Tool for Morphological classification of neurons (D2.6.1 – SGA2)

Images produced by the microscope

Automatic Segmentation
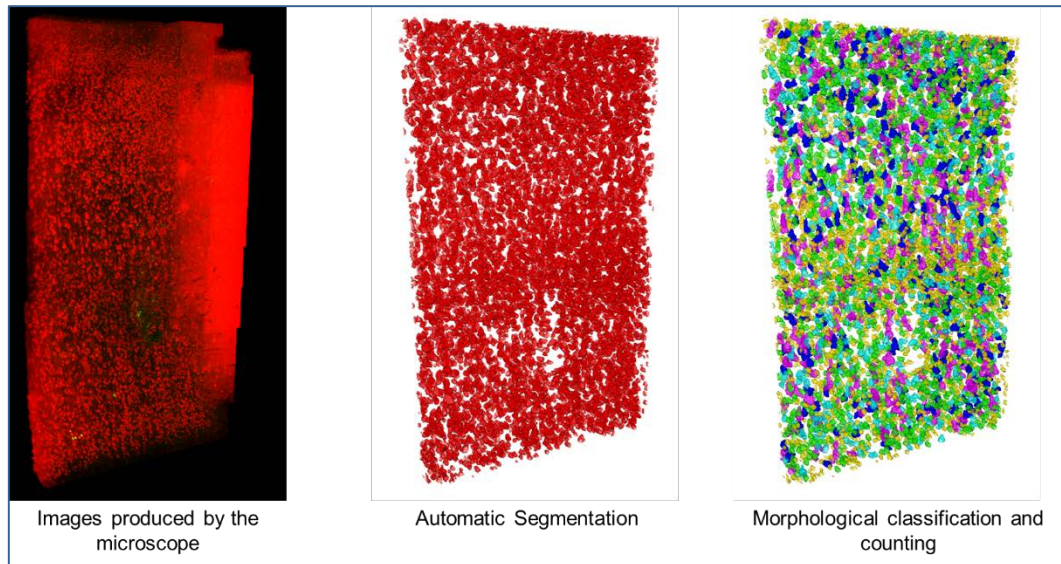
Morphological classification and counting

Figure 1: A pipeline for automatic segmentation and counting of neurons.

Left: slice of human brain cortex acquired with the Two Photon Fluorescence Microscope, stained with anti-NeuN antibodies. Centre: Result of automatic segmentation from the trained CNN model. Right: neurons classified based on their size.

Project Number:	785907	Project Title:	Human Brain Project SGA2

Document Title:	Tool for Morphological classification of neurons
Document Filename:	D2.6.1 (D13.1 D62) SGA2 M18 ACCEPTED 201005.docx
Deliverable Number:	SGA2 D2.6.1 (D13.1, D62)
Deliverable Type:	Other
Work Packages:	WP2.6 and WP2.3
Dissemination Level:	PU = Public
Planned Delivery Date:	SGA2 M18
Actual Delivery Date:	SGA2 M19 / 31 Oct 2019; resubmitted 21 Sep 2020; accepted 5 Oct 2020
Author(s):	Giacomo MAZZAMUTO, LENS (P40); Irene COSTANTINI, LENS (P40)
Compiled by:	Giacomo MAZZAMUTO, LENS (P40); Irene COSTANTINI, LENS (P40)
Contributor(s):	
SciTechCoord Review:	Martin TELEFONT, EPFL (P1)
Editorial Review:	Annemieke MICHELS, EPFL (P1)
Description in GA:	Tool for Morphological classification of neurons (T2.6.4) made available; storage of software on HBP Collab
Abstract:	Quantitative analysis of brain cytoarchitecture requires effective and efficient segmentation of the raw images. Here, we present a machine learning approach for 3D segmentation of neurons acquired with two-photon fluorescence microscopy. This pipeline enables quantitative evaluation of neuronal volume and density. The novel framework for segmentation is working at the level of local visual pattern rather than of single pixels, relying on deep convolutional neural network architectures.
Keywords:	Machine learning, segmentation, classification, neuronal morphology, digital histology
Target Users/Readers:	Neuroscientific community, neuroimaging community, histologists

Table of Contents

Table of Figures

History of Changes made to this Deliverable (post Submission)

