Convolutional neural networks for cytoarchitectonic brain mapping at large scale

Christian Schiffer [a,b,*], Hannah Spitzer [c], Kai Kiwitz [d], Nina Unger [d], Konrad Wagstyl [e], Alan C. Evans [f], Stefan Harmeling [g], Katrin Amunts [a,d], Timo Dickscheid [a,b]

[a] *Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Germany*
[b] *Helmholtz AI, Research Centre Jülich, Germany*
[c] *Institute of Computational Biology, Helmholtz Zentrum München, Germany*
[d] *Cécile & Oscar Vogt Institute for Brain Research, University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, Germany*
[e] *Wellcome Centre for Human Neuroimaging, University College London, London, United Kingdom*
[f] *Department of Neurology & Neurosurgery, Montréal Neurological Institute (MNI), McGill University, Montréal, Canada*
[g] *Institute of Computer Science, Heinrich-Heine-University Düsseldorf, Germany*

ARTICLE INFO

ABSTRACT

Human brain atlases provide spatial reference systems for data characterizing brain organization at different levels, coming from different brains. Cytoarchitecture is a basic principle of the microstructural organization of the brain, as regional differences in the arrangement and composition of neuronal cells are indicators of changes in connectivity and function. Automated scanning procedures and observer-independent methods are prerequisites to reliably identify cytoarchitectonic areas, and to achieve reproducible models of brain segregation. Time becomes a key factor when moving from the analysis of single regions of interest towards high-throughput scanning of large series of whole-brain sections. Here we present a new workflow for mapping cytoarchitectonic areas in large series of cell-body stained histological sections of human postmortem brains. It is based on a Deep Convolutional Neural Network (CNN), which is trained on a pair of section images with annotations, with a large number of un-annotated sections in between. The model learns to create all missing annotations in between with high accuracy, and faster than our previous workflow based on observer-independent mapping. The new workflow does not require preceding 3D-reconstruction of sections, and is robust against histological artefacts. It processes large data sets with sizes in the order of multiple Terabytes efficiently. The workflow was integrated into a web interface, to allow access without expertise in deep learning and batch computing. Applying deep neural networks for cytoarchitectonic mapping opens new perspectives to enable high-resolution models of brain areas, introducing CNNs to identify borders of brain areas.

1. Introduction

Human brain atlases provide a spatial framework for localizing information retrieved from neuroscientific studies of different brains, addressing brain organization from different angles and including different data modalities. The cerebral cortex of the brain is organized into cortical areas, which each have a specific functional role. They can be identified in cell body stained sections based on cytoarchitecture. Regional differences in the spatial arrangement and composition of the cells co-vary with changes in connectivity and function Goulas et al. (2018). Cytoarchitectonic borders can be identified in microscopic scans of histological brain sections, based on the analysis of the arrangement and distribution of cells, their different morphology and size, as well as differences in the appearance and relative thickness of cortical lay-

ers. Such criteria have been formulated for the first time more than a century ago to map the cerebral cortex, and still serve as guidelines for cytoarchitectonic analysis Amunts and Zilles (2015). Different approaches have been proposed in the past to identify positions of borders in a reliable manner Annese et al. (2004); Schleicher et al. (1999); Schmitt and Böhme (2002). The de-facto standard for identifying borders of cytoarchitectonic areas in the human cerebral cortex is a method based on multivariate statistical image analysis Schleicher et al. (1999), which has been applied for the identification of more than 200 areas to date Amunts et al. (2020). To map the whole extent of an area in both hemispheres, and to capture its intersubject variability through studies in large samples, however, is extremely time- and labor-intensive: Cytoarchitectonic maps need to aggregate properties across many histological sections and multiple brains. To address this challenge, mapping

includes a subset of histological sections (every 15–60ths section, i.e. 0.3 mm to 1.2 mm distance between sections) of ten human postmortem brains resulting in analyses of several hundred sections per area, which corresponds to a workload in the order of one or even several person years per area Amunts et al. (2020).

Recent high-throughput scanning devices and powerful compute resources enable a much higher degree of automation in digitalization and analysis of whole human brain sections at microscopical resolution. Technological progress has made it possible to 3D-reconstruct a complete postmortem brain at 20 micron spatial resolution with more than 7000 sections - the BigBrain Amunts et al. (2013). This high-resolution brain model opens the possibility to produce complete maps of cytoarchitectonic areas at full microscopic resolution, and to cover large image stacks with brain areas extending across thousands of sections. Hereby, each section image has up to 120,000 × 80,000 pixels image size each. In order to address these challenges, a method is required, which

1. automatically classifies brain areas based on cytoarchitectonic criteria,
2. handles series with thousands of 2D images of histological sections with data in the Giga- to Terabyte range,
3. is robust against histological artefacts, which are inevitable in large section series,
4. provides stable results independently of the cutting plane, e.g. when changes in the cutting direction relative to the brain tissue prevents analysis of the 6-layered structure of the cerebral cortex (in the following referred to as *oblique cuts*), and
5. can be operated and supervised by neuroscience experts without requiring advanced computer science skills.

Previous experience in cytoarchitectonic mapping has shown that the identification of brain areas considers multiple parameters. This is true for traditional visual inspection using a light microscope, as well as for automated mapping approaches. It involves complex multi-scale texture patterns, from the level of neurons up to a level of cortical layers and areas. However, several parameters that can be used for identification of cortical areas heavily depend on the cutting plane of the histological sections with respect to the orientation of cortical columns. The highly folded cerebral cortex of the human brain hereby poses particular challenges, since brain areas may appear in a very different way in dependence on the cutting angle. Thus, brain mapping needs to operate in a variable data space, where no restrictions should be made on the orientation of the cutting plane relative to the course of cortical layers and the brain surface. In addition, automated brain mapping needs to consider variation in tissue quality and staining, as well as histological artefacts. Finally, automated mapping methods must take into account variations in cytoarchitecture between different brains and lead to identical parcellations, even if interindividual differences in cytoarchitecture are large.

Previous work on automated cytoarchitectonic area segmentation (Spitzer et al., 2017; 2018) proposes to use Convolutional Neural Networkss (CNNss) for automatic segmentation of multiple cytoarchitectonic areas across multiple human brains. This is a remarkably challenging task, as the model needs to be robust against the considerable interindividual variability of the human brain, inevitable histological artefacts, variations in staining, and oblique cuts, to name only a few of the constraints. At the same time, it has to be highly sensitive to variations of cytoarchitecture in different brain areas, which may be subtle. This may result in a need for large amounts of training data, which is difficult to cover. Consequently, such generalized segmentation models are still subject to active research.

We here propose a new workflow for cytoarchitectonic mapping of a *target area* across large or complete series of histological human brain sections with a high degree of automation. The workflow is illustrated in Fig. 1. Following a "divide & conquer" approach, the full extent of a target brain area *a* is subdivided into intervals of sections, which are enclosed by annotations created at approximately regular section intervals.

Separate CNNss are then trained for each interval, using the enclosing annotations as training data. This results in a set of *local segmentation models*, each specialized to automatically map only the tissue sections which fall into the corresponding interval. By training local models for each interval of target area *a*, an interactive workflow is obtained that allows an expert to label cytoarchitectonic areas in full stacks of histological sections with minimal manual annotation, aided by Deep Learning, and at a speed that matches high throughput image acquisition.

In this work, we

1. introduce a method to automatically map cytoarchitectonic brain areas across large series of histological human brain sections (Section 2),
2. evaluate its precision on 18 cytoarchitectonic areas from the BigBrain dataset Amunts et al. (2013) to investigate its applicability to a wide range of different brain areas,
3. assess its precision for two areas in three brains with variable staining protocols Amunts et al. (2013, 2000); Ding et al. (2016) to investigate robustness against interindividual differences and different staining procedures, and
4. create highly detailed and complete 3D maps of four areas in the BigBrain dataset[1] and evaluate their anatomical plausibility.[2]

2. Materials and methods

2.1. GLI-based mapping of cytoarchitectonic areas for training and validation

Our proposed method requires annotations of the target area at roughly regular intervals in approximately 1% of sections in the stack. Such annotations consist of localizations of areal borders in the section, and are defined using the well-established GLI-based mapping procedure described in Schleicher et al. (1999). This approach starts by scanning the histological images and by building a Gray Level Index (GLI) image Schleicher et al. (1999). The GLI is a measure of the volume fraction of cell bodies Wree et al. (1982). In a next step, profiles extending from the cortical surface to the white matter border are extracted along Laplacians, which reflect laminar changes in the volume fraction of cell bodies, and thus encode cytoarchitecture. These Laplacians reflect an important feature of cortical cytoarchitecture, i.e. its columnar structure Schleicher et al. (2000). The cortical surface and the white matter border are manually identified. Using a sliding window procedure across the cortical ribbon, the similarity of blocks of profiles is being estimated by the Mahalanobis distance, a multi-variate distance measure, at each position, that is combined with a Hotelling's *t*-test for checking significance. Borders between areas are indicated by significant peaks in the Mahalanobis distance function. The positions of borders are then labeled in the image. These borders are then used as a basis for the network training and validation.

2.2. Datasets

The datasets used in this study comprise image series of histological sections of three human brains, which have been stained for neuronal cell bodies Amunts et al. (2020); Ding et al. (2016). The brains vary in terms of cytoarchitecture and folding pattern, as well as staining properties, presence of histological artifacts and other features (Fig. 2). Areas have been mapped in the past (cf. Section 2.1) using at least every 60th section of the series. These maps provide the basis to train the neural network models and to perform automatic segmentation in previously unseen, close by sections.

---

[1] https://www.bigbrainproject.org
[2] The maps are released in the public domain as part of the multilevel human brain atlas in the EBRAINS platform https://www.ebrains.eu
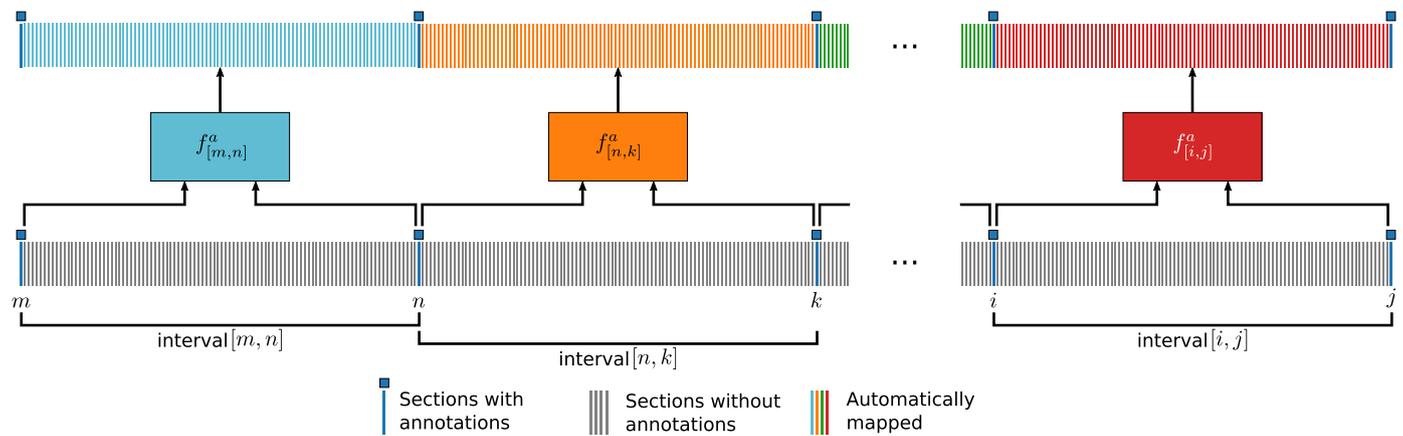
Fig. 1. Setup of our workflow. Images of histological sections are depicted as thin vertical bars, neural network models are depicted as colored boxes. The full extent of sections containing a target brain area a (sections m to j, bottom row) is subdivided into section intervals, which are defined by annotations at regular intervals (blue squares, $m, n, k, \ldots$). One local segmentation model $f_{[m,n]}^a$ is trained for each interval enclosed by a pair of annotations $[m, n]$. After training, each model is applied to automatically map sections falling into the corresponding interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
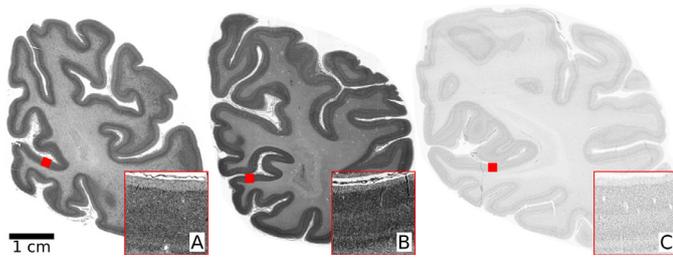

