

# Cell nuclei segmentation in mm-scale x-ray holographic nanotomography images of mouse brain tissue

Andrea Nathansen<sup>2</sup>, Matthis Clausen<sup>2</sup>, Manuel Berning<sup>1</sup>, Ethan MacKenzie<sup>3</sup>, Yuxin Zhang<sup>1,4</sup>,  
Alexandra Pacureanu<sup>5</sup>, Andreas T Schaefer<sup>1,4</sup>, Norman Rzepka\*<sup>2</sup>, Carles Bosch\*<sup>1</sup>

1 Sensory Circuits and Neurotechnology Lab, The Francis Crick Institute, London, UK.

2 scalable minds GmbH, Potsdam, Germany.

3 Phospho Biomedical Animation, The Greenhouse Studio 6, Unit 17 Millmead Industrial Estate,  
London, N17 9QU, UK

4 Department of Neuroscience, Physiology and Pharmacology, University College, London, UK.  
5 ESRF, The European Synchrotron, Grenoble, France.

\*Correspondence: norman.rzepka@scalableminds.com; carles.bosch@crick.ac.uk

**Keywords:** image segmentation, brain, Xray, neuroscience, biological tissues, holographic nanotomography, connectomics, olfaction

## ABSTRACT

Biological soft tissues are functional agglomerates of cells. They constitute the microenvironment where intercellular communication occurs. In turn, their woven structure underlies mechanical properties that contribute to their roles in the context of the organs and the organisms that contain them. Therefore, determining the density and spatial distribution of cells within the tissue offers key information for understanding its physiological properties and its state. X-ray holographic nanotomography is a non-destructive imaging technique capable of resolving subcellular details in biological tissues that has shown promising advantages to study the structure of neuronal circuits. However, the dimensions of the datasets required – covering volume landscapes of  $\sim\text{mm}^3$  – make manual annotation of individual nuclei an unrealistic task. We developed and trained an automated image segmentation classifier that accurately detects and segments cell nuclei in mouse brain tissue imaged with X-ray holographic nanotomography, and that generalises to similar datasets obtained from biological replicates with minimal additional ground truth. It provides the spatial locations and morphologies of the  $\sim 80\text{k}$  nuclei per dataset with a high recall. It harnesses the strengths of a high-performance computing cluster and embeds the curated results in two main simplified outcomes: a data table and explorable image segmentations and meshes associated with the original dataset, in a browser-compatible format that simplifies proofreading by multiple users. The classifier we present here can be readily integrated into an automated analytical pipeline for histological datasets obtained with synchrotron X-ray holographic nanotomography in the context of systems neuroscience as well as broader tissue life science studies.

## 1. INTRODUCTION

Biological soft tissues are functional substrates at the intersection between the cell and the organ-organism levels. From a functional perspective, they play critical roles at both scales: On one hand, they constrain how neighbouring individual cells can interact with each other. On the other hand, their woven structure provides emerging mechanical characteristics that become a critical part of their roles, e.g. defining essential properties in smooth, skeletal or cardiac muscles. When looking at the brain, histology becomes an integral part in the study of neuronal circuits – ensembles of interconnected neurons constrained in space that perform a certain computation. However, while tissues have long been studied at the macroscopic [1] and at the microscopic level [2], a multiscale study of biological tissues has not been technologically feasible until recently: resolving subcellular features across  $1\text{ mm}^3$  requires leveraging 3D image datasets in the terabyte range.

Automation of volume electron microscopy [3-6] has unlocked the potential to bridge those two scales and already provided insights on how different tissues operate [7-9]. More recently, other nondestructive approaches have shown the capacity to operate in this regime, either employing light [10, 11] or X-rays [12, 13] as illumination source. In particular, X-ray holographic nanotomography (XNH) [14] has proven capacity to resolve subcellular features across landscapes that extend to the mm scale in samples with preserved ultrastructure and can be compatible with follow-up volume electron microscopy [12, 15, 16].

In turn, automated image segmentation algorithms have seen an unprecedented improvement in recent years and high-performance compute clusters have become widely accessible to the materials and life sciences domains [17, 18]. Pioneering efforts in volume electron microscopy have provided essential datasets that facilitated training the first generation of connectome-delivering classifiers [19-21].

However, while obtaining mm-scale volume datasets at subcellular resolution is becoming a routine approach, instance classifiers that generalise well across independently sourced datasets of the same tissue are not common, resulting in a bottleneck in the data analysis pipeline. Here we aim to address this bottleneck by developing a nuclei classifier for mouse olfactory bulb tissue.

Cells are the essential components of tissues, and therefore the distribution and shape of their nuclei defines essential properties of the tissue itself. But their manual annotation is unfeasible already at the mm-scale: 1 mm<sup>3</sup> of mouse brain tissue contains on the order of 2\*10<sup>5</sup> nuclei and a 0.5 cm<sup>3</sup> mouse brain contains 100\*10<sup>6</sup> [20, 22]. Moreover, cell nuclei display a characteristic set of morphological traits that make them good targets for automated image classifiers: they have a compact shape, with volumes within a narrow range, and the ultrastructure of their interior - the nucleoplasm - often displays a distinct pattern to the adjacent cytosol of the cell that contains it. Therefore, classifiers aimed at segmenting nuclei can be trained to recognise not only the thin nuclear membrane but also the nucleoplasm.

While several tools have been developed to detect subcellular features in volume EM images [23-25], specific solutions to automatically retrieve nuclei across mm-scale datasets acquired by XNH are not yet developed. Those classifiers will provide unique insights on the typical counts, sizes and distributions of cells within the context of tissues.

Here, we developed and trained an image classifier that reliably segments cell nuclei in mouse brain tissue volumes acquired with XNH. It operates integrated on a high performance computing cluster, providing straightforward scalability to larger volumes within the terabyte scale. It detects the shapes, volumes and positions of cell nuclei in tissue volumes from the mouse olfactory bulb with high precision and recall across multiple cell layers and generalises well to analogous datasets. This makes it an essential tool integrated in a broader image analysis pipeline devoted to mammalian systems neuroscience addressing how neuronal circuits compute information given their connectivity [16, 26]. Furthermore, it is also applicable to the study of the structure-function of other biological tissues.

## 2. RESULTS

We first aimed to automatically detect nuclei in a 3D volume of mouse brain olfactory bulb acquired with X-ray holographic nanotomography. The dataset spanned across 0.456 mm<sup>3</sup> with voxels of (100 nm)<sup>3</sup>, containing a continuous tissue volume of 0.212 mm<sup>3</sup> and was obtained after a non-rigid stitching of 28 individual tiles (**Fig. 1a, Table 1**) [27]. We developed an image segmentation pipeline that provided an automated volumetric labelling for each nucleus inside a 0.197 mm<sup>3</sup> mask that covered the vast majority of the tissue and avoided non-tissue areas (e.g. those regions occupied by bare resin or empty space around the sample) (**Fig. 1b,c**), computed individual meshes for each nucleus (**Fig. 1d**) and returned an indexed table with essential morphometric parameters of each segmented item: its volume, its surface and its sphericity (**Fig. 1d**) [28]. This segmentation pipeline consists of two major steps in sequence (**Fig. 1e**): an automated instance segmentation relying on a trained U-Net architecture [29] followed by a heuristic filtering of segmented instances based on the expected volumes and sphericities of nuclei.