Date	Change Requested / Change Made / Other Action
31 Oct 2019	Deliverable submitted to EC
29 Jul 2020	Resubmission with specified changes requested in Review Report Main changes requested: • Change 1: The deliverable currently presents a DL based cell segmentation method instead of morphological classification of neurons • Change 2: The deliverable should contain a systematic description of the data used as training test and validation datasets and provide an overview on how ground truth has been generated and used • Change 3: A description of the network architecture as well as further details on evaluation strategy and results should be added • Change 4: Links to other components of the HBP should be specified • Change 5: While there is report on the results presented at various conferences, there is no indication of publication being prepared
16 Sep 2020	Revised draft sent by SP/CDP to PCO. The Deliverable was rewritten with the exception of Section 2, Section 5.2, the last paragraph of Section 3.2, Figure 2 and Figure 3. Main changes made, with indication where each change was made: • Change 1: In Sections 3.2 and 3.3 it is explained that the neural network actually performs pixel classification and that neuron morphology, together with other useful quantities, can be easily extracted as a post processing step. The advantages and the rationale behind this more general approach are described in the aforementioned sections. • Change 2: The requested information is described in Section 3.1 • Change 3: The requested information is described in Section 3.2 and 3.3 (for the network architecture) and Section 4.1 (for our results). • Change 4: In Sections 1, 3.1, 4.1 and 4.2 it is specified how this method is made possible by other upstream results that we obtained within the HPB and how it can be further applied within the context of HBP. A components annex was added. • Change 5: We have added references to our preprint on bioRxiv which is in the process of being published in a peer-reviewed journal and which describes this method in greater detail
21 Sep 2020	Revised version resubmitted to EC by PCO via SyGMa

1. Summary

In this document, we present the progress on the development of our neural network model for segmentation of neurons in 3D images of the human brain cortex acquired with a Two-Photon Fluorescence Microscope (TPFM). The developed method is not specific to TPFM but can be used on different imaging modalities (e.g. Confocal Microscopy, Light Sheet Fluorescence Microscopy - LSFM) after retraining with specific ground truth. The aim of this work is to automatically analyse large quantities of information, such as the big data produced with advanced microscopy performing high-resolution three-dimensional reconstruction.

This analysis method is part of a framework with as aim to map large portions of the human brain at sub-cellular resolution. During HBP - SGA2 various tasks were dedicated to achieve this challenging goal. In particular, specific sample preparation optimisations (clearing and staining procedures) have been performed (Component C2347: "Multiple staining for characterisation of different neuronal types"), as well as advanced microscopy development (Component C2349: "Integration of ultra-fast large area fluorescence microscope with Switch clearing techniques"). After that, the automatic segmentation step (Component C2354: "Morphological classification of neurons with immunohistological staining"), which we describe here, provides new possibilities to investigate the cytoarchitecture of human brain tissue at mesoscale. Once the 3D meshes of identified neurons are automatically extracted, it is possible to evaluate quantitative parameters such as neuronal density and mean volume, as well as neuron morphology, we have already performed this analysis on mouse brain samples (Component C1745: "Whole-brain maps of different neuronal types").

The automatic segmentation approach that we present here consists of a Convolutional Neural Network that actually performs pixel-based classification, i.e. pixels are assigned to two classes depending on whether they belong to the visual pattern of a neuron or to the background. After pixel classification, standard segmentation algorithms (e.g. isosurface search algorithms such as Marching Cubes) can then be employed to extract the 3D meshes of identified neurons. Once the 3D meshes are available, all other forms of processing and quantitative evaluation are left to a relatively simple post-processing step, where neuron morphology, volume and density can be extracted. In practice, our approach for automatic neuron segmentation not only leads to cell counting but, more importantly, to the volumetric assessment of neuronal size and shape that was simply impossible to obtain with state-of-the-art technologies.

We tested our method on four samples of human brain cortex originating from individuals with different health conditions (healthy and pathological) and different age groups. From these samples, we were able to observe the cortical layers, which can be clearly distinguished because of their different neuronal density and volume.

The current version of the neural network model that we have developed — together with our ongoing effort to port it to TensorFlow and Python — is available on GitHub at this link: https://github.com/lens-biophotonics/neuron_segmentation_model_hbp (under curation https://github.com/bweyers/HBPVisCatalogue/issues/414). The raw images acquired with TPFM on four samples of human brain cortex, together with the corresponding ground truth that was manually annotated to train the neural network, have been curated and are available on EBRAINS at this link: https://doi.org/10.25493/SNWB-YQR. Finally, the whole experimental and segmentation pipeline is described in detail in our comprehensive paper which is in the process of being published in a peer-reviewed journal and is available as a preprint on bioRxiv at this link: https://doi.org/10.1101/2020.08.06.219444.

