UNIVERSITÀ
DEGLI STUDI
FIRENZE

Università degli Studi di Firenze
Dipartimento di Ingegneria dell'Informazione (DINFO)
Corso di Dottorato in Smart Computing

Detection of Neuronal Soma from 3D Large-Scale Light-Sheet Microscopy Data

Candidate
Curzio Checcucci

Supervisor
Prof. Paolo Frasconi

PhD Coordinator
Prof. Stefano Berretti

ciclo XXXVI, 2020-23

Università degli Studi di Firenze, Dipartimento di Ingegneria dell'Informazione (DINFO).

To all those who can't
even dream of a doctorate

Acknowledgments

I would like to acknowledge the efforts, patience and input of my supervisor, Prof. Paolo Frasconi, whose incomparable expertise always made all projects go one step further. Thanks to Dr. Ludovico Silvestri, whose science-enthusiasm was contagious, he actually gave birth to the idea of this PhD and made me meet Prof. Frasconi. In particular my thanks go to Dr. Irene Costantini who collaborated on the main part of my research work. I would like to thank also Dr. Anna Kreshuk for her kind support and hospitality during my stay at EMBL, Heidelberg.

My gratitude to all colleagues of the AI lab, especially to Andrea Gemelli and Luca Bindini, whose motivation lightened the heavy moments of research frustration and with whom sharing scientific thoughts has always been helpful. I can not avoid mentioning Dashara Shullani and Daniele Baracchi, two pillars of our Department, a certainty amidst this uncertain world of academic research and of course, their lab, our favorite canteen. A special thank to Roberto Verdecchia, having him just at the next door was the best and funny distraction of last years. I would like to thank also Imad Zaza for having solved so many technical problems I had and for all the not so innocent laughs we had.

My special thanks go to the best university classmates ever, Chiara Camerota, Eugenio Palmieri and Giacomo Duroni, (almost) always by my side and whose endless discussions (almost) never went anywhere but taught me everything. Thanks to Alberto Zingoni for hosting so many board-games evenings, his games collection surprises me every time.

Last, but not least, to my family whose unconditional love has gone beyond any scientific comprehension and whose strength in overcoming so many difficulties is inspirational. I love you, Mum; I love you, Dad. Thank you. Thanks to my big sister, Simona, an unbreakable yet soft cornerstone, both the pillow and the hammer of our family. You have illuminated so many paths I have traveled; this PhD is just one of them, and not even the most intricate.

Abstract

Learning spatial distribution of neurons in specific brain areas is crucial for advancing the knowledge of brain structures and functions. Thanks to the advancements brought by light-sheet fluorescence microscopy, large-scale brain tissues are nowadays available at sub-cellular resolution. This comes with the challenge of processing Tera-bytes of data in a reasonable time and with strong performances. The work described in this thesis presents a reliable and accurate two-step approach for cell detection from vast and highly variable 3D image datasets. Multiple convolutional neural network variants are implemented in order to extract truthful probability maps from raw images and to facilitate the cell localization performed by a subsequent blob detector. The efficacy and scalability to huge data of the proposed technique is demonstrated through extensive validation and application on a cohort study of whole mouse brains and the entire Broca's area of a human. The automatic detection of neuronal soma from whole mouse brains allowed for brain-wide quantitative analyses, confirming biologically accepted theories of brain activations and connectivity on fear memory experiments, but also suggesting novel neuroscientific research directions. An extensive comparison with other state-of-the-art algorithms and with stereology, the gold-standard for large-scale neuronal counts, is presented on an entire human Broca's area. Results therefore prove the adaptability and effectiveness of deep-learning approaches in a high variety of contexts, hoping to provide many life-science laboratories worldwide with a tool to advance their researches.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cell detection is a highly common and fundamental task in many biological experiments. Having a complete mapping of the neurons in a brain region or even the whole brain could reveal important insights on its organization and functions. Systematic sampling procedures, as employed by stereology, have been applied for decades allowing for unbiased count estimates on whole brain regions. However, only average densities are retrieved, homogeneity assumption and a-priori subregion segmentation are necessary and moreover each tissue sample needs individual analysis requiring new manual annotations and subregion segmentation. Some automation were introduced by software like Fiji [68] and CellProfiler [11] that made classical image processing pipelines available to a large audience. However, relying on user defined image filters and thresholds, finding the parameters of such operations can be time-consuming and their results are not guaranteed to be optimal on all conditions the data may provide. Machine-learning approaches [19, 75] tried to overcome this issue by letting an automatic classifier to choose the optimal combination of features for the given task, but the set of input features must be a-priori selected, thus possibly avoiding considering other more important aspects of the data under study. Moreover and importantly, all above mentioned tools are not designed for large-scale image analyses, hence their application to high resolution images of whole brain regions could take years, hampering their adoption in this field. Deep-learning on the other hand, could overcome all these limitations by automatically learn the most discriminative features and providing highly efficient GPU implementations. Here we present BCFind-v2, a two-step approach for cell localization from

3D images. A fully-convolutional neural network is exploited to obtain 3D probability maps of neuronal soma from which a subsequent blob detector can easily recover the cell coordinates. Extensive application to whole mouse brains and to the entire Broca's area of a human together with a comprehensive comparison with two state-of-the-art deep-learning models for cell segmentation (StarDist [69, 83] and CellPose [77]) and with stereology prove the effectiveness of the proposed approach and its adaptability to various imaging properties.

1.1 The objective

The primary aim of this research project is the development of a robust and highly accurate deep-learning model for the precise cell localization from 3D light-sheet fluorescence microscopy images. This rapidly expanding imaging technique allows worldwide laboratories to acquire biological tissue samples of unprecedented scale at sub-cellular resolution. This possibility has the potential of unveiling previously unknown mechanisms within highly complex organs, contributing therefore to the improvement of treatments even for severe diseases. However, in order to make it possible highly accurate, automatic and scalable techniques are needed for systematic quantitative analyses. Our research moves towards this direction. We aim at developing a deep-learning model that provides state-of-the-art results on a wide range of distinct datasets with specific voxel resolutions and image properties. Moreover, since variability does not only come from different datasets but also from within the same dataset, the approach should be robust and reliable even after being specialized on that same data. It is indeed hard to train a model on a truly representative subset of such vast and variable data. Therefore, proper validation must be carried out in order to verify its generalization capability. Moreover, scalability is another key factor that we would like to address: processing even huge amount of data should be done in a reasonable time. Finally, we would also like to develop a user-friendly software which takes care of all needed steps, from data I/O, its preprocessing, the model training and validation, debugging protocols and visualization, to the final large-scale prediction, and last but not least, such software should be portable in order to be used in a large variety of environments even by people without extensive deep-learning expertise.

1.2 Contributions

The contributions of this thesis are manifold. A two-step approach that adopts a fully-convolutional neural network (FCNN) followed by a blob detector is exploited in order to maximize cell localization accuracy and adaptability to the unique challenges posed by 3D light-sheet fluorescence microscopy. Multiple variants of FCNNs are studied, implemented and tested both on mice and human brain datasets. An efficient GPU implementation of a classical blob detector is provided together with an automatic selection of its parameters. A fast and reliable localization pipeline is therefore presented and successfully applied to a cohort study of 29 whole mouse brains. Moreover, we further apply our technique to an entire Broca's area of a human, comparing its performances both with other state-of-the-art deep learning methods for cell localization and with the gold-standard for neuronal counts that is stereology. All these extensive applications and validations prove the reliability of our method and its scalability to large-scale data.

The developed software is available at `https://codeberg.org/curzio/BCFind-v2`. Easy-to-use command line commands to train and predict on large-scale data are provided, experiment settings can be specified by configuration file and a Docker image can be built from provided Dockerfile for consistent deployment.

The organization of this thesis follows the evolution of the above mentioned contributions. Starting by giving an overview of the problem under study and reviewing the principal related works in Chapter 2, we then describe the adopted approach and all considered variants, finally presenting experimental results on two challenging datasets, in Chapter 3. Part of this chapter have been also published in [74]. The application of developed localization pipeline and its biologically relevant results are described in Chapter 4 and published in [21]. Finally the extensive comparison between multiple deep-learning approaches and stereology is presented in Chapter 5 and will be submitted soon.

Chapter 2

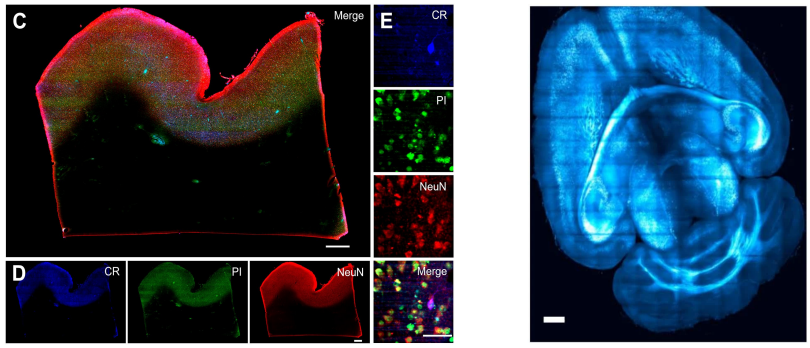
Brain cell counting techniques

This chapter gives a brief survey of related work on the analysis of biological images. The first part introduces the problem of cell counting and characterization from high-resolution light-sheet fluorescence microscopy images. Then we present, in order of increasing level of automation, various techniques usually adopted in the field to solve this problem. From histology and stereology for counts estimates, to classic computer vision approaches for feature extraction, to most modern machine-learning and deep-learning algorithms for pixel or image classification, a comprehensive review of the main software and tools available for brain cell analyses is given.

2.1 Introduction

Understanding the cellular composition of the brain is essential for unraveling its intricate functions and addressing neuroscientific questions. Cell counting, the process of quantifying the number and distribution of specific cells, has long been a fundamental technique in this endeavor. Over the years, several methodologies have been employed to count brain cells, ranging from traditional histological techniques to modern, cutting-edge technologies.

Light-sheet fluorescence microscopy (LSFM) [37] has transformed our ability to visualize and quantify neural cells in three dimensions with unprecedented clarity and efficiency. By illuminating and capturing specific fluorescent markers in brain tissues, LSFM provides subcellular-resolution



(a) (b)

Figure 2.1: LSFM imaging. (a) Slice of the Broca's area of a human brain [18]. (b) Maximum intensity projection of a whole mouse brain [74].

images of very large samples while minimizing photobleaching and phototoxicity. These advantages make it an invaluable tool for both basic research and clinical studies. However, light scattering arises when imaging deep within thick tissues and higher-resolutions also imply lower speeds, requiring a trade-off between acquisition time, image quality and tissue size. Separate acquisitions can also reveal large variations in image brightness, contrast and noise sources. Examples of LSFM imaging of a whole mouse brain and a slice of human brain are depicted in Figure 2.1.

Coupling this high-throughput imaging technique with accurate cell counting methodologies could be therefore a cornerstone in understanding the complex organization of the brain and its evolution across various developmental stages of diseases or experimental conditions. We here review some of the most common counting tools in the field, classifying them according to their underlying processes and presenting them in order of increasing level of automation.

2.2 Mathematical methods

Traditional methods rely on manual counts and spatial distribution assumptions to obtain cell density estimates in specific regions of interest. In the early years of XX century, scientists used to manually counts cells from 2D

sections and infer volume counts with mathematical models taking into account cell diameter and section thickness [1, 85]. However, such approaches find applications only on small regions due to the human effort required and the inherent biases of large-scale 3D extrapolation. Later, stereological method has been introduced to obtain unbiased estimates of true cell counts [76]. By working on 3D probes and assuring that cells were counted once and only once, stereology provides a more reliable and scalable tool for cell counting. Based on systematic sampling procedures, 3D boxes are selected from a grid subdivision of evenly spaced volume sections. Manual annotations are then performed on these boxes avoiding cells overlapping the bottom-left borders in order to assure cells are only counted once (Figure 2.2). Then denoting with c_{si} the number of counted cells in frame i of section s, x, y and h the sizes of annotation frames, T the section thickness, x_{step} and y_{step} the grid horizontal and vertical steps and with ssf the section sampling fraction (i.e. the ratio between the number of sampled sections and the total number of sections in the volume of interest), volume counts are estimated as follows:

$$asf = \frac{x \cdot y}{x_{step} \cdot y_{step}}$$
$$N = \frac{T}{h} \cdot \frac{1}{asf} \cdot \frac{1}{ssf} \cdot \sum_s \sum_i c_{si}$$

Section sampling fraction and grid steps are selected in order to minimize the coefficient of error as defined by Gundersen (1988) [31].

Due to its several applications and its unbiased estimates, stereology is commonly considered, especially by the neuroscientific community, as the gold standard for cell counting. However it still requires a large amount of human labor, needs to be applied on regions with homogeneous density and repeated for different samples.

2.3 Classic computer vision approaches

Early efforts in trying to automate cell detection procedures exploit classical computer vision algorithms. Tools as ImageJ [68] and CellProfiler [11] provide the users with customizable pipelines of image processing, thresholding and morphological transformations in order to fit a wide range of needs. Image filters are commonly adopted as initial steps to remove noise, equalize

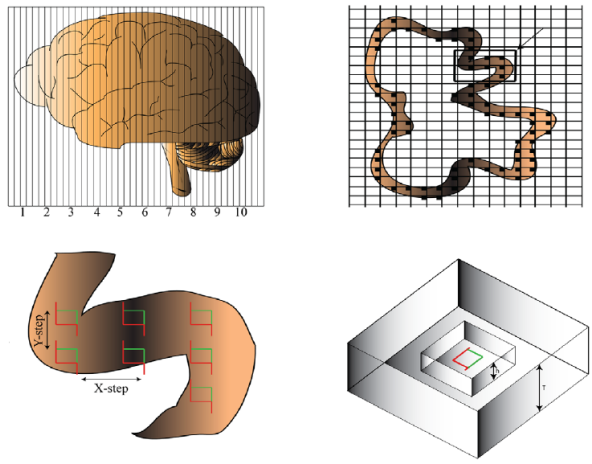


Figure 2.2: Optical fractionator sampling scheme for stereology. Sections of whole volume are evenly sampled from the tissue, a grid is superimposed to each section and annotations are performed on 3D frames of the grid.

contrast and enhance objects of interest. Then, a threshold is manually, or automatically [56,67,87], selected to separate the most likely relevant objects from the rest of the image. Finally, morphological transformations are then applied to refine cell segmentation and remove non biologically meaningful regions. On top of these operations watershed algorithm [79] is also commonly applied to separate and locate individual instances. However, when the objects of interest are defined more by texture and context than raw intensity many classical image processing techniques may fail. As a matter of fact these tools are mainly adopted for data exploration and quantification of known biologically-related features and are characterized by a high degree of human intervention.

