Generalist 3D Cell Phenotyping for All-Type Tissues

by

Xinyi Gu

B.S., the University of Arizona (2019)

Submitted to the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Aug 2021

Author . . . . . Xinyi Gu

Department of Electrical Engineering and Computer Science

Aug 20, 2021

Certified by. . . . . .

Kwanghun Chung

Associate Professor of Chemical Engineering

Institute for Medical Engineering and Science (IMES)

Picower Institute for Learning and Memory

Thesis Supervisor

Accepted by. . . . . .

Leslie A. Kolodziejski

Professor of Electrical Engineering and Computer Science

Chair, Department Committee on Graduate Students

Generalist 3D Cell Phenotyping for All-Type Tissues

by

Xinyi Gu

Submitted to the Department of Electrical Engineering and Computer Science

On Aug 27, 2021, in partial fulfillment of the

requirements for the degree of

Master of Science in Computer Science and Engineering

Abstract

Tissue-clearing methods, light-sheet microscopy, and antibody labeling enable extracting cellular and subcellular information, producing large amount of image data needs to be analyzed. Hundreds of heterogeneous cell types were detected through the data obtained across species and types of tissues. We developed a novel approach that is generally applicable to a wide range of cell types in the large-scale 3D brain datasets, using a pipeline that performs accurate detection of cells regardless of image resolution, labeling pattern, and tissue processing techniques used. The pipeline is compatible with various labeling techniques including IHC, Fluorescence in situ hybridization (FISH), and genetic labeling and can be used for cellular level quantification in all types of tissues.

Thesis supervisor: Kwanghun Chung

Title: Associate Professor of Chemical Engineering, Institute for Medical Engineering and Science (IMES), and Picower Institute for Learning and Memory

Acknowledgments

I would like to thank my advisor, Prof. Kwanghun Chung for his strong support, leadership and guidance, which makes me a better researcher. He provided me with so many great insights and opportunities in my research life. Thanks to Dr. Juhyuk Park, Dr. Shaoyu Lin and Seo Woo Choi for providing me the testing data for my pipeline and helping me with the biological and chemical questions. Thanks to the peers in computational subgroup in the Chung Lab: Minyoung Eveyln Kim and Lee Kamentsky for mentoring and help in computational parts, Nick DiNapoli, Webster Guan and Dr. Joha Park for discussion during the meetings. I also want to thank all members in the Chung Lab: Nick Evans, Dae Hee Yun, Nicholas Haas, Srinu Pujari, Yuxuan Tian, Lauren DeLorenzo, Dr. Ji Wang, Dr. Jose Vargas Asencio, Vamsi Mangena, Jiho Shin, Jisoo Kim, Chuanxi Victor Zhao and Allison Christiansen, for being supportive and helpful labmates and friends. Finally, deepest thanks to my family for their unconditional love. Their encouragement and unwavering support have made it possible for me to become who I am today.

Contents

List of Figures

Chapter 1

Introduction

Structural and functional mapping of the brain is one of the most popular in various studies in neuroscience. A critical prerequisite in brain mapping is the precise identification and quantification of cells across the intact brain. Advancement in microscopy and sample preparation techniques provide an approach to study the molecular and cellular complexity of the central nervous system. State-of-the-art tissue-clearing methods, light-sheet microscopy and antibody labeling enable extracting cellular and subcellular information on complex mammalian bodies and large human specimens [1]. Such technologies left researchers with a large amount of image data which is laborious or impossible to perform manual analysis, creating a high demand for automated cell quantification.

Quantification based on cell body segmentation and classification methods using cell markers has been developed, most of which trade-off between flexibility and automation. These methods range from fully manual labeling [2], to non-learning-based algorithms [3], to learning-based algorithms trained for every specific task [4][5], or even as one sub-task in a fully automated deep learning approach [6]. Automation reduces the cumbersome human effort to increase the scalability in big datasets as well as the efficiency of the analysis. However, most of the existing fully automated methods are trained with massive amounts of annotated data. Their application has been largely restricted to the specialized data as the training sets.

We seek to address this generalization problem by avoiding completely relying on learning-based approaches. A non-learning-based algorithm that is generally applicable to detecting and categorizing the cells in images from multi-round immunostaining is developed. Composed by a nuclei detection module and a cell phenotyping step, our approach is capable to perform accurate detection and quantification of cell subtypes based on the corresponding markers, regardless of the staining patterns in most of the tissue types, including mouse and human brain tissue as well as human brain organoids.

The main contributions of this thesis are as follows:

- We introduce the Generalist 3D Cell Phenotyping (GCP) approach, a pipeline composed of the nuclei segmentation and an expression level quantification algorithm that detects cells with different immunohistochemistry (IHC) subcellular expression patterns, tissue regional cell density.

- Our approach is compatible with FISH and IHC 3D images of different tissue types, tissue processing techniques, and image resolutions.

- GCP is capable of processing TB-size datasets.

Roadmap Chapter 2 presents a brief overview of central concepts in cellular and molecular profiling of the brain, focusing on its applications in neuroscience. We describe automated approaches to image analysis combined with tissue-clearing and light-sheet microscopy that enable 3D imaging of immunostained intact tissues to perform scalable tissue histology. We also outline

previous methods for cell detection, which is one of the most critical steps of cellular and molecular profiling of the brain. Chapter 3 introduces GCP, specifying the structure of this pipeline. Chapter 4 reports three experiments comparing GCP with previous approaches to demonstrate the generality, robustness, and scalability of GCP across different subcellular expression patterns, tissue regional cell density, tissue types, tissue processing techniques, and image resolutions. Finally, we conclude and discuss possible applications in Chapter 5 and suggest the direction for future work.

Chapter 2

Related Works/ Background

2.1 Tissue-Clearing and Light-Sheet Microscopy

Classification/phenotyping cells

Structural and functional mapping of the brain is crucial in various neuroscience studies, especially in understanding brain function and dysfunction. A critical prerequisite in brain mapping is the precise quantification of cells by cell identification and classification across the whole brain. Scientists obtain plenty of datasets with advanced microscopy and sample preparation techniques to study molecular and cellular complexity in the central nervous system. State-of-the-art tissue-clearing methods, light-sheet microscopy, and antibody labeling enable extracting cellular and subcellular information in complex mammalian bodies and large human specimens [1]. Such technologies leave researchers with a large amount of image data which is laborious or impossible to perform manual analysis, creating a high demand for automated cell quantification.

2.2 Automated Image Analysis

Challenge in cell detection and classification

Quantification methods based on cell body segmentation and classification for cell fluorescent images have been developed, most of which trade-off between flexibility and

automation. These methods range from fully manual labeling [2] to non-learning-based algorithms [3] to learning-based algorithms trained for every specific task [4][5]. Some fully automated deep learning approaches are developed for more complicated problems but cover cell detection as one sub-task [6]. Fully manual labeling using Corel Draw software or Cell Counter plugin in ImageJ [26] is usually used to obtain ground truth for training and evaluation. While retaining the highest accuracy, manual labeling is the least time-efficient approach. Aiming to enable fast cell and nuclei detection, non-learning-based algorithms include distance transform, morphology operation, Laplacian of Gaussian filtering, and radial symmetry-based voting [7] provide seed-seeking metrics near the object centroids. While nuclei detection tasks are accomplished with high performances, accuracy in cell detection remains relatively low due to the complex nature of cell images. Thus, supervised learning methods with both high automation and accuracy have attracted much attention recently. SVM, random forests, and deep neural networks have become popular approaches to nuclei or cell detection problems [7].

Automation reduces cumbersome human efforts to enhance the efficiency of the analysis and increase the scalability in big datasets. However, delineate choices of detection approaches are demanded to specific datasets. At the same time, the precision of detection results mainly depends on the morphology features in the data. For instance, non-learning-based methods usually fail for densely packed regions in the IHC images when the cell-type-specific markers located in the cell body. Similar problems emerge when dyes only stain the cell cytoplasm due to inaccurate features of unstained nuclei areas inside the cells. Fig.1 shows an example of this kind of staining. This problem can be relieved by manual adjustment of the parameters, which trades off the automation.

More and more scientists try to find alternative solutions in learning-based, fully automated approaches. Nevertheless, most fully automated methods are trained with massive amounts of

specific annotated data. Their application has been restricted mainly to the specialized data used as the training sets. These methods fail when the morphology of the objects is significantly different from the training data. Furthermore, the enormous quantity of the data generated for large tissues becomes a bottleneck [8] with the intensive computation requirement of learning-based approaches that expands the processing time for the whole-tissue analysis.

That is to say, a novel non-learning-based approach needs to be developed in consideration of the throughput and the fact that heterogeneous staining patterns hinge the generality of the learning-based techniques.