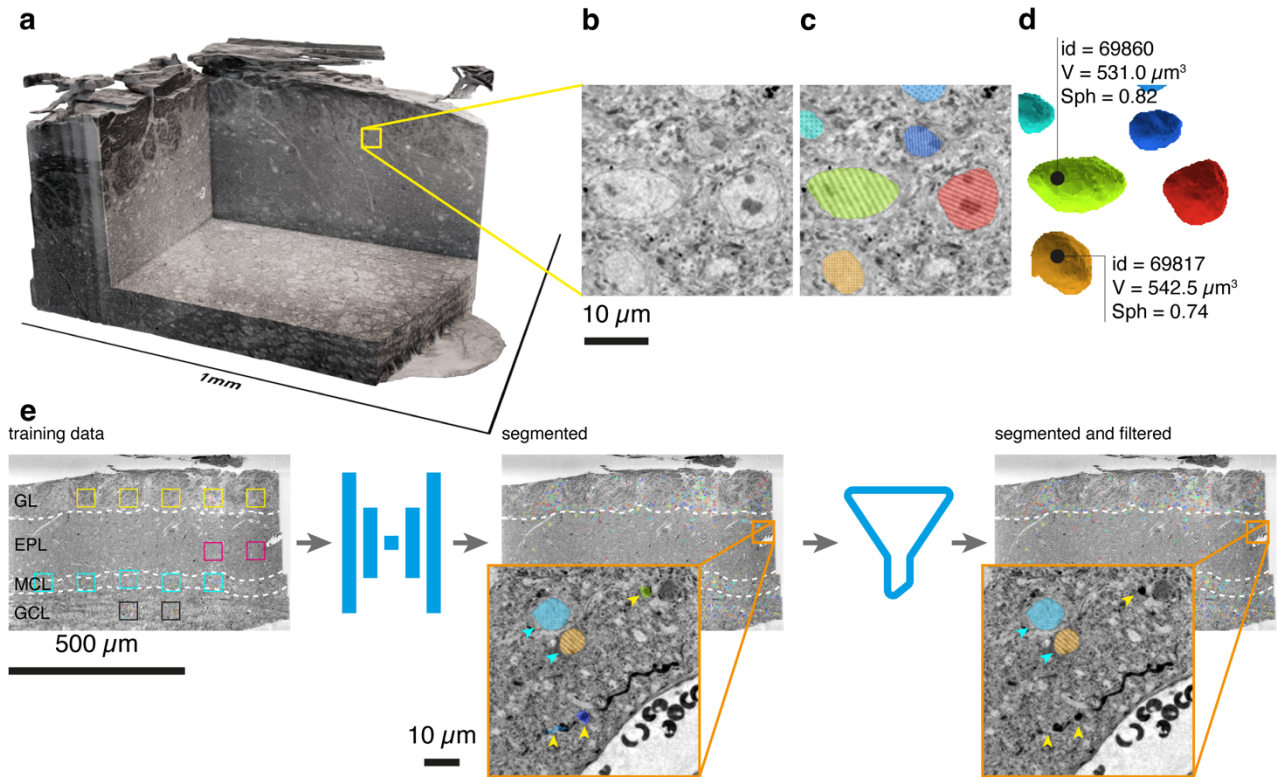


Figure 1. Workflow to segment cell nuclei in tissue volumes obtained with X-ray holographic nanotomography.

(a) XNH can provide continuous volumes mapping mm-scale landscapes with subcellular resolution. (b-c) Nuclei can be detected and segmented automatically, providing a quantitative insight on their morphologies and distribution in space (d). (e) The classifier used consists of a trained U-Net followed by a filter based on prior knowledge on the morphology of the targeted features. The arrowheads indicate correctly segmented nuclei (cyan) and segmentation artefacts discarded by the filtering step (yellow). *GL*, glomerular layer; *EPL*, external plexiform layer; *MCL*, mitral cell layer; *GCL*, granule cell layer.

The volume mapped in this dataset contains all histological layers in the olfactory bulb - each displaying a characteristic pattern of nuclei in terms of their individual size as well as in their spatial arrangement (density and distribution). We defined subvolumes where all nuclei contained were manually annotated at a 2-fold downsampled version of the datasets:  $(200 \text{ nm})^3$  voxels. A total of 30 subvolumes sampled 4 histological regions of interest: glomerular layer (*GL*, 10 subvolumes), external plexiform layer (*EPL*, 5 subvolumes), mitral cell layer (*MCL*, 10 subvolumes) and granule cell layer (*GCL*, 5 subvolumes) (**Fig. 2a, b**). Each of those subvolumes was  $(50 \text{ }\mu\text{m})^3$  in size to ensure large  $20 \text{ }\mu\text{m}$ -wide nuclei would be fully contained in the training data even in the regions where their density is the lowest, and therefore minimise split errors of the classifier. We make all the training and test data openly available for further reuse (see **Data Availability** section). The segmentation workflow consists of a 3D U-Net [29] to obtain a distance prediction, followed by a watershed algorithm [30] to generate an instance segmentation of cell nuclei (**Fig. 2c**). Due to the large size of the datasets, the prediction and segmentation was performed in smaller chunks which were merged afterwards.

Nuclei were positively identified in most cases (**Fig. 2d, f**) and rarely missed. However, certain features not constituting nuclei were also often erroneously classified as such (**Fig. 2e**). Manual inspection of the false positive segments revealed that these were features of diverse nature but with morphologies very distinct from nuclei. An unbiased analysis of detection accuracy confirmed this insight: truly detected nuclei clustered together in a parameter space defined by the volume and sphericity of the segmented features (**Fig. 3e**). An optimal double-threshold boundary therefore allowed filtering segmented nuclei from the pool of all segmented features: we defined both high-pass thresholds (volume and sphericity) by applying a grid-search in 2-fold cross-validation over 12 subvolumes of size  $(50 \text{ }\mu\text{m})^3$  (**Methods**). Together, the filtered result of the classifier provided a final segmentation for this dataset, consisting of 64166 nuclei (**Fig. 3g**) with

a precision and recall of above 90% across most regions (Fig. 3f). Each detected nucleus was stored with an associated identifier, a computed mesh and a summary of its key morphological characteristics (volume, surface and sphericity) (Fig. 3g, h, i).

