] [8] [9] [10], and powerful, efficient computation resources. Deep learning refers to learning methods that often start from raw data get to a more abstract level [11]. Convolutional Neural Networks (CNNs) have shown

significant success in challenging tasks in image classification and recognition [12] [13]. In this paper, we use a CNN based architecture for medical image segmentation known as Unet [14]. This architecture is a simple, fast, and end-to-end fully convolutional network that contains contraction and expansion paths to capture context and learn precise localization.

In this paper, we propose a method that iteratively utilizes deep learning with human-in-the-loop and an existing unsupervised algorithm (ASA) which eliminates human data labeling entirely (creating masks) of NeuN stained images to quantify the number of cells ROI. This approach uses the state-of-art deep learning architecture in which the ASA verified results on EDF images are used to train a convolution neural network (CNN) model to segment and make automatic neuron counts on test images. Meanwhile, a set of deep learning predicted masks are verified by a human-in-the-loop and fed back to the train set. The main innovation is: i) elimination of human labeling effort (creating masks) using an existing unsupervised algorithm (ASA) to generate masks to train a CNN, ii) using deep learning iterative process also to reduce human effort in data labeling, where user only verifies the correctness of segmentation, and iii) improving deep learning stereology cell counting by adding correctly labeled images (EDF images and their corresponding masks) to the training set for next iteration.

II. Unbaised Stereology

Unbiased stereology approach is the state-of-the-art for biological objects quantification in tissue sections [15]. An essential component of this approach is unbiased sampling (i.e., systematic-random) that avoids all source of biased assumption such as shape, size, and orientation [15] [3]. Unbiased stereology uses virtual disector box to quantify the number of cells in a region-of-interest (ROI). Counting cells is based on their location within an ROI and disector box. For instance, cells touching disector inclusion-line (i.e., disector upper and right line) or inside the disector box are counted. However, cells that touch exclusion line (i.e., disector lower and left line) are not counted. Example of the disector box counting procedure is shown in Fig. 1a , where the green line represents inclusion line, and the red line represents the exclusion line. Counted cells are marked with green dots.

III. Data set

The data set used in this work was sampled from the neocortex brain region of Tg4510 mice. As described by Mouton et al. in [2], animals and the process usedin this study and process were approved by the University of South Florida (USF) Institutional Animal Care and Use Committee which follows NIH guidelines. The data set includes both genetically modified mice and control mice. Mice neurons change while expressing mutant tau. These neuron changes include neuron degeneration and neuroglia cells activation [2] [16] [17]. Mice samples were stained with NeuN staining from which counting was performed manually using an optical fractionator [4]. Disector stacks were captured and saved using the Stereologer system [2]. Table I shows the number of sections from which multiple stacks were obtained and converted into extended depth of field (EDF) images. The total number of EDF image we have is 496 with their corresponding ASA masks.

TABLE I
Data set mice ids, number of sections per mouse and total number of stacks per mouse

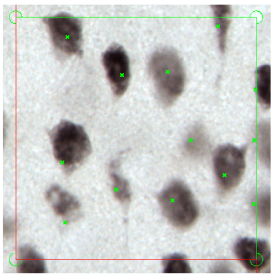
Mouse ID (LU)	Number of sections	Number of stacks
1 (LU2)	8	212
2 (LU3)	4	73
3 (LU14)	8	75
4 (LU17)	7	49
5 (LU29)	8	87

