Self-supervised representation learning for nerve fiber distribution patterns in 3D-PLI

Alexander Oberstrass[a,b], Sascha E.A. Muenzing[a], Meiqi Niu[a], Nicola Palomero-Gallagher[a,c], Christian Schiffer[a,b], Markus Axer[a,d], Katrin Amunts[a,c], Timo Dickscheid[a,b,e]

[a]Institute of Neuroscience and Medicine (INM- 1), Research Centre Jülich, Jülich, Germany

[b]Helmholtz AI, Research Centre Jülich, Jülich, Germany

[c]Cécile & Oskar Vogt Institute of Brain Research, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[d]Department of Physics, University of Wuppertal, Wuppertal, Germany

[e]Institute of Computer Science, Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Corresponding Author: Alexander Oberstrass (a.oberstrass@fz-juelich.de)

ABSTRACT

A comprehensive understanding of the organizational principles in the human brain requires, among other factors, well-quantifiable descriptors of nerve fiber architecture. Three-dimensional polarized light imaging (3D-PLI) is a microscopic imaging technique that enables insights into the fine-grained organization of myelinated nerve fibers with high resolution. Descriptors characterizing the fiber architecture observed in 3D-PLI would enable downstream analysis tasks such as multimodal correlation studies, clustering, and mapping. However, best practices for observer-independent characterization of fiber architecture in 3D-PLI are not yet available. To this end, we propose the application of a fully data-driven approach to characterize nerve fiber architecture in 3D-PLI images using self-supervised representation learning. We introduce a *3D-Context Contrastive Learning* (CL-3D) objective that utilizes the spatial neighborhood of texture examples across histological brain sections of a 3D reconstructed volume to sample positive pairs for contrastive learning. We combine this sampling strategy with specifically designed image augmentations to gain robustness to typical variations in 3D-PLI parameter maps. The approach is demonstrated for the 3D reconstructed occipital lobe of a vervet monkey brain. We show that extracted features are highly sensitive to different configurations of nerve fibers, yet robust to variations between consecutive brain sections arising from histological processing. We demonstrate their practical applicability for retrieving clusters of homogeneous fiber architecture, performing classification with minimal annotations and query-based retrieval of characteristic components of fiber architecture such as U-fibers.

Keywords: deep learning, contrastive learning, fiber architecture, polarized light imaging, occipital lobe, vervet monkey brain

1. INTRODUCTION

Decoding the human brain requires analyzing its structural and functional organization at different spatial scales, including cytoarchitecture and fiber architecture at microscopic resolutions (Amunts & Zilles, 2015; Axer & Amunts, 2022). Three-dimensional polarized light imaging (3D-PLI) (Axer, Graessel, et al., 2011) is an imaging technique that reveals the fine-grained configuration and 3D orientation of myelinated nerve fibers in both gray and white matter with micrometer resolution. 3D-PLI thus establishes a link between microscopic myeloarchitecture and dMRI-based structural connectivity at the

macro- and mesoscopic scale (Caspers & Axer, 2019; Zilles et al., 2016). 3D-PLI images provide detailed visual information for obtaining maps of fiber architecture at different scales. Based on 3D-PLI images, previous work demonstrated the detection of myelinated pathways and delineation of subfields in the human hippocampus (Zeineh et al., 2017) as well as the identification of fiber tracts and visual areas in the vervet monkey visual system (Takemura et al., 2020).

Polarized light imaging allows processing of whole-brain tissue sections and enables scanning of large tissue stacks (Axer, Gräßel, et al., 2020; Axer, Poupon, et al., 2020; Howard et al., 2023). However, interpretation and analysis of the complex information provided by 3D-PLI require substantial expertise that cannot scale to the vastly increasing amount of data produced by recent high-throughput devices. Moreover, *automated* large-scale analysis of fiber architecture at the resolution provided by 3D-PLI is challenging due to the complexity and high dimensionality of the data. In order to use data analysis algorithms, a suitable lower dimensional feature representation of 3D-PLI textures is needed. Ideally, features in this representation are highly expressive for different fiber configurations while being robust against other sources of variation, such as histological processing effects and the relative 3D orientation of image patches. Such features, however, are difficult to derive, and we hypothesize that an efficient representation cannot be manually engineered.

Over the last years, deep learning methods have become prevalent in analyzing images in related fields such as histopathology (de Matos et al., 2021), as they are able to learn representations from pure data. While annotations of fiber configurations in 3D-PLI are not yet available at the scale required for supervised deep learning, we do have access to large amounts of unlabeled data. Recent advances in self-supervised representation learning suggest using contrastive learning (Hadsell et al., 2006; van den Oord et al., 2018) to learn distinctive representations from unlabeled training data. The training objective here is to represent similar instances (positive pairs) as close points in the embedding space while pushing dissimilar instances (negative pairs) apart to prevent representational collapse. While other methods to prevent representational collapse have been proposed as well, such as clustering (Caron et al., 2020), distillation (X. Chen & He, 2021; Grill et al., 2020), information maximization (Zbontar et al., 2021), or variance preservation (Bardes et al., 2022), contrastive learning of visual representations by application or adaptation of SimCLR (T. Chen et al., 2020) and MoCo (He et al., 2020) is still popular in medical image analysis (X. Chen et al., 2022; Krishnan et al., 2022). Due to its simplicity, we build on

the SimCLR framework (T. Chen et al., 2020). An application of SimCLR in cytoarchitectonic brain mapping was recently performed for histological images (Schiffer et al., 2021). A main challenge they discovered was the tendency of models to focus more on anatomical landmarks than on features descriptive of cytoarchitecture when creating positive pairs based on data augmentations of the same image. To overcome this effect, they employ a supervised contrastive loss (Khosla et al., 2020) by defining positive pairs based on same labels and sample pairs within each brain area.

Several self-supervised learning methods were proposed to learn image representations based on spatial context, which can be used to create correlated views for contrastive learning (T. Chen et al., 2020; van den Oord et al., 2018; Van Gansbeke et al., 2021) or to define pretext tasks (Doersch et al., 2015; Noroozi & Favaro, 2016; Pathak et al., 2016). For microscopic imaging, predicting the geodesic distance between image patches along the brain surface (Spitzer et al., 2018) or the sequence of multiresolution histopathology images (Srinidhi et al., 2022) has been proposed as pretext tasks. Other approaches leverage the spatial continuity of images by maximizing mutual information between neighboring patches in histological images (Gildenblat & Klaiman, 2019) or satellite images (Ji et al., 2019). They assume textures in spatial proximity to be similar and, therefore, aim to contrast them with textures in more distant parts of images.

In the present study, we explore self-supervised contrastive learning for inferring descriptive features of local nerve fiber distribution patterns from raw 3D-PLI measurements. To generate positive pairs of 3D-PLI texture examples, we assume that fundamental properties of local fiber architecture are typically consistent between nearby image patches. While this assumption is likely violated at boundaries between distinct structural brain areas, we assume that it holds for the largest share of nearby image patches. In contrast to previous work (Gildenblat & Klaiman, 2019; Ji et al., 2019), instead of utilizing in-plane similarity of images, we use a 3D reconstructed histological volume to access the spatial coordinates of image patches in 3D. More precisely, we extract positive pairs of image patches at nearby coordinates across tissue sections. This sampling strategy is motivated by the idea that positive pairs from different tissue sections show independently measured tissue and thus encourage the learning of features that are robust to random variations in the measurement process not descriptive of fiber architecture. We denote this 3D-informed self-supervised learning strategy as *3D-Context Contrastive Learning* (CL-3D).