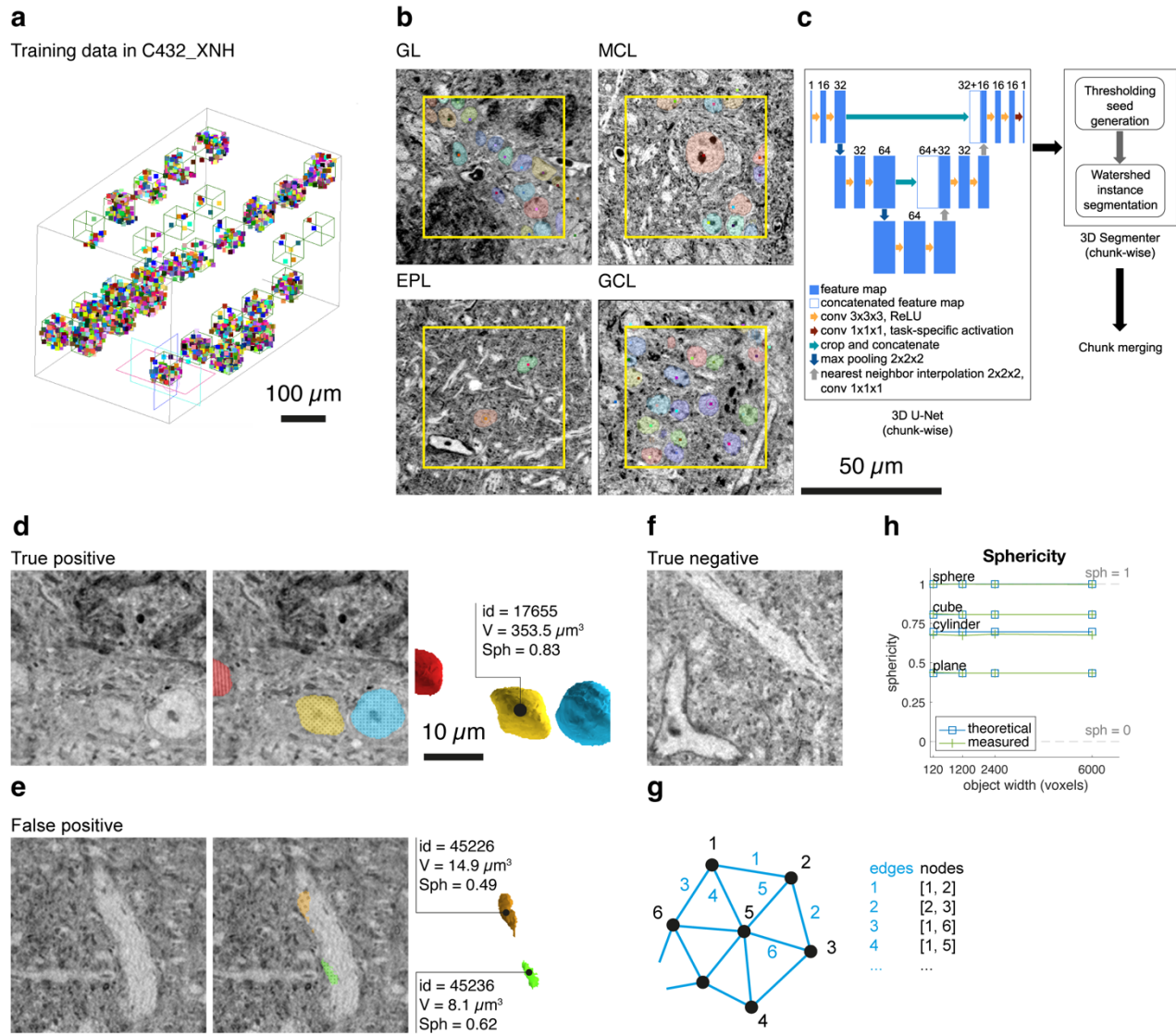


Figure 2. Architecture of the classifier

(a) Training data was obtained by manually annotating all nuclei present in 30 regions of interest each covering  $(50 \mu\text{m})^3$ , distributed across all relevant histological layers of the olfactory bulb tissue in the dataset. (b) Examples of manually annotated nuclei in the four layers mapped. (c) Diagram of the U-Net [29] and following segmentation steps.

(d-f) This classifier provided a good result detecting nuclei (d), while some voxels not belonging to nuclei in very rare occasions were also classified as such (e). Similarly, while most non-nuclear features were correctly ignored (f) in very rare occasions true nuclei were missed. (g) The 3D segmentations were later converted to standard meshes and stored as such, associated with the original dataset. (h) The measured sphericity of closed meshes extracted from segmented objects matched expected theoretical values for diverse regular polygon shapes across a broad range of sizes. *GL*, glomerular layer; *EPL*, external plexiform layer; *MCL*, mitral cell layer; *GCL*, granule cell layer.

Systems neuroscience - and tissue biology more generally - challenges hypotheses through experiments whose readouts are often vulnerable to sampling biases and sample preparation procedures. This vulnerability can be sometimes overcome

by incorporating sufficient biological replicates in the experimental design. Two more independent datasets reporting continuous tissue volumes of 0.269 mm<sup>3</sup> and 0.297 mm<sup>3</sup> of the same brain region were acquired in a similar way with XNH, providing a cohort of 3 comparable datasets (**Fig. 3d, j, p, Table 1**). We applied the segmentation pipeline without modifications to the second dataset (Y391, 0.261 mm<sup>3</sup> mask covering most tissue) (**Fig. 3j, Table 1**) and computed new dataset-specific thresholds for volume and sphericity using the same method as for C432 (see **Methods**) (**Fig. 3k, Table 1**). This provided an automated segmentation of 76820 nuclei with a precision and recall above 75% in most regions (**Fig. 3l, m**).

In order to provide a generalisable approach for further datasets, nuclei in the third dataset (Y489) (**Fig. 3p, Table 1**) were segmented using a battery of different classifiers incorporating new additional local ground truth (**Fig. 3c**). A total of 29 subvolumes each covering (50 μm)<sup>3</sup> and distributed across layers of interest (10 GL, 5 EPL, 10 MCL and 4 GCL) were manually annotated for nuclei (**Fig. 3a-b**). Multiple classifiers were trained using different sets of training data: from the unmodified classifier to versions adding groups of 5 additional subvolumes sampling different histological layers from the Y489 dataset, to a classifier only trained with the new 29 subvolumes. All classifiers were used to independently detect all nuclei in the same Y489 dataset (**Fig. 3c**). All classifiers showed high precision (>95%) when evaluated on a test set of 14 subvolumes of (50 μm)<sup>3</sup>. While the unmodified classifier on the new dataset showed only a 90% recall, addition of 5 subvolumes to the subset of training data raised the overall recall rate to >95%. The GL and GCL are the two most demanding tissue layers for a nuclear classifier: there, cell nuclei are very densely packed, often with little cytoplasmic volume in cells whose cell bodies sit directly adjacent to each other (**Fig. 3b**). The amount of features available to avoid misclassification are therefore minimal - often only the nuclear and plasma membranes (as opposed to the histological layers EPL and MCL, where cytoplasmic content and neurites of the same and other neurons are often present between different nuclei). When observing the performance of those classifiers at the most stringent regions, the addition of all 29 new subvolumes with locally sourced new training data did not harm performance and in some cases it would seem to improve the segmentation of densely packed nuclear clusters in the GCL. Therefore, the final classifier for Y489 used was the one trained on all previous and new ground truth data (59 subvolumes (7.35\*106 μm<sup>3</sup>) of training data). It provided an automated segmentation of 81779 nuclei (**Fig. 3s**), with a precision above 95% and recall above 90% across most regions (**Fig. 3r**).