ClearMap [62] is another open-source software specifically developed for LSFM imaging of cFOS$^+$ neurons in iDISCO [63] cleared tissues. In their workflow, cells are detected by consecutive application of an illumination correction pipeline, a background removal pipeline, a custom-designed equalization filter, a difference of Gaussians filter and a final local maxima detection.

2.4 Machine-learning methods

In the search of more flexible, generic and automated methods for cell detection and characterization in microscopy images, tools have been developed exploiting the advances in machine learning programs. iLastik [7, 75] offers an easy-to-use graphical user interface (GUI) to train a random forest classifier [9] from scribble annotations provided by the user. The random forest model is used to "capture highly non-linear decision boundaries in the feature space" [75] and for its generalization capabilities due to bootstrap sample (bag) of the training-data in each decision tree [10], optimization of the *out-of-bag* classification error and random feature selection at each leaf of the trees. However, by the very nature of machine learning models, users must select a fixed set of image features (color-based, edge-based and texture-based in iLastik) and this can hamper the accurate prediction of highly variable images, typically encountered in LSFM acquisitions.

Only recently CellProfiler developers have introduced CellProfiler Analyst [19] and in particular the *Classifier* application programming interface (API) for training classification algorithms. Unlike iLastik, their API allows for a variety of machine learning models, specifically random forest, adaptive boosting [24], support vector machine [17], gradient boosting [26], logistic regression, linear discriminant analysis [16], nearest neighbors and gentle boosting [25]. Being highly linked with CellProfiler, CellProfiler Analyst uses features previously extracted with the functionalities of its parent software. Classification is here more intended for cell type discovery rather than pixel-wise discrimination as in iLastik, i.e. it classifies image patches and uses more raw intensity than texture or context features. Moreover, only MySQL and SQLite databases can be accessed thus limiting more general uses.

2.5 Deep-learning methods

The last decade has seen the emergence of a broad family of algorithms known as deep-learning (DL) also in the field of biomedical image analysis. These algorithms automatically learn the most discriminative features thus greatly reducing the need of human intervention and feature engineering. DL models have also proven to reach unprecedented levels of accuracy gaining the attention of many scientists.

In a pioneering work, Frasconi et al. (2014) [23] proposed a neural network approach to 3D cell detection (BCFind). The model consisted of two steps: a 3 layers feedforward network (FFN) [14,66] for *semantic deconvolution* (i.e. non-linear transformation of input images to enhance semantically coherent objects) followed by mean-shift algorithm for the prediction of cell centers. Due to the high computational cost of MLPs, especially when applied to images, very small patches were considered and a strided sliding window was applied to predict all voxels more efficiently. Moreover, to account for spatial correlation of neighbors voxels, MLP outputs had the same shape of the inputs so that each voxel prediction could be obtained by averaging a window around it. They demonstrated high accuracy and scalability of this model by predicting the location of neurons in a whole mouse cerebellum.

Afterwards, convolutional neural networks (CNN) [15, 28, 43, 45] have increased their popularity and have also been applied to image segmentation [48,64]. The UNet [64] was certainly the biggest breakthrough in modern image-to-image problems especially those related to biological images. Its encoder-decoder configuration together with skip connections between these two (see Figure 2.3) allowed the authors to train accurate models even with relatively small training-sets, thus motivating its popularity among scientists working with biomedical images.

In the context of cell segmentation from high-resolution microscopy images, worth to mention is StarDist [69, 83], a UNet-based model which exploits the a-priori knowledge of cell shape convexity and includes star-convex polyhedra as targets. By asking the CNN to predict both a probability map and the radial distances to the nearest cell center for those pixels with probability greater than zero, they demonstrate improved performances in detecting the marked nuclei, especially in crowded scenes. Following the CNN predictions, object approximation to star-convex polyhedra and non-maximum suppression are adopted in order to retrieve the most probable cell shapes and locations. Notably, they provide easy command line utilization of both 2D [69] and 3D [83] versions of their model.

In a similar fashion, CellPose [77] uses vertical and horizontal gradients to refine the shapes of predicted objects and improve differentiation between very close instances. However, unlike StarDist, the adoption of spatial gradients relaxes the shape convexity assumption, allowing the prediction also of non-convex shapes. Moreover, CellPose proposes various modifications to

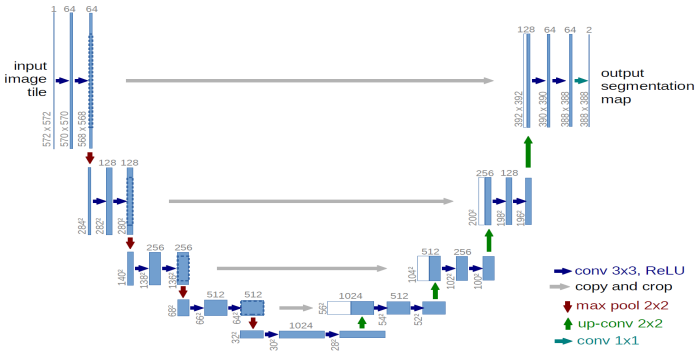


Figure 2.3: Schematic representation of UNet architecture. [64]

the original UNet architecture. Residual blocks [32] are used instead of the standard convolutional ones, an *image style* is extracted from the bottleneck of the UNet and then passed to subsequent decoding residual blocks, and encoder feature maps are added instead of concatenated to that of the decoder. By developing a new dataset of highly varied images of cells, they show high performances whether training and testing the model on specific types of data or training on all kind of available images and testing on specific data, demonstrating high transfer capabilities. Worth to notice that, up to date, CellPose is uniquely implemented in 2D and for 3D data relies on merging predictions over each plane view.

Chapter 3

BCFind-v2: Soft Semantic Segmentation for 3D Brain Cell Detection

In this chapter we describe a supervised algorithm for cell detection in 3D fluorescence microscopy images. We follow a two-step approach consisting of an initial soft-segmentation of raw images followed by a blob detector to extract cell coordinates. Soft-segmentation is used to train a fully convolutional network (FCN) from point annotations, a much easier labeling method compared to the highly time-consuming hard masks. This step serves to denoise the images and standardize cell shapes, luminosity and contrast. Localization is then performed by a blob detector based on the difference of two Gaussian kernels. A model-based hyper-parameter tuning is also adopted to automatically select the optimal parameters for the blob detector in order to facilitate its usage and improve the accuracy. Experimental results are provided, along with a comprehensive comparison of different FCN configurations. Results show that the our proposed technique is efficient and effectively detects neurons in highly variable and noisy brain images. [1]

[1] Part of this chapter has been published in "Universal autofocus for quantitative volumetric microscopy of whole mouse brains", *Nature Methods, 2021* [74].

3.1 Introduction

Accurate quantification of neurons in high-throughput microscopy images is one of the essential goals in neuroscience. Having an automatic method to quantify or even localize and segment cells is paramount for many biological laboratories. However large variability in the data and expensiveness of manual annotations are the main obstacles in developing flexible, accurate and automatic methods to do so. Many newly developed tools rely on highly customizable pipelines of classic computer vision algorithms [11, 62, 68] or adopt machine learning techniques with user-defined features [7, 75]. These approaches though, even if highly flexible and adaptable to many imaging techniques, struggle to generalize on highly variable intensities and sources of noise, requiring intensive hand tuning on small subsets of the data. More advanced deep-learning (DL) techniques have been proposed [69, 77], but they all rely on hard-mask labeling, a highly time-consuming annotation method. Here we present a two-step approach to localize cells from 3D fluorescence microscopy images using point annotations, a much faster annotation method that allows for the generation of larger data-sets with relative ease. Strongly inspired by the work of Frasconi et al. [23] we brought some improvements and additional features to the original implementation. Firstly, we replaced the two fully connected hidden layers neural network with a more advanced fully convolutional 3D UNet [13]. Multiple variations of the basic building block are explored and made available for data-specific needs and performances. In particular, users can choose among three different attention mechanisms [27, 36, 81], a residual convolutional block [32] and a mixture of UNets [39, 70] to fulfill their specific needs. Data augmentation is also included for better generalization capabilities. Secondly, coordinates extraction is performed by a fast and efficient GPU implementation of a standard blob detector based on multi-scale Gaussian kernel differences [50] in place of the original iterative mean-shift algorithm. We also equipped the blob detector with model-based automatic hyper-parameter optimization [8] for improved ease of use and performances. A schematic representation of the whole pipeline is depicted in Figure 3.1. We report results on two mice and one human brain dataset showing the effectiveness of the method and its adaptability to multiple contexts.

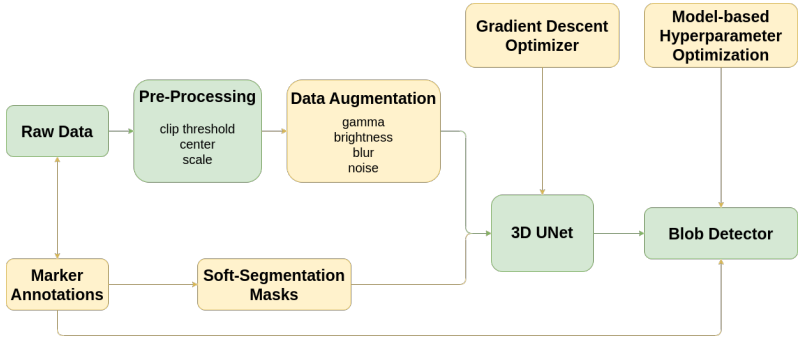


Figure 3.1: Pipeline of the proposed cell detection algorithm. Yellow boxes and edges are only needed for training.

3.2 Convolutional encoder-decoder: the UNet

Learning the probability that each pixel belongs to brain cells is the first step of our method. Given such a probability map it would be indeed easy for a blob detector to find center mass coordinates. Probability maps in fact would show only cells in a homogeneous way with any background noise removed. To this end we adopted the well-known UNet architecture [64], a fully convolutional network (FCN) that have shown remarkable performances on biomedical image segmentation problems. However, standing apart from standard semantic segmentation tasks which use hard-masks as labels, in our work segmentation is *soft*, i.e. rather than asking to precisely determine membrane voxels, the network is trained using Gaussian spheres as target. This "weak" supervision allows the adoption of point rather than pixel-wise annotations, much easier to obtain.

In our work we explored different UNet configurations, from a wide base implementation to the adoption of residual blocks for a much deeper network, the inclusion of attention mechanisms and the implementation of an ensemble model. The following subsections will give the details of all these variants.

3.2.1 Base model

Our base model follows the original implementation [64] with the straightforward adaptation to 3D data. The contracting path comprises four convolutional blocks, while the expanding path as many transposed convolutional

blocks. We may mutually refer to blocks belonging to the contracting path as encoding blocks, while blocks in the expansive path as decoding blocks. Each of them are however similarly composed by a single $3 \times 3 \times 3$ convolutional layer, using transposed convolution in case of decoding blocks, followed by batch normalization and ReLU activation. Due to the small size of cells, the last two encoding blocks do not apply downsampling operations, here implemented with strided convolution. Due to the high variability of pixel intensity, we decided to build a network as wide as possible by using a large number of initial filters (see Section 3.5.1 and 3.5.2 for the specific values adopted) and exponentially increase them by a factor of 2 every encoding block and symmetrically decrease them every decoding block. In order to output a probability map at the end of the network, a final convolutional layer with only one filter followed by sigmoid activation is applied.

3.2.2 Residual model

Residual learning was firstly proposed by He et al. (2016) [32] as a way to solve the vanishing gradient problem [6, 29, 35] for which deeper layers were difficult to train. Backpropagation in fact propagates small gradients of little activated layers through all the network, inducing very small or even null parameter updates. The solution presented in [32] was to allow the gradients to follow an alternative pathway that could skip those little activated layers and hence continue to regularly update the parameters. This was achieved by introducing an additive skip connection between the inputs and the outputs of each convolutional block. In mathematical notation, denoting with f_θ the learnable operations of a neural network building block, they did the following:

$$y = f_\theta(x) + x$$

that could also be written as

$$f_\theta(x) = y - x$$

from which the name of residual learning. Thanks to this trick, the authors were able to train an unprecedentedly deep neural network (152 layers) still reaching state-of-the-art results on image classification problems. Moreover, studying a method to visualize loss landscapes, Li et al. (2018) [46] found

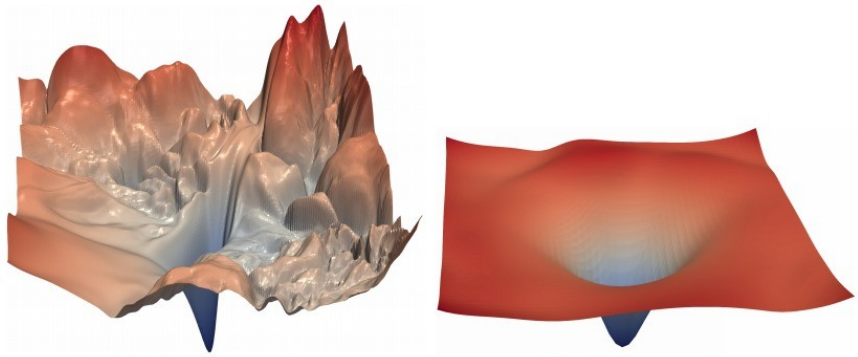


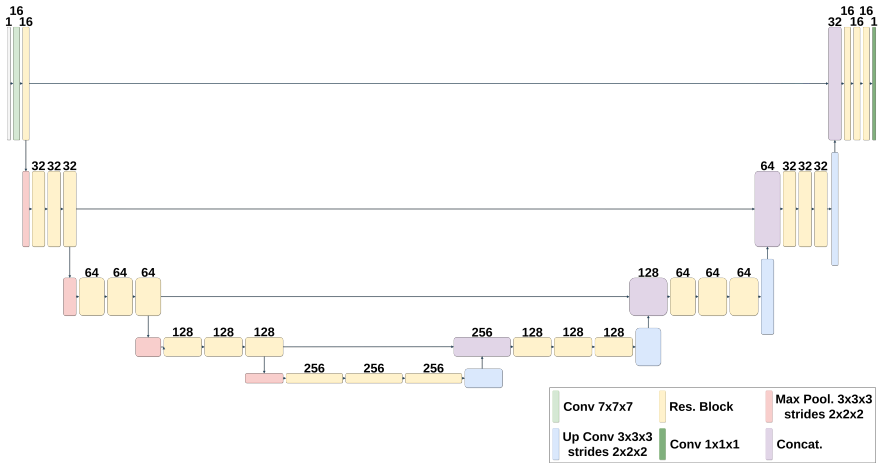
Figure 3.2: Loss landscape of plain vs residual neural network. [46]

that the above mentioned residual blocks actually prevent the loss from becoming chaotic and indeed maintain it smooth (Figure 3.2).