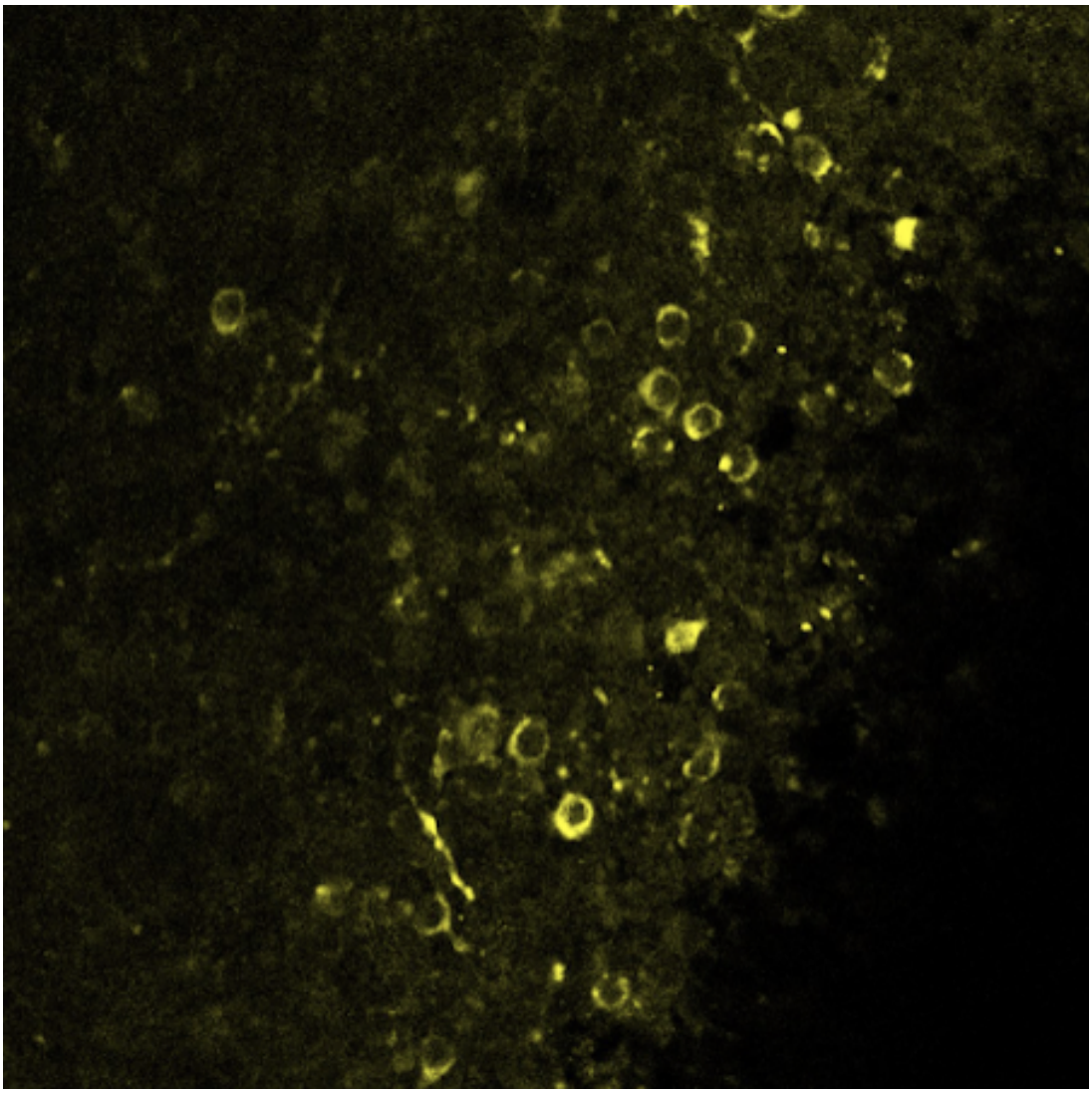

Fig. 1 oRGC stained by in human brain organoid tissue

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Generalist 3D cell phenotyping pipeline

In this chapter, we describe GCP, our novel, versatile and generalizable cell detection, and classification algorithm. First, we outline the overall components of our pipeline, which performs nuclei detection and cell classification to achieve a generalizable cell detection. We then provide details about the two core steps and introduce the parameters that need to be decided ahead of the whole pipeline. We further show that the automated steps can be manually adjusted as well. Lastly, we demonstrate the user interface of the pipeline.

3.1 Overall components of the pipeline

Most of the existing fully automated cell detection and classification methods are trained with massive amounts of annotated data based on the morphology. Their application has been largely restricted to specialized training data. We seek to address this generalization problem by avoiding completely relying on learning-based approaches. A non-learning-based algorithm that is generally applicable to detecting and categorizing cells in the images from multi-round immunostaining is developed.

In accompany with our available IHC datasets that collect three channels for three specific staining, our approach stands on an crucial fact that the existence of membrane segregates each

cell. That is to say, the cell components inside an individual cell never cross the boundary of this cell. Fig 1-1 shows this concept, in which there is always a shallow shell outside each nucleus that only contains its own corresponding cell components, and this rationalize our idea of phenotyping cells based on the quantitative measurement of markers inside this region. Fig 1-2 shows the schematic of GCP. The generalizable pipeline is composed of 2 objectives:

Nuclei Detection and Segmentation Algorithm First, we perform nuclei detection to locate all cells in the tissue volume. We then use the probability map to segment the accurate edges of each nucleus and define regions of interest (ROI) based on the segmentation result. Defining ROI based on nuclei detection and segmentation is a crucial part of this pipeline. A seeded watershed algorithm based on the geometric morphological properties using curvatures [3] that provides fine segmentations even in densely packed regions is used here.

Cell Classification Algorithm Second, we classify the types of cells according to multi-channel co-expression in the ROI enabled by the boundary of each segmented nucleus. The cell classification is achieved by quantifying the intensity of each voxel inside the ROI based on cell-type-specific protein/mRNA expression. An automatic thresholding approach is used to classify the detected nuclei into specific cell types without laborious training.