To verify the validity of the proposed approach, we compare texture representations by different methods on

the 3D reconstruction of the occipital lobe of a vervet monkey brain. We evaluate features based on their descriptive power for different fiber configurations, as well as their robustness to other sources of variation. Furthermore, we study the relationship between texture features and morphological measures at the macroscopic scale, using a precise automatic cortex segmentation which we developed specifically for 3D-PLI images based on a U-Net model (Ronneberger et al., 2015). To demonstrate the applicability of the learned CL-3D features, we show that the features form clusters that reflect different types of fiber architecture, enable classification tasks with minimal labels, and are suitable for query-based retrieval of U-fiber structures.

The main contributions of the present study are the following:

- We propose a novel *3D-Context Contrastive Learning* (CL-3D) strategy to learn a powerful feature embedding for microscopic resolution image patches from 3D-PLI.
- We present specific image augmentations for maximizing invariance of learned features with respect to typical variations in 3D-PLI images, increasing feature quality and robustness.
- We show a high sensitivity of the resulting 3D-PLI feature embeddings to fundamental configurations of nerve fibers, such as myelinated radial and tangential fibers within the cortex, fiber bundles, crossings, and fannings, as well as cortical morphology.
- Using a dataset from a vervet monkey brain, we demonstrate that the learned features are well suited for exploratory data analysis, specifically for

finding clusters of similar fiber architecture and retrieving locations with specific architectural properties based on interactively chosen examples.

2.   MATERIALS AND METHODS

2.1.   3D-PLI measurements from the occipital lobe of a vervet monkey brain

2.1.1.   Tissue samples

For this study, we use a 3D reconstruction of 234 coronal sections from the right occipital lobe of a 2.4-year-old adult male vervet monkey brain (ID 1818) measured with 3D-PLI (Takemura et al., 2020). The brain sample was obtained postmortem after flush with phosphate-buffered saline in accordance with the Wake Forest Institutional Animal Care and Use Committee (IACUC #A11-219) and conforming the AVMA Guidelines for the Euthanasia of Animals. It was perfusion fixed with 4% paraformaldehyde, immersed in 20% glycerin for cryo-protection, and frozen at -70°C. Sectioning of the frozen brain was performed coronally at 60 µm thickness using a large-scale cryostat microtome (Poly-cut CM 3500, Leica, Germany). Before each cutting step, *blockface images* (Axer, Graessel, et al., 2011) were taken as an undistorted reference for image realignment using a CCD camera (Fig. 1A).

2.1.2.   3D-PLI acquisition

For 3D-PLI measurement (Axer & Amunts, 2022; Axer, Amunts, et al., 2011; Axer, Graessel, et al., 2011), brain sections were scanned using a polarizing microscope
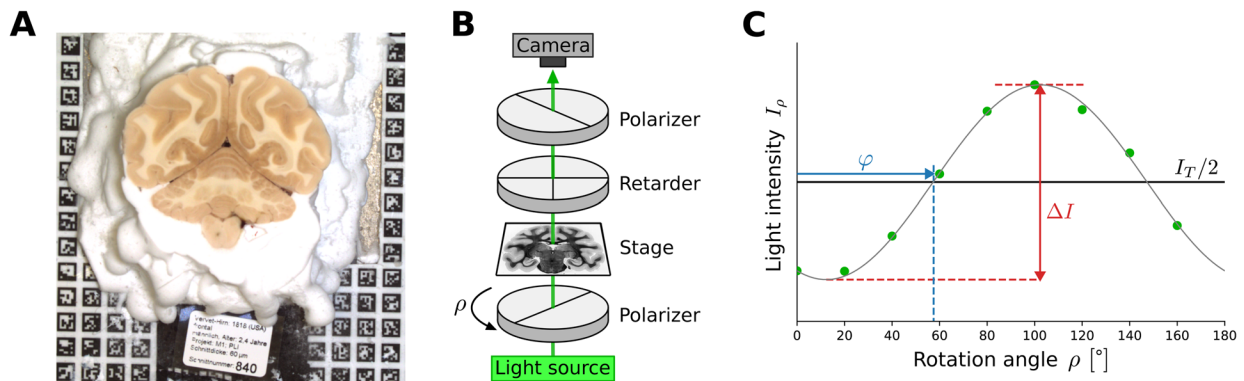

Fig. 1.   Overview of the 3D-PLI data acquisition. (A) Blockface images are taken for every section before slicing the mounted tissue block, providing a distortion-free reference for 3D volume reconstruction. ARTag markers positioned on the cryotome in the background are used for precise image alignment. (B) 3D-PLI measurement setup for the polarizing microscope (LMP-1) consisting of a coherent green light source, a rotating linear polarizer, a specimen stage, a stationary circular analyzer (quarter-wave retarder, linear polarizer), and a CCD camera to capture transmitted light intensities. (C) Example intensity profile recorded by a single pixel of the CCD camera at nine polarizer rotation angles ρ. The profile can be described by a sinusodial curve, parameterized by three modalities: transmittance I_T, direction φ, and retardation $|\sin\delta| = \Delta I / I_T$.

(LMP-1, Taorad, Germany), which provides a detailed view of nerve fiber architecture at 1.3 μm resolution. In this microscope setup, sections are placed on a stage between a rotating linear polarizer and a stationary circular analyzer consisting of a quarter-wave retarder and a second linear polarizer (Fig. 1B). The setup is illuminated by an incoherent white light LED equipped with a band-pass filter of 550 ± 5 nm half-width. Variations in transmitted light intensity are captured using a CCD camera for 9 equidistant rotation angles ρ of the rotating linear polarizer covering 180° of rotation. The recorded light intensity variations feature sinusoidal profiles at each pixel (Fig. 1C), which are determined by the spatial orientation of myelinated nerve fibers. Using Jones calculus, a physical description for these profiles can be derived as

$$I_\rho = \frac{I_T}{2} \cdot \left(1 + \sin(2\rho - 2\varphi) \cdot \sin\delta\right). \quad (1)$$

Harmonic Fourier analysis can be applied to retrieve parameter maps of transmittance I_T, retardation $|\sin\delta|$, and direction φ from the profiles (Axer, Amunts, et al., 2011). The phase shift δ between the ordinary and the extraordinary ray can be further decomposed as

$$\delta \approx 2\pi \frac{t \cdot \Delta n}{\lambda} \cos^2\alpha, \quad (2)$$

with the cumulative thickness of birefringent tissue t, birefringence Δn, the wavelength of the light source λ, and nerve fiber inclination angle α. While birefringence Δn and wavelength λ are kept constant for all pixels, the number of myelinated nerve fibers, reflected by t, varies. To resolve Equation (2) for inclination α, a transmittance-weighted model (Menzel et al., 2022) can be used to estimate t. Fiber inclination and direction information can then be jointly visualized in fiber orientation maps in HSV color space (Axer, Graessel, et al., 2011), where the hue value corresponds to in-plane fiber direction φ, while saturation and value reflect the out-of-plane fiber inclination α (both zero for vertical fibers at $\alpha = 90°$).