Fig. 2. Example images of cell body stained histological human brain sections taken from datasets B20(A), B01(B) and AAHB (C). All sections were sampled from a comparable region of the occipital lobe. Differences arise from inter-subject variability and variations in staining and histological processing protocols. Locations of detail views ($2\,\text{mm} \times 2\,\text{mm}$) are marked with red squares. For B20 and B01, only the right hemisphere is shown. AAHB only includes a single hemisphere. Cerebellum was removed from B20 and AAHB for visualization. Scale bar: 1cm (same for all three sections). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The first dataset - denoted as B20- is based on the original histological sections of the publicly available microscopic 3D model BigBrain (Amunts et al., 2013). The dataset consists of images of 7404 coronal sections with a thickness of 20 µm. A modified Merker stain Merker (1983) was used to stain cell bodies. A subset of sections was scanned at 1 µm resolution using a high-throughput light-microscopic scanner (TissueScope HS, Huron Digital Pathology Inc.). Annotations based on the GLI-based method (Section 2.1) at an interval of approximately 60 sections ($\approx 1.2\,\text{mm}$) were obtained for 18 cortical areas, belonging to different functional systems:

1. Visual areas $hOc1$, $hOc2$ (Amunts et al., 2000), $hOc3v$ (Rottschy et al., 2007) and $hOc5$ (Malikovic et al., 2007). Additional annotations at an interval of approximately 30 (0.6 mm) sections were created for $hOc5$, as well as on a small set of sections containing $hOc3v$ (Kiwitz et al., 2019a; 2019b; 2020a; 2020b).
2. Areas of the frontal operculum $Op5$, $Op6$ and $Op7$ Unger et al. (2020a, 2020b, 2020c).
3. Areas 44 and 45 of Broca's region (Amunts et al., 1999; 2004) in the inferior frontal gyrus.
4. Areas $hIP5, hIP6, hIP7$ and $hIP8$ (Richter et al., 2019) in the intraparietal sulcus.

5. Supplementary motor area SMA and pre-supplementary motor area $preSMA$ Ruan et al. (2018).
6. Premotor areas $6d1$, $6d2$ and $6d3$ Sigl (2018); Sigl et al. (2019a, 2019b, 2019c).

The BigBrain dataset has been fully reconstructed at 20 µm Amunts et al. (2013) and therefore opens the possibility to investigate the 3D consistency of the computed maps after transformation into the reconstructed space.

Brain areas differ in cytoarchitecture, as well as in size and in how much the morphology of an area changes across a series of consecutive brain sections. This has implications for the amount of annotations required to capture the relevant properties of certain areas. For example, $hOc1$ is large and shows only moderate changes across consecutive sections. In comparison, $hOc5$ is considerably smaller, and $hOc3v$ changes considerably across consecutive sections (see Fig. 10, C-F), resulting in a need for more annotations to capture their structure.

The second dataset - B01- has also been used for mapping in the past, whereby every 15th section of the whole series of sections was stained and digitized. This brain was 3D reconstructed with a spatial resolution of 1 mm isotropic Amunts et al. (2020). Annotations for visual areas $hOc1$ and $hOc2$ at an interval of approximately every 60th section (Amunts et al., 2000) in a subset of sections have been used. This dataset serves to investigate robustness against intersubject variability, while the lab protocol is similar to the one used for B20.

The third dataset - AAHB-, comes from the *Allen Adult Human Brain Atlas* Ding et al. (2016). It includes 106 unevenly spaced, publicly available sections. In contrast to the first two series of images, it differs in thickness (50 µm), and the staining method (Nissl staining). Annotations are provided for cortical and subcortical gray matter according to a modified Brodmann scheme on one hemisphere (cf. (Ding et al., 2016)). This dataset is used to investigate robustness of the proposed method against variable lab protocols and delineation criteria with respect to areas $hOc1$ and $hOc2$, which correspond to *"primary visual cortex (striate cortex, area V1/17)"* (identifier 10269) and *"parastriate cortex (area V2, area 18)"* (identifier 10271), respectively, in the Allen ontology.

2.3. Local segmentation models

Annotations of cytoarchitectonic areas based on GLI mapping (Section 2.1) were used to train CNNss, which we refer to as *local segmentation models*. Each local segmentation model $f_{[s_1, s_2]}^a$ was trained on two sections s_1 and s_2 (the *training sections*) with available annotations for a target area a. Trained local segmentation models were then applied
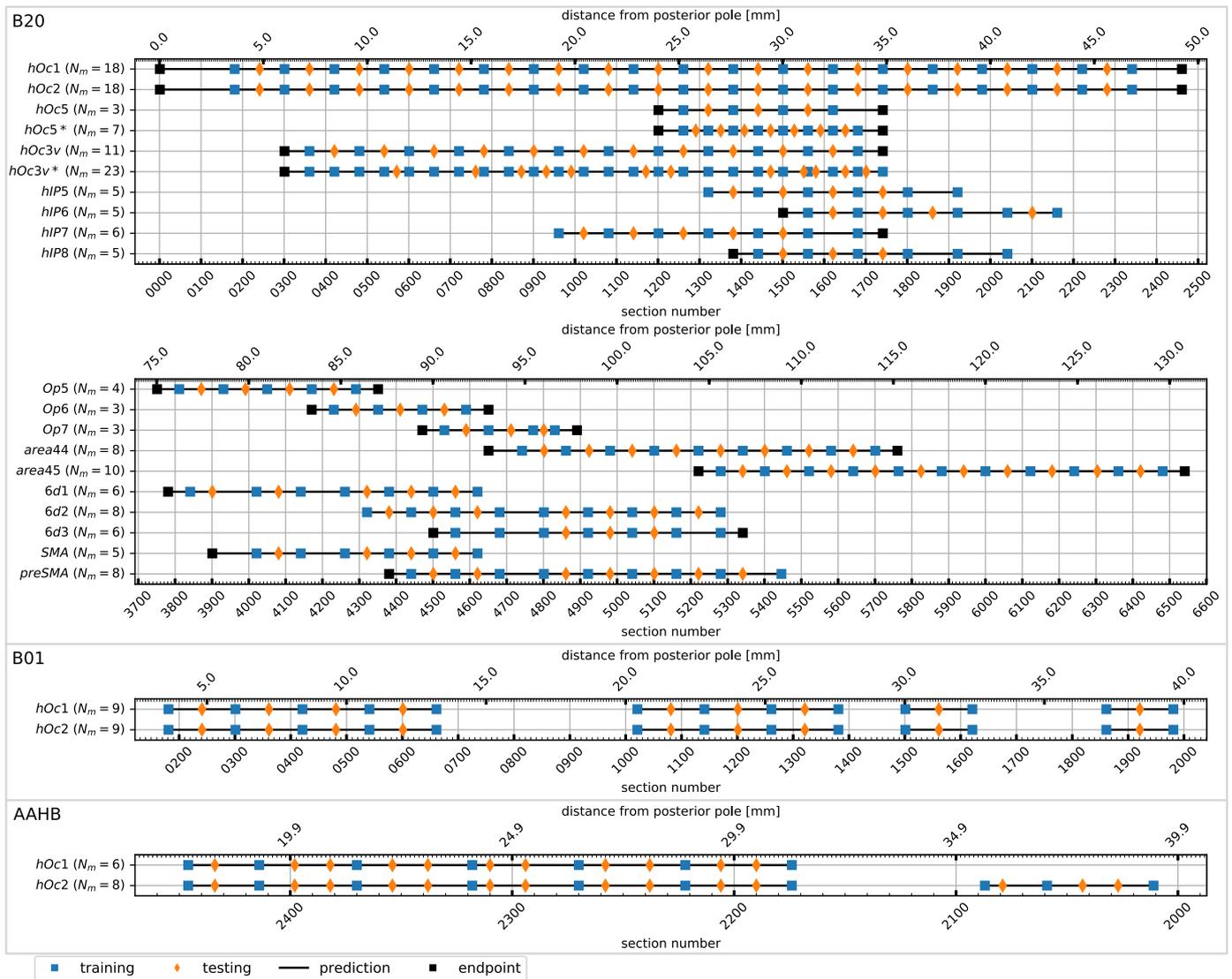
Fig. 3. Training and test sections from available annotations across stacks of histological sections. Consecutive pairs of training sections (blue squares) induce one local segmentation model. For example, model $f_{[181,301]}^{B20-hOc1}$ was trained on sections 181 and 301 of dataset B20, segments area $hOc1$ in the full interval $[182, 300]$, and was tested on Section 2.1. N_m denotes the number of trained local segmentation models. * marks experiments performed with a smaller training interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to "fill the gaps", i.e. to automatically segment the target area in sections enclosed by the respective training sections s_1 and s_2 (Fig. 1). The focus on a single target area and a spatially restricted stack of consecutive sections reduces cytoarchitectonic and morphological variations that need to be captured by the respective models, which we expect to result in improved performance compared to training models for multiple areas or a wider range of sections as proposed in Spitzer et al. (2017).

We trained local segmentation models for 18 cytoarchitectonic areas in B20 and two areas in each of B01 and AAHB. Fig. 3 gives an overview of sections used for the individual areas. Most local segmentation models were trained on two training sections with annotations at $\sim 2.4\,\mathrm{mm}$ distance, corresponding to ~ 120 sections for B20 and B01 and 48 sections for AAHB. Additional local segmentation models with a reduced interval size of 60 sections ($1.2\,\mathrm{mm}$) were trained for areas $hOc3v$ and $hOc5$ to account for highly variable morphology ($hOc3v$, see Fig. 10, C-F) and small area size ($hOc5$). For B01 and AAHB, local segmentation models were trained only for ranges of sections where annotations were available at the required interval. Segmentations of the outer most parts of cytoarchitectonic areas which were not enclosed by training sections (i.e. Sections 1 to 181 for $hOc1$ in B20) were processed using the closest

available local segmentation model. For example, model $f_{[181,301]}^{B20-hOc1}$ was also applied to the section interval $[1, 181]$.

