



**HAL**  
open science

# A General framework for Neuron Instance Segmentation based on Deep Learning

Huaqian Wu

► **To cite this version:**

Huaqian Wu. A General framework for Neuron Instance Segmentation based on Deep Learning. Computer Vision and Pattern Recognition [cs.CV]. Université Paris-Saclay, 2022. English. NNT : 2022UPAST174 . tel-04354820

**HAL Id: tel-04354820**

**<https://theses.hal.science/tel-04354820>**

Submitted on 20 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A General framework for Neuron Instance Segmentation based on Deep Learning

*Un cadre général pour la segmentation d'instance de neurones basé sur l'apprentissage profond*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n° 575  
Electrical, Optical, Bio-Physics and Engineering (EOBE)  
Spécialité de doctorat: Imagerie et Physique Médicale  
Graduate School : Sciences de l'Ingénierie et des Systèmes  
Réfèrent: Faculté des sciences d'Orsay

Thèse préparée au **Laboratoire des Maladies Neurodégénératives : mécanismes, thérapies, imagerie** (Université Paris-Saclay, CEA, CNRS),  
sous la direction de **Thierry DELZESCAUX**,  
directeur de recherche au CEA-MIRCen

**Thèse soutenue à Fontenay-aux-Roses,  
le 19 décembre 2022, par**

**Huaqian WU**

## Composition du jury

Membres du jury avec voix délibérative

### **Cyril POUPON**

Directeur de recherche, CEA, Université Paris-Saclay

### **Grégoire MALANDAIN**

Directeur de recherche, INRIA, Université Côte d'Azur

### **Cédric WEMMERT**

Professeur des universités, Université de Strasbourg

### **Cédric CLOUCHOUX**

Directeur général, WITSEE

### **Christophe DESTRIEUX**

Professeur des universités - praticien hospitalier, Université de Tours

Président

Rapporteur & Examineur

Rapporteur & Examineur

Examineur

Examineur

**Title:** A general framework for neuron instance segmentation based on Deep Learning

**Keywords:** Deep Learning, instance segmentation, brain, neuronal cells, microscopy

**Abstract:** Advances in virtual microscopy allow scanning whole slide images, capturing details at the cellular level and revealing the complexity of brain structures. Unbiased quantification of the distribution and features of neuronal cells is essential for understanding the mechanisms involved in brain development, aging, and neurodegenerative diseases. A crucial prerequisite for such studies is the identification of individual neurons.

In this thesis, we developed a deep learning-based segmentation framework for identifying individual neurons stained with NeuN in histological brain sections. Considering the tremendous manual labeling workload at the cellular scale to build training datasets, we first proposed a pixel-level mask synthesis pipeline based on point annotations. The synthetic masks contain three classes: neurons, tissue, and inter-cell contours. They were used as the ground truth to train and compare different network architectures. EfficientUNet was chosen as it has the best trade-off between accuracy and

efficiency. The prediction of the neural network produced a semantic segmentation. To further separate touching and overlapping neurons, we developed an original post-processing scheme based on ultimate erosion and dynamic reconstruction. We evaluated the proposed framework in a macaque brain dataset containing major anatomical regions and demonstrated its effectiveness, with an average F1 score of 0.92. Furthermore, we adapted the inference of the segmentation network for processing whole slide images using hybrid multi-CPU-GPU resources. An entire brain section of the mouse was processed via a user-friendly interface.

Additionally, to improve the robustness of the segmentation framework against color inconsistencies in microscopic images, we investigated two generative adversarial networks-based color normalization approaches to correct the color variations between images. Evaluation results show the superiority of these methods over conventional methods.

**Titre:** Un cadre général pour la segmentation d'instance de neurones basé sur l'apprentissage profond

**Mots clés:** Apprentissage profond, segmentation d'instance, cerveau, cellules neuronales, microscopie

**Résumé:** Les progrès de la microscopie virtuelle permettent de scanner des images de lames entières, de capturer des détails au niveau cellulaire et de révéler la complexité des structures cérébrales. La quantification non biaisée de la distribution et des caractéristiques des cellules neuronales est essentielle pour comprendre les mécanismes impliqués dans le développement, le vieillissement du cerveau ainsi que les maladies neurodégénératives. Un prérequis crucial pour de telles études est l'identification individuelle des neurones.

Dans cette thèse, nous avons développé un environnement de segmentation basé sur l'apprentissage profond pour identifier individuellement les neurones marqués avec NeuN dans des coupes histologiques du cerveau. Compte tenu de la charge de travail que représente l'étiquetage manuel des cellules pour construire un ensemble de données d'entraînement, nous avons d'abord proposé une méthode de synthèse de masques au niveau des neurones basée sur des points annotés manuellement. Les masques synthétiques contiennent trois classes : les neurones, le tissu et les contours intercellulaires. Ils ont été utilisés comme étant la vérité terrain pour entraîner et comparer différentes architectures de réseau. EfficientUNet a été choisi car il représente le meilleur com-

promis entre précision et efficacité. La prédiction du réseau des neurones a produit une segmentation sémantique. Afin de séparer les neurones qui se touchent ou se chevauchent, nous avons développé une approche originale de post-traitement basée sur l'érosion ultime et une reconstruction dynamique. Nous avons évalué la méthodologie proposée sur un ensemble de données provenant d'un cerveau de macaque et contenant les principales régions anatomiques. L'efficacité de cette méthodologie a été démontrée avec un score F1 moyen de 0,92. De plus, nous avons adapté l'inférence du réseau de segmentation pour le traitement d'images entières de lames en utilisant des ressources hybrides multi-CPU-GPU. Une coupe entière du cerveau d'une souris a été traitée via une interface ergonomique et simple d'utilisation.

Dans une optique d'amélioration de la robustesse de notre méthode de segmentation contre les variations de couleur dans les images microscopiques, nous avons étudié deux approches de normalisation des couleurs basées sur des réseaux adverses génératifs pour corriger les variations de couleur entre les images. Les résultats de l'évaluation sont prometteurs et montrent la supériorité de ces méthodes sur les méthodes conventionnelles.



## Acknowledgement

The past three years that I have spent at MIRCen have been a rewarding journey for me. I am so grateful that I could work on such an interesting research topic and develop tools that may be potentially helpful in neuroscience applications.

First and foremost, I would like to thank my supervisor Dr. Thierry Delzescaux, for leading me to this research field and always being there, guiding me, and encouraging me throughout my Ph.D. study. Your passion and dedication to research have always been inspiring to me. I would also like to thank Dr. Cédric Clouchoux for following my work during these three years and for all our insightful discussions. I want to express my gratitude to Nicolas Souedet for generously offering me his knowledge and technical assistance, which are indispensable to the completion of this thesis.

I sincerely thank my thesis committee — Dr. Grégoire Malandain, Prof. Cédric Wemmert, Dr. Cyril Poupon, and Prof. Christophe Destrieux, for spending their valuable time reviewing this thesis and their feedback.

I am grateful to the histology platform of MIRCen, in particular Dr. Caroline Jan and Camille Mabillon, for generating the data that is the cornerstone of this thesis and without whom it would not have been accomplished. I would like to express my gratitude to Dr. Marc Dhenain for providing your support and constructive comments on my paper, as well as to all the members of MIINDt.

I would like to thank all the current and past members of the image processing group. Especially Sébastien for giving me various help during the first two years of this journey. Many thanks go to my colleagues in Office 114: Lilian, François, Marina, Suzanne, . . . , for the good working atmosphere.

I acknowledge Région Ile-de-France, program DIM ELICIT, and Institute Paster for the financial support during my Ph.D. study. I would also like to thank my supervisor at the doctoral school, Prof. Riccardo Méssina, for his patience and constructive comments whenever I have questions.

Finally, I must thank my beloved parents for their continuous support and encouragement throughout the years of my study, as well as my brother for his presence and good humor. Special thanks to Frédéric for his companionship and support.



# Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>Résumé</b>	<b>iii</b>
<b>Acknowledgement</b>	<b>v</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xiv</b>
<b>Glossary</b>	<b>xvii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Context</b>	<b>5</b>
1.1 Brain anatomy . . . . .	5
1.1.1 Overview . . . . .	5
1.1.2 Main structures . . . . .	5
1.1.3 Major regions of interest . . . . .	7
1.1.4 Cell types . . . . .	8
1.1.5 Alzheimer’s disease . . . . .	13
1.1.6 Animal models . . . . .	16
1.2 Histology . . . . .	17
1.2.1 Principles of histology . . . . .	17
1.2.2 Stains for brain tissues . . . . .	18
1.2.3 Microscopy imaging . . . . .	19
1.3 Machine Learning . . . . .	24
1.3.1 Classical Machine Learning algorithms . . . . .	24
1.3.2 Deep Learning . . . . .	28
1.4 Motivation and challenges . . . . .	32
<b>2 Literature review</b>	<b>35</b>
2.1 Automatic cell segmentation methods . . . . .	35
2.1.1 Conventional methods . . . . .	35
2.1.2 Deep learning-based methods . . . . .	37
2.1.3 Whole Slide Imaging segmentation . . . . .	42
2.2 Color normalization for histological images . . . . .	44
2.2.1 Conventional methods . . . . .	44
2.2.2 Deep learning-based methods . . . . .	45



TABLE OF CONTENTS

---

2.3	Conclusion . . . . .	49
<b>3</b>	<b>Automatic generation of large-scale synthetic annotations and evaluation of Topcoders</b>	<b>51</b>
3.1	Introduction . . . . .	51
3.2	Biological dataset: a healthy macaque brain . . . . .	51
3.3	Semi-automatic pixel-level mask synthesis . . . . .	54
3.3.1	Random Forest binary segmentation . . . . .	54
3.3.2	Competitive region-growing . . . . .	54
3.3.3	Three-class-mask generation . . . . .	55
3.4	Evaluation of Deep Learning Topcoders Method for Neuron Individualization in Histo- logical Macaque Brain Section . . . . .	56
3.4.1	Introduction . . . . .	56
3.4.2	Materials and Methods . . . . .	57
3.4.3	Results . . . . .	61
3.4.4	Discussion . . . . .	63
3.4.5	Conclusion . . . . .	64
<b>4</b>	<b>A General Deep Learning framework for Neuron Instance Segmentation based on Efficient UNet and Morphological Post-processing</b>	<b>67</b>
4.1	Introduction . . . . .	67
4.2	A General Deep Learning framework for Neuron Instance Segmentation based on Effi- cient UNet and Morphological Post-processing . . . . .	67
4.2.1	Introduction . . . . .	68
4.2.2	Material and methods . . . . .	70
4.2.3	Results . . . . .	80
4.2.4	Discussion . . . . .	91
4.2.5	Conclusion . . . . .	93
<b>5</b>	<b>Development of a segmentation inference pipeline to process whole-slide images</b>	<b>101</b>
5.1	Introduction . . . . .	101
5.2	Whole-slide image dataset . . . . .	101
5.3	Software development for whole-slide image processing . . . . .	102
5.3.1	Decomposition of WSI into overlapping patches and weighted assembling . . .	102
5.3.2	Development of a parallel computing pipeline involving bothGPUs and CPUs .	104
5.4	Application of the proposed pipeline for segmenting whole-slide images . . . . .	112
5.4.1	Processing time under different configurations . . . . .	112
5.4.2	Segmentation results of the proposed pipeline on the mouse brain section . . .	114
5.5	Conclusion . . . . .	122
<b>6</b>	<b>Color normalization in microscopic images</b>	<b>123</b>
6.1	Color variation in microscopic image analysis . . . . .	123
6.2	Color normalization using Pix2Pix conditional GAN . . . . .	124
6.2.1	Training and test dataset for Pix2Pix . . . . .	125

6.2.2	Loss function of Pix2Pix . . . . .	125
6.2.3	Evaluation metrics to evaluate color normalization of Pix2Pix . . . . .	126
6.2.4	Conventional color normalization methods for comparison . . . . .	127
6.2.5	Evaluation of Pix2Pix and other compared methods . . . . .	127
6.2.6	Conclusion . . . . .	130
6.3	Adversarial Color Transfer to Study the Effect of Color Variation on Cell Instance Segmentation . . . . .	131
6.3.1	Introduction . . . . .	131
6.3.2	Material and Method . . . . .	133
6.3.3	Results and Discussion . . . . .	136
6.3.4	Conclusion and Perspectives . . . . .	139
<b>7</b>	<b>Discussion and Conclusion</b>	<b>141</b>
7.1	Contributions of the thesis . . . . .	141
7.2	Limitations of the thesis . . . . .	145
7.3	Future directions . . . . .	147
	<b>Résumé détaillé</b>	<b>149</b>



## List of Figures

1.1	Brain gross anatomy. . . . .	6
1.2	Regions of interest in the brain ( <i>e.g.</i> , rat brain): cortex, hippocampus, striatum, pallidum, and thalamus. The cortex is removed in the coronal view to better visualize other regions. . . . .	7
1.3	Basic parts of the neuron. . . . .	9
1.4	Basic forms of neurons. . . . .	10
1.5	Neurons and glial cells [103]. . . . .	12
1.6	The physiological structure of the brain and neurons in (a) healthy brain and (b) AD brain [13]. . . . .	15
1.7	Coronal section from mouse brain stained for neurons using different biomarkers. Left: Nissl and right: NeuN. . . . .	19
1.8	Digitalization of a histological specimen with a slide scanner for virtual microscopy. . . . .	20
1.9	Tile and linear scanning patterns. (a) Tile scanning with “focus every field”, (b) tile scanning with “focus every nth field” and (c) linear scanning. Dots mean focus points [55]. . . . .	21
1.10	WSI as a pyramid of image data [118]. . . . .	22
1.11	Left: Axio Scan Z.1 and right: image viewing software Zen. 1. Microscope, 2. objectives, 3. illumination, 4. cameras and 5. software. . . . .	22
1.12	Decision Tree example. . . . .	24
1.13	Simplified Random Forest example. . . . .	25
1.14	Node splitting in a decision tree model and Random Forest model. . . . .	26
1.15	A simple example of gradient boosting [166]. . . . .	27
1.16	Comparison of feature extraction between ML and DL. . . . .	29
1.17	The structure of an encoder and decoder network. The green (left) represents the encoder network and the purple (right) represents the decoder network. . . . .	30
1.18	General GAN architecture. . . . .	31
2.1	Flow diagram of structured preserving color normalization [132]. . . . .	46
2.2	StainGAN framework [114]. . . . .	47
2.3	Pix2Pix framework [107]. . . . .	47
3.1	Illustration of extracted coronal histological section (sagittal view of macaque brain surface). . . . .	52
3.2	Two section examples with color variations (90: dark and 91: brown). . . . .	52
3.3	Section 91 of macaque brain. The red squares show the images of $5000 \times 5000$ pixels [159]. . . . .	53
3.4	Training dataset examples: (a) caudate, (b-c) cortex, (d) hippocampus, (e) subiculum and (f) thalamus. . . . .	58

## LIST OF FIGURES

---

3.5	Flowchart of pixel-level label synthesis: (a) raw image, (b) RF segmentation, (c) point labels, (d) region-growing result (red contours) and (e) synthesized masks (black: background, blue: neurons and green: inter-cell contours). . . . .	59
3.6	Comparison of detection performance (det-F1) on the test set (several outlier values are below the vertical scale range). . . . .	62
3.7	Results of pixel-level label synthesis and segmentation in three anatomical regions. Top row: caudate; Middle row: cortex; Bottom row: hippocampus. (a) Raw images, (b) manual segmentation, (c) synthesized pixel-level masks and (d, e) results of $\text{Topcoders}_{\text{bowl}}$ and $\text{Topcoders}_{\text{neuron}}$ respectively. . . . .	63
4.1	Dataset examples with magnifications. (a) caudate, (b,c) cortex, (d) hippocampus, (e) subiculum and (f) thalamus. . . . .	71
4.2	Pixel-level mask synthesis. (a) original image, (b) manual point annotations, (c) RF binary segmentation, (d) fusion of colored connected components and (e), (f) labeled image produced by region growing and (f) final three-class-masks, blue: neurons, green: inter-cell contours and black: tissue. . . . .	73
4.3	Top: structure of EfficientNet-B5, it consists of 7 building MBConv blocks (represented with different color) and bottom: structure of our neural network using EfficientNet-B5 as encoder, named EfficientUNet-B5, the encoder is concatenated with the decoder at five different resolution (Block 2, Block 3, Block 4, Block 6 and Block 7). . . . .	75
4.4	Heatmaps of t-SNE projection. Left to right: all data, training data and test data. . .	81
4.5	Comparison of manual segmentations and synthetic masks. The white squares show the inconsistencies between manual segmentations. . . . .	83
4.6	Boxplot of F1-seg and AJI for different methods for comparative analysis. Left panel: F1 scores of segmentation. Right panel: AJI scores (outlier values of UNet are below the vertical scale range). . . . .	84
4.7	Original images and segmentation results of our method and other approaches in different regions with different neuron density. (a) raw images, (b) synthetic labels, (c-j) results of You <i>et al.</i> [160], EfficientUNet, UNet, HRNet, ResNet-34, DPN-soft, the ensemble model and the entire proposed framework. Top to Down: (1) thalamus, (2) caudate, (3) cortex and (4) hippocampus. The green contours represent segmentations that overlap the synthetic ground truth ( $\text{IoU} > 0.5$ ), the blue contours represent over-segmentations and the red contours represent missing and under-segmentations . . . . .	87
4.8	Intermediate results of the proposed post-processing. Top to down: (1) caudate, (2) cortex and (3) hippocampus. Left to right: (a) original images, (b) ground truth, (c) probability map of deep network, (d) binary mask of neuron channel, (e) ultimate residues, (f) reconstructed residues and (g) final segmentation after WS. The red square highlights the neurons that the neural network failed to separate but being fully segmented through the proposed post-processing. . . . .	90
4.9	24 images of the training set. . . . .	95
4.10	Supplementary cortex neuron dataset for assessing the generalizability of the proposed method. (a) Macaque 1. (b) Macaque 2. (c) Microcebus. (d) Mouse. . . . .	96

4.11	Segmentation results of supplementary cortex neuron dataset. Segmentation contours are shown in red and manual point annotations are shown in white. (a) Macaque 1. (b) Macaque 2. (c) Microcebus. (d) Mouse. . . . .	97
4.12	Segmentation results on three anatomical regions of a mouse brain. From left to right: raw images, semantic segmentations of EfficientUNet, and instance segmentations of the proposed post-processing. . . . .	98
4.13	Comparison of post-processing approaches based on distance transformation and ultimate erosion. (a) raw images, (b) ground truth, (c) probability map of EfficientUNet, (d) instance segmentation of our post-processing, and (e) instance segmentation of post-processing based on distance transformation. . . . .	99
5.1	Examples of inconsistent prediction at the boundaries. The boundaries and inconsistencies are highlighted with red dash lines and white squares, respectively, for better visualization. . . . .	102
5.2	Image patch extraction with overlap. . . . .	103
5.3	Overview of inference with overlapping extraction and assembling. For better illustration, the size of image and patch is defined as $1344 \times 1344$ and $512 \times 512$ pixels, respectively, with an overlap of 20%. . . . .	104
5.4	Comparison on three assembled predictions (size: $512 \times 512$ pixels) without and with the weight map. From left to right: original NeuN-stained images, synthetic ground truth, the assembled prediction without and with using the weight map. . . . .	105
5.5	Possible uses of BrainVISA software for image analysis. . . . .	106
5.6	Overview of Soma-workflow GUI [71]. . . . .	107
5.7	Overview of the proposed GPU-CPU parallel computing strategy. . . . .	108
5.8	Patch assembling scheme with operations on one single patch highlighted in blue. . . . .	108
5.9	Magnified stitched weight map and the center tile. . . . .	109
5.10	The original weight map, the center tile of the stitched weight map and the normalized weight map. . . . .	110
5.11	Patch assembling with the normalized weight map. . . . .	110
5.12	Graphic interface of the inference pipeline. . . . .	112
5.13	Processing time versus the number of GPUs for one mouse brain section. . . . .	114
5.14	A NeuN-stained mouse brain section. Below: 1. thalamus, 2. cortex and 3. dentate gyrus. . . . .	115
5.15	Neuron probability map of the mouse brain section. Below: 1. thalamus, 2. cortex and 3. dentate gyrus. . . . .	116
5.16	Inter-cell contour probability map of the mouse brain section. Below: 1. thalamus, 2. cortex and 3. dentate gyrus. . . . .	117
5.17	Non-stained tissue probability map of the mouse brain section. Below: 1. thalamus, 2. cortex and 3. dentate gyrus. . . . .	118
5.18	Class map of the mouse brain section. The black, gray and white colors represent non-stained tissue, inter-cell contours and neurons, respectively. Below: 1. thalamus, 2. cortex and 3. dentate gyrus. . . . .	119

## LIST OF FIGURES

---

5.19	Instance segmentation of the mouse brain section (rainbow colormap). Below: 1. thalamus, 2. cortex and 3. dentate gyrus. . . . .	121
6.1	Image patches extracted from sections 90 and 91. From left to right: putamen 1 (sparse), putamen 2 (dense), and cortex. . . . .	124
6.2	Architecture of Pix2Pix framework. Image modified from [107]. . . . .	125
6.3	Color normalization results of different methods. From left to right: target image, source images, recoloring results of [102], [82], [132] and Pix2Pix, respectively. . . . .	128
6.4	Neuron microscopic images with five nickel concentrations. Left: image examples of different nickel groups. Right: correlation of histogram difference and nickel concentration (with the nickel-free group as reference). . . . .	133
6.5	ReHistoGAN architecture. . . . .	135
6.6	Stain color transfer results from different methods. The first row presents recolored results of a 1.6% nickel image using a nickel-free image as reference. The second row shows recolored results after switching the source and target. . . . .	137
6.7	Macaque segmentation dataset recolored by ReHistoGAN. First row: target images, second row: source macaque image and recolored results. From left to right: GAN-recolored images using the macaque image (control), nickel-free (recolored-1), 0.4%, 0.8%, 1.2% and 1.6% nickel mouse images (recolored-2-5) as targets. . . . .	138

## List of Tables

2.1	Current cell segmentation methods based on Deep Learning. . . . .	40
3.1	Detection and segmentation performance of $\text{Topcoder}_{\text{bowl}}$ , $\text{Topcoder}_{\text{neuron}}$ and its constitutive models. Best and second best results are in bold with the best also underlined.	62
4.1	Average IoU score (mIoU) and Dice coefficient between the synthetic masks and three manual annotations (denoted as e1, e2 and e3) respectively. . . . .	82
4.2	Comparison of neuron detection, counting, instance segmentation and computational complexity of different automatic methods. Best and second best results are in bold with the best also underlined. FLOPs are computed for a $224 \times 224$ input RGB image.	85
4.3	Instance segmentation performance using different post-processing techniques. The best results are in bold. . . . .	89
4.4	Processing time for one image of $5000 \times 5000$ using different post-processing approaches.	90
4.5	Object-level segmentation performance of the proposed method on the supplementary test set. . . . .	90
5.1	Computation details of inference for various images on 1GPU (overlap ratio: sim 10%).	113
5.2	Time in minutes for processing one mouse brain section with different overlap ratios and GPU resources. . . . .	113
6.1	Comparison of color transfer methods. The $H_{\text{dist}}$ and $KL_{\text{div}}$ are calculated in RGB color space. . . . .	129
6.2	Comparison of color transfer methods. The $H_{\text{dist}}$ and $KL_{\text{div}}$ are calculated in $uv$ color space. . . . .	129
6.3	Comparison of color transfer methods. The $H_{\text{dist}}$ and $KL_{\text{div}}$ are reported to show the color difference between recolored results and target images. . . . .	138
6.4	Comparison of segmentation performances (F1, AJI and Dice) on segmentation dataset without and with color variations inherited from the different groups in the color transfer dataset. Color variations between the two datasets are quantified using $H_{\text{dist}}$ and $KL_{\text{div}}$ . . . . .	139





## Glossary

**2D** Two Dimensional. 17, 64, 70, 71, 148

**3D** Three Dimensional. 23, 148

**AD** Alzheimer's Disease. 13, 14, 32

**AJI** Average Jaccard Index. 78, 79, 81, 83, 84, 86, 88, 92, 136, 139

**API** Application Programming Interface. 106

**CAD** Computer Aided Diagnosis. 33, 44, 129, 148

**cGAN** conditional Generative Adversarial Network. 123, 130

**CNN** Convolutional Neural Network. 24, 30, 38, 39, 43, 60, 69, 70, 72–74, 91, 102

**CNS** Central Nervous System. 9, 10, 12, 14

**CPU** Central Processing Unit. 3, 61, 79, 107, 112, 113, 122, 142, 145

**DAB** Diaminobenzidine. 128, 133

**DAPI** 4',6-Diamidino-2-phenylindole. 57

**DL** Deep Learning. xi, 1, 2, 24, 28, 29, 32, 33, 36–38, 42–44, 48, 49, 51, 54–58, 60, 61, 64, 69, 70, 77, 79, 81, 83, 92, 93, 110, 123, 131–133, 139, 141, 142, 145, 149–151

**DSB** Data Science Bowl. 57, 58, 63, 64

**DT** Decision Tree. 24–26, 28, 54

**FCN** Fully Convolutional Network. 30, 37

**FLOPs** Floating Point Operations. xv, 81, 83–85, 91, 142

**FN** False Negative. 60, 78

**FP** False Positive. 78, 79, 92

**GAN** Generative Adversarial Network. xiv, 3, 29–31, 44–46, 49, 123, 124, 130, 132, 135, 136, 138–140, 143, 144, 153

**GBDT** Gradient Boosting Decision Tree. 24, 26, 28, 38, 55, 67

- GFAP** Glial Fibrillary Acidic Protein. 18, 92
- GPU** Graphics Processing Unit. viii, xiii, xv, 2, 3, 29, 32, 57, 60, 61, 71, 79, 102, 104, 107, 108, 111–114, 122, 142, 151
- GUI** Graphical User Interface. 104
- H&E** Hematoxylin and Eosin. 1, 17, 18, 32, 33, 37, 45, 56, 57, 123, 124, 127, 132, 140, 143
- H\_dist** Hellinger Distance. 127–130, 134, 136, 137
- HPC** High Performance Computing. 64, 65, 101, 147
- Iba1** Ionized Calcium Binding Adaptor Molecule 1. 18, 92
- IHC** Immunohistochemistry. 1, 18, 32, 33, 51, 128, 130, 132, 133, 137, 139, 143
- IoU** Intersection over Union. xii, 60, 70, 78, 79, 81, 86, 87, 92
- KL\_div** Kullback-Leibler Divergence. 127–130, 134, 136, 137, 139
- mIoU** Average Intersection over Union. 81
- ML** Machine Learning. xi, 2, 23, 24, 28, 32, 151
- MRI** Magnetic Resonance Imaging. 32, 147, 148, 154
- NeuN** Neuronal Nuclear Protein. 1, 18, 32, 33, 51, 54, 56, 57, 61, 64, 68, 70, 72, 92, 93, 101, 124, 127, 128, 130–134, 140, 141
- NMF** Non-Negative Matrix Factorization. 45
- NMS** Non-Maximum Suppression. 37
- OD** Optical Density. 45, 128, 132
- RAM** Random Access Memory. 57, 60, 61, 71, 79, 104
- RCE** Relative Count Error. 60, 61, 78, 81, 83, 88
- RF** Random Forest. 24–26, 28, 37, 54, 57, 58, 70, 72, 73, 81, 91, 141, 145
- ROI** Region of Interest. 32, 38, 43, 49, 141
- SVD** Singular Value Decomposition. 45, 127
- t-SNE** t-distributed Stochastic Neighbor Embedding. 71, 80

**TP** True Positive. 60, 78

**WS** Watershed Segmentation. 70, 76, 80, 88, 92

**WSI** Whole Slide Imaging. 19–21, 32, 33, 42, 43, 101, 102, 131, 151–153

## Introduction

The investigation of neuron population is essential in neuroscience to study brain function, development, aging, and neurodegeneration. A prerequisite for quantitative analysis at the cellular scale is to identify individual neurons. In practice, biologists mainly rely on stereology and manual segmentation to estimate the cell population in regions of interest. However, this technique is too cumbersome to be conducted on large-scale images and cohorts. For decades, many researchers have attempted to tackle this task in an automatic way. Nevertheless, these methods require specific settings adapted to the characteristics of cells (species, sizes, stains, *etc.*) and are sensitive to noise and artifacts on the image. Recently, Deep Learning (DL) based approaches have demonstrated remarkable potential in terms of accuracy and robustness in many computer vision fields, including cell instance segmentation. Once the DL model is trained, it can segment cells without human intervention. One of the main challenges is that most DL methods are supervised, and large annotated datasets are mandatory for the training. However, performing the annotation manually at the pixel-level can be extremely laborious and time-consuming. Moreover, most DL-based approaches have mostly been developed for Hematoxylin and Eosin (H&E) stained images with the cell nuclei stained as bluish and extracellular stained as pink. The neurons are often stained using immunohistochemical (IHC) staining with the NeuN antibody, which is usually dark brown. In addition, segmenting neurons is challenging due to the large variety in shape, size, and density across anatomical regions, especially in the dentate gyrus of the hippocampus, which is composed of thousands of highly aggregated neurons.

Although DL segmentation approaches are considered more robust than conventional methods, the color variation in the histological images exists as a supplementary critical issue. All relevant features for segmentation that the DL model has learned come from the training images. The segmentation DL model may underperform when applied to images with a color different from the training data. A solution is to standardize the color of test images using the training image color as the reference prior to the inference. In fact, the inconsistency of stain color is also problematic for the visual diagnosis of pathologists. Several color normalization approaches (both conventional and DL-based) have been proposed, but they are intended to correct color variations on H&E-stained images. The transfer on IHC-stained images and, more specifically, on NeuN-stained images remains an open question, especially for conventional approaches, which convert images to other color space from RGB

## INTRODUCTION

---

space to match the color statistics of the source image and the reference.

The objective of this thesis is to develop a DL-based approach to automatically individualize neuronal cells in various anatomical regions in animal models studied in preclinical research. Accuracy, robustness, and calculation cost are major criteria and constraints for developing this approach. The last feature aims to evaluate the scalability of the proposed approach for large-scale images. The robustness of the proposed segmentation method to color variations can be further improved by applying the color normalization method as a preprocessing step.

The work of this thesis took place in the Molecular Imaging Research Center (MIRCent) in the neurodegenerative diseases laboratory UMR9199 (LMN), a research unit of the French Alternative Energies and Atomic Energy Commission (CEA) in association with the National Center for Scientific Research (CNRS), the University of Paris-Saclay and in collaboration with WITSEE company. All the data used in this study are microscopic brain images produced at the histology platform of MIRCent.

This manuscript is subdivided into seven parts.

Chapter 1 presents the biological context of this study, including the basic organization of the brain, the characteristics of neuronal cells, the mechanisms involved in neurodegenerative diseases (*e.g.*, Alzheimer’s Disease), and animal models studied for neurodegeneration research. A brief description of the histology technique and image digitalization is provided. The advent of these technologies has increased the access to vast amounts of high-resolution data, which require adapted software solutions to be analyzed. In this context, Machine Learning (ML) algorithms, especially deep neural networks, are thought to be among the most effective tools for the automatic analysis of massive amounts of data. Several ML algorithms employed in this thesis study are also covered in this chapter.

Chapter 2 starts with a comprehensive review of current approaches developed for cell instance segmentation. Both conventional and DL-based approaches are presented and categorized according to their principles and the nature of ground truth, respectively. This chapter also investigates scalable strategies for DL-based approaches to process whole-slide images since they face the limitation of GPU memory compared to conventional methods. In addition, commonly used approaches for color normalization are also reviewed in the second part of this chapter, as color variation exists as a critical issue in microscopic image analysis.

Chapter 3 describes a macaque brain dataset produced at MIRCent for investigating neuron instance segmentation. Large labeled datasets are mandatory

---

for training neural networks, while manually performing the annotation at the pixel-level is extremely labor-intensive. A strategy is developed to generate synthetic masks with minimal manual effort. This strategy is applied to annotate a large macaque dataset. The evaluation of a state-of-the-art segmentation method using the macaque dataset and synthetic annotations is presented as a conference paper:

Wu, Huaqian, Nicolas Souedet, Zhenzhen You, Caroline Jan, Cédric Clouchoux, and Thierry Delzescaux. “Evaluation of Deep Learning Topcoders Method for Neuron Individualization in Histological Macaque Brain Section.” In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 2985-2988. IEEE, 2021.

Chapter 4 presents a general framework for neuron instance segmentation. Various neural networks are compared to identify an efficient architecture with the optimum trade-off of accuracy and computational cost. A novel post-processing design based on mathematical morphology is introduced to refine the segmentation of the neural network. The work of this chapter is presented as a journal paper:

Wu, Huaqian, Nicolas Souedet, Caroline Jan, Cédric Clouchoux, and Thierry Delzescaux. “A general deep learning framework for neuron instance segmentation based on efficient UNet and morphological post-processing.” *Computers in Biology and Medicine (2022)*: 106180.

Chapter 5 focuses on implementing the proposed neuron instance segmentation framework on whole slide images. A parallel computing strategy involving both CPU and GPU resources is established to accelerate inference processing using a partial image reading and writing library. A new pipeline dedicated to segmenting large images is integrated into BrainVISA software<sup>1</sup>. Additionally, this chapter also presents a validation of the pipeline on the entire mouse brain section.

Chapter 6 describes two attempts to correct color variations in microscopic images. They can be used as a preprocessing step of the neuron instance segmentation framework for test images whose colors differ from the training set. The first method is based on generative adversarial networks (GAN), with a generator mapping image to the color distribution of the training set. The second is also GAN-based, but the generator is trained to recolor images using the reference color histogram. With such flexibility, this chapter also quantifies the impact of color variations on neuron instance segmentation. The work of the second method is presented as a conference paper:

Wu, Huaqian, Nicolas Souedet, Camille Mabillon, Caroline Jan, Cédric Clou-

---

<sup>1</sup><https://brainvisa.info>

## INTRODUCTION

---

choux, and Thierry Delzescaux. “Adversarial Stain Transfer to Study the Effect of Color Variation on Cell Instance Segmentation.” In *International Workshop on Medical Optical Imaging and Virtual Microscopy Image Analysis (MOVI)*, pp. 105-114. Springer, Cham, 2022.

Finally, Chapter 7 summarizes the main contributions of the thesis. Limitations of the neuron instance segmentation framework, remaining challenges for processing whole-slide images, and issues of color normalization evaluation are discussed. This chapter also presents future directions and perspectives of the proposed neuron instance segmentation and color normalization approaches.



# 1 - Context

## 1.1 . Brain anatomy

### 1.1.1 . Overview

The brain is the most complex organ in the body of a vertebrate, and it serves as the core of the nervous system in all vertebrate animals, as well as the majority of invertebrate animals. Despite sustained research for centuries, the brain remains less understood among all the organs. Although the functioning of each individual brain cell is now very well understood, the details surrounding how millions of them work together are still unclear. With the introduction of new technologies over the past few decades, including medical imaging techniques, computational neuroscience, and artificial intelligence, our understanding of the brain has experienced a revolutionary change. This section briefly presents the components of the brain that will be addressed in this thesis work, including the basic structures, the anatomical regions, as well as the cell types found in the brain.

### 1.1.2 . Main structures

The brain consists of three main structural divisions (see Figure 1.1):

#### 1.1.2.1 . Cerebrum

The cerebrum (front of the brain) is the largest region of the brain. The outer layer is the cerebral cortex. The cerebrum controls temperature regulation and movement coordination. It also enables higher functions, including touch, vision, and hearing interpretation, as well as communication, reasoning, emotions, learning, and problem-solving.

The cerebrum is made up of two cerebral hemispheres, the right and left, joined by the corpus callosum, which allows communication between the two sides of the brain. Each hemisphere receives information from the contralateral side of the body and controls its movement. The hemispheres are divided into four lobes, as shown in Figure 1.1, with different colors [120, 90].

The **frontal lobe**, located in the front of the brain, is the largest lobe of the brain. Reasoning, motor abilities, higher-level cognition, and linguistic ability are all associated with it. The motor cortex is located near the central sulcus in the back of the frontal lobe.

The **parietal lobe** is the middle part of the brain. It contributes not only to spatial and visual perception but also to the processing of tactile sensory information such as pressure, touch, and pain.

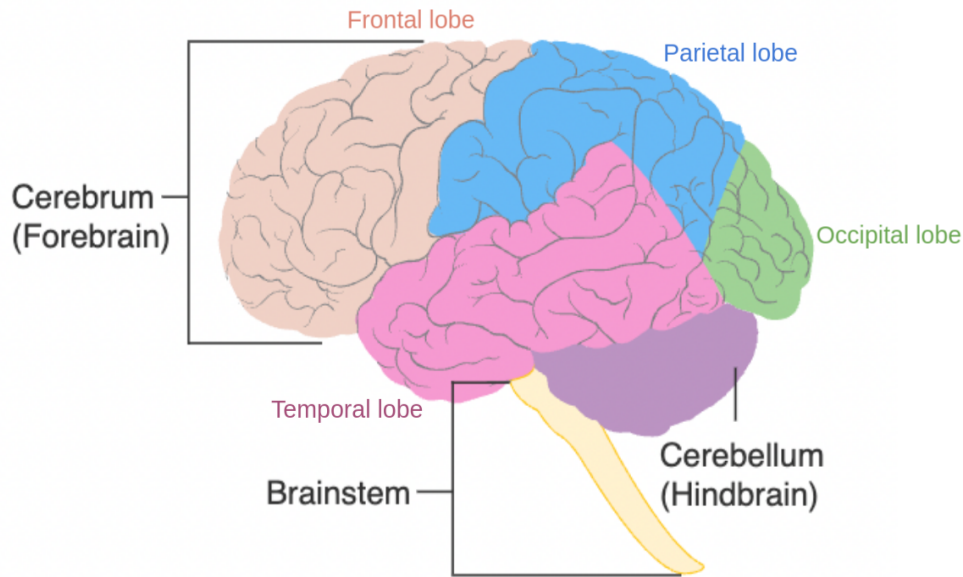


Figure 1.1: Brain gross anatomy<sup>1</sup>.

The **occipital lobe** is located in the back of the brain. It is associated with analyzing visual information (color, light, shape, and movement). The occipital lobe contains the primary visual cortex, which receives and processes information from the retinas of the eyes [120, 90, 61].

The **temporal lobe** is located on the bottom side of the brain. The main auditory cortex, which is vital for interpreting sounds and understanding language, is also located in this lobe, as well as the hippocampus, which is highly involved with memory [120, 90, 61].

#### 1.1.2.2 . Cerebellum

The cerebellum (“little brain”) is located at the back of the brain, above the brainstem and below the temporal and occipital lobes. The inner area communicates with the cerebral cortex, while the outer layer contains neurons. The cerebellum is just about 10% of the total size of the brain, yet it contains more than half of the total number of neurons. Recent studies estimate 101 billion neurons in the human cerebellum [5], which is about five times the number of neurons in the cerebral cortex [96]. It is essential for maintaining posture, balance, and equilibrium by coordinating voluntary muscle movements.

---

<sup>1</sup>[https://www.physio-pedia.com/Brain\\_Anatomy](https://www.physio-pedia.com/Brain_Anatomy)

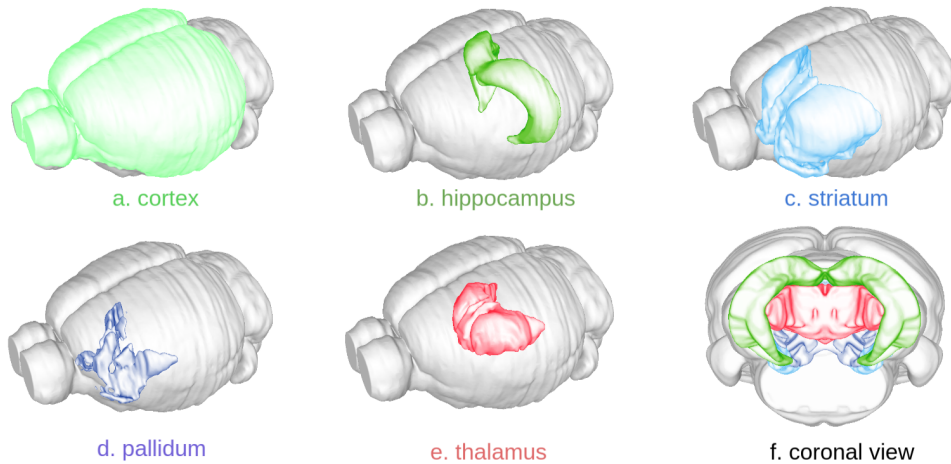


Figure 1.2: Regions of interest in the brain (*e.g.*, rat brain): cortex, hippocampus, striatum, pallidum, and thalamus (image from Blue Brain Cell Atlas<sup>2</sup>). The cortex is removed in the coronal view to better visualize other regions.

### 1.1.2.3 . Brainstem

The brainstem (middle of the brain) connects the cerebrum and cerebellum to the spinal cord and serves as a relay hub. The size of the brainstem is small, accounting for only 2.6 percent of the total weight of the brain [41]. It has a critical role in autonomic functions such as breathing, cardiac rhythms, body temperature and sleep cycle regulation, digestion, sneezing, coughing, vomiting, and swallowing [120, 61].

### 1.1.3 . Major regions of interest

Figure 1.2 shows five anatomical regions which have crucial roles in this study: cerebral cortex, hippocampus, striatum, pallidum, and thalamus. They are particularly important for studying neurodegenerative diseases.

#### 1.1.3.1 . Cortex

The cerebral cortex (Figure 1.2 a) is the outer layer of neural tissue of the cerebrum. It primarily contains gray matter (neural tissue that is mainly constituted of neurons). Despite being only a few millimeters thick, the cerebral cortex accounts for over half of the brain's overall weight. There are 16 billion neurons in the cerebral cortex of the human brain [47]. It is linked to our most advanced brain capacities.

<sup>2</sup><https://bbp.epfl.ch/nexus/cell-atlas/>

### 1.1.3.2 . Hippocampus

The hippocampus (Figure 1.2 b) is a complex region located deep within the temporal lobe. The hippocampus is involved in memory consolidation and cognitive learning. It is here where short-term memories are transformed into long-term memories. In humans, damage to the hippocampus can result in significant problems in long-term episodic memory [10]. Hippocampal atrophy is linked to several neurodegenerative diseases [4].

### 1.1.3.3 . Striatum

The striatum (Figure 1.2 c) is a crucial component of the reward and motor systems. It contains the caudate nucleus and the putamen. Multiple aspects of cognition are coordinated by the striatum, including motor and action planning, decision-making, and reward perception. Atrophy of the striatum is involved in Parkinson's disease [97], Huntington's disease [129], and other movement disorders.

### 1.1.3.4 . Pallidum

The globus pallidus (Figure 1.2 d) is an essential part of the basal ganglia. It receives the major afferents from the striatum, which is its principal input station. The globus pallidus is involved in voluntary movement regulation.

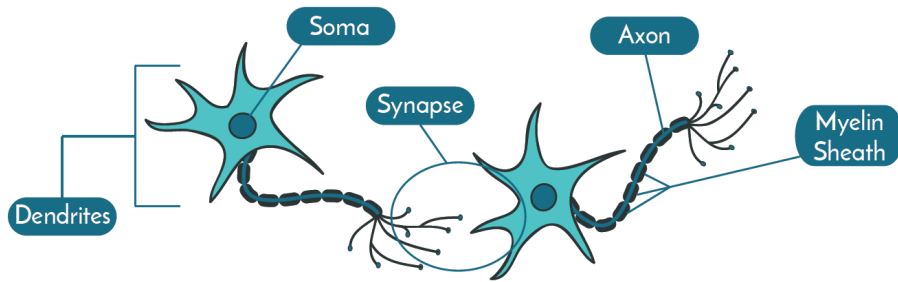
### 1.1.3.5 . Thalamus

The thalamus (Figure 1.2 e) is a vital gray matter structure located deep within the frontal lobe, near the brain's center. It is associated with important neurological and sensory functions. It serves as a relay station, receiving sensory impulses and transmitting them to the relevant cortical areas. The thalamus is also involved in the regulation of sleep and wakefulness states [120, 61].

This section briefly presented the major anatomical regions in the brain. In practice, neurobiologists often need details at the cellular level (*e.g.*, cell population) of regions of interest to further study physiological or pathological mechanisms and examine the effect of treatments.

## 1.1.4 . Cell types

The brain is made up of two types of cells: nerve cells (neurons) and glial cells (astrocytes, microglia, and oligodendrocytes). The average adult male human brain contains 86 billion neurons and about ten times more glial cells [6]. Although neurons are the most well-known brain cells, effective brain function requires both neurons and glial cells.

Figure 1.3: Basic parts of the neuron<sup>3</sup>.

#### 1.1.4.1 . Neurons

Neurons are specialized in processing and transmitting cellular signals. The cell body, or soma, branching dendrites that receive messages from other neurons, and the axon, which sends information out to other neurons through the axon terminal, are three essential elements of the neuron (see Figure 1.3). The neurons communicate with one another through synapses. When a neuron emits an action potential, signals, partly electric and partly chemical, go through the synapse from one neuron's axon to the dendrites of another neuron. Neurons are usually divided into three categories based on their function: sensory, motor neurons, and interneurons.

**Sensory neurons**, also called afferent neurons, have their cell bodies located in the dorsal ganglia outside of the spinal cord. They are nerve cells that are triggered by environmental sensory information. They transmit signals to the CNS via the receptors in response to specific types of stimuli that affect the cells of the sensory organs, including smell, taste, vision, sound, and temperature. This process is called sensory transduction.

**Motor neurons**, or efferent neurons, can be divided into upper motor neurons and lower motor neurons. The upper motor neurons are located in the motor cortex, with axons synapsing onto interneurons in the spinal cord and sometimes straight onto lower motor neurons. Lower motor neurons can be found in both the brainstem and the spinal cord. The axons are efferent nerve fibers that carry signals from the spinal cord to effector organs such as muscles and glands.

**Interneurons**, or association neurons, are situated between sensory and motor neurons. Interneurons are the core nodes of neural circuits that allow sensory or motor neurons to communicate with the CNS. Interneurons correspond roughly to 20-30% of neurons in the neocortex (which makes up about 80% of the human brain) [86]. The huge number of interneurons reflects the complex-

<sup>3</sup><https://dana.org/article/cells-of-the-brain/>

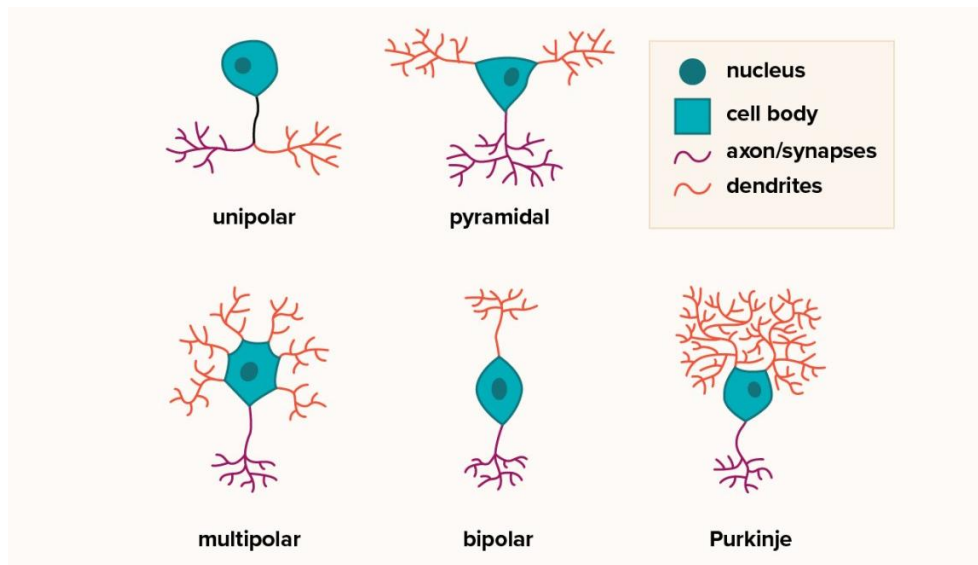


Figure 1.4: Basic forms of neurons (image from healthline<sup>4</sup>).

ity of integrating the CNS's sensory and motor parts, as well as the diversity of functions found in the brain and spinal cord.

Neurons vary in shape and size. Based on their anatomical characteristics, most neurons can be divided into five classes: multipolar, unipolar, bipolar, pyramidal and purkinje neurons (see Figure 1.4).

**Multipolar neurons** have a single axon that is surrounded by symmetrical dendrites. This is the most common type of neuron in the CNS. Multipolar neurons contain motor neurons and interneurons. They can be found in the cortex, the spinal cord, and the autonomic ganglia.

**Unipolar neurons** contain a single axon and are usually exclusively seen in invertebrate animals. They are sensory neurons located in the spinal and cranial nerve ganglia.

**Bipolar neurons** have two extensions from the cell body. Many bipolar cells are sensory neurons with specialized functions in sense transmission. They can usually be found in the eye's retina, the vestibular nerve, the cerebral cortex, and the spinal ganglia during embryonic development.

**Pyramidal neurons** have a single axon but several dendrites, forming a pyramid shape. Most of them are located in the cerebral cortex, the hippocampus, and the amygdala. They play a crucial role in cognitive ability and visually guided motor function [108, 27].

<sup>4</sup><https://www.healthline.com/health/neurons#anatomy>

**Purkinje neurons** are some of the biggest neurons in the brain, with numerous dendrites that branch out from the cell body. They are inhibitory neurons, the only output of all motor coordination in the cerebellar cortex, releasing inhibitory neurotransmitters to other neurons.