2.1.3. 3D registration

To access the three-dimensional context of images, registration of 3D-PLI parameter maps is performed on 234 sections of the right occipital lobe from section 841 to 1,083 (Fig. 2). Nine sections heavily deformed by histological processing are sorted out and replaced by their nearest neighbors to ensure high-quality 3D reconstruction. Before registration of 3D-PLI parameter maps, a volume reconstruction of blockface images for the complete brain is performed to serve as an undistorted reference space (Fig. 2A). We use this reference space to correct for distortions from histological processing such as shrinkage or expansion of tissue, and to anchor the occipital lobe in the whole-brain context. For reconstruction of blockface images, ARTag markers are positioned on the cryotome along with the mounted tissue block (Fig. 1A). By identification of the markers using the ARToolKitPlus library (Wagner & Schmalstieg, 2007), blockface images are aligned using affine transformations (Schober et al., 2015). Subsequently, nonlinear transformation fields are estimated for alignment of 3D-PLI transmittance maps using the blockface volume as reference, which yields a reconstruction of the overall anatomical shape and topology of the occipital lobe in the 3D-PLI volume space. The same transformation is used to align all 3D-PLI parameter maps. However, blockface images do not contain sufficient structural detail for precise alignment of fine structures visible in 3D-PLI such as single fiber bundles and small blood vessels. Therefore, the blockface alignment is used as initialization for an additional registration step between adjacent 3D-PLI sections to reconstruct coherent 3D fiber tract transitions. In this step, each section is aligned to its successor and predecessor by symmetric normalization, which combines affine and deformable transformation, maximizing cross-correlation between joint retardation and transmittance images. The registration is performed iteratively forward and backward through the stack of sections, with the first and last sections remaining fixed. All registrations are performed using the *ELAS-TIX* (Klein et al., 2010; Shamonin et al., 2014), *ANTs* (Avants et al., 2010, 2011), and *ITK* (McCormick et al., 2014) software packages. The computed transformation fields are used to warp all 3D-PLI parameter maps into a common volume space (Fig. 2C) with a resolution of 31,077 × 28,722 × 243 voxels and a voxel size of 1.3 × 1.3 × 60 μm³. In-plane orientation information reflected by direction maps φ is preserved by adjusting the 2D rotation components at each pixel estimated by the curl of the transformation field.

2.2. Data augmentations for 3D-PLI

Data augmentations are crucial for increasing the diversity of training data. Self-supervised contrastive learning methods in particular rely on augmentations to learn representations that are more generalizable and robust to variations in the input data (T. Chen et al., 2020). It is crucial that augmentation schemes model the expected variability adequately. In microscopy, for example, similar tissue can exhibit different orientations or intensities between scans. Since mounted brain sections are not perfectly flat, sharpness variation can occur within a scan, resulting in slightly out-of-focus areas.
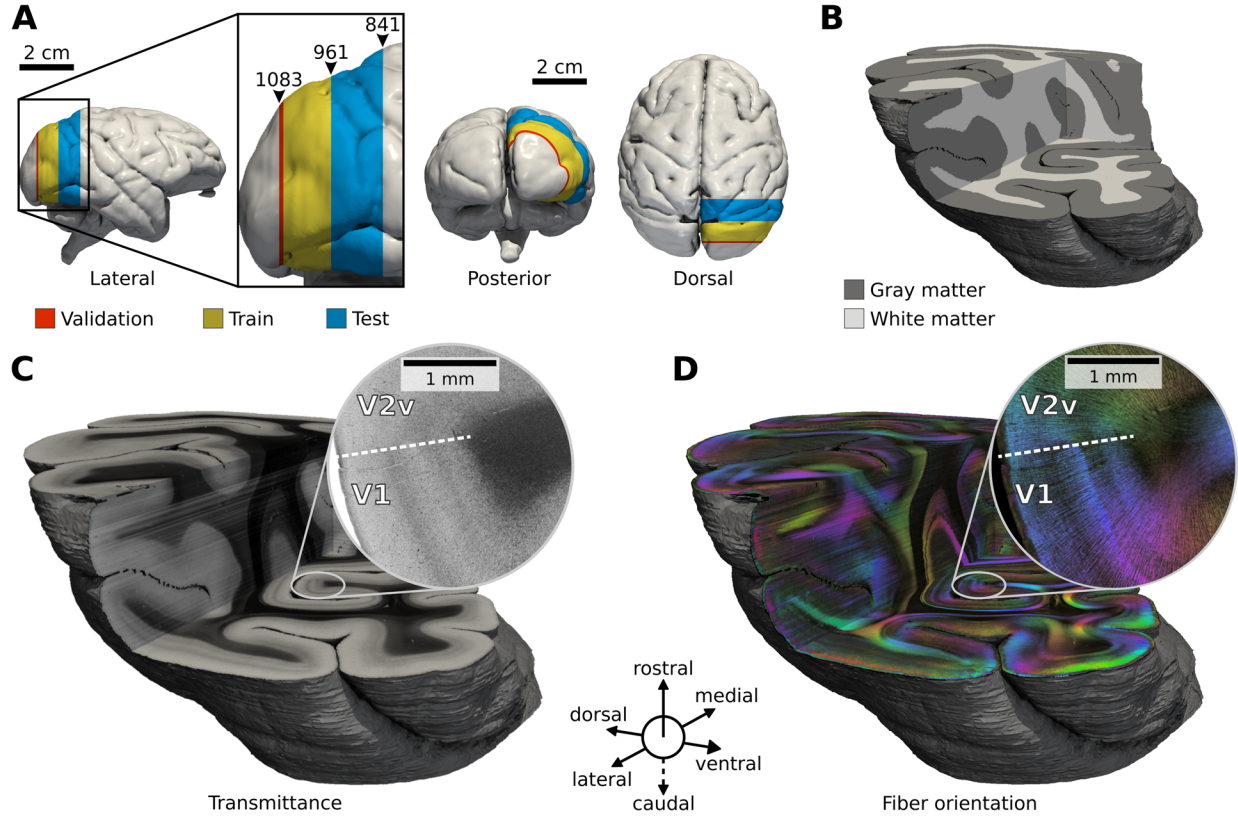
Fig. 2. 3D reconstructed occipital lobe of the right hemisphere of a vervet monkey brain measured with 3D-PLI. (A) Localization of the occipital lobe on the surface of the 3D blockface reconstruction. Sections used for training (yellow), validation (red), and testing (blue) are color coded. Numbers indicate section numbers. (B) 3D volume rendering of segmented cortical gray matter and white matter of the lobe. (C) 3D volume rendering for transmittance maps I_T, and (D) fiber orientation in HSV color space (hue: direction φ; saturation, brightness: inclination α). Zoom-ins highlight the fiber architecture at the border between primary visual cortex (V1) and secondary visual cortex (V2). All volumes are masked at the pial boundary shown as a gray surface.

In this section, we introduce a set of augmentations specifically designed to reflect typical variations in 3D-PLI images. We scale attenuation and thickness parameters in the physical model of the measured 3D-PLI signal, perform geometric affine and flip transformations, as well as Gaussian blur. All augmentations are performed for joint transmittance, direction, and retardation parameter maps. Figure 3 shows example applications of the derived augmentations.

2.2.1. Modulation of signal parameters

Parameters of the tissue in the physical model of 3D-PLI might vary (e.g., transparency, thickness) depending on postmortem time, tissue processing, or storage time of the mounted sections. Here, we provide transformations that can be implemented into 3D-PLI-specific data augmentations that approximate typical variations in the parameters.