2.4. Neural network architecture

For local segmentation models, the modified U-Net architecture (Ronneberger et al., 2015) proposed by Spitzer et al. (2017) was extended into a multi-scale neural network model (Fig. 5, C). U-Nets have proven to be very powerful for many applications in biomedical image segmentation (e.g. (Çiçek et al., 2016; Milletari et al., 2016)). They consist of an encoder and decoder branch, which are linked by skip-connections between layers of corresponding spatial resolution to allow recovery of fine-grained details during upsampling. Compared to the U-Net (Ronneberger et al., 2015), the modified architecture (Spitzer et al., 2017) employs additional encoder layers and a different number of filters to make processing of large image patches computationally tractable. To show the benefit of using a multi-scale variant of U-Nets, three network variants were used: A high-resolution encoder network (HR), a low-resolution network (LR), and a combined multi-scale architecture (MS).
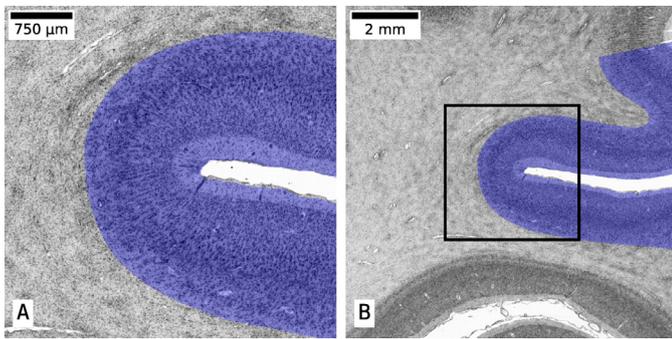
Fig. 4. Typical input patches for the proposed MSarchitecture. A high-resolution image patch (A, 2 μm per pixel) resolves fine-grained microstructural texture, while a lower resolution image patch (B, 16 μm per pixel) provides more information on macroanatomical context. The black rectangle indicates the position of patch (A) inside patch (B). Expert annotations of area $hOc2$ are overlayed in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

High-resolution encoder architecture (HR)

The architecture proposed in Spitzer et al. (2017) was used as base architecture (Fig. 5, A). A *high-resolution encoder* E_{HR} receives high-resolution input patches with a size of 2025×2025 pixels at 2 μm pixel resolution (4.05×4.05 mm^2, Fig. 4, A) and enables recognition of fine-grained microstructural textures. It consists of six convolutional blocks, with the number of filters set to $\{16, 32, 64, 64, 128, 128\}$ respectively. All but the last block are followed by a max-pooling operation with pool size 2 and stride 2. The first layer of the first block in E_{HR} uses a filter size of 5 and a stride of 4, which increases the receptive field while keeping memory consumption and computational effort tangible. All remaining convolutional layers of E_{HR} use a kernel size of 3 and stride 1. Following (Spitzer et al., 2018), we initialize E_{HR} from a network that has been pre-trained on a self-supervised task, specifically on predicting the geodesic distance along the brain surface between image patches from the BigBrain dataset. This auxiliary task has been shown to promote extraction of distinctive cytoarchitectonic features. The decoder consists of four convolutional blocks with the number of filters set to $\{128, 64, 64, 32\}$ respectively. Each block is preceded by an upsampling block, which consists of a nearest neighbor upsampling with kernel size 2 and stride 2, followed by a zero-padded convolutional layer with kernel size 2 and stride 1. All convolutional operations in the network are followed by batch normalization Ioffe and Szegedy (2015) and Rectified Linear Unit (ReLU) non-linearity.

Multi-scale network architecture (MS)

The multi-scale network architecture was obtained by attaching a *low-resolution encoder* E_{LR} as a second branch to HR, which receives lower resolution image patches with a size of 682×682 pixels at 16 μm pixel resolution (10.912×10.912 mm^2), centered at the same location as E_{HR} patches (Fig. 4). This branch allows to learn features at the scale of local cortical folding patterns. Although such macroscopic features are not generally representative of cytoarchitecture in human brains, as they vary largely between individuals (Amunts and Zilles, 2015), they are appropriate in the present setting due to the locality of the network models. E_{LR} is based on E_{HR}, and composed of six convolutional blocks with the same number of filters as E_{HR}. All convolutional filters use a filter size of 3 and a stride of 1. Convolutional layers in the first block use a dilation rate of 1, while all other convolutional layers within E_{LR} use a dilation rate of 2 to enlarge the receptive field.

Low-resolution encoder architecture (LR)

The third architecture is based on HR, but replaces the encoder E_{HR} with E_{LR} (Fig. 5, B). By design, this model can only recognize macroscopic tissue features, and no detailed cytoarchitectonic properties at the level of cell bodies.

2.5. Training strategy

Stochastic gradient descent with Nesterov momentum (Sutskever et al., 2013) was used as optimizer for training the neural network models. Training was performed for 3000 iterations. The learning rate was initially set to 0.01 and decreased by a factor of 0.5 after 1000, 1400, 1800, 2200 and 2600 iterations. Momentum was set to 0.9. Categorical cross-entropy with a weight decay of 0.0001 was used as loss function.

Background class labels

Spitzer et al. (2017) reported convergence problems when training models with a single background class that includes both white and gray matter components, resulting in a mix of tissue parts with very high and very low similarity to the target area under the same classification label. Thus, the general background class was split into separate labels for gray matter (cor) and white matter (wm), resulting in a semantic segmentation problem with the four classes bg, wm, cor, and the target area a. For splitting the background class into wm and cor, different strategies were used for each dataset:

1. For B20, a volumetric tissue classification presented in Lewis et al. (2014) was projected onto the 2D histological sections using transformations provided by the authors of Amunts et al. (2013).
2. For B01, the gray white matter segmentation described in Spitzer et al. (2017) was used.
3. For AAHB, the respective delineations available from the Allen ontology Ding et al. (2016) were used.

Patchwise training

The full resolution scans of the whole-brain sections are by far too large to be used for training. Thus, a patchwise training procedure as also proposed in Ronneberger et al. (2015), Spitzer et al. (2017, 2018) was employed. However, due to the locality of local segmentation models, patches were sampled only in the direct proximity of the target brain area a, to effectively teach the models to distinguish a from its immediate surroundings. Only pixels with a distance of 5 mm or less to any pixel annotated as a were considered as potential center points for training patches.

Data augmentation

The following data augmentations were employed to simulate most frequently observed variations in the data: Both at test and training time, images were rotated by multiples of 90 degrees so that the y axis of coronal sections matches approximately the cranial direction. Random rotation by an angle sampled from a uniform distribution with support $[-45, 45]$ were applied to account for small differences in rotation angle. Intensity variations were addressed by random pixel intensity augmentation with the function $f(x) = \alpha x^{\gamma} + \beta$. The same intensity transformation is applied to all pixels of a training patch. Parameters were chosen from uniform distributions with $\alpha \sim U[0.9, 1.1]$, $\beta \sim U[-0.2, +0.2]$ and $\gamma \sim U[0.8, 1.214]$. The range of each parameter was empirically chosen to reflect natural variations occurring in the data.

Implementation

Training was performed on the supercomputer *JURECA*[3] at the Jülich Supercomputing Centre at Research Centre Jülich (JSC) Krause and Thörnig (2018). Each compute node was equipped with four NVidia K80 GPUs with 12 Gigabyte of VRAM, 2 Intel Xeon E5-2680 v3 Haswell CPUs (12 2.5 GHz cores with hyperthreading each) and 128 Gigabyte of RAM (Krause and Thörnig, 2018). Training of one model occupied one GPU node, using all 4 GPUs and all 48 threads. Of the available 48 threads, 4 were assigned to one GPU each to coordinate the training process, while the remaining 44 threads read training patches from disk in a streaming fashion, applied data
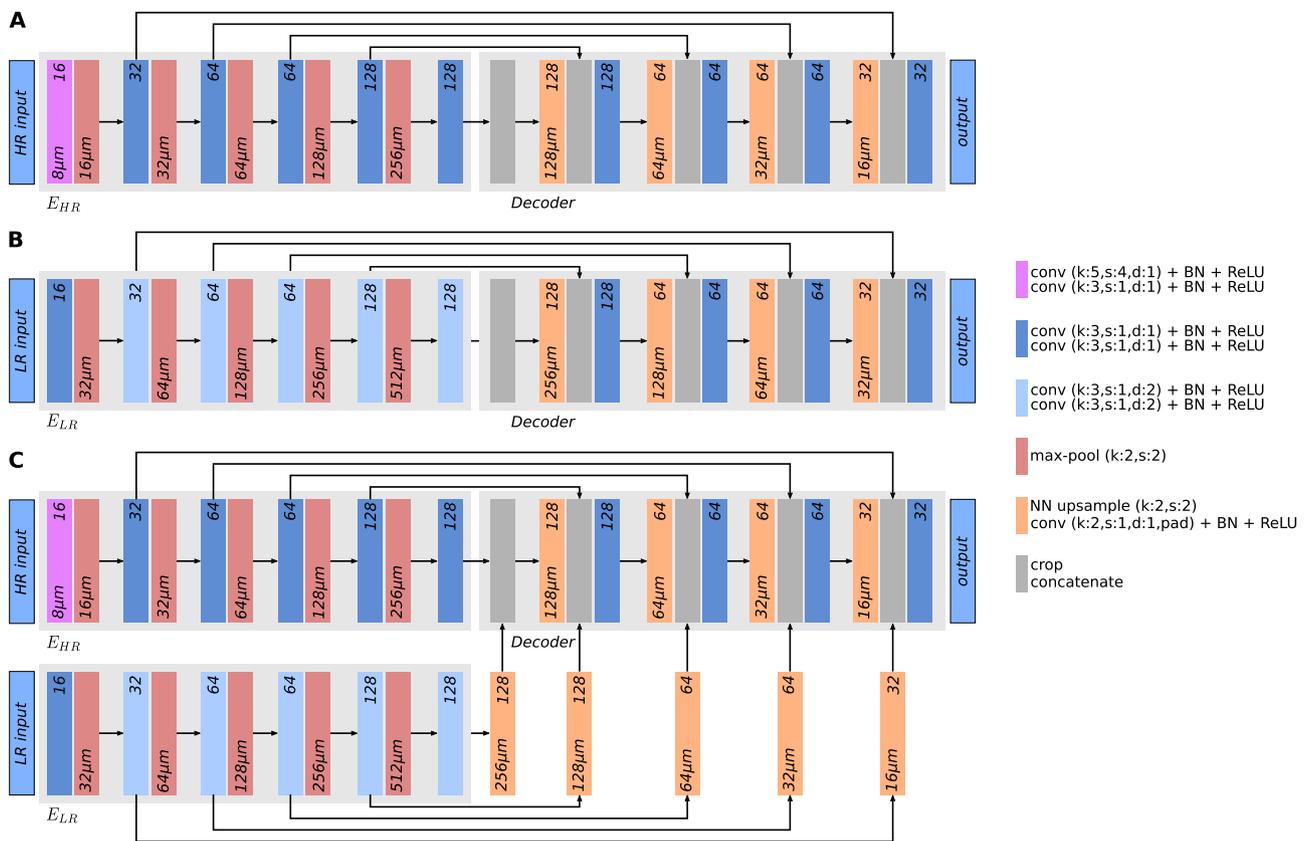

Fig. 5. Illustration of investigated neural network architectures. *A:* High-resolution architecture (HR) from (Spitzer et al., 2017), which can capture fine-grained microstructural textures. *B:* Low-resolution architecture (LR), which can capture macroscopic tissue features. *C:* Proposed multi-scale architecture (MS) to capture both fine and coarse grained tissue features. E_{HR} is pre-initialized with weights of the self-supervised network proposed in Spitzer et al. (2018). Numbers at the top of each block denote the number of filters used in the convolutional layers of this block. Numbers at the bottom denote the physical output spacing in µm per pixel for layers which change the physical spacing of the features.

augmentation and sent data to the training threads. Inter-process communication was implemented based on Message Passing Interface (MPI) using *mpi4py* (Dalcin et al., 2011). Training was implemented using *TensorFlow* (Abadi et al., 2016). Distributed training was performed using *Horovod* (Sergeev and Del Balso, 2018) and synchronous distributed stochastic gradient descent. Batch size was set to 16 image patches per GPU, resulting in a total effective batch size of 64 patches per iteration. The linear learning rate scaling rule for distributed training proposed in Goyal et al. (2017) was employed, scaling the learning rate by the number of GPUs.[4] Batch normalization statistics were computed independently for each GPU and not averaged during training. Software code is publicly available [5].

2.6. Web-based interactive workflow for efficient cytoarchitectonic mapping

The proposed workflow was implemented as an interactive web application (Fig. 6) to provide direct user control over the segmentation workflow through a web browser.[6] The application allows entering annotations in a sparse set of reference sections, controlling the training workflow on a remote cluster, and efficiently inspecting predicted segmentations in the complete stack of histological sections. It does not require in-depth expertise in Deep Learning and/or batch computations.
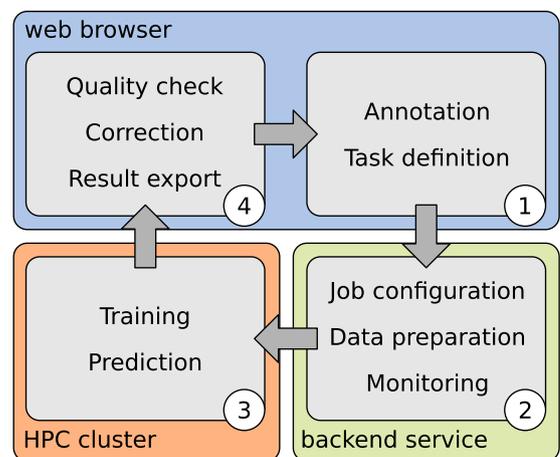

Fig. 6. Overview of the mapping workflow. The user starts by creating annotations (*1*) of a brain area *a* using the web-based annotation tool *microdraw*, and defines training tasks by specifying which annotations should be used to train local segmentation models (Section 2.5). Annotations and task definitions are then submitted to a backend web service (*2*) which prepares the data for training and submits a job to a HPC cluster. Training and subsequent prediction are performed on the HPC system (*3*). Obtained results can be viewed directly in *microdraw* for quality control (*4*). The user may decide to export results of sufficient quality for subsequent processing steps (e.g. 3D reconstruction), manually refine the predictions directly in *microdraw*, or repeat the workflow with additional annotations to improve performance.