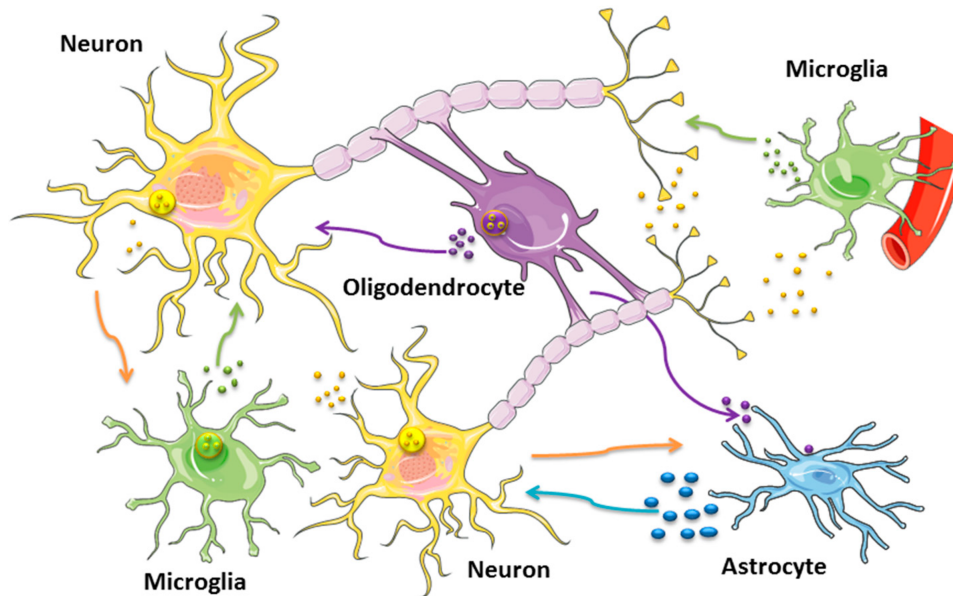


Figure 1.5: Neurons and glial cells [103].

#### 1.1.4.2 . Glial cells

Glial cells are non-neuronal cells that can be found in both the central and peripheral nervous systems. They are critical in providing protection, as well as physical and metabolic sustenance for neurons, such as neuronal insulation and communication, as well as nutrient and waste delivery. Glial cells can be divided into three categories: microglia, astrocytes, and oligodendrocyte lineage cells (see Figure 1.5).

**Microglia** can be found throughout the CNS. They are the immunocompetent and phagocytic cells that serve as the first and important line of active immune defense in the CNS [36]. They play a crucial role in maintaining brain homeostasis.

**Astrocytes** are star-shaped glial cells in the CNS. They play a vital role in the formation of the blood–brain barrier [124], the provision of nutrients to nervous tissue, the maintenance of extracellular ion balance, the regulation of cerebral blood flow, and inflammatory reactivity after injury [32].

**Oligodendrocytes** exist exclusively in the CNS. Its principal role is to create the myelin sheath [17], which provides support and insulation to axons in the CNS of several vertebrates, similar to the function of Schwann cells in the peripheral nervous system.

Studying the number, morphology, and distribution of brain cells, especially neurons in healthy brains, is essential for learning about brain development



and aging. It is also crucial for the study of brain dysfunction and neurodegenerative diseases since neuronal degeneration is directly related to the development of neurodegenerative diseases (*e.g.*, Alzheimer, Parkinson, and Huntington's diseases).

#### **1.1.5 . Alzheimer's disease**

Neurodegenerative diseases are a diverse group of disorders characterized by the progressive degeneration of the structure and function in the central or peripheral nervous system. Neurodegeneration can be found at several levels of neuronal circuitry in the brain, from molecular to systemic. These diseases are considered incurable since there is no known mechanism to stop this progressive neuron degeneration. Alzheimer's disease (AD) is the most common neurodegenerative disease and is particularly studied at MIRCen in preclinical research.

##### **1.1.5.1 . Signs and symptoms**

Memory loss is the most common early symptom of Alzheimer's disease. As the disease progresses, language problems, confusion, mood swings, loss of motivation, self-neglect, and behavioral problems are all possible symptoms. It is associated with 60–70% of dementia cases. Studies suggested that there were about 50 million people worldwide with AD by 2020 [13], which has become a heavy problem in our society.

The first symptoms of AD are memory and cognitive impairment, which are frequently misdiagnosed as aging or stress [141]. A person in the early stage of AD may appear healthy, but he/she may struggle with memory loss (especially short-term memory), decision-making, misplacing objects, and learning impairment [30]. As the disease progresses, the patient in the moderate stage may have long-term memory loss, confusion [140], difficulties with language and logical reasoning, anxiety, and anger. At this point, more rigorous observation and care are required [30]. During the final stage, people with severe AD are fully dependent on others. Inability to communicate [31], apathy, exhaustion, weight loss, seizures, skin infections, loss of mobility, and increased sleeping are common symptoms in this stage. External factors, such as pressure ulcer infection or pneumonia, are more often the cause of death rather than the disease itself [30].

### 1.1.5.2 . Causes

Although the cause of most AD cases is unclear, three main hypotheses for AD have been proposed: cholinergic, amyloid, and tau hypotheses.

The cholinergic hypothesis is the oldest, on which most therapies are based. It suggests that AD is caused by inadequate synthesis of the neurotransmitter acetylcholine (ACh). ACh is essential for cognitive function, which is involved in various physiological processes such as memory, attention, sensory information, learning, and other vital abilities. Cholinergic neuron degeneration has been found to occur in AD, causing an impact on cognitive function and memory loss [42].

The amyloid hypothesis is based on a strong correlation between the abnormal deposition of amyloid-beta peptide ( $A\beta$ ) in the CNS with dementia (as shown in Figure 1.6).  $A\beta$  is accumulated extracellularly as amyloid plaques (AP), which causes neurotoxicity and tau pathology, which leads to neuronal cell death and neurodegeneration. Despite the fact that AP also accumulates in healthy brains with aging, the amyloid hypothesis is still the most widely accepted pathogenic explanation underlying inherited Alzheimer's disease (IAD) [95, 62].

Tau hypothesis. Recent research suggests that tau protein, instead of  $A\beta$ , is the key factor in the development of AD. Tau proteins are essential for carrying nutrients and other essential components within the neuron's internal support and transport system. The shape of tau proteins is altered and deposited into neurofibrillary tangles. The tangles affect the transport system, triggering a toxic cascade that eventually results in cell death and brain shrinkage [62].

Other risk factors of AD include aging, genetics, environmental and medical factors. Old age is the primary factor for AD, and the incidence of the disease is increasing dramatically as the population ages worldwide [149]. The majority of cases of AD are late-onset AD (LOAD), which begin beyond the age of 65 [52]. In younger people (30-60), the disease can be classified into early-onset AD (EOAD), with just 1-6 percent of cases and most being familial. Genetic factors have been discovered to play an important role in the development of the disease. 70% of AD cases are associated with genetic factors [60]: the majority of EOAD cases are inherited in an autosomal dominant pattern, and mutations in the dominant genes are also related to AD. However, the aforementioned factors do explain all AD cases. Air pollution, nutrition, metals, infections, and a variety of other environmental risk factors may generate oxidative stress and inflammation, raising the risk of developing AD. Cardiovascular disease, obesity, diabetes, and other medical disorders are common in the elderly with AD, which are linked to an increased risk of AD as well.

The disorder is currently incurable, which increases the need to create and

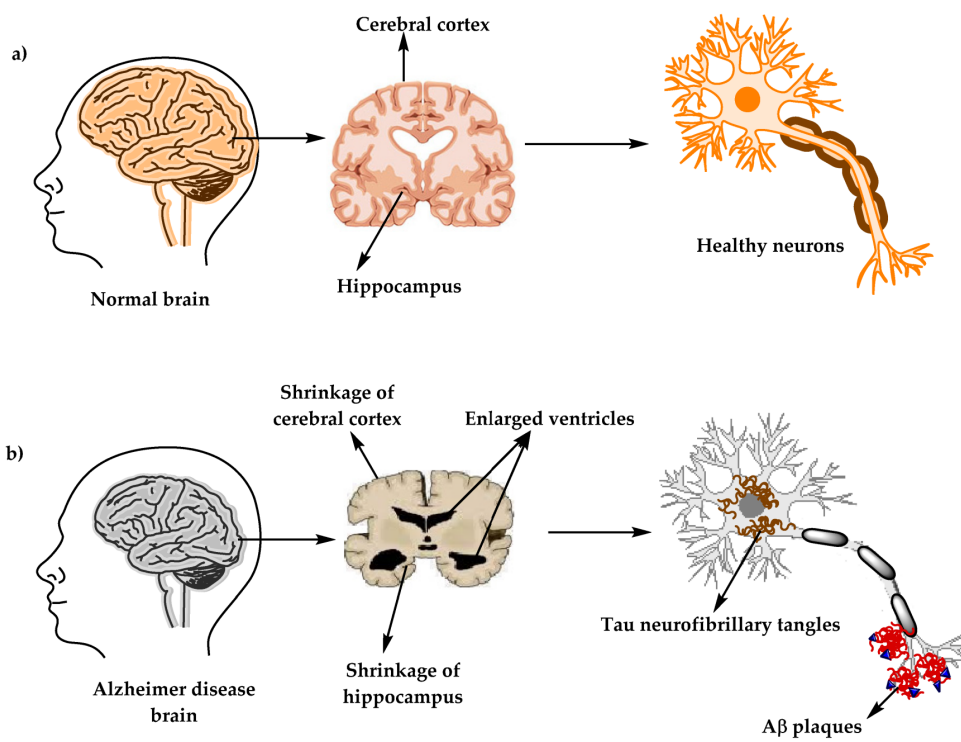


Figure 1.6: The physiological structure of the brain and neurons in (a) healthy brain and (b) AD brain [13].

characterize relevant animal models to aid translational research and preclinical drug development. In this case, the study of cell death (neurons and glial cells) is particularly relevant and critical due to the lack of automated methods. Many of the treatments that are now undergoing clinical trials were first carried out in animal models, which have played a significant role in defining important disease-related mechanisms and being at the forefront of evaluating novel therapeutic approaches [70].

### **1.1.6 . Animal models**

Translational research evaluates novel therapy strategies using animal models that replicate human disease. Therapeutic tests must be successful in animal models before being applied to humans in a clinical trial. At MIRCen, we use mainly rodent (mouse and rat) and primate (microcebus and macaque) models to conduct preclinical research on neurodegenerative diseases. About 90% of human genes have an equivalent in mice [14], proving the phylogenetic proximity (study of relatedness) of mice and humans. Additionally, rodent models have the benefit of being easily available because of their high reproductive capability, size, and short lifespan, which shortens the time required for studies. Primate models are more challenging to study due to their low rate of reproduction, ethical constraints, and longer lifespans, which extend the duration of experiments. Primate models, however, provide a greater understanding of brain functions since they are more structurally and functionally similar to human brains. In fundamental preclinical research, studies are first conducted in rodents, then in primates, and if the study is successful, a transition to the clinic may further be envisioned.

Animal models make possible the study of brain tissues post-mortem at the microscopic level, which constitutes a major asset to understanding both pathological and therapeutic mechanisms. The data used in this manuscript are microscopic images of histological sections of mouse and macaque brains since they provide the possibility of capturing neuron information at the cellular level. The following section describes the histology technique, biomarkers for neuron staining, as well as virtual microscopic imaging.

## 1.2 . Histology

Histology is a branch of biology that investigates post-mortem the anatomy of biological tissues at the microscopic level. The primary aim of histology is to figure out how tissues are organized at different levels of structure, from cells to intercellular substances to organs. Histology has a vital role in both our current understanding of life and medical practice. Histopathology is one branch of histology that is extremely helpful in assessing a diagnosis as well as the severity and progression of a disease through biopsy or postmortem analyses.

### 1.2.1 . Principles of histology

After the extraction of tissues, the preparation of histological sections can be generally conducted in the following steps: fixation, sectioning, staining, and imaging.

#### 1.2.1.1 . Fixation

Fixation is the preservation of biological tissues against deterioration due to autolysis or putrefaction and is performed during the euthanasia step in animal models by injection of paraformaldehyde solution. Tissue fixation is an essential phase in the production of histological sections, intending to preserve cells and tissue components while allowing for the preparation of stained, thin sections.

#### 1.2.1.2 . Sectioning

Sectioning is often performed with a coronal incidence on a microtome with a blade, which is equipped with a cooling system that keeps the sample frozen. This sectioning stage delivers a series of 2D sections of several micrometers in thickness.

#### 1.2.1.3 . Staining

Staining is used to enhance the contrast of the tissue as well as emphasize specific areas and components of interest since biological tissues contain low contrast naturally. Afterward, the sections are mounted on glass slides for protection, preservation, and microscopic observation.

#### 1.2.1.4 . Imaging

The stained histological sections can be observed directly with a microscope but can also be digitized with a high spatial resolution scanner, making it possible to further investigate with digital image processing.

A wide range of stains can be employed during the staining process, from dyes and metals to tagged antibodies [115]. Hematoxylin and eosin (H&E) is one of the most widely used stainings, especially in the medical diagnosis of cancer. H&E is employed as a routine stain: the cell nuclei are stained a purplish blue

by hematoxylin, while the extracellular matrix and cytoplasm are stained pink by eosin. On the other hand, special stains or immunohistochemistry (IHC) detection are used when biologists want to investigate specific tissue structures or cell types. The following section will present the common stains for neuronal cells in brain tissues.

### 1.2.2 . Stains for brain tissues

The routinely used methods for staining neuronal cells include Nissl and NeuN.

#### 1.2.2.1 . Nissl

In contrast to H&E, which is a general staining, many techniques selectively stain cells, cellular components, and specific substances. For instance, Nissl staining (see Figure 1.7 left) is used to identify Nissl substance (clumps of rough endoplasmic reticulum and free polyribosomes) with cresyl violet (purple). The Nissl technique allows for the identification and differentiation of all brain cell types by staining neurons and glial cells differently within the same nervous tissue. This characteristic makes it the most appropriate technique for analyzing the cytoarchitecture of different structures throughout the brain.

#### 1.2.2.2 . NeuN

Antibodies have recently been employed to identify proteins, carbohydrates, and lipids. This approach is called Immunohistochemistry (IHC), which has substantially improved the ability to distinguish between cell types. NeuN (see Figure 1.7 right) is a protein, a neuronal nuclear antigen that is commonly used as a differentiation marker for neurons. NeuN immunoreactivity has been widely utilized to detect neurons in tissue culture and in histological sections in order to evaluate the neuron/glial cell ratio in different brain regions and to investigate neuron development and differentiation [48].

Other common biomarkers for brain histology contain Ionized calcium binding adaptor molecule 1 (Iba1) for microglia and Glial fibrillary acidic protein (GFAP) for astrocytes. It is important to underline that, in addition to the above-mentioned markers, there exists a very large diversity of markers that allow the precise study of biological tissues. The microscope makes it possible to observe tissue directly at the cellular level. And digitization and automated systems pave the way for many possibilities: storage, sharing and data exchange, and access to visualization and analysis software.

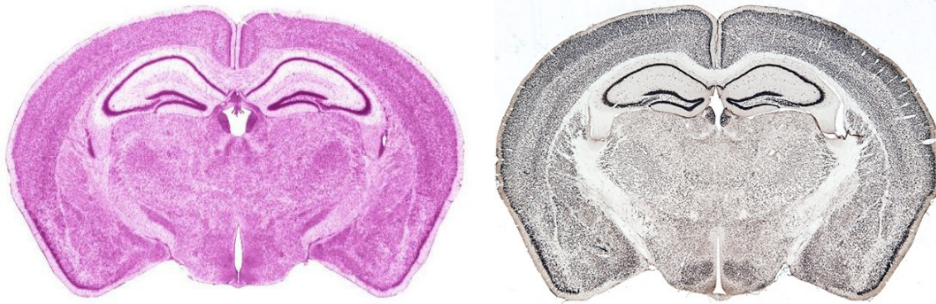


Figure 1.7: Coronal section from mouse brain stained for neurons using different biomarkers. Left: Nissl (image from Melbourne Brain Center<sup>5</sup>) and right: NeuN (image from Harvard Medical School<sup>6</sup>).

### 1.2.3 . Microscopy imaging

Biologists have long used optical microscopy to examine microscopic details and cellular structure in histological slices, which involves magnifying a sample by passing visible light transmitted through or reflected from it through a single lens or many lenses. In contrast to conventional transmitted light microscopy, the sample in a fluorescence microscope is illuminated by a specific range of light wavelengths through the objective lens. The fluorophores in the sample respond to this light by emitting longer wavelengths of light. Advances in digitization techniques have brought innovative approaches to improving histological image acquisition, such as Whole slide imaging (WSI) or virtual microscopy. This technique involves high-speed, high-resolution digital image acquisition of whole stained tissue sections, allowing analysis using a computer rather than a microscope. The use of WSI in digital pathology has continuously gained in popularity because it offers a practical way to assist disease diagnosis through interactive visualization and navigation of tissue slide images [49]. There are several key steps to digitize histology slides with WSI: scanning, storage, display, and analysis (see Figure 1.8).

With a high-resolution objective lens, tissue is scanned in a tile or linear pattern, and the digital representations are then coordinated and stitched together to form a seamless image of the entire slide [1]. In the tile pattern, the section is scanned as a series of square tiles (Figure 1.9 a and b). Image acquisition in a linear pattern is performed in long thin strips (Figure 1.9 c). Most scanners allow users to scan at different magnifications, and the most commonly used magnifications are  $20 \times (0.22 \times \mu m^2 / \text{pixel})$  and  $40 \times (0.11 \mu m^2 / \text{pixel})$ . Scanning at multiple focal planes (z-axis) is also possible and makes it possible to improve focus through tissue section thickness.

<sup>5</sup><https://www.melbournebraincentre.edu.au/>

<sup>6</sup><https://www.hms.harvard.edu/research/brain/atlas.html>

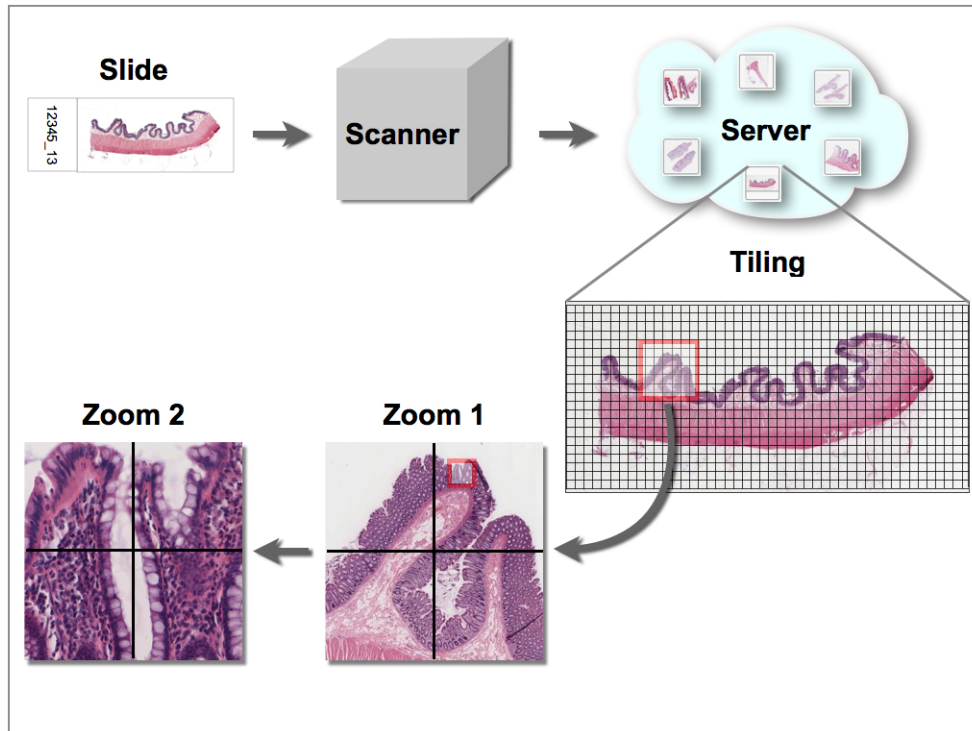


Figure 1.8: Digitalization of a histological specimen with a slide scanner for virtual microscopy<sup>7</sup>.

The high resolutions and potential multiple z-plane acquisitions increase the file size and acquisition duration tremendously. Image compression methods are frequently utilized to reduce file size, but they may introduce information loss during the conversion that cannot be recovered. Even with compression, the compressed image of the entire slide often surpasses 1 GB, making it difficult to display, share, or process. In practice, loading images at the highest resolution is not required to visualize the whole image. When biologists need to examine the slide at the maximum resolution, they generally focus on a specific region of interest, which does not require the entire image to be loaded [162]. Saving WSI data in a pyramid structure is an efficient way to store and transmit the image (see Figure 1.10): the WSI data is made up of multiple images of various resolutions (the altitude of the pyramid corresponds to the zoom level). The pyramid base represents the baseline image with full resolution. The thumbnail image is a low-resolution representation of the full image that can be displayed instantaneously. At intermediate resolutions, one or more intermediary layers of the pyramid may be established to enable the retrieval of image data at any resolution<sup>8</sup>.

<sup>7</sup><https://www.virtual-microscopy.net/>

<sup>8</sup><https://www.dicomstandard.org/>



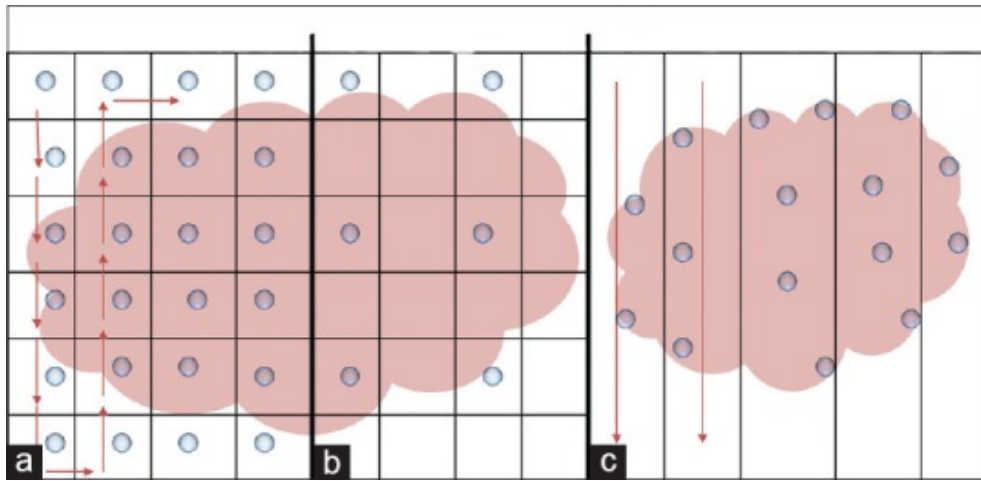


Figure 1.9: Tile and linear scanning patterns. (a) Tile scanning with “focus every field”, (b) tile scanning with “focus every  $n$ th field” and (c) linear scanning. Dots mean focus points [55].

Many WSI systems include image-viewing software that can be installed locally on computers. Annotating images, storing regions of interest, and exporting images to different formats are commonly supported. In addition to the functions mentioned above, other software solutions have been developed which allow applying image analysis algorithms automatically to whole-slide images, such as ImageJ [113] and QuPath [8].

CEA-MIRCen is equipped with a WSI Axio Scan Z.1 scanner<sup>9</sup> (see Figure 1.11) with an automated system that can scan up to 50 double or 100 single glass slides in a single session (for primate and rodent tissue section digitization, respectively). This system allows the scanning of the sections in both bright field and fluorescence modes. The Zeiss ZEN software, which operates the scanner and captures the images in CZI (Carl Zeiss Imaging) format, is a pyramidal file containing imaging data with relevant meta information. ZEN also supports exporting low-resolution images in other formats such as TIFF, JPEG, PNG, and GIF formats.

A medical image is an array of picture components called pixels or voxels that depicts the internal structure or function of an anatomical region. Medical image files consist of both metadata and pixel data. Metadata is the description of the image, which is typically saved as a header at the beginning of the file and contains the image dimensions, the spatial resolution, the pixel depth, the photometric interpretation, *etc.* The numerical values of the pixels are stored as pixel data. Pixel data are stored as integers or floating-point numbers using the fewest number of bytes required to represent the values. Neuroimaging

<sup>9</sup>Carl Zeiss Microscopy GmbH, Germany

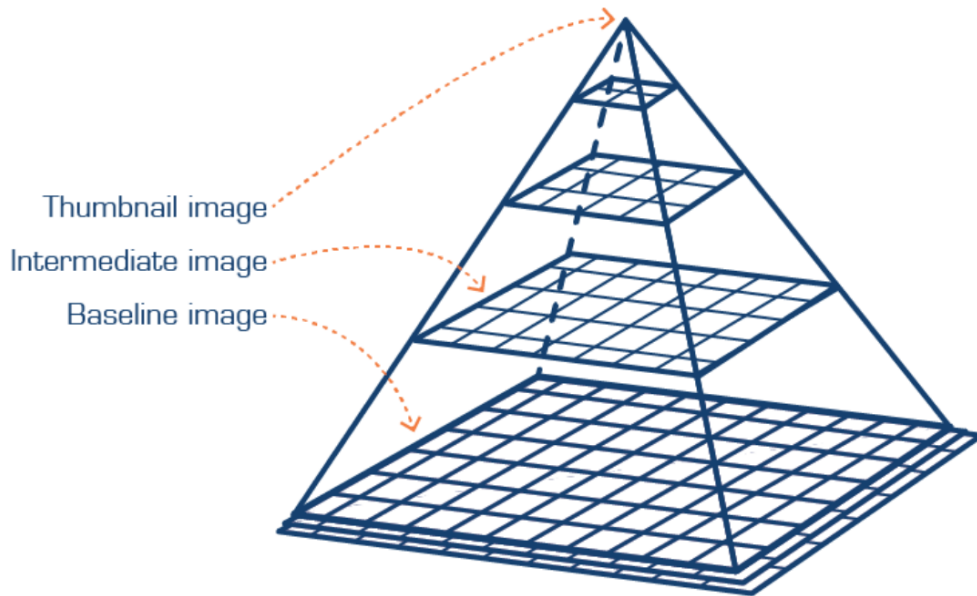


Figure 1.10: WSI as a pyramid of image data [118].



Figure 1.11: Left: Axio Scan Z.1 and right: image viewing software Zen. 1. Microscope, 2. objectives, 3. illumination, 4. cameras and 5. software.

Informatics Technology Initiative (Nifti) and Digital Imaging and Communications in Medicine (Dicom) are two popular file formats in medical imaging. Nifti allows the storage of the header and pixel data separately as .hdr and .img, and they are merged and stored as a single “.nii” file. In Dicom, meta-data and pixel data are combined into a single file, and the Dicom header also includes details on the image matrix and the most thorough description of the acquisition procedure and scan parameters used to generate the image [72].

At CEA, we use another in-house image format called GIS which is a simplified format compared to DICOM or Nifti formats. It is used for software development and data analysis. This format also supports 3D images. A GIS file contains three sub-files: IMA, DIM, and MINF:

The IMA file contains the basic pixel value matrix (equivalent to a raw file).

The DIM file is a minimal header file, which includes only the main properties of the image (size, resolution, type, *etc.*).

The MINF file is a header file that contains more detailed and additional properties of the image according to modality or image information (dimension, resolution, type, value range, specific metadata, *etc.*).

The advent of histology and virtual microscopy have made massive, high-resolution data available, and this has led to an explosion of methodological developments for the quantitative analysis of histological images. Segmentation of regions of interest, tumor detection and classification, and nuclei segmentation are important topics in this area. Machine Learning (ML) algorithms, and more recently, deep neural networks, are considered to be among the most effective tools in computer vision tasks. This thesis work investigates their application and performances in the context of neuron instance segmentation. Various ML techniques used in this work are introduced in the following section.

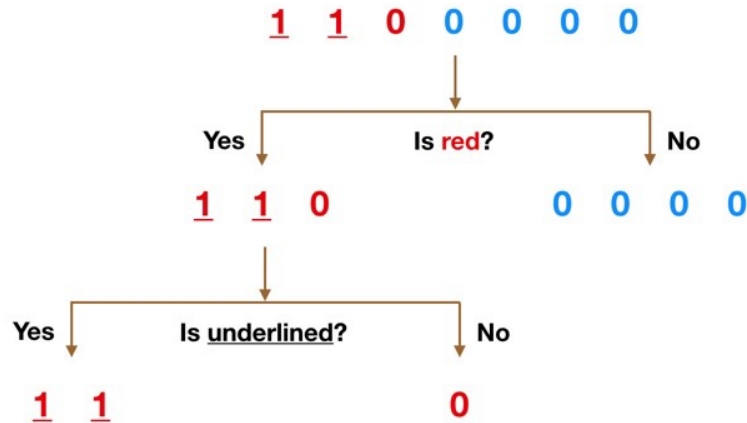


Figure 1.12: Decision Tree example (image from Towards Data Science<sup>10</sup>).

### 1.3 . Machine Learning

As part of artificial intelligence, ML algorithms are approaches that customize models to perform specific tasks by learning from training data. A well-trained model is able to make predictions or decisions without human intervention. ML methods are employed in a variety of speech recognition and computer vision applications. In this study, we will use both classical ML algorithms and Deep Learning (DL) methods, especially convolution neural networks (CNN).

#### 1.3.1 . Classical Machine Learning algorithms

According to the need for labeled training data, the learning process can be divided into supervised and unsupervised. Supervised learning is the most widely used method: it forms the predictions via a learned mapping  $f(x)$ , which produces an output  $y$  given an input  $x$ . A variety of ML algorithms has been proposed to estimate the mappings. Widely used ML algorithms include K-Nearest Neighbor, Support Vector Machines, Decision Trees (DTs), Logistic Regression, *etc.*

We used both Random Forest (RF) and Gradient Boosted Decision Trees (GBDT) in this study because of their good performance in classification and regression problems. Both of them are based on DTs. DTs are a non-parametric supervised learning method. The objective is to learn simple decision rules from data features and to build a model that predicts the value of a target variable. Similar to the flowchart, each internal node in the DT represents a “test” of a feature, each branch represents the result of the test, and each leaf node represents a category label. The classification rule is represented by the path from the root to the leaves. A simple decision tree example

<sup>10</sup><https://towardsdatascience.com/>

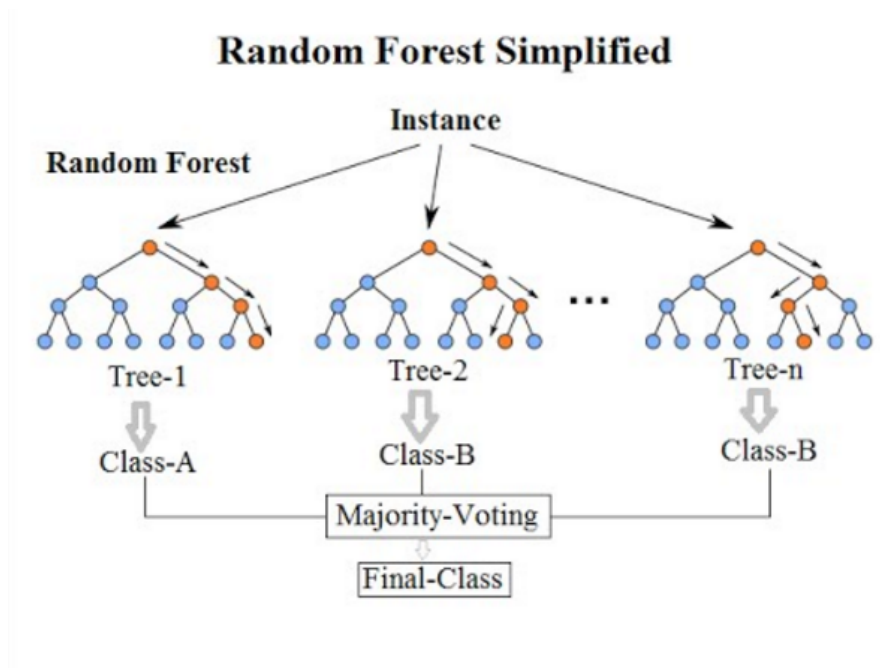


Figure 1.13: Simplified Random Forest example.

is shown in Figure 1.12. The ideal decision tree is one that accounts for the majority of the data while limiting the number of levels.

### 1.3.1.1 . Random Forest

Random Forest (RF) was first proposed in [50]. The idea is to use an ensemble of individual DTs that operate together. Each individual tree in the RF predicts a class label and the class, with the class with the highest votes becoming the final prediction. Figure 1.13 illustrates an example of the RF model, in which the final prediction will be “class B” based on majority voting. The term “random” reflects two aspects: bagging and feature randomness.

Bagging, also called bootstrap aggregating, is a technique used in RF to improve stability and accuracy. Each DT in the RF model is trained on a randomly selected subset of the training dataset with replacement. This procedure improves model performance since it decreases the variance without raising the bias. In particular, a single DT can be highly sensitive to noise in the training set, whereas the average of many trees, as long as the DTs are not correlated, is more robust. Bootstrap sampling can effectively de-correlate the DTs by training them with different sets.

RF also includes another type of bagging scheme called feature randomness, or feature bagging: selecting a random subset of features at node splitting in the learning process. A standard DT examines all possible features before

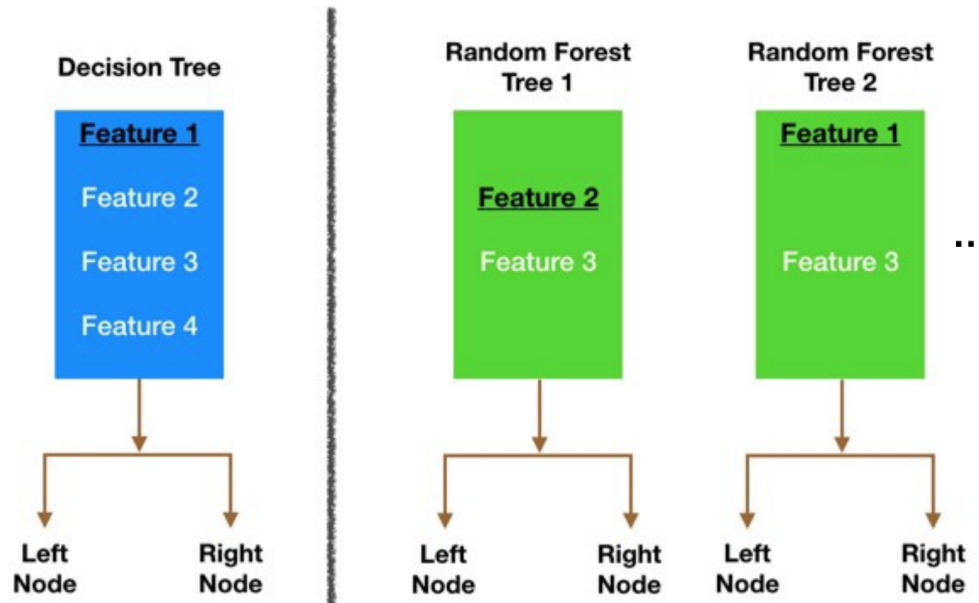


Figure 1.14: Node splitting in a decision tree model and Random Forest model (image from Towards Data Science<sup>11</sup>).

selecting the one that creates the greatest separation between observations in the left and right nodes. In a RF, on the other hand, each tree can only choose from a random features subset. This introduces even more variance among trees, leading to lower correlation and greater diversification. The difference in node splitting in RF and DT is shown in Figure 1.14.

### 1.3.1.2 . Gradient Boosted Decision Trees

Same as RF, Gradient Boosted Decision Trees (GBDT) [33, 34] is also strong ensemble learner of many weak DTs. However, they are highly different.

The main distinction between RF and GBDT is how decision trees are integrated. Unlike the bagging used in RF, GBDT uses a boosting technique to build the ensemble model. The boosting can be considered an iterative gradient descent algorithm. Each decision tree is additively concatenated with the goal of minimizing the error of the previous tree, as shown in Figure 1.15.

Suppose a GBDT model contains  $M$  subtrees, the output of stage  $m$  ( $1 < m < M$ ) is  $F_m$ , and the difference between the actual value and the predicted value is  $y - F_m$ . The objective of the next model,  $h_{m+1}$ , is thus to fit the residuals, which means:

$$F_{m+1}(x) = F_m(x) + h_{m+1}(x) = y \quad (1.1)$$

<sup>11</sup><https://towardsdatascience.com/>

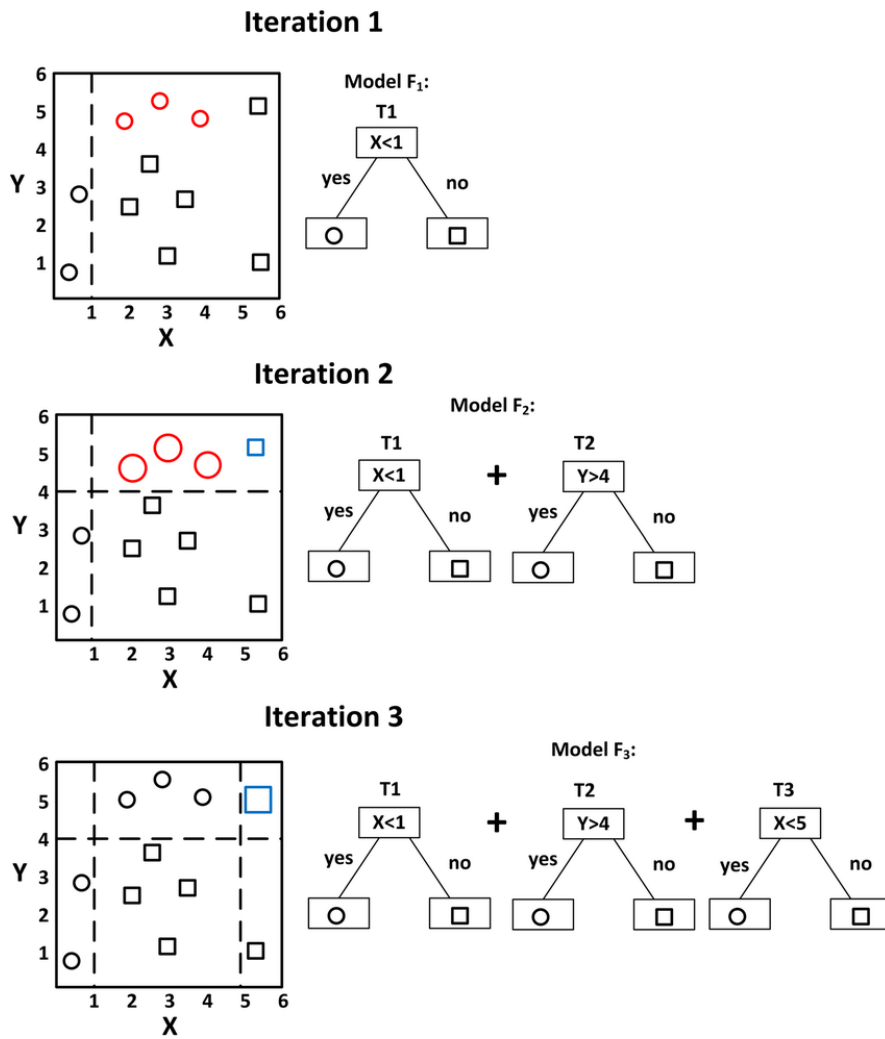


Figure 1.15: A simple example of gradient boosting [166].

Assume the loss function is the mean square error (MSE):

$$L_{mse} = \frac{1}{n}(y - F_m(x))^2 \quad (1.2)$$

where  $n$  is the number of samples. The gradient of  $L_{mse}$  equals:

$$\frac{\partial L_{mse}}{\partial F_m} = -\frac{2}{n}(y - F_m(x)) = -\frac{2}{n}h_{m+1}(x) \quad (1.3)$$

Therefore, for a given model, the residuals  $h_{m+1}(x)$  are proportional to the negative gradients of the MSE. In GBDT, the DTs do not match the complete data set. Instead, each tree fits the residuals of the previous tree, which gradually increases the accuracy and robustness of the model. Since the data used in DTs are different from one to another, all trees are uncorrelated. Both bagging (RF) and boosting (GBDT) reduce the variance. On the other hand, boosting keeps the focus on narrowing the gap between predicted and actual values by lowering residuals, which also reduces the bias.

### 1.3.2 . Deep Learning

Recently, as part of ML based on artificial neural networks (ANN), DL approaches have made many remarkable advances in computer vision, speech recognition, natural language processing, machine translation, *etc.* The term “deep” is related to the fact that the structure of artificial neural networks consists of multiple input, output, and hidden layers. DL has various advantages compared to classical ML.

#### 1.3.2.1 . Feature extraction

DL requires little human intervention. Unlike classical ML models, where the features are identified manually, the multiple layers in DL models extract progressively high-level features from the raw input in an automatic way. The different feature extraction of classical ML and DL methods is shown in Figure 1.16.

#### 1.3.2.2 . Accuracy

DL outperforms traditional ML when large datasets are available (*e.g.*, a ResNet-32 requires 20k training images to obtain an accuracy of 0.75 when performing classification on CIFAR-10) [75]. However, traditional ML algorithms are preferable for dealing with small datasets. To this end, many data augmentation techniques have been proposed to boost the performance of DL models [158, 87].



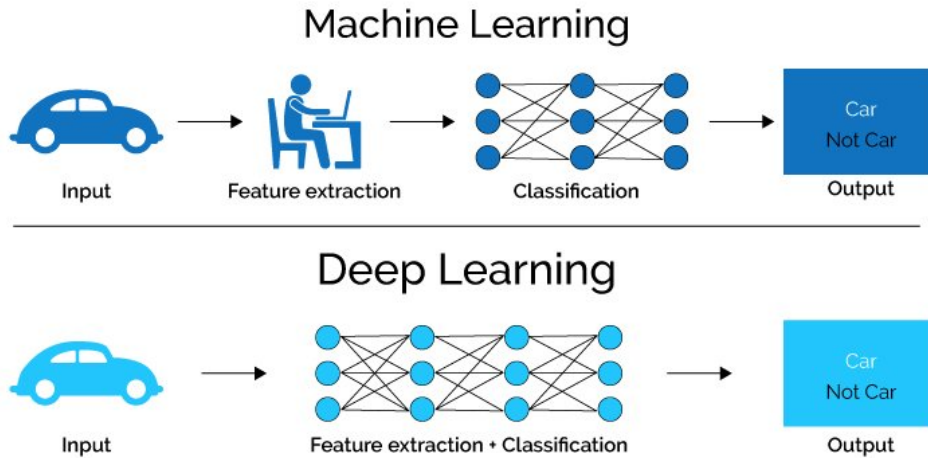


Figure 1.16: Comparison of feature extraction between ML and DL. (image from IEEE computer society<sup>12</sup>).

### 1.3.2.3 . Complex problem solving

When it comes to complicated issues like image classification, natural language processing, and speech recognition, deep learning is preferable since deep architecture can model non-linear relationships between input and output.

However, in addition to the need for large datasets, DL also faces other challenges, such as longer training time and high hardware dependencies (*e.g.*, GPU).

This thesis work focuses on DL models for medical image analysis. In this section, we will briefly introduce basic DL structures for computer vision, including the encoder networks that gradually reduce input resolution, the encoder-decoder networks that recover high-resolution representations, and Generative Adversarial Networks (GAN).

### 1.3.2.4 . Encoder networks

Most classification networks share the same design rules of LeNet-5 [143, 74], which we refer to here as the encoder, as shown in Figure 1.17 left. As the network gets deeper, the size of the feature map decreases gradually, and the depth (or the number of channels) increases, which is reflected by the thickness of the block in Figure 1.17. The encoder network learns low-resolution representations by concatenating convolutions from high resolution to low resolution, which can be regarded as encoding. The neural networks of this design include AlexNet [67], VGGNet [117], GoogleNet [126], ResNet [45], *etc.*

<sup>12</sup><https://www.computer.org/>

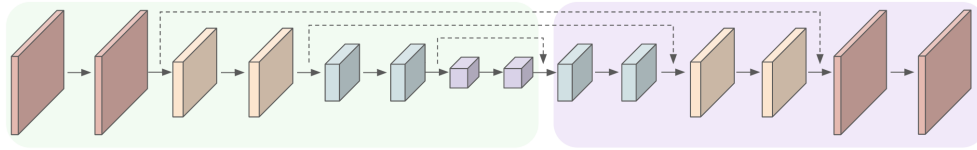


Figure 1.17: The structure of an encoder and decoder network. The green (left) represents the encoder network and the purple (right) represents the decoder network.

Encoder networks are suitable for tasks that do not require high-resolution information, such as classification, diagnosis, *etc.* On the other hand, for tasks like image detection and segmentation that require high-resolution representation, encoder architecture design is no longer applicable.

### 1.3.2.5 . Encoder-decoder networks

The decoder works in the opposite way to the encoder, as shown in Figure 1.17 right. It learns to read these compressed codes generated by the encoder and recover high-resolution representations needed for location-sensitive tasks such as object detection, semantic segmentation and instance segmentation. The decoder branch generally consists of deconvolution layers or transposed convolutional layers to upsample the resolution of the feature map. The typical networks which have the encoder-decoder structure include SegNet [7], DeconvNet [92], and Fully convolutional network (FCN) [80]. In addition, skip connections in UNet [105] concatenate the encoder and the decoder in order to retrieve context information at each resolution level. Most neural networks used in this study for semantic segmentation and instance segmentation are variants of UNet.

### 1.3.2.6 . Generative Adversarial Networks

Generative Adversarial Networks (GAN) was proposed by [38], consisting of a generative network and a discriminative network which are trained in an “adversarial” manner. Figure 1.18 illustrates classic GAN architecture. The generative network learns to generate a data distribution of interest (*e.g.*, images) from the latent space (a compressed, hidden representation of data points) or random noise. On the other hand, the discriminative network distinguishes data synthesized by the generator from the real data distribution. The objective of the generative network is to fool the discriminative network by synthesizing images that are similar to the real data distribution.

The discriminative network is generally an encoder-like CNN since its job is to distinguish the generated data from the real one, which can be seen as a classification task. A classical generative network, such as in the first [100, 38] has a decoder-like structure with transposed convolutional layers to generate images from a noise distribution. Recently, GAN has been applied to solve

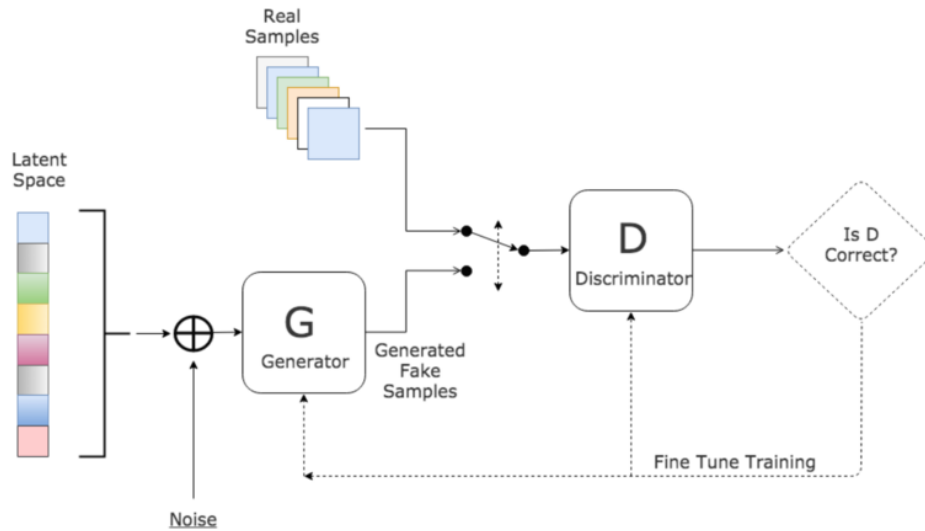


Figure 1.18: General GAN architecture (image from medium<sup>13</sup>).

image-to-image translation problems, and variants of GAN with an encoder-decoder-like generator have been developed. For example, the generators of Pix2Pix [57] and CycleGAN [169] can be applied to learn color mapping from input images to output images.

In this study, different GAN variants have been used to normalize the color variation on the histological images.

<sup>13</sup><https://medium.com/>

### 1.4 . Motivation and challenges

The brain is a complex organ, and scientists have been investigating it for decades to understand how it develops and ages. More recently, the rapidly increasing number of patients suffering from neurodegenerative diseases such as AD has prompted researchers to focus on better understanding this pathology. However, the cause of many neurodegenerative diseases is still unclear. One important characterization is the brain atrophy caused by the death of neurons, which can be observed directly with Magnetic Resonance Imaging (MRI) at advanced stages. However, the resolution of MRI images does not allow subtle quantitative analysis. With the advances in microscopy and WSI, we can now capture details at the cellular level.

Studies have shown that the number, distribution, and morphometric information of neurons are all important features in studying brain aging and neurodegeneration [131, 63, 138]. Such information cannot be obtained without the individualization of neurons. In practice, biologists rely on manual counting or segmentation to estimate neuron population in a region of interest (ROI), for example, the stereology technique [142, 148]. Nevertheless, manual identification of individual cells is extremely laborious and time-consuming, making it hard to undertake on large scales or in dense regions such as the hippocampus. Furthermore, studies based on manual segmentation are not reproducible since the results are closely related to the human experience. As a result, automatic segmentation methods are in high demand.

Recently, DL has shown great potential in medical image analysis. The results of nuclei segmentation competitions [16, 68] have revealed that DL-based methods outperform traditional ML methods in terms of accuracy and robustness. Most DL approaches, on the other hand, have been developed and validated for processing H&E-stained images, whereas the neurons are usually stained with NeuN by using immunohistochemistry (IHC). Neuron individualization using DL is a less explored research area. One limitation of DL methods lies in their demand for large annotated datasets, which are not always available. In addition, the prediction of neural networks is a probability map, which needs a post-processing step to obtain the individualized instances. Recent attention has focused on improving network architecture, while post-processing, in contrast, has not been well investigated.

WSIs contain several resolution layers, and the data with the highest resolution is extremely large. A macaque brain section, for example, is roughly  $200,000 \times 200,000$  pixels at resolution  $20\times$ , and the file size is about 150 GB. However, because of the GPU memory constraints, DL-based methods often handle small images (usually not larger than  $1024 \times 1024$  pixels). But images with this dimension represent only a small region, which is not sufficient for many

analytic purposes. Cell segmentation on large-scale WSI images remains a challenge.

Another issue in histological image analysis is color variation, which can be introduced during the staining process or the digitalization. This color inconsistency bothers the diagnosis of pathologists, and it may degrade the performance of Computer-Aided Diagnosis (CAD) systems. As a result, techniques of color normalization are often applied as a preprocessing step for DL-based methods. Nevertheless, these techniques themselves may also have an impact on image analysis, which has not been adequately examined. Moreover, most existing color normalization methods [132, 82] have been explicitly designed for H&E-stained images, which differ significantly from IHC in terms of staining type. The applicability of these methods on IHC images and particularly NeuN-stained images, remains an open question.