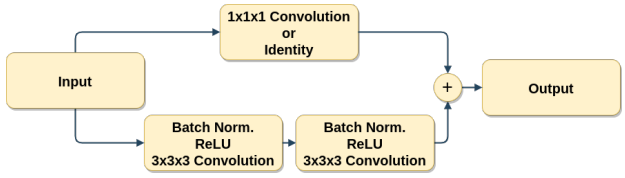
A residual version of the UNet is therefore implemented. This version adopts residual blocks in place of previously described convolutional blocks. Following the work of He et al. (2016) [33] we chose to adopt full pre-activation blocks (Figure 3.3b) and apply an initial $7 \times 7 \times 7$ convolutional block followed by $3 \times 3 \times 3$ residual convolutional block to the inputs. Since the high capacity of residual networks we decided to go deep too and apply three residual blocks in each contracting and expansive step while maintaining the same number of encoding and decoding blocks as the base model. Probability maps are here computed by a single $1 \times 1 \times 1$ convolutional layer with sigmoid activation, leading to a total of 52 layers. A representation of this architecture is depicted in Figure 3.3.

3.2.3 Attention models

Since their introduction in natural language processing (NLP) [78], attention mechanisms have received much consideration in the scientific literature and many adaptations to computer vision (CV) have been proposed [27,36,54,81]. The main contribution of these mechanisms is to model long-range interactions between input components to focus the most important ones. While in [78] and [27] attention is achieved by linearly combining each component accordingly to its correlation with the others, in [36] and [81] attention



(a)



(b)

Figure 3.3: Deep residual UNet as implemented in BCFind-v2. Whole architecture (a) and (b) pre-activated residual block.

weights are estimated by two fully-connected and a convolutional layer respectively (Figure 3.4). In CV there are two types of interactions that can be exploited: position-wise or channel-wise. In our work however we considered channel attentions only. This choice is mainly motivated by two reasons. On one hand, we believed that long-range spatial dependencies would not really help the detection of small objects that can appear everywhere in the image like cells do. On the other hand, being able to select really discriminative features and more importantly, being able to differentiate them across different inputs could be beneficial in wide networks with highly variable inputs. We therefore explored three different channel attention mechanisms: the squeeze-and-excite (SE) module [36], the efficient-channel-attention (ECA) [81] and the channel attention (CA) proposed by Fu et al. [27]. We also tried two different ways of including attention mechanisms into a UNet-like architecture. The first one adopts either the SE or ECA module into each block of the base model described in Section 3.2.1 in order to put more emphasis on the most important features of each convolutional layer. The other one instead employs the CA module in between the encoder-decoder skip connections, in a similar fashion to [54]. This latter in fact, uses the interactions between encoder and decoder features to better exploit the subsequent combination of the two. In particular, denoting as f_i the flattened encoder output at level i and with g_{k-i} the corresponding flattened decoder output at level $k-i$, where k denotes the total number of blocks in the UNet, our CA gate can be written as:

$$Q = linear(g_{k-i}), \qquad K = linear(f_i), \qquad V = linear(f_i)$$
$$Attn = softmax\left(Q^T \cdot K\right)$$
$$h_i = f_i + (V \cdot Attn)$$

h_i is then concatenated to g_{k-i} as usual.

3.2.4 Mixture model

Light microscopy imaging comes with lots of variability, either within the same acquisition (luminosity and contrast can drastically change in the Z dimension), or between different acquisitions and subjects. Modeling such diversity is therefore essential for the model to be accurate on all possible conditions. To do so we here treated the conditional distribution $P(Y|X)$ as

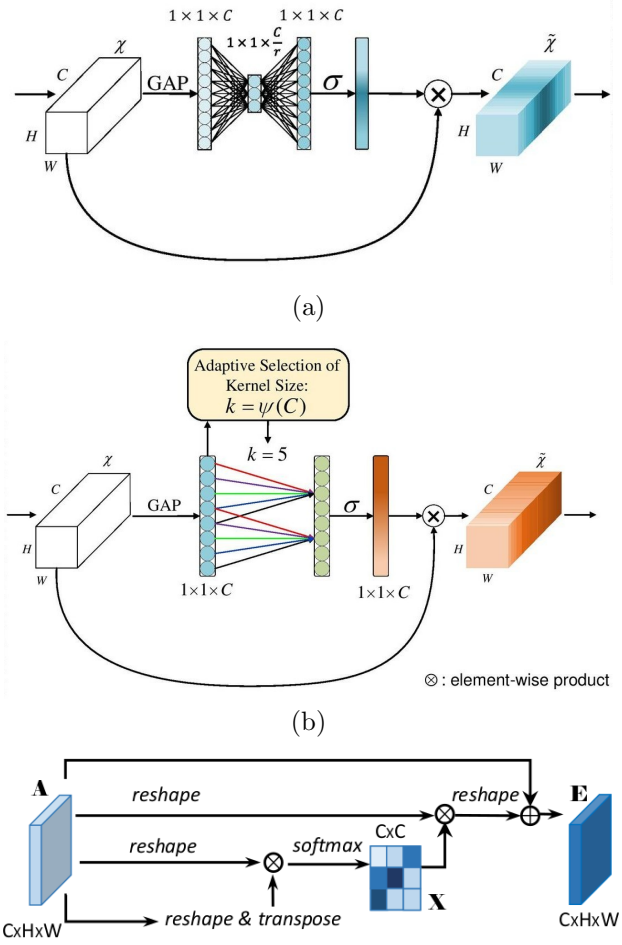


Figure 3.4: Attention modules implemented in BCFind-v2. (a) Squeeze-and-Excite [36], (b) Efficient Channel Attention [81] and (c) the channel attention proposed in [27].

a multimodal function and represented it with a mixture model. Following the seminal paper of Jacobs et al. (1991) [39], a fixed number of neural networks is depicted to model each unimodal distribution in the data, while a gating network estimates the mixing coefficients of the mixture. Denoting with g the gating network, with f_k the *expert* for the k^{th} component and with K the total number of considered components we can write the model as follows:

$$y = \sum_{k=1}^{K} g(x)^{(k)} f_k(x)$$

Thanks to this configuration each input is dispatched to one or more different experts which become more and more specialized on the given subset of the data. Even if training could be done by directly minimizing the difference between the linear combination of each expert and the target, this usually tends to distribute each input to many experts without really inducing a clustering of the data. Therefore a more convenient formulation of the loss would be if competition between experts is encouraged by making each expert learn the whole target rather than the residual of the others. In practice this is translated by defining the error as follow:

$$\mathcal{L}(x, y) = \sum_{k=1}^{K} g^{(k)}(x) || f_k(x) - y||^2$$

From this formulation, a smoother and better performing version of this same loss has been proposed in [39] and in fact this is what we also used:

$$\mathcal{L}(x, y) = -\log \left[\sum_{k=1}^{K} g(x)^{(k)} e^{\frac{1}{2} bce(f_k(x), y)} \right]$$

where *bce* denote the binary cross-entropy.

However, competitive training could lead to a self-reinforcing behavior in which the best performing expert is recursively selected and hence improved while the others are most of the time left apart. To avoid this trivial solution and encourage a more balanced expert selection, Shazer et al. (2017) [70] proposed an additional expert importance loss, defined as:

$$imp_k = \sum_x g(x)^{(k)}$$
$$\mathcal{L}^{imp} = \alpha \left[\frac{sd\left(imp_{k=1,\dots,K}\right)}{mean(imp_{k=1,\dots,K})} \right]^2;$$

and Fedus et al. (2022) [20] a load balancing loss:

$$load_k = \frac{1}{n} \sum_x \mathbf{1}\left[argmax(g(x)) = k\right]$$
$$\mathcal{L}^{load} = \gamma K \sum_{k=1}^{K} load_k \cdot \frac{1}{n} imp_k$$

Both were tested, however results in Section 3.5.2 refer to this latter, since we found inducing more balance.

The gating network is actually implemented as in [70] in which only the top J experts are selected and an additional learnable noise component is added to the predicted gate weights.

In our implementation we set $K = 5$, $J = 1$ and $\gamma = 0.2$. Moreover, to increase as much as possible the batch size and believing that even a simpler expert could effectively learn the within-cluster conditional distribution we select an initial number of filter equal to 32 and use the same architecture of the base model. Figure 3.5 graphically represents a mixture of UNet experts.

3.2.5 Data augmentation

Data augmentation is a technique to artificially enlarge the training set by randomly transforming existing data. This is particularly useful in biological images where labels are scarce. Moreover, data augmentation serves also as a regularizer and in contexts with small training-sets overfitting is particularly easy to happen. For all these reasons and through extensive experimentation we also included it in our pipeline. In particular, we found the following transformations particularly useful:

- Gaussian noise: $\sigma \in (0, 0.03)$
- Gaussian blur: $\sigma \in (0, 0.5)$

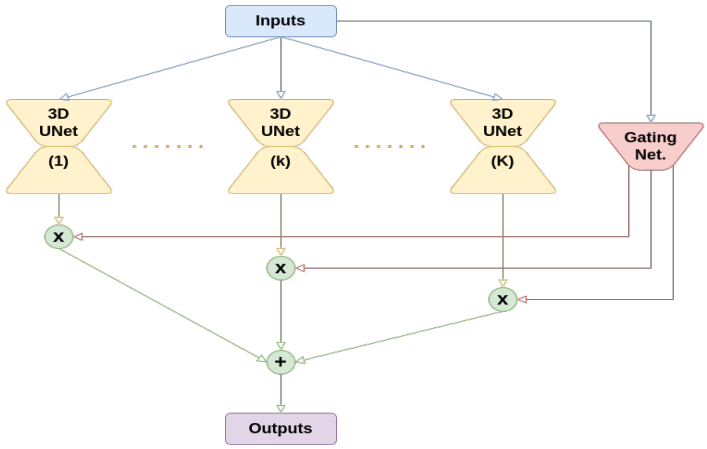


Figure 3.5: Mixture of UNets

- gamma: $\gamma \in (0.8, 1.2)$
- brightness: $\delta \in (-0.06, 0.06)$

Images are always rescaled in the interval $(0, 1)$ before transformation. Above augmentations are applied in random order, all with a probability of 0.4.

3.2.6 Point annotations and soft-masks

In order to train a segmentation model from point annotations, soft-masks are generated by placing a Gaussian sphere on each annotated coordinate. This method is particularly suited for light microscopy data in which only the nuclei are stained and cell membranes are not clearly recognizable but on the contrary, are rather blurry. The width of each Gaussian is selected in order to avoid overlaps and hence depends on the distance to the nearest cell. Starting with a default value of 3.5 for the σ, if the lowest distance is lower than 3.5^2 it will be decreased to the ratio of the distance and 3.5. Anisotropy is also considered by rescaling it on the dimensions with lower resolution. Soft-masks are then rescaled by their maximum value, ranging in the interval $(0, 1)$.

3.3 Blob detection with difference of Gaussians

Blob detection is the second step of our model: once the original images are transformed into probability maps through the UNet, cell coordinates are extracted by a blob detector based on the difference of two Gaussian kernels [50]. Such kernel is in fact an approximation of the Laplacian of a Gaussian whose shape is particularly suited to enhance blob-like objects. We chose to adopt this technique, in place of the original mean shift [23], because of its faster convolutional operations (computed on the GPU in our implementation) compared to the iterative mean shift algorithm and for its recognized performances on detecting (spherical) objects at multiple scales [5, 47, 52, 62, 72]. The algorithm works by blurring the input image with multiple Gaussian kernels whose σ increases by a fixed factor. Afterwards, consecutive filtered images are subtracted to obtain a transformation equivalent to the application of a kernel representing the difference of two Gaussians. Once the spherical objects are thus highlighted, a maximum filter retrieves the local maxima of the image and hence is able to return our desired cell coordinates.

3.3.1 Model-based hyperparameter optimization

In our experiments we found that selecting the optimal parameters is however not that straightforward and different values could lead to very different results in terms of F1-score. We therefore chose to make the selection automatic both obtaining more accurate results and making the blob detector usage easier. To do so a model-based hyperparameter optimization algorithm is adopted. In particular we used the tree-structured Parzen estimator (TPE) [8] to select the parameters that maximize the expected improvement (EI) on a surrogate function of the F1-score. Denoting with y the observed value of the objective function f, with y^* some quantile γ of the y values and with θ the set of parameters to be optimized, the TPE actually models $p(\theta|y)$ as follow:

$$p(\theta|y) = \begin{cases} l(\theta) & \text{if } y < y^* \\ g(\theta) & \text{if } y \geq y^* \end{cases}$$

where $l(\theta)$ is the non-parametric density estimated from those observed $\theta^{(i)}$ whose $f(\theta^{(i)}) < y^*$ and $g(\theta)$ the density formed by using the remaining

observations. By using this parametrization, Bergstra et al. (2011) [8] found that

$$EI_{y^*}(\theta) \propto \left(\gamma + \frac{g(\theta)}{l(\theta)}(1-\gamma)\right)^{-1}$$

and hence, thanks to the adopted forms of l and g (details in the paper [8]), it would be easy to draw many samples from l and evaluate them according to $\frac{g(\theta)}{l(\theta)}$ to maximize the EI.