**Table 1.** Compute time used to run the instance segmentation pipeline and post-processing on each dataset at (200 nm)<sup>3</sup> voxels (ie, 2-fold downscaling applied to original data). The classifier pipeline was run on a Slurm cluster on several CPU nodes equipped with an AMD EPYC 7763 CPU (hyperthreading enabled) of which we used up to 12 cores and 50 GB RAM each. The mentioned times include queuing times for sub-tasks.

dataset	total volume	volume of tissue	volume processed	Avg precision	Avg recall	voxel prediction time	segmentation	meshes	sphericity	Total
C432_XNH	0.456 mm <sup>3</sup>	0.212 mm <sup>3</sup>	0.197 mm <sup>3</sup>	0.975	0.962	2h 39min	4h 38min	0h 30min	0h 40min	8h 27min
Y391_XNH	0.685 mm <sup>3</sup>	0.269 mm <sup>3</sup>	0.261 mm <sup>3</sup>	0.934	0.916	2h 30min	4h 50min	1h 49min	0h 25min	9h 34min
Y489_XNH	0.713 mm <sup>3</sup>	0.297 mm <sup>3</sup>	0.282 mm <sup>3</sup>	0.943	0.980	10h 44min	14h 19min	6h 41min	0h 26min	32h 10min

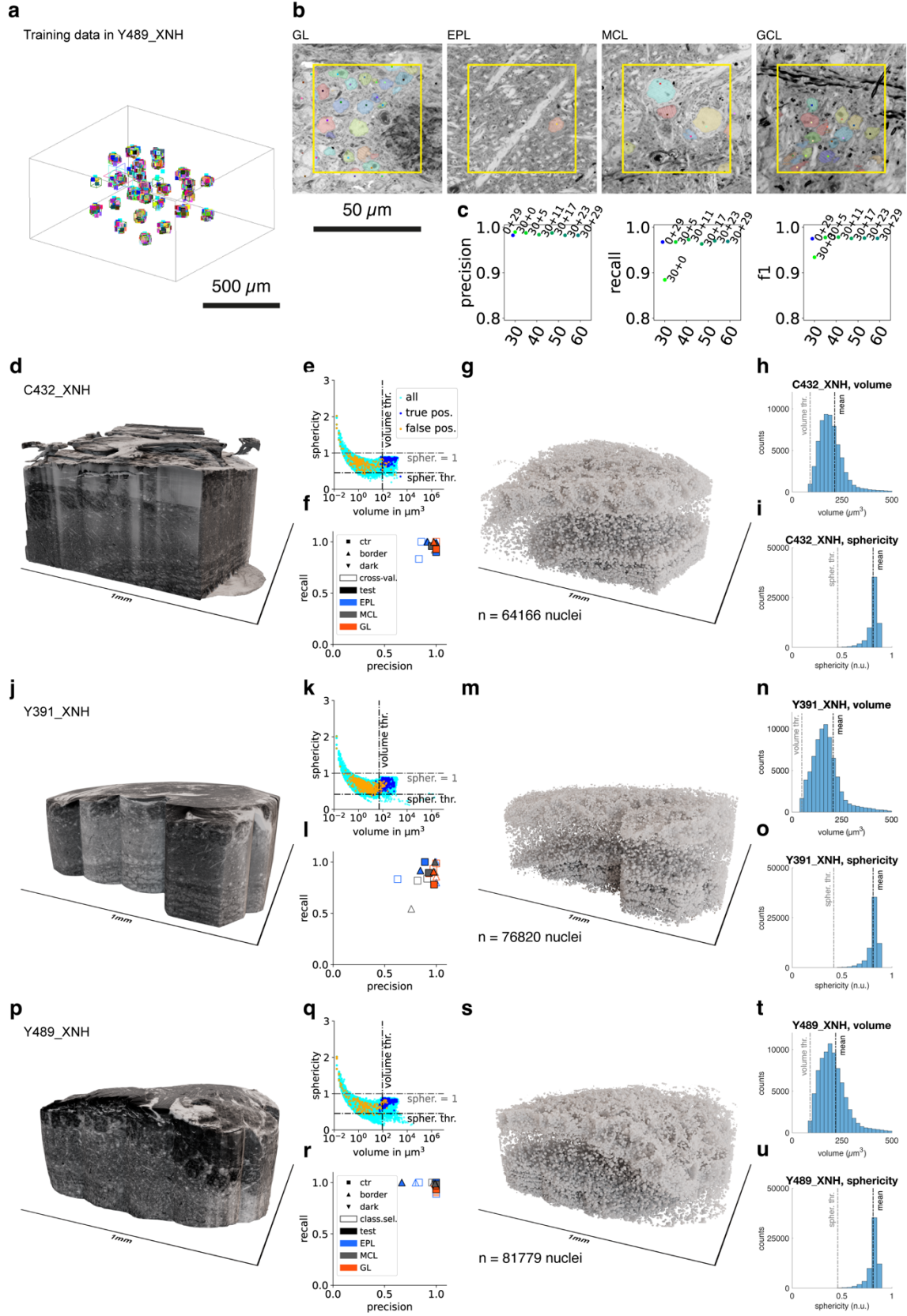


Figure 3. Generalisation to other datasets of the same imaging modality and similar characteristics

(a) Training data was also obtained from a third independently acquired dataset from a new sample that mapped the same brain region. (b) Examples of manually annotated nuclei in the four layers targeted. (c) Precision, recall and F1 score of the classifier operating on the new dataset when trained with increasing amounts of data sourced from the new dataset. (d) Volume rendering of the first dataset. (e) Objects segmented, distributed according to each item's volume and sphericity. Prior knowledge allowed describing thresholds for both metrics, providing a filtered result that detects nuclei with high accuracy. (f) Precision and recall on the filtered results at multiple regions of the dataset, including high-resolution tile centers (ctr) as well as tile border regions (border) from all histological regions. (g) Volume rendering of the meshes of all nuclei segmented from this dataset. (h-i) Volumes (h) and sphericities (i) of the segmented nuclei in this dataset. (j-o) Same as d-i for the second dataset, segmented with the same classifier without any additional training data. (p-u) Same as d-i for the third dataset, using a classifier that incorporates all the newly sourced training data (performance shown in (c), 30+29 subvolumes). *GL*, glomerular layer; *EPL*, external plexiform layer; *MCL*, mitral cell layer.

### 3. DISCUSSION

We have developed an image classifier that accurately segments cell nuclei in 3D maps of mouse brain tissue obtained with X-ray holographic nanotomography. This classifier not only provides satisfactory results within the dataset it has been trained on, but also generalises its performance to analogous datasets obtained independently with minimal addition of locally sourced training data. In this line, we observed a drop in recall when segmenting the Y391 dataset, which could be attributed to two factors: poorer signal-to-noise ratio (SNR) of the dataset itself (attributable to being reconstructed using tomograms taken at three instead of four distances) and no addition of locally sourced training data. In dataset Y489, where both factors were amended, we obtained F1 scores >95%. When comparing its performance within the same dataset, across regions with high SNR (tile center) and low SNR (tile border), as expected better performance is achieved where SNR is best while providing in all cases a sufficient yield for the intended analyses. Therefore, this classifier will be a useful component of an integrated analytical pipeline devoted to analysing mm-scale 3D maps of tissue ultrastructure in the mouse brain.

It operates on datasets that map the dense histology of samples with preserved ultrastructure: the contrast in the images is driven by the original distribution of proteins and lipids in space. Its performance on analogous independent datasets suggests that the classifier is robust enough to cope with variations of signal and noise patterns in the data. One would therefore expect this classifier to provide acceptable results on volumetric datasets reporting tissue ultrastructure obtained with any imaging technique that shares the essential features with XNH, such as the data being aligned within < 100nm precision in all directions, having isotropic voxels, and the contrast being driven by the distribution in space of all proteins and lipids, with an additional set of locally sourced training data comparable to the one reported in our third dataset. While focussed ion beam scanning electron microscopy and ptychographic X-ray tomography would most likely fit these requirements, other volumeEM techniques such as serial block-face electron microscopy could also be eligible provided that the quality of image registration meets the required precision. On the other hand, its performance on datasets obtained with imaging techniques that target serial sections (such as ssTEM, gridTape or ATUM-mSEM) will have to be assessed independently since these will incorporate slicing-specific artefacts such as cracks and folds that have not been included in the training data.