Considering all remaining issues, the objective of my thesis work is to develop a framework based on DL for automatically individualizing neurons in various anatomical brain regions on a large scale with minimal manual effort. This method will be precise, computationally affordable, and robust to color changes.

Specifically, the thesis aims to:

1. Investigate a strategy for semi-automatically generating large amounts of annotations to minimize manual efforts.
2. Propose a general framework consisting of a neural network and post-processing to individualize neurons in different anatomical regions and species.
3. Adapt the framework to process large-scale images on HPC resources.
4. Examine existing color normalization techniques on NeuN-stained images.
5. Quantify the degradation of segmentation performance due to color variation.

Related research works will be presented in the next chapters and through several publications (two international conference papers and a journal article).



## 2 - Literature review

### 2.1 . Automatic cell segmentation methods

Advances in microscopy, especially virtual microscopy, provide us with rich information at the cellular level. Single-cell identification on microscopic images is a prerequisite for many quantitative studies [16]. The accuracy of segmentation is critical for tracking single cells and retrieving characteristics such as cell number, size, and morphological information. As manual segmentation is extremely laborious, numerous automated methods have been proposed to address this issue.

#### 2.1.1 . Conventional methods

##### 2.1.1.1 . Thresholding

By turning a gray-scale image into a binary image based on a threshold, thresholding [94] is the simplest and fastest approach for segmenting cells from tissue. Pixels with an intensity greater than the threshold are assigned a value of 1, while the rest are assigned a value of 0. Only one global threshold is estimated automatically over the entire image, so lightly stained cells may be missed. Later methods have brought improvement by adding more thresholds and browsing an image with a sliding window instead of handling the entire image at once [116, 15]. However, thresholding methods only distinguish objects from the background. They are not able to identify touching or overlapping cells.

##### 2.1.1.2 . Morphological operations

Several studies have investigated cell segmentation using morphological operations [112, 144]. Specifically, erosion operations are performed on a binary image with increasing scales to obtain the markers for each cell (touching cells are well-separated), and then the markers are labeled and expanded with iterative dilation to restore the original cell shapes. Recently, mathematical morphology became popular as a preprocessing step to facilitate later segmentation [156]. Dorini *et al.* [26] use a multiscale morphological toggle operator to regularize contours before applying image segmentation [137].

### 2.1.1.3 . Watershed transform

Watershed transform [85, 104] was the most popular approach for cell instance segmentation before the advent of DL segmentation methods. Watershed transforms the image into a landscape, with the intensity reflecting elevation. The landscape is flooded with water from regional minima, which correlate to low-elevation locations, and creates dams to prevent water from different basins from combining. The boundaries of the dams are known as watershed lines, and they divide the landscape into regions corresponding to cells.

Because of intensity differences in the foreground and background, the watershed transform may be prone to over-segmentation [156]. Watershed can significantly alleviate over-segmentation with controlled markers as regional minima. Thus, watershed transform is often combined with automatic nuclei detection methods [101, 164] since manual markers are not always available. Yang *et al.* [157] used feature-preserving non-local means filtering to reduce noise in gray-scale images so that the markers used to initialize the watershed transform could be easily detected. Another approach to address over-segmentation is to define prior criteria to merge miss-segmented regions into real nuclei, for example, a measurement score to compare regions before and after merging [76, 77]. More recently, watershed transform has been applied as a popular post-processing technique of DL-based segmentation methods [16, 68].

### 2.1.1.4 . Graph-cut

Graph-cut methods address image segmentation as a graph partitioning problem that can be formulated in terms of energy minimization [153, 12, 29]. These methods represent an image as a weighted graph, with each node representing a pixel or superpixel in the image, and each edge is a weight between two nodes representing the similarity between neighboring pixels or superpixels. The graph is divided into many sets, each representing an object in the image, based on a set of criteria.

Al-Kofahi *et al.* [3] proposed a two-stage graph-cut method for nuclei segmentation. First, the image foreground is automatically extracted using a graph-cut-based binarization. Then, nuclei are detected using Laplacian-of-Gaussian filtering and distance maps. A second graph cut and graph coloring are applied to obtain the final segmentation. Lou *et al.* [81] demonstrated significant quantitative improvement by integrating a “blob”-like shape prior to the graph-cut framework. He *et al.* [46] proposed a normalized cut algorithm using a weighting matrix, which is calculated based on spatial location, intervening, and concave contours to separate the aggregated cells into individual cells. However, graph-cut methods have been used only for simple cases with-



out massive clustering of cells. In addition, these methods are not suitable for samples with varying cell sizes, as an a priori fixed cell size is required.

### 2.1.1.5 . Other methods

You *et al.* [160] proposed a framework for individualizing size-varying and touching neurons. First, a multi-scale series of Gaussian filters are applied for denoising. The centroids of neurons are then identified using a min-max filter. Finally, a competitive region-growing algorithm is used to find instance borders, with expansion constrained by the binary segmentation of a trained RF. The process is applied iteratively until the best filters are found.

### 2.1.2 . Deep learning-based methods

Recently, DL-based segmentation methods show better accuracy and robustness compared to traditional methods. Table 2.1 shows a summary of current state-of-the-art methods. According to the definition of loss function: they can be divided into the following categories:

#### 2.1.2.1 . Distance map regression

In literature, distance transform is a simple cell detection method that assigns a distance to each pixel of the cell class to the nearest feature point, which is commonly chosen as the edge pixels in a binary image. As a result, the local maxima of the computed distance map are usually matched to the cell centroids [156]. More recent attention has focused on predicting distance transform using neural networks [69]. Naylor *et al.* [91] addressed the cell segmentation in H&E-stained images as a distance map regression task (DIST). Given an input histological image, a fully convolutional network (FCN) is trained to estimate the distance of each pixel of the cell class from the nearest background pixel. The distance map is then normalized to the range  $[0, 1]$  to handle the size differences between cells. The final instance segmentation relies on two thresholds: one for classifying the foreground from the background and the other for controlling the size of segmented objects. Liu *et al.* [79] improved this method by adding a network branch to predict cell contours, and this additional information led to more accurate boundary predictions. Schmidt *et al.* presented StarDist to simulate the regular roundish cell shapes with star-convex polygons [111]. In order to define star-convex polygons, a neural network (UNet) is trained to predict the object probability and the distance in 64 radial directions simultaneously. After that, duplicate shapes are pruned using non-maximum suppression (NMS). This approach can accurately identify cells, although it is not pixel-accurate. In their further study, this issue was addressed by using stricter bounds during NMS [147].

### 2.1.2.2 . Pixel-wise classification

Another popular strategy is to tackle cell segmentation as a pixel-wise classification or semantic segmentation task, which can be extended to cell instance segmentation with post-processing. For example, binary classification of cells and tissue, while it is unable to separate touching cells. More recently, researchers have addressed this problem by adding a third class of cell contour, so that cell segmentation becomes a ternary classification of cell interiors, background, and cell boundaries. Cui *et al.* [23] proposed a convolutional neural network (CNN) that predicts nuclei and boundaries simultaneously, resulting in prediction maps of nuclei and boundaries. The nuclei map is then thresholded, labeled, and dilated to get the final segmented nuclei. The winning method of nuclei segmentation competition Data Science Bowl 2018 [16] showed that segmentation performance could be further improved by predicting only the boundaries between cells rather than entire cell contours. Still, the output of the network is a prediction map of three classes on which complex post-processing is applied for individualizing cells. Three thresholds are selected to generate three sets of cell instance candidates, and a GBDT model is trained to choose the best of them.

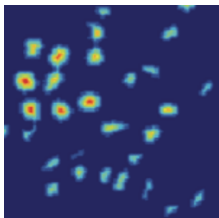
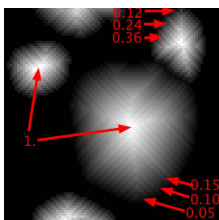
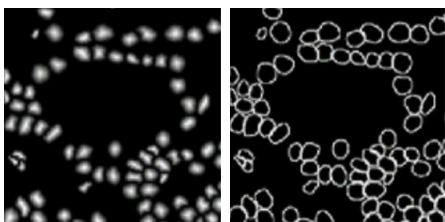
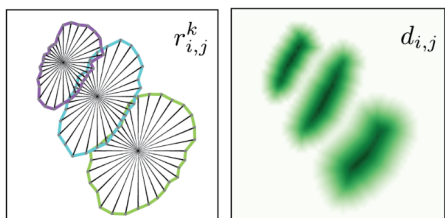
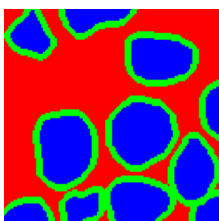
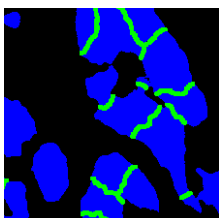
### 2.1.2.3 . Fortright instance segmentation

One landmark in DL-based instance segmentation is Mask R-CNN (Region-Based Convolutional Neural Network) [44]. The network consists of two stages. The first stage involves a Region Proposal Network, which offers multiple ROIs in a given image with bounding boxes. In the second stage, Mask R-CNN predicts a binary segmentation for each ROI in addition to predicting the class and box offset. Several studies have attempted to adapt this method for cell segmentation and achieved state-of-the-art results [51]. One advantage of Mask R-CNN over pixel-level classification is that it allows the segmentation of overlapping objects since each object is processed independently. On the other hand, Mask R-CNN is computationally more expensive for the same reason. Stringer *et al.* [123] designed a multi-task strategy to perform cell instance segmentation. A UNet is trained to predict vertical and horizontal gradients and cell probability. Pixels that flow to the same centroid are aggregated, and individual cell instances are defined by tracking gradient paths. Furthermore, cell shapes are refined with the cell probability map.


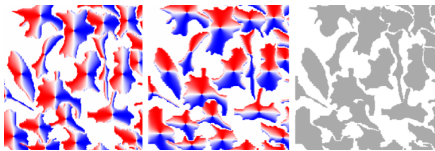
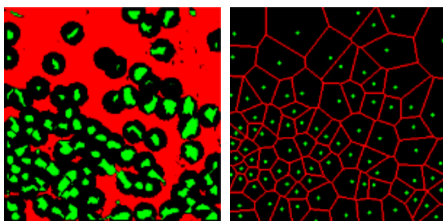
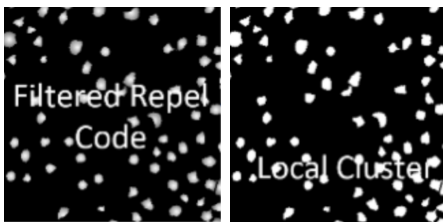
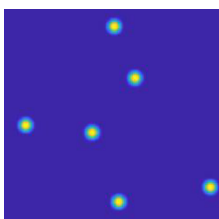
### 2.1.2.4 . Center detection and label extraction

The good performance of the aforementioned methods requires large datasets and pixel-level masks for training, which is both time-consuming and labor-intensive if carried out manually. Researchers have investigated cell segmentation using only point annotations to alleviate the labeling process. These methods can usually be divided into centroid detection and pixel-level label extraction. Qu *et al.* [98] proposed a framework that consists of two training stages for detection and segmentation, respectively. Cell locations can be detected with the trained detection model, and the voronoi diagram and k-means clustering are applied to automatically generate coarse labels for the segmentation network. Likewise, Chamanzar *et al.* [18] developed a multi-task training scheme that simultaneously learns cell detection and instance segmentation. Point annotations are fed both directly to guide the detection task and indirectly to guide the segmentation task by combining local pixel clustering and repel codes. You *et al.* [161] proposed a multiscale CNN to detect the center of neurons with different sizes. Detected cell centers are then used as seeds to initialize a competitive region-growing with a binary segmentation as the mask to obtain the final instance segmentation.

Table 2.1: Current cell segmentation methods based on Deep Learning.

Methods	Type	Ground Truth	Network	Comments
Kumar <i>et al.</i> [69]	Distance map regression		Two CNN	First CNN to detect nuclei and second CNN to classify patch-wise label
DIST [91]	Distance map regression		FCN	FRegression of the distance map
MCDNet [79]	Distance map regression		Two branched UNet	Dilated convolution, residual operations
StarDist [111]	Distance map regression		UNet	Prediction of star-convex polygons
Cui <i>et al.</i> [23]	Pixel-wise classification		UNet	Prediction of three classes (cell, background and contour)
Topcoders [16]	Pixel-wise classification		UNet	Prediction of three classes (cell, background and inter-cell contour)

## 2.1. AUTOMATIC CELL SEGMENTATION METHODS

nucleALzer [51]	Instance segmentation		Mask R-CNN and UNet	Mask R-CNN for segmentation, UNet for post-processing (correction)
CellPose [123]	Instance segmentation		UNet	Residual blocks, prediction of vertical, horizontal gradients and cell probability
Qu <i>et al.</i> [98]	Center detection and label extraction		UNet with ResNet-34 as encoder	Two networks for semi-supervised nuclei detection, followed with a segmentation network
Chamanzar <i>et al.</i> [18]	Center detection and label extraction		UNet with ResNet-50 as encoder	Segmentation network trained with repel code and local cluster (generated from point annotations)
You <i>et al.</i> [161]	Center detection and label extraction		Multiscale fully convolutional regression neural network	Detection network with segmentation post-processing

### 2.1.3 . Whole Slide Imaging segmentation

Whole slide images (WSI) are very large images that contain a high-resolution digitization of an entire tissue section or biopsy. Cell segmentation in whole slide images presents a number of challenges due to the large size and high resolution of the images. Some specific challenges of nuclei segmentation in whole slide images include:

1. Large image size: whole slide images can be several gigapixels in size, making them much larger than traditional microscopy images. This can make it difficult to process the entire image in a reasonable amount of time, especially if the method being used is computationally intensive.
2. Large number of nuclei: whole slide images often contain a large number of nuclei, with tens of thousands or even millions of cells present in a single image. This can make it difficult to accurately segment all cells, especially if there is overlap between cells or if the cells are irregularly shaped. For methods that iteratively process each object, the computational cost will increase exponentially.
3. Variability between regions: several anatomical regions appear in whole slide images, while cells can vary significantly from one region to another in appearance, with differences in size, shape, and intensity. Therefore, a robust segmentation method that works well for cells from different anatomical regions with different morphologies is needed.

Many segmentation methods cannot be directly applied to WSI. Deep learning-based (DL) segmentation methods are designed to work with images that are typically not larger than  $1024 \times 1024$  pixels and require the use of graphics processing units (s) to process the images efficiently. However, when it comes to segmenting large-scale images such as whole slide images, which can be several gigapixels in size, simply downsampling the images to a smaller size is not always an effective approach. This is because downsampling results in the loss of information, which can negatively impact the accuracy of the segmentation.

One alternative approach that has been proposed for handling large-scale images is patch-based segmentation. This involves dividing the original image into smaller patches, running the segmentation operations on each patch independently, and then stitching the patches back together to reconstruct the original large-scale image. This allows the segmentation to be performed on smaller, more manageable chunks of the image rather than the entire image at once. This can be an effective approach for segmenting large-scale images, although it may be more computationally intensive than simply downsampling the image.

Another issue with DL segmentation is the inaccurate prediction of image boundary regions. Because the predictions of the same pixel near the border of adjacent patches are not always consistent, this problem may generate a mosaic effect in the stitching results at the boundaries [23]. Currently, there are two common strategies to address this: overlapping patch extraction and decreasing the weight of boundaries.

Guo *et al.* [39] and Khened *et al.* [66] extracted image patches with an overlap of 50% of the patch size and averaged multiple predictions in the overlapped region after inference. Such a large overlap can effectively reduce the border effect, but it also results in a large number of redundant computations. For large-scale images, such as entire histological slices, this supplementary cost may become significant. In a probability map of CNN, the closer one pixel is to the edge, the less accurate its prediction is. Prediction weighting aims to lower the impact of border pixels while increasing the contribution of center pixels. Isensee *et al.* [56] proposed applying a Gaussian importance weight map before the stitching process. A similar scheme can be found in [23]. Rather than assigning equal weight to each pixel, assigning appropriate weight to each pixel better matches the nature of the probability map. In this circumstance, 50% of overlap may no longer be necessary. Nevertheless, the method of [23] can only segment cells on images with a resolution of  $1000 \times 1000$  pixels, which is insufficient for anatomical region-level analysis. The approaches of [39, 66, 56] are designed to perform semantic segmentation at the ROI level. So far, segmentation at the cellular level on large-scale WSI images remains a challenge.

## 2.2 . Color normalization for histological images

One challenge in histological image analysis is the undesirable variation in color, which can affect the diagnosis of pathologists as well as the accuracy of Computer-Aided Diagnosis (CAD) systems. A lot of factors can be associated with this issue, including the concentration of staining solution, staining duration, room temperature, sample volume, different manufacturers, *etc.* Thus in practice, the standardization of staining is difficult to manage. Furthermore, variables during the digitization process, such as different scanners, settings, and specifications of the scanner, may also introduce variations. Data augmentation and color normalization are two common ways to improve the robustness of DL methods to color variation. The former is used throughout the training process: data with diverse color appearances is fed into the network to improve its generalization capacity. Vasiljević *et al.* [135] proposed an unsupervised method based on GANs to enhance the diversity of stainings in the dataset. They demonstrated the benefit of this approach in improving the staining invariance of a glomerular segmentation network. However, performance may suffer in specific stains, and it may still fail when dealing with colors that were not present during the training. The latter, color normalization, is widely used as a preprocessing step in inference to ensure the model performance: the test set is mapped into the same color space of a reference image (generally from the training set) before applying the trained model.

### 2.2.1 . Conventional methods

#### 2.2.1.1 . Color matching methods

Histogram specification, or histogram matching, is the transformation of an image in order to match a reference histogram [40]. After this procedure, the source image will have similar intensity and color statistics as the target image. Histogram specification would produce exact results in the continuous case, but in practice, images are usually restricted to 8-bit values, which is a discrete case. Malandain *et al.* [83] normalized the intensity variations under different acquisition settings using histogram matching. In contrast to methods based on discrete histograms, a windowing methodology is introduced that allows matching continuous probability density functions and achieves more robust results. Instead of blind registration between histograms, Dauguet *et al.* [25] proposed to detect the range of intensities corresponding to different tissues before matching histograms. In order to improve robustness, an iterative algorithm is used to update these detections until a consistent tissue classification is achieved on consecutive slices. Exact histogram specification [21] transforms images into K-dimensional space, a continuous-valued space, and small random values are added to induce a strict ordering. However, histogram matching-based methods may bring artifacts in the processed image since it treats the color histogram independently from the image content.



## 2.2. COLOR NORMALIZATION FOR HISTOLOGICAL IMAGES

---

It may especially fail if the color distributions of the source and target are significantly different.

Another popular strategy is the color transfer algorithm [102], which was originally proposed to alter color in natural scene images. They show that the  $l\alpha\beta$  color space with decorrelated axes can be a powerful tool for manipulating image color. The color channels are aligned to match that of the reference in this space. More specifically, the goal is to match the color statistics (means and standard deviations) of two images. However, in histological images, because the transformation is applied at the image scale without separating the stain color and the tissue color, this method may cause incorrect color mapping.

### 2.2.1.2 . Stain-separation methods

Several stain-separation methods have been proposed for normalizing color in histology images, which often have multiple stains (*e.g.*, H&E). Non-negative matrix factorization (NMF) based algorithms address color normalization as an optimization problem. In [99], images are factorized into color appearance matrix and stain depth matrix since they are both non-negative, and the distance between source and target is reduced for decomposed matrices separately. Macenko *et al.* [82] proposed to transform images from RGB color space to optical density (OD) space, where color vectors can be easily separated. Each staining channel is then normalized independently using singular value decomposition (SVD). Khan *et al.* [65] proposed an image-specific color deconvolution method: a supervised classification framework is trained to label each pixel in the image with the appropriate stain, which results in a deconvolution stain matrix. Further, the statistics of the deconvolved source and target are mapped nonlinearly for each stain channel. By adding a sparseness regularization component to the stain depth matrix, the structured preserving color normalization (SPCN) with sparse NMF (SNMF) outperformed the standard NMF and SVD methods [132]. Figure 2.1 illustrates the diagram of SPCN. Janowczyk *et al.* proposed a sparse autoencoder to partition source and target images into tissue segments. Thereafter, Histogram matching and alignment are performed for each cluster rather than all pixels of the image.

### 2.2.2 . Deep learning-based methods

Recent studies have explored color normalization in histological images using neural networks. Generally, based on generative models (GAN), color normalization is addressed as an image-to-image style transfer task. Bentaieb *et al.* [9] proposed combining a generative model with a downstreaming task-specific network (classification or segmentation). The generative model is a style transfer model trained to alter source images with staining properties learned from the reference image set. The task-specific model is a discriminative model that is trained to assess how realistic the generated image is, as well as perform

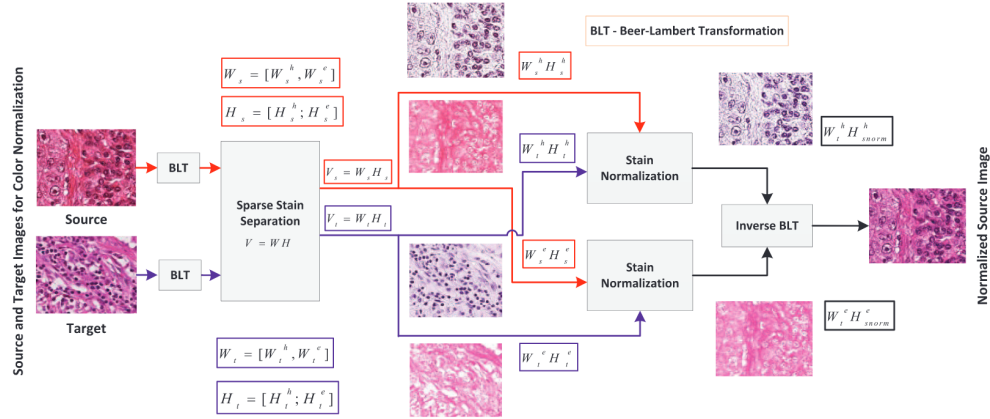


Figure 2.1: Flow diagram of structured preserving color normalization [132].

the auxiliary work. StainGAN, proposed by Shaban *et al.* [114] employed the Cycle-Consistent Adversarial Networks (CycleGAN) [169] to perform the color transfer as an unpaired image-to-image translation. Figure 2.2 shows the framework of StainGAN. The networks consist of two generators and two discriminators that perform non-linear mapping in two directions (source to target and adverse). Generator A is trained to alter the source color with the target color, and generator B is trained to map recolor images (generated by A) into the original color distribution. The additional task in [9] and the second generator and discriminator in [114] both have the same purpose of enforcing the generative models to preserve the semantic information of source images when altering the stain color.

Salehi *et al.* [107] built a stain-to-stain framework based on Pix2Pix [57], a conditional GAN that learns to apply the desired style to input images. As shown in Figure 2.3, the training set is the required reference template, which is used to generate gray-scale images and is also the ground truth that guides the generator to recolor the gray-scale images. At the same time, the discriminator learns to distinguish between a real pair (gray-scale image and the original color image) and a false pair (gray-scale image and the image recolored by the generator). The fundamental distinction of this approach is that the generative model maps from the gray-scale image rather than the source domain to the target domain. This strategy is less computationally expensive than the aforementioned GAN-based methods since it does not require an auxiliary task network [9] nor a second generator and discriminator [114]. Another advantage is that it does not have a prerequisite of source domain: the trained model can be applied to recolor images with different colors as they will all be converted into gray-scale. However, on the other hand, the gray-scale conversion may result in information loss, which has not been quantified and can be problematic for subsequent analysis.

## 2.2. COLOR NORMALIZATION FOR HISTOLOGICAL IMAGES

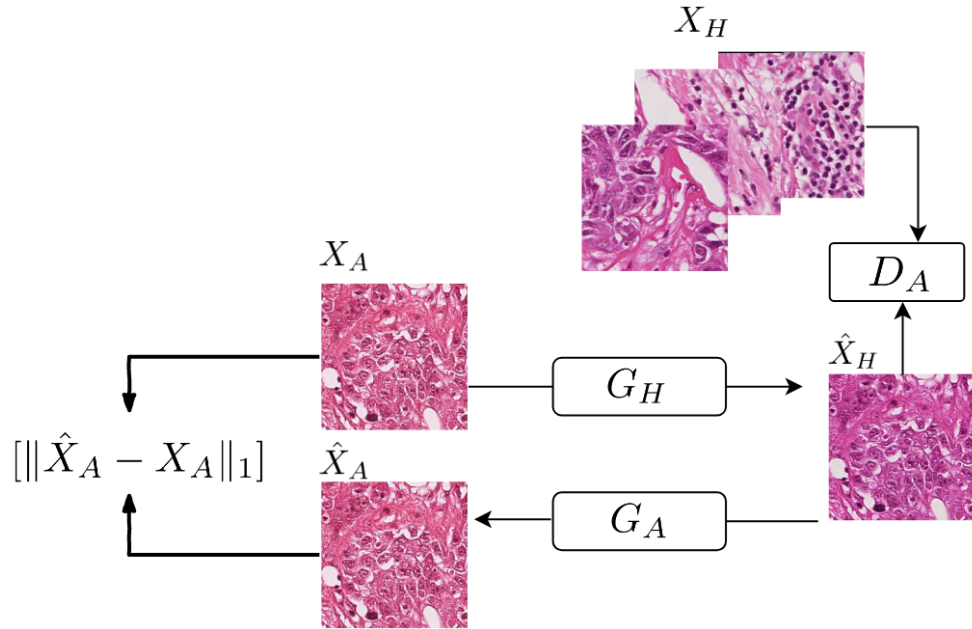


Figure 2.2: StainGAN framework [114].

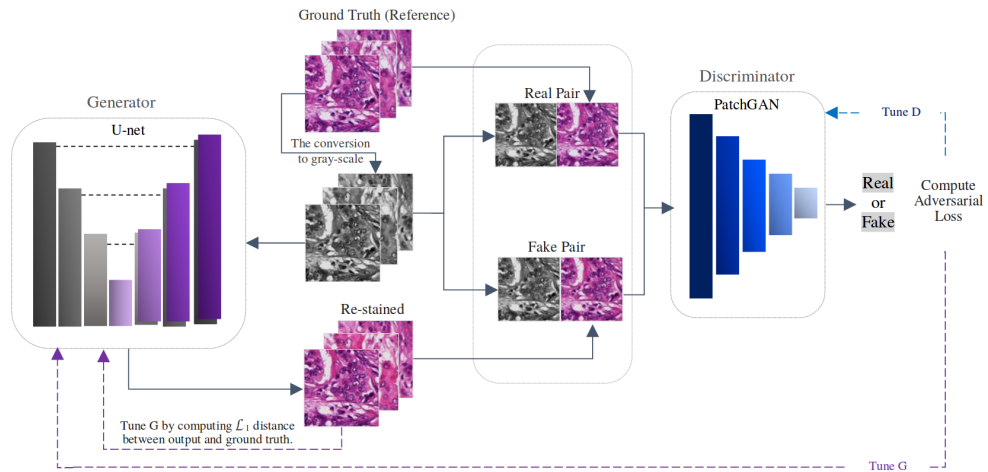


Figure 2.3: Pix2Pix framework [107].

## CHAPTER 2. LITERATURE REVIEW

---

The benefit of DL-based algorithms is that they learn color information from the entire training set rather than a single target image, resulting in more robust results. Furthermore, they eliminate the requirement for a target image, which is often picked manually by experts.

### 2.3 . Conclusion

In this section, we first reviewed automated methods for cell instance segmentation, both conventional and DL-based. Graph-cut and watershed were the two most widely-used methods until the revolution of DL. Recent competitions in cell instance segmentation [16, 68] have demonstrated the superiority of DL-based methods in both accuracy and robustness. We classified current DL-based methods into four groups according to the type of ground truth required for the training. Most of them are fully supervised and rely on large annotated training datasets to achieve good performance. However, annotated datasets in the biomedical field are rare, and manual annotation is extremely laborious, especially at the cellular scale. Moreover, the post-processing step in most works has not been detailed, although it plays a vital role in the final segmentation. In addition, we also reviewed current studies on whole-slide image processing. Several works have investigated the segmentation of ROIs on whole-slide Images with a workflow based on patch extraction, DL inference, and prediction assembling. To date, performing cell instance segmentation on whole-slide Images is still a challenge.

Segmenting neuronal cells in large-scale microscopic images is the primary objective of this thesis. Additionally, we are also interested in the correction of color variations in microscopic images, which can have a negative effect on segmentation. The second section reviewed current color normalization methods for microscopic images. Often used as a preprocessing step in DL-based cell instance segmentation, color normalization aims to match test images to the same color distribution as training images. Conventional approaches require a reference image as the target and try to apply the same color distribution to test images. DL-based approaches are often based on GAN. Generative networks are trained to map one color distribution to another. One advantage of these methods is that the target image is no longer necessary. However, one trained generator is limited to one specific color mapping—normalizing color to another color reference necessitates additional training time and effort.



## 3 - Automatic generation of large-scale synthetic annotations and evaluation of Topcoders

### 3.1 . Introduction

A challenge of DL-based methods is the demand for large annotated training datasets, which are not always available. In practice, the annotation is often performed manually. Manual segmentation is not only tedious and time-consuming but also requires human expertise. To this end, it is preferable to annotate datasets in an automated manner in order to accelerate the labeling process and minimize manual effort. This chapter presents a semi-automatic approach developed for generating synthetic annotations for a macaque neuron dataset, and an evaluation of a state-of-the-art segmentation method named Topcoders using the annotated dataset.

### 3.2 . Biological dataset: a healthy macaque brain

The dataset of this study came from a nine-year-old macaque brain. In total, 134 histological coronal sections stained by IHC using NeuN were produced (as shown in Figure 3.1). Due to a large amount of data, sections were divided into two series (even and odd-numbered) and stained separately, which resulted in a color variation between the two series: the even-numbered sections are dark, while the odd-numbered sections are brown. Figure 3.2 illustrates the color variation between series. And then, they were scanned by an AxioScan.Z1 (Zeiss) with the in-plane resolution of  $0.22 \mu m/\text{pixel}$  ( $20 \times$  magnification). One section at this resolution is approximately 120 GB.

Most neuron instance segmentation work of this thesis (Chapter 3 and 4) was conducted in section 91 (as shown in Figure 3.1 and Figure 3.3). This section is a representative section that contains the major anatomical regions of the brain. We chose to begin with this section as it presents a large diversity in terms of population and distribution of neurons in the brain. In a previous study [160], 50 images of  $5000 \times 5000$  pixels were extracted covering the main anatomical regions of this part, including the cortex, hippocampus, caudate, striatum, thalamus, claustrum, and putamen. The image locations are presented in Figure 3.3. We retrieved 30 images out of 50 to construct our dataset for neuron instance segmentation. Further, 24 images were selected and cropped into 11 616 image patches of  $224 \times 224$  pixels to build our training set. The remaining six images were utilized as the test set to evaluate the segmentation performance.

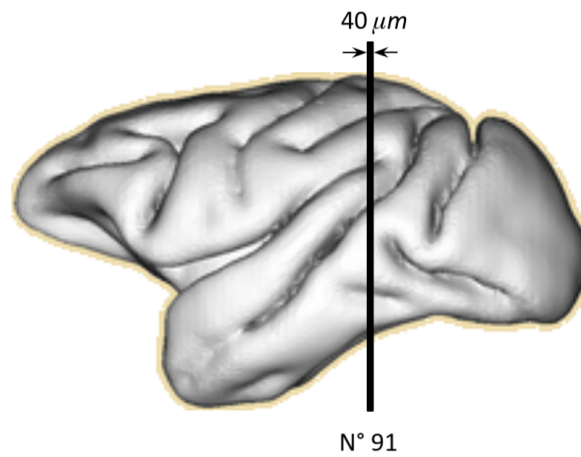


Figure 3.1: Illustration of extracted coronal histological section (sagittal view of macaque brain surface).

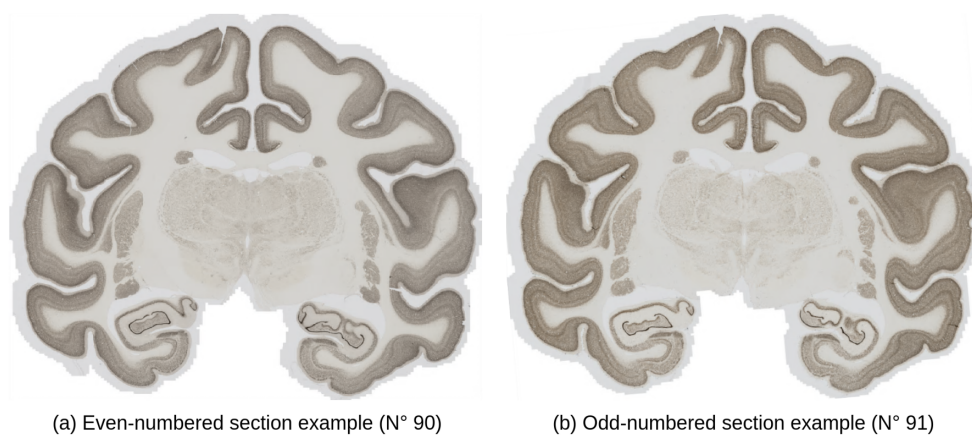


Figure 3.2: Two section examples with color variations (90: dark and 91: brown).



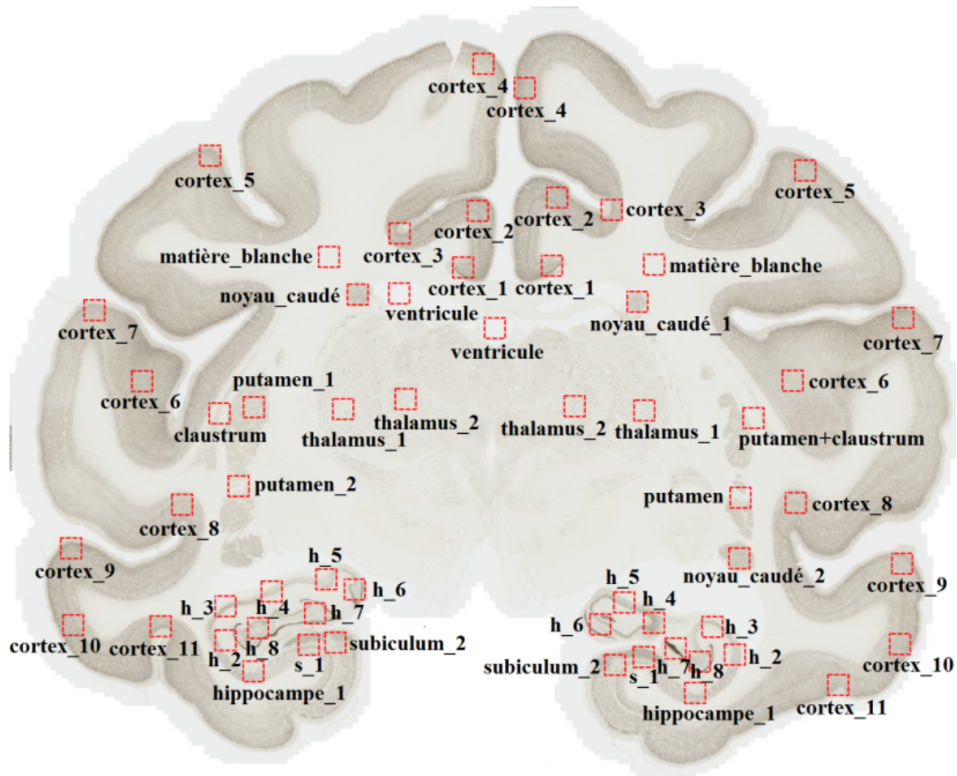


Figure 3.3: Section 91 of macaque brain. The red squares show the images of  $5000 \times 5000$  pixels [159].

This study will be extended to other sections of the brain in the next step. However, one issue that needs to be addressed first is the color variation between the two section series. Thus, the second objective of this thesis is to normalize colors in microscopic images, which will be investigated in Chapter 6 of the thesis.

### **3.3 . Semi-automatic pixel-level mask synthesis**

The capacity of the DL model depends on labor-intensive, time-consuming pixel-level annotations. In order to minimize manual labeling effort, we designed a semi-automatic strategy based on a competitive region-growing algorithm [160] to synthesize pixel-level annotations. The only manual work required with this approach was to locate the neuron centroids that were used as seeds to initialize the expansion, which was constrained with a RF binary mask. Furthermore, we applied morphological transformations to generate pixel-level masks of three classes, including neurons, background, and inter-cell contours. This pipeline will be briefly presented in Section 3.4 as part of a conference paper. The following subsections present this method in greater detail.

#### **3.3.1 . Random Forest binary segmentation**

The RF model of binary segmentation was retrieved from prior work [11]. A hundred DTs of the RF model were trained to classify neurons from the tissue. In contrast to DL models, which require a large dataset with annotations, the RF model was trained on a small dataset, including 100 representative NeuN images of  $512 \times 512$  pixels. Additionally, a feature selection strategy was applied to determine a small optimal feature subset (both color and texture) from a pre-selected feature family, leading to a faster and more robust segmentation.

#### **3.3.2 . Competitive region-growing**

A contour-based model was used to perform the region growth process [160]. The expansion was initialized with point annotations that were manually identified and labeled. During the process, the expanding rate was controlled by a penalty map of intensity and contour curvature. After each iteration, an average filter was used to smooth the contours. The expansion terminated when the contour reached the RF segmentation border. The majority filter ( $3 \times 3$  pixels) then assigned unlabeled pixels to a neighboring label.

#### 3.3.3 . Three-class-mask generation

The aim of this process was to synthesize masks to train neural networks. As mentioned in the Chapter 2, several types of ground truth exist in the literature. Inspired by the winning approach of the nucleus segmentation challenge [16], we addressed neuron instance segmentation as a three-class semantic segmentation task. In particular, the inter-cell contour class aimed to separate touching neurons in dense regions. Each neuron in the image was given a unique label after the region-growing process, and the non-stained tissue was labeled as zero. The boundary between two touching neurons was defined as the inter-cell contour class. The thickness of the inter-cell contour should be an even number given that it was equally contributed by two touching neurons. A thickness of four pixels ( $\sim 1 \mu m$ ) was defined empirically.

The synthesized three-class masks were used to train the DL segmentation models. We referred to the nuclei segmentation challenge of Data Science Bowl 2018 [16], which attracted 3 891 participants as a reference. The winning method, called Topcoders, contains eight neural networks and a GBDT model to refine the segmentation. The following section presents the evaluation of this method on our dataset using the synthesized masks as the ground truth.

### 3.4 . Evaluation of Deep Learning Topcoders Method for Neuron Individualization in Histological Macaque Brain Section

The remainder of this chapter comprises the following manuscript:

Wu, Huaqian, Nicolas Souedet, Zhenzhen You, Caroline Jan, Cédric Clouchoux, and Thierry Delzescaux. “Evaluation of Deep Learning Topcoders Method for Neuron Individualization in Histological Macaque Brain Section.” In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 2985-2988. IEEE, 2021.

#### **Abstract**

Cell individualization has a vital role in digital pathology image analysis. Deep Learning is considered as an efficient tool for instance segmentation tasks, including cell individualization. However, the precision of the Deep Learning model relies on massive unbiased dataset and manual pixel-level annotations, which is labor intensive. Moreover, most applications of Deep Learning have been developed for processing oncological data. To overcome these challenges, i) we established a pipeline to synthesize pixel-level labels with only point annotations provided; ii) we tested an ensemble Deep Learning algorithm to perform cell individualization on neurological data. Results suggest that the proposed method successfully segments neuronal cells in both object-level and pixel-level, with an average detection accuracy of 0.93.

#### **3.4.1 . Introduction**

Lack of information about neuron population, distribution and morphology at cell level has existed as a critical problem for the study of brain development and aging for many years. Traditionally, neurobiologists estimate manually the number of neurons in the region of interest. However, this method is tedious and subjective because its accuracy relies on the complexity of images. Several automated cell individualization methods have been proposed such as Watershed [22] and iCut [46]. However, these methods have several limitations. Watershed algorithm can be easily affected by noise in the images, often resulting in over- and under-segmentations. The iCut algorithm proposed to segment touching cells, fails in the regions where massive cells aggregate, and does not take into account size-varying cells such as neurons. Recently, the development of Deep Learning (DL) has revolutionized computer vision. Integrating DL models in cell detection and cell instance segmentation improves accuracy compared to traditional approaches [28, 23, 91, 69]. To achieve robust and rigorous segmentation, a large training dataset is mandatory. However, pixel-level annotation is laborious and time-consuming. In addition, aforementioned methods are designed mainly for analyzing oncological data, in particular H&E staining, whereas in neuroscience, the study of neurons usually relies on NeuN staining. Moreover, the task of neuron seg-

mentation is extremely challenging due to the variety in neuron shape, size and density in the brain. To the best of our knowledge, few studies have specifically investigated cell individualization for neurons [160].

In this paper, we propose a weakly supervised DL method for neuron instance segmentation, which requires only point annotations. The main contributions of this work are as follows: i) inspired by [160], we developed a new strategy to synthesize pixel-level labels using Random Forest (RF) segmentation and a competitive Region Growing algorithm; ii) we tested different configurations of DL networks on NeuN-stained images. The DL architecture used in this work is Topcoders, which ranked first during the nucleus segmentation competition: Data Science Bowl 2018 (DSB) and demonstrated promising performances on H&E staining images and fluorescent images stained with DAPI and Hoechst [16] and iii) by employing an overlapping patch extraction/assembling method [23], we were able to process large high-resolution images despite the limitation of GPU random access memory (RAM).

### **3.4.2 . Materials and Methods**

#### **3.4.2.1 . Dataset**

Dataset was derived from a previously published study [160], a representative histological section with thickness of  $40 \mu m$  was obtained from a healthy macaque brain and scanned by an AxioScan.Z1 (Zeiss) with the resolution of  $0.22 \mu m/\text{pixel}$  ( $20 \times$  magnification) ( $\sim 150$  GB). Based on 30 images of  $5000 \times 5000$  pixels, 24 images were chosen to create the training dataset including 11k patches ( $224 \times 224$  pixels), 6 other images were chosen as the test dataset. The datasets contained the main anatomical regions where the neuron distribution is investigated including cortex, hippocampus, caudate, *etc.* The datasets presented a rich diversity in terms of neuron shape and size, with both sparse and highly aggregated neuron distributions, as illustrated in Figure 3.4.

#### **3.4.2.2 . Pipeline of pixel-level label synthesis**

Traditionally, cell instance segmentation is addressed as a binary classification problem, the output of the classifier contains two classes: cells and tissue background. Recent works indicate that a more accurate segmentation can be achieved by classifying pixels into three classes: inter-cell contours, interior of cells and background [16]. We designed an algorithm for constructing pixel-level labels including these three classes, based on centroid point labels provided by the expert for each neuron and binary semantic segmentation results produced in previous works [160] (as shown in Figure 3.5. i) The binary segmentation of neurons and background was generated by applying a RF model of 100 decision trees, which was trained and optimized with following features: H, S, V color channels and local intensity [160]. ii) Manually

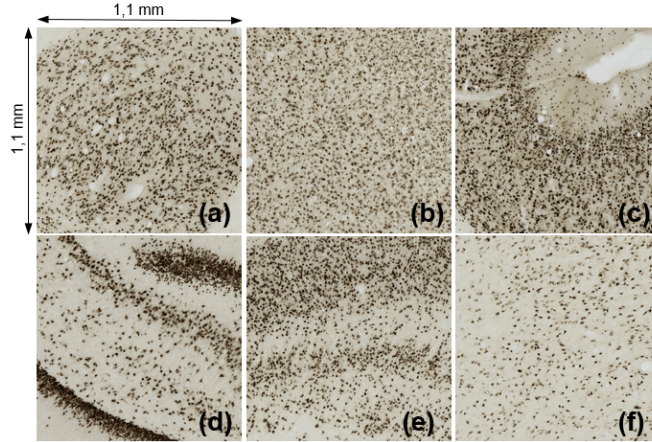


Figure 3.4: Training dataset examples: (a) caudate, (b-c) cortex, (d) hippocampus, (e) subiculum and (f) thalamus.

pinpointed centroid labels were identified by an expert, a disk with a radius of 5 pixels in the center of the neuron was marked as a point label (size and form adapted to visual checking and region-growing initialization). iii) A competitive region-growing process was applied to define pixel-level labels, the expansion of sub-region was constrained using the segmentation result of RF. Each pixel inside the neuron had the same label as its centroid. By applying the contours generated by region-growing on raw image Figure 3.5 (d), we visually assessed the quality of synthesized labels; iv) based on the pixel-level labels, three-classes masks were generated using morphological operations, including background, neurons and the border of touching neurons (thickness of 4 pixels), which is proved to enhance the segmentation result [16].

### 3.4.2.3 . Neural network architecture

We applied the winner algorithm of the 2018 Data Science Bowl [16], an ensemble model of 8 UNet-like encoder-decoder architectures, with encoders pre-trained on ImageNet database, including three ResNets (34, 101, 152) [45], two Dual Path Networks (DPN) 92 [20], two DenseNets (121, 169) [53] and one Inception-ResNet [125]. Such strategy also enabled us to compare the performance of the different neural networks individually and to evaluate the effectiveness of the entire model. Pretrained model derived from DSB (Topcoders<sub>bowl</sub>) [16], model trained on the neuron dataset (Topcoders<sub>neuron</sub>) and its constituents (8 models) were all tested. Training dataset was randomly divided into two groups ( $\frac{3}{4}$  and  $\frac{1}{4}$ ) for training and validation respectively. A heavy data augmentation was applied to prevent over-fitting, including rotation, flipping, channel shuffling, color inversion, *etc.* The training set was expanded to 6 times compared to its original size. Once the training of DL models was accomplished, a post-processing step aiming to optimize the seg-

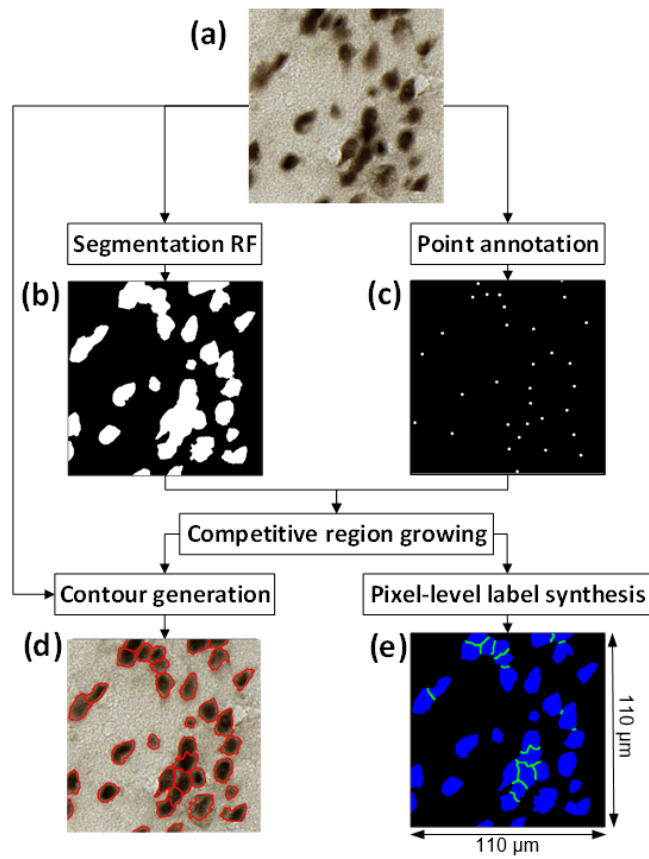


Figure 3.5: Flowchart of pixel-level label synthesis: (a) raw image, (b) RF segmentation, (c) point labels, (d) region-growing result (red contours) and (e) synthesized masks (black: background, blue: neurons and green: inter-cell contours).

mentation results was applied: a regression model (gradient-boosted trees) was trained to predict Intersection-over-Union (IoU) for all cell candidates, candidates with small predicted IoU ( $< 0.3$ ) were removed in order to decrease false predictions.

#### 3.4.2.4 . Overlapping extraction & assembling

The test dataset contained 6 large high-resolution images ( $5000 \times 5000$  pixels) which included various anatomical regions. However, most DL based segmentation algorithms cannot process such large-scale images due to GPU RAM limitation. Moreover, one constraint of CNN is that the prediction at the border of the input image is not accurate due to the lack of context information. To address this problem, we adopted the strategy of overlapped patch extraction and assembling proposed in [23]. The patches were extracted from raw images by a sliding window of  $1340 \times 1340$  pixels (determined according to the GPU RAM resources), a stride of 1220 pixels in height and width produced an overlap of 120 pixels. The prediction results of patches were seamlessly stitched to reconstruct the final result ( $5000 \times 5000$  pixels) using the same settings as well as a weight map, which was applied to each predicted patch so that the pixels closer to the edge of the patch have lower weights. The use of the weight map reduced the impact of inaccurate prediction of pixels at the border of the patches.

#### 3.4.2.5 . Evaluation metrics

To evaluate the proposed method, we computed F1 score (F) for both detection (det-F1) and instance segmentation (seg-F1) tasks:

$$P = \frac{TP}{TP + FP}; R = \frac{TP}{TP + FN}; F = 2 \frac{P \times R}{P + R} \quad (3.1)$$

where True Positive (TP), False Positive (FP) and False Negative (FN) represent the numbers of true, false and missing detection/instance segmentation respectively. For the detection task, a neuron detected was considered as a true positive when it was superimposed with exactly one centroid defined by the expert. As for the instance segmentation, the true positive was defined as the IoU greater than 0.5 between the detected neuron and the corresponding label.

Dice coefficient was calculated to evaluate the semantic segmentation. Another evaluation criterion was the relative count error (RCE):

$$\varepsilon = \frac{|N_a - N_e|}{N_e} \quad (3.2)$$

Where  $N_a$ ,  $N_e$  are the number of neurons detected by the proposed method



and the expert respectively.

#### 3.4.2.6 . Training details

Models were trained using PyTorch, Keras and Tensorflow frameworks. Each model needed different epoch number to converge (from 17 to 70 depending on network depth), with Adam optimizer and a starting learning rate of 1e-4 which decreased during the training. DL models with ResNet101, ResNet152 and one of DPN encoders used sigmoid activation, the other models used softmax activation. For loss calculation, the combination of soft dice and binary cross-entropy/categorical cross-entropy was chosen for sigmoid/softmax activation respectively.