2.   Introduction

The reconstruction of the cellular and molecular architecture of the human brain cortex is one of the main goals of the HBP. From a technological perspective, the production of such a map is highly challenging for at least three reasons: a) the intact brain tissue must be properly labelled with specific probes for different cell types; b) a macroscopic specimen must be imaged with micrometric resolution in a reasonable time; and c) imaging data must be analysed in an automatic fashion to extract quantitative information about the sample. Quantitative analysis of brain cytoarchitecture requires effective and efficient segmentation of the raw images. This task is highly demanding from an algorithmic point of view, because of the inherent variations of contrast and intensity in the different areas of the specimen, and of the very large size of the datasets to be processed. To achieve this goal, we have trained a Convolutional Neural Network (CNN) for the near real-time segmentation of neurons in three-dimensional images. This method, together with high-throughput sample preparation and imaging, lay the basis for a quantitative revolution in neuroanatomical studies.

3.   Materials and Methods

3.1   Human brain specimens

We developed our method on images obtained from different samples, we selected four specimens of human brain cortex, originating from individuals with different health conditions and ages, as specified below:

- Two portions of the left prefrontal cortex of an adult (male, sample 1) and an elderly subject (female, 99 years old, no Alzheimer's disease but initial cognitive decline, no hypertension; sample 2). These are the healthy control subjects.

- One dysplastic brain sample of the left temporo-occipital cortex of an adult (male, 29 years old, operated to treat drug resistant epilepsy due to focal cortical dysplasia Type IIa (FCDIIa); sample 3).

- One dysplastic brain sample of the left temporo-parietal cortex of a child (male, 8 years old, operated to treat drug resistant epilepsy due to hemimegalencephaly (HME); sample 4).

The samples were treated following the SWITCH/TDE protocol (implemented in SGA2, C2347) to clear and label the human brain tissues with anti-NeuN antibodies. Following the clearing protocol, Two-Photon Fluorescence Microscopy (TPFM) was used to perform 3D mesoscopic reconstruction of the different samples. Detailed information on the procedures is available in the bioRxiv preprint.

After imaging with TPFM, the acquired 3D tiles have been stitched together using ZetaStitcher, a software that we developed to deal with large-volume tomographies and made available on GitHub at https://github.com/lens-biophotonics/ZetaStitcher (under curation https://github.com/bweyers/HBPVisCatalogue/issues/402). The resulting stitched volumes consisted of several GB of data. In particular, the fused volumes of the four samples acquired in this work were sized 19, 50, 57 and 52 GB. To generate the ground truth, manual annotation was performed by two distinct operators. The training dataset was composed of 112 images of 512 x 512 px, corresponding to 450 x 450 μm^2, for a total of 7,312 manually annotated neurons; 14 images (1,505 neurons) were used to validate the CNN and 14 (1,208 neurons) to test it. The raw stitched images and the manually annotated ground truth are available for download from EBRAINS (https://doi.org/10.25493/SNWB-YQR).

3.2   The machine learning method

The developed 2D Convolutional Neural Network model, (available here: https://github.com/lens-biophotonics/neuron_segmentation_model_hbp), performs pixel-based classification. 32 x 32 x 2

sized patches (i.e. considering red and green channels, corresponding to NeuN and DAPI respectively) are extracted from the stitched volume, and fed to the CNN model after a preprocessing step consisting of a single 5 x 5 Gaussian kernel filtering stage with σ = 3. This operation replicates the intrinsic blurring caused by resampling during data augmentation in the training phase of the CNN. The neural network architecture consists of three convolutional layers, the first two of which are followed by 2 x 2 max-pooling downsampling, and three fully connected layers, the last of which makes use of a two-class softmax activation function. More detailed information, such as trainable parameters and the block diagram of the overall network structure, is available in the bioRxiv preprint.