3.4 Evaluation metrics

Due to the object-wise nature of our annotations, models are evaluated by their capability of finding those objects. If our labels are points, those same points should be ideally our predictions. Considered metrics therefore look at the correct matches between predicted and annotated coordinates. Unique matching is firstly computed through Hungarian algorithm [44], a method to solve unique assignment problems by mean of matching distance minimization. Predicted cell coordinates are therefore matched to their closest annotation. We then classified the predictions within a distance d to their matched annotation as true positives (TP) and as false positives (FP) otherwise. On the other hand, annotations whose assigned prediction is further than d are denoted as false negatives (FN). Considered metrics are then defined as follow:

$$\begin{aligned} Precision &:= \frac{|TP|}{|TP| + |FP|} \\ Recall &:= \frac{|TP|}{|TP| + |FN|} \\ F1 &:= 2\frac{Precision \cdot Recall}{Precision + Recall} \end{aligned}$$

where $|A|$ denotes the cardinality of set A. Compared to pixels-wise or correlation-based [55] metrics we found these criteria better suited for point annotations. On one hand, pixel-wise metrics tell us if shapes are well predicted, but our UNet targets have just proxy shapes which do not necessarily reflect reality and are not biologically accurate. Evaluating a model from these approximated shapes would be in fact biologically misleading. On the

other hand, we find correlation too much coarse and inaccurate. Firstly, it considers counts only and not how close predicted cells are to the annotated ones. Secondly, even very different counts can be correlated, in fact, correlation does not tell us anything about the actual values of compared variables. Paradoxically, if a model consistently obtains equal but very low values of precision and recall, it will predict counts perfectly correlated with, even identical to, the ground truth, but highly wrong coordinates.

3.5 Experimental results

Models have been tested and applied on two main data-sets, kindly provided by two neuroscientific groups of LENS. For the experiments, the data-sets have been randomly split in three non-intersecting sets: one for training (70%), one for validation (10%) and one for test (20%). UNet models have been trained to minimize the binary cross-entropy loss between predictions and soft-masks of training-set volumes, however weights are only saved when the validation loss improves. DoG parameters are then optimized on the UNet predictions of validation volumes. Reported results refer, of course, to the test-set. To train the UNet models we found particularly effective the adoption of stochastic gradient descent (SGD) optimizer with Nesterov momentum [53] set to 0.9 and the cosine decay with warm restarts [49] schedule for the learning rate. All UNets have been trained for 3000 epochs with an initial learning rate of 0.1. While DoG optimization has been carried out for 50 TPE steps.

3.5.1 Test on mice brain tissue

We here test the base model configuration (Section 3.2.1) on two different mice datasets whose brain cells were stained with different reagents. SST^+ neurons dataset comprises 327 volumes of $312 \times 312 \times 320 \mu m^3$ each at a resolution of $0.65 \times 0.65 \times 2.0 \mu m^3$ per voxel with a total of 28264 manually annotated cells. On the other hand, $cFos^+$ neurons dataset comprises 278 volumes of shape $312 \times 312 \times 320 \mu m^3$ each at a resolution of $0.65 \times 0.65 \times 2.0 \mu m^3$ per voxel with a total of 20952 manually annotated cells. Some examples of raw volumes and corresponding predictions for both datasets can be seen in Figure 3.7.

Staining	Model	Prec. (%)	Rec. (%)	F1 (%)
SST	UNet [74]	83.0	90.0	86.0
cFos	UNet [21]	84.0	74.0	78.0
	SE-UNet	75.1	73.3	74.2
	ECA-UNet	74.8	71.3	73.0

Table 3.1: Localization metrics on different mice brain datasets.

3.5.2 Test on human brain tissue

Here we compare all described UNet variants applied to images of NeuN$^+$ neurons in the human brain. This dataset comprises 54 volumes of $360 \times 360 \times 180 \mu m^3$ at a resolution of $3.6 \times 3.6 \times 3.6 \mu m^3$ per voxel with a total of 26596 annotated cells. Some examples of raw volumes and corresponding predictions for the Res-Unet configuration (Section 3.2.2) can be visualized in Figure 3.7. While Figure 3.6 shows the MIP of some volumes taken from the clusters induced by the Mixture of Unets (Section 3.2.4).

Model	Prec. (%)	Rec. (%)	F1 (%)
UNet	68.3	75.9	71.9
SE-UNet	70.5	79.1	74.6
ECA-UNet	71.2	**81.1**	75.8
Attn-UNet	75.1	76.3	75.7
Res-UNet	79.0	76.0	**77.4**
MoUNets	**80.8**	69.5	74.7

Table 3.2: Localization metrics for various UNet configurations on human NeuN$^+$ neurons data. Bold values highlight the best model for each metric.

3.5.3 Discussion

BCFind-v2 has been trained and validated on both mouse and human brain data at two different pixel resolutions ($0.65 \times 0.65 \times 2.0 \mu m^3$ for the mouse and $3.6 \times 3.6 \times 3.6 \mu m^3$ for the human). While on both mouse datasets we achieved satisfactory results already with the base UNet, on the human tissue dataset a wider architectural exploration has been deemed necessary.

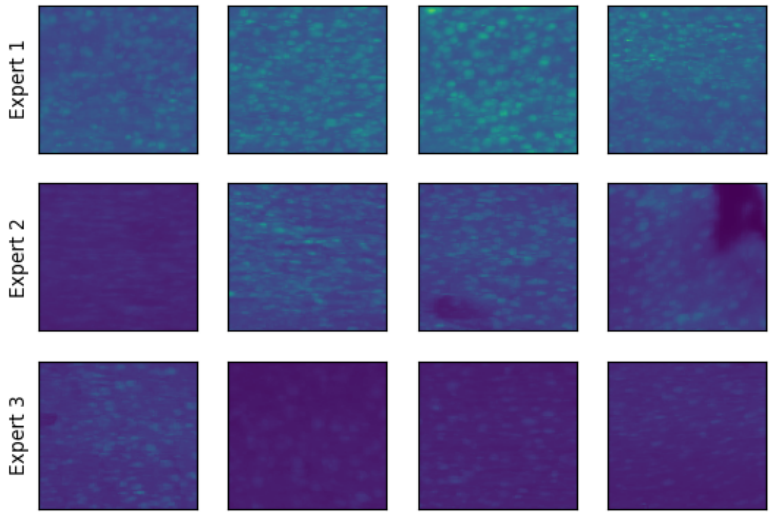


Figure 3.6: Clusters induced by Mixture of UNets. Images are the MIP over whole volume depth.

In fact, the lower spatial resolution of human brain images, the higher cell density, the lower contrast of the cells compared to the background and the less clear cell boundaries (Figure 3.7) added difficulties to the detection task. Six different CNN have been therefore compared trying to overcome such increased complexities. As a result, the best model improved the F_1-score by 5.5 points with respect to the base UNet model.

Model architecture has however been found to be data-specific, indeed the SE-UNet and the ECA-UNet had better results than the UNet on the human brain, but worse on the cFos mouse brain dataset.

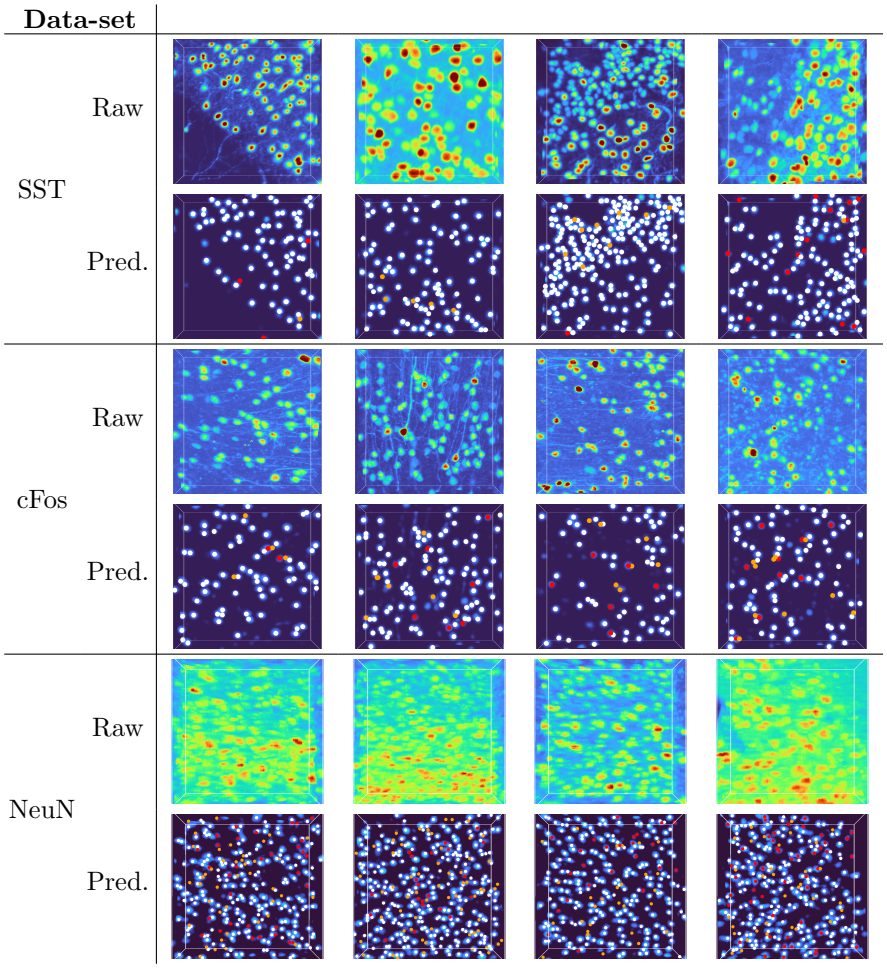


Figure 3.7: Raw and predicted volumes for each considered data-set. Predictions for human NeuN$^+$ neurons refer to Res-UNet model, while for mice SST$^+$ and cFos$^+$ neurons to the base UNet model. White dots represent true positive, red dots false positive and orange dots false negative detections.

Chapter 4

BRANT: Brain-Wide Neuron Quantification Toolkit to Study Fear Memory in Mice

In this chapter we introduce brain-wide neuron quantification toolkit (BRANT) for mapping whole-brain neuronal activation at micron-scale resolution, combining tissue clearing, high-resolution light-sheet microscopy, and automated image analysis. A detailed knowledge of the neural circuitry modulating fear memory could be the turning point for the comprehension of this emotion and its pathological states. A comprehensive understanding of the circuits mediating memory encoding, consolidation, and retrieval presents the fundamental technological challenge of analyzing activity in the entire brain with single-neuron resolution. The tool presented in this chapter allows for robust and scalable quantification of activity patterns across multiple phases of memory in mice. The methodology presented here paves the way for a comprehensive characterization of the evolution of fear memory. [1] [2]

[1] This chapter has been published as "Brain-wide neuron quantification toolkit reveals strong sexual dimorphism in the evolution of fear memory" in *Cell Reports, 2023* [21].

[2] *Acknowledgments:* this work was largely developed in collaboration with and supported by the European Laboratory for Non-Linear Spectroscopy (LENS), Sesto Fiorentino (FI), Italy.

4.1 Introduction

Fear responses are functionally adaptive behaviors that can be induced by a direct encounter with a threat or with situations previously associated with a threat. Fear induces many changes at different levels, from molecular and cellular to circuit ones [40, 41]. Typically specific brain areas, as the hippocampus, the amygdala and the prefontal cortex, are identified as the centers of memory processing. However, several studies highlighted the involvement of many other regions [59, 73], supporting the hypothesis that memory is distributed and dispersed across the entire brain. Unfortunately, available analysis tools [62] are not capable of routinely handling Tera-bytes sized datasets,limiting brain-wide activation analysis to anti-cFos immunostaining. Here, we present BRANT (brain-wide neuron quantification toolkit), a new pipeline for whole-brain mapping, exploiting TRAP mice [30], high-resolution light-sheet microscopy (LSM), and terabyte-scale image processing. Using BRANT, we analyze the evolution of whole-brain neural circuits recruited upon aversive memory in 14 females and 15 males, allowing for brain-wide cohort study. In order to study the evolution of fear memory, mice learn to associate a particular context (i.e. black box) with an aversive event (i.e. mild foot shock). The latency time to enter the the dark compartment is used as a direct measurement of memory. We selected three experimental groups based on three different memory phases describing the evolution of fear memory, from encoding to retrieval: a training group was selected to study fear encoding, while test groups at 24h and 7 days after training were selected to explore the recent and long-term fear memory retrieval, respectively. As expected, statistical analyses revealed significant differences of recorded latency times between these experimental groups. Moreover, after brain atlas registration through advanced normalization tools (ANTs) [3], cell densities have been estimated on 48 brain regions and partial least squares (PLS) [42] analysis has been performed to evaluate the activation pattern differences between males and females revealing gender-specific cFos expressions. Widely accepted theories also affirm that memory is distributed across multiple brain regions that are functionally connected [22, 65]. Therefore a functional connectome for each experimental group has been estimated and network statistics, as the nodes degree, nodes betweenness and small-word coefficient, have been computed. Results confirm once again the evolution of activation patterns over time and their differences between males and females.

4.2 BCFind-v2 on large-scale analysis

Working with TB-sized datasets is a complex task that requires a well-defined and efficient pipeline starting from data storage and loading to the analysis process. Each brain reconstruction comprises about 16 Terabytes of raw data with a voxel size of $0.65 \times 0.65 \times 2.0 \mu m^3$. This data size is incompatible with a cohort study, as it would require storage capabilities in the order of 1 Petabyte for a single study. For this reason, the acquired datasets were first compressed by a factor of 20 using the 16-bit lossy JPEG-2000 format, thus reducing disk usage while still retaining overall good image quality and detail level. Images were then stitched using ZetaStitcher (https://lens-biophotonics.github.io/ZetaStitcher/), a custom-made Python software developed at LENS for large volumetric stitching specifically developed for LSM. An important feature of ZetaStitcher is VirtualFusedVolume, an application programming interface (API) that provides seamless and effective access to high-resolution data by simply providing the spatial coordinates of the subvolume of interest within the virtually fused volume. In this way, large volumes can be programmatically processed in smaller chunks in a distributed environment and without user intervention, a key requirement to process the large datasets produced by high-resolution LSM. Automatic cell detection is achieved using BCFind-v2 (see Chapter 3) whose components (UNet and blob detector) have been separated in order to make them run in parallel. In particular, we used a queue to store multiple chunks of raw data coupled with a single threaded worker to run UNet predictions, and a second queue to store the UNet predictions coupled with a multiple threaded worker to retrieve and pass them to the blob detector for cell coordinates extraction. It was also on this occasion that we moved from CPU to GPU implementation of the blob detection. Efficient GPU memory allocation, both for the UNet and the blob detector, was also essential to avoid undesired overheads. Such implementation allows for fast and scalable analysis, reaching an average inference time of 240Gb/h, about one order of magnitude faster than the reported speed of ClearMap [62] for datasets of this size (Figure 4.1). BCFind-v2 predictions are done on overlapping substacks of the whole volume and then merged by removing those coordinates inside a frame with half of the overlap size. Figure 4.2 shows the axial, coronal and sagittal views taken from a predicted point cloud of a whole mouse brain.