IV. Adaptive Segmentation Algorithm

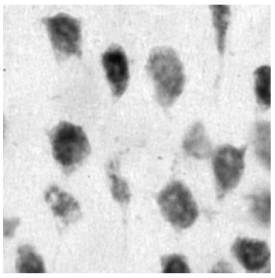
As shown in [2], the adaptive segmentation algorithm (ASA) consists of multiple steps optimized to segment cells at high magnification (63 to 100x oil immersion) microscopy. The ASA includes a Gaussian Mixture (GMM), morphological operations, Voronoi diagrams, and watershed segmentation. It starts with EDF images to segment NeuN stained cells within a region of interest (ROI) using GMM; where GMM uses pixel intensity for the Expectation Maximization algorithm (EM) to estimate its components followed by thresholding and morphological operations to get separate cells. A processed EDF image using opening then closing by reconstruction is used in the watershed foreground and background markers extraction. These foreground and background markers used in applying the watershed segmentation are followed by segmentation approximation using Voronoi diagrams. ASA uses smoothing process to enhance cell boundaries using a Savitzky-Golay filter [18]. The reason to use ASA is that manual annotation does not provide mask information, but instead, it provides a mark of what cell is being counted based on unbiased stereology approach. An example of manual annotation is shown in Fig. 1a.

V. Method

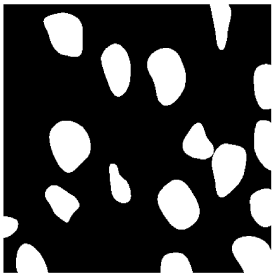
In unbiased stereology, where labeling cells is tedious, time-consuming, and subject to errors, iterative deep learning approach can leverage the data labeling process and generate correctly labeled examples that could help in building a more robust model. The EDF algorithm was applied to each stack of images to produce a single in-focus image as shown in Fig. 1b. Using EDF images, initial labels (masks) for our data set have been created using the ASA algorithm as shown in Fig. 1c, followed by a manual verification step to identify ASA accepted masks/images (i.e., ASA masks which match manual annotation) from ASA rejected masks/images (i.e., ASA masks do not match manual annotation) as illustrated in Fig. 2a. An example of manual annotation marks is shown in Fig. 1a, where green dots correspond to counted cells. For instance, the image shown in Fig. 1c was accepted by a user because of every cell mask (white blobs) inside or



(a) Manual annotation



(b) EDF image

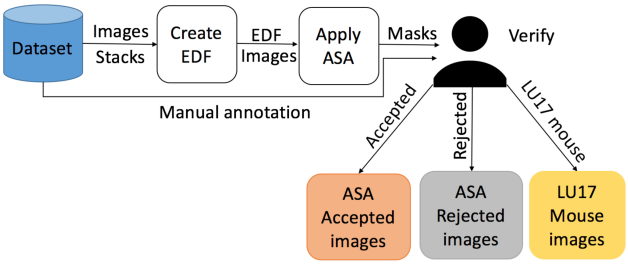


(c) ASA mask

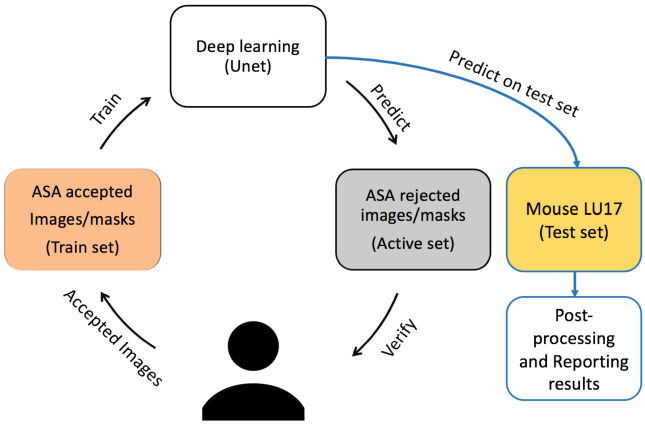
Fig. 1. example from our data set, where a) is the manual annotation (counted neurons have green dots), b) is the EDF image, and c) is the ASA mask for the EDF image shown in (b)

touching inclusion line (i.e., upper and right green line shown in Fig. 1a) correspond to a green dot (counted cells) in the manual annotation shown in Fig. 1a. This user verification step exempt cells that touch the exclusion line (disector left and lower line) are ignored for the purpose of training deep learning model.