The neural network was developed on top of the ALIQUIS framework for Machine Vision, which has proven to be particularly suitable for rapid prototyping of CNN architectures. ALIQUIS is developed by Bioretics srl, a private company (SME) involved in this project, and is freely available for research purposes (https://www.bioretics.com/aliquis). Efforts to port the network to TensorFlow and Python are ongoing and available on GitHub (https://github.com/lens-biophotonics/neuron_segmentation_model_hbp/tree/master/tensorflow_2.5D_model). The trained neural network model as well as the CNN structure, described in the ALIQUIS™ language, and the ground truth used for training are available for download from EBRAINS (https://doi.org/10.25493/SNWB-YQR).

Compared to a previous preliminary version of the neural network, several improvements have been introduced during SGA2. In particular, in the previous version of the CNN described above, only the red channel (NeuN staining) was used for the purposes of pixel classification and segmentation. In the new version of the network, we also take the green channel (DAPI staining) into consideration. This improves the performance of the network and increases its segmentation capabilities in the presence of clusters of neurons that were previously segmented as a single entity and are now correctly segmented as separate cells (Figure 2). From a more quantitative point of view, the performance of the new model can also be inferred from the FROC (Free-response Receiver Operating Characteristic) curves (Figure 3) which show how the new model achieves a higher sensitivity (or recall) with fewer cases of false positives.
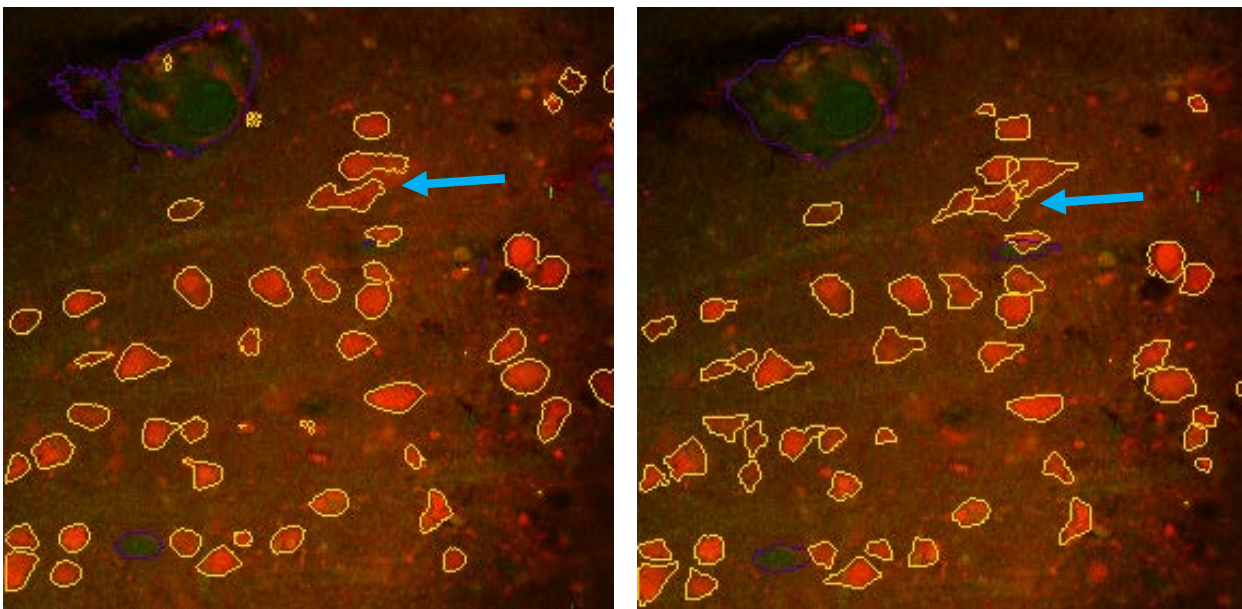

Figure 2: Segmentation in the presence of neuron clusters.