---

[4] Since we use a relatively small number of employed GPUs however, we do not apply the initial learning rate warm up phase described in Goyal et al. (2017).

[5] Code available at https://jugit.fz-juelich.de/c.schiffer/atlas

[6] Code available at https://jugit.fz-juelich.de/c.schiffer/atlasui

Technically, it is designed as an extension of the web based annotation tool *microdraw* [7], combined with a novel backend service that controls data exchange and job supervision on an ssh-accessible compute cluster. We used the workflow on the JURECA supercomputer at the Jülich Supercomputing Center (JSC). It uses common and freely available software components, and is portable to other sites, potentially requiring site-specific adjustments to account for differences in the software stack, scheduling system and data access.

The workflow typically iterates through the following steps:

1. The user enters annotations for a target brain area a in two tissue sections s_1 and s_2, enclosing a local stack interval of ≈ 100 sections using *microdraw*, and this way defines a local segmentation model $f^a_{[s_1, s_2]}$ (Section 2.3).

2. A training task for the local segmentation model is submitted as a job to a GPU cluster at the push of a button, using default parameters (Section 2.5). It does not require any further configuration. Training typically takes 70 min on one compute node of the JURECA supercomputer. Multiple jobs can be submitted in parallel, if the cluster allows.

3. After training, predictions for all sections in the interval are automatically generated. For a large area like $hOc1$, this takes approximately 30 min for 120 sections. Computed segmentations are automatically displayed in the web frontend once they become available. Data synchronization between the web server and compute nodes is handled by the backend service.

4. After inspecting the segmentation quality, the user can choose to enter additional training data, either reducing the size of the current interval or initiating the next interval in the stack.

2.7. Validation framework and strategy

Additional sections with annotations in between the training sections were used for validating performance of local segmentation models on sections that were not seen during training (orange diamonds in Fig. 1). Segmentations of these test sections were quantitatively evaluated using the F1 score (also known as Dice score or Sørensen-Dice index), computed as the harmonic mean of precision of recall. Auxiliary labels added to ensure convergence (Section 2.5) were excluded from F1 score calculation, as the focus lies on segmentation performance for target area a.

Similar to the proximity sampling strategy employed for training (Section 2.5), segmentations on sections not seen during training were only created and evaluated in the approximate region containing a on the respective sections. These approximate regions were determined by projection of the closest reference annotations for a to the image in question using conventional linear image registration based on robust image features as in Dickscheid et al. (2019).

The benefit of a multi-scale architecture was investigated by training separate local segmentation models with neural network architectures HR, LR and MS for all areas in B20. For HR and MS, the high-resolution encoder E_{HR} was initialized with the weights of the network from Spitzer et al. (2018). Furthermore, the performance of multiple local segmentation models, each trained on a local subset of sections as described in Section 2.3, was compared to the performance of one single model trained on all annotations available for a target area a in the following way: For each target area in the B20 dataset, one model was trained using the union of all training sections of the local segmentation models (blue squares in Fig. 1), using the same training strategy as for local segmentation models. We conducted these experiments using HR, LR and MS architectures, and denote models trained on the whole stack as HR (ALL), LR (ALL) and MS (ALL), respectively, again pre-initializing the high-resolution encoder E_{HR} with weights from Spitzer et al. (2018)

_____
[7] http://microdraw.pasteur.fr

The robustness of the proposed method against intersubject variability in brain structure and differences in staining protocols was investigated by training local segmentations models (with MS architecture) for areas $hOc1$ and $hOc2$ in datasets B01 and AAHB. To better understand the roles of the low and high-resolution branches (E_{LR} and E_{HR}) in the MSarchitecture, an experiment similar to occlusion sensitivity analysis (Zeiler and Fergus, 2014) was performed: Using model $f^{hOc2}_{[901,1021]}$ which implements the MSarchitecture, we investigated how predictions change when we set the input patch for either E_{LR} or E_{HR} to zero, effectively preventing information extraction using the respective branch.

2.8. Generating high-resolution 3D cytoarchitectonic maps in the BigBrain dataset

Non-linear transformations described in Amunts et al. (2013); Omidyeganeh et al. (2020) from 2D histological sections into 3D reconstructed space available for the BigBrain dataset Amunts et al. (2013) were used to generate 3D maps for areas $hOc1$, $hOc2$, $hOc3v$ and $hOc5$ from 2D segmentations produced by our method. Segmentations were obtained using the workflow described in Section 2.3 and checked for quality by an expert (e.g. plausibility and consistency across consecutive sections). For areas $hOc3v$ and $hOc5$, results of segmentation models trained with a training interval size of 1.2 mm were used for reconstruction (marked with ∗ in Fig. 3). Between 8% ($hOc3v$) and 23% ($hOc1$) of sections containing the investigated areas were not used for reconstruction due to histological artifacts (e.g. resulting from long-term storage or staining inhomogeneities). Segmentations that passed the quality check were transformed into the 3D reconstructed space. Excluded sections were replaced by interpolations from neighboring sections, using Laplacian fields as proposed in Schober et al. (2016).

Resulting 3D maps were smoothed using a median filter with kernel size $11 \times 11 \times 11$ pixel to compensate for small artefacts. The size of the filter was chosen to match the expected precision of annotations at boundaries (not higher than 100 μm), translating to 5 voxels at the target resolution of 20 μm. Furthermore, connected component analysis on the smoothed volume was performed to determine and remove spurious false positive predictions outside the target area, relying on the assumption that cytoarchitectonic areas are continuous in 3D. Only components with a minimum volume of 27 mm³ (3 mm × 3 mm × 3 mm) were kept. Effects of median filtering and connected component filtering are illustrated in Fig. 7.

To assess the improvement in 3D consistency and anatomical plausibility gained by the proposed workflow, a reference reconstruction of area $hOc1$ was computed, which performs a direct 3D interpolation between reference annotations obtained by GLI mapping. This reference reconstruction does not use the local segmentation models, and relies only on reference annotations and 3D reconstruction. It was computed by transforming the annotations of the training sections (blue squares in Fig. 3) into the 3D reconstructed space, and filling the gaps by Laplacian field interpolation (Schober et al., 2016).

The anatomical consistency of 3D reconstructed maps was further evaluated by computing their volume and surface area, which were then compared to reference values from Amunts et al. (2000). The volume of each area was computed by counting the total number of labeled voxels and multiplying the result by the physical size of each voxel.

The surface area was computed by first extracting a closed surface mesh of each area using the marching cubes algorithm Lewiner et al. (2003). The subset of mesh vertices lying on the pial surface was then determined by including all triangles where the cortical depth Bok (1929) was smaller than 0.25. To obtain the cortical depth of each mesh vertex, the procedure described in Leprince et al. (2015) was applied to the cortical ribbon defined by the gray and white matter segmentation provided with the BigBrain model (Lewis et al., 2014). The result was a volumetric dataset with voxels in the white matter labelled 1, voxels outside the brain labelled
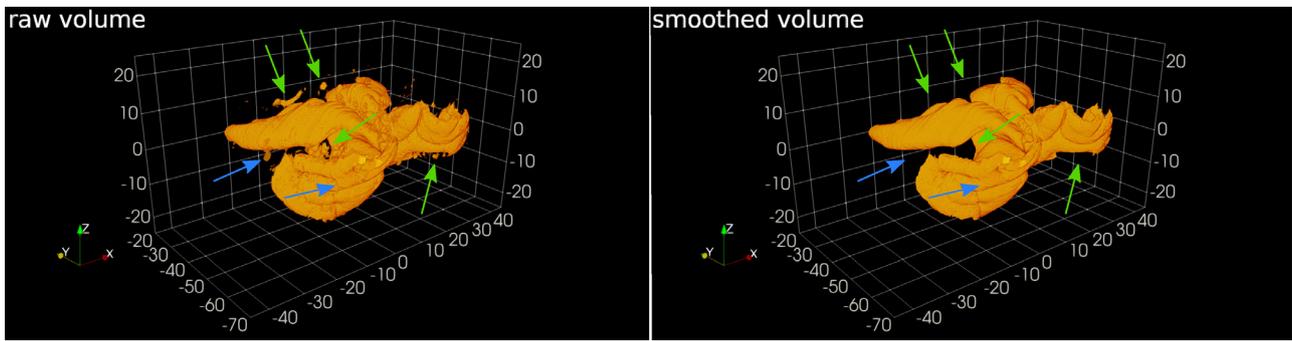
Fig. 7. Effects of median filtering and connected component filtering using the example of $hOc1$. Median filtering smooths the volume and removed small errors originating from registration errors or incorrect predictions (blue arrows). Filtering of small connected components removed small clusters of false positives from the volume (green arrows). Axes x, y and z correspond to left-to-right, posterior-to-anterior and ventral-to-dorsal directions, respectively. Axis labels are specified in mm and correspond to positions in the 3D reconstructed BigBrain space. See Fig. 14 for more images of $hOc1$ from different viewing angles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

0, and voxels inside the isocortex labelled with values between 0 and 1, representing their cortical depth according to the equivolumetric model (Bok, 1929). Cortical depths of mesh vertices were then looked up in this volume. Finally, the surface area of the pial surface for each cytoarchitectonic area was computed by summing up the area of all triangles associated to the pial surface.

Both volume and surface area measurements were corrected for tissue shrinkage Amunts et al. (2000). The volume-based shrinkage factor for B20has been determined in Amunts et al. (2005) based on the fresh weight and the volume after histological processing as $f_V = 1.931$. From this, an area-based (2D) shrinkage factor of $f_A = f_V^{2/3} = 1.551$ was derived.

3. Results

Differences in performance were observed depending on the network architecture (HR, LR, MS), the training setting (global vs. local segmentation models), the considered brain area, as well as the distance between annotated brain sections.

All architectures except for LR (ALL) show comparably good performance for $hOc1$. For most areas however, LR and MS achieved higher performance than other investigated models. For areas $hOc3v$ and $hOc5$, where additional models were trained with reduced distance between training sections (indicated by $*$ in Fig. 8), performance of LRand MSincreased when reducing the distance between training sections, while only minor improvements were observed for the remaining architectures.

Representative image patches segmented by the MSarchitecture for each investigated area extracted from test sections of B20are shown in Fig. 9. True positive, false positive and false negative predictions are indicated in green, red and blue, respectively. A large share of incorrectly classified pixels belonged to cortical regions with highly oblique cutting angles (Fig. 10 B, C). While large rifts tended to be excluded from the prediction (Fig. 10, A), smaller rifts or tissue foldings were correctly segmented as surrounding area (Fig. 9, A, D, E, G, O).

Scores obtained for areas $hOc1$ and $hOc2$ were overall consistent across different brain samples (Fig. 11). In all three cases, scores obtained for $hOc2$ were lower compared to $hOc1$. Lowest median F1 score for $hOc2$ was obtained for B20, along with an increased variance. Example patches showing the border between $hOc1$ and $hOc2$ on test sections extracted from approximately identical brain regions in the three datasets are shown in Fig. 12.

Models trained on all sections (HR (ALL), LR (ALL), MS (ALL)) obtained lower mean and median F1 scores than their locally trained counterparts HR, LR, and MS (Table 1). LR (ALL) and HR (ALL) showed comparable performance, MS (ALL) performed slightly better. The lowest

Table 1

F1 score statistics computed across all areas and test sections in the B20dataset obtained by the different network architectures HR, LRand MS(trained on local intervals), as well as HR (ALL), LR (ALL)and MS (ALL)(trained on all annotated sections per area). Higher mean/median values and lower standard deviation mean better performance.

model	median	mean	std
HR(all)	0.5319	0.5680	0.2075
LR(all)	0.5648	0.5533	0.1723
MS(all)	0.5869	0.6020	0.1973
HR	0.6294	0.6130	0.2105
LR	0.7439	0.7036	0.1865
MS	0.7469	0.7200	0.1825

scoring local model HR performs better than the highest scoring global model MS (ALL). Both LR and MS resulted in higher mean and median F1 scores than HR, with lower standard deviations. Highest mean and median performance was obtained by MS (Table 1, Fig. 8).