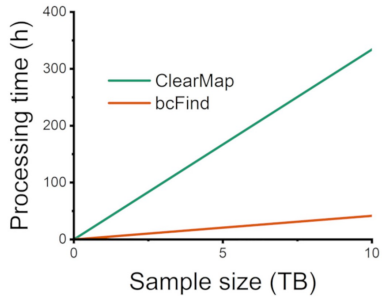
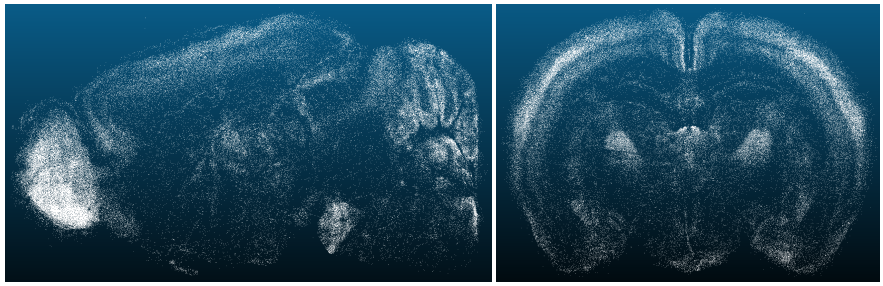
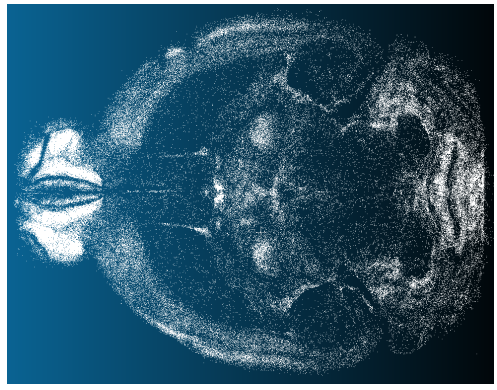


Figure 4.1: Inference speed of BCFind-v2 and ClearMap on whole mice brain analysis.



(a) (b)



(c)

Figure 4.2: Brain-wide BCFind-v2 inference. (a) Sagittal, (b) coronal and (c) axial views of predicted mouse brain point cloud.

4.3 Predicted densities differentiate memory phases and gender

By looking at latency times and predicted densities in each of the 48 brain regions considered we first studied correlation between neuronal activation and memory. As expected, the amygdala, the hippocampus, and the prefrontal cortex are correlated with the step-through latency times, but also, the activation of areas such as the pallidum, the striatum, the pons, and other regions less known to be involved in fear memory correlates with the time mice spent in the bright cage before stepping through into the dark compartment (where the aversive event happens). Moreover and importantly, brain regions found to be correlated with behavior are different for both sexes, underlining a sexual dimorphism (Figure 4.3a). A result confirmed also by PLS analysis. PLS is a statistical technique mainly used in high-dimensional contexts (i.e. the number of variables is much higher than the number of observations). Variables are firstly projected in a lower dimensional space whose components maximize the correlation with the response variable. Such projection allows for standard linear regression, previously unfeasible due to the *curse of dimensionality*, but also for evaluating the contributions that each variable has on the components used for the actual regression. Therefore, by looking at the contributions on those components that maximally correlate with the response variable we can measure the importance that each variable has on the predictions. Considering sex as response variable and brain region densities as predictors, the analysis highlighted significant gender specificity of some region densities. Those regions were also found to change between different memory phases. While the hippocampus and the amygdala are the areas that most differentiate males and females during training, in both 24h and 7d tests is the prefrontal cortex that gains more importance (Figure 4.3b-d).

4.4 Network analysis of cell densities

Whole-brain connectivity analysis is fundamental to understand the complex interactions between functionally coherent regions and their role in the development of fear memory. We therefore considered the strongest cross-correlations ($p < 0.05$) between the predicted brain region densities and built a whole-brain functional network for each experimental group (Figure 4.4).

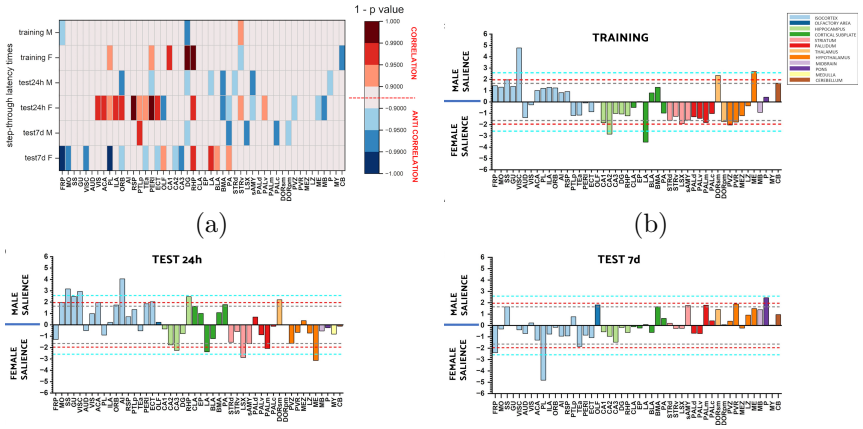


Figure 4.3: Neuronal activity analysis. (a) Pairwise correlation of latency time with neuronal density. (b-d) Standardized PLS contributions of each brain region density divided by experimental group. The gray, red, and blue lines reflect, respectively, contribution scores of 1.64 ($p < 0.1$), 1.96 ($p < 0.05$), and 2.58 ($p < 0.01$).

Quantitative comparisons of the estimated networks are then carried out by extracting global and local connectivity features. Density, ρ, of a graph is defined as the probability of observing an edge between two random nodes. From this statistic we note the high connectivity of females network on the 24h test ($\rho = 0.1285$), much higher than that of males ($\rho = 0.0585$). If we restrict this statistic to positive vs negative correlations we find even higher differences, with females that tend to increase the probability of positive correlations on the 24h test ($\rho^{+}_{train} = 0.0284$, $\rho^{+}_{t24h} = 0.0975$) while males do the opposite ($\rho^{+}_{train} = 0.0629$, $\rho^{+}_{t24h} = 0.0346$). On the other hand, negative correlations remain similar across sexes and memory phases, with the only exception of few negative links in training males. Removing isolated components, we also look at small-world coefficient

$$\sigma = \frac{C/C_r}{L/L_r},$$

where the subscript r refers to a random graph, C to the clustering coefficient and L to the average shortest path length. This statistic indicates that most nodes can reach any other node with a small number of steps. Specifically a small-word network is defined to be a network where the average shortest distance between random nodes grows proportionally to the logarithm of the total number of nodes, i.e. is characterized by high average clustering coefficient and short average shorted path length. By sampling 100 random graphs, we found that whether males network shows small-world features in all memory phases, as also found in [84], conversely, females network at 24h test largely resembles a random graph, suggesting different reorganization pathways between males and females.

Moreover, looking at local statistics such as the node degree and the betweenness, we were able to select central nodes (hubs) in each memory phase, highlighting those regions responsible for controlling the exchange of information. Defining as hubs those regions above the 80^{th} percentile of both degree and betweenness distributions, we found a cortical transition from sensorimotor cortices to associative areas in males. A finding in line with standard theory of memory consolidation [38]. Conversely, in female subjects, network hubs persist in subcortical regions across the entire time span investigated, encouraging the sexual dimorphism findings of this work.

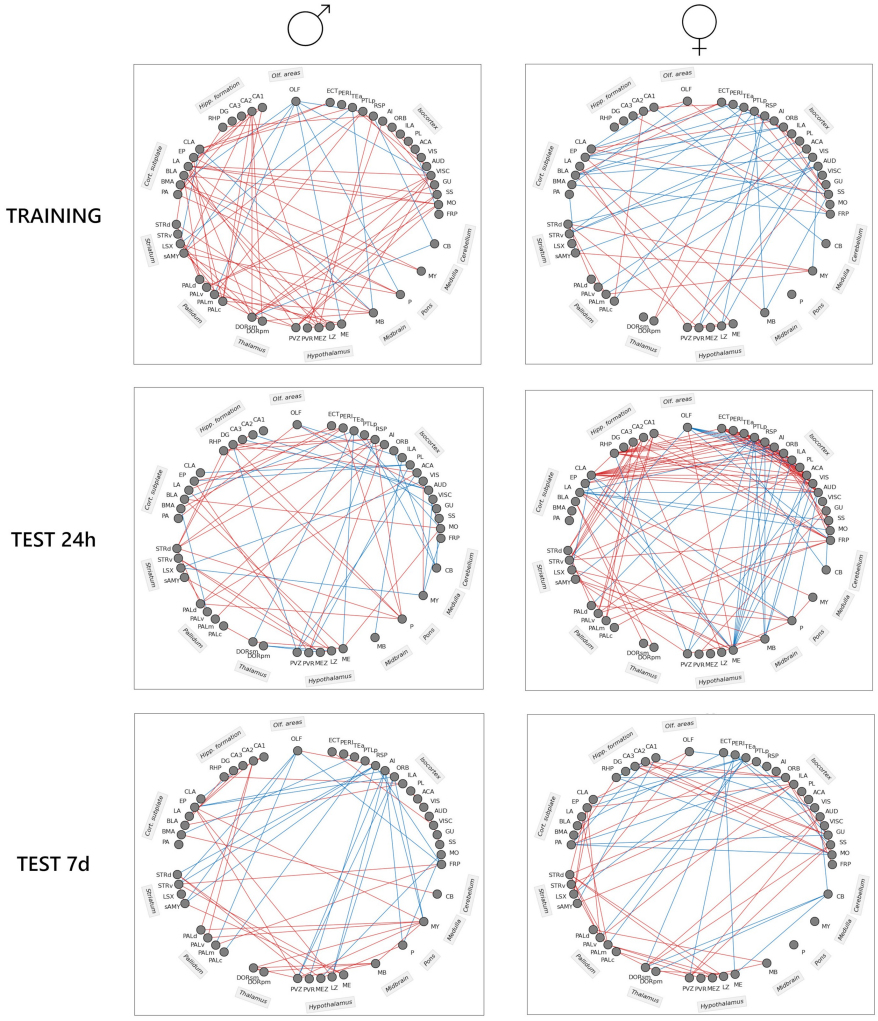


Figure 4.4: Fear memory networks in male and female mice. Gray circles represent the 48 selected regions, while lines the significant ($p < 0.05$) positive (red) or negative (blue) correlations.

4.5 Conclusion

The collaborative effort between scientists of multiple disciplines allowed the definition of a well-defined high-throughput pipeline able to acquire multiple whole-brain high-resolution images, efficiently store them and retrieve them, obtain accurate brain-wide cell locations, align the predicted point clouds to a reference atlas and finally extract brain region counts for in-depth study of brain functions. Thanks to all the adopted steps we confirmed previous studies and suggested novel research directions for better understanding fear memory.

Chapter 5

The Human Broca's Area: Comparing Automatic Detections with Stereological Estimates

This study presents a comprehensive comparative analysis of three deep-learning (DL) methods (BCFind-v2, StarDist and CellPose) and stereology for neuron quantification in the human Broca's area. We firstly evaluate the ability of DL methods to correctly detect cell coordinates both on unseen brain slabs and on annotations made by two different groups of experts with two different software. Then, we look at the predictions on whole brain slices, their visible quality and the comparison between densities predicted by DL models and those predicted by stereology. Inference time of DL methods is also taken into account and evaluated. Lastly, we discuss about the manual annotation effort each method requires and the granularity of the information it retrieves. [1] [2]

[1] Part of this work was conducted while the author was a visiting Ph.D. student at the European Molecular Biology Laboratory (EMBL), Heidelberg (Germany), from September to the end of December 2022 (working with. Dr. Anna Kreshuk).

[2] The work presented in this chapter is ready for submission with the title "Stereology or Deep-Learning? On the reliability and extrapolation power of deep-learning methods applied to large-scale human brain tissue" to *Scientific Reports*.

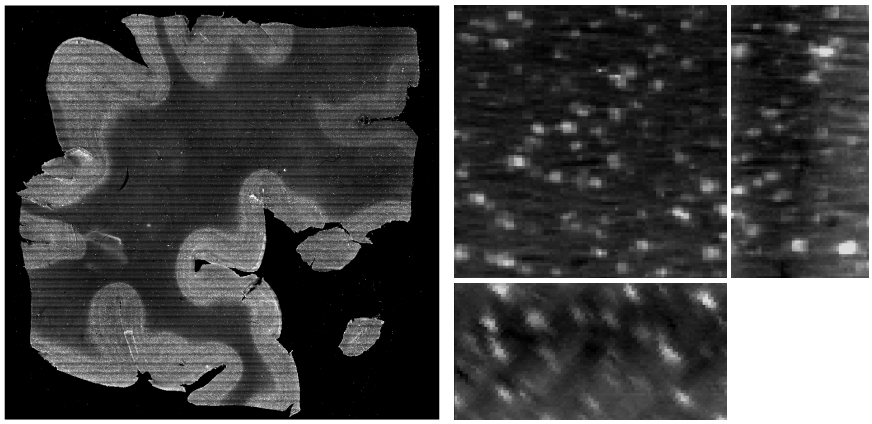
5.1 Introduction

The accurate quantification of neurons in specific brain regions is of utmost importance for understanding the intricate organization and function of the human brain. Although comprehensive cell census in animal models are now possible [51], there is no such correspondence in the human brain due to unavoidable distortions introduced by the slicing, clearing and staining processes of actual imaging protocols. This study uses an innovative high-throughput LSFM-based imaging pipeline [18] that allows for subcellular visualization of the entire Broca's area of a human brain with minimum distortions. This cutting-edge imaging technique has therefore made it possible to apply the most modern image analysis tools and hence obtain information details previously unavailable. However, the images thus obtained are still characterized by high brightness and contrast variability and poor signal-to-noise ratio (see Figure 5.1). Our comparative analysis evaluate the possibility of applying DL models on this difficult yet fundamental, domain.