In this section, we describe our iterative deep learning approach which can be described in five steps: 1) we train a deep learning model (Unet) on EDF images, and their corresponding ASA accepted masks that match manual annotation images. 2) A prediction was made on EDF images of ASA rejected masks that do not match manual annotation, and we refer to this set of images as the "active set". 3) Another set that does not overlap with either the training nor the active set is the "test set" which contains EDF images of different sections of a unique mouse (LU17 mouse). 4) The results of testing on the active set were verified by the user to compare the predicted mask and manual annotation similarity (as described in the previous paragraph). If an image matches the manual annotation image (i.e., cells marked for counting on a manual annotation image were segmented correctly using deep learning), then that image is moved to the training set for deep learning. Meanwhile, it gets removed from the active set. On the other hand, if an image does not match with its corresponding manual annotation image, then it remains in the active set. 5) the iterative process was performed until the number of images accepted is small ($\leq$ 5). Fig. 2b demonstrates the proposed method. It is important to note that human-in-the-loop involvement means accepting or rejecting a mask based on its corresponding manual annotation as



(a) Initial masks created using ASA followed by human verification



(b) Iterative human-in-the-loop verification of deep learning predicted masks

Fig. 2. Proposed method in two folds: a) creating EDF images, and applying ASA, then human verification, and then b) iterative process using accepted ASA masks/images for training, and ASA masks/images as an active set. Human verification (i.e., accept or reject) on every predicted mask. Test set is LU17 mouse images

shown in Fig. 1a. Therefore, no relabeling was made by the human-in-the-loop.

For each iteration of our method, we trained a deep learning architecture (Unet) for 100 epochs using Keras and Tensorflow backend [19] [20]. The Adam optimizer was used where the learning rate was set to $1e^{-4}$, while exponential decay rates for the moment estimates hyperparameters $\beta 1$ (first moment) and $\beta 2$ (second moment) were set to 0.9 and 0.999 respectively [21].

Based on the unbiased stereology method, cells counted in an ROI are those stained cells that are located inside the ROI or touching the inclusion line (i.e., top and right green line) but not touching the exclusion line (i.e., bottom and left red lines) as shown in Fig. 1a. For training purposes, we have kept all cells ignoring the unbiased stereology constraint. However, prior to reporting the results on the test set (LU17 mouse images), a preprocessing step was applied to remove small noise on the predicted mask, separate touching cells, and to impose unbiased stereology criteria of counting by

removing cells that touch the exclusion disector line.

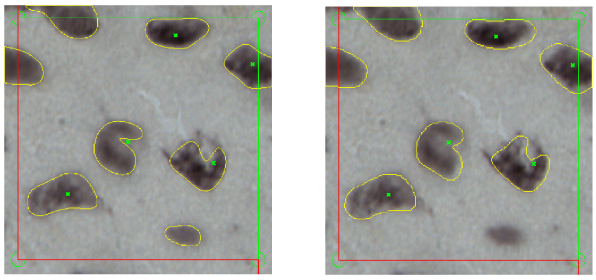
VI. Experiments and Results

Our data set has 496 stacks from 5 different mice. The EDF algorithm was used to create EDF images for each stack to get an in-focus image. The number of images in the initial train set (no augmentation) is 124 images, the number of images in the initial active set is 323 images, and the number of images in the test set (LU17 mouse images) is 49 images. Data augmentation used in this experiment was rotation 15° of elastic augmentation [22]. For example, for an image M, we apply an elastic algorithm with two different random seeds, which yields two elastic images M_1, M_2. Then for each of M_1 and M_2 we apply rotation augmentation of 15°. The total number of images generated by applying elastic then rotation augmentation of a single image is 49 images (including original image). This elastic and rotation augmentation is applied to the whole slide EDF image, and then we cropped images 20 pixels around the disector line as shown in Fig. 1b. We have used error rate to report results on the test set as shown in Equation 1, where y_true is the number of counted cells on ground truth (manual annotation), and y_pred is the number of counted cells on a predicted deep learning mask. For iteration 1, training on ASA accepted only images, and testing on LU17 mouse images (i.e., test set) resulted in 4.97 % error rate, and the user accepted 51 images from the active set. Increasing training images helped to reduce the error rate on the second iteration to 2.71 %; furthermore, 27 images were accepted by the user and moved to the training set. The lowest error rate on the test set was 1.58 % with the highest number of training images.