Setting the input patch of either E_{LR} or E_{HR} to zero provides indication on the influence of different scales in the proposed MSarchitecture (Fig. 13): Having access to only low-resolution image information, the model still identifies the approximate location of area $hOc2$, but with poorly defined borders. Using only high-resolution information, the model captures finer details, but has difficulties localizing the area correctly. When the model has access to both high- and low-resolution information, this results in better agreement with the reference annotations.

Locations, orientations and shapes of reconstructed 3D maps (computed using steps described in Section 2.8) were anatomically plausible and consistent (Fig. 14). The 3D map of $hOc5$ showed partially missing extremal ends along the posterior anterior axis. Volume and surface estimates from the 3D maps reported in Table 2 corresponded well with the numbers reported in Amunts et al. (2000). Surface areas of $hOc1$, $hOc2$ and $hOc5$ were largely confirmed with the reference values, as well as the volumes derived from automatic segmentations of areas $hOc1$ and $hOc2$. The reconstructed volume of area $hOc5$ stood out by being considerably smaller than the reference volume.

Comparison of corresponding 3D reconstructions of area $hOc1$ (Fig. 14 E vs. F) showed that the proposed approach provided anatomically more consistent results than direct spatial interpolation of GLI-based annotations, while both build on the same annotation effort. 3D interpolation produced abrupt transitions in anterior-posterior direction
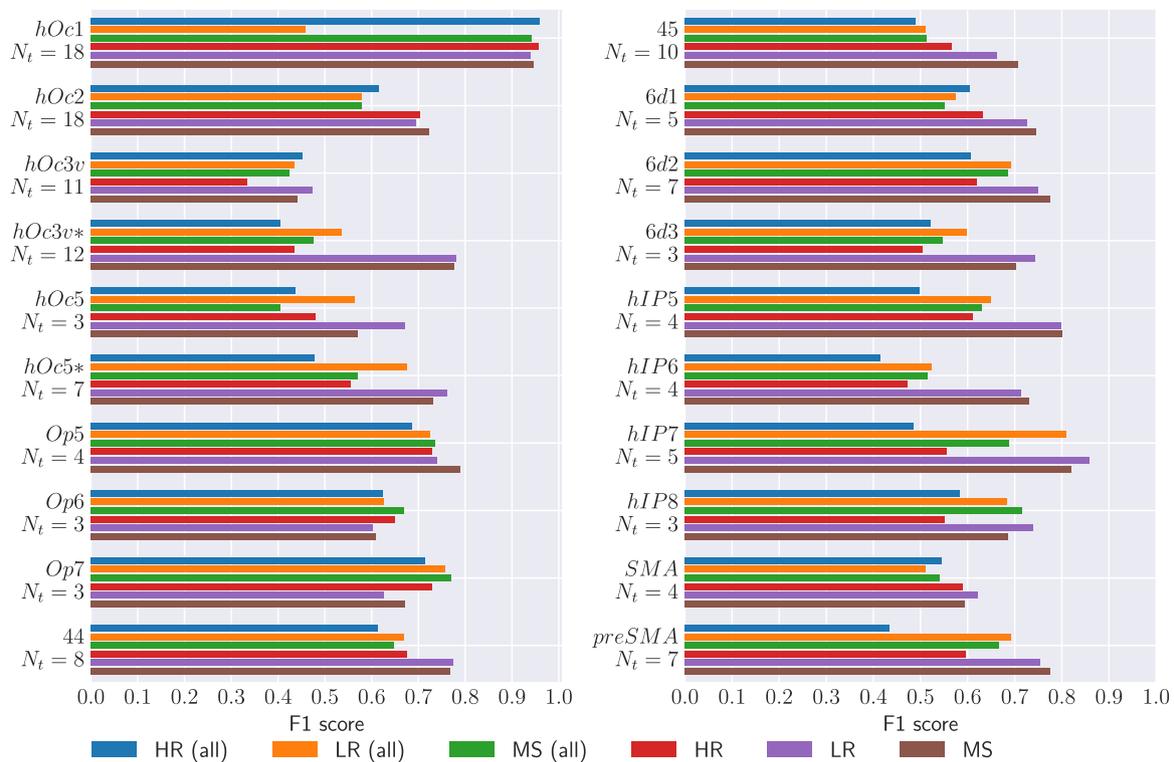
Fig. 8. Median F1 scores for HR (ALL), LR (ALL), MS (ALL), HR, LRand MS per investigated brain area in dataset B20. N_t denotes the number of test sections for which F1 scores were computed for a particular area. ∗ indicates where training of local segmentation models was performed with reduced distance between training sections. Higher values denote better performance.
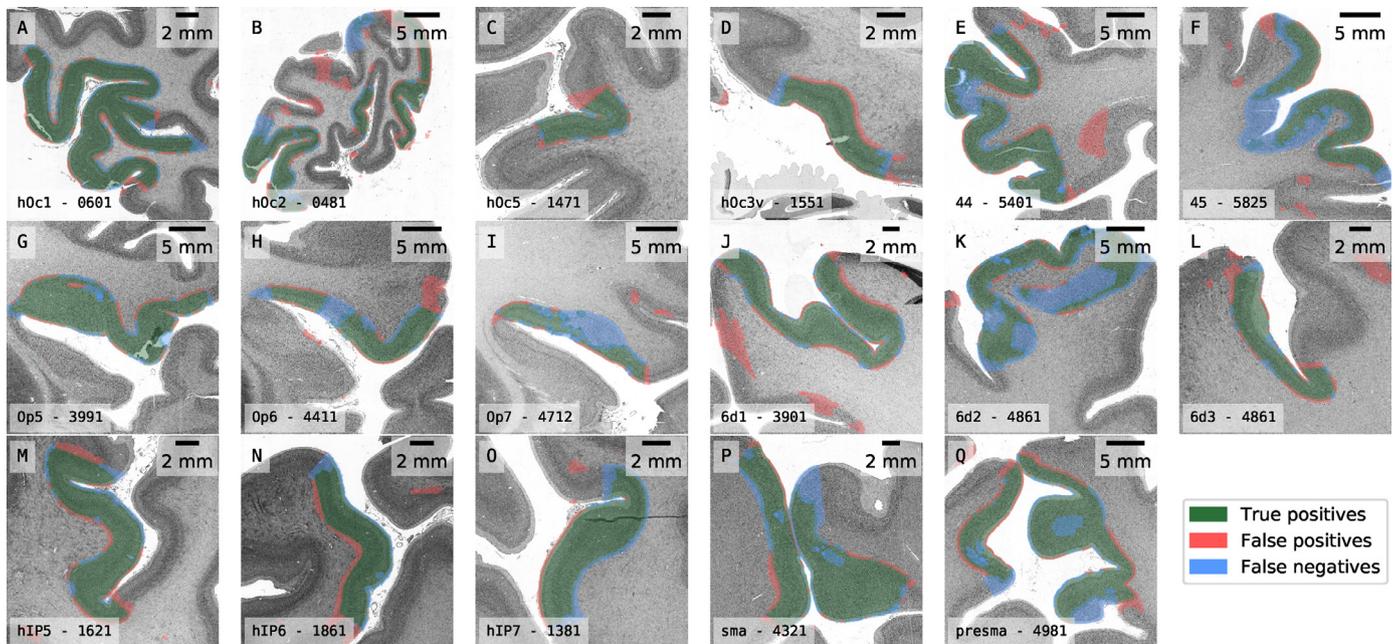

Fig. 9. Example image patches and corresponding model predictions extracted from test sections of B20segmented using the proposed MSarchitecture. One image patch is shown for each investigated cytoarchitectonic area. Colors green, red and blue indicate true positive, false positive and false negative predictions, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 14, F, 1) and only captured structures already contained in the reference annotations, leading to inconsistencies near fine-grained morphological structures (e.g. Fig. 14 F, 2 and 3). The proposed method often produced reasonable segmentations for sections outside the training interval (Fig. 14, E, 1), which interpolation cannot provide by definition.

4. Discussion

In this work, we proposed a novel Deep Learning based workflow to create segmentations of cytoarchitectonic areas in large series of histological human brain sections using only a limited set of manually created annotations. We evaluated this approach across different cytoar-
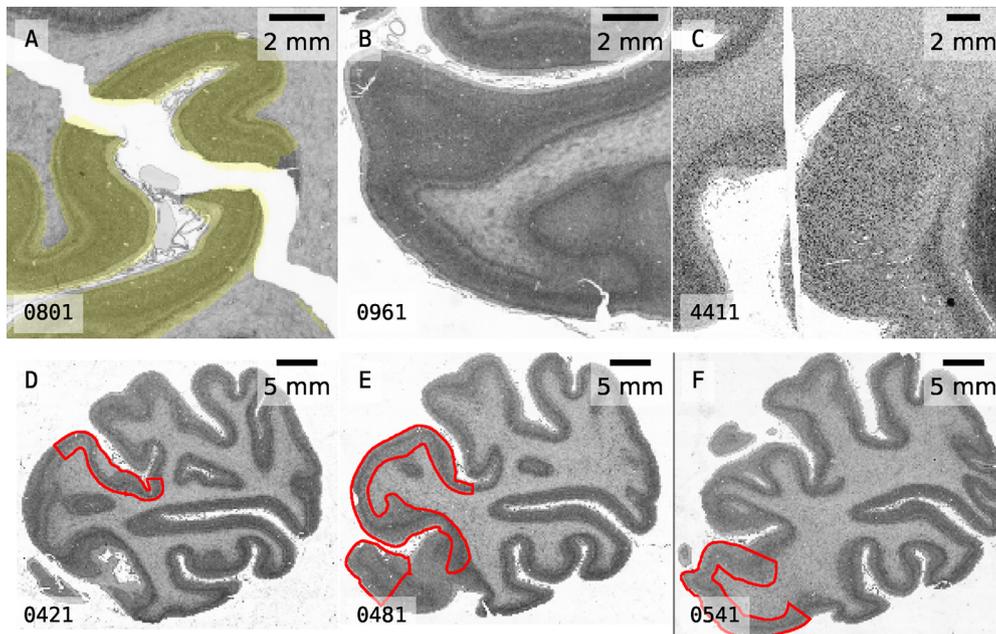
Fig. 10. Image patches extracted from B20 showing common challenges encountered during manual and automated cytoarchitectonic mapping. *A:* Mechanical damages resulting from histological processing. Prediction for $hOc1$ shown in yellow demonstrate handling of larger mechanical damages. *B:* Region where the cutting angle is highly oblique, leading to partial or full occlusion of cortical layer structure (*oblique cuts*). *C:* Mechanically damaged and obliquely cut tissue. *C-F:* Example illustrating highly variable morphology of area $hOc3v$ (highlighted in red) across 120 histological sections in B20. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
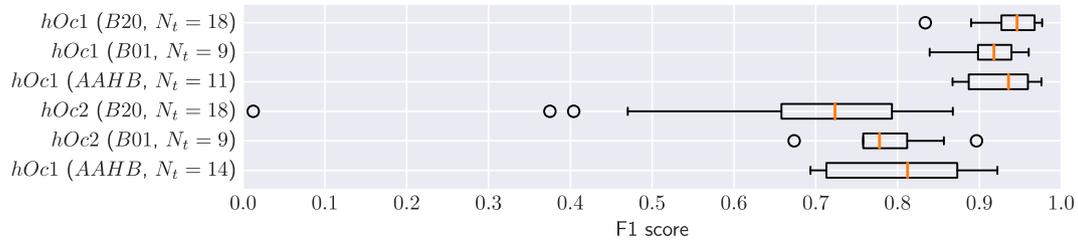