Attenuation coefficient. The transmitted light intensity of the tissue can vary across image acquisitions due to a

change in light attenuation. Assuming uniform attenuation for simplicity, transmittance I_T can be described by Bouguer–Lambert's law as

$$I_T = I_0 e^{-t\mu} \qquad (3)$$

for the intensity of incident light I_0, section thickness t, and attenuation coefficient μ. Scaling the attenuation coefficient by linear scaling factor γ_a as $\mu' = \gamma_a \mu$ results in a scaled transmittance

$$I_T' = I_0 e^{-t\mu'} = I_0 \left(\frac{I_T}{I_0} \right)^{\gamma_a}. \qquad (4)$$

Note that this equation is only an approximation of a real change in μ due to the simplifying assumption of uniform attenuation.

Section thickness. Although the section thickness was held constant throughout all brain sections used in this study, it might vary between data acquisitions of other samples. To reflect a linear change in thickness parameter
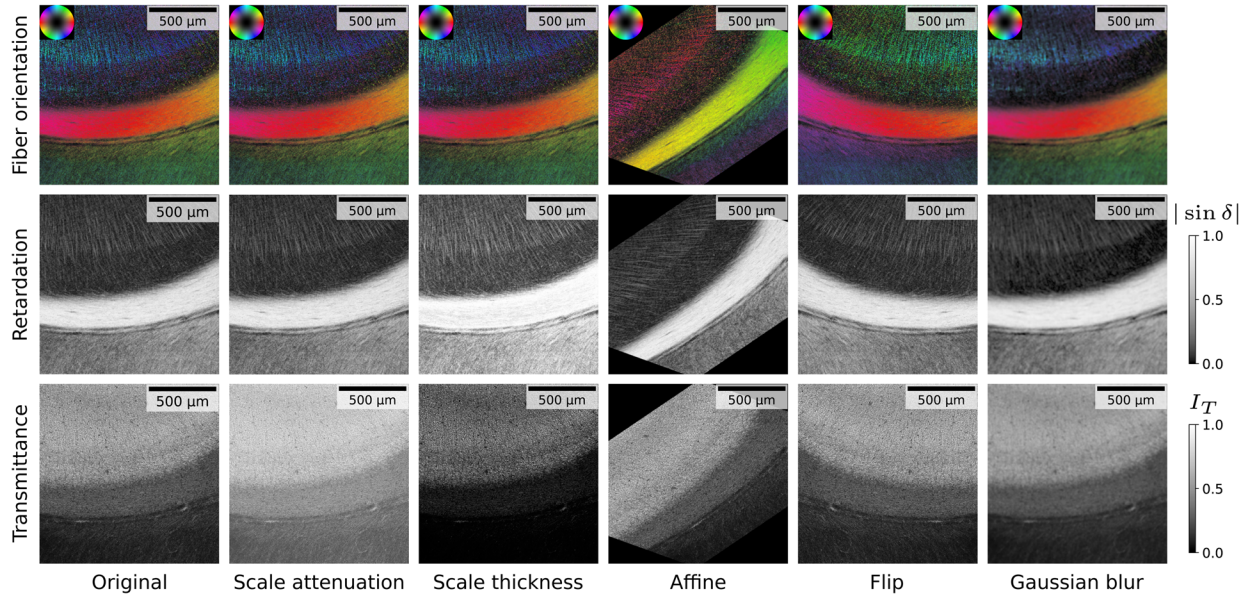
Fig. 3. Illustration of implemented 3D-PLI data augmentations for an example patch from the calcarine sulcus. Images show transmittance I_T, retardation $|\sin\delta|$, and fiber orientation in HSV color space (hue: direction φ; saturation, brightness: inclination α). The colormap for retardation is scaled with a gamma correction for visibility. Parameters for the transmittance weighted model to compute fiber orientations are kept constant for all augmentations.

t in Equation (2), we scale phase retardation $\delta \propto t$ by a linear scaling factor γ_t. For retardation $r = |\sin(\delta)|$ and $\delta' = \gamma_t\delta$ we obtain a scaled retardation

$$r' = \sin(\gamma_t \arcsin(r)). \tag{5}$$

To adjust the light transmittance I_T, we compute scaled I_T' analog to Equation (4) for scaled thickness $t' = \gamma_t t$ as follows:

$$I_T' = I_0 e^{-t'\mu} = I_0\left(\frac{I_T}{I_0}\right)^{\gamma_t}. \tag{6}$$

Note that this augmentation is also only an approximation to a real change in thickness t, as it does not add or remove tissue components from the measurement.

2.2.2. Resampling

Many image transformations require resampling of image intensity values. For 3D-PLI, resampling of the measured intensities from Equation (1) can be performed as

$$I_\rho' = \sum_i w_i I_{\rho,i} \tag{7}$$

by a weighted mean of intensity values $I_{\rho,i}$ with corresponding weights w_i, where $\sum_i w_i = 1$. With 3D-PLI parameter maps, however, we work with derivations of the originally measured image intensities and cannot

directly resample values for retardation $r = |\sin\delta|$ and direction φ. By representing Equation (7) as Fourier series and due to the linearity of the Fourier transformation, resampling of the 3D-PLI parameter maps can be performed through

$$I_T' = \sum_i w_i I_{T,i} \tag{8}$$

$$r' \cdot e^{i2\varphi'} = \frac{1}{I_T'}\sum_i w_i r_i I_{T,i} \cdot e^{i2\varphi_i} \tag{9}$$

by computing I_T' via Equation (8) before obtaining r' and φ' from Equation (9) via decomposition of the right-hand side into magnitude and phase, which correspond to r' and $2\varphi'$, respectively. The equations are used for all geometric transformations and filters, such as affine transformations or Gaussian blur, that require resampling of 3D-PLI parameter maps. We use geometric transformations to account for different orientations or distortion of tissue and Gaussian blur to mimic slightly out-of-focus areas.

2.2.3. Direction correction

Since 3D-PLI measures the absolute in-plane orientation of nerve fibers, any transformation that changes the geometry of image pixels requires a subsequent correction of direction values. For applications in diffusion MRI, Preservation of Principal Directions (PPD) (Alexander

et al., 2001a, 2001b) was introduced to preserve directional information undergoing nonrigid transformations. While proposed for 3D diffusion tensors, a similar correction mechanism can be introduced for 3D-PLI, where we restrict the transformations to in-plane transformations for simplicity. We convert direction angles φ to cartesian coordinates normalized to one as

$$\vec{d} = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} \tag{10}$$

in order to generate corrected direction angles φ′ via

$$\overline{d'} = J_f \vec{d}$$
$$\varphi' = \text{atan2}\left(d_2', d_1'\right) \tag{11}$$

for nonlinear image transformation function $f : \mathbb{R}^2 \to \mathbb{R}^2$, which maps pixel coordinates in the source domain to coordinates in the target domain, and Jacobi matrix J_f of function f. The correction is applied before application of function f to transform the image. For specific transformations, Equation (11) can be simplified to more convenient forms.