This classifier structure proved extremely efficient and robust to detect nuclei in XNH datasets. We expect other features mapped in these datasets - such as volume occupied by tissue, resin or blood vessels - to be detectable by specifically-trained classifiers with a similar architecture, possibly requiring adjustments related to the biological nature of the features. Moreover, promising classifiers that segment mitochondria in vEM datasets [24] could be leveraged and incorporated in the analytical pipeline to further broaden the insights one can quickly gather automatically from newly acquired XNH datasets.

However, segmentation of features in tissue is fundamentally limited by resolution: biological tissues are composed by cells and organelles, and both features are delimited by a 20 nm-thick bi-layered lipidic membrane. While resolving this essential 20 nm feature enables segmenting any cell or subcellular component [7, 31-33], relevant merging errors should be expected when aiming to segment organelles with coarser resolutions.

The analysed areas of the three datasets reported cover mouse brain olfactory bulb tissue volumes of  $0.197 \mu\text{m}^3$ ,  $0.261 \mu\text{m}^3$  and  $0.282 \mu\text{m}^3$ , and contain 64166, 76820 and 81779 nuclei with an average size of  $212.9 \pm 0.5 \mu\text{m}^3$ ,  $207.9 \pm 18.9 \mu\text{m}^3$  and  $217.7 \pm 0.4 \mu\text{m}^3$  and an average sphericity of  $0.8114 \pm 2 \cdot 10^{-4}$ ,  $0.7871 \pm 3 \cdot 10^{-4}$  and  $0.7849 \pm 2 \cdot 10^{-4}$ , respectively (mean  $\pm$  SEM). Addressing these annotations manually would have been unfeasible: allowing 10 seconds to find, annotate and proof-read each individual nucleus would involve  $\sim 200$  human-hours per dataset only to find their locations. Fully segmenting them in 3D (to extract their volumes and sphericities), allowing 24 min/nucleus, would take  $\sim 30 \cdot 10^3$  human-hours per dataset. Recent developments in X-ray imaging and in volumeEM are boosting the capacity to generate these datasets at scale – enabling experimental timeframes compatible not only with the acquisition of singular proof-of-concept datasets but of controlled dataset cohorts like the one reported in this study. This technological capacity is already delivering novel insights on the function of biological soft tissues, led largely by the field of connectomics, focussed on the study of neuronal circuits. In mammals, neuronal circuits are embedded in volumes not smaller than  $1 \text{ mm}^3$  [34], and a whole mouse brain at connectomic resolution is regarded as a feasible aim within a decade [35, 36]. Processing the datasets that will be acquired in the upcoming years will necessarily require a robust data architecture that uses computational resources efficiently. This classifier ran at an average rate of 387 min per processed  $0.1 \text{ mm}^3$  (**Table 1**). We anticipate it being operative at this regime for up to  $1 \text{ mm}^3$  datasets. For larger volumes, there are opportunities for optimization, such as reducing scheduling overheads on the Slurm cluster and utilising GPUs for predictions. With these optimizations, we expect processing volumes of  $10 \text{ mm}^3$  (equivalent to the entire olfactory bulb of an adult mouse) to take a few days on a similarly equipped Slurm cluster.

Exploring, annotating and proof-reading terabyte-scale datasets requires a data architecture that ensures robustness of the original dataset over long periods of time, that facilitates sharing and access by many users from diverse locations worldwide, that streams efficiently the data users require using the minimal bandwidth necessary, that integrates well with high-performance image analysis computational routines, and that relies on an open-access data format ensuring long-term support for both datasets and metadata. OME-Zarr has been suggested as the consensus data format that can meet all these requirements: this chunked data format allows any user to explore a petascale dataset from their browser, it is optimised for distributed computing routines addressed at high performance clusters and is open-source [37, 38]. We employed a Zarr-compatible dataset architecture, WEBKNOSSOS [39] to store, explore and annotate our datasets. We release all the densely annotated training data and tagged test data in this format, enabling us and others to develop more accurate segmentation algorithms in the future.

Altogether, we have developed an image classifier that accurately segments nuclei in 3D mouse brain datasets obtained with X-ray holographic nanotomography. This classifier generalises well to other datasets of independently processed and imaged samples with minimal additional training data and operates natively on Zarr-compatible datasets stored in a WEBKNOSSOS architecture. It makes it an integral tool of a scalable image analysis pipeline for future datasets exploring mouse brain tissue with X-ray holographic nanotomography at the  $\text{mm}^3$  scale. This classifier will provide biologically relevant insights on terabyte-scale volume datasets mapping biological tissues early in the analytical pipeline, and it will pave the way to establish robust image segmentation solutions for even larger datasets.

## 4. MATERIALS AND METHODS

### Animals

Animals used in this study were 9-12 week old wildtype mice of C57Bl/6 background of either sex. All animal protocols were approved by the Ethics Committee of the board of the Francis Crick Institute and the United Kingdom Home Office under the Animals (Scientific Procedures) Act 1986.

### Sample preparation

A detailed description of these samples and datasets will be published in a separate manuscript currently under preparation. Briefly, fresh tissue slices  $0.5 \text{ mm}$  thick,  $\sim 3 \cdot 3 \text{ mm}^2$  wide were dissected, fixed with paraformaldehyde and glutaraldehyde in ice-cold isotonic buffer (sodium cacodylate  $0.15 \text{ M}$ ,  $\text{pH } 7.40$ ), stained with heavy metals following a ROTO protocol for volume electron microscopy, dehydrated and embedded in Epon812 or equivalent epoxy resin [26, 40, 41]. Samples were scanned with a Zeiss Versa 512 microCT to verify correct staining throughout and further imaged at propagation-based microtomography beamlines at Diamond I13-2 and PSI TOMCAT. These datasets guided the targeted trimming of the



parent specimen down to a <1mm diameter, <1mm height sample that contained all relevant tissue regions using either ultramicrotomy tools or a femtosecond laser [42].

### **X-ray holographic nanotomography**

The X-ray holographic nanotomography datasets were acquired at the ID16A beamline of the European Synchrotron in Grenoble, France. The beam energy used was 17 keV for C432 and 33 keV for Y391 and Y489. C432 and Y391 were acquired at cryogenic conditions, with 300 ms exposure time per rotation angle as the sample rotated through 180° in 1600 or 2000 steps, respectively. To reduce ring artefacts, the samples were displaced laterally with a random jitter at each rotation angle. Sample Y489 was acquired at room temperature with continuous scanning technique, where the sample rotated non-stop though 180° while 2000 projection images were taken with 150 ms exposure time per frame. Each sample location was imaged twice with a 5 µm displacement in x, y and z. This allowed a Noise2Noise [43] based denoising algorithm to be applied using a 3D U-Net architecture [29]. For tomogram reconstruction, for each angle, projections from all distances were first aligned then used to compute a phase map [14, 44, 45]. A 3D tomogram was then obtained by combining phase maps from all angles using filtered back-projection [46]. Each reconstructed tomogram had a high-resolution core corresponding to the overlap in field of view of all distances ((200 µm)<sup>3</sup>) and a progressively lower resolution edge that extends to the full field of view ((321.6 µm)<sup>3</sup>). To cover all regions of interest, 28, 32 and 36 overlapping tomograms were acquired, reconstructed and then stitched with NRstitcher [27] for datasets C432, Y391 and Y489, respectively.

### **Instance segmentation workflow**

The segmentation pipeline is built with Voxelytics (scalable minds, Potsdam, Germany) and runs on Python 3.11 with PyTorch 2.1.1.