Fig. 11. F1 scores for segmentations of $hOc1$ and $hOc2$ obtained by the MSarchitecture on test sections of datasets B20, B01 and AAHB. N_t denotes the number of test sections for which F1 scores were computed for a particular area.
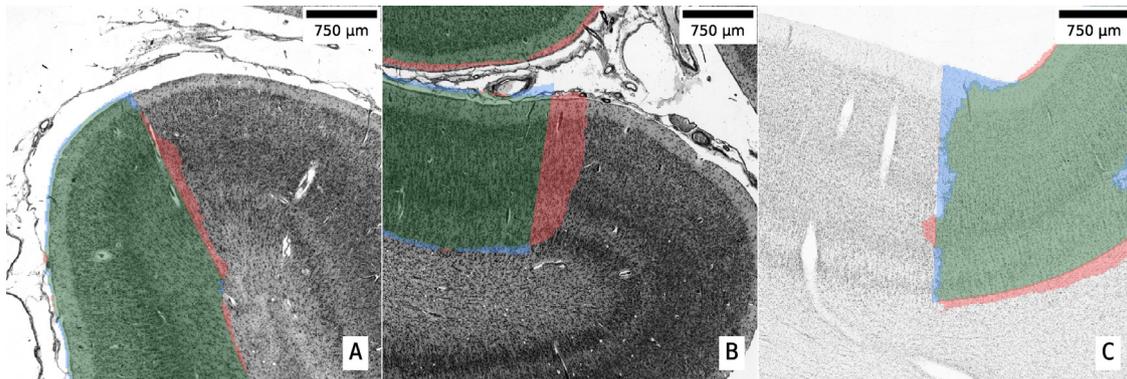

Fig. 12. Example patches and typical segmentation results extracted from test sections in datasets B20 (A), B01 (B) and AAHB (C). All three patches show the segmentation of $hOc2$ obtained by a local segmentation model with MSarchitecture. Patches were extracted at the border between $hOc1$ and $hOc2$ and in comparable regions of the respective brain. Colors green, red and blue indicate true positive, false positive and false negative predictions, respectively (see also legend in Fig. 9). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

chitectonic areas, brain samples and staining protocols. As a concrete use case, we then applied it to create high-resolution 3D maps of areas $hOc1$, $hOc2$, $hOc3v$ and $hOc5$ in the BigBrain Amunts et al. (2013).

4.1. Quality of derived 3D maps in the BigBrain

The proposed method produced 3D maps with a high degree of anatomical consistency and identified cytoarchitectonic areas precisely in the histological brain sections. Partially missing extremal ends re-

main a challenge, as seen in anterior-posterior direction of $hOc5$. Such parts are often difficult to identify even using manual methods. Therefore, training data for such extremal ends is difficult to obtain. The segmentation of extremal ends could potentially be addressed by providing additional GLI-based mappings (at the cost of additional annotation effort), or by an explicit shape-based inference step on top of the pixel segmentation. The 3D map of $hOc1$ created with the proposed method is superior to the map obtained by direct spatial interpolation between GLI-based annotations. Methods based on 3D interpolation inherit any
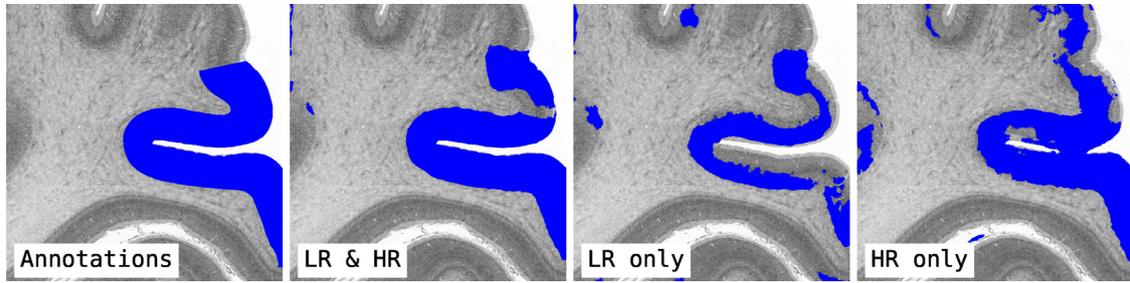
Fig. 13. Reference annotations and predictions of area $hOc2$ (blue) for an example patch from section 961 of the B20dataset. Predictions were obtained by using both low- and high-resolution information (LR & HR), only low-resolution (LR only), and only high-resolution information (HR only). The input image patch for E_{HR} and E_{LR} was set to zero to investigate the role of low- and high-resolution image information, respectively. Predictions were created with model $f_{[901,1021]}^{hOc2}$ using the MSarchitecture. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
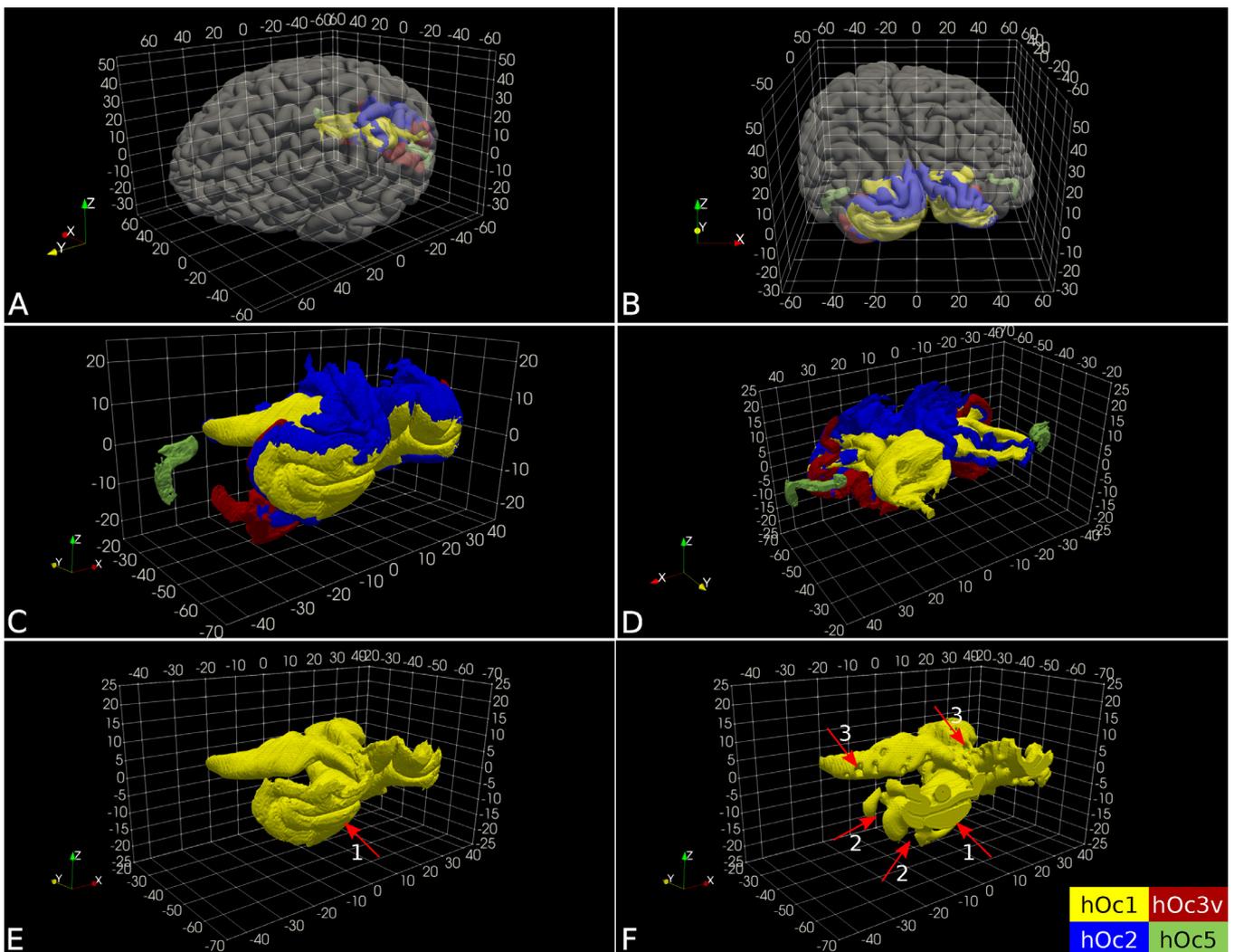

Fig. 14. 3D maps of visual cytoarchitectonic areas $hOc1$ (yellow), $hOc2$ (blue), $hOc3v$ (red) and $hOc5$ (green), obtained by transforming the independent 2D segmentations generated by the proposed method into the 3D reconstructed space of the B20dataset. $A + B$: Spatial embedding of reconstructed areas into the 3D reconstructed BigBrain volume. $C + D$: Detailed view of reconstructed cytoarchitectonic areas. $E + F$: Comparison of $hOc1$ reconstructed based on our proposed method (E) and based on an interpolation between annotations in the reconstructed space, using Laplacian fields as proposed in Schober et al. (2016) (F). Arrows in F) mark example locations demonstrating shortcomings of the interpolation based reconstruction. Axes x, y and z correspond to left-to-right, posterior-to-anterior and ventral-to-dorsal directions, respectively. Axis labels are specified in mm and correspond to positions in the 3D reconstructed BigBrain space. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Estimated volumes (in mm³) and surface areas (in mm²) of brain areas derived from the full 3D maps in the 3D reconstructed space of the B20 dataset. Reference mean μ and standard deviation σ were computed based on male subjects from (Amunts et al., 2000). Shrinkage corrected of volumes and surface areas was performed using correction factors $f_V = 1.931$ and $f_A = 1.551$ respectively (Amunts et al., 2005).

area	volume	corrected	μ	σ	z-score
hOc1	9019.30	17416.27	18042.2	2464.39	−0.25
hOc2	6448.60	12452.26	12634.2	2862.84	−0.06
hOc3v	1974.76	3813.26	n.a.	n.a.	n.a.
hOc5	304.10	587.21	1144.4	406.53	−1.37
area	surface	corrected	μ	σ	z-score
hOc1	6891.03	10685.76	12213.0	2225.55	−0.69
hOc2	6749.64	10466.52	10390.4	2925.37	0.03
hOc3v	2142.04	3321.62	n.a.	n.a.	n.a.
hOc5	319.79	495.89	450.2	135.92	0.34

error in the alignment of consecutive sections, making them inappropriate for stacks with only linear or no 3D reconstruction. The proposed method does not assume any prior 3D reconstruction - in fact its outputs might be used to guide image registration with landmarks.

4.2. Practical usefulness of the implemented workflow

The presented method showed good robustness against intersubject variability and different histological processing protocols. Thus it largely overcomes the need for brain or area specific parameter adjustments, which makes it well suited to be used as a self-contained tool for neuroscientists. Consequently, it was possible to implement it into a web application that provides a practical mapping workflow for end users from different disciplines. The web application is currently used by five neuroscientists in our institute for their research projects, without requiring support from a computer scientist. The interactive workflow enables efficient mapping of brain areas across full series of histological sections, en par with high throughput microscopy. Such efficiency of mapping was previously impossible in our experience. To give a concrete example, we consider that a trained expert typically needs 30–60 min to identify cytoarchitectonic borders for one cortical area on a single tissue section. Using the established GLI-based mapping approach (see Section 2.1), this would translate to an approximate effort of 150 work days (8 h per day) to map hOc1 across the whole stack of 2461 sections. In comparison, the proposed method required annotation of only 18 sections to generate precise segmentations of the complete stack, corresponding to approximately 9 working hours. Altogether, including quality checks and computations, the presented workflow allows precise mapping of a large brain area in the order of 1–2 weeks - a task that would require almost a year of work with previously established methods.

Although the workflow provides a high degree of automation, we still recommend final inspection of results by an expert to ensure optimal quality. The interactive web application presented in Section 2.6 assists users with such quality control, by displaying predictions and allowing immediate correction of remaining errors with significantly less effort than annotating images from scratch. Of course, the typical amount of necessary manual corrections is an important indicator for the usefulness of the tool in practice. In our experience from mapping a whole range of different human brain regions, quality control and manual corrections typically take in the order of few hours per brain area, which may include several thousand sections.

Nevertheless, the need for manual supervision could be further reduced by investigating into methods for identifying prediction errors. Such methods could directly inform the user where additional annotations could help to further optimize the results, thereby realizing an active learning (Settles, 2009) scheme.