Rotation. For an example of counter-clockwise rotation by arbitrary angle θ, Equation (11) can be simplified to

$$\varphi' = \varphi + \theta. \tag{12}$$

Affine transform. For pixel coordinates $\vec{p} = \left[x, y\right]^T$ and an affine transformation composed of translation vector $\vec{t}$ and matrix A, the transformation function is given as

$$f(x, y) = A\vec{p} + \vec{t}. \tag{13}$$

If inserted into Equation (11), a simplified correction mechanism for the affine transformation can be derived as

$$\vec{d} = A\vec{d}$$
$$\varphi' = \text{atan2}\left(d_2', d_1'\right). \tag{14}$$

2.3. Cortex segmentation

To access brain morphology and distinguish gray and white matter locations, a U-Net model (Ronneberger et al., 2015) is trained for segmentation of pixels into background (BG), gray matter (GM), and white matter (WM) classes and applied to every 3D-PLI section of the lobe. For training the model, we create a dataset representing a large variety of textures in 3D-PLI images with minimal labeling effort by employing an active learning strategy in the annotation process. Rather than annotat-

ing complete sections, we manually select 58 square regions of interest (ROIs) of size 2,048 pixels (2.66 mm) from several sections, including sections outside the occipital lobe for a higher variety of examples. We train a U-Net model using these ROIs as inputs and apply the model to all available sections. We subsequently select new ROIs based on the most severe misclassifications in the model outputs. This process is repeated to obtain a growing dataset of 58, 119, 183, 301, and finally 369 ROIs of highly diverse patches capturing different textures across the entire brain.

It should be noted that large parts of the cortex can be segmented at acceptable quality using simple thresholding of transmittance and retardation values (Menzel et al., 2022). Challenging parts, such as an oblique cut border between gray and white matter or an intersecting pial surface within narrow sulci, form only a small fraction of the data, but have a significant impact on matching inner and outer cortical boundaries. We, therefore, apply a multiclass implementation of focal loss (Lin et al., 2017) for the training objective to increase emphasis for the model on challenging examples. We use 3D-PLI-specific augmentations, as described in Section 2.2, for training the cortex segmentation model.

Segmentations of the final model are corrected manually by removing small tissue fragments, extrapolating broken tissue, and filling holes to obtain a topologically correct cortex segmentation. As a last step, the resulting segmentations for individual 3D-PLI sections are stacked to form a segmented volume of the entire cortex in 3D (Fig. 2B).

2.4. 3D context contrastive learning

Contrastive learning aims to learn robust and descriptive representations of data samples by contrasting similar and dissimilar pairs. The goal is to learn an encoding function f that groups similar samples closely together in representation space while pushing dissimilar samples apart from each other. Assuming a reasonable measure of similarity that can be efficiently derived from the data, similar samples are generated as *positive pairs* consisting of an *anchor sample* and a *positive sample* with high similarity, while for *negative pairs* the anchor is combined with a dissimilar *negative sample*.

In this study, we derive similarity from the spatial neighborhood of image patches in a *3D-Context Contrastive Learning* objective. Given a random location p^a for the anchor sample, we obtain a positive sample from neighborhood location $p^+ = p^a + \Delta p$, where Δp is chosen based on two variants of spatial *context sampling* (Fig. 4A):

(i) **CL-2D**, where Δp is sampled on an in-plane circle with radius r from the same tissue section. Setting $r = 0$ is considered a special case, where the anchor sample is also used as positive sample and no context sampling is performed (**Same**).

(ii) **CL-3D**, where Δp is sampled on a sphere with radius r and p^+ is taken across sections by rounding the sampled coordinate to the nearest available section, but excluding the section from which the anchor sample was taken (i.e., p^a and p^+ are always located on different sections). Setting $r = 0$ is considered a special case, where p^+ refers to the nearest neighbor (**NN**) at the same in-plane coordinates as p^a, but in a random adjacent section.

For p^a, only locations with visible tissue are considered to avoid sampling positive pairs containing background only, using the segmentation masks obtained in Section 2.3. As CL-3D requires access to the spatial relationship of sampling locations between sections, we perform context sampling in the undistorted blockface 3D reference space. We utilize estimated transformation fields from performing the 3D registration in Sec-

tion 2.1 to warp locations in the blockface volume to individual 3D-PLI sections. Using patches from original 3D-PLI parameters maps instead of registered ones directly has the advantage of including additional variation between positive samples for the contrastive learning objective, such as different orientations of texture. In addition to the spatial sampling, we perform random augmentations for all samples as detailed in Section 2.2 (Fig. 4B).

For training encoder f, we build on the SimCLR contrastive learning framework (T. Chen et al., 2020) (Fig. 4C). In this specific framework, N augmented positive pairs $(\tilde{x}_i, \tilde{x}_j)$ are randomly sampled for each training step and stored in minibatches of $2N$ total examples $\{\tilde{x}_k\}$. We refer to the set containing indices (i, j) for all N positive pairs in the minibatch as Ω. For each positive pair, all $2(N-1)$ other random samples $\{\tilde{x}_k\}_{k \neq i,j}$ are considered negative samples, which originate from random sections and locations. When the training volume is large relative to sampling radius r of positive pairs, it is unlikely that a negative sample lies in the same spatial neighborhood as the positive one. Encoder f typically refers to a deep learning model, which yields network activation vectors $h_k = f(\tilde{x}_k)$ as representations. An
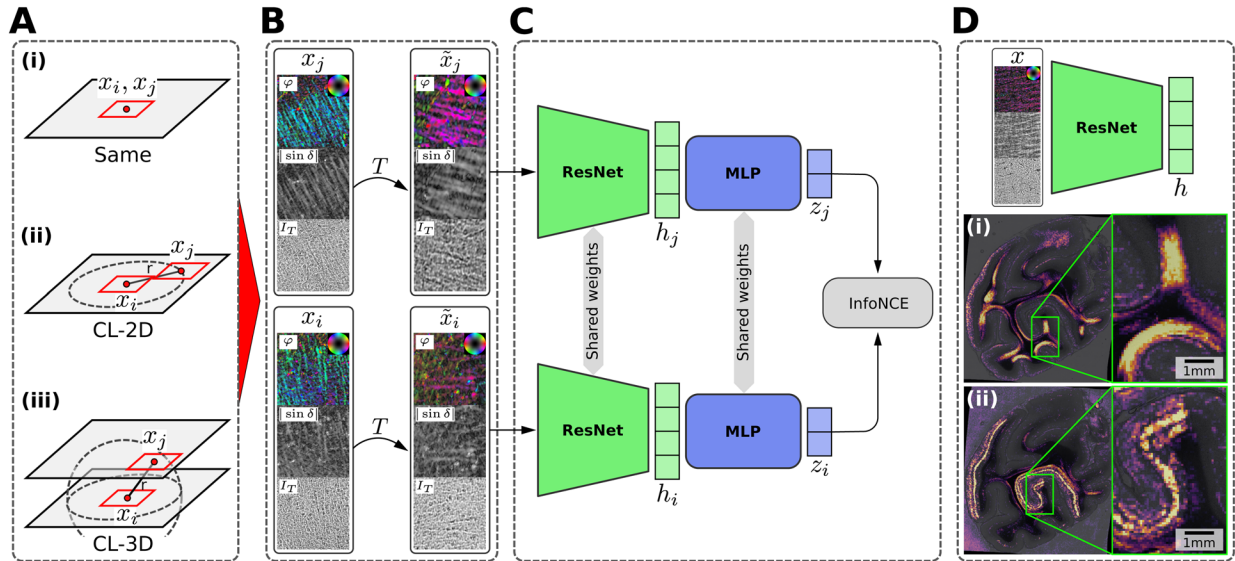