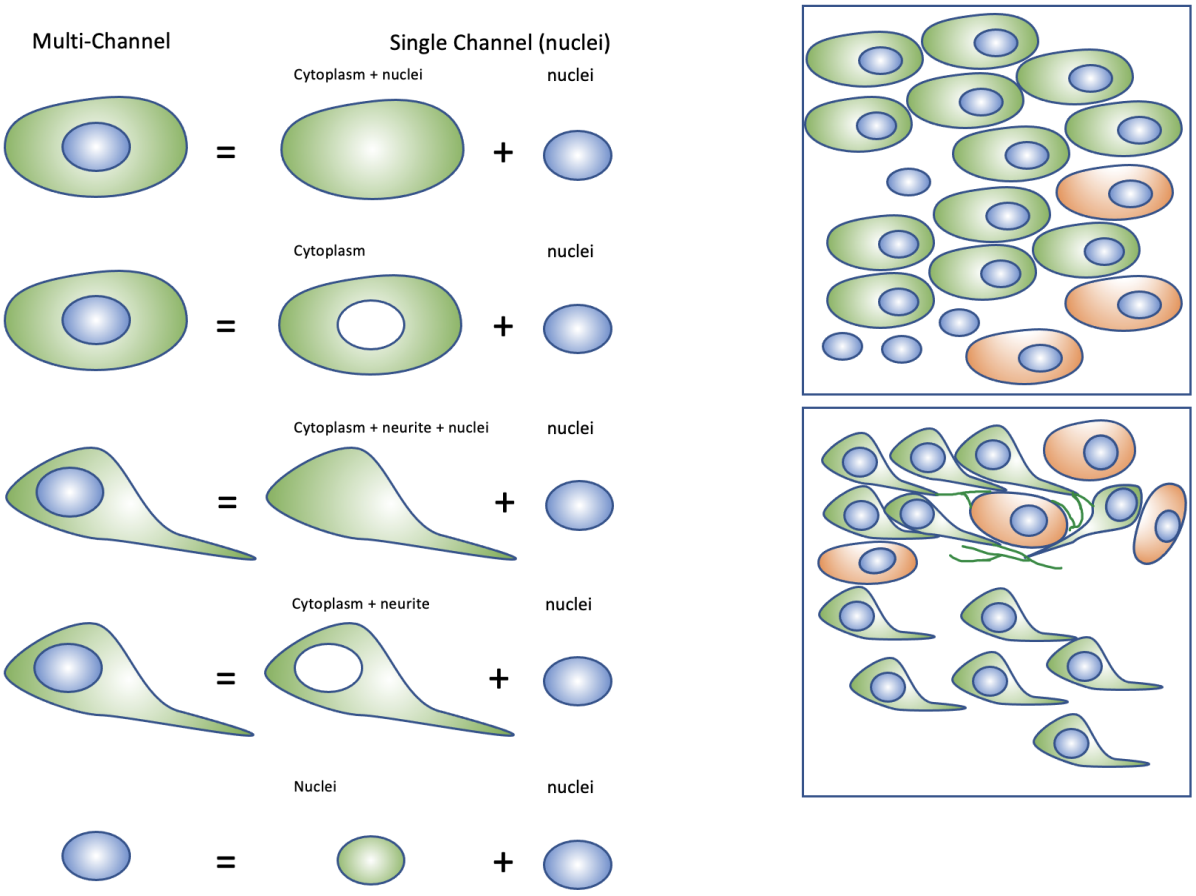

Fig. 2 Conceptual design of the pipeline

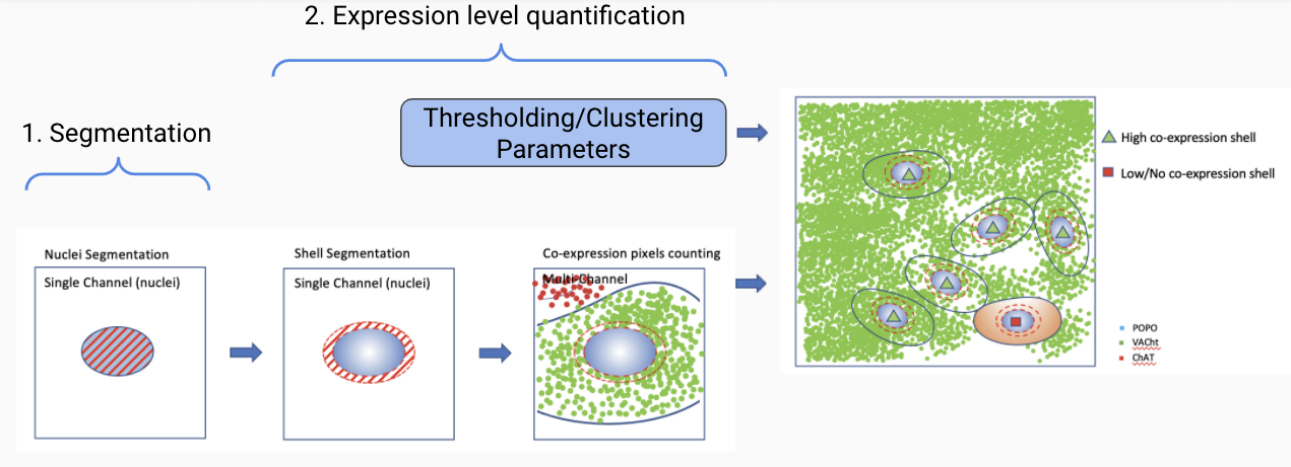

Fig. 3 Schematic of GCP

3.2 Nuclei detection and segmentation

Nuclei detection and segmentation algorithms for IHC images have been continuously improved in past years. Plenty of learning and non-learning-based nuclei detection and segmentation algorithms are practicable for datasets of various types of cells [3][7][9][10][11]. A notable advantage in precision for nucleus data is usually observed comparing to the detection and segmentation results in cell images.

We compared available approaches based on accuracy and efficiency and deliberately chose one of each from CNNs, supervised learning algorithms, and non-learning-based algorithms: (1) Cellpose: Cellpose is a generalist, deep learning-based segmentation method newly developed in 2021 [12]. The network trained on a highly varied 2D cell image dataset predicts the spatial gradients based on U-Net architecture. To extend the network to 3D segmentation, Cellpose predicts and averages the gradient on XY, XZ, and YX slices independently with the trained 2D model. The prediction precision of Cellpose outperforms Mask R-CNN, Stardust, and U-Net. (2) SVM: Among supervised learning algorithms, the SVM-based methods achieve relatively high precision. These approaches detect redundant candidate regions and apply SVM-based supervised learning to assess the candidate regions [13][14][15][16]. (3) Curvature-based seeded watershed: Atta-Fosu et al. [3] compute eigenvalues of the shape matrix to generate accurate seeds that inherit the original shape of their respective cells. Their non-learning-based algorithm outperforms popular approaches like MINS [17], SMMF [18], and CellSegm [19] and reaches a Rand Index score of 97.05%.

As the nuclei are usually similar in shape under a specific resolution, both non-learning-based and learning-based algorithms can be used for nuclei segmentation. In relatively low-

resolution datasets with more uniform nuclei morphology, non-learning-based algorithms possess comparable performance to CNN-based methods. Here in this thesis, we use the curvature-based seeded watershed for the datasets collected by 4x objective.

3.3 Cell classification

The second step of GCP is cell classification based on molecular phenotyping. Immunohistochemistry is widely used in research to understand the distribution, localization, and morphology of the cells. Under the consideration of the typical morphology of IHC data, we use the 1-voxel-thick shell out of the segmented nucleus to define the ROI of the individual cell in IHC image data. The average intensity representing the cell-type-specific protein expression level of the voxels in the ROI, will be calculated respectively.

$$I_{avg} = \frac{I_1 + I_2 + \ldots\ldots + I}{\#\, of\ voxels\ in\ ROI} \tag{1}$$

The automated thresholding method is applied to classify the cells into two groups: Cells with high average intensities are classified as positive cells expressing the marker in the channel, and cells that do not express this specific marker exhibit low average intensities. We expect to automatically find the thresholds to distinct no marker-expressing, low marker-expressing, and high marker-expressing groups among cells, so a revised valley-emphasis method [20] is used to the statics without a precise bimodal distribution. The Valley-emphasis method is capable of handling multi-level thresholding. For M-1 thresholds, the optimal thresholds $\{t_1^*, t_2^*, \ldots, t_{M-1}^*\}$are given as:

$$\{t_1^*, t_2^*, \ldots, t_{M-1}^*\} = argmax_{0<t_1<t_2<\ldots<t_{M-1}<L-1}\{(1 - \sum_{j=1}^{M-1} h(t_j))(\sum_{k=1}^{M} p_k \mu_k^2)\} \quad (2)$$

Where $1 - \sum_{j=1}^{M-1} h(t_j)$ is the weight of each between-class variance of image term calculated by $\sum_{k=1}^{M} p_k \mu_k^2$ in equation (2).

Besides the automated approach, a manual adjustment of the threshold is available in the pipeline to improve the accuracy.

While the IHC signal being sensitive and reflective enough of analyte concentrations over a broad range [21], Fluorescence in situ hybridization is also widely used as a cytogenetic method in tissue profiling. DNA fragments incorporated with fluorophore-coupled nucleotides effectively map genes and polymorphic loci onto metaphase chromosomes to construct a physical genome map [22][23]. To pursue the generality of our method, we extend the pipeline to be compatible with IHC and FISH datasets without tedious parameters adjustment.

Image data collected with FISH has the cell-type-specific markers expressed in nuclei. We will use a 2-voxel-thick shell covering the 1-voxel-thick part outside the nuclei and a 1-voxel-thick part overlapping the nuclei.

3.4 User interface

The GCP pipeline is initially written in Python. Users can run their data through the pipeline with Jupyter as well as the command line.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Experiments

4.1 Human brain cell classification with immunohistochemistry staining

4.1.1 Data

One of the recent image volumes obtained using a customized oblique selective plane illumination microscopy (oSPIM) with 2x objective in the Chung Lab is for mELAST (MAP-ELAST) processed human brain tissue multiplexed with YOYO1, NeuN and Somatostatin. The tissue was expanded 2x and resulted in a 4x expanded image dataset. NeuN only stains cytoplasm, which results in a varied morphology in the data compared to those markers located in both nuclei and cytoplasm. The unstained nuclei located inside the cytoplasm make it difficult for traditional non-learning-based algorithm to detect each NeuN+ cells.

The 4x expanded dataset maintains an even texture inside the nuclei, which results in a sufficiency of defining a shell outside the nucleus with non-learning-based nuclei segmentation methods. We used the curvature-based seeded watershed to segment nuclei, and the boundary of each nucleus helps determine the ROI in step 2 of our pipeline.

The data we have for this experiment has a size of 2048x5762x14203. We cropped out a sample volume of size 16x256x256 and performed manual labeling to generate ground truth to evaluate the precision.

4.1.2 Baseline Methods

We choose curvature-based seeded watershed as our benchmark since it is one of the state-of-the-art non-learning-based algorithms that enable fast and accurate object detection and segmentation in biomedical image datasets. The algorithm improves the traditional intensity-based watershed and succeeds in most situations, including tissues with closely packed cells [3].

4.1.3 Evaluation

Manual labeling performed in imageJ with the cell counter plugin are used as ground truth for cell detection. Fig. 4-1-1,2 and3 visualize the detected result on one of the 2D planes inside the 3D volume, qualitatively comparing the performance of different approaches. The average precision metric (AP) computed from $AP = \frac{TP}{TP+FP+FN}$ is 0.5 for the test volume processed by GCP and 0.4 by curvature-based seeded watershed.

4.1.4 Results

GCP is capable of performing relatively high accurate cell detection in the human brain tissue datasets. The detection result is still reliable even in the dataset with unusual cell morphology, which causes the traditional non-learning-based methods to fail. We proved GCP is more generalizable in the cell detection task for the data with different staining patterns than other

algorithms. Which is to say, it can detect cells in the datasets where the cells are phenotyped by the markers located in cytoplasm only, or nucleus only, or both.