This work was conducted on a workstation equipped with bi-processors (operating system: Ubuntu 16.04 LTS 64-bits, CPU: Intel Xeon E5-2630 v3 at 2.4 GHz, RAM: 128 GB, GPU: NVIDIA GTX 1080Ti).

#### 3.4.3 . Results

The results of  $\text{Topcoders}_{\text{bowl}}$ ,  $\text{Topcoders}_{\text{neuron}}$  and its constitutive models (named according to the encoder network) were evaluated on a test dataset including  $\sim 16\text{k}$  neurons. Table 3.1 reports the performance of neuron detection (det-F1), instance segmentation (seg-F1), semantic segmentation (Dice) and neuron counting (RCE). Although the training dataset of  $\text{Topcoders}_{\text{bowl}}$  did not include neuron data, it was able to detect most neurons correctly (det-F1 score of 0.83 and RCE of 0.22). Nevertheless, it performed less well in both instance (seg-F1: 0.71) and semantic (Dice: 0.75) segmentation. A significant improvement of 10% was obtained by training with NeuN-stained data.  $\text{Topcoders}_{\text{neuron}}$  achieved the highest detection accuracy (det-F1: 0.927), it was also one of the best models for instance segmentation (seg-F1: 0.87). Among the constitutive models, ResNet101 outperformed others in neuron detection (det-F1: 0.926) and counting (RCE: 0.037). DPN softmax was the best model for instance and semantic segmentation (seg-F1: 0.88 and Dice: 0.95), followed by ResNet34 (seg-F1: 0.87 and Dice: 0.93).

To better compare the performance of each model, the detection accuracy (det-F1) distribution of the 10 models are plotted in Figure 3.6.  $\text{Topcoders}_{\text{neuron}}$ , which had the best average detection accuracy on the test dataset, was also one of the most robust models, it performed well for all tested anatomical regions. The best constitutive model, ResNet101, was also robust, whilst it performed less well for most tested anatomical regions than  $\text{Topcoders}_{\text{neuron}}$ .

Figure 3.7 shows cropped examples of synthesized masks and segmentation results in three anatomical regions. Neurons are presented by colored labels. The results are displayed according to an increasing density of neurons (first row: caudate, sparse; second row: cortex, dense and last row: hippocampus,

Table 3.1: Detection and segmentation performance of Topcoder<sub>bowl</sub>, Topcoder<sub>neuron</sub> and its constitutive models. Best and second best results are in bold with the best also underlined.

Model	det-F1	seg-F1	Dice	RCE	
Topcoder <sub>bowl</sub>	0.825	0.705	0.751	0.219	
Topcoder <sub>neuron</sub>	<b><u>0.927</u></b>	<b>0.870</b>	0.928	0.040	
Constituents	DenseNet121	0.910	0.783	0.830	0.063
	DenseNet169	0.912	0.839	0.869	0.062
	DPN sigmoid	0.869	0.803	0.885	0.130
	DPN softmax	0.918	<b><u>0.880</u></b>	<b><u>0.951</u></b>	0.059
	Inception-ResNet	0.901	0.848	0.910	0.116
	ResNet34	0.914	<b>0.870</b>	<b>0.934</b>	0.088
	ResNet101	<b>0.926</b>	0.868	0.921	<b><u>0.037</u></b>
	ResNet152	0.923	0.865	0.923	<b>0.038</b>

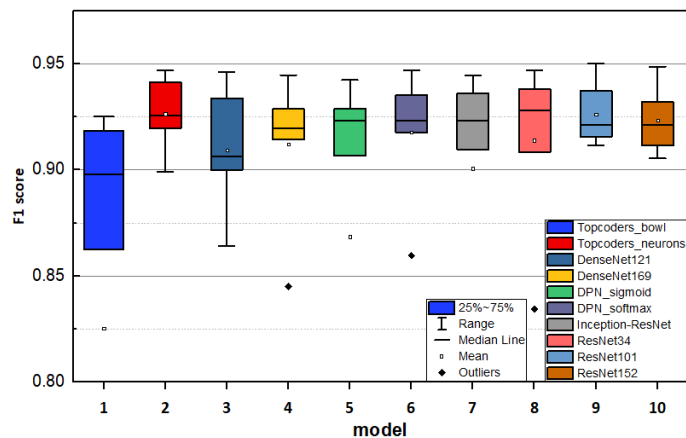


Figure 3.6: Comparison of detection performance (det-F1) on the test set (several outlier values are below the vertical scale range).

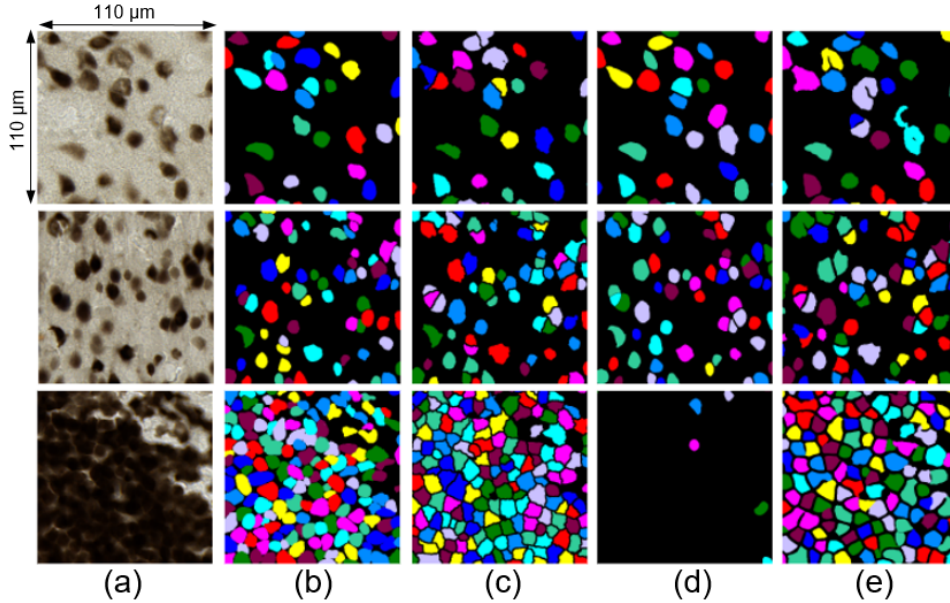


Figure 3.7: Results of pixel-level label synthesis and segmentation in three anatomical regions. Top row: caudate; Middle row: cortex; Bottom row: hippocampus. (a) Raw images, (b) manual segmentation, (c) synthesized pixel-level masks and (d, e) results of  $\text{Topcoders}_{\text{bowl}}$  and  $\text{Topcoders}_{\text{neuron}}$  respectively.

very dense). For better illustrating the results, the neurons in these regions were segmented manually, as shown in Figure 3.7 (b). Figure 3.7 (c) shows the synthesized pixel-level masks, compared to the manual annotations, the proposed method produced satisfying masks for most regions (both in distribution and shape based on visual evaluation). Figure 3.7 (d) illustrates the results of  $\text{Topcoders}_{\text{bowl}}$ , the neurons were correctly detected in the sparse regions, but the contour of neurons were often distorted. Moreover, the region of massive touching neurons was wrongly considered as the background. Figure 3.7 (e) presents the predictions of  $\text{Topcoders}_{\text{neuron}}$ , it was a solid model performing very well for all tested anatomical regions, at both object-level and pixel-level.

#### 3.4.4 . Discussion

$\text{Topcoders}_{\text{bowl}}$  was believed to be a well generalized model based on DSB 2018 results. It successfully separated neurons in regions with sparse distribution while it did not respect the neuron shape and it failed in regions where massive neurons aggregated. This is probably due to the absence of cells with various forms and highly clustered distribution such as neurons in the dataset of DSB.

The performance of  $\text{Topcoders}_{\text{neuron}}$  demonstrated the superiority of the ensemble model. Although the results of  $\text{Topcoders}_{\text{neuron}}$  were obtained by combining the predictions of the constitutive models, it achieved better results

than most constituents for all the tasks.

Among all the constitutive models, all ResNet backbone models performed well, including ResNet34, the best model for semantic and instance segmentation, and ResNet101, the best constitutive model for neuron detection. Generally, a deeper network can capture more complex features. While the deepest model ResNet152 did not achieve the best performance in any task, it might be related to the fact that deeper networks are generally more difficult to train owing to the vanishing gradient problem [127]. Another interesting finding is that the choice of activation had an important influence on segmentation results. Although two DPN models had exactly the same architecture, the one with softmax activation performed better than that with sigmoid activation. Since most tested neural networks achieved good results, we believe that it is feasible to apply DL techniques for neuron counting. However, compared to stereology [148], which takes into account the thickness of tissue for possible superimposition of cells and provides unbiased quantifications, DL methods can only deal with 2D images and provide a valuable estimation of cell counting. A specific dataset and study need to be designed to quantify discrepancies between DL methods and stereology.

#### **3.4.5 . Conclusion**

In this work, we investigated the ability of a weakly supervised method to specifically detect and segment neurons in NeuN-stained histology images. By applying state-of-the-art DL architectures, this study provides the first comprehensive assessment of different neural networks for neuron individualization. An optimal model trained using neuron data was obtained, it was able to separate size, shape and density-varying neurons successfully. Experimental results in the main anatomical regions demonstrated the effectiveness of the proposed method against the default DSB model. The current study was carried out with the default settings, further optimization in training parameters and architecture need to be investigated. Besides, developments in high-performance computing (HPC) are also ongoing to test the efficiency of cross-validation. Further work is required to establish a comparative analysis of Topcoders and other deep learning-based instance segmentation methods, as well as stereology – the reference method used in biomedical analysis. An exciting perspective will be to extend this study to whole sections and brains, which will improve our understanding of brain development, aging and neurodegeneration.

#### **Compliance With Ethical Standards**

The experimental procedures involving animal models described in this paper were approved by the Institutional Animal Care and Ethics Committee.

#### **Acknowledgment**

This work was supported by DIM ELICIT's grants from Région Ile-de-France. This work was granted access to the HPC resources of TGCC under the allocation 2019-(A0040310374) made by the GENCI.



## 4 - A General Deep Learning framework for Neuron Instance Segmentation based on Efficient UNet and Morphological Post-processing

### 4.1 . Introduction

In the previous chapter, we demonstrated the effectiveness of a state-of-the-art approach called Topcoders for neuron instance segmentation. This approach involves training an ensemble model consisting of eight U-Net-like networks, followed by a post-processing phase using a gradient boosting decision tree (GBDT) model to reduce segmentation errors. While the ensemble model achieved superior detection accuracy compared to the individual networks, it required a significant amount of computational resources and time for the training and the inference. Additionally, some of the individual networks performed similarly to the ensemble model, but the post-processing step added additional computational overhead by requiring the prediction of the best threshold for each cell. These factors made the Topcoders approach less suitable for processing large images.

To address these limitations and enable the efficient processing of large-scale images, we present a novel end-to-end neuron instance segmentation framework in this chapter. Our approach utilizes a single neural network with an optimal balance between precision and computational efficiency, as well as an original post-processing scheme. We demonstrate that our approach outperforms Topcoders and other methods with a simpler model and easier parameter settings.

### 4.2 . A General Deep Learning framework for Neuron Instance Segmentation based on Efficient UNet and Morphological Post-processing

The remainder of this chapter comprises the following manuscript:

Wu, Huaqian, Nicolas Souedet, Caroline Jan, Cédric Clouchoux, and Thierry Delzescaux. “A general deep learning framework for neuron instance segmentation based on efficient UNet and morphological post-processing.” *Computers in Biology and Medicine* (2022): 106180.

**Abstract**

Recent studies have demonstrated the superiority of deep learning in medical image analysis, especially in cell instance segmentation, a fundamental step for many biological studies. However, the excellent performance of the neural networks requires training on large, unbiased dataset and annotations, which is labor-intensive and expertise-demanding. This paper presents an end-to-end framework to automatically detect and segment NeuN-stained neuronal cells on histological images using only point annotations. Unlike traditional nuclei segmentation with point annotation, we propose using point annotation and binary segmentation to synthesize pixel-level annotations. The synthetic masks are used as the ground truth to train the neural network, a UNet-like architecture with a state-of-the-art network, EfficientNet, as the encoder. Validation results show the superiority of our model compared to other recent methods. In addition, we investigated multiple post-processing schemes and proposed an original strategy to convert the probability map into segmented instances using ultimate erosion and dynamic reconstruction. This approach is easy to configure and outperforms other classical post-processing techniques. This work aims to develop a robust and efficient framework for analyzing neurons using optical microscopic data, which can be used in preclinical biological studies and, more specifically, in the context of neurodegenerative diseases.

**keywords:** Neuron instance segmentation; Deep Learning; Mathematical morphology; Histological images; Optical microscopy.

**4.2.1 . Introduction**

Advances in microscopy allow scanning whole slide images, capturing details at the cellular level and revealing the complexity of brain structures. It provides the opportunity to quantitatively analyze cell populations, morphology and distribution to answer biological questions. For example, the number, distribution [63, 54, 131] and morphometric information [138] of neurons are important features in studying brain aging, including neurodegenerative diseases. A crucial prerequisite for such studies is cell instance segmentation, which plays a crucial role in digital pathology image analysis. Neuron segmentation is exceptionally challenging because the size, density, and intensity of neurons differ a lot from one anatomical region to another. Since manual identification of single cells is extremely laborious and time-consuming, several automatic segmentation algorithms have been proposed: thresholding [94], graph cut [81, 46] and watershed [22, 136, 137]. These methods need to be specifically adapted for different configurations (species, cell types, stainings). Furthermore, noise or other technical artifacts can easily influence the segmentation results. Under-segmentation and over-segmentation often occur when they deal with touching or overlapping cells like neurons. You *et al.* [160] proposed a framework based on gaussian, min-max filter and region-growing



algorithm to deal with such data, but it is computationally expensive due to numerous iterations and performed poorly on light-stained regions.

Recently, deep learning (DL) has achieved remarkable progress in many fields [73], especially in medical image analysis. Neural networks have been successfully applied to detect abnormal signals [167], segment lesion areas [145, 128, 43, 165] for clinical diagnosis. DL-based methods have also shown superiority in cell segmentation competitions [16, 68], achieving better segmentations with stronger robustness than traditional algorithms. Naylor *et al.* [91] addressed this problem as a regression task of estimating the nuclei distance map. A more common strategy is to address this problem as a semantic segmentation task, such as the pixel-wise binary classification of cells and background [119], or more recently, the ternary classification of the interior of cells, background and cell boundaries [16, 23, 19]. Most convolutional neural networks (CNNs) such as AlexNet [67], VGGNet [117] and ResNet [45] learn representations by gradually reducing the size of feature maps, but the high-resolution features are lost during this process. These networks are not suitable for pixel-wise tasks like cell segmentation. Several networks add a resolution-recover process to address segmentation problems. For example, SegNet [7] and DeconvNet [92] use unpooling and deconvolution layers to recover original resolution; UNet [105], a breakthrough in the field of medical image segmentation, with skip connections concatenate the high resolution features of the encoder path to the upsampled output of the decoder path. Our previous work [151] evaluated an ensemble model of eight UNet-like neural networks with different backbone CNNs. Wang *et al.* proposed HRNet [143], which maintains high-resolution representations and assembles features from multi-resolution streams. It outperformed other state-of-the-art networks on several tasks of semantic segmentation, object detection and instance segmentation.

The good performance of CNNs relies on large datasets and the quality of pixel-level annotations, which are tedious and labor-intensive to be carried out manually. To facilitate the labeling process, researchers investigated several weakly-supervised methods using point annotations: Qu *et al.* [98] trained a CNN to predict the cell center location, and generated pixel-level labels using Voronoi transformation and k-means clustering. Based on a similar strategy, Chamanzar *et al.* [18] proposed a multi-task learning integrating repel encoding to enhance segmentation performance. However, [98, 18] are not straightforward and involve multiple networks or branches to achieve the final segmentation.

Moreover, CNNs without post-processing often failed to handle touching objects [91]. Researchers mainly focus on improving the performance of CNNs, while the post-processing part is generally not explicitly described, although it is a critical step to obtain good segmentations. Applying watershed seg-

mentation (WS) on the cell probability map derived from DL is the most common way [68, 154]. Graph partition [121] and distance transform [155] are also popular techniques. The winning method of [16] proposed a more tricky technique: a regression model was trained to predict the intersection-over-union (IoU) for cell candidates produced by applying different thresholds on the probability map. With this method, only the candidate with the highest IoU was preserved for each object. One drawback of these methods is that well-configured parameters are required to ensure the performance. Thus, this requires redesign of the parameter settings for applications on novel data. Generic post-processing methods for cell instance segmentation are scarce and worth investigating.

In this paper, we propose an end-to-end framework based on CNNs for neuron instance segmentation, with the following contributions: 1) we establish and validate a semi-automated pixel-level-mask synthesis pipeline using only point annotations and Random Forest (RF) binary segmentations. This approach allows to generate a large labeled dataset with minimal manual cost, 2) inspired by the instance segmentation challenge [134], we integrate EfficientNet-B5 [127] into a UNet-like encoder-decoder architecture, this new model is spatially and semantically precise and 3) we propose a novel strategy for post-processing probability map based on ultimate erosion, dynamic reconstruction and WS. Our framework is generic and easy to configure: the mask synthesis pipeline is independent of the neural network, the synthetic masks can be easily derived to validate other supervised methods. The only parameter needed at the post-processing stage is the size of the structuring element, which equals the size of the smallest cell in the dataset. We compared our network to various state-of-the-art CNNs methods [160, 23, 151, 143], as well as the proposed post-processing to other classical approaches. The findings demonstrate the superiority of our framework in terms of accuracy and efficiency, allowing us to develop a powerful tool for evaluating neurons in preclinical research and neurodegenerative diseases.

## 4.2.2 . Material and methods

### 4.2.2.1 . Dataset

The data of this study are a set of two-dimensional (2D) light microscopy images. A representative histological section with a thickness of 40  $\mu m$  was produced from a healthy 9-year-old male macaque brain, stained by immunohistochemistry using the neuronal nuclei (NeuN) antibody, and scanned by an AxioScan.Z1 (Zeiss) with the in-plane resolution of 0.22  $\mu m$ /pixel ( $\times 20$  magnification). The animal study was reviewed and approved by the Comité d'éthique approved by the MESR belonging to the EU: CETEA DSV – Comité n° 44. Thirty images of 5000  $\times$  5000 pixels were selected to represent the heterogeneity of neuron distribution in the central anatomical regions, including

## 4.2. NEURON INSTANCE SEGMENTATION

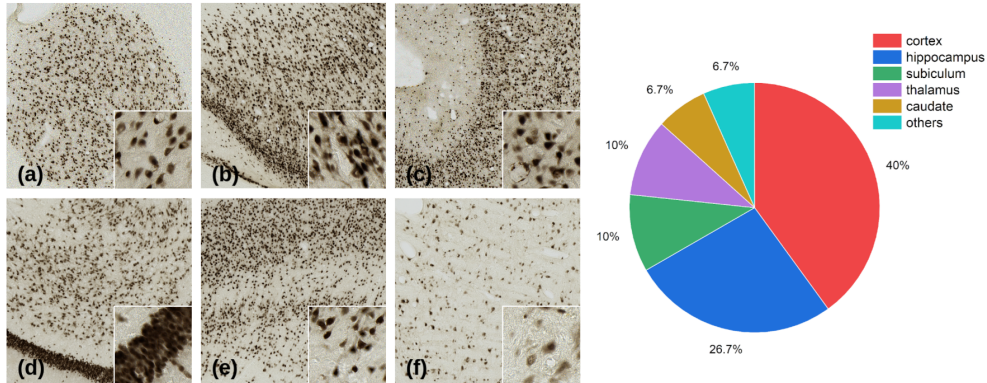


Figure 4.1: Dataset examples with magnifications. (a) caudate, (b,c) cortex, (d) hippocampus, (e) subiculum and (f) thalamus.

cortex, hippocampus, thalamus, subiculum, *etc* (see Supplementary Figure 4.9). These images showed a large diversity in terms of neuron size, shape, contrast and density, with both sparse (*e.g.*, caudate and thalamus) and highly aggregated regions (*e.g.*, cortex and hippocampus), as illustrated in Figure 4.1. The pie chart presents the dataset composition in terms of anatomical regions, with about 67% of images coming from the cortex and the hippocampus since the cortex is the largest brain structure (76%) [4] and hippocampal atrophy is linked to several neurodegenerative diseases [133].

To verify the representativeness of our test set, we extracted features of the dataset using a ResNet101 model [45] pre-trained on ImageNet and embedded the high-dimensional features into 2D space with t-distributed stochastic neighbor embedding (t-SNE) projection [133].

All images were divided into two subsets: 24 images for the training set and 6 images for the test set. The test set contains the following regions: caudate, cortex ( $\times 2$ ), hippocampus, subiculum and thalamus. Training images were cropped into 11k patches of  $224 \times 224$  pixels for which  $\frac{1}{4}$  of them were used to validate our neural networks at the end of each training epoch. To prevent overfitting and increase the robustness of the model, we applied data augmentation, including random rotation, vertical and horizontal flipping, RGB channel shuffling, color inversion, *etc*. The training set was expanded to 6 times the original size.

The size of images in the test set is  $5000 \times 5000$  pixels, which requires more memory than the GPU RAM. They were firstly cropped into smaller patches ( $1344 \times 1344$  pixels, can be adjusted according to the GPU RAM) with an overlap (120 pixels,  $\sim 10\%$ ) in both vertical and horizontal directions. The size of patches was constrained by the GPU memory (16 GB). A weighted map [23] was applied to the probability map of each patch to reduce the impact of

inaccurate prediction at the border area. The weighted probability maps were then seamlessly assembled to reconstruct the probability map of the original large-scale image.

In order to assess the generalizability of the proposed method, we added a test set that is independent of our training set. As shown in Supplementary Figure 4.10, this dataset contains four cortex images of  $1024 \times 1024$  pixels from various animal subjects, including two macaques, a microcebus [35], and a mouse. Manual point annotation is conducted to evaluate the object-level segmentation of the proposed method. This study focuses on the cortex, which is not only the largest brain structure but also the region of greatest interest to neuroscientists. Compared to the training set (see Figure 4.1), the additional macaque images are less brown and have greater contrast, the microcebus image is rosy brown, and the mouse image is gray with lighter stain intensity.

#### 4.2.2.2 . Pixel-level mask synthesis

Traditional nuclei segmentation methods based on point annotations involve center detection and pixel-level label extraction. In this study, we addressed neuron instance segmentation as a semantic segmentation task, which is to classify each pixel in the image into three following classes: neurons, background and contours of touching neurons. However, this strategy required instance annotations at the pixel-level, which would have been extremely labor-intensive to achieve. We designed a pipeline to synthesize pixel-level masks to alleviate the manual labeling effort. Our strategy is shown in Figure 4.2, which consists of two stages: the first stage is to segment neurons from the tissue and perform point annotation, and the second stage is to further separate each neuron instance under the guidance of point annotations for initialization and the constraint of the binary segmentation. Figure 4.2 (a) shows a NeuN image, where the neuron centroids were annotated manually by a disk with a radius of 5 pixels, as shown in Figure 4.2 (b). Figure 4.2 (c) presents the binary segmentation of neurons and the tissue, which is generated automatically with a RF model. As described in the previous work [160], the RF model contains 100 decision trees. It is trained using an optimized subset of automatically selected features on a small binary segmentation dataset of 100 images of  $512 \times 512$  pixels [11]. A CNN may be less effective with a small training dataset of this size. Figure 4.2 (d) shows the connected components of the binary segmentation superimposed with point annotations. Multiple centroids can be found inside one component since the binary segmentation could not identify individual neurons. Thus, the second phase involved a competitive region-growing algorithm separating touching neurons, using centroids as the seed to initialize the growing process, and the binary segmentation of RF to constrain the expansion. Figure 4.2 (e) shows the results of the region growing, where neuron instances are separated and assigned with a unique label.

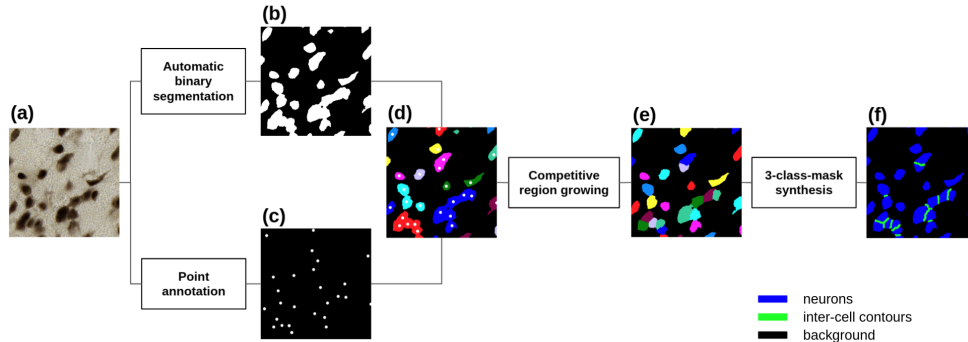


Figure 4.2: Pixel-level mask synthesis. (a) original image, (b) manual point annotations, (c) RF binary segmentation, (d) fusion of colored connected components and (c), (e) labeled image produced by region growing and (f) final three-class-masks, blue: neurons, green: inter-cell contours and black: tissue.

The generated annotations were used as ground truth since the process was guided by the point annotation. Morphological operations were applied to instance annotations to obtain masks of three classes: in addition to the neuron and tissue classes, the pixels that connect different labels were automatically identified as the third class of inter-cell contour to separate touching or overlapping neurons [151]. The inter-cell contour has a constant thickness of 4 pixels, which was defined empirically. Figure 4.2 (f) shows the final result of the pipeline, which are synthetic masks of three classes that were used to train the segmentation neural networks. Although manual efforts are still required, such as the point annotations and the binary masks to train the RF, they are far less time-consuming and labor-intensive than massive instance annotations required to train a CNN for instance segmentation.

It is worth noting that the objective was to generate pixel-level annotations to complete the dataset needed to train the segmentation neural networks. The use of this pipeline is finished once the annotations are synthesized, but the same strategy can be used to annotate other datasets.

To validate the synthetic mask generation process, three experts performed manual segmentation on a small dataset containing five patches of  $500 \times 500$  pixels, including caudate, cortex ( $\times 2$ ), hippocampus and subiculum. The average manual segmentation time for 5 images was 2.5 hours.

### 4.2.2.3 . Neural networks

In our previous work [151], we showed the efficiency of UNet-like architecture for neuron instance segmentation. Hence, we decided to keep the same strategy with a more recent neural network as the backbone. The family of models called EfficientNets was proposed by Tan *et al.* [127], which showed superiority in accuracy and efficiency against previous CNNs. The baseline model, EfficientNet-B0, was generated with neural architecture search [127]. Its main building block is mobile inverted bottleneck MBConv [109]. Scaling up one network dimension of width, depth and input image resolution can improve accuracy. In particular, compound scaling of three dimensions can provide a significant gain [127]. This approach brought seven scaled-up versions, named EfficientNet-B1 to EfficientNet-B7. EfficientNet-B5 was chosen in this work as the result of a trade-off between accuracy and training cost. Figure 4.3 (a) presents the architecture of EfficientNet-B5, consisting of stem layers, seven main building blocks of MBConv and final layers. The resolution of the feature map was reduced five times gradually (from  $224 \times 224$  to  $7 \times 7$  pixels) after the stem layers, block 2, block 3, block 4 and block 6, respectively. Based on EfficientNet-B5 as the encoder, we gradually recovered the original high resolution through the decoder path, which consists of deconvolution and convolution layers. Figure 4.3 (b) shows the structure of our network, named EfficientUNet-B5, skip connections concatenate encoder and decoder at five different resolutions.

### 4.2.2.4 . Loss function

Since the neuron instance segmentation was addressed as a semantic segmentation of three classes, we used the compound loss of categorical cross-entropy (CE), two soft dice losses ( $D_{neuron}$  and  $D_{contour}$ ) [151, 168] for neuron class and inter-cell contour class respectively to train the network. The global loss function  $L$  is defined as:

$$L = 0.5CE + 0.3D_{neuron} + 0.2D_{contour} \quad (4.1)$$

$$CE = -\frac{1}{nc} \sum_{i,j} \sum_k^c t_{i,j,k} \log(p(i, j, k)) \quad (4.2)$$

$$D_k = 1 - \frac{2 \sum_{i,j} t_{i,j,k} \log(p(i, j, k)) + 1}{\sum_{i,j} t_{i,j,k} + \sum_{i,j} p(i, j, k) + 1} \quad (4.3)$$

where  $k$  denotes one class among  $c$  classes ( $c = 3$ ),  $t_{i,j,k}$  is equal to 1 if the pixel  $(i, j)$  belongs to class  $k$ ,  $p(i, j, k)$  denotes the probability of pixel  $(i, j)$  of being class  $k$ ,  $n$  is the number of pixels in the patch. Cross-entropy is the most popular loss function for classification tasks. Soft dice loss [88]

## 4.2. NEURON INSTANCE SEGMENTATION

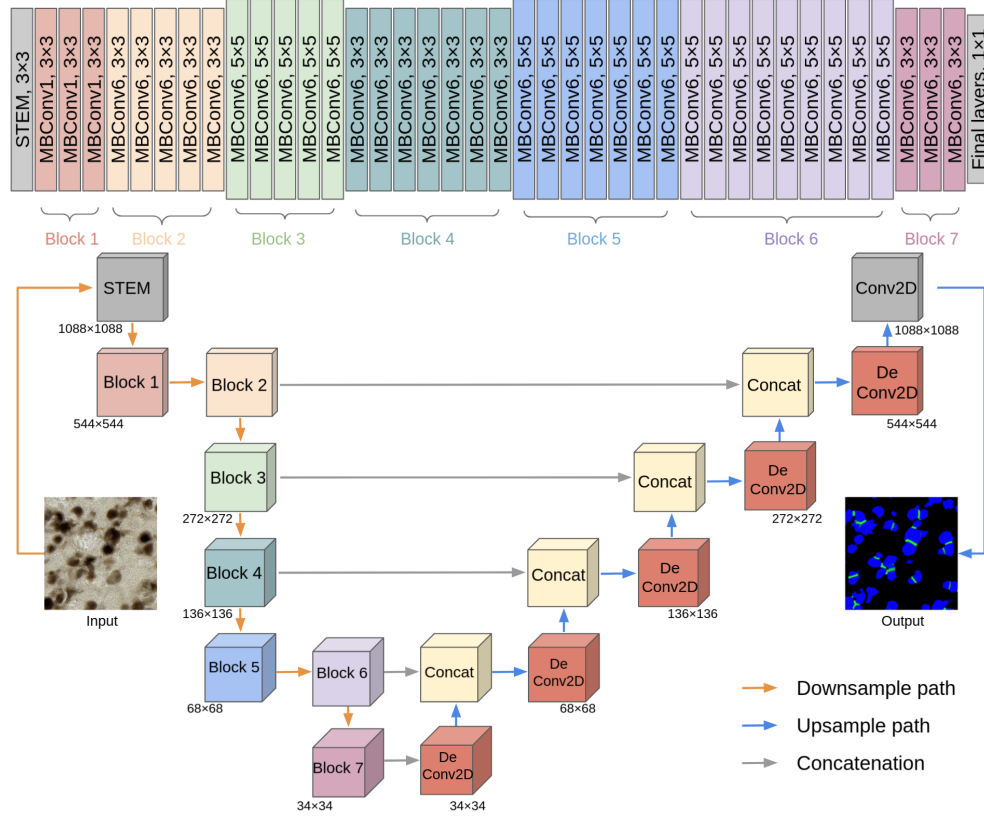


Figure 4.3: Top: structure of EfficientNet-B5, it consists of 7 building MB-Conv blocks (represented with different color) and bottom: structure of our neural network using EfficientNet-B5 as encoder, named EfficientUNet-B5, the encoder is concatenated with the decoder at five different resolution (Block 2, Block 3, Block 4, Block 6 and Block 7).

was adapted from the Dice coefficient to calculate the similarity between two images. One was added in 4.3 to ensure that the function is not undefined when  $t_{i,j,k} = p(i, j, k) = 0$  [58]. Here, we associated  $CE$  with  $D$  for neurons and inter-cell contours to force the network to distinguish the two classes from the tissue. The weighting factors (0.5, 0.3 and 0.2) indicate the contribution of each item to the compound loss, were empirically defined constants and used to deal with class imbalance:  $CE$  and the sum of  $D$  (neuron and contour) had the same weight, while the weight of  $D_{neuron}$  was slightly higher than that of  $D_{contour}$  because the neuron class was the most important in our case.

#### 4.2.2.5 . Post-processing

The most straightforward post-processing is to apply a threshold to the probability map, yet one threshold will not fit all tested neurons with varying sizes, shapes and intensities. Better segmentation can be achieved using more sophisticated methods such as graph partitioning or distance transform. However, these methods increase the computation cost considerably and introduce several hyperparameters that need to be defined empirically. In this study, we propose an efficient and generic post-processing approach, as presented in Algorithm 1. The output of our network is a 3-channel probability map, presented with an RGB image. Each channel corresponds to one class: channel R represents the background, channel G represents the contours between touching neurons and channel B represents neurons. First, we extracted pixels that were most likely to be neuron class ( $argmax(P) = neuron$ ), denoting these pixels to 1 and other pixels to 0 to create a binary image. Second, we applied the ultimate erosion on the binary image with a disk-shaped structuring element, whose radius was equal to 10 pixels, the same as the radius of the smallest neuron expected [93], ensuring that no more than one neuron would be erased during erosion to prevent under-segmentation. We hypothesized that the inter-cell class could separate entirely or partially the touching neurons. The second case often occurs in dense regions where the inter-cell class is not sufficient to cut touching neurons entirely. However, it could create an initial concavity between cells, which would provide an optimal condition for performing ultimate erosion to complete the separation process. We labeled the ultimate residues, each residue representing an individualized neuron. Then, we proposed a dynamic dilation reconstruction using the same structuring element: each residue was dilated with the same number of erosion applied to produce an approximation of their original size. Due to the disk-shaped structuring element, the dilated residues might have an unnatural smooth shape. We used WS to restore the refined morphologic information: the dilated residues were used as seeds to initialize the WS expansion, constrained by a binary mask which merged neuron and inter-cell contour classes (1 if one pixel belonged to neuron or inter-cell contour, 0 otherwise). The merging of inter-cell contour and neuron channels aimed to restore the cell pixels lost due



to our artificially created inter-cell class.

The only parameter of our post-processing to be set is the size of the structuring element of ultimate erosion, which is equal to that of the smallest neuron in the dataset. It is easy to configure and generic, it can be applied independently to other DL-based nuclei segmentation methods using a similar strategy.

---

**Algorithm 1:** Post-processing using mathematical morphology

---

**Input:** A three-channel probability map

**Output:** Neuron instance segmentation

- 1 Create binary mask based on the neuron channel
  - 2 Ultimate erosion with a disk  $S(r = 10pixels)$
  - 3 **for** *each ultimate residue*  $U_i$  **do**
  - 4      $N_i \leftarrow$  number of erosion before  $U_i$  being removed
  - 5     dilation using  $S$ ,  $U_i \leftarrow dil(U_i)$ ,  $N_i \leftarrow N_i - 1$
  - 6     repeat dilation until  $N_i = 0$
  - 7  $M \leftarrow$  fusion of neuron channel and inter-cell channel
  - 8 Apply WS, markers  $\leftarrow U$ , mask  $\leftarrow M$
  - 9 **return** segmented neurons
-

#### 4.2.2.6 . Evaluation metrics

We aim to establish a framework which consists of the neural network and post-processing. We performed comprehensive comparisons for neural networks and post-processing approaches respectively. Four tasks, including detection, instance segmentation, semantic segmentation, and counting, were evaluated using five metrics. They are the F1 score (F1-det) for detection, the relative count error (RCE) for counting, the Dice score for semantic segmentation, and the F1 score (F1-seg) and Aggregated Jaccard Index (AJI) for instance segmentation.

The tasks of detection and counting were evaluated at the object-level by matching the predicted neurons with the point annotation. A predicted neuron was considered as a true positive (TP) when it was superimposed with precisely one point annotation. Otherwise, it was defined as a false positive (FP: not superimposed with any centroid) or a false negative (FN: superimposed with more than one centroid). FN also included the case that no neuron was detected in the location of a centroid. With TP, FP and FN, we computed precision (P), recall (R) and F1-det, and RCE as follows:

$$P = \frac{TP}{TP + FP}; R = \frac{TP}{TP + FN}; F1 = 2 \times \frac{P \times R}{P + R} \quad (4.4)$$

$$RCE = \frac{|FP - FN|}{TP + FN} \quad (4.5)$$

RCE is the ratio of count error over the number of neurons identified with centroids.

Furthermore, we estimated how well-segmented neurons matched synthetic masks at the pixel-level. The IoU score  $\frac{|A \cap B|}{|A \cup B|}$  was computed for all pairs of objects, where  $A$  is a predicted neuron and  $B$  is the corresponding ground truth mask. When segmentation of a neuron covered the mask completely, the IoU score was 1. Since it is almost impossible to perform two identical segmentations, even for an expert, we selected 0.5 as the threshold of minimum IoU to identify correct segmentation. In this case: the TP was defined as the IoU greater than 0.5 between the predicted neuron and the synthetic mask. Otherwise, it was a FP or a FN. We computed the P, R and F1 score (F1-seg) to evaluate the segmentation performance with this new criterion. Dice coefficient and Aggregated Jaccard Index (AJI) [69] were also calculated to evaluate the segmentation at the pixel-level. The AJI is defined as:

$$AJI = \frac{\sum_{i=1}^N |G_i \cap S(G_i)|}{\sum_{i=1}^N |G_i \cup S(G_i)| + \sum_{u=1}^U |S_u|} \quad (4.6)$$

Where  $G_i$  is one ground truth object,  $S(G_i)$  is the segmented object that maximizes the IoU with  $G_i$ .  $U$  is the set of segmented objects that have not

been assigned to any ground truth object. AJI is the most stringent among all evaluation metrics. It aims to penalize errors at both object and pixel-level. It would also help us to distinguish methods that score similarly on other metrics.

#### 4.2.2.7 . Implementation details

The network was implemented using Tensorflow and Keras. The encoder was pre-trained on ImageNet. The learning rate started from  $1e-4$  and decreased gradually during the training. The model was trained for 100 epochs during 40h, with Adam optimizer. We monitored the training and validation loss of each epoch and saved the model with the lowest validation loss (at 44th epoch). Test Time Augmentations of flipping ( $\times 2$ ) and rotation ( $\times 4$ ) were applied. This work was conducted on a workstation equipped with bi-processors (operating system: Ubuntu 16.04 LTS 64-bits, CPU: Intel Xeon gold 5218 at 2.3 GHz, RAM: 128 GB, GPU: NVIDIA Quadro RTX 5000 with 16 GB memory).

#### 4.2.2.8 . Compared methods and parameter settings

We performed comparisons of both neural networks and post-processing approaches. For neural network benchmarks, we compared the proposed EfficientUNet with an unsupervised method proposed in [160] and the following state-of-the-art networks, most were UNet-like: UNet [23, 105], HRNet [143], an ensemble model [151] of eight encoder-decoder networks and each constitutive network. The encoders included three ResNets (34, 101, 152), two Dual Path Networks DPN-92 (with sigmoid and softmax activation, respectively), two DenseNets (121, 169) and one Inception-ResNet. It is worth noting that we named the eight encoder-decoder networks after their encoder for simplicity.

The method [160] was unsupervised and did not require any additional training. All DL methods were trained on the same dataset with synthetic annotations. UNet was trained from scratch with the same configuration described in [23]. HRNet was trained from scratch, the training details were the same as EfficientUNet, see 2.7. The eight constituents of the ensemble model were trained separately with encoders pre-trained on ImageNet using the configuration described in [16]. In this part, we applied the same post-processing [16] to eliminate the effect of different post-processing for DL methods. It introduced a second-stage training procedure, training a regression model with neuron morphological information to predict the IoU. We created three groups of cell candidates using three thresholds (0.6, 0.7, and 0.8), and only the candidate with the highest predicted IoU was retrained for each object. The segmentation would be removed if the predicted IoU was too low, in order to reduce the number of FPs.

Another contribution of this work is the post-processing scheme. Ablation experiments were carried out to demonstrate the benefits of the proposed post-processing. We compared the proposed post-processing scheme to four classical methods: the first, which is the most often used method, includes applying WS to the probability map's thresholded ( $\geq 0.5$ ) cell channel, which served as a baseline for this section. The second is the winning method of [16], which was the same approach that we used for the comparison of networks. Additionally, we assessed the post-processing scheme of [163], which refines the neuron class by removing ambiguous pixels that are likely to be the contour class, followed by a dilation process to recover the neuron form. The last is a widely-used distance-transformation-based technique [91, 154]. Distance transformation is applied to the neuron class to obtain the distance map. With a minimum allowed distance being 20 pixels (minimum diameter of neurons), the local maximums of the distance map are further used as the seeds to initialize WS. Only the ultimate erosion and the dynamic dilation in Algorithm 1 are replaced by the distance transformation to fairly compare the separation capacity of the two techniques. All approaches were applied on the same probability maps, the prediction results of EfficientUNet.

### 4.2.3 . Results

In this section, we first show the representativeness of the test set versus the training set, see 4.2.3.1. In 4.2.3.2, we quantitatively evaluate the quality of the synthetic mask by comparing it with three manual segmentations. In 4.2.3.3, we present a comprehensive comparison of the proposed EfficientUNet and several state-of-the-art methods. In 4.2.3.4, we illustrate the advantage of our post-processing strategy in comparison with other classical techniques.

#### 4.2.3.1 . Qualitative evaluation of datasets composition

As shown in Figure 4.4, we plotted the heatmaps (number of bins = 25) based on the t-SNE projection of the whole dataset, the training set and the test set, respectively. X and Y were normalized to [0,1], representing the two dimensions of the embedded space. Since test data have similar feature distribution (t-SNE plots of two distributions present overlap) compared to that of training data, we can consider the test set as representative of the whole dataset.

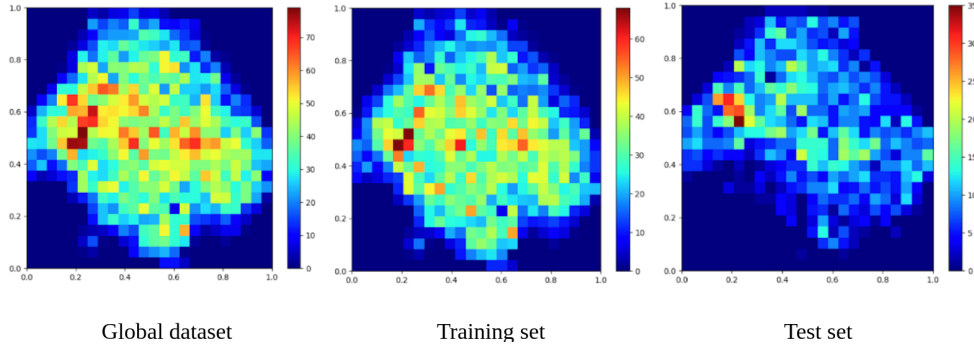


Figure 4.4: Heatmaps of t-SNE projection. Left to right: all data, training data and test data.

#### 4.2.3.2 . Quantitative validation of synthetic masks

The objective of our mask synthesis pipeline was to minimize manual effort, the expert labeled only the centroid of neurons instead of identifying the contour of neurons. We considered the point annotations unbiased, and evaluated the binary segmentation as well as the region-growing process. Table 4.1 reports the average IoU score (mIoU) and Dice coefficient between the synthetic masks and three manual annotations (denoted as e1, e2 and e3) respectively. We also computed the scores between manual annotations to show the inter-expert variability. We observed that the evaluation scores varied when we compared the synthetic masks to different manual annotations. Overall, our synthetic masks were of good quality, in particular, with the e1 group as the reference. Figure 4.5 shows inconsistent manual segmentations between experts in three of five images. On average, the scores of synthetic annotations were at the same level of the inter-experts scores, with a tiny difference on Dice (-0.6%) and a slight decrease (-2.4%) of mIoU. On the other hand, the experts spent 2.5 hours on annotating five small images ( $500 \times 500$  pixels). The entire training dataset (11k images of  $224 \times 224$  pixels) would take over two months to annotate manually, demonstrating the need of the semi-automatic mask synthesis pipeline.

#### 4.2.3.3 . Comparison of segmentation methods

Table 4.2 reports the performance of the tasks of detection (F1-det), counting (RCE), instance segmentation (F1-seg and AJI) and semantic segmentation (Dice), as well as the computational complexity (#Params: number of trainable parameters and number of FLOPs: Floating-Point Operations) of the proposed EfficientUNet and other 12 methods as mentioned in Section 4.2.2.8. We did not compute the Dice for the unsupervised method [160], because it used the same binary RF segmentation that constrained our synthetic masks. All DL methods except UNet had a very high F1-det ( $> 0.92$ ) with low vari-

Table 4.1: Average IoU score (mIoU) and Dice coefficient between the synthetic masks and three manual annotations (denoted as e1, e2 and e3) respectively.

Metrics	Synthetic vs experts				Inter-experts			
	ours vs e1	ours vs e2	ours vs e3	mean	e1 vs e2	e1 vs e3	e2 vs e3	mean
mIoU	0.774	0.706	0.724	0.735	0.753	0.769	0.755	0.759
Dice	0.914	0.888	0.888	0.897	0.898	0.91	0.9	0.903

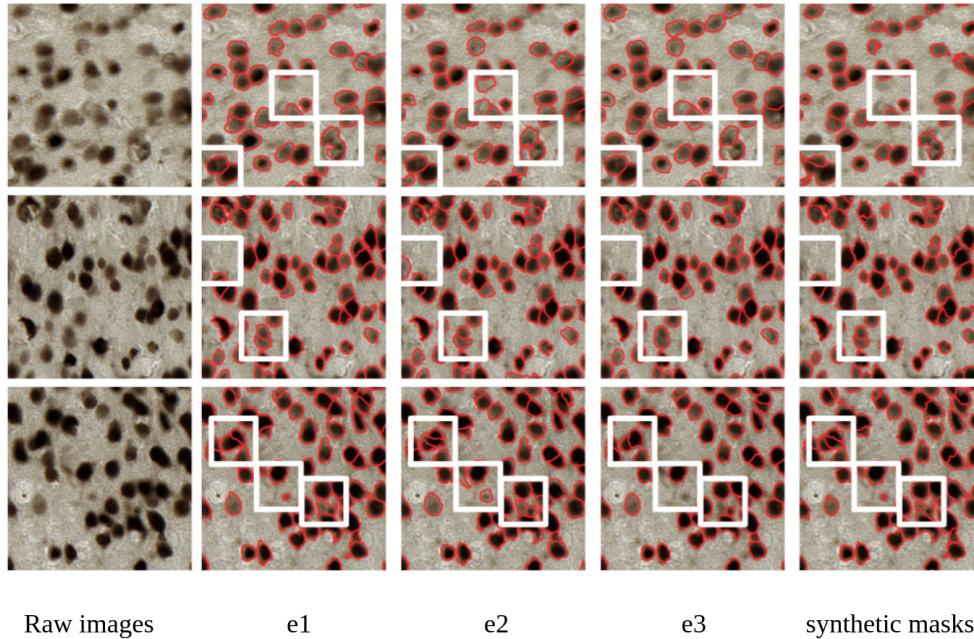


Figure 4.5: Comparison of manual segmentations and synthetic masks. The white squares show the inconsistencies between manual segmentations.

ance. The ensemble model outperformed the others on the tasks of detection (F1-det: 0.931) and counting (RCE: 0.026). In contrast, the best model for the tasks of semantic and instance segmentation was one constitutive model, DPN-soft, with the highest scores of Dice (0.951), F1-seg (0.907) and AJI (0.759). The proposed EfficientUNet was the second best model for segmentation tasks and it also performed well on the tasks of detection and counting, with similar scores to the best model for each metric (F1-det: -0.007, RCE: +0.018, Dice: -0.017, F1-seg: -0.006, AJI: -0.017). Followed by two constitutive ResNet backbone networks: ResNet-34 had good performance in the tasks of detection and semantic segmentation, while the performance declined on the task of counting. ResNet-101 was the best single model at the object-level (F1-det and RCE), but it performed poorly on segmentation tasks. The scores of HRNet were comparable to those of EfficientUNet, but the scores on segmentation tasks were slightly lower (Dice, F1-seg and AJI). The scores of UNet were the lowest in all the tasks, at both object and pixel levels. The unsupervised method [160] achieved good results on segmentation tasks: 0.87 in F1-seg and 0.729 in AJI, while it performed worse on the tasks of detection and counting. Further considering the computational complexity of the DL models, the proposed EfficientUNet was the best among all single models. The FLOPs of the proposed EfficientUNet (15.13 GB) were similar to that of the lightest model (UNet, FLOPs: 15.29 GB), regardless its model size was six times that of UNet. By contrast, although the performance of DPN-soft

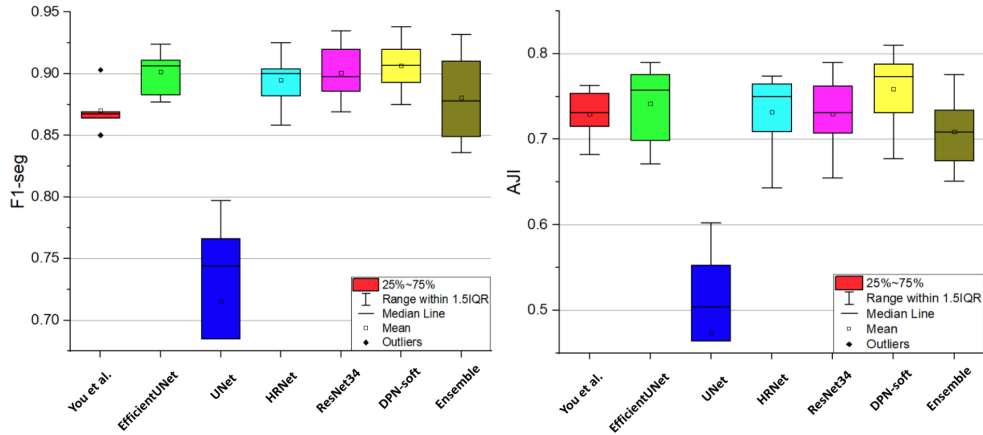


Figure 4.6: Boxplot of F1-seg and AJI for different methods for comparative analysis. Left panel: F1 scores of segmentation. Right panel: AJI scores (outlier values of UNet are below the vertical scale range).

was slightly better than that of EfficientUNet, the complex structure (Params: 133.5 MB, FLOPs: 51.64 GB) makes it complicated to be trained and applied efficiently to larger-scale images.