Fig. 4.   Illustration of the proposed 3D context contrastive learning scheme. (A) Context sampling performed to obtain correlated views of similar nerve fiber architecture (x_i, x_j) as (i) identical patches (Same), (ii) in-plane shifted patches on a circle with radius r (CL-2D), or (iii) patches on a sphere with radius r across sections (CL-3D). (B) Data augmentations T for 3D-PLI are randomly applied to sampled patches to promote learning representations that are robust to typical variations in 3D-PLI measurements. Patches are visualized as transmittance I_T, retardation $|\sin\delta|$, and fiber orientation map (FOM) in HSV color space (hue: direction φ; saturation, brightness: inclination α). (C) SimCLR contrastive learning framework (T. Chen et al., 2020) consisting of a ResNet encoder, hidden features h_i and h_j, a fully connected MLP projection head, projections z_i and z_j, and InfoNCE loss. (D) For inference, the trained encoder is applied on unaugmented patches x to extract 3D-PLI texture features h. Whole sections are converted to feature maps using a sliding window approach. Two example feature maps are shown on top of transmittance maps for reference highlighting (i) U-fibers and (ii) primary visual cortex (V1).

additional projection head g is introduced to map the activations h_k to a lower dimensional space of projections $z_k = g(h_k)$, on which contrastive loss is applied. The training objective is given in terms of the InfoNCE loss (van den Oord et al., 2018):

$$\ell_{i,j} = -\log \frac{\exp\big(S_c\big(z_i, z_j\big)/\tau\big)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp\big(S_c\big(z_i, z_k\big)/\tau\big)}, \tag{15}$$

where $z_k = g\big(f\big(\tilde{x}_k\big)\big)$ and similarity metric S_c chosen as the cosine similarity with temperature parameter τ. The former per-sample loss is accumulated into a total loss as

$$\mathcal{L} = \frac{1}{2N} \sum_{(i,j) \in \Omega} \ell_{i,j} + \ell_{j,i}. \tag{16}$$

After training, the projection head is discarded and for inference only representations $h = f(x)$ on unaugmented samples x are used (Fig. 4D).

2.5. Model training

2.5.1. Model architecture

For encoder f, we use a ResNet-50 (He et al., 2016) model with 3 input channels, removing the last fully connected layer. The original ResNet-50 encoder outputs 2,048 feature channels, which is well evaluated on natural images (Deng et al., 2009). However, training the model at full capacity on our data systematically resulted in high activations to infrequent but highly pronounced structures such as tangential cut radial fibers. Such structures made it trivial to solve the contrastive learning objectivate based on spatial similarity. Therefore, we reduce the number of features for all blocks in the ResNet-50 architecture to 1/8 to limit the encoder capacity, which results in 256-dimensional hidden representations, preventing the model from overfitting to these specific structures. We choose this dimensionality as a trade-off between preventing overfitting and maintaining reasonable model capacity, as bigger models can learn more general features (X. Chen & He, 2021). For projection head g, we use a two-layer MLP with ReLU activations, hidden feature size of 90 and outputs z of size 32. To feed 3D-PLI images to the ResNet, parameter maps transmittance I_T, direction φ, and retardation $|\sin\delta|$ are stacked as $x = (I_T, \sin\delta \cdot \cos(2\varphi), \sin\delta \cdot \sin(2\varphi))$ to the channel dimension, which resolves the cyclic nature of direction values φ. We standardize the input channels by running mean and standard deviation over the first 1,024 batches during training.

2.5.2. Implementation

We use *PyTorch* (Paszke et al., 2019), *PyTorch Lightning* (Borovec et al., 2022), and *Hydra* (Yadan, 2019) frameworks using the *Quicksetup-ai* (Mekki et al., 2022) template for building our model. Data augmentations (cf. Section 2.2) are implemented using the *Albumentations* (Buslaev et al., 2020) framework. Training is conducted using a distributed data-parallel strategy on 4 Nvidia A100 GPUs with synchronized batch normalization statistics (Ioffe & Szegedy, 2015) on the supercomputer JURECA-DC at the Jülich Supercomputing Centre (JSC) (Thörnig, 2021).

2.5.3. Data sampling

We sample square patches of size 192 pixels (253 µm) as anchor samples from random locations within the training volume, excluding background using the previously generated cortex segmentation. For each anchor sample, we take positive samples from a random location in spatial proximity, depending on the chosen definition of spatial similarity. As we sample patch locations on the fly, we do not have a fixed dataset size but define an *epoch* as the sampling of $512 \times 512 = 262{,}144$ positive pairs. Per training step, we take 512 anchor samples and positive samples and process them evenly split on the 4 GPUs.

2.5.4. Data augmentation

For all samples, we apply an affine transformation (scaling from [0.9, 1.3] on each axis, rotation from [-180°, 180°], and shearing from [-20°, 20°] on each axis) with linear interpolation and subsequent center cropping to crops of size 128 pixels (169 µm) to eliminate padding effects. Subsequently, we perform random flipping on center crops and scale relative thickness t (Eq. (4)) and the attenuation coefficient µ (Eq. (3)) each by random scaling from a logarithmic distribution with basis 2 from [-1, 1]. As a last augmentation, we perform Gaussian blur with a probability of 50% and σ from [0.0, 2.0].

2.5.5. Training

All augmented crops are fed to the encoder model and projection head in order to minimize the loss in Equation (16). We use Adam optimizer (Kingma & Ba, 2017) with a learning rate of 10^{-3}, a weight decay of 10^{-6}, and default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. For the choice of temperature parameter τ in Equation (15), we follow the optimal choice of $\tau = 0.5$ reported by T. Chen et al. (2020) when training until convergence. We apply

the same loss for training and validation. All models are trained until convergence if the validation loss does not reduce for more than 50 epochs, which takes between 195 (2D context) and 400 (3D context) epochs.

2.5.6. Inference

After training, model weights are frozen and inference is performed on complete sections using the trained ResNet encoder, discarding the projection head. Each section is converted into feature maps using a sliding window approach by dividing 3D-PLI parameter maps into tiles of size 128 pixels (169 μm) with 50% overlap. The overlap is chosen to better represent pixels at the edges of patches that would otherwise lie on the boundary between adjacent patches. We extract a 256-dimensional feature vector for each tile without applying the data augmentation used in training. Extracted feature vectors are reassembled into feature maps with reduced in-plane resolution of 84.4 μm per pixel compared with original input parameter maps, but 256 feature channels characterizing the local texture content. Compared with the section thickness of 60 μm, this makes the feature voxels approximately isotropic.

2.6. Classical texture features

The present approach enables learning of texture features specifically for 3D-PLI parameter maps. As baselines for texture analysis, we apply classical first-order histogram features (mean, variance, skewness, kurtosis, entropy), Grey-Level Co-occurrence Matrices (GLCM) (Haralick et al., 1973), Local Binary Patterns (LBP) (Ojala et al., 2002), and a combined set of all of their features. These are well-established approaches to represent textures in medical imaging such as CT, MR, PET (Scalco & Rizzo, 2017), and histopathology (de Matos et al., 2021). We compute texture features for whole sections using the same sliding window approach introduced in Section 2.5 by dividing 3D-PLI parameter maps into tiles of size 128 pixels (169 μm) with 50% overlap.