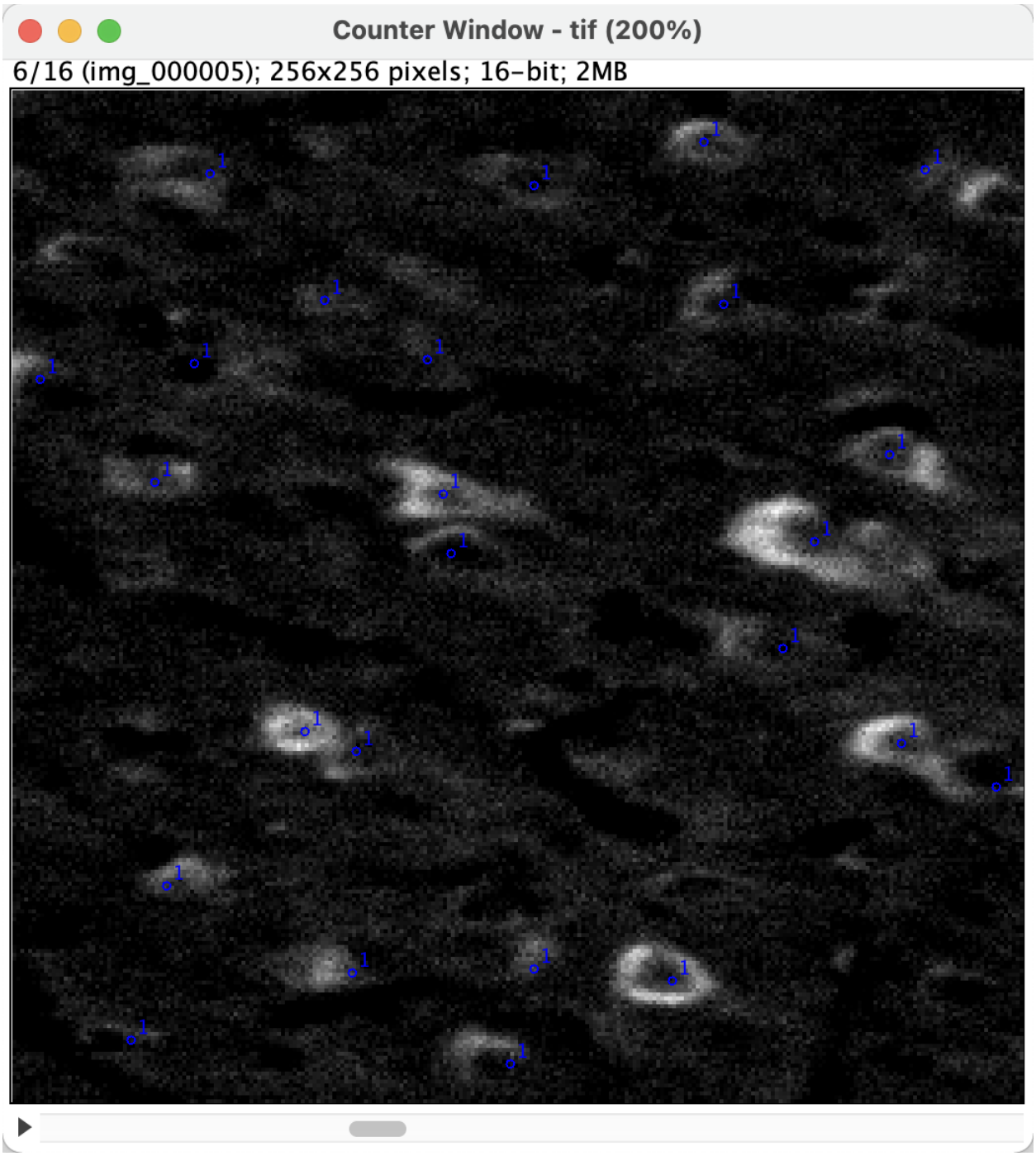

Fig. 4-1-1 Ground truth for NeuN+ cell detection, manual labeling, number of labeled NeuN+ cells (blue ‘o’): 25

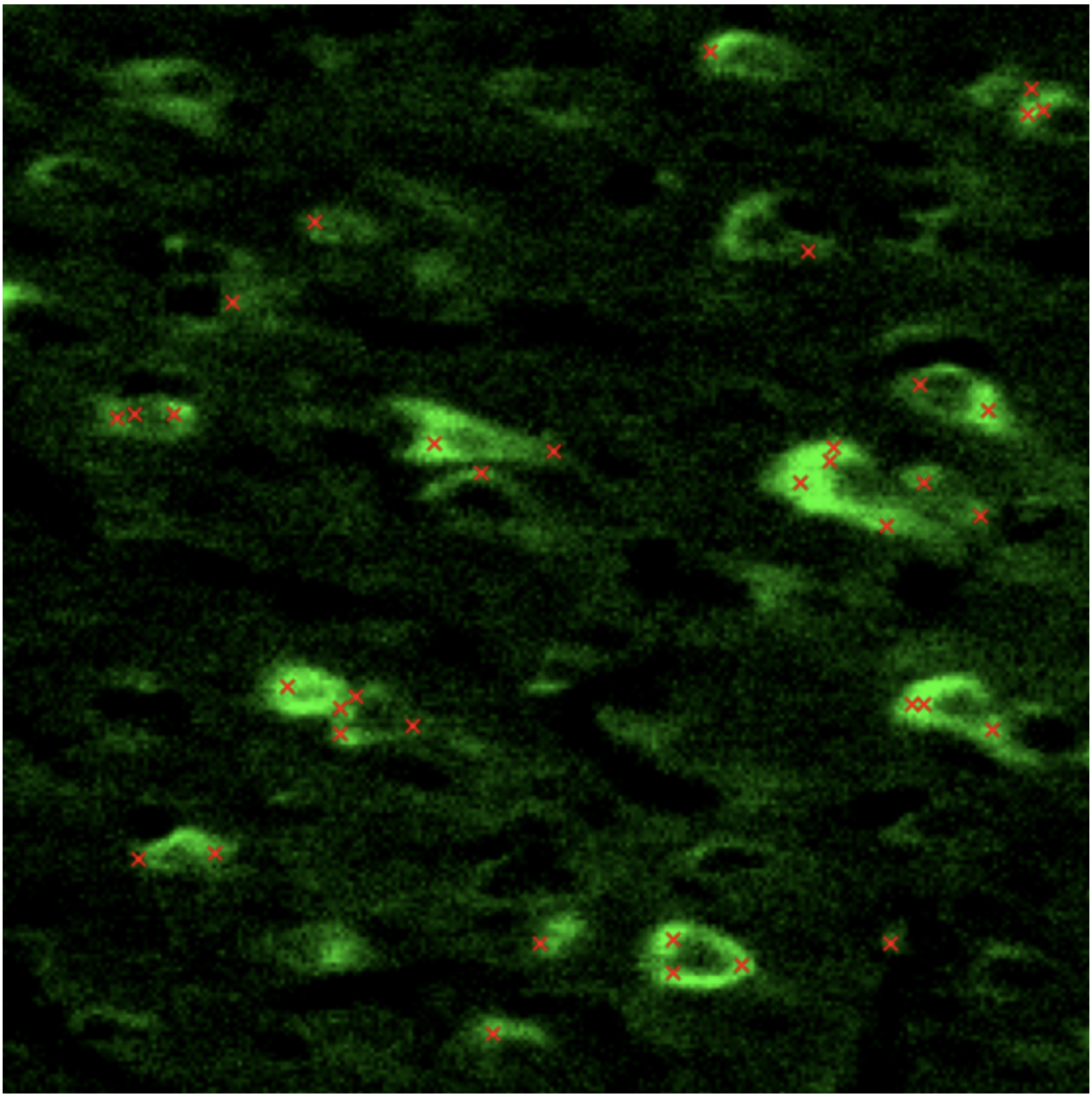

Fig. 4-1-2 Curvature-based seeded watershed detection results of NeuN+ cells, threshold of probability map: 0.9, number of detected NeuN+ cells (red 'x'): 37