To properly validate the automatic methods, we not only verify localization metrics on a (small) test-set, but we also compare large-scale predictions with stereology. The latter, relying on human counts and systematic sampling procedure, produces unbiased estimates and in fact, it is considered the gold standard for neuronal counts estimate on such complex and large-scale data. In fact, if properly validated, DL models are able to extract much more granular information (i.e.individual cell coordinates or even segmentation) and, unlike stereology, are completely data-driven allowing the researchers to relax a-priori assumptions and possibly highlighting subregional anomalies, undetectable by simple stereology.

Unlike previous studies [2,55] that validate a single model and mainly rely on predicted counts and their correlation with ground truth annotations, we take into account three automatic DL methods (BCFind-v2, see Chapter 3, the 3D implementation of StarDist [83] and CellPose [77]) and evaluate them on multiple facets. We firstly test the models on their capability of correctly locating cell coordinates, their main contribution with respect to standard stereology, with a particular focus on testing their predictions on unseen brain slices, arguably the most difficult task on this kind of data. Then, we look up at whole slice predictions, both visual inspection and direct comparison with stereological density estimates are considered, discussing both qualitatively and quantitatively their reliability on large-scale extrapolations. Later, since dealing with huge amount of data, inference speed is considered.

Finally, we evaluate the human effort required by each method related also to the granularity of information retrieved.



(a) (b)

Figure 5.1: LSFM imaging of the Broca's Area of a human brain. (a) Maximum intensity projection (MIP) of an entire slab, $482.4 \times 38775.6 \times 44197.2 \mu m^3$. (b) The axis projections of a smaller volume, $180 \times 360 \times 360 \mu m^3$.

5.2 Cell localization on multiple brain slabs

To acquire the human Broca's area with LSFM the tissue has been sliced into 48 400μm-thick sections (in the text referred as slabs or slices interchangeably) which, even if they underwent identical treatment, reveal strong variability (see histograms in Fig. 5.3). Testing models' resiliency to this changes is therefore essential to evaluate their reliability in this common scenario in human brain LSFM 3D reconstruction.

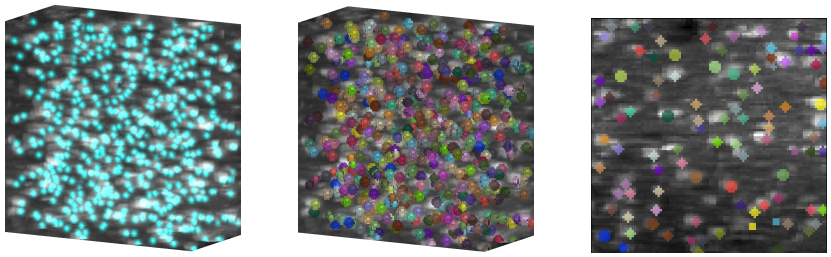
Available volumes for training DL-models were derived from six distinct slices (1–6), therefore we carried out a leave-one-slab-out cross validation procedure where each fold corresponds to all volumes belonging to a specific brain slice. We here evaluate the models on the correctness of predicted locations looking at object-wise metrics (see Section 3.4). Table 5.1 reports mean and standard deviation of precision, recall and F1-score on this 6-fold

experiment.

Hard-masks for StarDist and Cellpose are generated by thresholding the soft-masks used by BCFind-v2 at 0.06. Moreover, to adapt the 2D nature of CellPose to our 3D data we trained this model on 9 XY-planes MIPs of the original volumes. An example of inputs and targets given to StarDist and CellPose for training is depicted in Figure 5.2. During inference we did not use the 3D adaptation proposed by the authors due to a distortion on XZ-planes coming from the 45° inclination of the microscope (see Figure 5.1b) which is not properly taken into account by the adopted spherical masks. Moreover, their proposed method for 3D prediction would require three forward passes of the whole model, greatly increasing the inference time (see Section 5.4). We therefore only predict on XY slices and subsequently merge the predicted instances of adjacent planes that have an intersection-over-union higher than 0.5. Coordinates are finally extracted by taking the mean location of each instance. Hyperparameters are kept as close as possible to the official implementations, only adding a dropout rate of 0.3 for StarDist and setting the average cell diameter to 5 pixels for CellPose.

As we can see from Table 5.1, BCFind-v2 achieves the highest mean recall and F1-score, however due to the high standard deviations this model cannot be uniquely identified as the best, especially on the recall metric. StarDist in fact, obtains similar F1-score, but with higher precision and lower recall. On the other hand, Cellpose shows the lowest F1-score (with 99% confidence interval in $[33.2, 58.0]$), mainly driven from the very low recall, a result we also share with Oltmer et al. (2023) [55]. We however need to be aware of the challenge CellPose faced by using 2D hard-segmentation targets obtained from 3D point annotations.

We then compared models' predictions with the annotations made for stereology, here available only for brain slices 6, 18, 30 and 42. In fact, if DL methods could accurately predict stereological annotations, their predictions would obtain near identical stereological estimates, therefore any differences in predicted densities should be explored more in detail. Worth to notice that these annotations were made by a different group of experts with a different software specifically designed for stereological purposes (Stereo Investigator® by MBF bioscience). To take into account the peculiarity of stereological annotations which exclude cells overlapping the bottom-left border of the annotation box, we used the predicted coordinates on a 3 pixels top-right shifted box. Overall (last row of Table 5.2), BCFind-v2 achieves



A **B** **C**

Figure 5.2: Examples of input-target pairs adopted in (A) BCFind-v2, (B) StarDist and (C) CellPose training.

Method	**Prec.** (%) (st. dev.)	**Rec.** (%) (st. dev.)	**F1** (%) (st. dev.)
BCFind-v2 (ResUNet)	81.2 (5.8)	74.7 (8.9)	77.4 (4.2)
StarDist (ResNet)	85.1 (3.5)	67.9 (6.2)	75.3 (3.7)
StarDist (UNet)	85.3 (5.7)	67.4 (11.7)	74.5 (6.4)
CellPose	79.7 (3.7)	32.4 (6.2)	45.6 (5.9)

Table 5.1: Localization metrics on LENS annotations. Mean and standard deviation of precision, recall and F1 metrics on 6-fold leave-one-slab-out training procedure. Here only volumes available to train DL-models, taken from slabs 1–6, are considered.

best F1-score and recall, with StarDist (ResNet) not too far behind. CellPose, again, still struggles in detecting all cells, yet improving from the previous test, however it now reaches the highest precision. From this experiment we mainly note two facts: StarDist and BCFind-v2 predictions experience a sharp drop on the recall metric on Slab 18; the recall metric is almost always (except for Slab 18 and CellPose predictions) higher than the precision metric, an inverted relationship compared to Table 5.1. While the first one can be better understood by looking at Figure 5.3, the second one we speculate could be caused by a more conservative annotation approach

adopted by stereological experts compared with the more comprehensive one adopted at LENS. A conjecture corroborated also by the substantial recall improvement of CellPose predictions.

Slab n.	Tot. markers	Method	Prec. (%)	Rec. (%)	F1 (%)
6	379	BCFind-v2 (ResUNet)	69.0	**84.4**	75.9
		StarDist (ResNet)	74.7	79.7	**77.1**
		StarDist (UNet)	**81.5**	70.7	75.7
		CellPose	**81.5**	36.2	50.1
18	746	BCFind-v2 (ResUNet)	76.9	**68.7**	**72.6**
		StarDist (ResNet)	76.5	61.8	69.6
		StarDist (UNet)	**81.2**	39.9	53.5
		CellPose	78.9	47.0	58.9
30	626	BCFind-v2 (ResUNet)	72.0	81.6	76.5
		StarDist (ResNet)	73.6	81.8	77.5
		StarDist (UNet)	**77.0**	**81.9**	**79.4**
		CellPose	76.1	56.4	64.8
42	494	BCFind-v2 (ResUNet)	76.4	82.6	79.4
		StarDist (ResNet)	74.8	**84.6**	79.4
		StarDist (UNet)	**78.0**	81.2	**79.6**
		CellPose	83.8	47.2	60.4
Tot.	2245	BCFind-v2 (ResUNet)	73.8	**78.0**	**75.8**
		StarDist (ResNet)	75.6	75.4	75.5
		StarDist (UNet)	78.9	65.9	71.8
		CellPose	**79.3**	47.8	59.7

Table 5.2: Localization metrics on the annotations made for stereology, grouped by slab. Bold values are the per slab best metrics. Stereological estimates have been done on these four slices only.

5.3 Large-scale predictions

Enlarging the view, but decreasing the granularity of performance metrics, we here look at predictions on whole brain slabs, those in which also stereology has been applied: slab 6, 18, 30 and 42. We would like to point out that this work is the first one, to our knowledge, comparing so many methods on such large-scale data.

The human cortical Broca's area is organized into five main layers, parallel to the surface of the brain, which differentiate each other by the size, shape and function of the neuronal bodies: molecular (plexiform) layer, external granular, external pyramidal, internal pyramidal and multiform (fusiform)

layer. These are also referred for brevity as layer I, II, III, V and VI respectively. To note that while most of the cerebral cortex is organized into six layers, the Broca's area and, in general the motor cortex, is usually divided into four layers only being layer IV very thin and considered only as a pathway from the thalamus without specialized neurons [4, 71].

Table 5.3 shows the predicted densities on layer III, V and VI of the four above mentioned slabs for all considered methods. Since stereology reports layer estimates, we here group the results by layer. Average results per layer are considered more robust estimates of real densities.

On Layer III StarDist (UNet) obtains the closest to stereology estimate, but also shows high variability, as we can understand from the bad predictions on Slab 18 (see Figure 5.3). On Layer V is StarDist (ResNet) to achieve the closest estimate, but still high variability is also detected. However, stereology itself is affected by high variability, obtaining density estimates ranging from 12162 to 20484 $\frac{\#cells}{mm^3}$. BCFind-v2 densities are, on the other hand, the closest to stereology on Layer VI, moreover with a much lower standard deviation. Overall, BCFind-v2 and StarDist (ResNet) obtain similar results on all layers and slabs.

Qualitatively, looking at Figure 5.3, we firstly notice the low brightness on the imaging of Slab 18 which also decreases along the Y axis, explaining the poor recall performances of BCFind-v2 and StarDist on this slice (Table 5.2). However, BCFind-v2 predictions seem less affected by stripe artifacts and the poor contrast on the upper part of the image. While StarDist, especially with UNet backbone, displays higher rate of false negatives and stripe artifacts all over the considered section. Secondly, the high fluorescence of Slab 30, even if it does not reveal clear artifacts on the predictions, mainly visible with high rates of false negatives, can explain the decreased precision metrics of BCFind-v2 and StarDist (ResNet) (Table 5.2), hence possibly hiding some false positive predictions. Finally, CellPose low recall is here evident with less dense predictions, strongly affected by changes in the image brightness. By an overall inspection of Figure 5.3, BCFind-v2 seems to return more consistent predictions, only marginally affected by variations in input brightness. We would also like to point out how clearly visible the cell density changes between cortical layers are in the DL predictions (Figure 5.4), paving the way for a possibly automatic layer segmentation and a more in depth inspection of brain cytoarchitecture.

Layer	Slab	BCFind-v2 (ResUNet)	Stardist (ResNet)	StarDist (UNet)	Cellpose	Stereology	Volume (mm^3)
III	6	16299.80	13826.54	**11435.43**	6235.90	10646.75	84.9555
	18	14413.17	**13245.13**	9713.07	9866.65	13412.64	113.2260
	30	13656.05	15475.09	14465.01	**10517.81**	11649.91	102.6240
	42	18052.03	18546.95	16531.34	**9643.21**	12251.03	51.2406
	mean (st. dev.)	15605.26 (1709.50)	15273.43 (2377.96)	**13036.21** (3496.18)	9065.89 (1922.80)	11990.08 (1156.44)	
V	6	18017.30	14864.79	**11970.96**	6380.82	12162.03	47.1015
	18	**15780.55**	13827.55	9843.87	9875.26	17037.78	62.9554
	30	14428.27	**16011.94**	14808.08	10118.20	16708.95	56.7181
	42	**20371.37**	20665.24	18194.80	9541.95	20484.54	27.4817
	mean (st. dev.)	17149.37 (2259.02)	**16342.38** (3016.84)	13704.43 (3485.60)	8979.06 (1748.19)	16598.32 (3415.10)	
VI	6	15839.03	**13978.84**	11517.38	7664.54	13890.03	8.6068
	18	**13480.22**	11848.20	8439.84	10217.72	17602.81	79.6373
	30	13316.53	**14374.16**	13235.04	10340.26	15462.57	59.0898
	42	17273.42	**17402.98**	15689.57	10576.47	23231.62	35.0406
	mean (st. dev.)	**14977.30** (1659.38)	14401.05 (2288.20)	12220.46 (3074.82)	9699.75 (1364.95)	17546.76 (4083.97)	

Table 5.3: Predicted densities $\left(\frac{\#cells}{mm^3}\right)$ and corresponding volumes on whole human brain slabs, grouped by layer. Bold values are the estimates based on DL predictions the closest to stereology.

5.4 Inference time

Dealing with huge amount of data, as in this case, could require impracticable machine-time to perform the analyses, therefore it is essential for a model to keep its inference time as short as possible. We therefore measure the times each DL method employs to make predictions on volumes of $100 \times 100 \times 50$ voxels. To avoid hardware and implementation impacts on the reported times, we rescaled the execution times by resource percentage usage: 60% of the GPU for CellPose neural network and post-processing, 10% of the CPU for CellPose 3D adaptation of 2D predictions, 80% of the GPU for for BCFind-v2 neural network and post-processing, 10% of the GPU for StarDist neural network and 90% of the CPU for StarDist post-processing. Moreover, since time is highly affected by the number of predicted cells, in Figure 5.5 we report the average results after binning the number of predicted cells in a sample of 8000 volumes.