Direction maps φ represent the absolute orientation of fibers within the imaging plane. Since we aim to find texture representations that are independent of their absolute orientation, we are more interested in local patterns of φ than in their absolute values. We use the Sobel operator as a first derivative filter to highlight image edges and eliminate absolute values of φ. To filter direction values φ, we need to resolve their circular nature. Therefore, we represent direction angles φ in polar form as complex numbers

$$z = \cos(2\varphi) + i\sin(2\varphi) \tag{17}$$

and apply Sobel filtering as

$$G_x = K_x * z \text{ and } G_y = K_y * z \tag{18}$$

with convolution operator $*$ and Sobel filter kernels K_x and K_y. We aggregate the filtered images as

$$\hat{\varphi} = \left| G_x + G_y \right| / 12, \tag{19}$$

where $|\cdot|$ extracts the magnitude of complex numbers, and dividing by 12 normalizes the filtered values to [0, 1].

For histogram features, we compute normalized histograms with 128 bins for each of the parameter maps I_T, $\sin\delta$, and $\hat{\varphi}$. From the histograms, we compute mean, variance, skewness, kurtosis, and entropy as features. Features for all parameter maps are concatenated, resulting in a total of 15 histogram features.

To extract LBP features, we compute local binary patterns for each patch by dividing the angular space into eight points with multiple radii [1, 2, 3] to define the local neighborhood of texture. We compute normalized histograms with 10 bins of LBP values for each radius and parameter map and concatenate them into a feature vector with 90 features.

We compute normalized and symmetric GLCMs for 32 equally spaced bins of parameter maps for distances [1, 2, 4] and angles [0, $\pi/4$, $\pi/2$, 3 $\pi/4$]. From each GLCM, we compute contrast, correlation, energy, and homogeneity as features (Haralick et al., 1973). We concatenate the features for all parameter maps and distances while averaging over the angles to make the features robust to rotations, resulting in 36 total features.

In addition, we include a comprehensive combined set of all classical texture features (Histogram, LBP, GLCM) as a baseline.

2.7. Pretrained encoder on ImageNet

In recent years, there has been increasing interest in using pretrained deep learning models as feature extractors in histopathology (de Matos et al., 2021), a domain very close to ours. Here, we observe that many recent studies analyzing histopathological images use encoders pretrained on ImageNet (Deng et al., 2009) for feature extraction and downstream analysis (Breen et al., 2024; Liu et al., 2024; Wu et al., 2024). Therefore, we complement the classical baselines with a pretrained ResNet-50 (He et al., 2016) encoder, which has been trained on images from the ImageNet dataset using the SimCLR contrastive learning objective (T. Chen et al., 2020).

As the ResNet-50 model was trained on natural RGB images, we use two types of images generated from

3D-PLI parameter maps to visualize the fiber architecture: (1) Transmittance maps, stacked in the color channel dimension, to create grayscale RGB images, and (2) fiber orientation maps (FOM), which can be directly fed to the model. In contrast to the width-reduced ResNet-50 architecture we use for CL-3D and CL-2D, this model has full capacity and produces 2,048 features for image patches of 128 pixels (169 µm) size. For the creation of feature maps, we use the same sliding window approach as in the inference of CL-2D and CL-3D (Section 2.5).

3. EXPERIMENTS AND RESULTS

We train CL-2D and CL-3D models using the data from the occipital lobe of a vervet monkey brain described in Section 2.1. We split the volume into sections for training (#962–#1077), sections for validation (#1078–#1083), and sections for testing (#851–#961) as shown in Figure 2A. We evaluate the feature representations produced by the models regarding their descriptive power for different fiber configurations, spatial consistency, and applicability for downstream tasks. In particular, we show that classification of texture into tissue classes requires less annotations based on CL-3D and CL-2D features than other methods. We compare the extent to which features of different approaches can be related to brain morphology and their robustness to variations between sections. We investigate the main factors of variation specifically for CL-3D features and demonstrate that they lend themselves to interactive data exploration and identification of nerve fiber architecture in large volumes of 3D-PLI data. All experiments are evaluated exclusively on features extracted from sections not included in training.

3.1. Linear evaluation of features with minimal labels

A common approach to assess the quality and robustness of feature representations is to perform linear evaluation (T. Chen et al., 2020; van den Oord et al., 2018) on a given classification task. For the linear evaluation protocol, a simple linear classifier is trained on top of features extracted for each data sample. Being able to perform the classification task with a simple linear model indicates a good discrimination of classes in feature space and thus high-quality features. In addition, we analyze the robustness of features in a weakly supervised setting by providing the classifier with increasing number of labeled training examples, starting with only a few per class. These examples scale with human annotation effort involved in creating the dataset. Self-supervised CL-3D and CL-2D encoders are still trained on the full amount of unlabeled training data, which does not require

human annotations and is, therefore, available in a much higher quantity.

For the classification task, we use the training data acquired for the cortex segmentation performed in Section 2.3, which segments 3D-PLI images into three classes: background (BG), gray matter (GM), and white matter (WM). We divide the annotated ROIs into 228 ROIs (from 42 sections located caudal to the central sulcus) for training and 141 ROIs (from 19 sections encompassing the prefrontal cortex) for testing. This train/test split ensures that section IDs for testing the classifier do not contain sections from the occipital pole used for training the self-supervised models and allows testing the generalizability of features across the brain. To perform classification on individual texture patches, we extract square patches of 128 pixels (169 µm) on a regular grid per ROI and assign each patch the label of the most frequent class in the segmentation mask. The resulting dataset comprises many different textures with slightly unbalanced class distributions for both training (10,696 WM, 32,613 GM, and 12,829 BG patches) and testing (7,065 WM, 19,577 GM, and 7,589 BG patches).

For the linear classifier, we use a logistic regression classifier in a one-versus-rest scheme. We compare the classification performance by computing macro F1 scores across classes, and calculate the significance of these scores by computing standard error over 50 independent fits of the classifier on random subsets of training examples.

3.1.1. Evaluation of features by different methods

We compare CL-3D and CL-2D features with classical texture features and a pretrained ResNet-50 encoder on ImageNet under the linear evaluation protocol. Additionally, we report the performance of a supervised ResNet-50 classifier as a reference, specifically trained on the classification task using the full training dataset. For this model, we used a class-weighted cross-entropy loss with Adam optimizer in the same setting as described in Section 2.5. The model was trained for 416 epochs until convergence of validation accuracy, using a random 80/20 train/validation split.