As shown in Table 4.2, most methods reach a high F1-det and Dice, which is not helpful for performance comparison. On the other hand, F1-seg and AJI provide distinct differences between methods, revealing performance at both object and pixel levels. Figure 4.6 compares F1-seg and AJI of the unsupervised method You *et al.* [160], the proposed EfficientUNet, UNet, HRNet, the two best constitutive models (ResNet-34 and DPN-soft) and the ensemble model. What stands out in the figure is that UNet performed poorly on the task of instance segmentation. In terms of F1-seg, both the proposed EfficientUNet and DPN-soft provided good and robust performance. In particular, the difference in EfficientUNet scores across regions was less than 4%, which indicates that it performed well on all anatomical regions tested. Although the median score of HRNet was similar to that of ResNet-34, the interquartile range was slightly lower than that of ResNet-34. The mean score of the unsupervised method You *et al.* [160] was at the same level as that of the ensemble model, but it was more robust, with the slightest variation between regions. AJI illustrates differences between methods in a more significant way, the scores of all the methods were decreased. DPN-soft and the proposed EfficientUNet remained the two best models. As well as HRNet and ResNet-34, their scores were similar but slightly lower than the best models. Surprisingly, You *et al.* [160] was the second-last method of F1-seg, showing better performance than the ensemble model with AJI: it reached the same level as HRNet and ResNet-34.



Table 4.2: Comparison of neuron detection, counting, instance segmentation and computational complexity of different automatic methods. Best and second best results are in bold with the best also underlined. FLOPs are computed for a  $224 \times 224$  input RGB image.

Model	(Mean $\pm$ Standard deviation)							
	#Params(M)	FLOPs(G)	F1-det	RCE	Dice	F1-seg	AJI	
<i>You et al.</i>	-	-	0.889 $\pm$ 0.019	0.076 $\pm$ 0.062	-	0.87 $\pm$ 0.017	0.729 $\pm$ 0.03	
EfficientUNet	39.25	15.29	0.924 $\pm$ 0.012	0.044 $\pm$ 0.033	<b>0.934<math>\pm</math>0.024</b>	<b>0.901<math>\pm</math>0.018</b>	<b>0.742<math>\pm</math>0.048</b>	
UNet	5.75	15.13	0.819 $\pm$ 0.061	0.264 $\pm$ 0.099	0.890 $\pm$ 0.017	0.715 $\pm$ 0.087	0.473 $\pm$ 0.137	
HRNet	9.5	26.55	0.924 $\pm$ 0.011	0.044 $\pm$ 0.041	0.933 $\pm$ 0.019	0.895 $\pm$ 0.023	0.732 $\pm$ 0.049	
DenseNet-121	17	30.26	0.920 $\pm$ 0.018	0.041 $\pm$ 0.033	0.830 $\pm$ 0.047	0.801 $\pm$ 0.09	0.615 $\pm$ 0.053	
DenseNet-169	18.5	44.26	0.924 $\pm$ 0.012	0.043 $\pm$ 0.041	0.883 $\pm$ 0.032	0.863 $\pm$ 0.034	0.682 $\pm$ 0.064	
ResNet-34	23.75	7.81	<b>0.929<math>\pm</math>0.014</b>	0.067 $\pm$ 0.034	<b>0.934<math>\pm</math>0.015</b>	<b>0.901<math>\pm</math>0.024</b>	0.729 $\pm$ 0.047	
ResNet-101	47.25	52.08	<b>0.929<math>\pm</math>0.013</b>	<b>0.036<math>\pm</math>0.034</b>	0.911 $\pm$ 0.023	0.875 $\pm$ 0.034	0.690 $\pm$ 0.049	
ResNet-152	62.25	58.99	0.926 $\pm$ 0.013	0.039 $\pm$ 0.039	0.918 $\pm$ 0.018	0.878 $\pm$ 0.026	0.704 $\pm$ 0.038	
Incep-ResNet	63.75	40.98	0.925 $\pm$ 0.014	0.081 $\pm$ 0.043	0.923 $\pm$ 0.026	0.894 $\pm$ 0.022	0.729 $\pm$ 0.053	
DPN-sig	133.5	51.64	0.921 $\pm$ 0.016	0.053 $\pm$ 0.03	0.910 $\pm$ 0.029	0.864 $\pm$ 0.05	0.679 $\pm$ 0.064	
DPN-soft	133.5	51.64	0.928 $\pm$ 0.011	0.043 $\pm$ 0.038	<b>0.951<math>\pm</math>0.011</b>	<b>0.907<math>\pm</math>0.022</b>	<b>0.759<math>\pm</math>0.048</b>	
Ensemble	-	-	<b>0.931<math>\pm</math>0.011</b>	<b>0.026<math>\pm</math>0.023</b>	0.918 $\pm$ 0.026	0.88 $\pm$ 0.036	0.709 $\pm$ 0.046	

Figure 4.7 shows comparative results of representative patches from different anatomical regions. Besides the methods mentioned above (in Table 4.2 and Figure 4.6), we also report the segmentation results of the complete proposed pipeline (network and post-processing) to visualize the improvement brought by the proposed post-processing. If the IoU between a segmented object and the ground truth is superior to 0.5, we highlighted the contours in green. Otherwise, the contours are displayed either in blue for over-segmentation or red for under-segmentation and missing detection respectively. You *et al.* [160] (column c) segmented most neurons correctly in sparse regions, while under-segmentation often occurred when the neurons aggregate. UNet (column e) suffered from under- and missing segmentation, but it was also ineffective in preserving the shape of neurons. The segmentations of EfficientUNet (column d), HRNet (column f), ResNet-34 (column g) were roughly comparable, performing well in all anatomical regions, from the thalamus (sparse) to the hippocampus (dense). Especially EfficientUNet caused fewer segmentation errors in caudate and cortex. DPN-soft (column h) and the ensemble model (column i) caused over-segmentations in thalamus and caudate, and under-segmentations in cortex and hippocampus. EfficientUNet with the proposed post-processing (column j) made fewer segmentation errors than the others in any anatomical region, suggesting that the proposed post-processing can successfully correct segmentation errors of the neural network (see columns d and j).

#### 4.2.3.4 . Comparison of post-processing schemes

Table 4.3 and Table 4.4 compare the performance of the instance segmentation task (F1-seg and AJI) and the processing time of different post-processing techniques respectively. Although the probability maps were the same (EfficientUNet), the results vary considerably between the different post-processing methods. The proposed method appeared to be the best for all anatomical regions, with an average F1-seg of 0.917 and AJI of 0.774. The distance transformation achieved comparable scores with slight decreases, and it was less computationally expensive than the proposed method. However, it over-segmented circle-shaped neurons (see Supplementary Figure 4.13). Compared to the baseline, the regression model of Topcoders [16] brought a slight improvement (2% in both F1-seg and AJI), whereas it took 460.6 s to process one image, which is 27 times that of the baseline. Zeng *et al.* [163] was the fastest method, and it required only 7.88s to process one image. However, it was also the method that performed the worst, with a decline of 9.2% and 14% in F1-seg and AJI compared to the proposed method. Among all tested anatomical regions, higher scores were found in the region with high contrast (subiculum) and in the region where only a few neurons were aggregated (caudate). In contrast, the scores dropped in the region with high neuron density (hippocampus) and the region with low contrast with the background (thala-

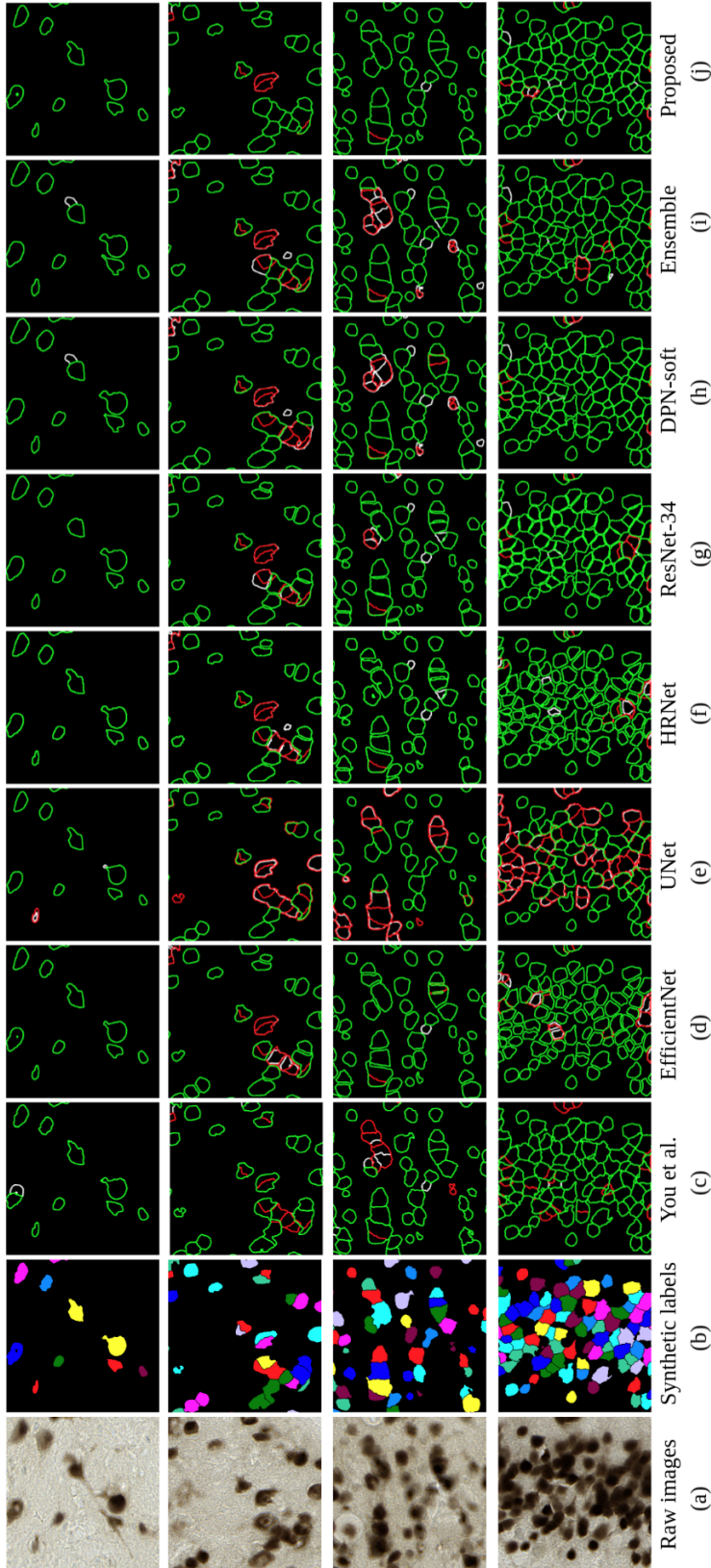


Figure 4.7: Original images and segmentation results of our method and other approaches in different regions with different neuron density. (a) raw images, (b) synthetic labels, (c-j) results of You *et al.* [160], EfficientUNet, UNet, HRNet, ResNet-34, DPN-soft, the ensemble model and the entire proposed framework. Top to Down: (1) thalamus, (2) caudate, (3) cortex and (4) hippocampus. The green contours represent segmentations that overlap the synthetic ground truth ( $\text{IoU} > 0.5$ ), the blue contours represent over-segmentations and the red contours represent missing and under-segmentations

mus). By comparing the proposed method and the baseline, ours significantly improved the instance segmentation, with an average gain of 4% and 4.6% in F1-seg and AJI, respectively. In particular, for dense regions like the cortex and the hippocampus, where lots of neurons aggregate, the proposed method increased AJI by 7.7% and 6.3%, respectively.

Figure 4.8 presents the results for each intermediate step of our post-processing scheme. The inter-cell class allowed our neural network to segment the sparse region like caudate and most neurons in the cortex. However, our network sometimes failed to separate touching neurons, as illustrated by the red square in Figure 4.8 (b, c). This problem became troublesome for the hippocampus region, where many neurons aggregated, which would cause the under-estimation in neuron population counting. In this particular case, the ultimate erosion process in our scheme could bring a critical advantage. Although the inter-cell class did not completely separate the touching neurons, it provided an optimal condition to apply further erosion: an initial concavity. Figure 4.8 (d) shows the ultimate residues computed. We observed that the neurons that the neural network had not separated were now fully individualized. By applying the same iteration number of dilation with the same structuring element as erosion on each ultimate residue, we restored the coarse morphological information of neurons, as illustrated in Figure 4.8 (e). Finally, the refined segmentation was obtained with the WS using dilated residues as seeds. A combination of inter-cell and neuron classes was used to constrain the expansion, allowing to eliminate the artificial gap between touching cells by reassigning the inter-cell pixels to neurons.

#### 4.2.3.5 . Assessment of the generalizability of the proposed method

Table 4.5 reports the object-level segmentation of the proposed method on the supplementary test set, including cortex images of two macaques, one microcebus and one mouse. The purpose of this experiment was to evaluate the generalizability of the proposed method on data different from the training set. The results on cortex images of the initial test set are illustrated as a reference. For the task of detection, the scores decreased on other animal subjects, but no significant differences were observed, with the largest decrease being 2.23% that was found in the mouse. As for the task of counting, the method made more errors on other animal subjects except for the microcebus, which had the lowest RCE of 0.03%. The RCE of the macaques and the mouse was approximately on the same level, with an increase of 5% compared to the reference.

Table 4.3: Instance segmentation performance using different post-processing techniques. The best results are in bold.

Region	F1-seg					AJI				
	baseline	Topcoders	Zeng <i>et al.</i>	Dist	Ours	baseline	Topcoders	Zeng <i>et al.</i>	Dist	Ours
caudate	0.885	0.903	0.82	0.906	<b>0.916</b>	0.754	0.775	0.641	0.787	<b>0.79</b>
cortex	0.87	0.896	0.822	0.909	<b>0.915</b>	0.689	0.72	0.598	0.759	<b>0.766</b>
hippocamp	0.857	0.877	0.828	0.899	<b>0.902</b>	0.661	0.671	0.614	0.722	<b>0.724</b>
subiculum	0.912	0.924	0.877	0.939	<b>0.94</b>	0.783	0.79	0.69	0.819	<b>0.821</b>
thalamus	0.876	0.911	0.781	0.896	<b>0.916</b>	0.76	0.776	0.662	0.758	<b>0.777</b>
overall	0.878	0.901	0.825	0.91	<b>0.917</b>	0.723	0.742	0.634	0.767	<b>0.774</b>

Table 4.4: Processing time for one image of  $5000 \times 5000$  using different post-processing approaches.

Method	baseline	Topcoders	Zeng <i>et al.</i>	Dist	Ours
Time(s)	17.17	460.6	7.88	28.3	118.57

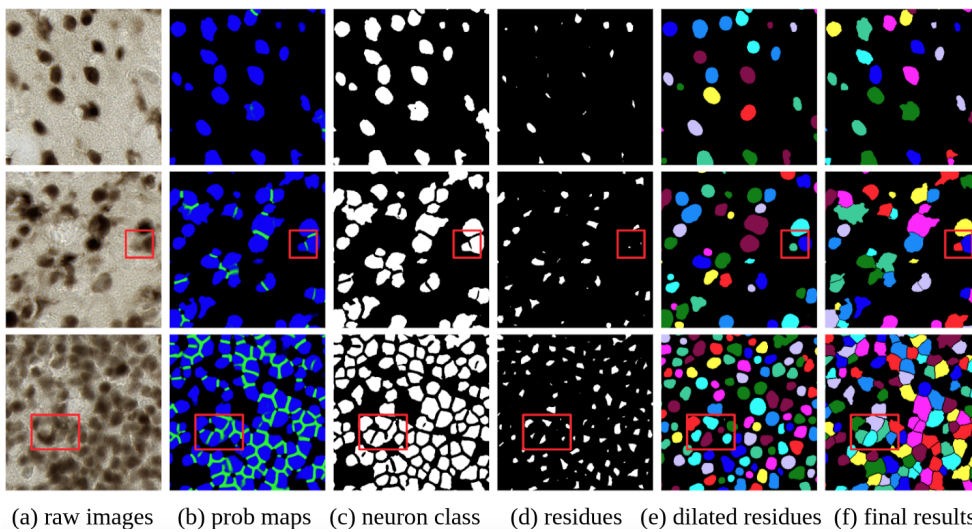


Figure 4.8: Intermediate results of the proposed post-processing. Top to down: (1) caudate, (2) cortex and (3) hippocampus. Left to right: (a) original images, (b) ground truth, (c) probability map of deep network, (d) binary mask of neuron channel, (e) ultimate residues, (f) reconstructed residues and (g) final segmentation after WS. The red square highlights the neurons that the neural network failed to separate but being fully segmented through the proposed post-processing.

Table 4.5: Object-level segmentation performance of the proposed method on the supplementary test set.

data supp	ref macaque	macaque 1	macaque 2	microcebus	mouse	mean
F1-det	0.93	0.918	0.927	0.921	0.908	0.918
RCE	0.02	0.075	0.073	0.003	0.071	0.056

#### 4.2.4 . Discussion

Unbiased quantification of individualized cells is essential for many biomedical analyses. One challenging application field is counting and individualizing the neuronal cells as their size, shape and density vary from one anatomical region to another. Recent studies demonstrated the importance of neuron morphology and distribution in studying cerebral functions and neurodegenerative diseases [63, 54, 131, 138]. Therefore, an automatic neuron segmentation method is a cornerstone for such research. In this work, we present an end-to-end framework aiming to improve neuron detection and instance segmentation in the major anatomical regions of the macaque brain. Our mask synthesis pipeline based on pin-pointed centroids and RF binary segmentation allowed us to generate large amounts of pixel-level annotations for training, which would have been impossible to achieve with manual cell segmentation. In addition, embedding the state-of-the-art CNN (EfficientNet-B5) into a UNet-like architecture increased segmentation accuracy. Although mathematical morphology techniques such as ultimate erosion for segmenting connected components [139] were proposed decades ago, it requires a strong concavity prerequisite between the connected components to obtain good results. This condition was satisfied by adding the inter-cell class to the probability map. Furthermore, we proposed dynamic reconstruction as a complementary step with the ultimate erosion to further improve the morphological reconstruction.

Table 4.1 and Figure 4.5 showed that pixel-level annotations produced by our pipeline were comparable to those of experts, which confirmed the good quality of the synthetic annotation. After an exhaustive comparison, EfficientUNet was chosen as it was the model with the best trade-off between accuracy and computational cost. The ensemble model was the best model for the tasks of detection and counting, but it was difficult to train and maintain because it consisted of eight independent networks. As for the tasks of instance and semantic segmentation, the best model was DPN-soft. It was the heaviest individual model, with the model size and FLOPs 3.4 times that of EfficientUNet, resulting in a considerably higher inference latency. The proposed EfficientUNet was not computationally expensive but outperformed most networks with a slight decrease over the best models. An efficient network is indispensable to extend this work to a larger scale image (entire brain section, even brains). Under this constraint, EfficientUNet seems to be the best choice among all tested networks. ResNet-34 and HRNet also showed a good trade-off between accuracy and efficiency, they could be potential candidates for developing an efficient segmentation network with a slight loss of accuracy. However, the models that reported good performance (HRNet, ResNet-34, EfficientNet and DPN-soft) caused a gap between the touching neurons, especially in the dense regions such as the hippocampus, as shown in Figure 4.7 d, f-h. One explanation is that all three classes were generally well classified

with these methods, including the inter-cell class. However, the applied post-processing scheme proposed in [16] used only the probability map of neuron class to constrain the WS expansion. Our post-processing results (last column of Figure 4.7) demonstrated that this contradiction could be eliminated by merging the neuron and the inter-cell classes as the WS mask. The results of EfficientUNet and the constitutive models suggested the superiority of UNet-like architecture on segmentation tasks. However, the basic UNet performed poorly in almost all regions because the depth and width of UNet were probably not sufficient to capture complex features of neurons.

In previous literature, researchers concentrated on enhancing network designs to improve cell segmentation, while the post-processing phase was usually under-investigated. This work highlights the significance of this process. Despite the fact that all post-processing techniques were applied to the same probability map, the proposed method outperformed [163] by 14% in AJI (see Table 4.3), which is more considerable than the majority of the neural network performance differences in Table 4.2. Distance transformation was less computationally expensive than ultimate erosion, and it achieved comparative results. However, it is not suited for neurons since it cannot correctly segment circle-shaped objects. The post-processing approach of [16] slightly improved the segmentation by using a better threshold and reducing FP segmentations. However, in addition to the extra training time of the regression model, it was also computationally expensive. For each object on the image, we needed to calculate the morphological information for three cell candidates and retain only the candidate with the highest predicted IoU. On the contrary, the most time-consuming step in the proposed post-processing method, the ultimate erosion, was applied at the image scale. It considerably improved accuracy while requiring less computing time than [16]. Moreover, it can be easily applied to other DL based methods and other nuclei data without expertise-demanding parameter settings. The only parameter that needs to be adjusted is the size of the structuring element used for morphological operations, which corresponds to the size of the smallest cell. Nevertheless, one prerequisite of the proposed method is that the cell contours need to be roughly smooth and without obvious concavity, as is the case with NeuN-stained neurons. Assume the cells have complex shapes with branches (*e.g.*, microglia with Iba1 and astrocytes with GFAP). In that case, in addition to retraining the deep model, modifications in post-processing will be required to preserve the particular morphological information.

A heavy data augmentation has been applied to increase the robustness of the neural network. Table 4.5 illustrates the object-level segmentation results of the proposed method on various animal subjects. Our findings suggest that the proposed method remained effective despite the difference between sam-



ples and species. The performance of detection slightly decreased in macaque and microcebus images which have distinct colors with the training set. On the other hand, a significant decline was observed in the mouse image, indicating that the light stain intensity may be a more crucial challenge than the color inconsistency between the test and training images. Hence, it could be conceivably hypothesized that stain intensity augmentation is required to further enhance the model robustness. Nevertheless, as a preliminary study, this dataset contained only four images of the cortex. Further work on larger datasets with other anatomical regions needs to be investigated to confirm this observation.

Taken together, our framework is competitive on both the tasks of detection and instance segmentation compared to other reference approaches. From point annotations to pixel-level neuron individualization, it performs well on all tested anatomical regions, with efficient architecture and more accessible parameter settings.

#### 4.2.5 . Conclusion

In this paper, we present an end-to-end DL framework to perform neuron detection and instance segmentation. A major problem with neural networks is the difficulty of producing precise annotations at large scales. We proposed a mask-synthesis pipeline to generate pixel-level labels using only point annotations, which considerably reduced the manual labeling effort and processing duration. This pipeline was applied to automatically generate annotations at large scale for the NeuN dataset, and the same strategy can be applied to other datasets. The efficiency of UNet-like design on segmentation was proved by a thorough comparison of networks based on the synthetic annotation. The proposed EfficientUNet, in particular, offered the optimum trade-off between accuracy and computation cost. It is, therefore, possible to be applied to large-scale biological studies. Using the probability maps of EfficientUNet, we compared various post-processing approaches and demonstrated the significance of this step on instance segmentation. The segmentation of the neural network was further enhanced by our post-processing method through ultimate erosion and dynamic reconstruction. In particular, the excellent performance in the cortex and the hippocampus enables us to envision further investigation related to brain functions and neurodegenerative diseases, for instance, quantitative assessment of neuronal loss to characterize animal models and to evaluate drug efficacy. More importantly, the proposed post-processing does not require ad-hoc parameter setting, which can be of great value in practice for non-expert users. The preliminary study on other animal subjects demonstrated the good generalizability of our framework. The decline in the mouse image demonstrated the impact of the staining intensity on the segmentation. Future work should consider integrating intensity changes during

the data augmentation to increase the robustness of the model toward intensity inconsistencies. The current evaluation focused on images of  $5000 \times 5000$  pixels. Patch extraction, patch prediction, and stitching were all intermediate processes that caused additional memory costs. Further works on partial image reading and writing and developments for high-performance computing are required to expand this research to whole histological sections and potentially entire brains. It would help us to better understand brain development and aging, and would also provide efficient tools to develop and validate new therapies.

### **Declaration of Competing Interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### **Acknowledgements**

The authors wish to thank the anonymous reviewers for their insightful comments and suggestions. This work was supported by DIM ELICIT grants from Région Ile-de-France. The authors would like to acknowledge Camille Mabillon for manual segmentation, Zhenzhen You for point annotation and Géraldine Liot, Suzanne Lam, Fanny Petit and Marc Dhenain for providing supplementary mouse and microcebus brain images.

### **Supplementary Figures**

The remainder of this chapter comprises supplementary figure for the following manuscript:

Wu, Huaqian, Nicolas Souedet, Caroline Jan, Cédric Clouchoux, and Thierry Delzescaux. “A general deep learning framework for neuron instance segmentation based on efficient UNet and morphological post-processing.” *Computers in Biology and Medicine* (2022): 106180.

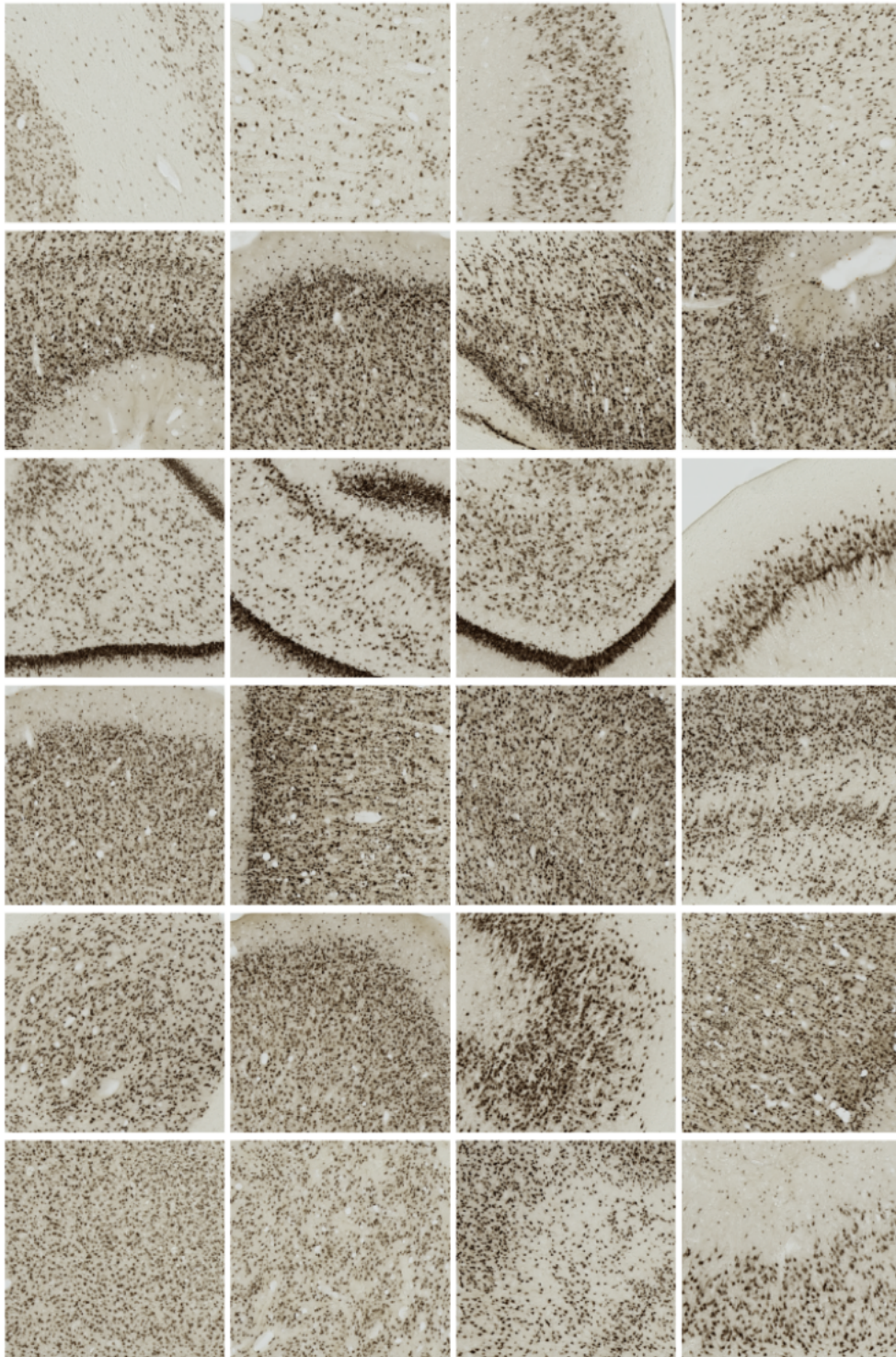


Figure 4.9: 24 images of the training set.

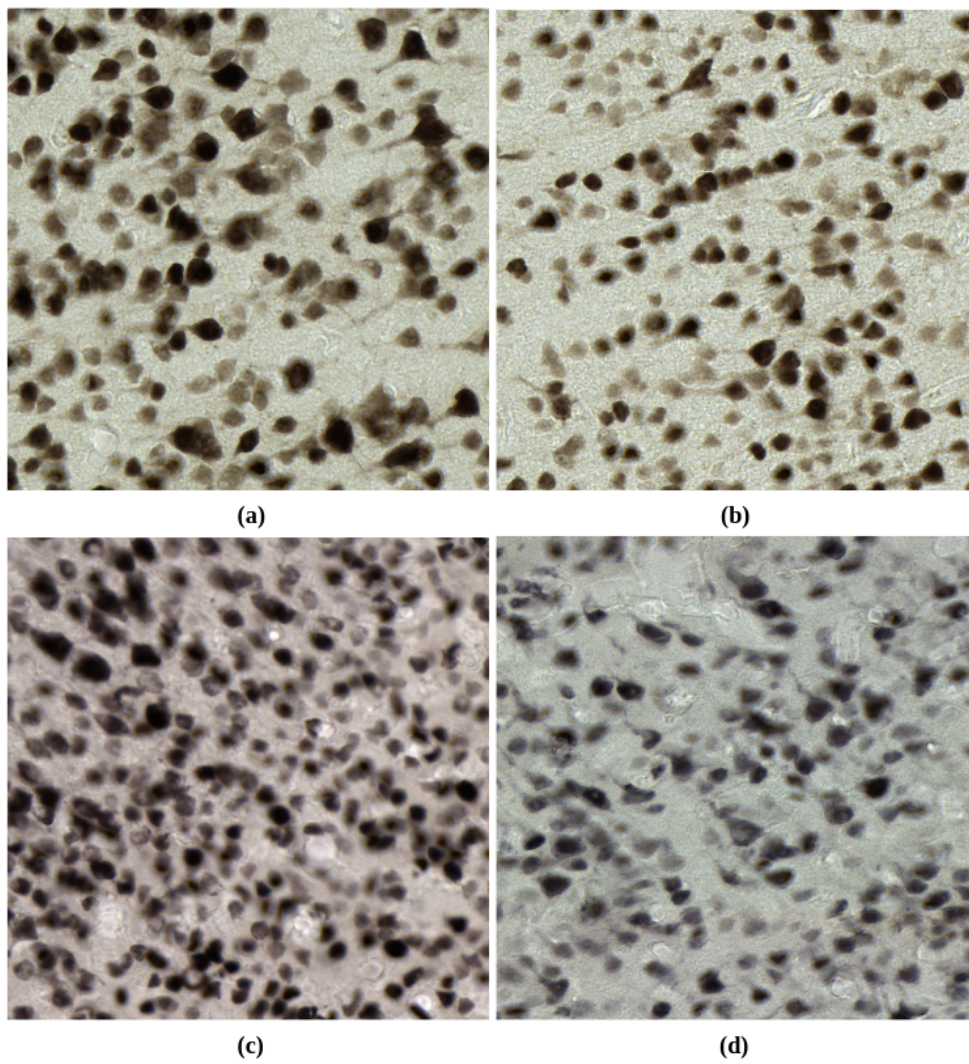


Figure 4.10: Supplementary cortex neuron dataset for assessing the generalizability of the proposed method. (a) Macaque 1. (b) Macaque 2. (c) Microcebus. (d) Mouse.

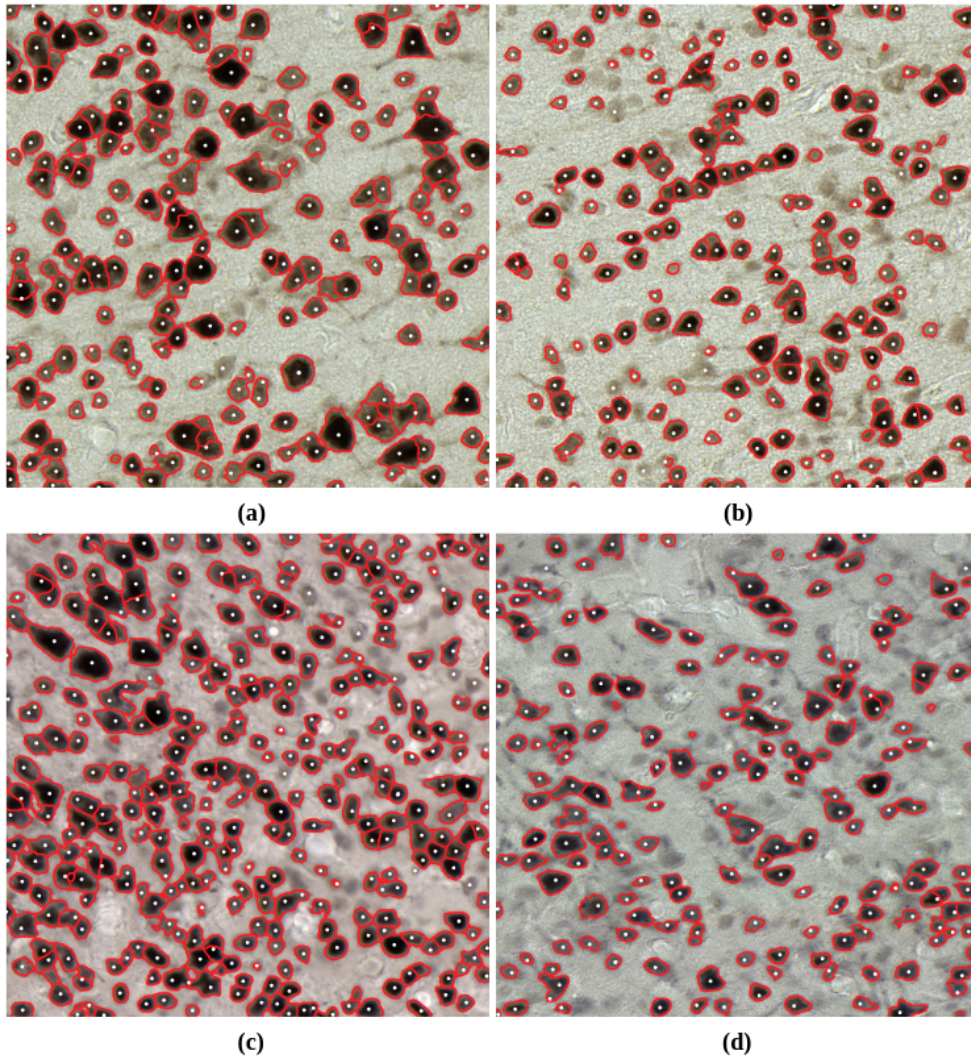


Figure 4.11: Segmentation results of supplementary cortex neuron dataset. Segmentation contours are shown in red and manual point annotations are shown in white. (a) Macaque 1. (b) Macaque 2. (c) Microcebus. (d) Mouse.

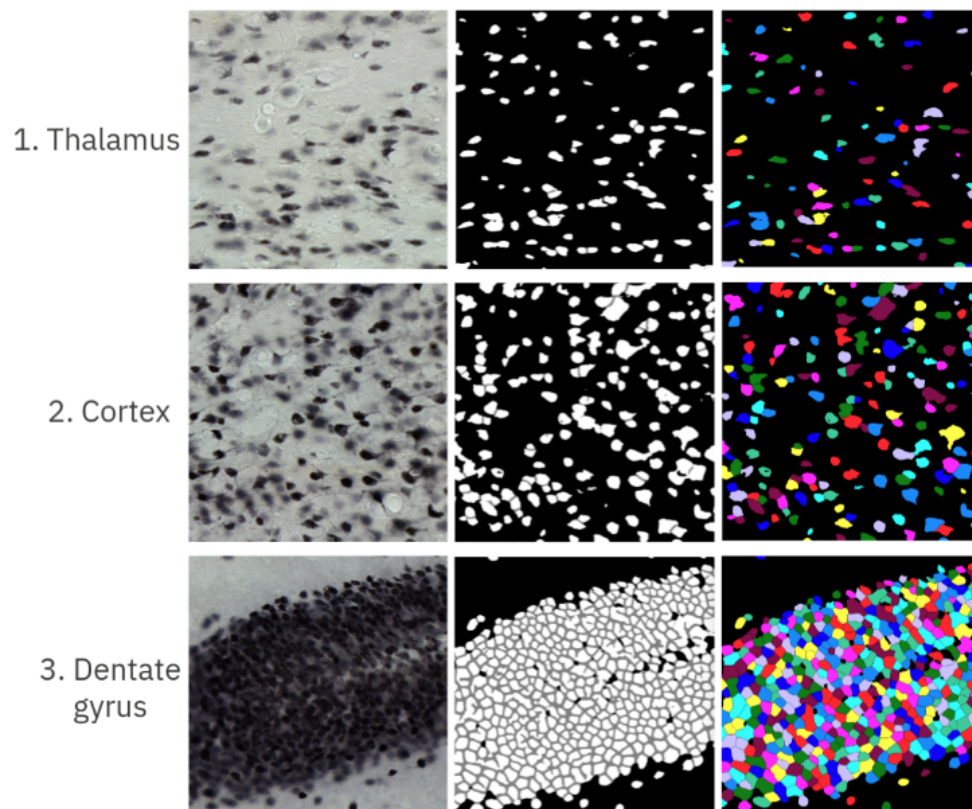


Figure 4.12: Segmentation results on three anatomical regions of a mouse brain. From left to right: raw images, semantic segmentations of EfficientUNet, and instance segmentations of the proposed post-processing.

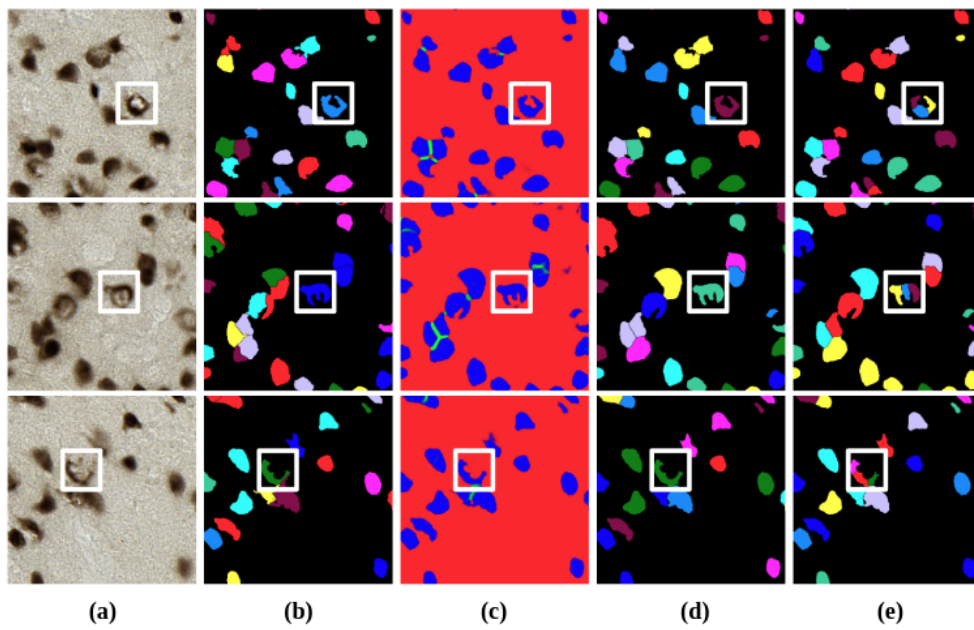


Figure 4.13: Comparison of post-processing approaches based on distance transformation and ultimate erosion. (a) raw images, (b) ground truth, (c) probability map of EfficientUNet, (d) instance segmentation of our post-processing, and (e) instance segmentation of post-processing based on distance transformation.





## 5 - Development of a segmentation inference pipeline to process whole-slide images

### 5.1 . Introduction

Chapter 4 presented an end-to-end neuron instance segmentation framework containing an efficient neural network with a complementary post-processing scheme. The proposed framework has been evaluated on images of  $5000 \times 5000$  pixels and demonstrated a high accuracy at a relatively low computational cost, showing good scalability for application to larger images. A tool that can handle WSI in a reasonable amount of time would be of great interest for analyzing massive microscopic data in biological studies. In this chapter, a software solution that enables the inference of the neuron segmentation network across whole histological sections is proposed (mouse and macaque brain). And it can be further implemented on high-performance computing resources (HPC).

### 5.2 . Whole-slide image dataset

The test data in this chapter consists of whole-slide images of a macaque brain and a mouse brain to test the generalizability of our segmentation method between species. The macaque brain section came from the same dataset presented in Chapter 3. The mouse brain section was produced with similar settings: stained by immunohistochemistry using the neuronal nuclei (NeuN) antibody and scanned by an AxioScan.Z1 (Zeiss) with the in-plane resolution of  $0.22 \mu m/\text{pixel}$  ( $20 \times$  magnification). Several differences must be noted between these two datasets: 1) the color representation of the mouse section and the macaque section is different, this color variation will be investigated in Chapter 6, and 2) the size, shape, and distribution of the neuron in the mouse brain may be different from those in the macaque brain. The size of the mouse brain section and the macaque brain section is about  $43k \times 24k$  pixels and  $231k \times 188k$  pixels, respectively. To better visualize the image details, we only illustrate the results of the mouse section since the size of the macaque section is extremely large.

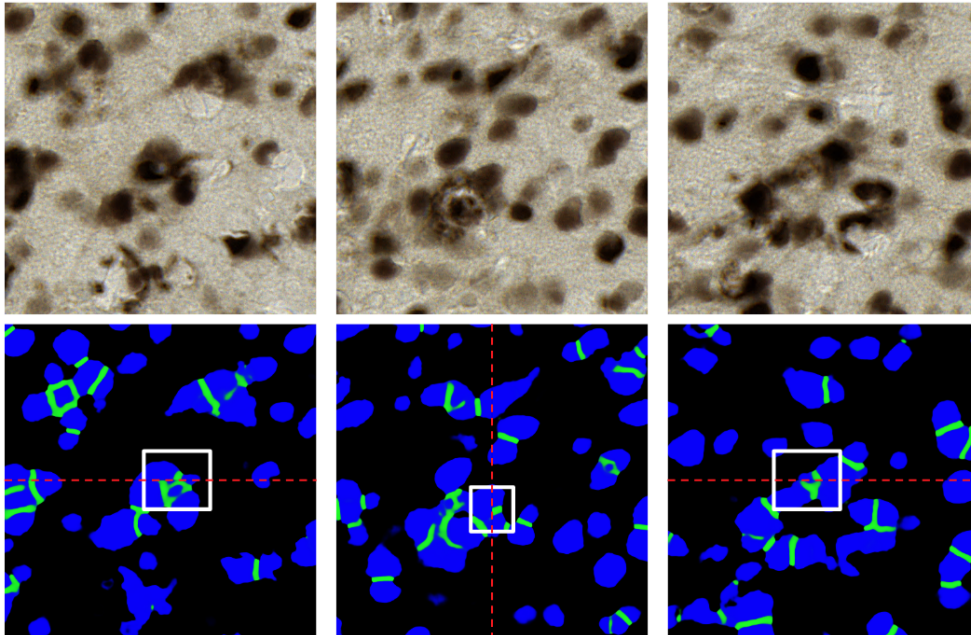


Figure 5.1: Examples of inconsistent prediction at the boundaries. The boundaries and inconsistencies are highlighted with red dash lines and white squares, respectively, for better visualization.

### 5.3 . Software development for whole-slide image processing

#### 5.3.1 . Decomposition of WSI into overlapping patches and weighted assembling

In CNNs, the convolution kernels go through the input image with a small sliding window instead of processing the full image at once. As the segmentation is a pixel-wise task, the prediction of one pixel should be the same in a small patch as in the original WSI. Mainly constrained by the GPU memory, the size of the input image for a neural network is usually no more than  $1024 \times 1024$  pixels. For an image larger than this size, a common strategy is to decompose it into smaller patches before the inference. Afterward, the prediction results of the patches are stitched together to obtain the prediction of the whole image. However, the prediction of the patch border area may be inaccurate due to insufficient context information. Stitching patches directly together may result in an inconsistent prediction at the boundaries, as shown in Figure 5.1.

This issue has been addressed in Chapters 3 and 4 by extracting patches with an overlap and applying a weight map to patches before the assembling [23]. In order to process large images ( $5000 \times 5000$ ) despite the constraints of GPU memory size, images were cropped into image patches with overlapping before

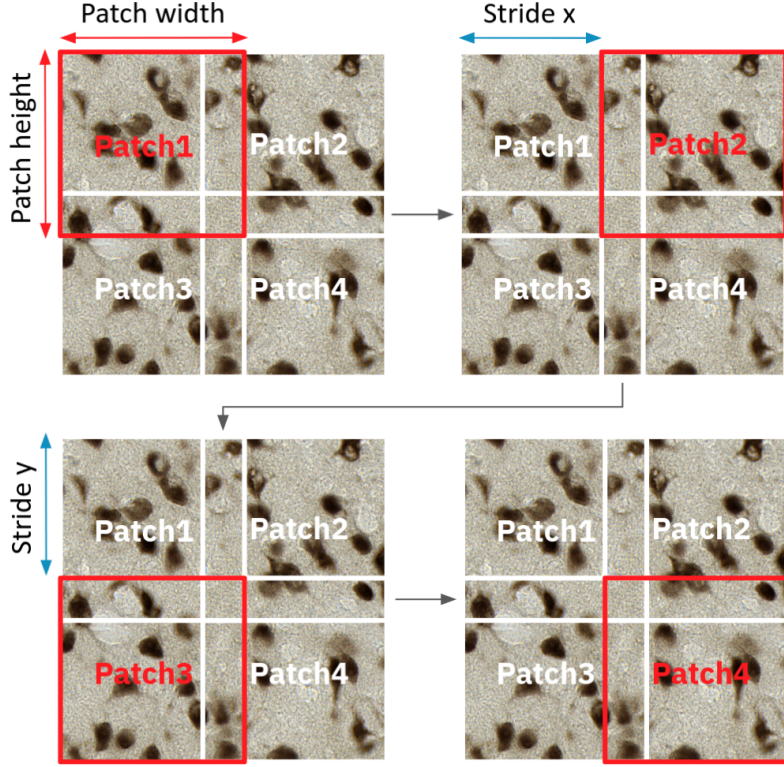


Figure 5.2: Image patch extraction with overlap.

the inference, as shown in Figure 5.2. The whole inference process using the weight map is illustrated in Figure 5.3. After the inference of patches, a weight map was applied to every predicted patch. The weight map was defined as [23]:

$$W_{i,j} = \alpha \frac{D_{i,j}^e}{D_{i,j}^c + D_{i,j}^e}; \alpha = \frac{h \times w}{\sum_{i=1}^h \sum_{j=1}^w \frac{D_{i,j}^e}{D_{i,j}^c + D_{i,j}^e}} \quad (5.1)$$

where  $i,j$  is the position of one pixel in a patch,  $W_{i,j}$  is the distance weight applied to the pixel,  $D_{i,j}^e$  and  $D_{i,j}^c$  are the distances from the nearest border and from the center, respectively.  $h$  and  $w$  are the height and width of the image patch. The weighted predictions were then seamlessly stitched together, and the stitched prediction was reweighted to obtain the final prediction (see Figure 5.3). The stitching consists of writing the patch content to its original position in the input image and summing the adjacent patches in the overlap area.

Figure 5.4 compares the assembled prediction without and with the weight map. In the first case, the predictions on the overlap area were simply averaged, which resulted in a mosaic effect at the border (see white squares in

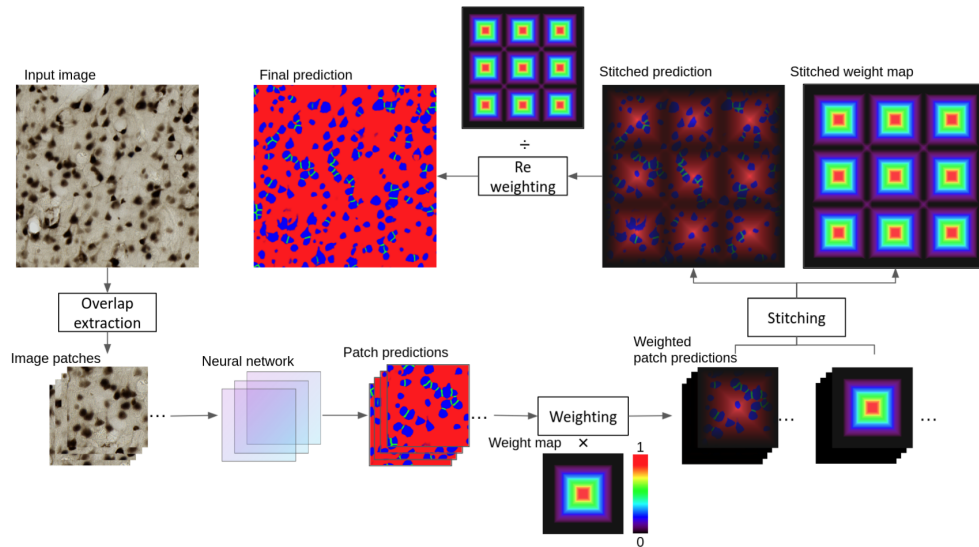


Figure 5.3: Overview of inference with overlapping extraction and assembling. For better illustration, the size of image and patch is defined as  $1344 \times 1344$  and  $512 \times 512$  pixels, respectively, with an overlap of 20%.