Left: results of neuron segmentation using the old neural network (only red channel). Right: result of neuron segmentation using the new improved version of the neural network (red and green channels). The blue arrow highlights a region of clustered neurons that are correctly segmented with the new version of the CNN. Blood vessels are segmented as well, shown here with a purple contour.
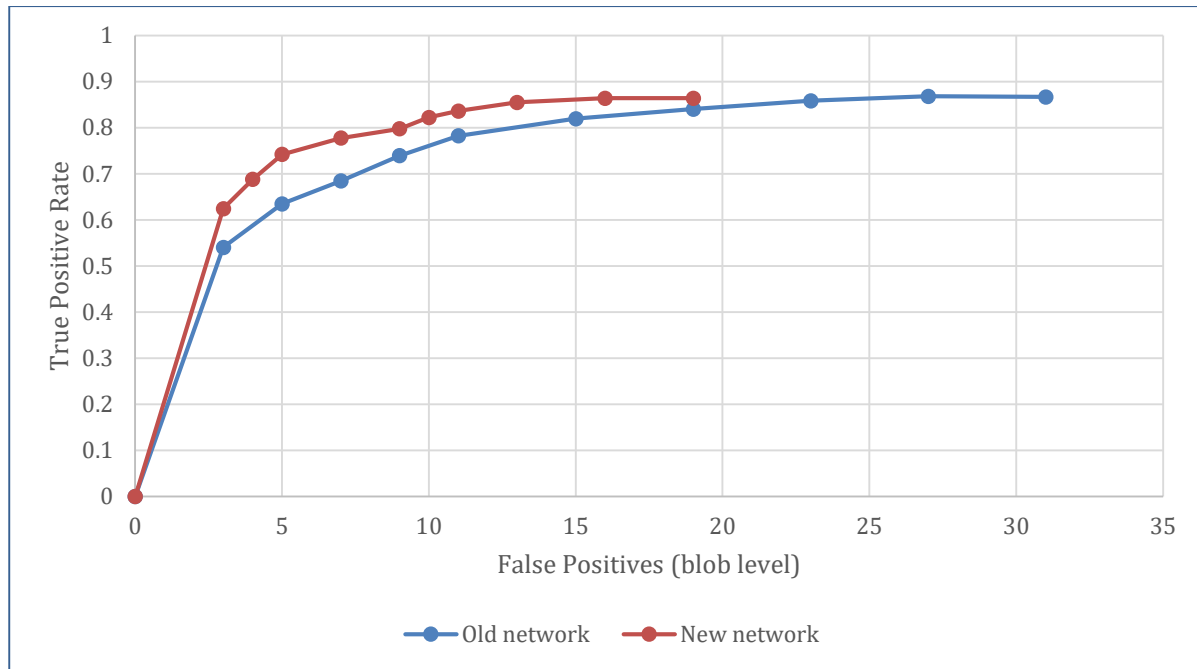
Figure 3: Comparison of FROC curves between old and new version of the neural network.

3.3   3D reconstruction using the 2.5D approach

The CNN model independently processes 2-channel acquisition frames and produces probability heatmaps. These two-dimensional maps are reassembled back into a 2.5D stack to obtain an estimate of the three-dimensional probability distribution of neuronal soma presence. On these reassembled 3D heatmaps, we apply isosurface search algorithms (such as Marching Cubes) to obtain the volumetric representation of identified neurons in the form of 3D meshes.

In practice, the neural network actually performs pixel classification in 2D. After extraction of the 3D objects using standard algorithms, the evaluation of whichever quantitative parameter and neuron morphology is left to a relatively simple post-processing step. In Section 4, we demonstrate that neuronal morphology, volume and density can be evaluated from the vectorial data. We chose this approach because it is easier to implement, requires less ground truth and is more general. Indeed, the neural network need only be trained once (for neuron segmentation), and then all sorts of data can be extracted at a later time (including, but not limited to, neuron morphology). Overall, this approach allows us to retrieve a three-dimensional vectorial reconstruction of the segmented objects in the entire z-stack.

All 2.5D computations have been performed on a standard Linux-based workstation by the ALIQUIS™ software ecosystem with Google TensorFlow as CNN back end.

4.   Validation and Impact

4.1   Actual Use of Output(s) / Exploitation

We applied our neural network model for neuron segmentation (C2354) on all four samples mentioned above. The resulting 3D meshes of the segmented neurons of the four samples are available for download from EBRAINS (https://doi.org/10.25493/SNWB-YQR). Three-dimensional renderings - which were prepared using Fiji - are shown in Figure 4.