Worth to notice that while both BCFind-v2 and StarDist directly ingest 3D volumes, CellPose needs to process 2D slices per time and subsequently merge them. Unfortunately, the official implementation does not allow for batch prediction (only large 2D images are internally tiled and batched) so we had to predict each z-plane separately. The 2D nature of this model is therefore its main speed bottleneck. BCfind-v2 and StarDist on the contrary, being 3D models, have faster neural network predictions.

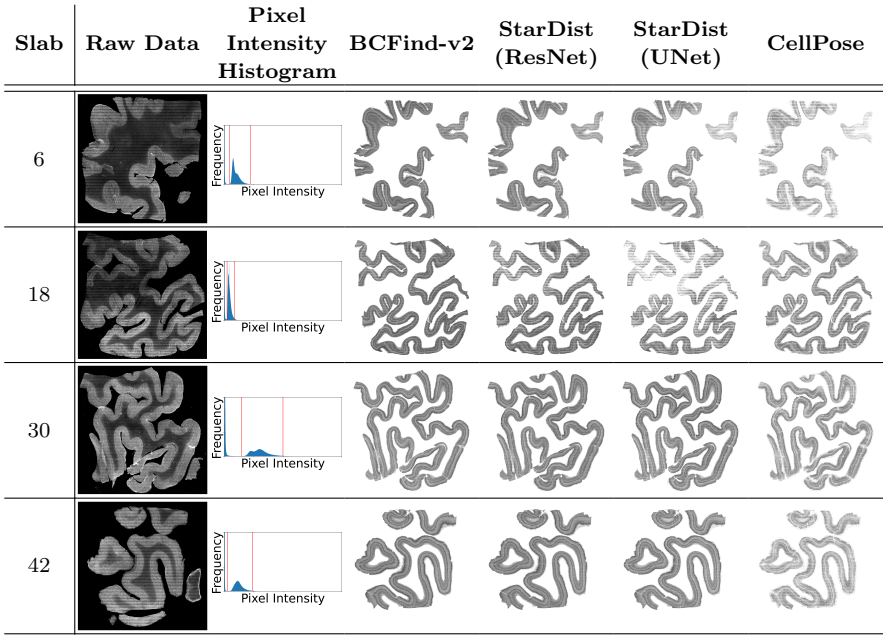


Figure 5.3: Maximum intensity projections (MIP) of four brain slabs from the considered human Broca's area, corresponding pixel intensity histograms and cell coordinate predictions of DL methods. As its clear from the histograms, the pixel dynamics of displayed MIPs (red vertical lines) cover very different ranges of intensities.

Low-weighted StarDist neural networks (400K parameters for the ResNet and 1.2M parameters for the UNet) compared to BCFind-v2 neural network (18M parameters) are faster. However, the CPU-implemented post-processing of StarDist greatly increases the prediction time, while the low-weighted GPU-implemented blob detector of BCFind-v2 maintains strong speed performances. Overall, considering the rescaled time, BCFind-v2 employed 10min to analyze the 8000 considered volumes, StarDist (ResNet) 1h 13min, StarDist (UNet) 1h 36min and CellPose 4h 10min.

Computations are made on a system with an Nvidia GeForce RTX 2080 Ti, eight cores Intel Xeon W-2123 and 126Gb RAM. The implementation of each model is taken from its official repository: https://github.com/stardist/stardist/tree/master for StarDist and https://github.com/

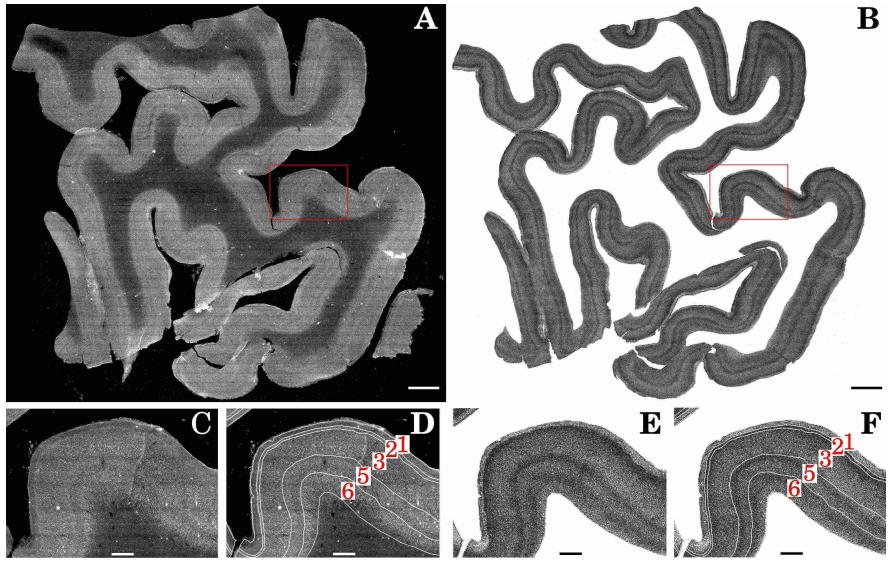


Figure 5.4: DL predictions identify cell density changes between cortical layers. (A) MIP of slab 30. (B) Corresponding BCFind-v2 predictions. The highlighted region of interest (RoI) (C) without and (D) with layer contours on the raw data MIP. The same RoI on the BCFind-v2 predictions (E) without and (F) with layer contours. Red numbers in D and F denote the cortical layer identifier. Scale bars in A–B are 3 mm long, while in C–F are 750 μm. Layer segmentation has been manually drawn on a central plane of the raw image. DL predictions, unaware of layer segmentation, delineate individual layers and even two known subregions of clustered neurons at the layer III-V interface and in the upper part of layer VI that appear as dense bands in these layers.

`MouseLand/cellpose` for CellPose.

5.5 Manual annotation effort

When analyzing 3D biological images three kind of information can be mainly extracted. From the finest to the coarsest we have: complete segmentation, centroid location and density or counts of objects of interest (in our case neurons). It is easy to understand how much complex and labor intensive is

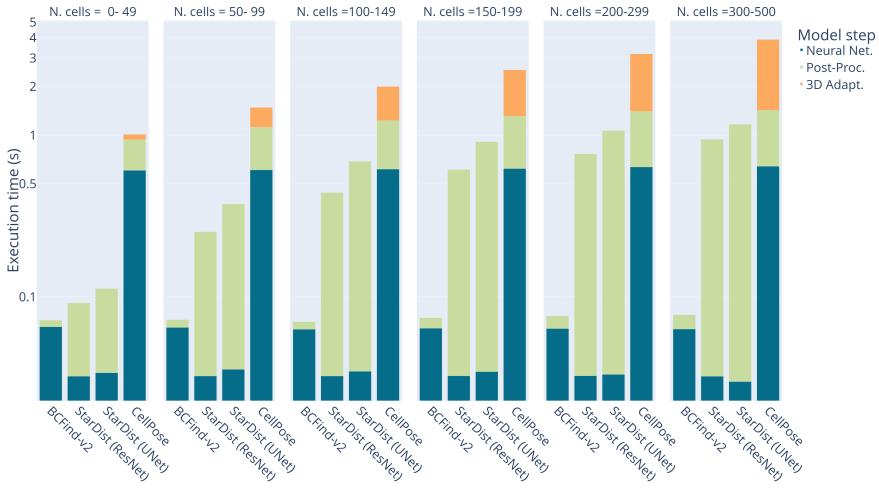


Figure 5.5: Execution times for different numbers of predicted cells. Predictions are made on volumes with identical shape of $360 \times 360 \times 180 \mu m^3$. All operations are performed on a machine with an Nvidia GeForce RTX 2080 Ti, eight cores Intel Xeon W-2123 and 126Gb RAM.

having to completely segment 3D objects, since highly irregular polyhedrons have to be drawn. On the other hand, having to locate object centers only is a much faster process, just requiring one point per object. For what concerns density/counts predictions, they actually need to rely on some sort of coordinate annotations, but since they only require the model a large context understanding of the scene, necessitate a minor number of manual annotations. Additionally, making assumptions on the context under analysis and thus reducing the need for the model to learn it, decreases further the number of annotations required. Stereology, with its assumption of homogeneous layer densities and its count predictions, falls into this category: the model with lowest demand for annotations and hence human effort required. Only 2245 cell markers were indeed needed for stereology to estimate layer densities in the Broca's area of a human, leading to an approximate labor time of 6-7 h. To which we should also add the time to correctly segment the cortical layers, estimated at 6-7 h. To note that layer segmentation is here uniquely done on a central 2D plane of each brain slab and then projected to the whole slab thickness. Conversely, DL models, to learn how to detect

cells in 3D images, needed in this case 22596 ground truth markers (10 times the number needed by stereology) requiring an approximate labor time of 81 h (< 6 times the time needed by stereology). We do not have information on the time that would be needed to segment at least the same amount of cells as needed for localization purposes, but we argue that this would be far beyond the acceptable time or price of many life-science laboratories.

In conclusion, despite complete cell segmentation being the most refined information obtainable, it is almost impracticable on large-scale, highly variable LSFM images. On the other hand, marker annotations for coordinate predictions are a much more feasible task, but still, quite laborious. This labor is however paid back with no underlying assumptions on cell distribution and object-wise information retrieval. Stereology instead, stands on the highest step of the podium when talking about manual annotation effort, but only average layer densities are retrieved. If homogeneity assumption within segmented layers holds and no finer information is deemed necessary, stereology would surely be the method of choice. Otherwise, if assumptions do not easily hold or you don't want to rely on a-priori layer segmentation, DL models could give you more fine-grained information without relying on spatial distribution assumptions. We also note that StarDist and CellPose, unlike BCFind, not only return cell coordinates, but also their segmentation, making them the methods with the most fine-grained predictions. However, when trained from markers only and on images where the cell membranes are not visible, as in our case, the quality of segmentation is not guaranteed and hence probably misleading.

5.6 Conclusion

Three DL models for cell detection (BCFind-v2, StarDist-3D and CellPose) have been extensively tested and evaluated under multiple aspects, each complementing the other. Object-wise metrics and large-scale comparison with stereology showed the high overall performances of StarDist and BCFind-v2, while CellPose, despite its claims of generality and adaptability, fails in detecting many cells from complex 3D LSFM images. Visual inspection of predicted point clouds revealed the impacts of imaging variability and artifacts, especially on StarDist (UNet) predictions, but also biologically coherent structures. The work therefore proves that when properly validated, automatic models con obtain reliable and, importantly, biologically inter-

pretable predictions. On the side of limitations however, supervised DL models still require very large training-sets, hence depending on a great human labeling effort. In this sense, the trade-off between human labor and prediction granularity is the real crossroads when it comes to choosing a model. Future research must surely go in the direction of reducing this human effort (Section 6.2).

Chapter 6

Conclusion

This chapter summarizes the contribution of the thesis and discusses avenues for future research.

6.1 Summary of contribution

The thesis presents an efficient and scalable software for accurate cell detection on large-scale light-sheet fluorescence microscopy data. We validated and applied the proposed method on two challenging, diverse (in voxel resolution, stained neurons and origin of biological samples) and vast 3D data. Multiple variants of deep-learning backbone are available for case-specific needs and performances. An easy-to-use implementation is freely available at `https://codeberg.org/curzio/BCFind-v2` provided with a Dockerfile for consistent deployment. The successful application to whole mouse brains led to two international journal publications [21, 74]. The extensive comparison of proposed technique applied to the entire Broca's area of a human is ready for submission under the title "Stereology or Deep-Learning? On the reliability and extrapolation power of deep-learning methods applied to large-scale human brain tissue" to *Scientific Reports*.

6.2 Directions for future work

Supervised-learning surely offers a reliable method for training accurate quantification methods, however the requirement of large training-sets hinder

its applicability to high-throughput biological experiment pipelines. Future researches will need therefore to reduce the burden of human annotation. Self-supervised pre-training techniques could be a real cornerstone to this end. Contrastive-learning indeed has proven to give huge opportunities in reducing the number of training labels [12, 82], exploiting them more efficiently [88] or even in image-to-image translation [58]. Similarly, generative models [34, 61, 89] can also be applied both to generate reliable artificial data [57] or to learn effective features without supervision [60, 86]. Moreover, model efficiency and scalability can be further improved through the adoption of end-to-end object detectors [80, 90] removing therefore the computational cost of the blob detection step of our pipeline.

Appendix A

Publications

This research activity has led to several publications in international journals and conferences. These are summarized below.

International Journals

1. A. Franceschini, G. Mazzamuto, **C. Checcucci**, L. Chicchi, D. Fanelli, I. Costantini, M. B. Passani, F. S. Pavone, L. Silvestri. "BRAin-wide Neuron quantification Toolkit reveals strong sexual dimorphism in the evolution of fear memory", *Cell Reports*, vol. 42, iss. 8, August 2023. [DOI: 10.1016/j.celrep.2023.112908]
2. L. Silvestri, M. C. Müllenbroich, I. Costantini, A. P. Di Giovanna, G. Mazzamuto, A. Franceschini, D. Kutra, A. Kreshuk, **C. Checcucci**, L.A. Toresano, P. Frasconi, L. Sacconi, F. S. Pavone. "Universal autofocus for quantitative volumetric microscopy of whole mouse brains", *Nature Methods*, vol. 18, pp. 953-958, 2021. [DOI: 10.1038/s41592-021-01208-1]

Submitted

1. **C. Checcucci**, M. Scardigli, J. Ramazzotti, N. Bradly, B. Wicinski, P. Hof, I. Costantini, F. S. Pavone, P. Frasconi. "Stereology or Deep-Learning? A Comparative Study on Fluorescence Human Brain 3D Reconstruction", Submitted to *Scientific Reports*, 2023.

International Conferences and Workshops

1. I. Costantini, M. Scardigli, J. Ramazzotti, N. Brady, F. Cheli, G. Mazzamuto, F. M. Castelli, **C. Checcucci**, L. Silvestri, P. Frasconi, F. S. Pavone.

"High-resolution human brain 3D reconstruction with light-sheet fluorescence microscopy", in *Proc. of Neural Imaging and Sensing* (SPIE BIOS 2023), San Francisco (United States), 2023.