Figure 5.4) when the prediction of the adjacent patches for the same area was not consistent. By decreasing the weight of border pixels, using the weight map effectively reduced the border mosaic effect. After the inference, the final assembled prediction was fed as input to the post-processing scheme to produce the instance segmentation. Constrained by the architecture design, EfficientUNet requires an input image size that should be a multiple of 32. Thus a patch size of  $1344 \times 1344$  pixels (defined according to our GPU RAM properties) was chosen to minimize the number of patches and to accelerate the inference process.

### 5.3.2 . Development of a parallel computing pipeline involving both GPUs and CPUs

#### 5.3.2.1 . BrainVISA: a parallel computing software

The inference is carried out patch-by-patch when using the patch-based strategy, which can be considered as a collection of independent chain operations that do not have to interact with one another. They can be handled in parallel using BrainVISA and Soma-workflow to speed up the process.

BrainVISA is an open-source software platform dedicated to neuroimaging research. It offers a wide range of capabilities, including toolboxes for neuroimaging, visualization tools, graphical user interfaces (GUI), and optimization for massive computation. BrainVISA is extendable and customizable, allowing users to quickly incorporate personal processes to meet their needs, as shown in Figure 5.5. BrainVISA automatically generates GUIs for non-expert users

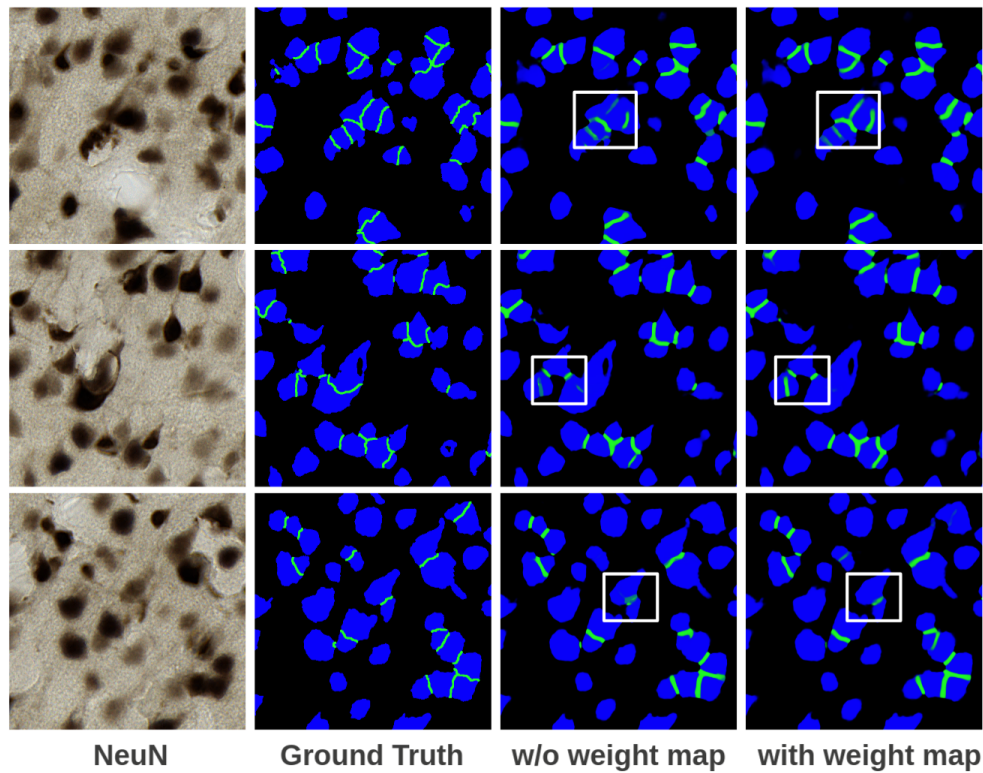


Figure 5.4: Comparison on three assembled predictions (size:  $512 \times 512$  pixels) without and with the weight map. From left to right: original NeuN-stained images, synthetic ground truth, the assembled prediction without and with using the weight map.

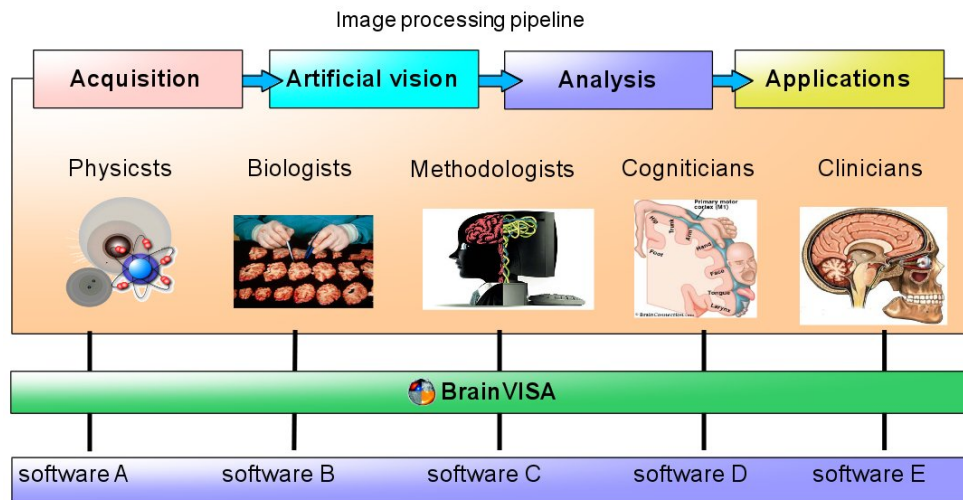


Figure 5.5: Possible uses of BrainVISA software for image analysis (image from<sup>1</sup>).

based on the parameter characteristics of processes.

Soma-workflow [71] is a library in BrainVISA software, which offers the possibility to submit, control, and monitor a set of tasks (jobs) for distributed computing using various computing resources, both locally and remotely (laptops, multiple core workstations, and clusters). The Python API for Soma-workflow can handle both independent and dependent processes. Additionally, a graphical user interface is available, as seen in Figure 5.6, providing non-expert users with a simple and quick way to track workflows.

The test dataset in previous Chapter 3 and 4 consists of  $5000 \times 5000$  pixels images. The patch size of  $1344 \times 1344$  pixels and stride of 1220 pixels yielded 16 patches per image. With the same setup, one histological section of the mouse brain ( $43k \times 24k$  pixels, 2.9 GB) corresponded to 720 patches, and that of the macaque brain ( $231k \times 188k$  pixels, 118 GB) corresponded to 29 796 patches. Accordingly, there would be 720 and 29 796 inference tasks to be processed, respectively. Such processing is extremely time-consuming with the classical inference workflow. Fortunately, BrainVISA and Soma-workflow provided the possibility to concurrently execute these tasks in parallel.

<sup>1</sup><https://brainvisa.info/>

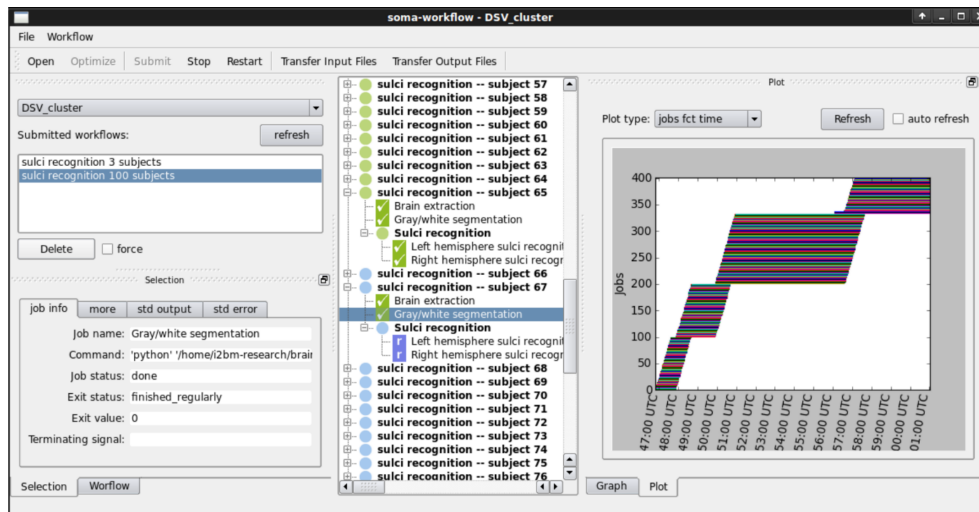


Figure 5.6: Overview of Soma-workflow GUI [71].

### 5.3.2.2 . Proposed parallel computing strategy

Figure 5.7 shows our CPU-GPU parallel computing strategy: the input image was divided into  $N$  bands, where  $N$  was the number of GPU-CPU pairs. Both the GPU and the CPU were used to process one band. The processing of one image band followed the sequential patch-based design (described in 5.3.1). This strategy has the advantage of maintaining a balanced data processing size in terms of computational resources (*i.e.*, the size of each band is the same except for the last band).

### 5.3.2.3 . Adapting the weight map to simplify the assembling

In Figure 5.3, the use of a weight map decreased the intensity of patch predictions. A reweighting process was therefore required after the stitching to recover the original intensity for each pixel. Figure 5.8 illustrates the weighting and reweighting processes for a patch located in the middle of the image. After segmentation inference, the patch prediction was weighted by the weight map to reduce the weight of the border area. Then, the patch prediction was stitched back to its original position, and the boundary was merged with the adjacent prediction. The reweighting process can be viewed as a normalization step, where the stitched prediction is divided by the stitched weight map. Both weighting and reweighting are pixel-level operations. The center tile of the stitched prediction was divided (normalized) by the center tile of the stitched weight map.

Figure 5.9 shows the stitching of nine weight map patches. The overlapping regions, covered with diagonal lines, were contributed by two or four patches. The center tile was the weight map applied in Figure 5.8 for normalizing (reweighting) the tile of the weighted prediction. The four corners and the

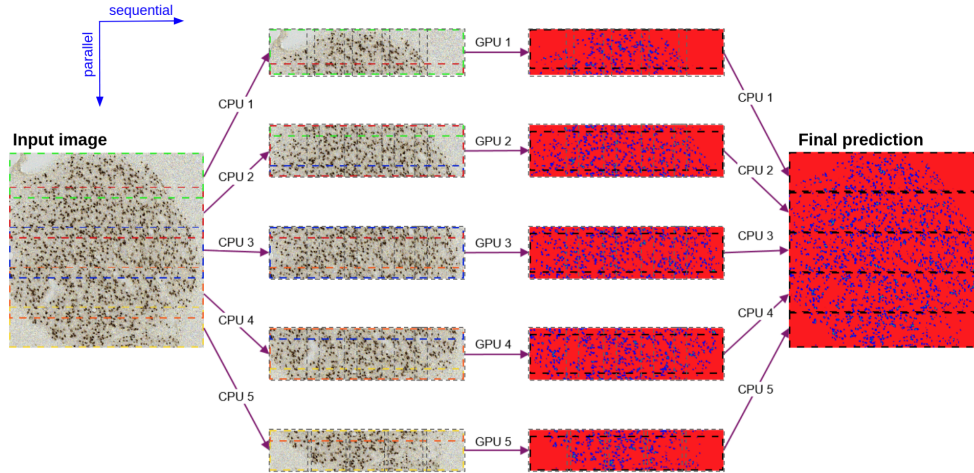


Figure 5.7: Overview of the proposed GPU-CPU parallel computing strategy.

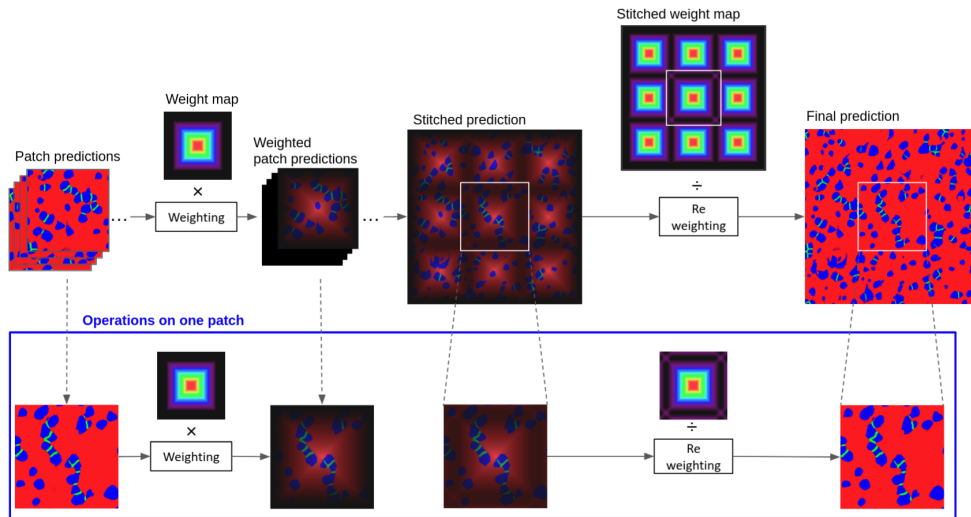


Figure 5.8: Patch assembling scheme with operations on one single patch highlighted in blue.



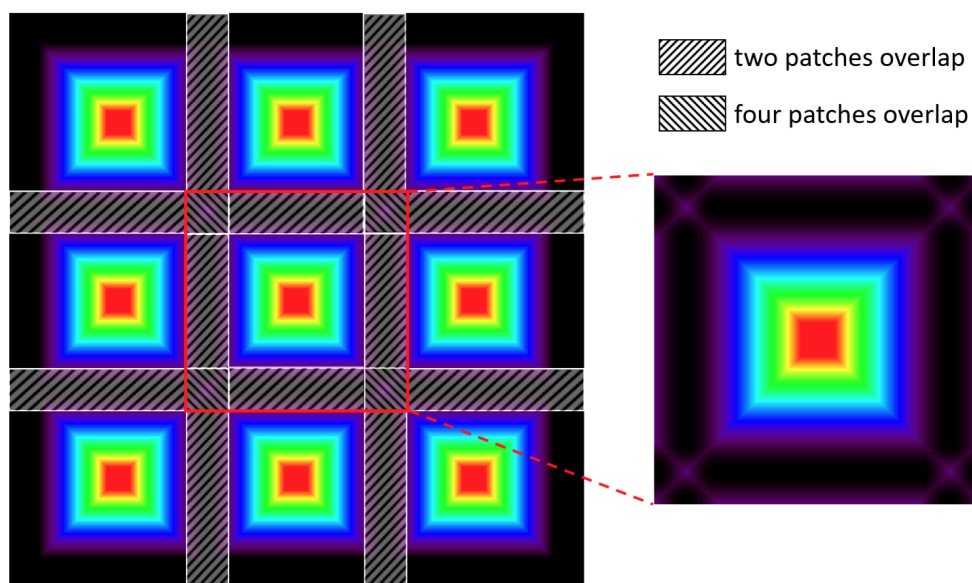


Figure 5.9: Magnified stitched weight map and the center tile.

rest of the borders were added up by four and two adjacent weight patches, respectively.

Figure 5.10 compares the weight map and the center tile of the stitched weight map. We can see that the two maps are identical except for the overlapping border area. The weighting and reweighting were just two inverse operations for most pixels in the patch. Thus, it was unnecessary to compute and apply the weight map for the region without overlapping. Therefore, we adapted the weight map to eliminate redundant calculations. As shown in Figure 5.10 right, the new weight map was generated from the previous weight map normalized (pixel-wise division) by the corresponding area of the stitched weight map. In the normalized weight map, the pixels in the center area without overlap have the same weight as one, and the weight of pixels in the overlap area is inversely proportional to the distance from the border.

Figure 5.11 shows the patching assembling using the normalized weight map. The assembling process is simplified since the reweighting step is no longer required. Moreover, the new weight map is only used to adjust the weight of the overlapping border; most pixels within the patch remain intact, which also decreases the calculation. The stitching result is shown in Figure 5.11 right. The mosaic effect in the previous stitching result is eliminated. However, the image borders have a lower intensity since they consist only of one single patch. In practice, this effect can be avoided by mirror padding the input image when extracting patches: the border of the input image is mirrored and duplicated with the size equaling that of the overlapped area; the black border of the

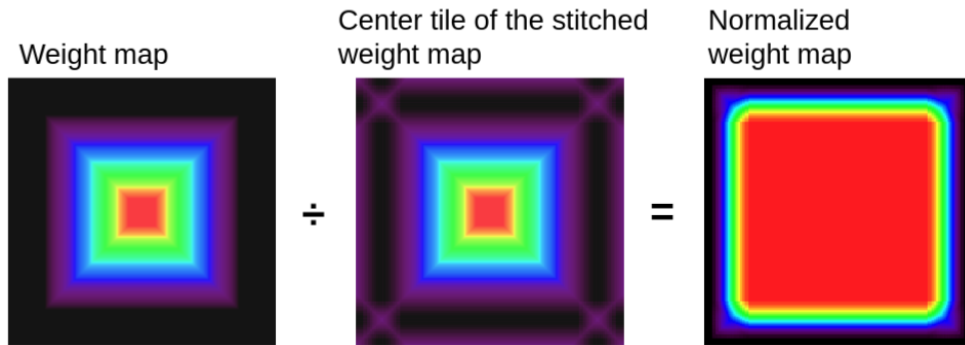


Figure 5.10: The original weight map, the center tile of the stitched weight map and the normalized weight map.

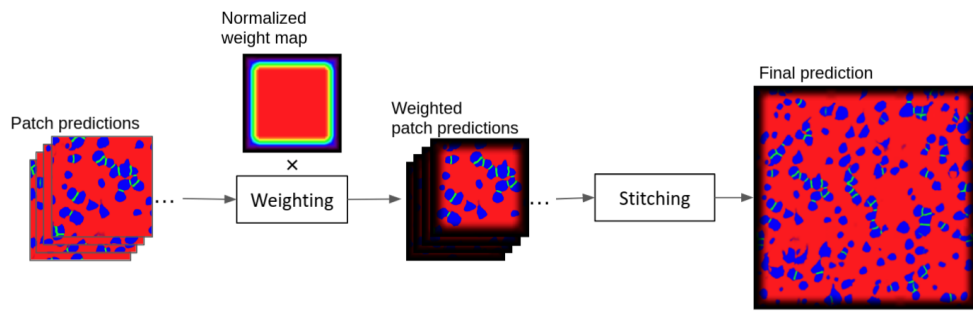


Figure 5.11: Patch assembling with the normalized weight map.

stitching result corresponds to the padding and needs to be cropped to obtain the final result.

#### 5.3.2.4 . Partial image reading and writing

As shown in Figure 5.3, the inference of overlap extraction involved several intermediate steps, and each step brought intermediate results. For example, image patches and their predictions. This could be cumbersome when processing the whole slide image, not only for loading the whole image in memory but also for storing the intermediate results. An effective way to address this problem is to execute the entire pipeline at once using partial reading and writing. Before reading the image content, we first created a three-channel empty image with the same dimension as the input image on the disk to reserve the space to store the final prediction. Then, the number and position of patches were calculated, taking into account the width and height of the stride. The extraction of the patches was performed by loading a portion of the image content in memory, which was fed as input to the DL model to generate the prediction for each patch. Finally, the weighted patches were written to the corresponding location of the output image on the disk. All operations were

carried out in memory, and only the final prediction was saved. In order to economize space, unnecessary intermediate results (image patches and their prediction) were not stored on the disk by default.

### 5.3.2.5 . Summary

The strategy can be summed up as follows :

1. N bands' coordinates are calculated according to the number of GPUs and CPUs.
2. Each band is loaded to memory by a CPU using partial reading.
3. Each band is processed using a pair of CPU-GPU and the patch-based design. Each patch of image band I is iteratively passed to GPU I, where the model has been initialized with the trained model. The prediction for each patch of image band I is stitched in memory using CPU I.
4. Once the whole image band is processed, the CPU I partially writes the output result to disk. To deal with concurrent access issues in overlap areas during parallel processing, CPU I does not write predictions to disk in the bottom overlap area of image band I. The bottom overlap area of image band I, which is also the top overlap area of image band I+1, will be written to disk by CPU I+1 after the processing of image band I+1.

### 5.3.2.6 . Interface of the segmentation inference pipeline for whole-slide images

The interface of our inference pipeline is illustrated in Figure 5.12. Three parameters are necessary, including the path of input, output image, and the trained model weight. Prediction classes, patch size, overlap, GPU settings, and other parameters have predefined values but can be easily modified.

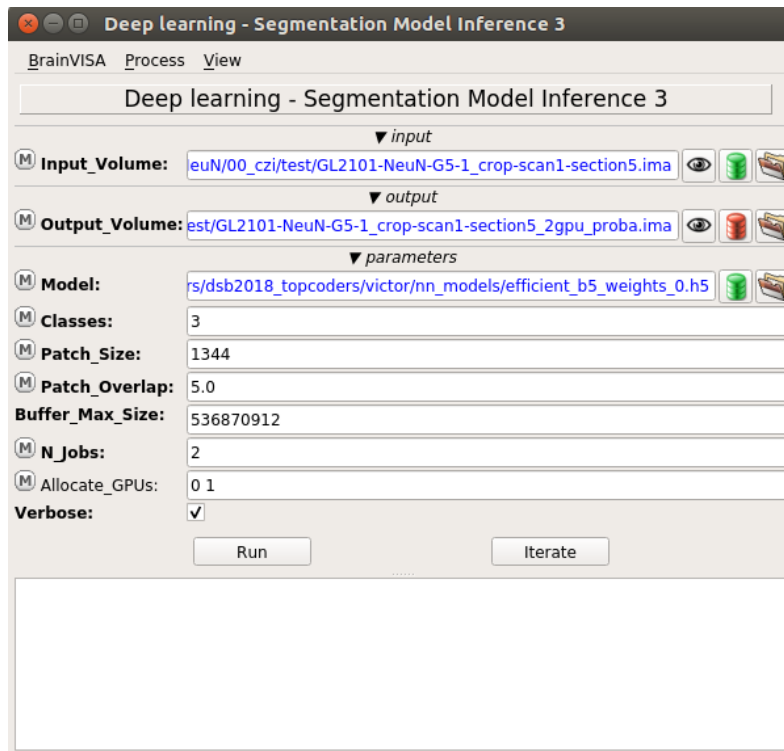


Figure 5.12: Graphic interface of the inference pipeline.

## 5.4 . Application of the proposed pipeline for segmenting whole-slide images

### 5.4.1 . Processing time under different configurations

Table 5.1 compares the processing time on different images on a single GPU. The processing time and the number of patches are linearly correlated to the size of the image. With 16 patches predicted in less than a minute, the prediction of small images ( $5000 \times 5000$  pixels) was relatively fast. Although it took over 40 times longer, processing the complete mouse brain section was still bearable (34 min). However, the processing time increased to 24 hours for an entire macaque brain section, which was 41 times larger than the mouse brain section. For the entire macaque brain, which consists of 134 sections, it would require 134 days to process on a single GPU (NVIDIA Quadro RTX A5000 with 24 GB memory, the following calculation is performed on the same GPU if not specified).

We evaluated the performance of the parallel computing pipeline on a mouse brain section with different overlap ratios and GPU settings, with the resulting processing times shown in Table 5.2. The number of GPUs below implies the number of GPU-CPU pairs since, in our strategy, one GPU is associated with

## 5.4. APPLICATION OF THE PIPELINE

Table 5.1: Computation details of inference for various images on 1GPU (overlap ratio: sim 10%).

Image type	Image size	File size	Number of patches	Processing time
Image of test set	5000 × 5000	72 MB	16	< 1 min
Mouse brain section	43k × 43k	2.9 GB	720	34 min
Image of test set	231k × 188k	118 GB	29 796	< 24 h

Table 5.2: Time in minutes for processing one mouse brain section with different overlap ratios and GPU resources.

Overlap ratio	Number of patches	Processing time (min)		
		1GPU	2GPUs	3GPUs
10%	720	34	18	13
50%	2 205	106	58	42

one CPU. As shown in Table 5.2, an overlap of 50% brought three times the workload (number of patches) and processing time compared to the overlap of 10%. Theoretically, the processing time should scale linearly with the number of GPUs because practically all operations—including partial image reading, prediction, and partial image writing—are distributed over CPU-GPU pairs. Figure 5.13 shows the relation between the processing time and the number of GPUs for different overlap ratios. By using a second GPU, the inference time was reduced by almost half, with 52.9% and 54.7% savings for overlaps of 10% and 50%, respectively. However, the benefit of additional GPUs became less significant when we increased the number of GPUs from two to three. Only 5 and 16 minutes were saved by the third GPU for processing the mouse brain section with a ratio of 10% and 50%, respectively. There are a variety of factors that might be related to this issue: technically speaking, the GPUs in our implementation were not independent of each other. They were connected via the same PCIe (Peripheral Component Interconnect Express) link, which limited the bandwidth available for each card in comparison to the case of a single GPU.

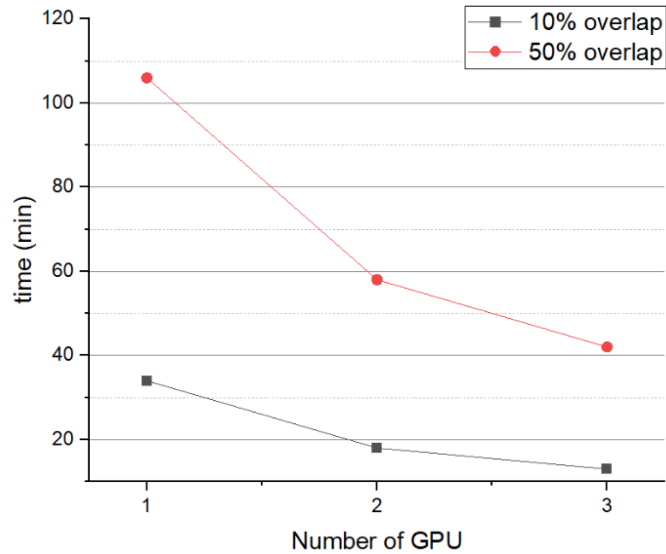


Figure 5.13: Processing time versus the number of GPUs for one mouse brain section.

#### 5.4.2 . Segmentation results of the proposed pipeline on the mouse brain section

The tested mouse brain section is presented in Figure 5.14, as well as three typical anatomical regions with different neuron distributions, which are the thalamus (sparse), cortex (dense), and dentate gyrus (very dense), respectively. The neurons in the mouse section are darker in color than those in the macaque section, despite the fact that both sections were dyed and scanned similarly. Additionally, the neurons in the three regions have different densities. They also vary significantly in size, shape, and intensity.

Due to the large size of the macaque brain, we present only the results for the mouse section to better visualize the segmentation.

Our inference pipeline provides two types of output according to the user's need: the probability map and the class map. By default, the pipeline saves the prediction of the neural network as a three-channel float image to preserve a high level of precision rather than rounding it into a three-channel U8 image, as in Chapter 3 and 4, as shown in Figure 5.15, Figure 5.16, and Figure 5.17. The pixel values in each channel vary between 0 and 1, representing the probability of being tissue, inter-cell contour, and neuron class, respectively. However, the size of three float images for the mouse brain section would have corresponded to approximately 12 GB, which could have been cumbersome to process or even to store on the disk. An alternative strategy to conserve memory usage is to store the class map, as shown in Figure 5.18. First, we created an empty U8 image with the same dimension as the input whole-slide

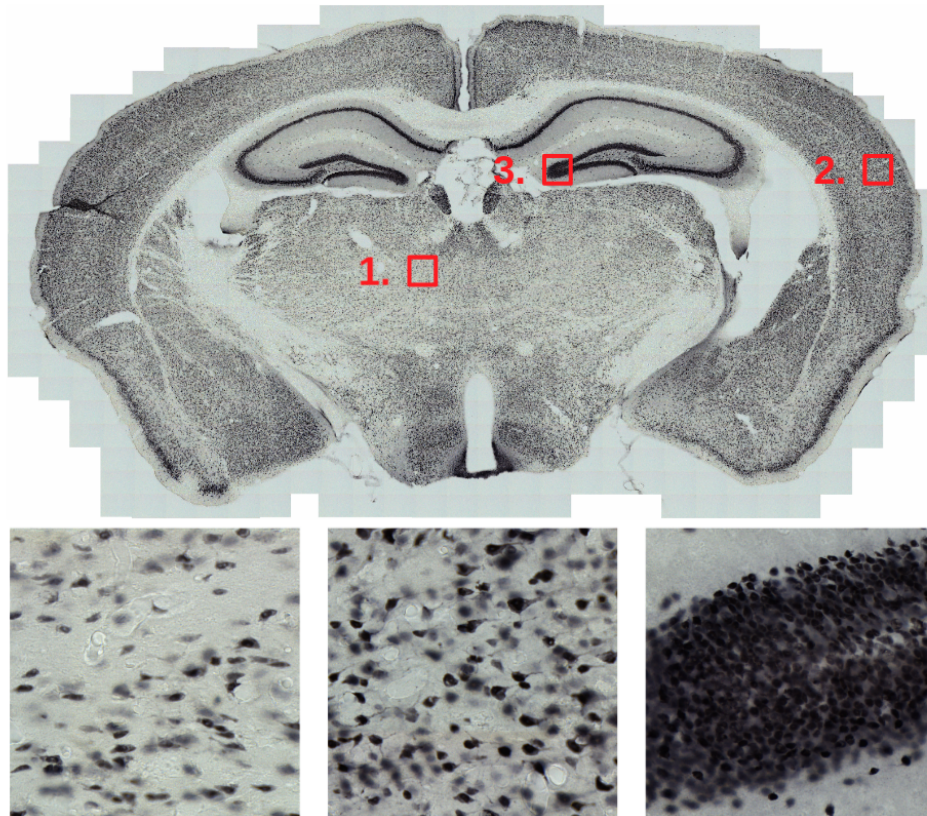


Figure 5.14: A NeuN-stained mouse brain section. Below: 1. thalamus, 2. cortex and 3. dentate gyrus.

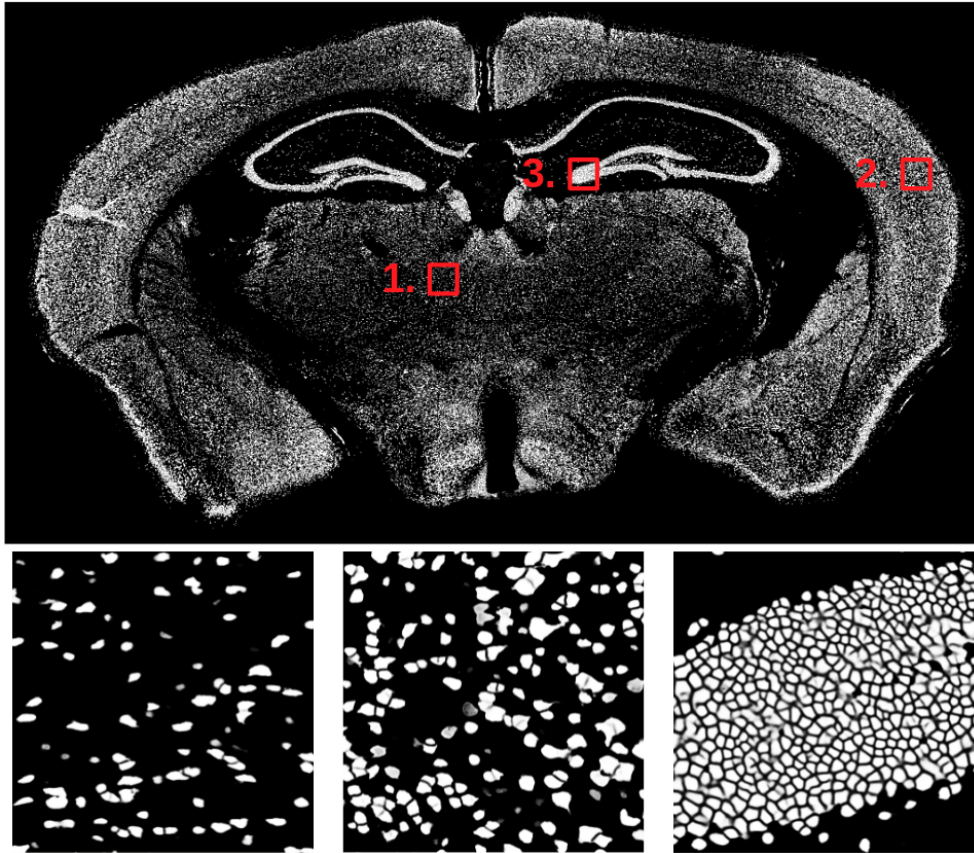


Figure 5.15: Neuron probability map of the mouse brain section. Below: 1. thalamus, 2. cortex and 3. dentate gyrus.

image on the disk. The dominant class of each pixel in the class map was then determined by comparing the three probabilities. The class map had three values (0, 1, and 2), which stood for the tissue, intercellular contours, and neurons, respectively. It could be stored as a single-channel U8 image. As a result, the size of the probability maps was decreased to 1/12 of the original three-channel float image (size of float type: 4 Bytes, number of channels: 3), which allowed for compression without sacrificing any crucial information for further processing.

The segmentation neural network used in the inference process was derived from Chapter 4 and trained on macaque neuron data rather than the mouse neuron dataset. Interestingly, the model was still robust, and segmentation results were promising, despite the fact that mouse neurons differ from macaque neurons in size and color. In the three anatomical regions, the model performed well in classifying the pixels into the corresponding class. However, the neural network showed its limitation in regions without neurons. The model misclassified the boundaries of the mouse brain section as the neuron



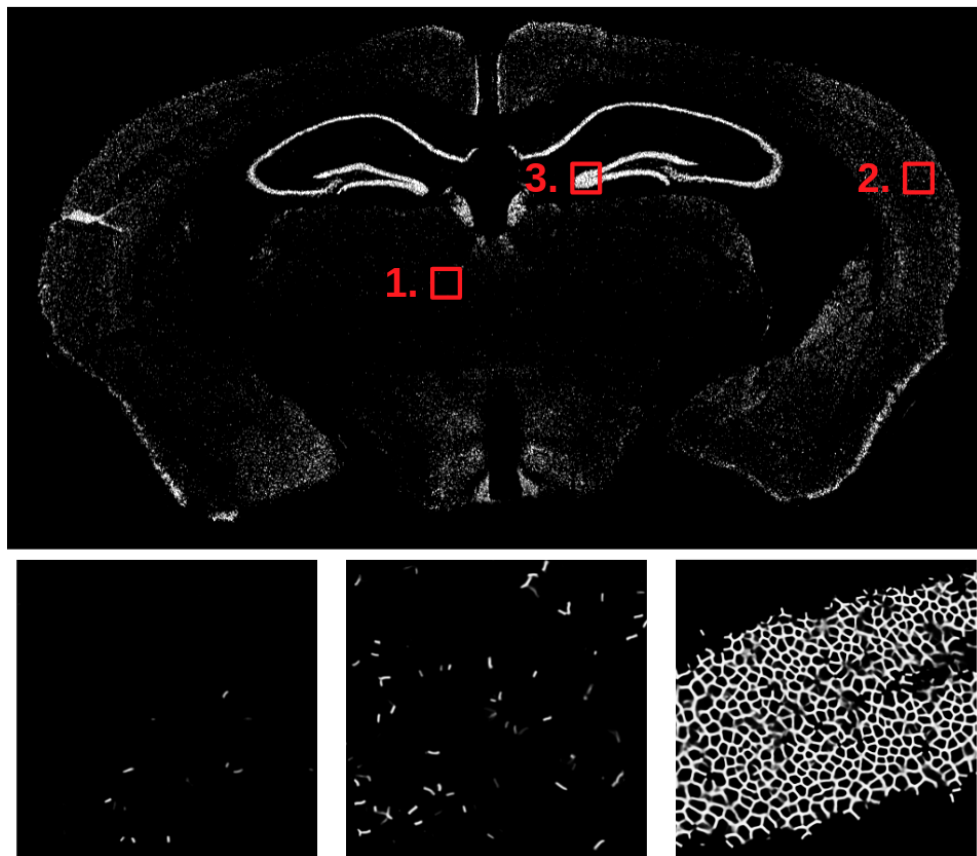


Figure 5.16: Inter-cell contour probability map of the mouse brain section. Below: 1. thalamus, 2. cortex and 3. dentate gyrus.

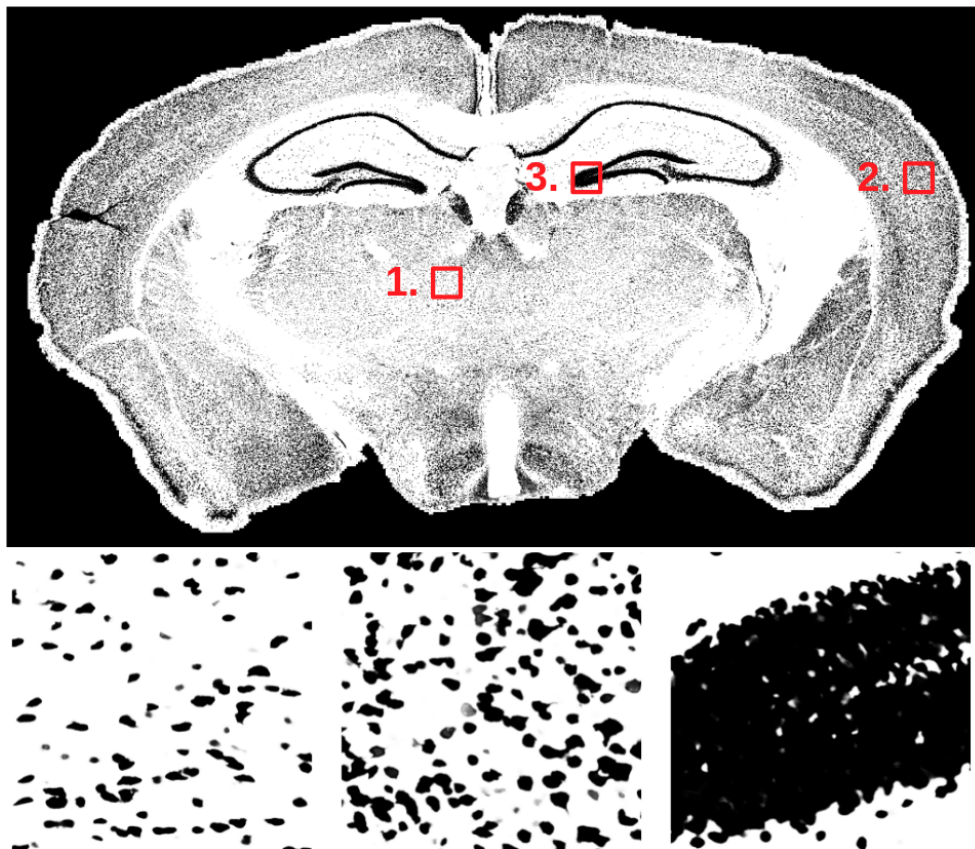


Figure 5.17: Non-stained tissue probability map of the mouse brain section. Below: 1. thalamus, 2. cortex and 3. dentate gyrus.

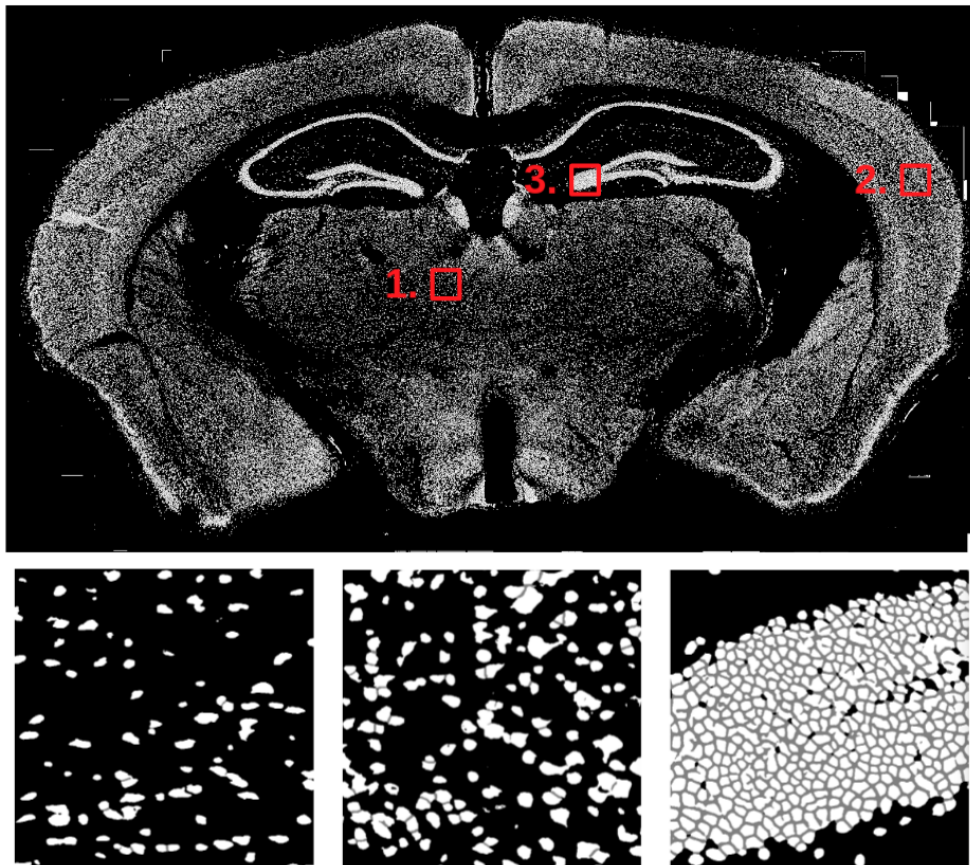


Figure 5.18: Class map of the mouse brain section. The black, gray and white colors represent non-stained tissue, inter-cell contours and neurons, respectively. Below: 1. thalamus, 2. cortex and 3. dentate gyrus.

class, while there was no neuron in these regions, as seen in Figure 5.18. This may be related to the lack of such data in the training set of the neural network. A possible solution is to fine-tune the network on a complementary dataset that comprises the missing non-stained area data.

The result of our segmentation inference pipeline is the semantic segmentation of neurons, non-stained tissue, and inter-cell contours, which requires further processing to separate and label each neuron. We applied the post-processing from Chapter 4 on the probability map of the entire mouse brain. The final instance segmentation is shown in Figure 5.19. The results indicated that the mouse brain section contained 95 765 neurons in total, which were labeled from left to right and from top to bottom with a rainbow colormap. The neurons in the three anatomical regions were recolored to better differentiate the touching neurons. The wrong classification in the tissue (Figure 5.18) further impacted the final instance segmentation: as shown in Figure 5.19, in particular in the lower right region, the border of the mouse brain section was wrongly considered as neuron instances. Fine-tuning the segmentation network with the non-stained tissue data would improve the precision of the network in these regions, but this would need new data collection and extra training. This problem can also be solved by adding a filter to the post-processing to remove non-disc-shaped and oversized objects. Although the intensity, density, and morphology of neurons varied across regions, the segmentation results were visually promising for all regions tested. However, quantitative assessments need to be investigated to rigorously validate the use of our segmentation model in different species.

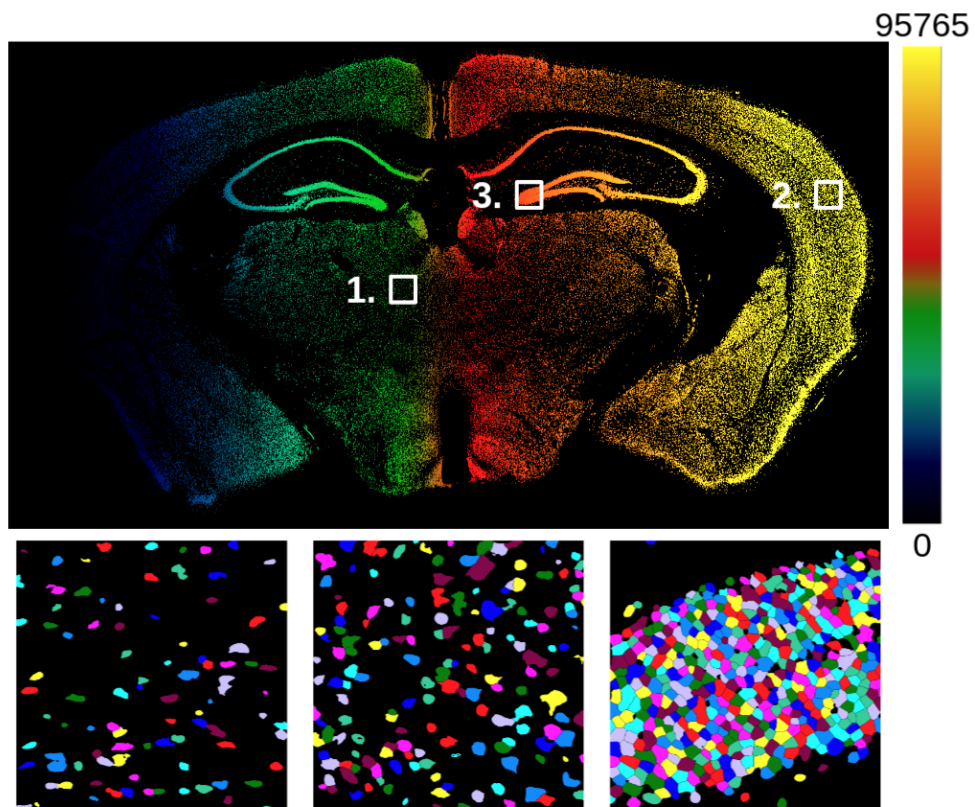


Figure 5.19: Instance segmentation of the mouse brain section (rainbow colormap). Below: 1. thalamus, 2. cortex and 3. dentate gyrus.

### 5.5 . Conclusion

This chapter presents a software solution for applying our segmentation framework to whole-slide images. By integrating GPU-CPU parallel computing and a partial image reading/writing library, the proposed process enables, for the first time, the inference of the segmentation network on whole-slide images. Based on the prediction of the pipeline, our post-processing scheme further individualized neurons over an entire mouse brain section. This can be of great value in biological research. Information about neuron number, size, and morphology of whole brain sections would be extremely beneficial for studying brain development, aging, and neuron degeneration, as well as evaluating novel treatments. Furthermore, given that the training set only contains data from one macaque brain section, the first segmentation results obtained on the whole mouse brain section are promising and let us envision generalizing our segmentation framework to other animal samples and cell types.

## 6 - Color normalization in microscopic images

### 6.1 . Color variation in microscopic image analysis

Stain color variation in histological images, caused by a variety of factors, is a challenge not only for the visual diagnosis of pathologists but also for cell segmentation algorithms. Many factors in the staining procedure are related to this issue. For example, dye type, solution concentration, staining protocol, and even room temperature. Even if the staining procedure is strictly standardized, variations may still occur during the digitization process, such as different scanners and scanning parameters. As mentioned earlier in Chapter 3, the macaque brain data used in this thesis work was divided into two series for the staining due to a large number of tissues. This resulted in a color inconsistency between the two series despite the staining and the scanning procedures were standardized. The training dataset of the neuron instance segmentation networks was built based on section 91 of the odd-numbered section, which is brown. However, the neurons in even-numbered sections are darker, as shown in Figure 6.1. Section 90 is the adjacent tissue of section 91, and similar neuron distributions can be found in the two sections. However, applying the segmentation network trained on images from section 91 to images from section 90 may suffer from this color inconsistency.

Most conventional color normalization methods in literature were designed for H&E-stained images and are unsuitable for images with immunohistochemical staining, as the color statistic mapping is performed on color spaces other than RGB. Therefore in this chapter, we focus on DL-based color normalization approaches. Two methods are identified and tested: Pix2Pix [57, 107] and ReHistoGAN [2]. Both of them are built on conditional GAN (cGAN). With no restrictions on the color of the input image, the former addresses color normalization as a problem of recoloring gray-scale images. The latter has the flexibility of applying arbitrary target colors. The work on ReHistoGAN is presented as a conference paper in Section 6.3.

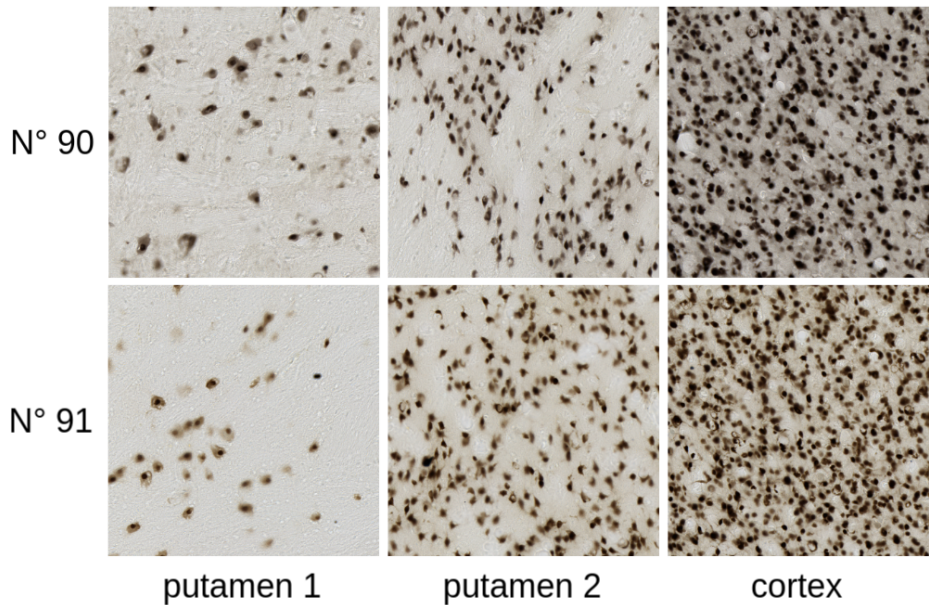


Figure 6.1: Image patches extracted from sections 90 and 91. From left to right: putamen 1 (sparse), putamen 2 (dense), and cortex.

## 6.2 . Color normalization using Pix2Pix conditional GAN

Pix2Pix was proposed by [57] as a general solution to the image-to-image translation problem. The networks can be trained to map input images from one space to another target space. For instance, creating photographs from label maps, reassembling items from edge maps, coloring gray-scale images, *etc.* Inspired by the last application, Salehi *et al.* [107] applied the Pix2Pix framework to normalize color variations in H&E histological images and revealed promising results. This preliminary work intended to test its capacity on NeuN-stained images of the macaque brain dataset presenting color variations.