From a more quantitative point of view, and to further demonstrate the full extent of the applicability of the techniques described in this report, we analysed the automatically segmented neurons of the mesoscopic reconstruction to extract anatomical information. In particular, we

obtained the average mean volume and density distributions of neurons in healthy vs pathological and young vs elderly individuals (detailed information is available in the bioRxiv preprint).
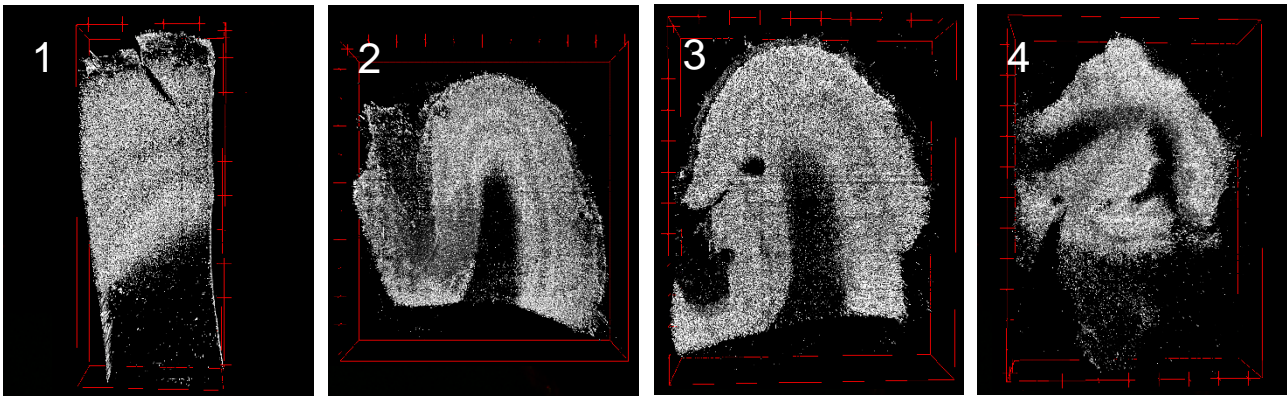

Figure 4: 3D rendering of the masks of segmented neurons in the 4 samples.

The neuronal segmentation made it possible to distinguish the different cortical layers through the cortex thanks to their different neuronal volume and density. Figure 5, obtained with Blender (www.blender.org), shows a 3D rendering of the neurons of sample 1. The objects are colour-coded depending on their volume. It is apparent that the different cortical layers can be clearly distinguished (from left to right of the figure: layer 1,2,3,4,5,6 and white matter). Indeed, the inner anatomical architecture of the six cortical layers can be reconstructed, with the different layers exhibiting different neuron morphologies and densities.
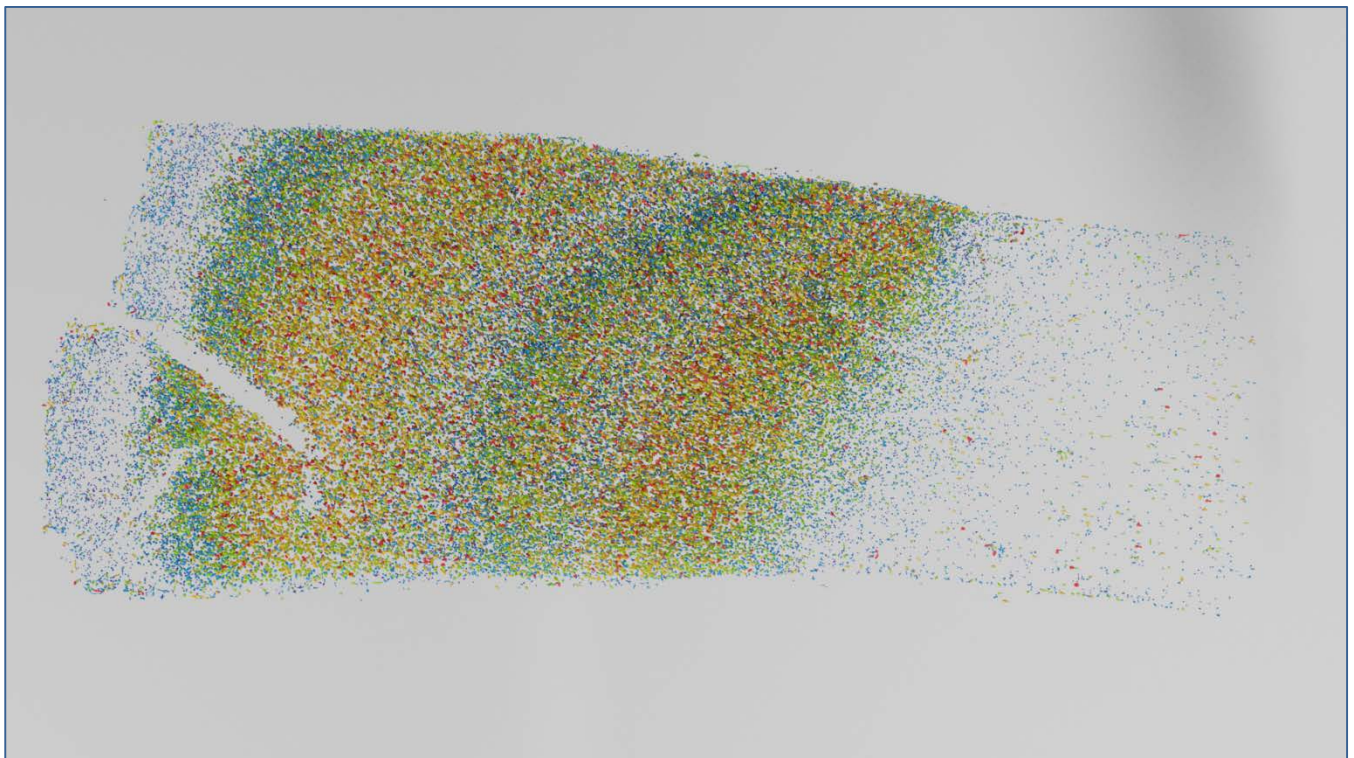