### **Image annotation**

Subvolumes were defined as bounding boxes at specific locations in the dataset in WEBKNOSSOS. Each subvolume was then submitted as an individual task to an expert annotator (Mindy Support), only displaying data downsampled two-fold to (200 nm)<sup>3</sup> voxels. Annotators were briefed to locate all nuclei in their assigned subvolumes with single-node skeletons, and to then complete full volume manual segmentation assigning each nucleus an independent segment identifier. It took on average 24 minutes to locate and manually segment each nucleus. Annotations in all subvolumes always retained the relationship to the original data (the volumes presented were a masked version of the original dataset and coordinates were always those of the whole dataset).

### **Data preparation**

Classifier training required initial conversion of training data to uint8, i.e., constraining the range of voxel values to 0-255. The conversion started with clipping the input data to a narrower range (C432: [52562.20, 54917.48], Y489: [-92.7594, 70.8168]), followed by rescaling to the uint8 range. Then, we applied normalisation on the uint8 data (C432: mean 130.97, std: 53.75; Y489: mean: 102.15, std: 35.77). We split the annotated segmentation labels into patches of 76x76x72 voxels. The related input data patches had a size of 116x116x112 voxels to account for the context required by the U-Net. If the context lied outside of the data bounding box, it was filled with black voxels. We used a train:validation ratio of 80:20 and a batch size of 4. Data augmentation included random mirroring along any axis, random rotation by multiples of 90 degrees around the z axis, and adding noise with a randomly chosen standard deviation between 0 and  $\frac{10}{dataset\ std}$ . Further, we augmented brightness and contrast by histogram manipulation with a randomly chosen shifting delta between 0 and  $\frac{40}{dataset\ std}$  and a scaling factor that was restricted to the range of  $\pm 2 * dataset\ std$  and sampled from a lognormal distribution with standard deviation of 0.1 .

For prediction on a complete dataset, we created a tissue mask beforehand, as a convex hull over a set of manually annotated polygon vertices, to avoid predicting noise in non-tissue areas. Voxels outside the mask were automatically classified as background. Contrary to training, prediction was not limited to uint8 data. Thus, we used the original data without uint8 conversion and normalised with dataset-specific mean and standard deviation (C432: mean 53739.84, std

588.81; Y391: mean: 42707.97, std: 709.68; Y489: mean: -30.04, std: 27.07). As the sizes of the datasets demand distributed processing, we conducted the prediction and subsequent segmentation in chunks.

### ***U-Net architecture and training***

Our U-Net classifier (**Fig. 2c**) is based on the 3D U-Net described in [29]. Each residual block consists of two 3x3x3 convolutional layers with ReLU [47] activation. In the downsampling path, the number of features is increased by the second convolution of the residual block. We used 2x2x2 max pooling for the two downsampling steps. Our approach deviates from [29] by computing the respective two upsampling steps via 2x2x2 nearest-neighbour interpolation, by using a 1x1x1 convolution instead of transposed convolution, and by not using batch normalisation. A final 1x1x1 convolution layer per prediction task, namely binary semantic segmentation and distance regression, was added to produce a single-channel output feature map. The resulting map of the semantic segmentation prediction contains logits which are converted by a subsequent sigmoid activation to obtain a map of voxel-wise nucleus probabilities.

To take advantage of the spherical nature of nuclei, distance-based prediction approaches have been developed [48]. Following this idea, we trained the classifier to predict a distance map in addition to the probability prediction. We used a linear activation for the respective 1x1x1 convolution layer. The distance map contains the distance to the closest background voxel for every nucleus voxel and the negative distance to the closest nucleus voxel for every background voxel, clamped at  $\pm 4000\text{nm}$ . Thus, the predicted distance of 0 constitutes the border between nucleus and background. We adapted the training data labels by separating each pair of annotated segments with a thin border of background voxels and subsequently applying Euclidean distance transform.

We used mean squared error loss on the distance map and unreduced binary cross-entropy loss combined with a sigmoid layer on the logits map. As background voxels are far more common than nuclei voxels, we applied class weights, inversely proportional to the class ratio and calculated over all training boxes, for the loss computation. The model's weights were initialised with the Kaiming uniform initialization [49]. We used the Adam optimizer [50] with a learning rate of 0.0001. For the initial classifier on the C432 dataset, we ran training until convergence and therefore set a maximum of 40000 epochs - i.e., iterations through the complete training dataset - with early stopping with a patience of 400 epochs. We found that the classifier converged after 342 epochs. Therefore, a smaller maximum epoch was used for the six classifiers trained on Y489. To enable comparison between those six classifiers, we trained each of them for a similar number of steps, thus adapting the maximum epoch inversely proportional to the amount of training data (0 boxes C432 + 29 boxes Y489: 770 epochs, 30+5: 635 epochs, 30+11: 540 epochs, 30+17: 470 epochs, 30+23: 425 epochs, 30+29: 380 epochs). The training was conducted in a Slurm cluster on a GPU node using one NVIDIA V100 32GB GPGPU with NVLink and an Intel Xeon Gold 6148 CPU, running at 2.4GHz, hyperthreading enabled, up to 32 cores used and 50 GB RAM allocated.

### ***Segmenting the U-Net prediction***

In the previous step, the U-Net predicted a probability map and a distance map. We considered the distance map to be inherently more suitable for thresholding with focus on avoiding mergers. Moreover, we determined the distance prediction quality good enough to not require combining with the predicted probabilities. Therefore, we selected the distance prediction to proceed with the segmentation.

The segmentation step is shown in **Fig. 2c**. We generated seeds by thresholding the prediction at a distance of 705.6 nm, slightly inside the predicted nuclei, to prevent merging nuclei that touch at their borders. Afterwards, we used a watershed algorithm [30] to grow out the segment instances until a distance of 0, i.e., the predicted border between nucleus and background.

After obtaining the instance segmentation in chunks, we merged segments that touch at the chunk borders with an intersection over union score of at least 75% to avoid chunking-induced splits.

## Segmentation post-processing

We computed the segment volumes and center positions chunk-wise with subsequent aggregation. As positions are non-aggregatable, we enlarged each chunk by 20µm at each side to enable determining the positions for segments spanning multiple chunks. Segment volumes were calculated via voxel counting, segment center positions via Euclidean distance transform.

## Mesh calculation

Aside from the meshes for visualisations (see *Rendered visualisations*), we generated meshes within our post-processing procedure with the marching cubes algorithm [51]. The resulting meshes consisted of connected vertices and the enclosed triangular faces (**Fig. 2g**). We applied mesh simplification using an implementation of the quadric mesh reduction algorithm [52] from the Python package pyfqmr to reduce the number of triangles by a factor of 100 while restricting the simplification error to a maximum of 16000 nm. The meshes were computed and stored in chunks, merging was performed when loading a mesh.

## Sphericity

We computed the surface in each mesh by summing up the areas of all the faces of the mesh. Then, for each mesh of a segmented object, we computed its sphericity - a metric of compactness, reporting the ratio of the surface-to-volume ratio of a 3D object to the one of a sphere of the same volume [28]. Accordingly, perfectly spherical objects will have a sphericity of 1, whereas highly eccentric objects will display sphericity values close to 0. Sphericity can be obtained from the particle's volume and surface:

$$Sphericity = \pi^{\frac{1}{3}} * (6 * volume)^{\frac{2}{3}} * \frac{1}{surface}$$

Note that by generating meshes with our mesh calculation process, mesh vertices do not match exactly the borders of voxels. This mismatch leads to slight inaccuracies in the calculated sphericities because we computed the surface from the meshes and the volume as voxel count from the segments. However, those inaccuracies become only relevant for small ( $\leq 3$  voxels-wide) 3D features which can then generate artifactual sphericity values larger than 1. Since nuclei are expected to present overall round structures and their diameter (10 µm) is typically  $>100$  times larger than the voxel size (100 nm), their characteristic sphericities can be reliably retrieved (**Fig. 2h**).