Figure 6.2 illustrates the architecture of Pix2Pix, containing a generator and a discriminator, which are unsupervised adversarially trained. The original RGB images were first converted to gray scale. The objective of the generator was to map the gray-scale images into the original color space using a UNet architecture with skip connections. The real pair in Figure 6.2 is made of a gray image and its original RGB image, while the fake one is the pair of the gray image and the image recolored by the generator. At the same time, the discriminator was trained to distinguish the fake pair from the real one. Pix2Pix has an advantage over other GAN-based color normalization approaches: it can be applied to source images no matter its original color, as input images are converted into gray-scale prior to be recolored by the generative network.



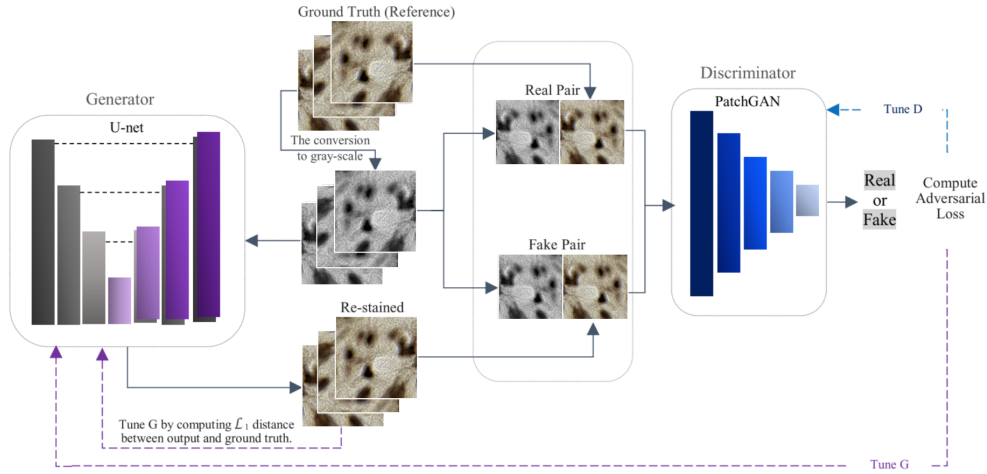


Figure 6.2: Architecture of Pix2Pix framework. Image modified from [107].

### 6.2.1 . Training and test dataset for Pix2Pix

The intention of this preliminary work was to correct the color variation between the two section series of the macaque brain in order to extend the neuron instance segmentation network presented in Chapter 4 to the even-numbered series. This segmentation network was trained using images from section 91. The objective of color normalization was to recolor test images extracted from section 90 using the color of section 91 as the reference before applying the trained segmentation network. Therefore, we created a training set based on section 91 to train Pix2Pix. This dataset contained 3249 images of  $256 \times 256$  pixels, for which  $\frac{1}{5}$  of them were used for validation (same setting as in [107]). Most images came from the cortex, as it is the anatomical region that interests most neurobiologists. The test set contained a large cortex image of  $5000 \times 5000$  pixels extracted from image 90, which corresponded to 361 patches of  $256 \times 256$  pixels.

### 6.2.2 . Loss function of Pix2Pix

The generator ( $G$ ) and the discriminator ( $D$ ) in Figure 6.2 were trained simultaneously. The loss function of the Pix2Pix was defined as:

$$L_{cGAN}(G, D) = E_{x,y}[\log D(x, y)] + E_{x,z}[\log(1 - D(x, G(x, z)))] \quad (6.1)$$

where  $x$  is the gray-scale image,  $z$  is a random noise vector, and  $y$  is the ground truth RGB image. The objective of the generator was to minimize 6.1, whilst that of the discriminator was to maximize it. In addition to confusing the discriminator, the generator was expected to generate images that were close to the ground truth. The second objective was constrained by the  $L_1$  loss since it encouraged less blurring than  $L_2$  [57], which was defined as:

$$L_{L1}(G) = E_{x,y,z}[\|y - G(x, z)\|_1] \quad (6.2)$$

Overall, the final loss function became:

$$L = L_{cGAN}(G, D) + \lambda L_{L1}(G) \quad (6.3)$$

where  $\lambda$  is a hyperparameter that controlled the balance between the two objectives, which was set to 100 [107].

During the inference, only the generator would be used to recolor the images. Finding a generator that satisfied the following criteria was the main goal of the entire training:

$$L^* = \arg \min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G) \quad (6.4)$$

### 6.2.3 . Evaluation metrics to evaluate color normalization of Pix2Pix

Visual evaluation is frequently used in the literature to qualitatively verify the performance of color normalization approaches [82]. Quantitative evaluations were often based on the structural similarity index measure (SSIM) [146] between recolored and target images [132, 107, 114], which was defined as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (6.5)$$

where  $x$  and  $y$  are recolored and target images.  $\mu$ ,  $\sigma$  are the average and the variance of the intensity of the image, respectively.  $c_1$  and  $c_2$  are small variables that are added to stabilize the division. SSIM of one color image is calculated for all color channels and averaged. One subtle difference in the image content can impact the value of SSIM, which limits its applicability. It is only pertinent when the source image and the target image have the same content. For instance, [132, 107, 114] used SSIM to evaluate color normalization on the same images digitized using different scanners. On the other hand, when the target and the source do not share the same image content, SSIM is no longer relevant.

The target image and the source image in our study came from two different histological sections. Although the two images represented the same anatomical region (cortex), the image contents were similar but not identical. Thus, we calculated the similarity of the color histogram between the recolored image and the target to evaluate the color normalization. A color histogram represents the color distribution of an image. We measured the color histogram for two color spaces: RGB space and log-chroma space [2]. In RGB space, the

value of each channel represents the intensity, allowing the color histogram to account for both color and illumination information. On the other hand, in log-chroma space, the value of one channel is normalized by other channels, which makes the histogram insensitive to illumination changes.

For each color space, Hellinger distance ( $H\_dist$ ) and Kullback-Leibler divergence ( $KL\_div$ ), also known as relative entropy, were used to measure the difference between color histograms.  $H\_dist$  measures the square root difference of two distributions, which is then summed by the Euclidean norm.  $H\_dist$  is symmetric and bounded between  $[0, 1]$ . It is defined as:

$$H\_dist(P, Q) = \sqrt{\frac{1}{2} \sum_{x \in X} \left( \sqrt{P(x)} - \sqrt{Q(x)} \right)^2} \quad (6.6)$$

where  $P$  and  $Q$  are color histogram distributions of target and recolored images, and  $X$  stands for the ensemble of bins of the histogram.

$KL\_div$  is asymmetric and unbounded. It is defined as:

$$KL\_div(P||Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \quad (6.7)$$

Both  $H\_dist$  and  $KL\_div$  were used to quantify the similarity between the two color histogram distributions. The similarity increased as the value approached zero.

#### 6.2.4 . Conventional color normalization methods for comparison

We compared Pix2Pix with other conventional color normalization approaches presented in Chapter 2, including color transfer algorithm [102], singular value decomposition (SVD)-based normalization [82] and structure-preserving color normalization (SPCN) [132]. These methods have been successfully applied to H&E-stained histological images. For the first time, this study evaluated their performance on NeuN-stained images.

#### 6.2.5 . Evaluation of Pix2Pix and other compared methods

Figure 6.3 shows the color normalization results of Pix2Pix and other state-of-the-art methods. As shown in Figure 6.3 (c), [102] altered the color of entire images from dark to brown, including the tissue color, which was even browner than the tissue in the target image. Moreover, the intensity of the recolored results has been changed. The results of [102] were brighter than both target and source images. [82] and [132] obtained similar recoloring results, which were visually better than those of [102]. However, they also encountered the issue of intensity change, especially [102]. This problem was

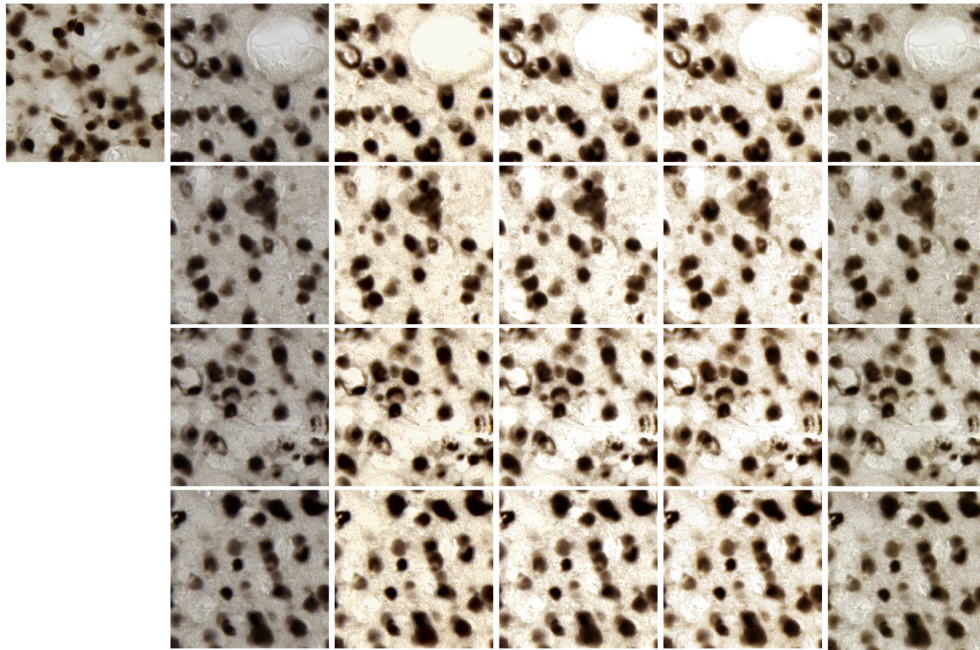


Figure 6.3: Color normalization results of different methods. From left to right: target image, source images, recoloring results of [102], [82], [132] and Pix2Pix, respectively.

already identified in a stain separation study that was conducted to separate the Hematoxylin (H) and DAB stain channels on an IHC image [132]. The luminosity of the results was more intense than the input image despite the H and DAB channels being clearly separated, and the density of the separated DAB did not correspond to the amount of DAB that was actually present. The authors [132] made the assumption that the RGB color space to the Optical Density (OD) color space conversion was invalid for DAB since it was a light scatterer rather than an absorber. Due to the fact that normalization was also carried out in the OD space, the intensity issue of [82] might be connected to the same cause. Taken together, all tested conventional methods were inadequate to handle color variations in NeuN-stained images. On the other hand, Pix2Pix demonstrated promising results, preserving both the target color and the source content while maintaining the illumination.

Table 6.1 shows the quantitative evaluation of different methods based on color histograms of RGB space. Pix2Pix outperformed others with the lowest H\_dist and KL\_div. The KL\_div indicated significant differences between Pix2Pix and other methods, with the value of other methods being four times superior to that of Pix2Pix, and the H\_dist of Pix2Pix being approximately half of those of other methods. The high similarity between the target histogram and the histogram of the recolored image is coherent with the obser-

## 6.2. COLOR NORMALIZATION WITH PIX2PIX

Table 6.1: Comparison of color transfer methods. The H\_dist and KL\_div are calculated in RGB color space.

metrics	Reinhard	Macenko	Vahadane	Pix2Pix
H_dist	0.5374	0.5481	0.5491	<b>0.2358</b>
KL_div	0.8645	0.8605	0.8706	<b>0.1765</b>

Table 6.2: Comparison of color transfer methods. The H\_dist and KL\_div are calculated in  $uv$  color space.

metrics	Reinhard	Macenko	Vahadane	Pix2Pix
H_dist	<b>0.1007</b>	0.1115	0.1381	0.1300
KL_div	<b>0.0433</b>	0.0485	0.0737	0.0687

variation in Figure 6.3. The recolored result of Pix2Pix presented the closest color to the target image. However, no substantial difference was observed among the conventional methods. [132] appeared to be the worst method in both H\_dist and KL\_div. [102] had slightly lower H\_dist than [82], while the opposite was true for KL\_div.

To help understand whether the poor results of the conventional methods were associated with the unnatural brightness variations shown in Figure 6.3, we compared the color histograms in the log-chroma space. The intensity of each color channel was normalized and the color histogram represented the proportional distribution of the R, G, and B color channels. Changes in illumination altered the intensity of the R, G and B channels, but the proportional distribution between channels remained intact. The quantitative evaluation in log-chroma space is illustrated in Table 6.2. The most interesting aspect of this table is how the H\_dist and KL\_div of the conventional approaches dropped once the effect of illumination was ignored. These results further supported the idea that the poor performance reported in Table 6.1 was due to the illumination issue illustrated in Figure 6.3. In this color space, all tested methods were at the same level, and Pix2Pix was no longer the best method, with H\_dist and KL\_div higher than [102] and [82]. The method which obtained the best scores was [102], followed by [82]. The performance of [132], however, remained the poorest.

By comparing Table 6.1 and Table 6.2, we can conclude that the poor performance of conventional methods in Figure 6.3 was mainly due to variations in illumination. Although these methods yielded correct proportional distributions of R, G, and B color channels, the problem of illumination may mislead the visual diagnosis of pathologists and affect the accuracy of CAD systems.

A process of illumination normalization would be mandatory in practice to extend the use of these methods in IHC-stained images.

The two measurements (H\_dist and KL\_div) were globally consistent, as shown by the tables above. In the cases when two color distributions were comparable, H\_dist tended to be more sensitive to variations. On the other hand, KL\_div might be a preferable option when two color distributions diverged significantly from one another because it showed greater sensitivity than H\_dist. Nevertheless, this study is a preliminary study, and quantitative analysis in more images needs to be investigated to confirm this finding.

### 6.2.6 . Conclusion

This chapter aimed to find a color normalization method suited for NeuN-stained histological images. Three conventional approaches and one conditional adversarial network-based method named Pix2Pix were examined with the goal of recoloring one image sample from section No. 90 using images from section No. 91 as reference. Overall, Pix2Pix demonstrated better performance both visually and quantitatively, with the color distributions of recolored images being fairly close to the reference. Contrary to expectations, none of the conventional methods appeared to be effective at correcting color variations on NeuN-stained images since they all resulted in an undesired illumination change. In addition, Pix2Pix was trained to learn the color distribution over the whole training dataset. The trained model was able to recolor test images with the color of the training data. In contrast to previous cGAN-based methods, which mapped colors from one specific distribution to another, Pix2Pix did not have constraints on source images. No matter the color of the source images, they were first converted into gray-scale images before being fed into the generator, which then recolored the gray-scale images into the desired color images. There is, however, a limitation of Pix2Pix, as well as most GAN-based normalization approaches. The images can only be recolored using one particular color distribution that the trained model has learned from the training set. A new model must be trained if another reference color is required, which would take a lot of development time and effort. Their application is thus constrained by a lack of adaptability. The following section presents another work of the thesis for color normalization: a more adaptable GAN-based normalization approach that can recolor images using different target colors.

### 6.3 . Adversarial Color Transfer to Study the Effect of Color Variation on Cell Instance Segmentation

The remainder of this chapter comprises the following manuscript:

Wu, Huaqian, Nicolas Souedet, Camille Mabillon, Caroline Jan, Cédric Clouchoux, and Thierry Delzescaux. “Adversarial Stain Transfer to Study the Effect of Color Variation on Cell Instance Segmentation.” In *International Workshop on Medical Optical Imaging and Virtual Microscopy Image Analysis (MOVI)*, pp. 105-114. Springer, Cham, 2022.

#### **Abstract**

Stain color variation in histological images, caused by a variety of factors, is a challenge not only for the visual diagnosis of pathologists but also for cell segmentation algorithms. To eliminate the color variation, many stain normalization approaches have been proposed. However, most were designed for hematoxylin and eosin staining images and performed poorly on immunohistochemical staining images. Current cell segmentation methods systematically apply stain normalization as a preprocessing step, but the impact brought by color variation has not been quantitatively investigated yet. In this paper, we produced five groups of NeuN staining images with different colors. We applied a deep learning image-recoloring method to perform color transfer between histological image groups. Finally, we altered the color of a segmentation set and quantified the impact of color variation on cell segmentation. The results demonstrated the necessity of color normalization prior to subsequent analysis.

**keywords:** Histological images; Microscopy; Stain transfer; Generative Adversarial Network; Cell Segmentation.

#### **6.3.1 . Introduction**

Cell segmentation is the first step of many biological applications in preclinical research. For example, neuron instance segmentation is the prerequisite for quantitative studies of neuron population, morphology and distribution to investigate brain aging and neurodegenerative diseases. Advances in Whole Slide Imaging (WSI) techniques allow scanning entire tissue sections at the cellular level, while processing such a massive amount of data is still challenging. To reduce the manual workload, many automatic segmentation approaches have been proposed. In particular, Deep Learning (DL) based methods have demonstrated higher accuracy and robustness than traditional segmentation methods [152, 78]. Although DL methods are considered more robust, the inconsistency of stain color in the histological images exists as a critical issue. This problem is related to many factors in the staining procedure, such as dye type, solution concentration, staining duration, and room temperature. Even

if the staining procedure is strictly standardized, variation may still occur during the digitization. The DL models often underperform when applied to images with colors different from the training dataset [130]. To this end, most DL-based segmentation methods systematically applied color normalization as a preprocessing step [23, 69]. In contrast, the impact of the color variation on the segmentation is still not fully understood and inadequately quantified.

Many stain normalization methods have been proposed, which can be divided into two categories: conventional and deep learning-based. Histogram matching [21] maps the histogram of the source image to that of the target image, treating color distribution independently of image content. The color transfer method [102] converts the image into  $l\alpha\beta$  color space and matches mean and standard deviation of histograms. However, they perform badly when the color of the source and target images differ significantly. The fringing method [82] and the structure-preserving color normalization [132] normalize the stain vectors in the optical density (OD) space separately for each staining channel. They were developed specifically for color variation in hematoxylin-eosin (H&E) staining histological images. Neurons are often stained with NeuN, which is a biomarker evidenced by using immunohistochemistry (IHC). Color normalization in IHC-stained histological images is a less explored research area. Recently, several DL-based approaches have been proposed. These approaches generally address the color normalization as a style transfer problem, using Generative Adversarial Networks (GANs) [114, 9] to learn the color distribution of the reference stain and apply it on the source image. Nevertheless, the trained model can only deal with the specific stain of the training set and it is not suitable for images with multiple stains. GANs were often used as “black boxes”, in which we had no explicit control over the image details until the invention of StyleGAN [64]. By feeding the generator gradually with latent style vectors, StyleGAN can control the style and appearance of the generated image. Furthermore, they show that the color-related features are mainly determined in the fine layers (superior layers). Inspired by StyleGAN, Afifi *et al.* proposed HistoGAN [2], which controls the color of GAN generated images by feeding the desired color histogram in the last two blocks of the generator. In particular, a variant of HistoGAN, named ReHistoGAN, is designed to generate a realistic image with the color of the target image and the content of the input source image.

In this paper, we created a dataset dedicated to exploring color transfer, containing five histological image groups of a mouse brain with controlled color variations. For the first time, we applied ReHistoGAN to recolor histological images while keeping the image content. The trained ReHistoGAN is not limited to a single stain target, which was the main drawback of previous GAN-based normalization methods. We recolored a cell segmentation test



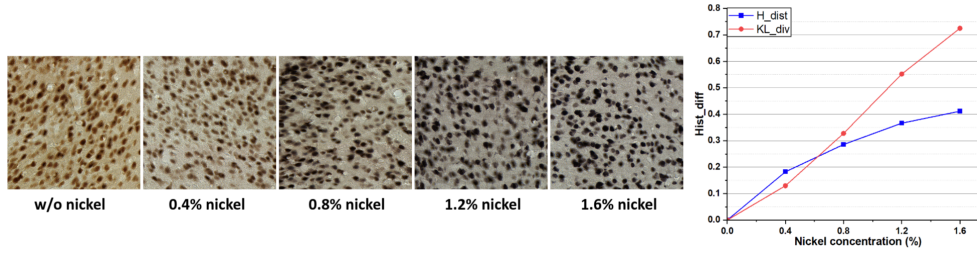


Figure 6.4: Neuron microscopic images with five nickel concentrations. Left: image examples of different nickel groups. Right: correlation of histogram difference and nickel concentration (with the nickel-free group as reference).

set of a macaque brain with the trained ReHistoGAN, using images of different colors as references, including five from the mouse set and one from the macaque set as a control group. To our knowledge, this is the first study to quantify the effect of color change on cell segmentation of the DL model. The results revealed that the segmentation degradation of the DL model was linked to the color difference between the training and test sets. The segmentation performance in groups with significant color variation was considerably reduced. On the contrary, the control group remains at the same level as the initial group. Thus, color normalization prior to inference is mandatory for robust and accurate segmentation.

### 6.3.2 . Material and Method

#### 6.3.2.1 . Dataset

The data used for the color transfer study are mouse neuron microscopic images. Histological sections of a mouse brain were stained using IHC: sections were incubated with NeuN antibody, then treated with Diaminobenzidine (DAB) chromogen and nickel (a DAB enhancer) to develop color. DAB staining is usually brown, and the addition of nickel helps the staining become darker and easier to recognize. The DAB amount stayed constant throughout the staining process, and the only variable was the nickel concentration. This procedure yielded five section groups with different color representations, one of which was nickel-free and the other four were enhanced with nickel concentrations of 0.4%, 0.8%, 1.2% and 1.6%, respectively. The nickel concentration in this study is estimated by the proportion of nickel solution to DAB solution. Figure 6.4 left illustrates image examples from the five groups. The higher the concentration of nickel, the darker the color of neurons. We manually extracted 10 images ( $3000 \times 3000$  pixels) of the cortex from each section group to build our dataset, including a training set of 13k  $256 \times 256$  images and a test set of five  $512 \times 512$  images, which represented five color groups.

The cell segmentation dataset and a trained DL segmentation model were

retrieved from [150]. Microscopic images were extracted from a macaque brain section stained with NeuN and nickel concentration of 1.6%, knowing that the distribution and staining of neurons in mouse and macaque cortex are similar. This dataset was independent of the mouse color transfer dataset. In this study, we used only a subset of cortex, containing 36  $1024 \times 1024$  images, on which we altered the stain color to examine the effect on segmentation.

### 6.3.2.2 . Histogram feature

We converted images from RGB to log-chroma space and calculated the histograms in this space. The intensity of each channel was normalized by the other two channels to make the histogram differentiable and insensitive to illumination changes. For example, the conversion of the R channel is defined as [2]:

$$I_{uR} = \log\left(\frac{I_R + \epsilon}{I_G + \epsilon}\right), I_{vR} = \log\left(\frac{I_R + \epsilon}{I_B + \epsilon}\right) \quad (6.8)$$

where  $I_K$  is the intensity of the  $K$  channel of image  $I$ ,  $\epsilon$  is added for numeric stability, and the  $u$  and  $v$  are the  $K$  channel normalized by the other two channels, respectively. The  $G$  and  $B$  channels are also projected into log-chroma space to compute  $I_{uG}$ ,  $I_{vG}$ ,  $I_{uB}$ ,  $I_{vB}$ . Binning is performed in  $u$  and  $v$  dimensions for each channel, resulting in a histogram with the shape of  $n \times n \times 3$ , where  $n$  equals 64, is the number of bins.

The difference between the two histograms was measured using Hellinger distance (H\_dist) and Kullback-Leibler divergence (KL\_div) [2]. They are usually used to quantify the similarity between two probability distributions. The more similar the distributions are, the closer the value is to 0. Previous studies suggested that H\_dist had a better sensibility to minor deviations, whereas KL\_div was more responsive to large deviations [2, 89]. In this study, they were used to assess the color variation in different nickel groups and compare the color transfer performance of different normalization methods. In particular, the H\_dist was also employed to train the neural network as the color matching loss.

### 6.3.2.3 . ReHistoGAN

The objective of ReHistoGAN was to generate a realistic image, preserving both the content of the source image and the color information of the target image. As shown in Figure 6.5, the network consisted of a generator to recolor the input image and a discriminator. The generator network was modified from a UNet-like structure with skip connections. The target histogram features were projected into latent space and inserted in the last two blocks of the decoder to control the color of the output image. The generator was trained to generate an image having the same content as the source image and a color distribution similar to the target image. It is worth noting that the ReHistoGAN was trained in a fully unsupervised manner since the only ground

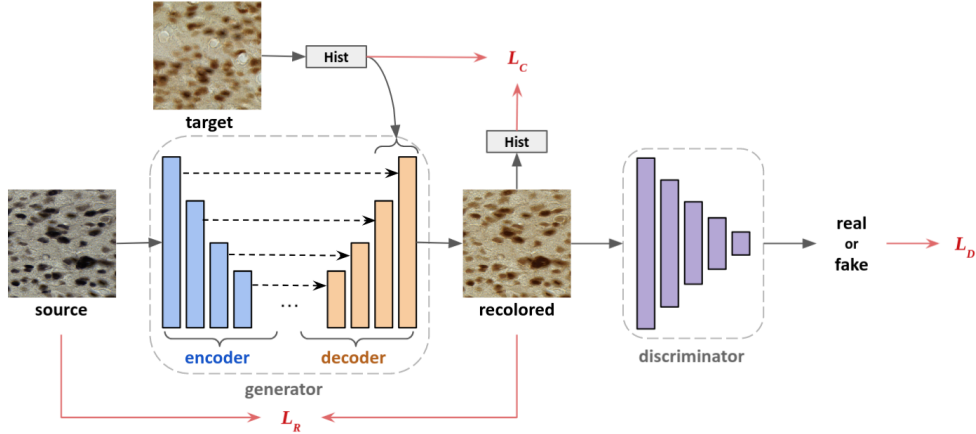


Figure 6.5: ReHistoGAN architecture.

truth needed was the histogram features, which were computed as described in 6.3.2.2. The goal of the discriminator was to distinguish the generated image from the real one. The loss function of the entire network is defined as [2]:

$$L = \alpha L_C(H_t, H_r) + \beta L_R(I_s, I_r) + \gamma L_D \quad (6.9)$$

where  $t$ ,  $s$  and  $r$  are target, source and recolored images respectively.  $H$  is the histogram.  $L_C$  is the color matching loss computed with Hellinger distance.  $I$  is perceptual detail (without color information), which was obtained with the Laplacian operator.  $L_R$  is the reconstruction loss computed with the L1 norm. The discriminator loss  $L_D$  is used to measure how realistic the generated image is.  $\alpha$ ,  $\beta$  and  $\gamma$  are hyperparameters equal to 32, 1.5 and 4, respectively, which were defined empirically in [2] to control the weight of each term.

The ReHistoGAN model was then applied to recolor each image of the cell segmentation dataset using itself and five images in the mouse color transfer test set as targets. This procedure yielded six supplementary groups with continuous color changes. The first was a control group where no color changes were expected (same source image as target image), which was added to measure possible artifacts brought by the GAN network during encoding and decoding. The remaining five groups inherited the color distributions from the color transfer dataset, which were different from the segmentation set. Six colors on the same set allowed us to study the effect of color variation on cell segmentation.

#### 6.3.2.4 . Cell segmentation

We utilized the segmentation method proposed in [150], which consisted of a neural network to predict the cell, the inter-cell contour and the background, followed by a post-processing scheme based on mathematical morphology to individualize each cell instance. Color augmentation techniques were applied during the training to improve the robustness of the model to color inconsistencies. After training, we applied the network on the segmentation test set as well as the six GAN-recolored sets to study the impact of color variation.

#### 6.3.2.5 . Evaluation metrics

In this study, the color variation between images was assessed by the difference between color histograms (H\_dist and KL\_div). A good color transfer method should provide a recolored image with low H\_dist and KL\_div matching the target.

The cell segmentation was evaluated using three metrics: F1 score, Aggregated Jaccard Index (AJI) and Dice [150]. F1 score evaluated the cell instance segmentation, a segmented cell was considered a true positive when it overlapped the ground truth with an intersection over union greater than 0.5. AJI was a stricter metric to evaluate the instance segmentation since it penalized false segmentations. We also computed the Dice score to evaluate the semantic segmentation.

### 6.3.3 . Results and Discussion

We measured the color histogram similarity between the nickel-free group versus other nickel groups using H\_dist and KL\_div. As shown in Figure 6.4 right, color variation becomes more significant as nickel concentration gets higher. In the tested range, the KL\_div curve was roughly linear, whereas the H\_dist saturated as nickel concentration increased.

#### 6.3.3.1 . Color transfer performance of ReHistoGAN and other methods

We compared ReHistoGAN with four state-of-the-art color transfer methods: Macenko *et al.* [82], Reinhard *et al.* [102], Vahadane *et al.* [132] and histogram matching [21]. Each image in the mouse test set was recolored using all color groups as color references, resulting in 25 combinations in total. Figure 6.6 shows the visual comparison of two extreme cases from the nickel-free group (brown) and the highest-nickel group (dark). Macenko *et al.* [82] changed the stain color while it introduced artifacts, especially when recoloring the dark stain to brown. Reinhard *et al.* [102] was unable to retain all color information of the target image, with the stain color in the results still affected by the source images. Vahadane *et al.* [132] achieved better results visually, however both Macenko *et al.* [82] and Vahadane *et al.* [132] suffered intensity issues, the recolored images were unnaturally too bright compared to the source and

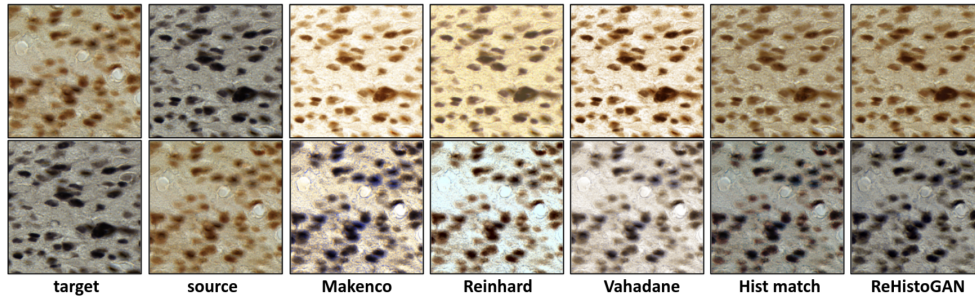


Figure 6.6: Stain color transfer results from different methods. The first row presents recolored results of a 1.6% nickel image using a nickel-free image as reference. The second row shows recolored results after switching the source and target.

target images. The possible explanation for this might be that the IHC stain is a scatterer of light rather than an absorber. Therefore the Beer-Lambert law of absorption used for color space conversion is no longer applicable [132]. The dark stain was successfully recolored into brown using the histogram matching [21]. On the other hand, it did not perform well in converting brown stains to dark (brown spots still exist, see Figure 6.6), and the resulting stain colors were absent from both the source and target images. Compared to other methods, ReHistoGAN showed better performance and robustness, and it worked well in all cases, transferring the target color without losing the source content.

Table 6.3 reports the quantitative comparison of ReHistoGAN and other methods on mouse data.  $H\_dist$  and  $KL\_div$  of color histograms were measured between the recolored images and their target images. Histogram matching had the best average  $H\_dist$  and  $KL\_div$  since the objective was to produce similar color histograms. However, the high standard deviation suggests that this method lacks robustness. This finding is consistent with visual results in Figure 6.6, where the method performed poorly when the source and target images had very dissimilar statistics. On the other hand, besides comparable quantitative results as histogram matching (same  $KL\_div$  and slightly higher  $H\_dist$ ), ReHistoGAN had the lowest standard deviations, which indicates its robustness. Despite bringing illumination changes, Macenko *et al.* [82], Reinhard *et al.* [102] and Vahadane *et al.* [132] were unable to correctly adjust color variation in IHC-stained images. Compared to Histogram matching [21] and ReHistoGAN, these methods resulted in recolored images with more significant color differences from the target images.

Table 6.3: Comparison of color transfer methods. The H\_dist and KL\_div are reported to show the color difference between recolored results and target images.

Metrics	Macenko	Reinhard	Vahadane	Hist match	ReHistoGAN
H_dist	0.239± 0.038	0.284± 0.223	0.383 ± 0.131	0.043± 0.031	<b>0.052± 0.006</b>
KL_div	0.258± 0.076	0.42± 0.458	0.557 ± 0.284	0.011± 0.012	<b>0.011± 0.002</b>

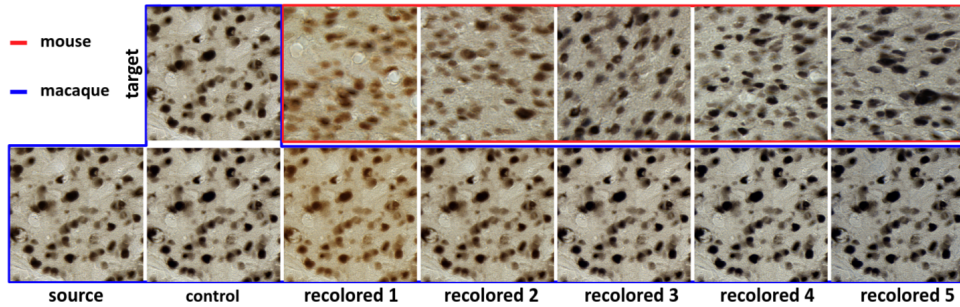


Figure 6.7: Macaque segmentation dataset recolored by ReHistoGAN. First row: target images, second row: source macaque image and recolored results. From left to right: GAN-recolored images using the macaque image (control), nickel-free (recolored-1), 0.4%, 0.8%, 1.2% and 1.6% nickel mouse images (recolored-2-5) as targets.

### 6.3.3.2 . Impact of color variation on cell segmentation

We applied ReHistoGAN to recolor images of the macaque segmentation test set, as illustrated in Figure 6.7. The second column is the control group recolored using the same macaque image as the target, which allowed us to investigate the possible effect of encoding and decoding. The following columns show recolored images using five nickel mouse groups as the target, respectively. In total, six additional sets were produced, including one set without color change (control) and five with continuous color variations from brown to dark (recolored 1-5). Surprisingly, experts estimated that the original macaque images were closest in color to recolored-3 (0.8%), although the nickel concentration used for macaque images was the same as recolor-5 (1.6%). This inconsistency may be due to factors other than nickel concentration during the staining process.

We evaluated the segmentation network on the original and six additional sets of macaque images. Table 6.4 reports the segmentation results and the color variation between the test sets and the training set. Despite the fact that the model was trained using color augmentation techniques, segmentation performance has been negatively impacted by color variations. The scores of the original set were the best since they came from the same data as the training

### 6.3. COLOR TRANSFER WITH REHISTOGAN

Table 6.4: Comparison of segmentation performances (F1, AJI and Dice) on segmentation dataset without and with color variations inherited from the different groups in the color transfer dataset. Color variations between the two datasets are quantified using H\_dist and KL\_div.

Metrics	original	control	recolored-1	recolored-2	recolored-3	recolored-4	recolored-5
H_dist	-	-	0.257	0.093	<b>0.05</b>	0.132	0.182
KL_div	-	-	0.27	0.035	<b>0.01</b>	0.075	0.135
F1	0.91	0.899	0.842	0.892	<b>0.898</b>	0.893	0.867
AJI	0.754	0.734	0.628	0.711	<b>0.733</b>	0.729	0.7
Dice	0.972	0.961	0.909	0.951	<b>0.962</b>	0.956	0.936

set. They were used as references to show the performance degradation due to GAN artifacts and color variation. The scores of the control group were comparable to the reference but slightly lower, indicating that almost all information of the original image was preserved during encoding and decoding. The group with the most degradation in segmentation was the recolored-1 set, which had the most significant color variation compared to the original set (KL\_div: 0.27). In particular, the AJI of this set decreased by 17% when compared to the reference. The recolored-3 set, on the other hand, scored similarly to the control group as it had similar color distributions both visually and quantitatively (KL\_div: 0.01). It suggests that the DL segmentation model is robust to slight color variation. Overall, the findings suggest that the segmentation performance decreases more in test sets with higher color variation. Surprisingly, the segmentation in the recolored-4 set was better than that in recolored-2, even though the former had a more visually distinct color variance. This might be due to the fact that despite the color variation, the contrast between cell and tissue is also essential for neural networks to segment cells. In our case, the darker stained images in the recolored-4 set presented a more significant contrast than lighter stained images in the recolored-2 set, providing a better condition for cell segmentation.

#### 6.3.4 . Conclusion and Perspectives

In this paper, we created an original dataset with five controlled color changes, which is well-suited for evaluating the performance of color transfer methods. We applied the ReHistoGAN method to perform the stain color transfer on IHC staining histological images. The results demonstrated its superiority and robustness compared with other state-of-the-art methods. Using ReHistoGAN, we intentionally altered the color of the macaque segmentation test set, which enabled us for the first time to quantitatively investigate the impact of color variation on IHC-stained cell segmentation (semantic and instance) using neural networks. Experiments showed that the cell segmentation negatively correlated to the color variation between the test and the training sets.

However, for a given color variation, the segmentation was better on images with darker stains. (see recolored-2 and recolored-4 in Table 6.4). Previous GAN-based approaches [9, 114] have not been assessed due to their lack of flexibility. Further studies are needed to validate the color transfer performance of ReHistoGAN compared to these methods. As a preliminary study, we only investigated NeuN staining images in the cortex region. Further work is needed to establish a strategy for the automatic selection of appropriate target images in various anatomical regions, which will allow us to expand this study to the entire brain. Moreover, additional study on images with other stainings (*e.g.*, H&E) would also be worthwhile.

### **Acknowledgements.**

This work was supported by DIM ELICIT grants from Région Ile-de-France, by the French National Research Agency (project SUMMIT ANR-21-CE45-0022-01) and by the European Union’s Horizon 2020 research and innovation program under the grant agreement No. 945539 (Human Brain Project SGA3).



## 7 - Discussion and Conclusion

### 7.1 . Contributions of the thesis

The objective of this thesis was to develop an automatic method for neuron instance segmentation in various ROIs in histological sections of brains. This work could be split into two parts: (1) **a robust neuron instance segmentation framework based on neural networks and morphological post-processing**, and (2) **GAN-based color transfer approaches to eliminate color variations between histological sections**.

In Chapter 1, we first presented a general introduction to the biological background of the central nervous system, the anatomical regions of the brain, the neurons, and Alzheimer’s disease. Then, we described the histology technique and the NeuN staining, which served as the foundation for the majority of the study. A number of traditional machine learning techniques and deep neural networks applied in this study were also reported. Chapter 2 reviewed the state-of-the-art cell instance segmentation and color normalization approaches for histological images, including conventional and Deep Learning-based methods. The main contributions were introduced in the following chapters, from Chapter 3 to Chapter 6.

The main topic of the first part of the manuscript (from Chapter 3 to Chapter 5) consisted of the development of neuron instance segmentation for macaque brain sections based on DL techniques. In Chapter 3, based on manual point annotations and a RF model trained for segmenting neuron class from the tissue, we proposed a mask synthesis pipeline to generate pixel-level annotations automatically in order to reduce the manual labeling workload. This strategy allowed us to annotate a large dataset of macaque neurons containing 11k patches of  $224 \times 224$  pixels (about 100,000 neurons in total). Based on this dataset, we assessed the winning method of the nuclei segmentation challenge Data Bowm Science 2018 [16], Topcoders, an ensemble model of eight UNet-like neural networks and a post-processing scheme to remove incorrect segmentations. This method demonstrated the excellent performance of UNet-like architecture on cell segmentation tasks. However, it is challenging to implement this strategy in practice because it requires inferring eight neural networks and applying post-processing iteratively to each predicted object.

Chapter 4 further explored the capacity of this architecture: we compared in total 13 segmentation approaches, 11 of which were UNet-like encoder-decoder networks, including the UNet, eight constituent models, and the en-

semble model from Chapter 3, and EfficientUNet, a UNet-like network with EfficientNet as the encoder. We also evaluate HRNet, which does not follow the encoder-decoder design and maintains high resolution throughout the convolutional layers. A conventional method previously developed in our lab to deal with neuron segmentation was tested as well. Heavy data augmentation was applied during the training to increase the robustness of the neural networks, including rotation, flipping, random cropping, RGB channel shuffling, color inversion, *etc.* Various aspects were considered for the evaluation, such as the accuracy in instance segmentation, semantic segmentation, detection, and counting, as well as the model weight and computational complexity for DL models. After a thorough comparison, EfficientUNet was identified as the model with the best trade-off between latency and accuracy: it was the second-best model for segmentation with model size and FLOPs three times lighter than the best model. Thus, subsequent developments were mainly based on this model. The proposed approach integrating the ultimate erosion and a dynamic dilation, as well as four other widely used post-processing techniques from the literature, were all applied to the probability maps of EfficientUNet. The proposed approach produced the best segmentation results. The proposed method was also evaluated in a pilot study with a supplementary cortex dataset from other animal brains, including two other macaques, a microcebus, and a mouse. The proposed framework remained effective despite the difference between samples and species, indicating the robustness and generalizability of the proposed framework. A performance decrease is observed in images with color variations and a light stain intensity. It implies that stain color and intensity normalization can be an effective tool to further improve the robustness of the framework and to extend the application of the framework to other animal samples and other cell types. Furthermore, the combination of EfficientUNet and the proposed post-processing allowed us to segment neurons accurately and efficiently, enabling us to envision the processing of larger-scale images. An extension of this work, an application for segmenting whole-slide images, has been subsequently carried out.

In Chapter 5, we have described our parallel mechanism that combines the CPU and GPU hardware resources to speed up the inference over whole-slide images. Thanks to the partial reading and writing library that was previously developed in our lab, our pipeline could load a part of an image in memory instead of the whole image content. Given an input image, each CPU-GPU pair was distributed with one image band. Up to 3 GPUs were successfully tested simultaneously, and this protocol could be extended to a larger number of GPUs. The GPU processed the patch content that was read by the CPU through a sliding window. An overlap between neighbor patches was considered to take into account possible border effects. Contrary to previous methods [23, 56], which require an additional reweighting process to normalize

the contribution of each pixel in the assembled prediction, the proposed novel weight map simplifies the process by directly computing the contribution of the processed patch in the final image. Throughout the entire process, only the final prediction was written to the disk. This work was integrated into BrainVISA software as a program with a user-friendly interface, making it easier to be used by biologists. The prediction result of the inference pipeline is a probability map. By combining our post-processing scheme described in Chapter 4, we successfully individualized neurons in an entire mouse brain section. The information about neuron number, distribution, and morphology in various anatomical regions, which can be potentially extracted from segmented sections, could be particularly beneficial for the neuroscience community to study cytoarchitecture in control brain or to quantify cell death in neurodegenerative animal models (decipher pathological mechanisms or test drug efficacy).

Before applying the segmentation pipeline to the whole brain, one issue that needs to be addressed is the correction of color inconsistency between sections. The second part of the manuscript (Chapter 6) attempted to correct the color variation with two approaches based on conditional GANs. Most widely used color normalization methods were developed specifically for H&E-stained images, but they were not well suited for IHC images. The first section of Chapter 6 tackled color normalization as a recoloring task of gray-scale images [107]. The training data were extracted from section 91 of the macaque brain, and the trained model was evaluated on images from section 90, where the neurons are darker than the training data. A further benefit is that, unlike earlier GAN-based methods [114, 9], which were trained to transfer one specific color distribution to another, the trained generator was able to apply the color of the training dataset to test images regardless of their original colors. More importantly, because the generator learns color characteristics from the entire training dataset, there is no need to select a target image as the color reference, a task that is frequently carried out manually by an expert for conventional approaches. On the other hand, a trained Pix2Pix network can only recolor test images with one specific target color, which is that of the training data. Extra training is necessary for other target colors.

The second section of Chapter 6 presented another strategy for color transfer between images. A variant of StyleGAN [64] called ReHistoGAN [2] was employed to recolor the input image using a target color histogram injected in the generator as the reference. Changing image color by modifying the color histogram is not novel in the literature. Traditionally, histogram specification [21, 59] alters the color histogram independently from the image content, which is, therefore, suboptimal for normalizing colors in images with different scenes. This issue was addressed in ReHistoGAN by training the generator

to apply the color target histogram while preserving the source image content. ReHistoGAN has the advantage of being flexible. In contrast to earlier GAN-based methods, which were restricted to one specific color of the training set, ReHistoGAN can be used to normalize images to various target colors, even colors that are different from the training data. Such design may be of great practical value for saving development time and effort. The evaluation was performed on an original dataset with five continuous controlled color variations produced from a mouse brain. The accuracy of this method was demonstrated with a detailed quantitative assessment of a total of 25 pairs of source and target combinations. Furthermore, the trained generator allowed us to investigate the effects of color variations on neuron instance segmentation by recoloring the test set of Chapter 4. This study demonstrated that the segmentation performance was directly related to the color difference between test and training images, which suggests the necessity of color normalization prior to the segmentation inference.

## 7.2 . Limitations of the thesis

Despite how well the results of the proposed segmentation framework aligned with the ground truth, the synthetic annotations were limited by the binary segmentation of the RF model, which was not perfect. Manual annotations are scarce and extremely time-consuming. It is, therefore, not feasible to annotate neuron instances manually in all 11k images of the training set. One possible way to improve the quality of synthetic annotations will be to enlarge the original binary segmentation dataset, which will allow employing a neural network rather than the RF model to further improve the segmentation accuracy and robustness.

The proposed post-processing scheme has one hypothesis, which is that most neurons are disk-shaped. The separation of touching and overlapping neurons by ultimate erosion relies on the concave point that exists exclusively between neurons. Its use is particularly adapted as the DL method has efficiently initiated full or partial cell splitting. For cells with irregular or branched shapes, such as astrocytes and microglia, which may have concave points within a single cell, our method could be adapted by extending the number of classes by including a branch class. In addition to extra training, it is critical to carefully specify the size of the structuring element for erosion to prevent cells from being divided into segments. Nonetheless, most parts of our pipeline could be easily reused.

Although the quantitative evaluation of the proposed segmentation framework in the cortex region of other species produced promising results, rigorous validation in other regions, especially in the hippocampus, is required before being applied to biological studies.

The post-processing part of the segmentation of the whole-slide image presents a challenge that has not yet been fully solved. Currently, only the inference phase has been parallelized, allowing the processing of a whole macaque brain section (118 GB). Post-processing still relies on a single CPU to load and process the entire image in memory. However, the post-processing of images of this size cannot be realized in a local workstation without partial reading and writing mechanisms. For post-processing, a different extraction and assembling technique is required. Without overlapping, a neuron that crosses multiple patches will be treated as separate small neurons, not as a single large neuron. Extra developments for processing neurons at the border are ongoing to further parallelize the entire process.

As for the color normalization part, it is essential that the content of the input image remains intact. However, due to the lack of color-invariant structural similarity measurements, the validation of this aspect relied on the visual assessment and the indirect evaluation by a downstream instance segmentation

task. Direct measurements of the structural change due to encoding and decoding need to be conducted, which requires examining color normalization approaches with datasets where the source and target images share the same image content, such as images digitized with different scanner devices. In addition, the target image for ReHistoGAN depends on manual identification, and the color histogram of an image from a sparse region may differ significantly from that of an image from a dense region. This constraint limits its application on a large scale, for example, an entire section, where it is mandatory to manually specify a target image for each possible distribution of neurons.

### 7.3 . Future directions

Notwithstanding these limitations, this study offers valuable insights into neuron distribution in the brain. Although this thesis work focused mainly on healthy animal samples, the proposed approach can be easily applied to pre-clinical data with pathological cases. In practice, in order to observe neuron changes under different conditions, biologists still rely on visually assessing MRI images or performing stereology in a region of interest to observe neuron changes in various situations. The former limits the observation at the macroscopic scale, and it only reveals the brain atrophy when it becomes significant. The latter provides details at the microscopic scale, but it is extremely laborious and time-consuming. Therefore, applying the proposed segmentation framework to pathological sections will be of great interest to precisely track the effect of novel therapies on specific anatomical regions and entire brain sections, which will be particularly beneficial to the neuroscience community. Moreover, assessing neuron morphology and distribution at a large scale could enable us to detect slight early changes due to pathologies that are extremely difficult to identify.

It is exciting to see how well the evaluation of the proposed segmentation framework is on primate (macaque and microcebus) and rodent (mouse) animal models. A transition of our approach to human brain samples provided through donation programs may further be envisioned. The ANR-funded SUMMIT (Small vessel diseases: Ultrastructure & Microvasculature computational Model to refine Individual Treatment) research project began in early 2022 to study human brain samples using advanced high-field MRI techniques and histology. This project is led by the CEA Neurospin imaging center, and the samples are collected and made available by the laboratory of anatomy of Tours. The objective of this project is to decode the cytoarchitecture of the human brain by MRI. The creation of this decoder is based on the MEDUSA cytoarchitecture simulator [37] developed at Neurospin, which allows the production of the corresponding theoretical MRI data with HPC. On the other hand, the production and analysis of histological samples (multiple stains produced from real human brain samples) will be performed at MIRCen and will provide gold standard information to interpret the MRI images that will have been produced on these real samples but also realistic quantified cell features to adjust the simulator parameters. For this, the method developed during this thesis will be exploited and extended to the different cellular elements of the brain.

An unbiased neuron counting across all sections of the brain is not possible because we are not sure to see all the neurons present in the slice under the focus of the microscope. The comparison between individuals is nevertheless possible and relies on relative count since we introduce the same measurement

bias. Compared to the stereological approach, our method allows us to process data massively and provide additional spatial information. The current work was conducted on two-dimensional (2D) microscopic images produced with histology, which resulted in the loss of the three-dimensional (3D) coherence of the brain. Several studies have attempted to restore this coherence through the 3D reconstruction of 2D sections. Malandain *et al.* [84] reconstructed a 3D volume with consistent geometry and intensity from 2D autoradiographic sections. The fusion of the 3D volume and MRI images ensured that the reconstructed volume could be superimposed on the reference anatomy. Dauguet *et al.* [24] suggested aligning histological sections of the baboon brain with the anatomical MRI images using an intermediate 3D consistent volume of “blockface” photographs generated during sectioning. A “semi-rigid” transformation was applied to each 2D section, and a 3D elastic transformation was subsequently applied to complete the registration of histological sections with the MRI data. The 3D reconstruction of 2D segmented sections is necessary for an accurate estimation. Recently, several 3D reconstruction techniques have been proposed, such as comparing the similarity of segments across 2D images to extract 3D cell shapes [122], continuous boundary tracing criterion to reconstruct 3D object surface [106] and graph-theoretic network-flow based association of 2D objects along the axial direction [110], *etc.* These works may give us insights to expand this study into samples or whole brains in 3D, which will deepen our understanding of brain anatomy at the cellular scale and provide a gold standard to evaluate the digital simulation of brain structures.