Figure 5: Colour-coded neuronal rendering of sample 1.

Volumes: Blue 400-2,500 μm^3, light blue 2,500-4,500 μm^3, green 4,500-8,500 μm^3, yellow 8,500-12,000 μm^3, red ≤12,000 μm^3

4.2   Potential Use of Output(s)

With respect to the results presented in the previous section and from a broader perspective, the workflow presented here, based on the machine learning approach, allows to scale up and create many more maps of cytoarchitectonic areas over the coming years, working towards a full mapping of the brain. While our approach was developed with fluorescence images, it is not restricted to

TPFM. It could be adapted to analyse images from different modalities (e.g. Confocal Microscopy, Light Sheet Fluorescence Microscopy) after proper re-training with specific ground truth. For example, this workflow could be applied to analyse the images obtained with the Dual-View Inverted Confocal Light-Sheet Fluorescence Microscope (di2CLSFM) optimised during HBP SGA2 (Component C2349: "Integration of ultra-fast large area fluorescence microscope with Switch clearing techniques"). Indeed, it could be used to analyse the hippocampal region acquired during SGA2 with the di2CLSFM (Component C2293: "Layer-specific excitatory and inhibitory neuronal maps of hippocampus").

The 3D reconstruction of the neuronal volume enables extracting morphological information that allows discriminating cell types using a general staining such as NeuN, reducing the number of labels necessary for the analysis (a critical point in human tissue preparation). Moreover, in the future, variability in volume could be used in pathological analyses to assess, more reliably and without visual artefacts (different orientation or distribution), the morphological alteration of neurons, even in the presence of sparse features. Indeed, the 3D evaluation increases the statistical accuracy and the sensitivity of the assessment, as already demonstrated for other tissues in tumour diagnostics and in stratifying patient prognosis (Tanaka *et al*. Nat Biomed Eng 2017).

Given the extent of the imaged samples and the number of cells involved, automated approaches such as the one presented here are the only possibility for performing quantitative analysis in a systematic fashion at a large scale. What was achieved here is a fundamental first step in what we envision as "3D digital histology". The data obtained will provide new and more complete information of the cellular organisation of the tissue at high resolution that could be used in brain models and simulation.

5. Dissemination

5.1 Publications

- P2581 Mazzamuto, G., Costantini, I., Neri, M., Roffilli, M., Silvestri, L., & Pavone, F. S., "Automatic Segmentation of Neurons in 3D Samples of Human Brain Cortex." Springer, Cham. Conference (pp. 78-85): 21st International Conference on the Applications of Evolutionary Computation (EvoApplications) Location: Parma, Italy Date: APR 04-06, 2018. DOI: https://doi.org/10.1007/978-3-319-77538-8_6.

- P2582 Costantini I., Mazzamuto G. et al. A combined pipeline for quantitative analysis of human brain cytoarchitecture. DOI: https://doi.org/10.1101/2020.08.06.219444 (*SGA3 publication*)

5.2 Posters, Lectures, Presentations and Talks

The work described here has been presented at the following conferences:

- 6 April 2018: "Automatic Segmentation of Neurons in 3D Samples of Human Brain Cortex" Poster at the International Conference on the Applications of Evolutionary Computation (EvoApplications), scientific community, Parma, Italy.