4.3. Ability to distinguish higher associative areas

In contrast to primary areas such as the primary visual cortex hOc1, so called higher associative areas have a less distinct cytoarchitecture, and less prominently differ from their neighbouring areas. Such observation lead Bailey and von Bonin to the conclusion that it is almost impossible to reliably distinguish such areas from each other, and to define borders between them (Bailey and von Bonin, 1951). This view is not supported any more due to the possibility to identify cytoarchitectonic borders in a reliable and reproducible way (for an overview see (Zilles and Amunts, 2010)). However, the fact that intersubject differences between identical areas of different brains may exceed cytoarchitectonic differences between two neighboring areas in one and the same brain creates challenges for modern brain mapping (Amunts et al., 1999).

Atzeni et al. (2018) also addressed automated mapping of histology. They segmented brain structures in a serial stack of human brain sections from the Allen Human Brain Atlas (Ding et al., 2016) (dataset AAHB used in our experiments). They used annotations from Ding et al. (2016) on a small set of sections at regular intervals, in order to train a probabilistic model that combines multi-atlas segmentation with a CNNs to segment the remaining sections. Compared to the present work however, their approach is restricted to brain structures that can be recognized at a resolution of 250 μm. The authors confirm in their paper that more subtle classes, in particular subdivisions of the isocortex, introduce excessive noise with their approach. The method presented here segmented both hOc1 and hOc2 in the same dataset with high accuracy by including more fine-grained texture features into the models, thus going clearly beyond this restriction.

4.4. Effect of the local segmentation models

Previous work on automatic cytoarchitectonic brain mapping using machine learning emphasized the importance of strategies for efficient exploitation of available training data and prior information. This includes incorporating probabilistic priors from brain atlases Spitzer et al. (2017) and self-supervised learning Spitzer et al. (2018). The key idea of the present paper is to use multiple local segmentation models, each of which focuses on a spatially restricted subset of sections in one specific brain area. In order to maximize practical benefit, we make an explicit design decision not to aim for a general classification model of multiple brain areas and brains. The benefit of such local segmentations models is confirmed by our experiments, which showed significantly improved performance of HR, LR, and MS compared to their globally trained counterparts HR (ALL), LR (ALL)and MS (ALL).

A major advantage of the local segmentation models is the ability to flexibly adjust the distance between training sections to account for regions with particularly simple or complex properties. This has been demonstrated for the challenging areas hOc5 and hOc3v, where a reduction of the distance between training sections from 120 (2.4 mm) to 60 (1.2 mm) improved precision to a satisfactory level while keeping the annotation effort tractable.

Distance reduction results in major performance gains when using local segmentation models LRor MS, but only minor gains when using globally trained models LR (ALL) or MS (ALL). This suggests that local segmentation models make more efficient use of the additional training data.

In a similar fashion, larger areas or areas with distinct cytoarchitectonic features (e.g. hOc1) can be segmented with a coarser set of training sections, in this case reducing annotation effort.

The availability of expert annotations limits our ability to evaluate the effect of reducing distance between annotated sections. The results of our experiments for areas hOc3v and hOc5 suggest, that the optimal distance between annotated sections depends on the cytoarchitectural and morphological complexity of a brain area. The proposed interactive

workflow allows users to add annotations incrementally until satisfied with the segmentation.

On the downside of local models, hyperparameter assessment (e.g. for learning rate or model architecture) is not straightforward when training multiple models on different training sets and evaluating them on individual test sets. Model performance needs to be evaluated across several areas, sections and brains, which can be computationally expensive and lead to a slow development process.

4.5. Effect of the multi-scale model architecture

Macroscopic features of the cerebral cortex (e.g. folding patterns) vary between individual brains. While the location of gyri and sulci can provide guidance for localizing brain areas, such coarse landmarks cannot generally be used to precisely predict cytoarchitectonic borders (Amunts and Zilles, 2015; Fischl et al., 2008; Im and Grant, 2019; Lebenberg et al., 2018). Consequently, established methods for identifying cytoarchitectonic areas rely on high-resolution microscopic information Schleicher et al. (1999); Spitzer et al. (2017, 2018). In contrast, the proposed local segmentation models are able to exploit macroscopic features for improving segmentation performance, thanks to their specialization on only a part of a specific brain area from one individual brain. This can be seen from the higher scores produced by the LR and MS architectures for local segmentations models (Table 1), and by the lack of such an effect for models trained on all sections of a brain area (HR (ALL), LR (ALL), MS (ALL)). Microscopic resolution features further contribute to the performance of the local segmentation models, as verified by a Wilcoxon signed-rank test (Wilcoxon, 1945) ($p = .0011$) which confirms that the multi-scale approach of MS further improves the performance compared to LR. However, this is a relatively small effect compared to the improvement between HR and MS. We can therefore assume that macroscopic information is more relevant than microscopic information in the strictly local setting.

For further understanding the influence of the two scales in the model, it is helpful to compare predictions obtained when occluding (i.e. setting to zero) either the microscopic or macroscopic inputs in a MS model: Using only low-resolution inputs, the model is still able to approximately locate the area, but fails to capture the fine details. On the other hand, a model restricted to see only high-resolution inputs struggles to correctly locate the area. This confirms our assumption that the MS model mostly uses low-resolution inputs for localization and high-resolution inputs for local refinement. While the results of this experiment cover only a specific setting (e.g. model, area and input location), they give us some confidence that the model works as expected.

4.6. Failure mode analysis

The predictions produced by the proposed model typically include some remaining errors (Figs. 9, 12), which require careful interpretation.

Close to brain region boundaries, it must be noted that the GLI-based reference annotations enforce straight lines to model the border, which reflects the vertical arrangement of neurons in columns as a major principle in cortical organization (Schleicher et al., 2000; 1999). This may lead to discrepancies with the present segmentations, that do not enforce such constraints (see for example Fig. 12). In fact the location of the boundary is not determined down to the single pixel, and as a consequence, pixel-level metrics have a somewhat limited significance there.

Many of the remaining classification errors further coincide with highly oblique cutting angles of the tissue. As also reported in (Schleicher et al., 1999; Spitzer et al., 2017; 2018), identification of cortical areas is almost impossible at such angles, because the laminar composition of the cortex is then almost invisible in the 2D section. In such cases, experts would consult adjacent sections to identify areas, which the proposed method cannot do. An extension of the method

considering multiple adjacent sections for classification might be able to overcome this issue.

Whether or not remaining segmentation errors are critical in practice depends on the availability of postprocessing methods for correcting them. In the experiments that we carried out, the precise 3D reconstruction of the BigBrain dataset could be used for removing spurious errors (Section 2.8). However, if no precise 3D reconstruction is available, the manual effort for quality control of predictions and any necessary error corrections increases. For such settings, it would be beneficial to develop additional heuristics to identify errors, e.g. for detecting inter-section inconsistencies from only approximate section alignments.

5. Conclusion

A novel method based on Convolutional Neural Networks (CNNss) was introduced for automated mapping of cytoarchitectonic areas in large series of histological human brain sections. Segmentation models were trained for segmentation of different cytoarchitectonic areas in histological stacks obtained from three different brain samples. A key idea is to train separate local segmentation models based on annotations of one specific target area in only two training sections, to focus the learning process on microscopic and macroscopic tissue features close to the training sections. Local segmentation models enable exploitation of low-resolution macroscopic information and significantly improve performance over globally trained models. After training, local segmentation models were able to accurately segment sections in between their respective training sections. By concatenating results from multiple local segmentation models, segmentations for complete brain areas can be obtained. The proposed method opens up new possibilities to map complete stacks of histological human brain sections in a highly automated fashion, and thus provides an important basis for building high-resolution human brain maps for datasets like BigBrain. To the best of our knowledge, the maps of areas $hOc1$, $hOc2$, $hOc3v$ and $hOc5$ computed for the BigBrain model using this method are the first high-resolution 3D maps of human cytoarchitectonic areas created from full stacks of histological sections at cellular resolution. These maps enable precise studies of area-specific morphological and columnar features at microscopic resolution, and in combination with existing cortical layer maps (Wagstyl et al., 2020) an investigation into layer-specific aspects of each region. Dense maps further enable straightforward mapping from the volume to the whole brain mesh surface, which in turn facilitates comparison with other modalities, especially in-vivo imaging. They represent an important contribution for using BigBrain as a microscopic resolution reference space, since they provide direct links to probabilistic cytoarchitectonic reference parcellations at the macroscopic scale (Amunts et al., 2020) that are widely used in neuroimaging studies. As such, our work makes an important contribution to linking neuroscientific findings across spatial scales.

Ethics Statement

The study carried out requires no separate ethical approvals. Postmortem brains were obtained in accordance to legal and ethical regulations and guidelines. Brain tissue for datasets B01 and B20 was obtained through the body donor program of the department of anatomy of the Heinrich Heine University Düsseldorf and with approval of the ethics committee of the medical faculty of the Heinrich Heine University Düsseldorf. Brain tissue for dataset AAHB was obtained from the University of Maryland Brain and Tissue Bank and with approval by the Human Investigation Committees and Institutional Ethics Committees of the University of Maryland.

Declaration of Competing Interest

The authors declare no competing interests.

Credit authorship contribution statement

Christian Schiffer: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Hannah Spitzer:** Conceptualization, Methodology, Software, Writing - review & editing. **Kai Kiwitz:** Data curation, Validation, Writing - review & editing. **Nina Unger:** Data curation, Validation, Writing - review & editing. **Konrad Wagstyl:** Software, Writing - review & editing. **Alan C. Evans:** Resources, Writing - review & editing, Funding acquisition. **Stefan Harmeling:** Writing - review & editing, Supervision. **Katrin Amunts:** Conceptualization, Investigation, Validation, Writing - review & editing, Supervision, Project administration, Resources, Funding acquisition. **Timo Dickscheid:** Conceptualization, Methodology, Supervision, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition.

Acknowledgments

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118327.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., 2016. TensorFlow: a system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation OSDI 16, pp. 265–283.

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N.J., Habel, U., Schneider, F., Zilles, K., 2005. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. Anat. Embryol. 210 (5–6), 343–352. doi:10.1007/s00429-005-0025-5.

Amunts, K., Lepage, C., Borgeat, L., Mohlberg, H., Dickscheid, T., Rousseau, M.-É., Bludau, S., Bazin, P.-L., Lewis, L.B., Oros-Peusquens, A.-M., Shah, N.J., Lippert, T., Zilles, K., Evans, A.C., 2013. BigBrain: an ultrahigh-resolution 3D human brain model. Science 340 (6139), 1472–1475. doi:10.1126/science.1235381.

Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., Zilles, K., 2000. Brodmann's areas 17 and 18 brought into stereotaxic space—where and how variable? Neuroimage 11 (1), 66–84. doi:10.1006/nimg.1999.0516.

Amunts, K., Mohlberg, H., Bludau, S., Zilles, K., 2020. Julich-Brain: a 3D probabilistic atlas of the human brain's cytoarchitecture. Science doi:10.1126/science.abb4588.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H.B.M., Zilles, K., 1999. Broca's region revisited: cytoarchitecture and intersubject variability. J. Comp. Neurol. 412 (2), 319–341. doi:10.1002/(SICI)1096-9861(19990920)412:2<319::AID-CNE10>3.0.CO;2-7.

Amunts, K., Schleicher, A., Zilles, K., 2004. Outstanding language competence and cytoarchitecture in Broca's speech region. Brain Lang. 89 (2), 346–353. doi:10.1016/S0093-934X(03)00360-2.

Amunts, K., Zilles, K., 2015. Architectonic mapping of the human brain beyond brodmann. Neuron 88 (6), 1086–1107. doi:10.1016/j.neuron.2015.12.001.

Annese, J., Pitiot, A., Dinov, I.D., Toga, A.W., 2004. A myelo-architectonic method for the structural classification of cortical areas. Neuroimage 21 (1), 15–26. doi:10.1016/j.neuroimage.2003.08.024.

Atzeni, A., Jansen, M., Ourselin, S., Iglesias, J.E., 2018. A probabilistic model combining deep learning and multi-atlas segmentation for semi-automated labelling of histology. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, vol. 11071. Springer International Publishing, pp. 219–227. doi:10.1007/978-3-030-00934-2_25.