$$Error\_rate = \frac{|y\_true - y\_pred|}{y\_true} * 100 \qquad (1)$$

TABLE II
Results of the proposed method that shows the number of accepted images from active set every iterartion, and the error rate (%) on a different mouse (LU-17)

Iteration number	Number of accepted images	Error rate on LU-17 (%)
1	51	4.97
2	27	2.71
3	18	6.56
4	9	3.16
5	5	**1.58**



(a) ASA (b) deep learning

Fig. 3. Example from our data set, where a) the ASA masks contour overlaid on manual annotation image (counted neurons have green dots), b) the active learning based deep learning predicted masks contour (accepted the fifth iteration of our iterative deep learning based unbiased stereology) overlaid on manual annotation image

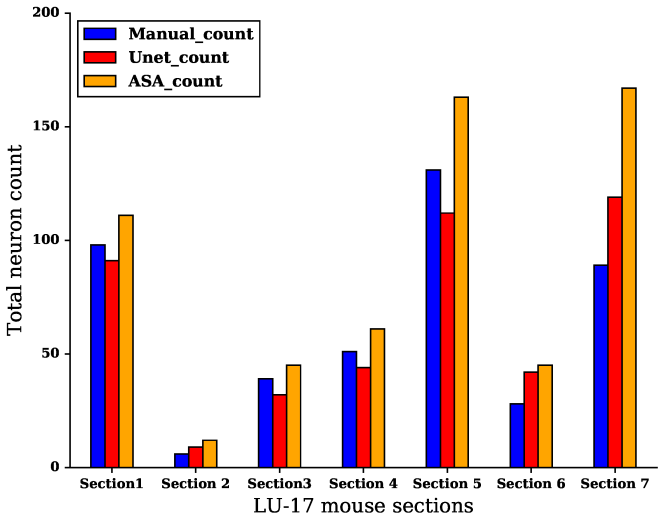


Fig. 4. LU-17 mouse cells count using manual, ASA, and Unet (deep learning)

TABLE III
LU-17 mouse cells count using manual method, Unet (deep learning), and ASA.

LU17 sections	Manual cells count	Unet cells count	ASA cells count
Section 1	98	91	111
Section 2	6	9	12
Section 3	39	32	45
Section 4	51	44	61
Section 5	131	112	163
Section 6	28	42	45
Section 7	89	119	167

In Table II, complete results of the iterative deep learning based unbiased stereology approach of five iterations are provided. Fig. 3 shows an example from our data set to compare ASA mask, and the iterative deep learning predicted mask. Fig. 3b shows an improved segmentation of cells on EDF image which was accepted on the fifth iteration. One unanticipated finding was that on the third iteration, the error rate increased to about 6.5 % on the test set. This could be caused by the rotation augmentation artifacts and user subjectivity on accepting new images/masks from the active set. In Fig. 4, we show a visual comparison between the final active learning iteration (iteration 5) results, ASA, and manual counting. Where manual cell count, ASA based cell count, and deep learning (Unet) based cell count are reported for the test set (LU-17 mouse images). The ASA error rate on the test set (LU-17 mouse images) was 36.56 %. Additionally, cells count of the LU17 mouse in different sections is shown in Table III.

The evidence from this study suggests that the iterative deep learning based unbiased stereology method presented herein is much faster than the state-of-the-art stereology since human involvement was mainly reduced. The state-of-the-art stereology takes about 2-3 hours per ROI; however, the proposed method herein estimated time was approximately 20-30 minutes per ROI (including preparing masks using ASA, human verification, training deep learning for a single iteration, predicting on the test set, post-processing, and counting). Human involvement reduction was as follows:

1) Instead of creating initial labels manually (creating masks), an unsupervised algorithm (ASA) was utilized to create initial masks, then a user verification step to merely accept or reject image/mask based on the match to manual annotation.
2) Instead of relabeling active set images/masks predicted by deep learning in each iteration of the iterative deep learning process, the human-in-the-loop was only verifying the correctness of predicting masks, that is accepting or rejecting based on the match to the manual annotation as described earlier.