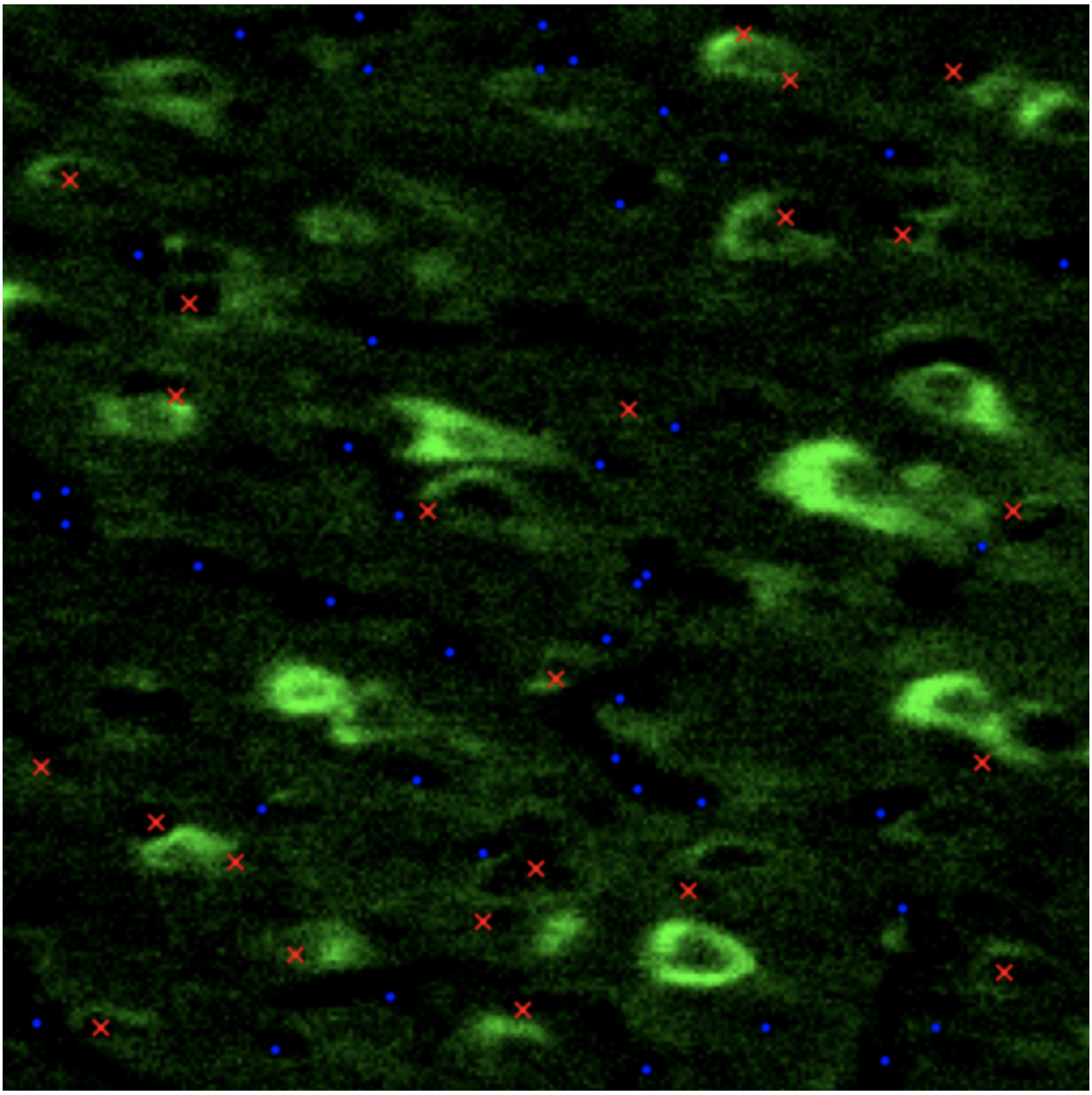

Fig. 4-1-3 GCP detection results of NeuN+ cells, threshold of average intensity: 64, number of detected NeuN+ cells (red ‘x’): 23

4.2 Fish dataset cell classification for mouse brain

4.2.1 Data

The FISH dataset we used to test the compatibility of our pipeline is a mouse brain labeled with Avp. Nuclei were labeled by Syto16 and taken under 405 nm channel accompanied with three other FISH labeling by oSPIM with 4x objective.

4.2.2 Evaluation

Ground truth of the cell detection tasks was generated through manual labeling with the cell counter plugin in ImageJ. Fig. 4-2-1,2, and 3 visualize the detected result on one of the 2D planes inside the 3D volume. Qualitatively comparing the performance of the approaches, the precision is similar between two approaches.

4.2.3 Results

Morphology of the FISH labeling is more even compared to some of the IHC datasets. Most of the markers are located inside the nuclei and are blob-like. Curvature-based seeded watershed generated a similar result as GCP, and the accuracy of both methods are high in the area with relatively low cell density area.

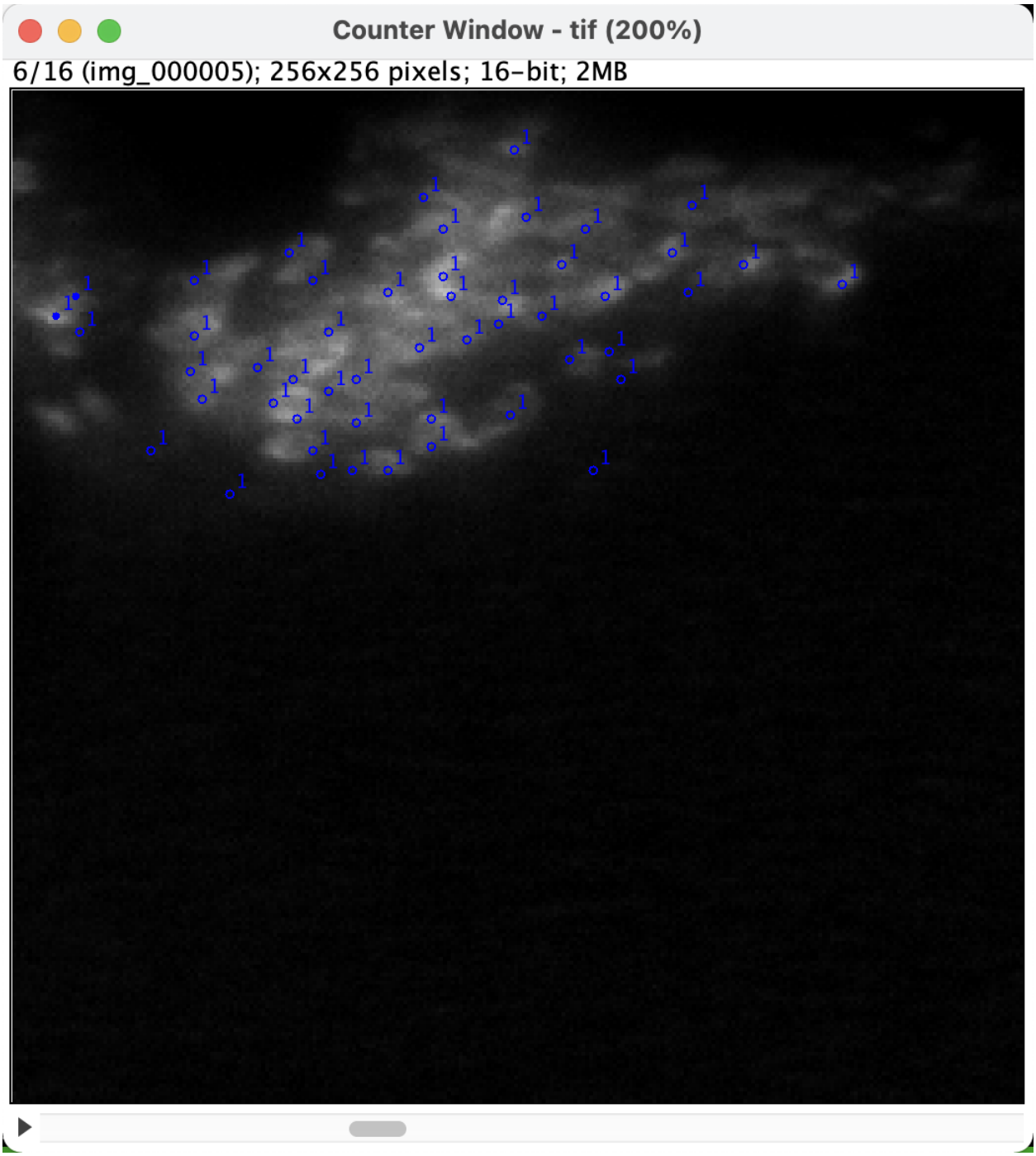

Fig. 4-2-1 Ground truth for Avp+ cell detection, manual labeling, number of labeled Avp+ cells (blue ‘o’): 50