Bailey, P., von Bonin, G., 1951. The isocortex of man. Urbana 3.

Bok, S.T., 1929. Der Einfluß der in den Furchen und Windungen auftretenden Krümmungen der Großhirnrinde auf die Rindenarchitektur. Zeitschrift für die gesamte Neurologie und Psychiatrie 121 (1), 682. doi:10.1007/BF02864437.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 424–432.

Dalcin, L.D., Paz, R.R., Kler, P.A., Cosimo, A., 2011. Parallel distributed computing using Python. Adv. Water Resour. 34 (9), 1124–1139. doi:10.1016/j.advwatres.2011.04.013.

Dickscheid, T., Haas, S., Bludau, S., Glock, P., Huysegoms, M., Amunts, K., 2019. Towards 3D reconstruction of neuronal cell distributions from histological human brain sections. Future Trends HPC Disruptive Scenario 34, 223.

Ding, S.-L., Royall, J.J., Sunkin, S.M., Ng, L., Facer, B.A.C., Lesnar, P., Guillozet-Bongaarts, A., McMurray, B., Szafer, A., Dolbeare, T.A., Stevens, A., Tirrell, L., Benner, T., Caldejon, S., Dalley, R.A., Dee, N., Lau, C., Nyhus, J., Reding, M., Riley, Z.L., Sandman, D., Shen, E., van der Kouwe, A., Varjabedian, A., Write, M., Zollei, L., Dang, C., Knowles, J.A., Koch, C., Phillips, J.W., Sestan, N., Wohnoutka, P., Zielke, H.R., Hohmann, J.G., Jones, A.R., Bernard, A., Hawrylycz, M.J., Hof, P.R., Fischl, B., Lein, E.S., 2016. Comprehensive cellular-resolution atlas of the adult human brain. J. Comp. Neurol. 524 (16), 3127–3481. doi:10.1002/cne.24080.

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T., Mohlberg, H., Amunts, K., Zilles, K., 2008. Cortical folding patterns and predicting cytoarchitecture. Cereb. Cortex 18 (8), 1973–1980.

Goulas, A., Zilles, K., Hilgetag, C.C., 2018. Cortical gradients and laminar projections in mammals. Trends Neurosci. 41 (11), 775–788. doi:10.1016/j.tins.2018.06.003.

Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y., He, K., 2017. Accurate, large minibatch SGD: training imagenet in 1 h. arXiv preprint arXiv:1706.02677.

Im, K., Grant, P.E., 2019. Sulcal pits and patterns in developing human brains. Neuroimage 185, 881–890.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456.

Kiwitz, K., Schiffer, C., Dickscheid, T., Amunts, K., 2019. Reference delineations of area hOc1 (V1, 17, CalcS) in individual sections of the BigBrain [Data set]. EBRAINS doi:10.25493/3GSV-T4A.

Kiwitz, K., Schiffer, C., Dickscheid, T., Amunts, K., 2019. Reference delineations of area hOc2 (V2, 18) in individual sections of the BigBrain [Data set]. EBRAINS doi:10.25493/8MKD-D77.

Kiwitz, K., Schiffer, C., Dickscheid, T., Amunts, K., 2020. Reference delineations of area hOc3v (LingG) in individual sections of the BigBrain [Data set]. EBRAINS doi:10.25493/DBZG-QZ.

Kiwitz, K., Schiffer, C., Dickscheid, T., Malikovic, A., Amunts, K., 2020. Reference delineations of area hOc5 (LOC) in individual sections of the BigBrain [Data set]. EBRAINS doi:10.25493/4027-K9Y.

Krause, D., Thörnig, P., 2018. JURECA: Modular supercomputer at Jülich supercomputing centre. J. Large-Scale Res. Facil. JLSRF 4, A132. doi:10.17815/jlsrf-4-121-1.

Lebenberg, J., Labit, M., Auzias, G., Mohlberg, H., Fischer, C., Rivière, D., Duchesnay, E., Kabdebon, C., Leroy, F., Labra, N., Poupon, F., Dickscheid, T., Hertz-Pannier, L., Poupon, C., Dehaene-Lambertz, G., Hüppi, P.S., Amunts, K., Dubois, J., Mangin, J.-F., 2018. A framework based on sulcal constraints to align preterm, infant and adult human brain images acquired in vivo and post mortem. Brain Struct. Funct. 223 (9), 4153–4168. doi:10.1007/s00429-018-1735-9.

Leprince, Y., Poupon, F., Delzescaux, T., Hasboun, D., Poupon, C., Rivière, D., 2015. Combined Laplacian-equivolumic model for studying cortical lamination with ultra high field MRI (7 T). In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 580–583. doi:10.1109/ISBI.2015.7163940.

Lewiner, T., Lopes, H., Vieira, A.W., Tavares, G., 2003. Efficient implementation of marching cubes' cases with topological guarantees. J. Graphics Tools 8 (2), 1–15. doi:10.1080/10867651.2003.10487582.

Lewis, L., Lepage, C., Fournier, M., Zilles, K., Amunts, K., Evans, A.C., 2014. BigBrain: initial tissue classification and surface extraction. 20th Annual Meeting of the Organization for Human Brain Mapping (OHBM).

Malikovic, A., Amunts, K., Schleicher, A., Mohlberg, H., Eickhoff, S.B., Wilms, M., Palomero-Gallagher, N., Armstrong, E., Zilles, K., 2007. Cytoarchitectonic analysis of the human extrastriate cortex in the region of V5/MT+: aprobabilistic, stereotaxic map of area hOc5. Cereb. Cortex 17 (3), 562–574. doi:10.1093/cercor/bhj181.

Merker, B., 1983. Silver staining of cell bodies by means of physical development. J. Neurosci. Methods 9 (3), 235–241. doi:10.1016/0165-0270(83)90086-9.

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). IEEE, pp. 565–571.

Omidyeganeh, M., Lepage, C., Wagstyl, K., Spitzer, H., Dickscheid, T., Amunts, K., Evans, A., 2020. Non-linear registration of 1 μm histology sections into 3D 20 μm BigBrain space. 26th Annual Meeting of the Organization for Human Brain Mapping.

Richter, M., Amunts, K., Mohlberg, H., Bludau, S., Eickhoff, S.B., Zilles, K., Caspers, S., 2019. Cytoarchitectonic segregation of human posterior intraparietal and adjacent parieto-occipital sulcus and its relation to visuomotor and cognitive functions. Cereb. Cortex 29 (3), 1305–1327. doi:10.1093/cercor/bhy245.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Rottschy, C., Eickhoff, S.B., Schleicher, A., Mohlberg, H., Kujovic, M., Zilles, K., Amunts, K., 2007. Ventral visual cortex in humans: cytoarchitectonic mapping of two extrastriate areas. Hum. Brain Mapp. 28 (10), 1045–1059. doi:10.1002/hbm.20348.

Ruan, J., Bludau, S., Palomero-Gallagher, N., Caspers, S., Mohlberg, H., Eickhoff, S.B., Seitz, R.J., Amunts, K., 2018. Cytoarchitecture, probability maps, and functions of the human supplementary and pre-supplementary motor areas. Brain Struct. Funct. 223 (9), 4169–4186. doi:10.1007/s00429-018-1738-6.

Schleicher, A., Amunts, K., Geyer, S., Kowalski, T., Schormann, T., Palomero-Gallagher, N., Zilles, K., 2000. A stereological approach to human cortical architecture: identification and delineation of cortical areas. J. Chem. Neuroanat. 20, 31–47. doi:10.1016/S0891-0618(00)00076-4.

Schleicher, A., Amunts, K., Geyer, S., Morosan, P., Zilles, K., 1999. Observer-independent method for microstructural parcellation of cerebral cortex: a quantitative approach to cytoarchitectonics. Neuroimage 9 (1), 165–177. doi:10.1006/nimg.1998.0385.

Schmitt, O., Böhme, M., 2002. A robust transcortical profile scanner for generating 2-D traverses in histological sections of richly curved cortical courses. Neuroimage 16 (4), 1103–1119. doi:10.1006/nimg.2002.1159.

Schober, M., Axer, M., Huysegoms, M., Schubert, N., Amunts, K., Dickscheid, T., 2016. Morphing image masks for stacked histological sections using Laplace's equation. In: Tolxdorff, T., Deserno, T.M., Handels, H., Meinzer, H.-P. (Eds.), Bildverarbeitung Für Die Medizin 2016. Springer, pp. 146–151. doi:10.1007/978-3-662-49465-3_27.

Sergeev, A., Del Balso, M., 2018. Horovod: fast and easy distributed deep learning in TensorFlow. arXiv preprint arXiv:1802.05799.

Settles, B., 2009. Active Learning Literature Survey. Computer Sciences Technical Report. University of Wisconsin–Madison.

Sigl, B., 2018. Zytoarchitektur, Netzwerke und Funktionen der Areale des menschlichen dorsolateralen prämotorischen Kortex - Komponenten motorischer Planung und Kandidat für das Frontale Augenfeld Dissertation.

Sigl, B., Bludau, S., Mohlberg, H., Eickhoff, S.B., Amunts, K., 2019a. Interpolated 3D map of area 6d1 (PreCG) in the BigBrain [Data set].. EBRAINS doi:10.25493/K0X6-KKB.

Sigl, B., Bludau, S., Mohlberg, H., Eickhoff, S.B., Amunts, K., 2019b. Interpolated 3D map of area 6d2 (PreCG) in the BigBrain [Data set].. EBRAINS doi:10.25493/8WSN-JQ8.

Sigl, B., Bludau, S., Mohlberg, H., Eickhoff, S.B., Amunts, K., 2019c. Interpolated 3D map of area 6d3 (PreCG) in the BigBrain [Data set].. EBRAINS doi:10.25493/B87N-ZDX.

Spitzer, H., Amunts, K., Harmeling, S., Dickscheid, T., 2017. Parcellation of visual cortex on high-resolution histological brain sections using convolutional neural networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, pp. 920–923. doi:10.1109/ISBI.2017.7950666.

Spitzer, H., Kiwitz, K., Amunts, K., Harmeling, S., Dickscheid, T., 2018. Improving cytoarchitectonic segmentation of human brain areas with self-supervised siamese networks. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Springer International Publishing, pp. 663–671.

Sutskever, I., Martens, J., Dahl, G., Hinton, G., 2013. On the importance of initialization and momentum in deep learning. In: International Conference on Machine Learning, pp. 1139–1147.

Unger, N., Bludau, S., Mohlberg, H., Caspers, S., Amunts, K., 2020a. Probabilistic cytoarchitectonic map of area OP5 (Frontal Operculum) (v2.0) [Data set]. EBRAINS doi:10.25493/TWFH-BJZ.

Unger, N., Bludau, S., Mohlberg, H., Caspers, S., Amunts, K., 2020b. Probabilistic cytoarchitectonic map of area OP6 (Frontal Operculum) (v2.0) [Data set]. EBRAINS doi:10.25493/41KE-8HT.

Unger, N., Bludau, S., Mohlberg, H., Caspers, S., Amunts, K., 2020c. Probabilistic cytoarchitectonic map of area OP7 (Frontal Operculum) (v2.0) [Data set]. EBRAINS doi:10.25493/T2M3-2ST.

Wagstyl, K., Larocque, S., Cucurull, G., Lepage, C., Cohen, J.P., Bludau, S., Palomero–Gallagher, N., Lewis, L.B., Funck, T., Spitzer, H., 2020. BigBrain 3D atlas of cortical layers: cortical and laminar thickness gradients diverge in sensory and motor cortices. PLoS Biol. 18 (4), e3000678.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bull. 1 (6), 80–83. doi:10.2307/3001968.

Wree, A., Schleicher, A., Zilles, K., 1982. Estimation of volume fractions in nervous tissue with an image analyzer. J. Neurosci. Methods 6 (1), 29–43. doi:10.1016/0165-0270(82)90014-0.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: European Conference on Computer Vision. Springer, pp. 818–833.

Zilles, K., Amunts, K., 2010. Centenary of Brodmann's map–conception and fate. Nat. Rev. Neurosci. 11 (2), 139–145. doi:10.1038/nrn2776.