To determine the volume and sphericity thresholds for C432 and Y391, we conducted 2-fold cross-validation over a set of 12 subvolumes spanning the different histological regions. We obtained the thresholds via grid search over the volumes and sphericities of the detected objects and selected the combination that resulted in the best F1 score. We valued higher recall over higher precision. Thus, we restricted the recall computed over all training boxes combined to be at least 0.8. Further, we performed aggregation over the folds by selecting the minimum for each threshold. For Y489, we applied the thresholds computed on C432.

## Evaluation

As described in (*Sphericity*), we calculated the thresholds for C432 and Y391 on 12 boxes. We further kept a held-out test set of 6 additional boxes. On Y489, we used 14 boxes for classifier selection and 6 boxes as test set. All boxes covered  $(50 \mu\text{m})^3$  and contained skeleton annotations of nuclei instances in the MCL, GL and EPL layer - three boxes in the centre and three boxes in the border region of each layer. For C432 and Y391, the boxes were evenly distributed among the cross-validation folds and the test set, each containing one box per region and layer. The test set for Y489 followed the same principle, the classifier selection set contained two boxes per region and layer with addition of one box from a darker region of each the MCL and the GL layer.

When calculating precision, recall, and F1 score per box, we aimed to minimise measurement artefacts due to nuclei being cut off at the box border. Therefore, we enlarged each box by 8µm in each direction beforehand, which was treated as a

tolerance margin that contained no ground truth annotations but predicted nuclei in the margin were not counted as false positives. The average precision and recall values in **Table 1** reflect the average score across subvolumes (**Fig. 3f,l,r**).

### **Rendered visualisations**

3D renders of datasets and nuclei were generated using SideFx Houdini 19.5.303 and Maxon Redshift 3.5.8. Datasets were exported at the relevant magnifications from WEBKNOSSOS using the webknossos Python API and imported into Houdini as image stacks, and geometry was masked using an exterior mesh generated in WEBKNOSSOS. Volume annotations of the segmented nuclei were read from the same location, including the unique identifier assigned to each segmented object. Segmented nuclei were converted to polygonal mesh using [53]. For each dataset, meshed nuclei with a volume larger than 10x the mean nuclear volume were considered data conversion artefacts and masked out.

## **ACKNOWLEDGEMENTS**

The authors are grateful to the scientific computing, biological research and electron microscopy science technology platforms of the Francis Crick Institute. We thank Jeroen Claus for advice and support in data rendering and critical reading of the manuscript; to Albane le Tournoux de la Villegeorges, Valentin Pinkau, Daniel Werner, Georg Wiese, Jonathan Striebel, Philipp Otto, Florian Meinel and Tom Herold for their work on Voxelytics and WEBKNOSSOS as well as discussions and guidance on training and evaluating the classifiers, to the team at Mindy Support for image annotation and to Joerg Lindenau and Carsten Waltenberg at ZEISS for support in targeted femtosecond laser milling.

This research was funded in whole, or in part, by the Wellcome Trust (FC001153 and 110174/Z/15/Z to A.T.S.). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. We acknowledge the Paul Scherrer Institut, Villigen, Switzerland, for provision of synchrotron radiation beamtime at the TOMCAT beamline (proposals 20190417, 20211896 and 20220191), the DIAMOND synchrotron, Didcot, UK, for beamtime at the I13-2 beamline (proposal MT20274) and the European synchrotron, Grenoble, France, for provision of synchrotron radiation beamtime at the beamline ID16A (proposals LS2918, LS3025 and LS3186).

This work was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001153 to A.T.S.), the UK Medical Research Council (FC001153 to A.T.S.), and the Wellcome Trust (FC001153 to A.T.S.). It was also supported by a Physics of Life grant (EP/W024292/1) to A.T.S. and A.P. funded by EPSRC and Wellcome and a UKRI MRC Business Engagement Fund (P2023-0024) to C.B. and N.R. A.P. acknowledges funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme (852455).

### **Data and Code Availability**

Training data used to generate all classifiers and code to replicate the main plots reported are accessible at <https://github.com/cboschp/segVolData>

### **Conflicts of interest**

N.R., A.N. and M.C. are employees at scalable minds. N.R. holds equity in scalable minds. scalable minds develops and provides image analysis software and services. E.M. is an employee at Phospho Biomedical Animation.

## REFERENCES

- [1] S. R. y. Cajal, [Texture of the Nervous System of Man and the Vertebrates] Springer Vienna, (1999).
- [2] E. G. Gray, "Electron microscopy of synaptic contacts on dendrite spines of the cerebral cortex," *Nature*, 183(4675), 1592-3 (1959).
- [3] W. Denk, and H. Horstmann, "Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure," *PLoS Biol*, 2(11), e329 (2004).
- [4] G. Knott, H. Marchman, D. Wall *et al.*, "Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling," *J Neurosci*, 28(12), 2959-64 (2008).
- [5] J. S. Phelps, D. G. C. Hildebrand, B. J. Graham *et al.*, "Reconstruction of motor control circuits in adult *Drosophila* using automated transmission electron microscopy," *Cell*, 184(3), 759-774 e18 (2021).
- [6] C. S. Xu, K. J. Hayworth, Z. Lu *et al.*, "Enhanced FIB-SEM systems for large-volume 3D imaging," *Elife*, 6, (2017).
- [7] C. S. Xu, S. Pang, G. Shtengel *et al.*, "An open-access volume electron microscopy atlas of whole cells and tissues," *Nature*, 599(7883), 147-151 (2021).
- [8] S. Looma, J. Straehle, V. Gangadharan *et al.*, "Connectomic comparison of mouse and human cortex," *Science*, 377(6602), eabo0924 (2022).
- [9] D. Witvliet, B. Mulcahy, J. K. Mitchell *et al.*, "Connectomes across development reveal principles of brain maturation," *Nature*, 596(7871), 257-261 (2021).
- [10] M. R. Tavakoli, J. Lyudchik, M. Januszewski *et al.*, "Light-microscopy based dense connectomic reconstruction of mammalian brain tissue," *bioRxiv*, 2024.03.01.582884 (2024).
- [11] T. W. Shin, H. Wang, C. Zhang *et al.*, "Dense, Continuous Membrane Labeling and Expansion Microscopy Visualization of Ultrastructure in Tissues," *bioRxiv*, 2024.03.07.583776 (2024).
- [12] A. T. Kuan, J. S. Phelps, L. A. Thomas *et al.*, "Dense neuronal reconstruction through X-ray holographic nanotomography," *Nat Neurosci*, 23(12), 1637-1643 (2020).
- [13] C. L. Walsh, P. Tafforeau, W. L. Wagner *et al.*, "Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography," *Nat Methods*, 18(12), 1532-1541 (2021).
- [14] P. Cloetens, and J. Baruchel, "Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation x rays," *Applied Physics Letters*, 75(2912), (1999).
- [15] A. Azevedo, E. Lesser, J. S. Phelps *et al.*, "Connectomic reconstruction of a female *Drosophila* ventral nerve cord," *Nature*, 631(8020), 360-368 (2024).
- [16] C. Bosch, T. Ackels, A. Pacureanu *et al.*, "Functional and multiscale 3D structural investigation of brain tissue through correlative in vivo physiology, synchrotron microtomography and volume electron microscopy," *Nat Commun*, 13(1), 2923 (2022).
- [17] M. Januszewski, J. Maitin-Shepard, P. Li *et al.*, "Flood-Filling Networks," *arXiv*, (2016).
- [18] M. Schmidt, A. Motta, M. Sievers *et al.*, "RoboEM: automated 3D flight tracing for synaptic-resolution connectomics," *Nat Methods*, 21(5), 908-913 (2024).
- [19] A. Motta, M. Berning, K. M. Boergens *et al.*, "Dense connectomic reconstruction in layer 4 of the somatosensory cortex," *Science*, 366(6469), (2019).
- [20] MICrONS-Consortium, J. A. Bae, M. Baptiste *et al.*, "Functional connectomics spanning multiple areas of mouse visual cortex," *bioRxiv*, 2021.07.28.454025 (2021).
- [21] M. Berning, K. M. Boergens, and M. Helmstaedter, "SegEM: Efficient Image Analysis for High-Resolution Connectomics," *Neuron*, 87(6), 1193-206 (2015).
- [22] S. Herculano-Houzel, B. Mota, and R. Lent, "Cellular scaling rules for rodent brains," *Proc Natl Acad Sci U S A*, 103(32), 12138-43 (2006).
- [23] A. Aswath, A. Alsahaf, B. N. G. Giepmans *et al.*, "Segmentation in large-scale cellular electron microscopy with deep learning: A literature survey," *Med Image Anal*, 89, 102920 (2023).
- [24] R. Conrad, and K. Narayan, "Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model trained on a diverse dataset," *Cell Syst*, 14(1), 58-71 e5 (2023).
- [25] H. Spiers, H. Songhurst, L. Nightingale *et al.*, "Deep learning for automatic segmentation of the nuclear envelope in electron microscopy data, trained with volunteer segmentations," *Traffic*, 22(7), 240-253 (2021).
- [26] Y. Zhang, T. Ackels, A. Pacureanu *et al.*, "Sample Preparation and Warping Accuracy for Correlative Multimodal Imaging in the Mouse Olfactory Bulb Using 2-Photon, Synchrotron X-Ray and Volume Electron Microscopy," *Front Cell Dev Biol*, 10, 880696 (2022).