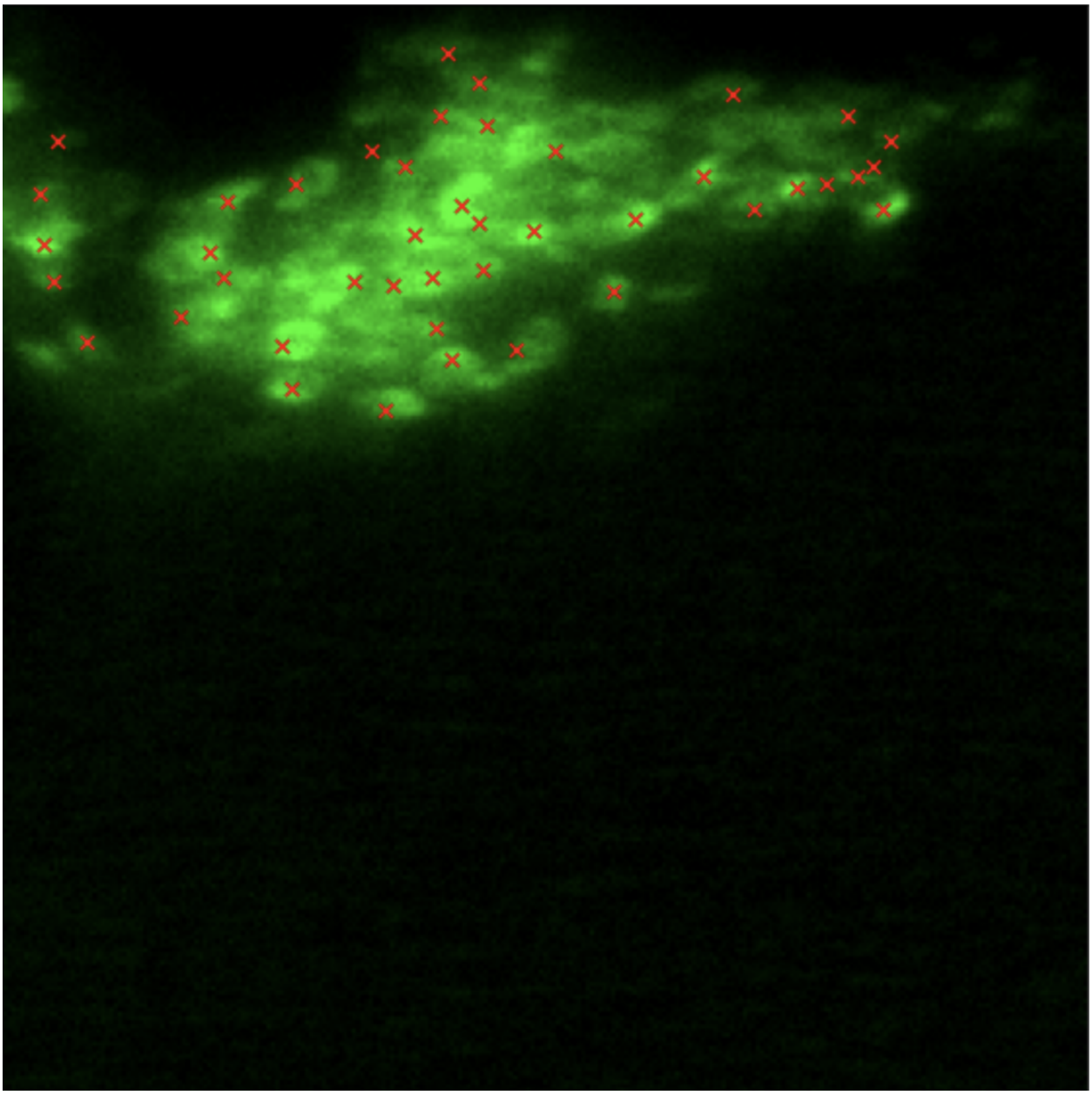

Fig. 4-2-2 Curvature-based seeded watershed detection results of Avp+ cells, threshold of probability map: 0.5, number of detected NeuN+ cells (red ‘x’): 43

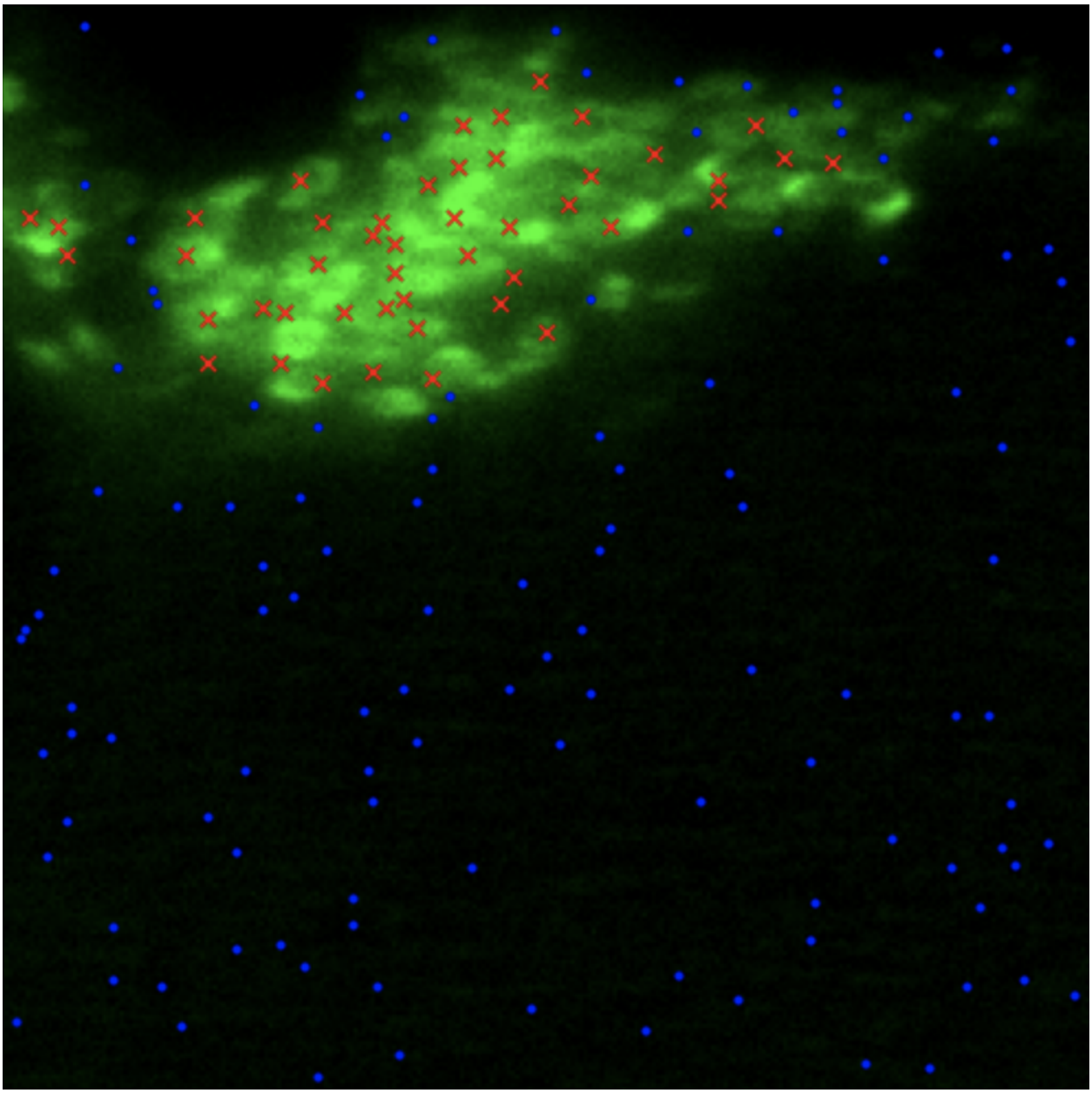

Fig. 4-2-3 GCP detection results of Avp+ cells, threshold of average intensity: 1324, number of detected Avp+ cells (red ‘x’): 46

4.3 Microglia detection in mouse brain with immunohistochemistry staining

4.3.1 Data

To demonstrate the compatibility of GCP to a wide range of tissue and tissue processing techniques, we further run our pipeline on an intact mouse brain dataset. Nuclei in this dataset are stained by Syto16, the other two channels are for NeuN and IBA1 staining. In the first section of the experiment, we already processed the NeuN. We analyzed microglia, which are IBA1+ cells in this part.

4.3.2 Evaluation and results

The evaluation relies on manual labeling and uses curvature-based seeded watershed algorithm as the baseline, which is the same as previous steps. The AP of GCP is 0.42, and curvature-based watershed segmentation has an AP of 0.36. Performances of both methods remain poor, indicating the challenge in detecting cells with heterogeneous morphology.