  Proceedings paper: Autors: Mazzamuto, G., Costantini, I., Neri, M., Roffilli, M., Silvestri, L., & Pavone, F. S. Title: "Automatic Segmentation of Neurons in 3D Samples of Human Brain Cortex." Springer, Cham. Conference (pp. 78-85): 21st International Conference on the Applications of Evolutionary Computation (EvoApplications) Location: Parma, Italy Date: APR 04-06, 2018.

- 3-6 April 2018 "Software Tools for Efficient Processing of High-Resolution 3D Images of Macroscopic Brain Samples", Poster at Optics and the Brain, Hollywood, Florida, United States.

  Proceedings abstract: Authors: Mazzamuto G., Silvestri L., Costantini I., Orsini F., Neri M., Roffilli M., Frasconi P. & Pavone, F. S. Title: "Software Tools for Efficient Processing of High-Resolution 3D Images of Macroscopic Brain Samples"; Optics and the Brain 2018, Hollywood, Florida, United States

- 9-14 July 2018 G. Mazzamuto et al. "A processing pipeline for big data in high-resolution microscopy" Presentation and poster by Giacomo Mazzamuto at the ESMI Imaging Technology Summer Workshop on "BIG DATA in IMAGING" TOPIM TECH 2018. Chania, Greece.

- 8-10 September 2018 "Effective management and analysis of ultra-terabyte brain images", Presentation by Ludovico Silvestri, International Workshop Whole-brain neuron reconstruction and applications - NRA'2018 Nanjing, China

- 15-18 October 2018 Posters and presentations at the HBP Open Day and HBP Summit, Maastricht, The Netherlands. "Distribution and morphological characterization of neurons in the human hippocampus" (SP2-SP6 meeting), "CDP1 Whole mouse brain atlas and model"

- 9-11 January 2019 "Scalable analysis of ultra-terabyte brain images with deep learning", Poster by Ludovico Silvestri, Quantitative BioImaging Conference, Rennes, France.

- 2-7 February 2019 "Processing Big Data in High-Resolution 3D Microscopy of Brain Samples", Presentation by Giacomo Mazzamuto, BIOS at SPIE Photonics West, San Francisco, California, United States.

- 14-17 April 2019 "3D stitching of ultra-terabyte brain images and feature extraction with deep learning", Poster by Ludovico Silvestri at Focus on Microscopy, London, United Kingdom

- 23-27 June 2019 "Deep learning strategies for scalable analysis of high-resolution brain imagery", presentation by Giacomo Mazzamuto, ECBO European Conference on Biomedical Optics, Munich, Germany.

  Proceedings Abstract: Authors: Mazzamuto G., Costantini I., Orsini F., Roffilli M., Frasconi P., Pavone, F. S. & Silvestri, L. Title: "Deep learning strategies for scalable analysis of high-resolution brain imagery", ECBO 2019, European Conference on Biomedical Optics, Munich, Germany.

- 16 June 2019 "Large-scale high-resolution imaging of biological samples with advanced microscopy Techniques". Invited lecture by Giacomo Mazzamuto at the V Summer School "Photonics Meets Biology", Heraklion, Greece.

Annex: Components

List of components mentioned in the text:

- C2347: "Multiple staining for characterisation of different neuronal types", Task T2.3.2, WP2.3. The component aims at implementing a new clearing protocol to perform 3D staining of volumetric human brain tissue sample.

- C2349: "Integration of ultra-fast large area fluorescence microscope with Switch clearing techniques, Task T2.3.2, WP2.3. Development of a new custom-made light-sheet fluorescence microscopy apparatus to perform 3D reconstruction of volumetric cleared samples.

- C2354: "Morphological classification of neurons with immunohistological staining", Task T2.6.4, WP2.6. Development of a neural network model for segmentation and morphological classification of neurons. The topic of this deliverable.

- C1745: "Whole-brain maps of different neuronal types", Task T1.3.1, WP1.3. Production of maps of different neuronal types across the entire mouse brain. Different types of neurons are discriminated based on the molecular marker expressed by the transgenic line.

- C2293: "Layer-specific excitatory and inhibitory neuronal maps of hippocampus", Task T2.3.1, WP2.3. Acquisition of a human brain hippocampus with an advanced fluorescent microscope to perform acquisition of excitatory and inhibitory neurons in 3D in combination with specific clearing and staining techniques.