- [27] A. Miettinen, I. V. Oikonomidis, A. Bonnin *et al.*, “NRStitcher: non-rigid stitching of terapixel-scale volumetric images,” *Bioinformatics*, 35(24), 5290-5297 (2019).
- [28] H. Wadell, “Volume, Shape, and Roundness of Quartz Particles,” *Journal of Geology*, 43, 250-280 (1935).
- [29] O. Çiçek, A. Abdulkadir, S. S. Lienkamp *et al.*, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” arXiv, abs/1606.06650, (2016).
- [30] P. J. Soille, and M. M. Ansout, “Automated basin delineation from digital elevation models using mathematical morphology,” *Signal Processing*, 20(2), 171-182 (1990).
- [31] D. P. Hoffman, G. Shtengel, C. S. Xu *et al.*, “Correlative three-dimensional super-resolution and block-face electron microscopy of whole vitreously frozen cells,” *Science*, 367(6475), (2020).
- [32] A. V. Weigel, C. L. Chang, G. Shtengel *et al.*, “ER-to-Golgi protein delivery through an interwoven, tubular network extending from ER,” *Cell*, 184(9), 2412-2429 e16 (2021).
- [33] L. K. Scheffer, C. S. Xu, M. Januszewski *et al.*, “A connectome and analysis of the adult *Drosophila* central brain,” *Elife*, 9, (2020).
- [34] M. Helmstaedter, “Cellular-resolution connectomics: challenges of dense neural circuit reconstruction,” *Nat Methods*, 10(6), 501-7 (2013).
- [35] L. F. Abbott, D. D. Bock, E. M. Callaway *et al.*, “The Mind of a Mouse,” *Cell*, 182(6), 1372-1376 (2020).
- [36] C. Bosch, G. S. Jefferis, L. Collinson *et al.*, [Scaling up connectomics: the road to a whole mouse brain connectome] Wellcome Trust, (2023).
- [37] J. Moore, D. Basurto-Lozada, S. Besson *et al.*, “OME-Zarr: a cloud-optimized bioimaging file format with international community support,” *Histochem Cell Biol*, (2023).
- [38] N. Rzepka, J. A. Bogovic, and J. A. Moore, [Chapter 14 - Toward scalable reuse of vEM data: OME-Zarr to the rescue] Academic Press, (2023).
- [39] K. M. Boergens, M. Berning, T. Bocklisch *et al.*, “webKnossos: efficient online 3D data annotation for connectomics,” *Nat Methods*, 14(7), 691-694 (2017).
- [40] Y. Hua, P. Laserstein, and M. Helmstaedter, “Large-volume en-bloc staining for electron microscopy-based connectomics,” *Nat Commun*, 6, 7923 (2015).
- [41] M. Pallotto, P. V. Watkins, B. Fubara *et al.*, “Extracellular space preservation aids the connectomic analysis of neural circuits,” *Elife*, 4, (2015).
- [42] C. Bosch, J. Lindenau, A. Pacureanu *et al.*, “Femtosecond laser preparation of resin embedded samples for correlative microscopy workflows in life sciences,” *Applied Physics Letters*, 122(14), 143701 (2023).
- [43] J. Lehtinen, J. Munkberg, S. Laine *et al.*, “Noise2Noise: Learning Image Restoration without Clean Data,” arXiv, (2018).
- [44] D. Paganin, S. C. Mayo, T. E. Gureyev *et al.*, “Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object,” *J Microsc*, 206(Pt 1), 33-40 (2002).
- [45] B. Yu, L. Weber, A. Pacureanu *et al.*, “Evaluation of phase retrieval approaches in magnified X-ray phase nano computerized tomography applied to bone tissue,” *Opt Express*, 26(9), 11110-11124 (2018).
- [46] A. Mirone, E. Brun, E. Gouillart *et al.*, “The PyHST2 hybrid distributed code for high speed tomographic reconstruction with iterative reconstruction and a priori knowledge capabilities,” *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 324, 41-48 (2014).
- [47] V. Nair, and G. E. Hinton, [Rectified linear units improve restricted boltzmann machines] Omnipress, Haifa, Israel(2010).
- [48] P. Naylor, M. Laé, F. Reyal *et al.*, “Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map,” *IEEE Transactions on Medical Imaging*, 38(2), 448-459 (2019).
- [49] K. He, X. Zhang, S. Ren *et al.*, [Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification] IEEE Computer Society, (2015).
- [50] D. P. Kingma, and J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv, (2017).
- [51] W. E. Lorensen, and H. E. Cline, [Marching cubes: A high resolution 3D surface construction algorithm] Association for Computing Machinery, (1987).
- [52] M. Garland, and P. S. Heckbert, [Surface simplification using quadric error metrics] ACM Press/Addison-Wesley Publishing Co., (1997).
- [53] K. Museth, “VDB: High-resolution sparse volumes with dynamic topology,” *ACM Trans. Graph.*, 32(3), Article 27 (2013).