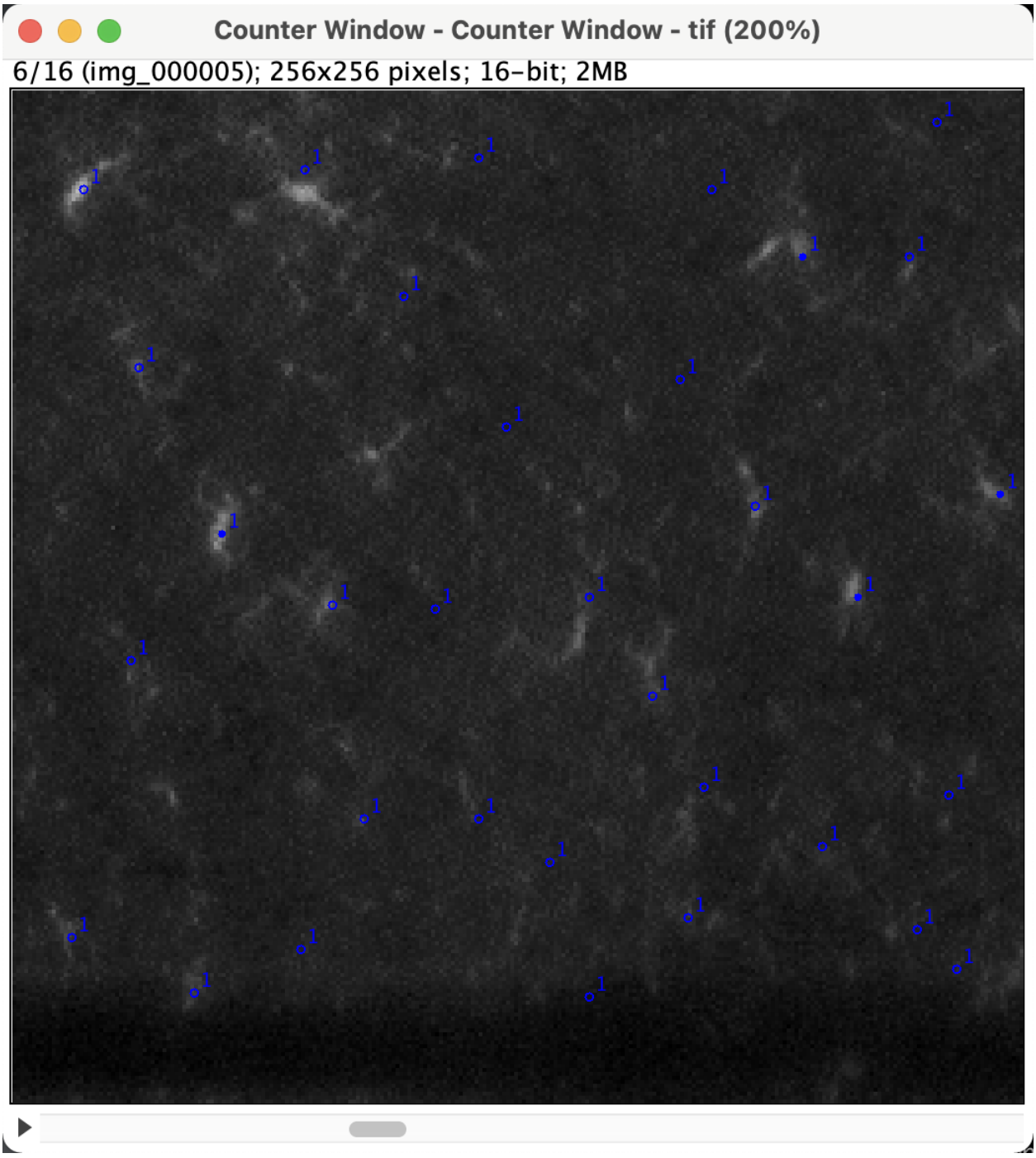

Fig. 4-3-1 Ground truth for microglia detection, manual labeling, number of labeled IBA1+ cells (blue ‘o’): 33

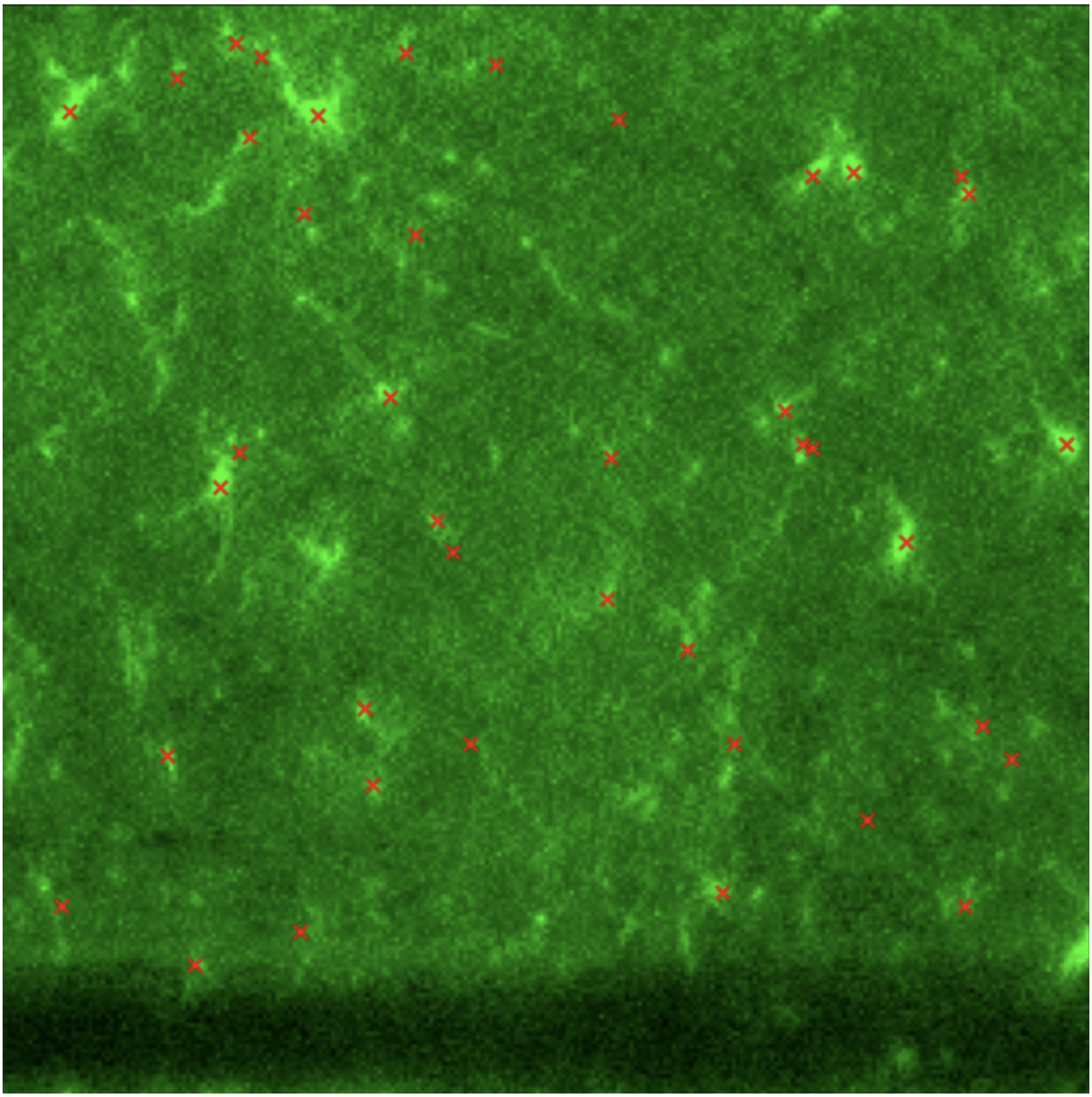

Fig. 4-3-2 Curvature-based seeded watershed detection results of IBA1+ cells (microglia), threshold of probability map: 0.8, number of detected NeuN+ cells (red ‘x’): 41

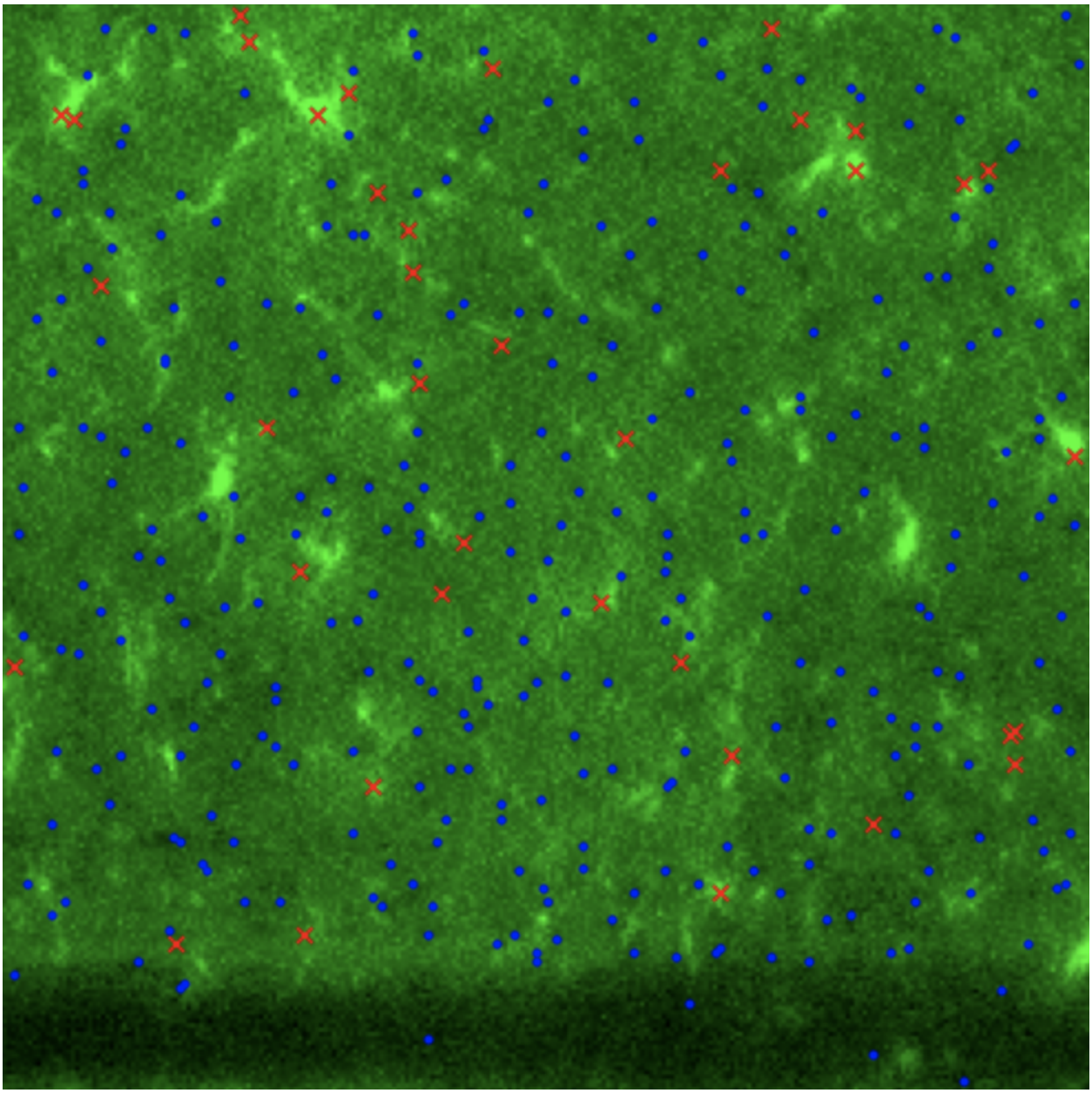

Fig. 4-3-3 GCP detection results of IBA1+ cells (microglia), threshold of average intensity: 528, number of detected IBA1+ cells (red ‘x’): 38

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 5

Conclusion

We presented the Generalist 3D Cell Phenotyping approach, a novel algorithm for cell detection and classification for microscopy image volumes that improves the generality and reliability of traditional cell detection methods. We demonstrated our pipeline could accurately identify cells in most situations without tedious training and laborious manual adjustment. Theoretically, our approach would further aid manual labeling in some cases when most of the other methods, even manual detection, failed, for example, when cells are crowded together, and the boundary of the cells is difficult to determine.