The generalisability of this study is subject to certain limitations. For instance, lack of enough data to best train deep learning models. Another limitation is user subjectivity in verifying predicted masks by deep learning in the active set. Notwithstanding the relatively limited data and user subjectivity constraints, this work offers valuable insights into using an existing unsupervised algorithms (ASA) to generate masks (labels) instead of human labeling (creating masks), then improving the model performance by iterative deep learning based unbiased stereology with a human-in-the-loop.

VII. Conclusion

This paper presents an iterative deep learning based unbiased stereology strategy that uses an existing unsupervised algorithm (ASA) masks as initial labels for training a deep convolutional neural network to segment and count cells on ROIs of NeuN stained images. The proposed method herein was able to achieve good results compared to the ASA cell counting, although ASA generated the initial labels (segmentation masks). Moreover, the proposed algorithm eliminates human effort in data labeling, where human work was merely to verify masks based on the corresponding manual annotation. Our approach has some drawbacks such as human-in-the-loop and user subjectivity which could be an obstacle with massive sets of images for verification. Iterative deep learning based unbiased stereology techniques in conjunction with initial labels (masks) from an existing algorithm showed encouraging results comparing to the current stereology counting method which is laborious, slow, and time-consuming to obtain for a significant amount of data.

VIII. 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. This research was partially supported by the National Science Foundation under award number (#1513126) and (#1746511).

References

[1] P. R. Mouton, *Unbiased stereology: a concise guide*. JHU Press, 2011.
[2] P. R. Mouton, H. A. Phoulady, D. Goldgof, L. O. Hall, M. Gordon, and D. Morgan, "Unbiased estimation of cell number using the automatic optical fractionator," *Journal of chemical neuroanatomy*, vol. 80, pp. A1–A8, 2017.
[3] M. Burke, S. Zangenehpour, P. R. Mouton, and M. Ptito, "Knowing what counts: unbiased stereology in the non-human primate brain," *Journal of visualized experiments: JoVE*, p. 27, 2009.
[4] M. West, L. Slomianka, and H. J. G. Gundersen, "Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator," *The Anatomical Record*, vol. 231, no. 4, pp. 482–497, 1991.
[5] A. G. Valdecasas, D. Marshall, J. M. Becerra, and J. Terrero, "On the extended depth of focus algorithms for bright field microscopy," *Micron*, vol. 32, no. 6, pp. 559–569, 2001.
[6] A. P. Bradley and P. C. Bamford, "A one-pass extended depth of field algorithm based on the over-complete discrete wavelet transform," in *Image and Vision Computing'04 New Zealand (IVCNZ'04)*. not found, 2004, pp. 279–284.
[7] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.
[8] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
[9] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
[10] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
[12] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
[15] C. B. Saper, "Any way you cut it: a new journal policy for the use of unbiased counting methods," *Journal of Comparative Neurology*, vol. 364, no. 1, pp. 5–5, 1996.
[16] K. Santacruz, J. Lewis, T. Spires, J. Paulson, L. Kotilinek, M. Ingelsson, A. Guimaraes, M. Deture, M. Ramsden, E. McGowan *et al.*, "Tau suppression in a neurodegenerative mouse model improves memory function," *Science*, vol. 309, no. 5733, pp. 476–481, 2005.
[17] T. L. Spires, J. D. Orne, K. SantaCruz, R. Pitstick, G. A. Carlson, K. H. Ashe, and B. T. Hyman, "Region-specific dissociation of neuronal loss and neurofibrillary pathology in a mouse model of tauopathy," *The American journal of pathology*, vol. 168, no. 5, pp. 1598–1607, 2006.
[18] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
[19] F. Chollet *et al.*, "Keras," https://github.com/fchollet/keras, 2015.
[20] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems,"

2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[21] D. Kinga and J. B. Adam, "A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[22] P. Y. Simard, D. Steinkraus, J. C. Platt *et al.*, "Best practices for convolutional neural networks applied to visual document analysis." in *ICDAR*, vol. 3, 2003, pp. 958–962.