2. M. Scardigli, I. Costantini, N. Brady, M. Baghdad, J. Ramazzotti, G. Mazzamuto, F. M. Castelli, **C. Checcucci**, L. Silvestri, P. Frasconi, F. S. Pavone. "3D molecular phenotyping of the human brain Broca's area using light-sheet fluorescence microscopy", in *Biomedical Spectroscopy, Microscopy and Imaging II* (SPIE Photonics 2022), Strasbourg (France), 2022

Other works

1. T. Kreuz, F. Senocrate, G. Cecchini, **C. Checcucci**, A. L. A. Mascaro, E. Conti, A. Scaglione, F. S. Pavone. "Latency Correction in Sparse Neuronal Spike Trains", *Journal of Neuroscience Methods*, vol. 381, 2022. [DOI: 10.1016/j.jneumeth.2022.109703]

2. B. Giannoni, F. Pollastri, C. Adembri, D. Straticò, P. Vannucchi, A. Stival, **C. Checcucci**, C. Bruno, R. Pecci. "Hearing outcomes and patient satisfaction after stapes surgery: local versus general anaesthesia" *Acta Otorhinolaryngologica Italica*, vol. 42, iss. 5, pp. 471-80, 2022/ [DOI: 10.14639/0392-100X-N2033]

3. G. Cecchini, A. L. A. Mascaro, A. Scaglione, **C. Checcucci**, E. Conti, I. Adam, D. Fanelli, R. Livi, F. S. Pavone, T. Kreuz. "Cortical Propagation as a Biomarker for Recovery after Stroke", *PLOS Computational Biology*, vol. 17, pp. 1-23, 2021. [DOI: 10.1371/journal.pcbi.1008963]

Bibliography

[1] M. Abercrombie, "Estimation of nuclear population from microtome sections," *The anatomical record*, vol. 94, no. 2, pp. 239–247, 1946.

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

[3] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ants similarity metric performance in brain image registration," *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.

[4] H. Barbas and M. Á. García-Cabezas, "Motor cortex layer 4: less is more," *Trends in neurosciences*, vol. 38, no. 5, pp. 259–261, 2015.

[5] D. J. Barry, C. Gerri, D. M. Bell, R. D'Antuono, and K. K. Niakan, "Giani–open-source software for automated analysis of 3d microscopy images," *Journal of cell science*, vol. 135, no. 10, p. jcs259511, 2022.

[6] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[7] S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy *et al.*, "Ilastik: interactive machine learning for (bio) image analysis," *Nature methods*, vol. 16, no. 12, pp. 1226–1232, 2019.

[8] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," *Advances in neural information processing systems*, vol. 24, 2011.

[9] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[10] L. Breiman, J. Friedman, and C. J. Stone, *Classification and regression trees*. CRC press, 1984.

[11] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat *et al.*, "Cellprofiler: image analysis software for identifying and quantifying cell phenotypes," *Genome biology*, vol. 7, pp. 1–11, 2006.

[12] K. Chaitanya, E. Erdil, N. Karani, and E. Konukoglu, "Contrastive learning of global and local features for medical image segmentation with limited annotations," *Advances in neural information processing systems*, vol. 33, pp. 12 546–12 558, 2020.

[13] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer, 2016, pp. 424–432.

[14] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, big, simple neural nets for handwritten digit recognition," *Neural computation*, vol. 22, no. 12, pp. 3207–3220, 2010.

[15] D. C. Cireşan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *Twenty-second international joint conference on artificial intelligence*. Citeseer, 2011.

[16] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge, 2013.

[17] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.

[18] I. Costantini, L. Morgan, J. Yang, Y. Balbastre, D. Varadarajan, L. Pesce, M. Scardigli, G. Mazzamuto, V. Gavryusev, F. M. Castelli *et al.*, "A cellular resolution atlas of broca's area," *Science Advances*, vol. 9, no. 41, p. eadg3844, 2023.

[19] D. Dao, A. N. Fraser, J. Hung, V. Ljosa, S. Singh, and A. E. Carpenter, "Cellprofiler analyst: interactive data exploration, analysis and classification of large biological image sets," *Bioinformatics*, vol. 32, no. 20, pp. 3210–3212, 2016.

[20] W. Fedus, B. Zoph, and N. Shazeer, "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity," *The Journal of Machine Learning Research*, vol. 23, no. 1, pp. 5232–5270, 2022.

[21] A. Franceschini, G. Mazzamuto, C. Checcucci, L. Chicchi, D. Fanelli, I. Costantini, M. B. Passani, B. A. Silva, F. S. Pavone, and L. Silvestri, "Brain-wide neuron quantification toolkit reveals strong sexual dimorphism in the evolution of fear memory," *Cell Reports*, vol. 42, no. 8, 2023.

[22] P. W. Frankland and B. Bontempi, "The organization of recent and remote memories," *Nature reviews neuroscience*, vol. 6, no. 2, pp. 119–130, 2005.

[23] P. Frasconi, L. Silvestri, P. Soda, R. Cortini, F. S. Pavone, and G. Iannello, "Large-scale automated identification of mouse brain cells in confocal light sheet microscopy images," *Bioinformatics*, vol. 30, no. 17, pp. i587–i593, 2014.

[24] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[25] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)," *The annals of statistics*, vol. 28, no. 2, pp. 337–407, 2000.

[26] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[27] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3146–3154.

[28] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[29] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.

[30] C. J. Guenthner, K. Miyamichi, H. H. Yang, H. C. Heller, and L. Luo, "Permanent genetic access to transiently active neurons via trap: targeted recombination in active populations," *Neuron*, vol. 78, no. 5, pp. 773–784, 2013.

[31] H.-J. G. Gundersen, "The nucleator," *Journal of microscopy*, vol. 151, no. 1, pp. 3–21, 1988.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[33] ——, "Identity mappings in deep residual networks," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. Springer, 2016, pp. 630–645.

[34] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[35] S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber *et al.*, "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," 2001.

[36] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[37] J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. H. Stelzer, "Optical sectioning deep inside live embryos by selective plane illumination microscopy," *Science*, vol. 305, no. 5686, pp. 1007–1009, 2004.

[38] I. Izquierdo, C. R. Furini, and J. C. Myskiw, "Fear memory," *Physiological reviews*, vol. 96, no. 2, pp. 695–750, 2016.

[39] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, no. 1, pp. 79–87, 1991.

[40] S. A. Josselyn and P. W. Frankland, "Memory allocation: mechanisms and function," *Annual review of neuroscience*, vol. 41, pp. 389–413, 2018.

[41] S. A. Josselyn, S. Köhler, and P. W. Frankland, "Finding the engram," *Nature Reviews Neuroscience*, vol. 16, no. 9, pp. 521–534, 2015.

[42] A. Krishnan, L. J. Williams, A. R. McIntosh, and H. Abdi, "Partial least squares (pls) methods for neuroimaging: a tutorial and review," *Neuroimage*, vol. 56, no. 2, pp. 455–475, 2011.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[44] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[45] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[46] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the loss landscape of neural nets," *Advances in neural information processing systems*, vol. 31, 2018.

[47] T. Lindeberg, "Feature detection with automatic scale selection," *International journal of computer vision*, vol. 30, pp. 79–116, 1998.

[48] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[49] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.

[50] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91–110, 2004.

[51] P. manuscript editors, A. coordination, I. data analysis Armand Ethan 42 Yao Zizhen 5, A. seq data generation, processing Fang Rongxin 45 Hou Xiaomeng 10 Lucero Jacinta D. 18 Osteen Julia K. 18 Pinto-Duarte Antonio 18 Poirion Olivier 10 Preissl Sebastian 10 Wang Xinxin 10 97, E. retro-seq data generation, processing Dominguez Bertha 53 Ito-Cole Tony 1 Jacobs Matthew 1 Jin Xin 54 99 100 Lee Cheng-Ta 53 Lee Kuo-Fen 53 Miyazaki Paula Assakura 1 Pang Yan 1 Rashid Mohammad 1 Smith Jared B. 54 Vu Minh 1 Williams Elora 54 *et al.*, "A multimodal cell census and atlas of the mammalian primary motor cortex," *Nature*, vol. 598, no. 7879, pp. 86–102, 2021.

[52] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[53] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate $\mathcal{O}\left(\frac{1}{k^2}\right)$," in *Dokl. akad. nauk Sssr*, vol. 269, no. 3, 1983, pp. 543–547.

[54] O. Oktay, J. Schlemper, L. L. Folgoc, M. J. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. G. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," *ArXiv*, vol. abs/1804.03999, 2018.

[55] J. Oltmer, E. W. Rosenblum, E. M. Williams, J. Roy, J. Llamas-Rodriguez, V. Perosa, S. N. Champion, M. P. Frosch, and J. C. Augustinack, "Stereology neuron counts correlate with deep learning estimates in the human hippocampal subregions," *Scientific Reports*, vol. 13, no. 1, p. 5884, 2023.

[56] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[57] S. Pandey, P. R. Singh, and J. Tian, "An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation," *Biomedical Signal Processing and Control*, vol. 57, p. 101782, 2020.

[58] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*. Springer, 2020, pp. 319–345.

[59] M. A. Pezzone, W.-S. Lee, G. E. Hoffman, and B. S. Rabin, "Induction of c-fos immunoreactivity in the rat forebrain by conditioned and unconditioned aversive stimuli," *Brain research*, vol. 597, no. 1, pp. 41–50, 1992.

[60] V. Purma, S. Srinath, S. Srirangarajan, A. Kakkar *et al.*, "Genselfdiff-his: Generative self-supervision using diffusion for histopathological image segmentation," *arXiv preprint arXiv:2309.01487*, 2023.

[61] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[62] N. Renier, E. L. Adams, C. Kirst, Z. Wu, R. Azevedo, J. Kohl, A. E. Autry, L. Kadiri, K. U. Venkataraju, Y. Zhou *et al.*, "Mapping of brain activity by automated volume analysis of immediate early genes," *Cell*, vol. 165, no. 7, pp. 1789–1802, 2016.

[63] N. Renier, Z. Wu, D. J. Simon, J. Yang, P. Ariel, and M. Tessier-Lavigne, "idisco: a simple, rapid method to immunolabel large tissue samples for volume imaging," *Cell*, vol. 159, no. 4, pp. 896–910, 2014.

[64] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[65] D. S. Roy, Y.-G. Park, M. E. Kim, Y. Zhang, S. K. Ogawa, N. DiNapoli, X. Gu, J. H. Cho, H. Choi, L. Kamentsky *et al.*, "Brain-wide mapping reveals that engrams for a single memory are distributed across multiple brain regions," *Nature communications*, vol. 13, no. 1, p. 1799, 2022.

[66] D. E. Rumelhart, G. E. Hinton, R. J. Williams *et al.*, "Learning internal representations by error propagation," 1985.

[67] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern recognition*, vol. 33, no. 2, pp. 225–236, 2000.

[68] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid *et al.*, "Fiji: an open-source platform for biological-image analysis," *Nature methods*, vol. 9, no. 7, pp. 676–682, 2012.

[69] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, "Cell detection with star-convex polygons," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*. Springer, 2018, pp. 265–273.

[70] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, and J. Dean, "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer," 2017. [Online]. Available: https://openreview.net/pdf?id=B1ckMDqlg

[71] S. Shipp, "The importance of being agranular: a comparative account of visual and motor cortex," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1456, pp. 797–814, 2005.

[72] S. A. Shuvaev, A. A. Lazutkin, A. V. Kedrov, K. V. Anokhin, G. N. Enikolopov, and A. A. Koulakov, "Dalmatian: an algorithm for automatic cell detection and counting in 3d," *Frontiers in neuroanatomy*, vol. 11, p. 117, 2017.

[73] B. A. Silva, S. Astori, A. M. Burns, H. Heiser, L. van den Heuvel, G. Santoni, M. F. Martinez-Reza, C. Sandi, and J. Graeff, "A thalamo-amygdalar circuit underlying the extinction of remote fear memories," *Nature Neuroscience*, vol. 24, no. 7, pp. 964–974, 2021.

[74] L. Silvestri, M. Müllenbroich, I. Costantini, A. Di Giovanna, G. Mazzamuto, A. Franceschini, D. Kutra, A. Kreshuk, C. Checcucci, L. Toresano *et al.*, "Universal autofocus for quantitative volumetric microscopy of whole mouse brains," *Nature Methods*, vol. 18, no. 8, pp. 953–958, 2021.

[75] C. Sommer, C. Straehle, U. Koethe, and F. A. Hamprecht, "Ilastik: Interactive learning and segmentation toolkit," in *2011 IEEE international symposium on biomedical imaging: From nano to macro*. IEEE, 2011, pp. 230–233.

[76] D. Sterio, "The unbiased estimation of number and sizes of arbitrary particles using the disector," *Journal of microscopy*, vol. 134, no. 2, pp. 127–136, 1984.

[77] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: a generalist algorithm for cellular segmentation," *Nature methods*, vol. 18, no. 1, pp. 100–106, 2021.

[78] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[79] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 13, no. 06, pp. 583–598, 1991.

[80] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7464–7475.

[81] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11 531–11 539, 2020.

[82] X. Wang, R. Zhang, C. Shen, T. Kong, and L. Li, "Dense contrastive learning for self-supervised visual pre-training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3024–3033.

[83] M. Weigert, U. Schmidt, R. Haase, K. Sugawara, and G. Myers, "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.

[84] A. L. Wheeler, C. M. Teixeira, A. H. Wang, X. Xiong, N. Kovacevic, J. P. Lerch, A. R. McIntosh, J. Parkinson, and P. W. Frankland, "Identification of a functional connectome for long-term fear memory in mice," *PLoS computational biology*, vol. 9, no. 1, p. e1002853, 2013.

[85] S. D. Wicksell, "The corpuscle problem: a mathematical study of a biometric problem," *Biometrika*, pp. 84–99, 1925.

[86] W. Xiang, H. Yang, D. Huang, and Y. Wang, "Denoising diffusion autoencoders are unified self-supervised learners," *arXiv preprint arXiv:2303.09769*, 2023.

[87] J.-C. Yen, F.-J. Chang, and S. Chang, "A new criterion for automatic multilevel thresholding," *IEEE Transactions on Image Processing*, vol. 4, no. 3, pp. 370–378, 1995.

[88] X. Zhao, R. Vemulapalli, P. A. Mansfield, B. Gong, B. Green, L. Shapira, and Y. Wu, "Contrastive learning for label efficient semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 623–10 633.

[89] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[90] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable detr: Deformable transformers for end-to-end object detection," *arXiv preprint arXiv:2010.04159*, 2020.