The color inconsistency between images troubles not only the performance of CAD systems but also the diagnosis of pathologists. ReHistoGAN has superiority in terms of flexibility and accuracy against the conventional approaches as well as Pix2Pix (see Chapter 6). Applying ReHistoGAN as a universal tool for biomedical image analysis will be of significant interest. The major limitation of this approach is that the recoloring performance directly relates to the reference color histogram provided, which is specified by experts. Future work will involve developing an automatic selection of reference images to alleviate manual effort, process large samples and make it more accessible for non-experts users.



## Résumé détaillé

Les neurones sont des cellules clés du système nerveux qui sont responsables de la transmission de l'information sensorielle et de la communication entre les différentes régions du cerveau. Ils sont impliqués dans de nombreuses fonctions importantes du cerveau, telles que la perception, la cognition, l'apprentissage et la mémoire, ainsi que le contrôle des mouvements et des comportements. La quantification du nombre, de la distribution et des caractéristiques des neurones est cruciale pour comprendre les mécanismes impliqués dans le développement du cerveau, le vieillissement et les maladies neurodégénératives. Une technique de référence couramment utilisée pour étudier les neurones est l'immunohistochimie (IHC), qui permet de marquer les neurones avec un anticorps spécifique appelé NeuN (neuronal nuclei) et de les observer à haute résolution grâce à un microscope.

En pratique, les biologistes utilisent des techniques manuelles telles que la stéréologie pour quantifier sans biais le nombre de neurones dans de petites régions d'intérêt, mais cette méthode est chronophage et nécessite une intervention manuelle très fastidieuse. Ainsi, elle n'est pas pratique pour mener des études à grande échelle. Pour réduire l'effort humain, de nombreuses méthodes de segmentation automatique ont été proposées, notamment celles basées sur l'intelligence artificielle et l'apprentissage profond (deep learning -DL-). Ces approches utilisent des algorithmes entraînés sur de grandes bases de données annotées pour identifier les cellules automatiquement dans de nouvelles images. Cependant, l'apprentissage supervisé nécessite de nombreuses annotations de neurones à l'échelle cellulaire. Cette tâche est difficile à réaliser manuellement, surtout pour l'ensemble de données d'entraînement et correspondent à une quantité importante de travail.

Il convient de noter que les approches basées sur DL ont principalement été développées et validées pour les images colorées avec l'hématoxyline et à l'éosine (H&E) en oncologie, tandis que les neurones sont souvent colorés avec le marquage NeuN qui est très différent. De plus, la forme, la taille et la densité des neurones peuvent varier de façon importante d'une région à l'autre du cerveau, ce qui rend leur segmentation précise et robuste difficile dans certaines régions, en particulier le gyrus denté de l'hippocampe qui est une zone particulièrement dense.

En comparaison avec les méthodes conventionnelles, les approches basées sur DL pour la segmentation de cellules sont considérées comme plus précises et robustes. Cependant, la variation de couleur présente dans les images his-

tologiques peut avoir un impact sur la performance des modèles DL de segmentation. Même si les protocoles et les moyens de production des coupes histologiques sont standardisés, il est fréquent d’observer des de couleur entre les coupes histologiques produites. Cette variation peut affecter le modèle de segmentation, car il est entraîné à apprendre des caractéristiques à partir d’une base d’entraînement spécifique qui peut ne pas intégrer toutes les variations de couleur existantes. Si la couleur des images de test est significativement différente de celle des images d’entraînement, le modèle peut ne pas être aussi performant. Pour résoudre ce problème, une solution est de normaliser la couleur des images de test en les modifiant pour qu’elles correspondent à la couleur des images du jeu d’entraînement.

La variation de couleur peut également poser problème pour les pathologistes qui réalisent leur diagnostic visuellement à partir des tissus en se basant sur les couleurs des cellules et d’autres caractéristiques des images. Pour y remédier, plusieurs approches de normalisation de couleurs ont été développées, incluant des méthodes conventionnelles et des approches basées sur DL. Les méthodes conventionnelles impliquent souvent la conversion des images dans d’autres espaces de couleurs afin de faire correspondre les statistiques des couleurs de l’image de source à celles de l’image de référence. Cependant, la plupart de ces approches ont été conçues pour corriger la couleur des images colorées avec H&E. Le transfert de ces approches aux images colorées par IHC, et spécifiquement aux images colorées par le NeuN, reste une question ouverte.

Dans ce manuscrit de thèse, nous présentons une méthode basée sur DL pour segmenter automatiquement les neurones dans différentes régions anatomiques de modèles animaux utilisés en recherche préclinique. Nous avons étudié les performances de la méthode proposée en termes de précision, de robustesse et de coût de calcul, nous avons évalué notre capacité à étendre l’analyse à des images de grandes dimensions et à plus grande échelle. Nous avons utilisé la normalisation de couleur comme une étape préliminaire pour améliorer la robustesse de la méthode face aux variations de couleur qui peuvent être présentes dans les images à traiter.

Les travaux de cette thèse ont été réalisés au Centre de Recherche en Imagerie Moléculaire (Molecular Imaging Research Center -MIRCen-) dans le laboratoire des maladies neurodégénératives (UMR9199 LMN), une unité de recherche du Commissariat à l’Energie Atomique et aux Energies Alternatives (CEA) en collaboration avec le Centre National de la Recherche Scientifique (CNRS) et l’Université Paris-Saclay. Toutes les données utilisées dans cette étude proviennent d’images cérébrales microscopiques réalisées au sein de la plateforme d’histologie de MIRCen.

Ce manuscrit est subdivisé en sept parties.

Dans le premier chapitre de cette étude, nous décrivons le contexte biologique dans lequel notre recherche s'inscrit. Nous fournissons une description détaillée de la structure et de l'organisation de base du cerveau, les caractéristiques des neurones ainsi que les mécanismes impliqués dans les maladies neurodégénératives comme la maladie d'Alzheimer. Nous abordons également l'utilisation des modèles animaux pour la recherche sur la neurodégénérescence et décrivons brièvement les techniques d'histologie et de numérisation des images.

Avec l'émergence de technologies avancées permettant de produire de grandes quantités de données à haute résolution (scanner de lames virtuelles entières, whole slide imaging -WSI-), il y a un besoin grandissant de solutions logicielles capables d'analyser efficacement ce type de données massives. Pour ce faire, les algorithmes d'apprentissage automatique (machine learning -ML-), en particulier les réseaux de neurones profonds, sont considérés comme l'un des outils les plus efficaces pour l'analyse automatique de grandes quantités de données. Nous décrivons également dans notre manuscrit les différents algorithmes ML de référence utilisés dans cette étude.

Le chapitre 2 de ce manuscrit décrit les différentes approches développées pour la segmentation des cellules individuelles, comprenant les méthodes conventionnelles et les méthodes basées sur DL. Ces méthodes sont classifiées en fonction de leurs principes de fonctionnement et de la nature de l'annotation des données. Nous explorons également les stratégies pour rendre les approches basées sur DL évolutives et capables de traiter des images de coupes de cerveaux entières, étant donné que ces approches sont limitées par la mémoire de la carte graphique (GPU) par rapport aux méthodes conventionnelles. Le chapitre présente également les approches couramment utilisées (conventionnelles et DL) pour normaliser la couleur des images microscopiques, car les variations de couleur se produisent fréquemment et elles représentent un défi important dans l'analyse d'images.

Le chapitre 3 de cette étude décrit une méthode pour générer des masques de segmentation synthétiques pour entraîner des réseaux de neurones profonds. Les données utilisées dans cette étude proviennent des images de cerveaux de macaques sains produites à MIRCen. Une grande quantité de données annotées est nécessaire pour entraîner les réseaux de neurones. Cependant, l'annotation manuelle au niveau cellulaire est un processus fastidieux. Pour résoudre ce défi, ce chapitre présente une stratégie pour générer des masques synthétiques avec un minimum d'intervention manuelle. Cette stratégie a permis d'annoter un grand nombre de neurones de macaque contenant 11 milles imagerie de  $224 \times 224$  pixels (soit environ 100 000 neurones au total). En utilisant cet ensemble de données annoté, nous avons évalué la méthode gagnante de la compétition de segmentation de cellules Data Bowm Science 2018, Topcoders, un modèle d'ensemble de huit réseaux neuronaux de type U-Net en

utilisant son schéma de post-traitement pour raffiner les segmentations. Les résultats de cette étude ont été publiés sous la forme d'un article de conférence internationale avec acte :

Wu, Huaqian, Nicolas Souedet, Zhenzhen You, Caroline Jan, Cédric Clouchoux, and Thierry Delzescaux. "Evaluation of Deep Learning Topcoders Method for Neuron Individualization in Histological Macaque Brain Section." In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 2985-2988. IEEE, 2021.

Bien que la méthode présentée dans le chapitre 3 ait montré des résultats prometteurs pour la segmentation des neurones, son application à grande échelle est difficile en raison de sa complexité de mise en œuvre nécessitant l'inférence de huit réseaux de neurones et une phase de post-traitement qui prédit la meilleure segmentation pour chaque cellule. Ces contraintes rendent cette approche inadaptée pour le traitement d'images de grande taille. En conséquence, le chapitre 4 vise à surmonter ces limitations en proposant un nouveau cadre efficace pour la segmentation de neurones à grande échelle. En utilisant la base de données annotées produite dans le chapitre 3, nous avons comparé les principaux modèles de réseaux de neurones pour identifier un modèle (EfficientUNet) qui présente le meilleur compromis en termes de précision et de coût de calcul. Nous présentons également une méthode de post-traitement originale basée sur l'érosion ultime et la reconstruction dynamique pour finaliser l'individualisation des neurones qui se touchent ou se chevauchent à partir des cartes de probabilités produites. Les résultats de cette étude ont été publiés sous la forme d'un article dans une revue internationale :

Wu, Huaqian, Nicolas Souedet, Caroline Jan, Cédric Clouchoux, and Thierry Delzescaux. "A general deep learning framework for neuron instance segmentation based on efficient UNet and morphological post-processing." *Computers in Biology and Medicine (2022)*: 106180.

Les images de lames entières (WSI) sont des images de très grande taille qui contiennent une numérisation à très haute résolution (x20, 0.22  $\mu\text{m}$  de résolution spatiale) d'une section de tissu ou d'une biopsie entière. Ces images peuvent atteindre des tailles de plusieurs gigapixels, ce qui fait de la segmentation de ces images un défi. Le chapitre 5 de ce manuscrit de thèse présente une approche pour implémenter la méthode de segmentation de neurones proposée sur des données WSI. Afin de rendre possible et d'accélérer la vitesse de l'inférence, une stratégie de calcul parallèle est mise en place en utilisant des ressources computationnelles hybrides multi-CPU-GPU avec une bibliothèque pour la lecture et l'écriture partielles d'images. Nous avons conçu une chaîne de traitement pour la segmentation de grandes images qui a été intégrée au

logiciel BrainVISA avec une interface graphique pour en faciliter l'utilisation. Enfin, nous avons validé ce pipeline sur des coupes entières de cerveau de souris ( 5 Go) et de macaque ( 120 Go).

Dans le chapitre 6 de ce manuscrit, nous présentons deux approches pour corriger les variations de couleur dans les images microscopiques de cerveaux en utilisant des réseaux adversaires génératifs (Generative Adversarial Network -GAN-). Ces méthodes peuvent être utilisées comme une étape de prétraitement dans le cadre de la segmentation d'instances de neurones, en particulier lorsque les images de test présentent des variations de couleur par rapport aux données d'entraînement. La première approche, appelée Pix2Pix, utilise un générateur pour recolorer une image en niveau de gris avec la couleur de l'image de référence, cependant un modèle de Pix2Pix entraîné est limité à une seule coloration cible. Une deuxième approche, inspirée de StyleGAN, appelée ReHistoGAN a été proposée. Dans ce cas, des histogrammes de couleurs de référence de plusieurs colorimétries produites artificiellement sur des données par la plateforme d'histologie de MIRCen (cinq niveaux de couleurs de référence différentes modulées par la concentration de nickel utilisée) peuvent être injectés dans le générateur, permettant ainsi une normalisation couleur des données. Enfin, nous avons évalué l'impact des variations de couleur sur la segmentation d'instance de neurones. Les résultats de cette deuxième approche sont présentés sous la forme d'un article de conférence :

Wu, Huaqian, Nicolas Souedet, Camille Mabillon, Caroline Jan, Cédric Clouchoux, and Thierry Delzescaux. "Adversarial Stain Transfer to Study the Effect of Color Variation on Cell Instance Segmentation." In *International Workshop on Medical Optical Imaging and Virtual Microscopy Image Analysis (MOVI)*, pp. 105-114. Springer, Cham, 2022.

Le chapitre 7 du manuscrit présente les principales contributions de cette étude. Il résume les limites de la méthode de segmentation des neurones et les défis persistants dans le traitement des WSI. Il aborde également les questions liées à l'évaluation de la normalisation des couleurs. Ce chapitre présente également les orientations futures et les applications potentielles des approches proposées pour la segmentation des neurones et la normalisation des couleurs.

En conclusion, ce travail de thèse propose dans une première partie un cadre complet pour la segmentation d'instance de neurones dans les images microscopiques ainsi qu'une stratégie de calcul intensif pour traiter des images de très grandes dimensions. Une deuxième partie décrit des approches de transfert de couleur basées sur les GAN pour corriger les variations de couleur pouvant exister entre des coupes histologiques.

Notre étude a montré un grand potentiel sur des données provenant de modèles

animaux primates (macaques et microcèbes) et rongeurs (souris). Bien que ce travail se soit concentré dans un premier temps sur des échantillons d'animaux sains, les méthodes développées peuvent être appliquées à des données précliniques dans des cas pathologiques avec des premières études initiées. Il serait aussi intéressant d'explorer les applications possibles de notre approche sur des échantillons de cerveaux humains recueillis via des programmes de dons de corps pour une analyse plus précise du cerveau humain ainsi que pour mieux interpréter les signaux produits par des systèmes d'imagerie telle que l'imagerie par résonance magnétique (Magnetic Resonance Imaging, -MRI).

## Bibliography

- [1] ABELS, E., PANTANOWITZ, L., AEFNER, F., ZARELLA, M. D., VAN DER LAAK, J., BUI, M. M., VEMURI, V. N., PARWANI, A. V., GIBBS, J., AGOSTO-ARROYO, E., ET AL. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the digital pathology association. *The Journal of pathology* 249, 3 (2019), 286–294.
- [2] AFIFI, M., BRUBAKER, M. A., AND BROWN, M. S. Histogan: Controlling colors of gan-generated and real images via color histograms. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2021), pp. 7941–7950.
- [3] AL-KOFAHI, Y., LASSOUED, W., LEE, W., AND ROYSAM, B. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering* 57, 4 (2009), 841–852.
- [4] ANAND, K. S., AND DHIKAV, V. Hippocampus in health and disease: An overview. *Annals of Indian Academy of Neurology* 15, 4 (2012), 239.
- [5] ANDERSEN, B. B., KORBO, L., AND PAKKENBERG, B. A quantitative study of the human cerebellum with unbiased stereological techniques. *Journal of Comparative Neurology* 326, 4 (1992), 549–560.
- [6] AZEVEDO, F. A., CARVALHO, L. R., GRINBERG, L. T., FARFEL, J. M., FERRETTI, R. E., LEITE, R. E., FILHO, W. J., LENT, R., AND HERCULANO-HOUZEL, S. Equal numbers of neuronal and non-neuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology* 513, 5 (2009), 532–541.
- [7] BADRINARAYANAN, V., KENDALL, A., AND CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 39, 12 (2017), 2481–2495.
- [8] BANKHEAD, P., LOUGHREY, M. B., FERNÁNDEZ, J. A., DOMBROWSKI, Y., MCART, D. G., DUNNE, P. D., MCQUAID, S., GRAY, R. T., MURRAY, L. J., COLEMAN, H. G., ET AL. Qupath: Open source software for digital pathology image analysis. *Scientific reports* 7, 1 (2017), 1–7.

## BIBLIOGRAPHY

---

- [9] BENTAIEB, A., AND HAMARNEH, G. Adversarial stain transfer for histopathology image analysis. *IEEE transactions on medical imaging* 37, 3 (2017), 792–802.
- [10] BIRD, C. M., AND BURGESS, N. The hippocampus and memory: insights from spatial processing. *Nature Reviews Neuroscience* 9, 3 (2008), 182–194.
- [11] BOUVIER, C., SOUDET, N., LEVY, J., JAN, C., YOU, Z., HERARD, A.-S., MERGOIL, G., RODRIGUEZ, B., CLOUCHOUX, C., AND DELZESCAUX, T. Reduced and stable feature sets selection with random forest for neurons segmentation in histological images of macaque brain. *Scientific reports* 11, 1 (2021), 1–17.
- [12] BOYKOV, Y., VEKSLER, O., AND ZABIH, R. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence* 23, 11 (2001), 1222–1239.
- [13] BREIJYEH, Z., AND KARAMAN, R. Comprehensive review on alzheimer’s disease: causes and treatment. *Molecules* 25, 24 (2020), 5789.
- [14] BRONNER, G., SPATARO, B., PAGE, M., GAUTIER, C., AND RECHENMANN, F. Modeling comparative mapping using objects and associations. *Computers & chemistry* 26, 5 (2002), 413–420.
- [15] CAI, H., YANG, Z., CAO, X., XIA, W., AND XU, X. A new iterative triclass thresholding technique in image segmentation. *IEEE transactions on image processing* 23, 3 (2014), 1038–1046.
- [16] CAICEDO, J. C., GOODMAN, A., KARHOHS, K. W., CIMINI, B. A., ACKERMAN, J., HAGHIGHI, M., HENG, C., BECKER, T., DOAN, M., MCQUIN, C., ET AL. Nucleus segmentation across imaging experiments: the 2018 data science bowl. *Nature methods* 16, 12 (2019), 1247–1253.
- [17] CARLSON, N. R., AND BIRKETT, M. A. *Physiology of behavior*, 10th edn, 2010.
- [18] CHAMANZAR, A., AND NIE, Y. Weakly supervised multi-task learning for cell detection and segmentation. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (2020), IEEE, pp. 513–516.
- [19] CHEN, H., QI, X., YU, L., DOU, Q., QIN, J., AND HENG, P.-A. Dcan: Deep contour-aware networks for object instance segmentation from histology images. *Medical image analysis* 36 (2017), 135–146.



- 
- [20] CHEN, Y., LI, J., XIAO, H., JIN, X., YAN, S., AND FENG, J. Dual path networks. *Advances in neural information processing systems* 30 (2017).
- [21] COLTUC, D., BOLON, P., AND CHASSERY, J.-M. Exact histogram specification. *IEEE Transactions on Image processing* 15, 5 (2006), 1143–1152.
- [22] COUSTY, J., BERTRAND, G., NAJMAN, L., AND COUPRIE, M. Watershed cuts: Minimum spanning forests and the drop of water principle. *IEEE transactions on pattern analysis and machine intelligence* 31, 8 (2008), 1362–1374.
- [23] CUI, Y., ZHANG, G., LIU, Z., XIONG, Z., AND HU, J. A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images. *Medical & biological engineering & computing* 57, 9 (2019), 2027–2043.
- [24] DAUGUET, J., DELZESCAUX, T., CONDÉ, F., MANGIN, J.-F., AYACHE, N., HANTRAYE, P., AND FROUIN, V. Three-dimensional reconstruction of stained histological slices and 3d non-linear registration with in-vivo mri for whole baboon brain. *Journal of neuroscience methods* 164, 1 (2007), 191–204.
- [25] DAUGUET, J., MANGIN, J.-F., DELZESCAUX, T., AND FROUIN, V. Robust inter-slice intensity normalization using histogram scale-space analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2004), Springer, pp. 242–249.
- [26] DORINI, L. B., MINETTO, R., AND LEITE, N. J. Semiautomatic white blood cell segmentation based on multiscale analysis. *IEEE journal of biomedical and health informatics* 17, 1 (2012), 250–256.
- [27] ELSTON, G. N. Cortex, cognition and the cell: new insights into the pyramidal neuron and prefrontal function. *Cerebral cortex* 13, 11 (2003), 1124–1138.
- [28] FALK, T., MAI, D., BENSCH, R., ÇIÇEK, Ö., ABDULKADIR, A., MARRAKCHI, Y., BÖHM, A., DEUBNER, J., JÄCKEL, Z., SEIWALD, K., ET AL. U-net: deep learning for cell counting, detection, and morphometry. *Nature methods* 16, 1 (2019), 67–70.
- [29] FELZENSZWALB, P. F., AND HUTTENLOCHER, D. P. Efficient graph-based image segmentation. *International journal of computer vision* 59, 2 (2004), 167–181.

## BIBLIOGRAPHY

---

- [30] FÖRSTL, H., AND KURZ, A. Clinical features of alzheimer’s disease. *European archives of psychiatry and clinical neuroscience* 249, 6 (1999), 288–290.
- [31] FRANK, E. Effect of alzheimer’s disease on communication function. *Journal of the South Carolina Medical Association (1975)* 90, 9 (1994), 417–423.
- [32] FREEMAN, M. R., AND ROWITCH, D. H. Evolving concepts of gliogenesis: a look way back and ahead to the next 25 years. *Neuron* 80, 3 (2013), 613–623.
- [33] FREUND, Y., AND SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences* 55, 1 (1997), 119–139.
- [34] FRIEDMAN, J., HASTIE, T., AND TIBSHIRANI, R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics* 28, 2 (2000), 337–407.
- [35] GARY, C., LAM, S., HÉRARD, A.-S., KOCH, J. E., PETIT, F., GIPCHTEIN, P., SAWIAK, S. J., CAILLIEREZ, R., EDDARKAOUI, S., COLIN, M., ET AL. Encephalopathy induced by alzheimer brain inoculation in a non-human primate. *Acta neuropathologica communications* 7, 1 (2019), 1–21.
- [36] GINHOUX, F., LIM, S., HOEFFEL, G., LOW, D., AND HUBER, T. Origin and differentiation of microglia. *Frontiers in cellular neuroscience* 7 (2013), 45.
- [37] GINSBURGER, K., MATUSCHKE, F., POUPON, F., MANGIN, J.-F., AXER, M., AND POUPON, C. Medusa: A gpu-based tool to create realistic phantoms of the brain microstructure using tiny spheres. *NeuroImage* 193 (2019), 10–24.
- [38] GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDEFARLEY, D., OZAI, S., COURVILLE, A., AND BENGIO, Y. Generative adversarial networks. *Communications of the ACM* 63, 11 (2020), 139–144.
- [39] GUO, Z., LIU, H., NI, H., WANG, X., SU, M., GUO, W., WANG, K., JIANG, T., AND QIAN, Y. A fast and refined cancer regions segmentation framework in whole-slide breast pathological images. *Scientific reports* 9, 1 (2019), 1–10.

- 
- [40] GURCAN, M. N., BOUCHERON, L. E., CAN, A., MADABHUSHI, A., RAJPOOT, N. M., AND YENER, B. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering 2* (2009), 147–171.
- [41] HAINES, D. E., AND MIHAILOFF, G. A. *Fundamental neuroscience for basic and clinical applications E-book*. Elsevier Health Sciences, 2017.
- [42] HAMPEL, H., MESULAM, M.-M., CUELLO, A. C., FARLOW, M. R., GIACOBINI, E., GROSSBERG, G. T., KHACHATURIAN, A. S., VERGALLO, A., CAVEDO, E., SNYDER, P. J., ET AL. The cholinergic system in the pathophysiology and treatment of alzheimer’s disease. *Brain 141*, 7 (2018), 1917–1933.
- [43] HE, B., HU, W., ZHANG, K., YUAN, S., HAN, X., SU, C., ZHAO, J., WANG, G., WANG, G., AND ZHANG, L. Image segmentation algorithm of lung cancer based on neural network model. *Expert Systems 39*, 3 (2022), e12822.
- [44] HE, K., GKIOXARI, G., DOLLÁR, P., AND GIRSHICK, R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2961–2969.
- [45] HE, K., ZHANG, X., REN, S., AND SUN, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.
- [46] HE, Y., GONG, H., XIONG, B., XU, X., LI, A., JIANG, T., SUN, Q., WANG, S., LUO, Q., AND CHEN, S. icut: an integrative cut algorithm enables accurate segmentation of touching cells. *Scientific reports 5*, 1 (2015), 1–17.
- [47] HERCULANO-HOUZEL, S. Numbers of neurons as biological correlates of cognitive capability. *Current Opinion in Behavioral Sciences 16* (2017), 1–7.
- [48] HERCULANO-HOUZEL, S., AND LENT, R. Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain. *Journal of Neuroscience 25*, 10 (2005), 2518–2521.
- [49] HIGGINS, C. Applications and challenges of digital pathology and whole slide imaging. *Biotechnic & Histochemistry 90*, 5 (2015), 341–347.
- [50] HO, T. K. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (1995), vol. 1, IEEE, pp. 278–282.

## BIBLIOGRAPHY

---

- [51] HOLLANDI, R., SZKALISITY, A., TOTH, T., TASNADI, E., MOLNAR, C., MATHE, B., GREXA, I., MOLNAR, J., BALIND, A., GORBE, M., ET AL. nucleaizer: a parameter-free deep learning framework for nucleus segmentation using image style transfer. *Cell Systems* 10, 5 (2020), 453–458.
- [52] HOU, Y., DAN, X., BABBAR, M., WEI, Y., HASSELBALCH, S. G., CROTEAU, D. L., AND BOHR, V. A. Ageing as a risk factor for neurodegenerative disease. *Nature Reviews Neurology* 15, 10 (2019), 565–581.
- [53] HUANG, G., LIU, Z., VAN DER MAATEN, L., AND WEINBERGER, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 4700–4708.
- [54] HUGHES, C., AND LANTOS, P. A morphometric study of blood vessel, neuron and glial cell distribution in young and old rat brain. *Journal of the neurological sciences* 79, 1-2 (1987), 101–110.
- [55] INDU, M., RATHY, R., AND BINU, M. “slide less pathology”: Fairy tale or reality? *Journal of oral and maxillofacial pathology: JOMFP* 20, 2 (2016), 284.
- [56] ISENSEE, F., JAEGER, P. F., KOHL, S. A., PETERSEN, J., AND MAIER-HEIN, K. H. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* 18, 2 (2021), 203–211.
- [57] ISOLA, P., ZHU, J.-Y., ZHOU, T., AND EFROS, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1125–1134.
- [58] JADON, S. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (2020), IEEE, pp. 1–7.
- [59] JAIN, A. K. *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.
- [60] JIANG, T., YU, J.-T., TIAN, Y., AND TAN, L. Epidemiology and etiology of alzheimer’s disease: from genetic to non-genetic factors. *Current Alzheimer Research* 10, 8 (2013), 852–867.
- [61] JONES, K. *Neurological Assessment E-Book: A Clinician’s Guide*. Elsevier Health Sciences, 2011.

- 
- [62] KAMETANI, F., AND HASEGAWA, M. Reconsideration of amyloid hypothesis and tau hypothesis in alzheimer’s disease. *Frontiers in neuroscience* (2018), 25.
- [63] KARLSEN, A. S., AND PAKKENBERG, B. Total numbers of neurons and glial cells in cortex and basal ganglia of aged brains with down syndrome—a stereological study. *Cerebral cortex* 21, 11 (2011), 2519–2524.
- [64] KARRAS, T., LAINE, S., AND AILA, T. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), pp. 4401–4410.
- [65] KHAN, A. M., RAJPOOT, N., TREANOR, D., AND MAGEE, D. A non-linear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE transactions on Biomedical Engineering* 61, 6 (2014), 1729–1738.
- [66] KHENED, M., KORI, A., RAJKUMAR, H., KRISHNAMURTHI, G., AND SRINIVASAN, B. A generalized deep learning framework for whole-slide image segmentation and analysis. *Scientific reports* 11, 1 (2021), 1–14.
- [67] KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM* 60, 6 (2017), 84–90.
- [68] KUMAR, N., VERMA, R., ANAND, D., ZHOU, Y., ONDER, O. F., TSOUGENIS, E., CHEN, H., HENG, P.-A., LI, J., HU, Z., ET AL. A multi-organ nucleus segmentation challenge. *IEEE transactions on medical imaging* 39, 5 (2019), 1380–1391.
- [69] KUMAR, N., VERMA, R., SHARMA, S., BHARGAVA, S., VAHADANE, A., AND SETHI, A. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging* 36, 7 (2017), 1550–1560.
- [70] LAFERLA, F. M., AND GREEN, K. N. Animal models of alzheimer disease. *Cold Spring Harbor perspectives in medicine* 2, 11 (2012), a006320.
- [71] LAGUITTON, S., RIVIERE, D., VINCENT, T., FISCHER, C., GEFFROY, D., SOUDET, N., DENGHIEN, I., AND COINTEPAS, Y. Soma-workflow: a unified and simple interface to parallel computing resources. In *MICCAI workshop on high performance and distributed computing for medical imaging* (2011).

## BIBLIOGRAPHY

---

- [72] LAROBINA, M., AND MURINO, L. Medical image file formats. *Journal of digital imaging* 27, 2 (2014), 200–206.
- [73] LECUN, Y., BENGIO, Y., AND HINTON, G. Deep learning. *nature* 521, 7553 (2015), 436–444.
- [74] LECUN, Y., BOTTOU, L., BENGIO, Y., AND HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 11 (1998), 2278–2324.
- [75] LEI, S., ZHANG, H., WANG, K., AND SU, Z. How training data affect the accuracy and robustness of neural networks for image classification.
- [76] LIN, G., ADIGA, U., OLSON, K., GUZOWSKI, J. F., BARNES, C. A., AND ROYSAM, B. A hybrid 3d watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. *Cytometry Part A: the journal of the International Society for Analytical Cytology* 56, 1 (2003), 23–36.
- [77] LIN, G., CHAWLA, M. K., OLSON, K., GUZOWSKI, J. F., BARNES, C. A., AND ROYSAM, B. Hierarchical, model-based merging of multiple fragments for improved three-dimensional segmentation of nuclei. *Cytometry Part A: The Journal of the International Society for Analytical Cytology* 63, 1 (2005), 20–33.
- [78] LIU, D., ZHANG, D., SONG, Y., HUANG, H., AND CAI, W. Panoptic feature fusion net: a novel instance segmentation paradigm for biomedical and biological images. *IEEE Transactions on Image Processing* 30 (2021), 2045–2059.
- [79] LIU, X., GUO, Z., CAO, J., AND TANG, J. Mdc-net: a new convolutional neural network for nucleus segmentation in histopathology images with distance maps and contour information. *Computers in Biology and Medicine* 135 (2021), 104543.
- [80] LONG, J., SHELFHAMER, E., AND DARRELL, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440.
- [81] LOU, X., KOETHE, U., WITTBRODT, J., AND HAMPRECHT, F. A. Learning to segment dense cell nuclei with shape prior. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (2012), IEEE, pp. 1012–1018.
- [82] MACENKO, M., NIETHAMMER, M., MARRON, J. S., BORLAND, D., WOOSLEY, J. T., GUAN, X., SCHMITT, C., AND THOMAS, N. E. A

- method for normalizing histology slides for quantitative analysis. In *2009 IEEE international symposium on biomedical imaging: from nano to macro* (2009), IEEE, pp. 1107–1110.
- [83] MALANDAIN, G., AND BARDINET, E. Intensity compensation within series of images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2003), Springer, pp. 41–49.
- [84] MALANDAIN, G., BARDINET, E., NELISSEN, K., AND VANDUFFEL, W. Fusion of autoradiographs with an mr volume using 2-d and 3-d linear transformations. *NeuroImage* 23, 1 (2004), 111–127.
- [85] MALPICA, N., DE SOLÓRZANO, C. O., VAQUERO, J. J., SANTOS, A., VALLCORBA, I., GARCÍA-SAGREDO, J. M., AND DEL POZO, F. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry: The Journal of the International Society for Analytical Cytology* 28, 4 (1997), 289–297.
- [86] MARKRAM, H., TOLEDO-RODRIGUEZ, M., WANG, Y., GUPTA, A., SILBERBERG, G., AND WU, C. Interneurons of the neocortical inhibitory system. *Nature reviews neuroscience* 5, 10 (2004), 793–807.
- [87] MIKOŁAJCZYK, A., AND GROCHOWSKI, M. Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop (IIPhDW)* (2018), IEEE, pp. 117–122.
- [88] MILLETARI, F., NAVAB, N., AND AHMADI, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)* (2016), IEEE, pp. 565–571.
- [89] MOCNIK, F.-B. Benford’s law and geographical information—the example of openstreetmap. *International journal of geographical information science* 35, 9 (2021), 1746–1772.
- [90] MTUI, E., GRUENER, G., AND DOCKERY, P. *Fitzgerald’s Clinical Neuroanatomy and Neuroscience E-Book*. Elsevier Health Sciences, 2020.
- [91] NAYLOR, P., LAÉ, M., REYAL, F., AND WALTER, T. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging* 38, 2 (2018), 448–459.
- [92] NOH, H., HONG, S., AND HAN, B. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (2015), pp. 1520–1528.

## BIBLIOGRAPHY

---

- [93] O'KEEFE, J. Hippocampal neurophysiology in the behaving animal.
- [94] OTSU, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* 9, 1 (1979), 62–66.
- [95] PARONI, G., BISCEGLIA, P., AND SERIPA, D. Understanding the amyloid hypothesis in alzheimer's disease. *Journal of Alzheimer's Disease* 68, 2 (2019), 493–510.
- [96] PELVIG, D. P., PAKKENBERG, H., STARK, A. K., AND PAKKENBERG, B. Neocortical glial cell numbers in human brains. *Neurobiology of aging* 29, 11 (2008), 1754–1762.
- [97] PICCONI, B., PISANI, A., BARONE, I., BONSI, P., CENTONZE, D., BERNARDI, G., AND CALABRESI, P. Pathological synaptic plasticity in the striatum: implications for parkinson's disease. *Neurotoxicology* 26, 5 (2005), 779–783.
- [98] QU, H., WU, P., HUANG, Q., YI, J., YAN, Z., LI, K., RIEDLINGER, G. M., DE, S., ZHANG, S., AND METAXAS, D. N. Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. *IEEE transactions on medical imaging* 39, 11 (2020), 3655–3666.
- [99] RABINOVICH, A., AGARWAL, S., LARIS, C., PRICE, J., AND BELONGIE, S. Unsupervised color decomposition of histologically stained tissue samples. *Advances in neural information processing systems* 16 (2003).
- [100] RADFORD, A., METZ, L., AND CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015).
- [101] RAIMONDO, F., GAVRIELIDES, M. A., KARAYANNOPOULOU, G., LYROUDIA, K., PITAS, I., AND KOSTOPOULOS, I. Automated evaluation of her-2/neu status in breast tissue from fluorescent in situ hybridization images. *IEEE Transactions on Image Processing* 14, 9 (2005), 1288–1299.
- [102] REINHARD, E., ADHIKHMEN, M., GOOCH, B., AND SHIRLEY, P. Color transfer between images. *IEEE Computer graphics and applications* 21, 5 (2001), 34–41.
- [103] RIVA, P., BATTAGLIA, C., AND VENTURIN, M. Emerging role of genetic alterations affecting exosome biology in neurodegenerative diseases. *International Journal of Molecular Sciences* 20, 17 (2019), 4113.



- 
- [104] ROERDINK, J. B., AND MEIJSTER, A. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta informaticae* 41, 1-2 (2000), 187–228.
- [105] RONNEBERGER, O., FISCHER, P., AND BROX, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (2015), Springer, pp. 234–241.
- [106] RUSZCZYCKI, B., PELS, K. K., WALCZAK, A., ZAMLYŃSKA, K., SUCH, M., SZCZEPANKIEWICZ, A. A., HALL, M. H., MAGALSKA, A., MAGNOWSKA, M., WOLNY, A., ET AL. Three-dimensional segmentation and reconstruction of neuronal nuclei in confocal microscopic images. *Frontiers in neuroanatomy* 13 (2019), 81.
- [107] SALEHI, P., AND CHALECHALE, A. Pix2pix-based stain-to-stain translation: A solution for robust stain normalization in histopathology images analysis. In *2020 International Conference on Machine Vision and Image Processing (MVIP)* (2020), IEEE, pp. 1–7.
- [108] SALIMI, I., FRIEL, K. M., AND MARTIN, J. H. Pyramidal tract stimulation restores normal corticospinal tract connections and visuomotor skill after early postnatal motor cortex activity blockade. *Journal of Neuroscience* 28, 29 (2008), 7426–7434.
- [109] SANDLER, M., HOWARD, A., ZHU, M., ZHMOGINOV, A., AND CHEN, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 4510–4520.
- [110] SARKAR, R., DARBY, D., FOUCAMBERT, H., MEILHAC, S., AND OLIVO-MARIN, J.-C. Nu3d: 3d nuclei segmentation from light-sheet microscopy images of the embryonic heart. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)* (2021), IEEE, pp. 929–933.
- [111] SCHMIDT, U., WEIGERT, M., BROADDUS, C., AND MYERS, G. Cell detection with star-convex polygons. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2018), Springer, pp. 265–273.
- [112] SCHMITT, O., AND HASSE, M. Morphological multiscale decomposition of connected regions with emphasis on cell clusters. *Computer Vision and Image Understanding* 113, 2 (2009), 188–201.

## BIBLIOGRAPHY

---

- [113] SCHNEIDER, C. A., RASBAND, W. S., AND ELICEIRI, K. W. Nih image to imagej: 25 years of image analysis. *Nature methods* 9, 7 (2012), 671–675.
- [114] SHABAN, M. T., BAUR, C., NAVAB, N., AND ALBARQOUNI, S. Staingan: Stain style transfer for digital histological images. In *2019 Ieee 16th international symposium on biomedical imaging (Isbi 2019)* (2019), IEEE, pp. 953–956.
- [115] SHOSTAK, S. Histology’s nomenclature: Past, present and future. *Biol Syst Open Access* 2, 122 (2013), 2.
- [116] SHU, J., FU, H., QIU, G., KAYE, P., AND ILYAS, M. Segmenting overlapping cell nuclei in digital histopathology images. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2013), IEEE, pp. 5445–5448.
- [117] SIMONYAN, K., AND ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [118] SINGH, R., CHUBB, L., PANTANOWITZ, L., AND PARWANI, A. Standardization in digital pathology: Supplement 145 of the dicom standards. *Journal of pathology informatics* 2, 1 (2011), 23.
- [119] SIRINUKUNWATTANA, K., RAZA, S. E. A., TSANG, Y.-W., SNEAD, D. R., CREE, I. A., AND RAJPOOT, N. M. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging* 35, 5 (2016), 1196–1206.
- [120] SNELL, R. S. *Clinical neuroanatomy*. Lippincott Williams & Wilkins, 2010.
- [121] SONG, Y., ZHANG, L., CHEN, S., NI, D., LEI, B., AND WANG, T. Accurate segmentation of cervical cytoplasm and nuclei based on multi-scale convolutional network and graph partitioning. *IEEE Transactions on Biomedical Engineering* 62, 10 (2015), 2421–2433.
- [122] STEGMAIER, J., AMAT, F., LEMON, W. C., MCDOLE, K., WAN, Y., TEODORO, G., MIKUT, R., AND KELLER, P. J. Real-time three-dimensional cell segmentation in large-scale microscopy data of developing embryos. *Developmental cell* 36, 2 (2016), 225–240.
- [123] STRINGER, C., WANG, T., MICHAELIS, M., AND PACHITARIU, M. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods* 18, 1 (2021), 100–106.

- 
- [124] SUZUKI, Y., HALONEN, S., WANG, X., AND WEN, X. Cerebral toxoplasmosis: pathogenesis and host resistance. In *Toxoplasma gondii*. Elsevier, 2007, pp. 567–591.
- [125] SZEGEDY, C., IOFFE, S., VANHOUCKE, V., AND ALEMI, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence* (2017).
- [126] SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCKE, V., AND RABINOVICH, A. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 1–9.
- [127] TAN, M., AND LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (2019), PMLR, pp. 6105–6114.
- [128] TANG, S., AND YU, F. Construction and verification of retinal vessel segmentation algorithm for color fundus image under bp neural network model. *The Journal of Supercomputing* 77, 4 (2021), 3870–3884.
- [129] TEICHMANN, M., DUPOUX, E., KOUIDER, S., BRUGIÈRES, P., BOISSÉ, M.-F., BAUDIC, S., CESARO, P., PESCHANSKI, M., AND BACHOUD-LÉVI, A.-C. The role of the striatum in rule application: the model of huntington’s disease at early stage. *Brain* 128, 5 (2005), 1155–1167.
- [130] TELLEZ, D., LITJENS, G., BÁNDI, P., BULTEN, W., BOKHORST, J.-M., CIOMPI, F., AND VAN DER LAAK, J. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical image analysis* 58 (2019), 101544.
- [131] THU, D. C., OORSCHOT, D. E., TIPPETT, L. J., NANA, A. L., HOGG, V. M., SYNEK, B. J., LUTHI-CARTER, R., WALDVOGEL, H. J., AND FAULL, R. L. Cell loss in the motor and cingulate cortex correlates with symptomatology in huntington’s disease. *Brain* 133, 4 (2010), 1094–1110.
- [132] VAHADANE, A., PENG, T., SETHI, A., ALBARQOUNI, S., WANG, L., BAUST, M., STEIGER, K., SCHLITTER, A. M., ESPOSITO, I., AND NAVAB, N. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE transactions on medical imaging* 35, 8 (2016), 1962–1971.

## BIBLIOGRAPHY

---

- [133] VAN DER MAATEN, L., AND HINTON, G. Visualizing data using t-sne. *Journal of machine learning research* 9, 11 (2008).
- [134] VAN ETTEN, A., AND HOGAN, D. The spacenet multi-temporal urban development challenge. *arXiv preprint arXiv:2102.11958* (2021).
- [135] VASILJEVIĆ, J., FEUERHAKE, F., WEMMERT, C., AND LAMPERT, T. Towards histopathological stain invariance by unsupervised domain augmentation using generative adversarial networks. *Neurocomputing* 460 (2021), 277–291.
- [136] VETA, M., HUISMAN, A., VIERGEVER, M. A., VAN DIEST, P. J., AND PLUIM, J. P. Marker-controlled watershed segmentation of nuclei in h&e stained breast cancer biopsy images. In *2011 IEEE international symposium on biomedical imaging: from nano to macro* (2011), IEEE, pp. 618–621.
- [137] VETA, M., VAN DIEST, P. J., KORNEGOOR, R., HUISMAN, A., VIERGEVER, M. A., AND PLUIM, J. P. Automatic nuclei segmentation in h&e stained breast cancer histopathology images. *PloS one* 8, 7 (2013), e70221.
- [138] VICAR, T., RAUDENSKA, M., GUMULEC, J., AND BALVAN, J. The quantitative-phase dynamics of apoptosis and lytic cell death. *Scientific reports* 10, 1 (2020), 1–12.
- [139] VINCENT, L., AND BEUCHER, S. *The morphological approach to segmentation: an introduction*. Centre de Morphologie Mathématique, Ecole Nationale Supérieure des Mines de Paris, 1989.
- [140] VOLICER, L., HARPER, D. G., MANNING, B. C., GOLDSTEIN, R., AND SATLIN, A. Sundowning and circadian rhythms in alzheimer’s disease. *American Journal of Psychiatry* 158, 5 (2001), 704–711.
- [141] WALDEMAR, G., DUBOIS, B., EMRE, M., GEORGES, J., MCKEITH, I., ROSSOR, M., SCHELTENS, P., TARISKA, P., AND WINBLAD, B. Recommendations for the diagnosis and management of alzheimer’s disease and other disorders associated with dementia: Efn guideline. *European Journal of Neurology* 14, 1 (2007), e1–e26.
- [142] WALLØE, S., PAKKENBERG, B., AND FABRICIUS, K. Stereological estimation of total cell numbers in the human cerebral and cerebellar cortex. *Frontiers in human neuroscience* 8 (2014), 508.
- [143] WANG, J., SUN, K., CHENG, T., JIANG, B., DENG, C., ZHAO, Y., LIU, D., MU, Y., TAN, M., WANG, X., ET AL. Deep high-resolution

- representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence* 43, 10 (2020), 3349–3364.
- [144] WANG, Q., NIEMI, J., TAN, C.-M., YOU, L., AND WEST, M. Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy. *Cytometry Part A: The Journal of the International Society for Advancement of Cytometry* 77, 1 (2010), 101–110.
- [145] WANG, S., CONG, Y., ZHU, H., CHEN, X., QU, L., FAN, H., ZHANG, Q., AND LIU, M. Multi-scale context-guided deep network for automated lesion segmentation with endoscopy images of gastrointestinal tract. *IEEE Journal of Biomedical and Health Informatics* 25, 2 (2020), 514–525.
- [146] WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.
- [147] WEIGERT, M., SCHMIDT, U., HAASE, R., SUGAWARA, K., AND MYERS, G. Star-convex polyhedra for 3d object detection and segmentation in microscopy. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2020), pp. 3666–3673.
- [148] WEST, M., SLOMIANKA, L., AND GUNDERSEN, H. J. G. Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. *The Anatomical Record* 231, 4 (1991), 482–497.
- [149] WINBLAD, B., AMOUYEL, P., ANDRIEU, S., BALLARD, C., BRAYNE, C., BRODATY, H., CEDAZO-MINGUEZ, A., DUBOIS, B., EDVARDSSON, D., FELDMAN, H., ET AL. Defeating alzheimer’s disease and other dementias: a priority for european science and society. *The Lancet Neurology* 15, 5 (2016), 455–532.
- [150] WU, H., SOUEDET, N., JAN, C., CLOUCHOUX, C., AND DELZESCAUX, T. A general deep learning framework for neuron instance segmentation based on efficient unet and morphological post-processing. *Computers in Biology and Medicine* (2022), 106180.
- [151] WU, H., SOUEDET, N., YOU, Z., JAN, C., CLOUCHOUX, C., AND DELZESCAUX, T. Evaluation of deep learning topcoders method for neuron individualization in histological macaque brain section. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (2021), IEEE, pp. 2985–2988.

## BIBLIOGRAPHY

---

- [152] WU, H., WANG, Z., SONG, Y., YANG, L., AND QIN, J. Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 11666–11675.
- [153] WU, Z., AND LEAHY, R. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 15, 11 (1993), 1101–1113.
- [154] XIE, L., QI, J., PAN, L., AND WALI, S. Integrating deep convolutional neural networks with marker-controlled watershed for overlapping nuclei segmentation in histopathology images. *Neurocomputing* 376 (2020), 166–179.
- [155] XING, F., XIE, Y., AND YANG, L. An automatic learning-based framework for robust nucleus segmentation. *IEEE transactions on medical imaging* 35, 2 (2015), 550–566.
- [156] XING, F., AND YANG, L. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE reviews in biomedical engineering* 9 (2016), 234–263.
- [157] YANG, L., QIU, Z., GREENAWAY, A. H., AND LU, W. A new framework for particle detection in low-snr fluorescence live-cell images and its application for improved particle tracking. *IEEE transactions on biomedical engineering* 59, 7 (2012), 2040–2050.
- [158] YANG, S., XIAO, W., ZHANG, M., GUO, S., ZHAO, J., AND SHEN, F. Image data augmentation for deep learning: A survey. *arXiv preprint arXiv:2204.08610* (2022).
- [159] YOU, Z. *Étude de la morphologie et de la distribution des neurones dans le cerveau de macaque par microscopie optique*. PhD thesis, Université Pierre et Marie Curie, 2017.
- [160] YOU, Z., BALBASTRE, Y., BOUVIER, C., HÉRARD, A.-S., GIPCHTEIN, P., HANTRAYE, P., JAN, C., SOUÉDET, N., AND DELZESCAUX, T. Automated individualization of size-varying and touching neurons in macaque cerebral microscopic images. *Frontiers in neuroanatomy* 13 (2019), 98.
- [161] YOU, Z., JIANG, M., SHI, Z., SHI, C., DU, S., LIANG, J., HÉRARD, A.-S., JAN, C., SOUÉDET, N., AND DELZESCAUX, T. Macaque neuron instance segmentation only with point annotations based on multiscale

- fully convolutional regression neural network. *Neural Computing and Applications* 34, 4 (2022), 2925–2938.
- [162] ZARELLA, M. D., BOWMAN, D., AEFNER, F., FARAHANI, N., XTHONA, A., ABSAR, S. F., PARWANI, A., BUI, M., AND HARTMAN, D. J. A practical guide to whole slide imaging: a white paper from the digital pathology association. *Archives of pathology & laboratory medicine* 143, 2 (2019), 222–234.
- [163] ZENG, Z., XIE, W., ZHANG, Y., AND LU, Y. Ric-unet: An improved neural network based on unet for nuclei segmentation in histology images. *Ieee Access* 7 (2019), 21420–21428.
- [164] ZHANG, C., HUBER, F., KNOP, M., AND HAMPRECHT, F. A. Yeast cell detection and segmentation in bright field microscopy. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* (2014), IEEE, pp. 1267–1270.
- [165] ZHANG, J., YU, J., FU, S., AND TIAN, X. Adoption value of deep learning and serological indicators in the screening of atrophic gastritis based on artificial intelligence. *The Journal of Supercomputing* 77, 8 (2021), 8674–8693.
- [166] ZHANG, Z., MAYER, G., DAUVILLIERS, Y., PLAZZI, G., PIZZA, F., FRONCZEK, R., SANTAMARIA, J., PARTINEN, M., OVEREEM, S., PERAITA-ADRADOS, R., ET AL. Exploring the clinical features of narcolepsy type 1 versus narcolepsy type 2 from european narcolepsy network database with machine learning. *Scientific reports* 8, 1 (2018), 1–11.
- [167] ZHOU, J., ZHANG, X., AND JIANG, Z. Recognition of imbalanced epileptic eeg signals by a graph-based extreme learning machine. *Wireless Communications and Mobile Computing* 2021 (2021).
- [168] ZHOU, Z., SIDDIQUEE, M. M. R., TAJBAKHSI, N., AND LIANG, J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging* 39, 6 (2019), 1856–1867.
- [169] ZHU, J.-Y., PARK, T., ISOLA, P., AND EFROS, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2223–2232.