In each dataset listed in experiments, we saw that GCP could achieve relatively reliable cell detection and classification for most cells even with rare and heterogeneous morphology. However, the accuracy still needs further improvement. There are two significant limitations for GCP:

(a) First, ROI in the average intensity calculation heavily relies on the nuclei segmentation results. A precise boundary for each nucleus is highly demanded to make sure we classify the cells strictly based on the markers located in each individual cell. While a segmentation results larger than the actual size of the nucleus results in counting markers inside other cells, a smaller one omits the markers in the cytoplasm. Inaccurate segmentation results cause an incorrect

identification of the individual cells and harm the automated thresholding based on statistics of the whole cell population.

(b) Second, chromatic aberration while obtaining the image dataset shifts the nuclei locations, resulting in inaccurate ROI. Deviation in phase or optical path length from the ideal wavefront and optical aberration cause diffractions in imaging [27][28][29]. Based on the types of chromatic aberration, the diffractions can be mitigated by shifting and scaling. Introducing machine-aided co-registration can be the next step to improve the precision of the pipeline.

More and more informative but complex data are obtained with the development of tissue preparation and imaging technology. The Advancement of a cell detection method is extremely important in understanding a large amount of data after acquiring it. As the increasingly heterogeneous cell morphology is being observed, it seems likely that generality will continue to act as one of the most desired prior in cell phenotyping algorithms.

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] Ueda, H. R. *et al.* Tissue clearing and its applications in neuroscience. *Nat. Rev. Neurosci.* **21**, 61–79 (2020).

[2] Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012).

[3] Atta-Fosu, T. et al. 3D clumped cell segmentation using curvature based seeded watershed. J. Imaging 2, (2016).

[4] Sommer, C., Straehle, C., Koethe, U. & Hamprecht, F. A. Ilastik: interactive learning and segmentation toolkit. IEEE International Symposium on Biomedical Imaging, 230–233 (2011).

[5] Berg, S. et al. ilastik: interactive machine learning for (bio)image analysis. Nat. Methods 16, 1226–1232 (2019).

[6] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. arXiv 18, (2019).

[7] F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review," *IEEE Rev. Biomed. Eng.*, vol. 9, no. c, pp. 234–263, 2016, doi: 10.1109/RBME.2016.2515127.

[8] Y. Frégnac, “Big data and the industrialization of neuroscience: A safe roadmap for understanding the brain?,” *Science (80-.).*, vol. 358, no. 6362, pp. 470–477, 2017, doi: 10.1126/science.aan8866.

[9] U. Adiga, B. Bell, L. Ponomareva, D. Taylor, R. Saldanha, S. Nelson, and T. Lamkin, “Mapping infected cell phenotype,” IEEETrans. Biomed. Eng., vol. 59, no. 8, pp. 2362–2371, Aug. 2012.

[10] P. Quelhas, M. Marcuzzo, A. Mendonca, and A. Campilho, “Cell nuclei and cytoplasm joint segmentation using the sliding band filter,” IEEE Trans. Med. Imag., vol. 29, no. 8, pp. 1463–1473, Aug. 2010.

[11] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, “Accurate segmentation of cervical cytoplasm and nuclei based on multi-scale convolutional network and graph partitioning,” IEEE Trans. Biomed. Eng., vol. 62, no. 10, pp. 2421–2433, Oct. 2015.

[12] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, “Cellpose: a generalist algorithm for cellular segmentation,” Nat. Methods, vol. 18, no. January, 2020, doi: 10.1038/s41592-020-01018-x.

[13] R. Bise and Y. Sato, “Cell detection from redundant candidate regions under non-overlapping constraints,” IEEE Trans. Med. Imag., vol. 34, no. 7, pp. 1417–1427, Jul. 2015.

[14] H. Su, F. Xing, J. D. Lee, C. A. Peterson, and L. Yang, “Automatic myonuclear detection in isolated single muscle fibers using robust ellipse fitting and sparse representation,” IEEE Trans. Comput. Biol. Bioinfo., vol. 11, no. 4, pp. 714–726, Jul. 2014.

[15] H. Su, F. Xing, J. D. Lee, C. A. Peterson, and L. Yang, “Learning based automatic detection of myonuclei in isolated single skeletal muscle fibers using multi-focus image fusion,” in Proc. IEEE Int. Symp. Biomed. Imag., Apr. 2013, pp. 432–435.

[16] K. Z. Mao, P. Zhao, and P. H. Tan, “Supervised learning-based cell image segmentation for p53 immunohistochemistry,” IEEE Trans. Biomed. Eng., vol. 53, no. 6, pp. 1153–1163, Jun. 2006.

[17] X. Lou, M. Kang, P. Xenopoulos, S. Muñoz-Descalzo, and A. K. Hadjantonakis, “A rapid and efficient 2D/3D nuclear segmentation method for analysis of early mouse embryo and stem cell image data,” Stem Cell Reports, vol. 2, no. 3, pp. 382–397, 2014.

[18] J. Cheng and J. C. Rajapakse, “Segmentation of Clustered Nuclei With Shape Markers and Marking Function,” in IEEE Transactions on Biomedical Engineering, vol. 56, no. 3, pp. 741-748, March 2009, doi: 10.1109/TBME.2008.2008635.

[19] E. Hodneland, T. Kögel, D. M. Frei, H. H. Gerdes, and A. Lundervold, “CellSegm - a MATLAB toolbox for high-throughput 3D cell segmentation,” Source Code Biol. Med., vol. 8, pp. 1–24, 2013, doi: 10.1186/1751-0473-8-16.

[20] J. L. Fan and B. Lei, “A modified valley-emphasis method for automatic thresholding,” Pattern Recognit. Lett., vol. 33, no. 6, pp. 703–708, 2012, doi: 10.1016/j.patrec.2011.12.009.

[21] D. E. Carvajal-Hausdorf, K. A. Schalper, V. M. Neumeister, and D. L. Rimm, “Quantitative measurement of cancer tissue biomarkers in the lab and in the clinic,” Lab. Investig., vol. 95, no. 4, pp. 385–396, 2015, doi: 10.1038/labinvest.2014.157.

[22] P. R. Langer-Safer, M. Levine, and D. C. Ward, “Immunological method for mapping geneson Drosophila polytene chromosomes,” Proc. Natl. Acad. Sci. U.S.A. vol. 79, no. 14, pp. 4381-5438, 1982, doi: 10.1146/annurev-genom-090413-025346

[23] J. B. Lichter, M. J. Difilippantonio, A. J. Pakstis, P. J. Goodfellow, D. C. Ward, and K. K. Kidd, “Physical and genetic maps for chromosome 10,” Genomics vol. 16, pp. 320–324, 1993, doi: 10.1006/geno.1993.1192

[24] T. Falk et al., “U-Net: deep learning for cell counting, detection, and morphometry,” Nat. Methods, vol. 16, no. 1, pp. 67–70, 2019, doi: 10.1038/s41592-018-0261-2.

[25] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, “NIH Image to ImageJ: 25 years of image analysis,” Nat. Methods, vol. 9, no. 7, pp. 671–675, 2012, doi: 10.1038/nmeth.2089.

[26] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, “NIH Image to ImageJ: 25 years of image analysis,” *Nat. Methods*, vol. 9, no. 7, pp. 671–675, 2012, doi: 10.1038/nmeth.2089.

[27] P. Törők, S. J. Hewlett, and F. Varga, “The role of specimen-induced spherical aberration in confocal microscopy,” *J. Microsc.*, vol. 188, no. 2, pp. 158–172, 1997, doi: 10.1046/j.1365-2818.1997.2440802.x.

[28] M. Schwertner, M. J. Booth, and T. Wilson, “Characterizing specimen induced aberrations for high NA adaptive optical microscopy,” *Opt. Express*, vol. 12, no. 26, p. 6540, 2004, doi: 10.1364/opex.12.006540.

[29] S. F. Gibson and F. Lanni, “Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy,” *J. Opt. Soc. Am. A*, vol. 9, no. 1, p. 154, 1992, doi: 10.1364/josaa.9.000154.