With minimal training examples, results in Figure 5 show a clear lead in classification performance by our CL-2D and CL-3D models. Both models achieve macro F1 scores of 0.94 with only 30 random samples per class to fit the linear classifier. By using 10,000 samples per class, CL-3D, CL-2D, and a combination of all classical texture features (Combined) match the F1 score of the supervised ResNet-50 model (0.97). All other methods, including each individual classical texture descriptor,
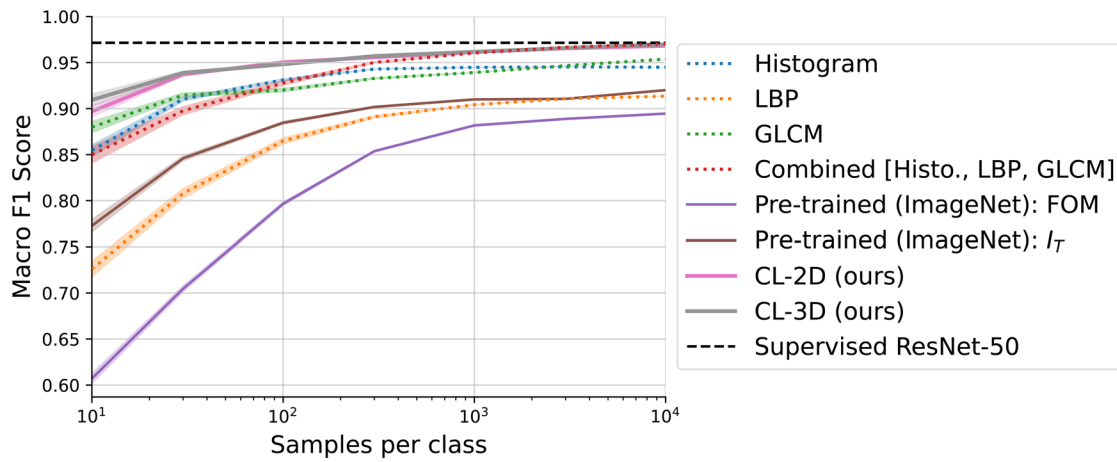
Fig. 5. Comparison of different feature extraction methods under the linear evaluation protocol. A simple linear classifier is fitted on extracted features with an increasing number of labeled samples per class to classify texture patches as gray matter, white matter, or background. With minimal samples per class provided, CL-3D and CL-2D features perform best, demonstrating highest robustness across the brain. Using the full number of samples per class, CL-3D, CL-2D, and a combination of classical texture features (Combined) all match the performance of a ResNet-50 model trained specifically on this task, indicating a high-quality feature space of these methods. Shaded areas mark standard error over 50 independent fits of the classifier on randomly selected samples.

show lower F1 scores throughout all number of samples per class.

3.1.2. Effect of data augmentations for 3D-PLI on texture features by CL-3D and CL-2D

In Section 2.2, we introduce data augmentations specifically designed for 3D-PLI parameter maps and use them in the training of CL-3D and CL-2D models. To test the effect of these augmentations on feature quality and robustness, we perform linear evaluation of CL-3D and CL-2D models trained without any augmentation, with every single augmentation, and with all augmentations combined.

As shown in Figure 6A, a CL-2D model trained without data augmentations clearly underperforms compared with models trained with any individual data augmentation, except the blur augmentation, which appears to degrade feature representations for this classification task. Among models trained with individual augmentations, color distortions (modulation of section thickness and the attenuation coefficient) yield the best results. A combination of all augmentations together performs the best overall.

Results in Figure 6B demonstrate that CL-3D benefits most from color distortions. Models trained with geometric transformations only, such as affine and flip augmentations, perform worse than a model trained without augmentations. Excluding them from the full set of augmentations, however, does not improve performance (see Appendix Fig. A1 in Appendix). Using all introduced

augmentations during training leads to the best results for CL-3D.

3.2. Main factors of variation in the learned representations

To gain insights into the main factors of variation captured by CL-3D features, we perform principal component analysis (PCA) on a random subset of 1 million voxels from the feature maps. We use the estimated principal axes to project feature channels for the entire dataset onto nine components with largest explained variance (64.2% cumulative explained variance), with at least 2.8% of variance explained per component. Explained variance refers to the amount of information in the dataset that is retained by each PCA component.

Figure 7D (see also Figure A2 in the Appendix) shows images for the first nine principal components, which in general terms reveal anatomically plausible structures. Component (1) shows a clear separation of white matter (WM) and gray matter (GM). Within GM, higher values indicate layers with a lower density of radial fibers. Within GM, values in component (2) distinguish between layers with high density of radial fibers (low values) and high density of tangential fibers (high values in the superficial layers, which contain mainly tangential fibers, and values around 0 in the Gennari stripe, which presents both radial and tangential fibers), while high values within WM highlight edges between fiber bundles. Low values in component (3) highlight layer I, which contains a high density of tangential fibers, as well as WM structures with high fiber density that
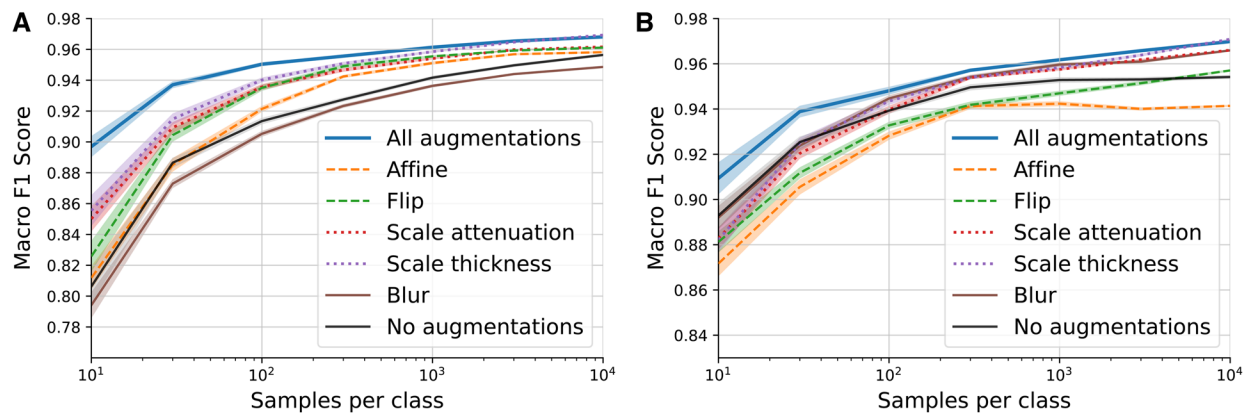
Fig. 6. Impact of different classes of data augmentations on quality and robustness of extracted features by the proposed models (A) CL-2D and (B) CL-3D. Quality and robustness are evaluated using the linear evaluation protocol under an increasing number of labeled samples per class. A simple linear classifier is fitted on extracted features to differentiate texture patches as gray matter, white matter, or background. Macro F1 scores are presented for each model across different augmentation sets. Models trained with all augmentations achieve the highest robustness and quality of features, surpassing those trained with individual augmentations or without any augmentations. Shaded areas indicate the standard error over 50 independent fits of the classifier on random training samples.

run in-plane, such as the tapetum. High values indicate tangentially sectioned radial fibers. Component (4) has low values toward out-of-plane fibers and highly oblique cortex and high values toward superficial GM layers with a low density of radial fibers and low obliqueness. Component (5) shows high values for layer IVc in the primary visual area (V1) and layer VI throughout the whole cortex. In component (6), high values show layer I throughout the cortex as well as layer IVb within V1 (Stria of Gennari), that is, they highlight GM layers with a high density of tangential fibers. High and low values in (7) mainly represent WM, with high values indicating U-fibers and other in-plane fibers and low values indicating steep fibers or crossings. The stratum sagittal (SS) has the lowest values, as here fibers emerge vertically from the plane. High values in GM mainly highlight tangentially sectioned radial fibers. High values in (8) also highlight tangentially sectioned radial fibers in GM. Additionally, they indicate abrupt twisting of flat, in-plane fibers that twist out of plane at the GM/WM transition, and show parts of U-fibers that are cut through the plane. In component (9), layers IVa, IVc, and VI within V1 are characterized by high values.

3.3. Encoding of brain morphology in texture features

Fiber architecture has mutual dependencies with cortical morphology (Striedter et al., 2015; Van Essen, 1997). To investigate to what extent different texture representations encode cortical morphology, we extract a range of morphological parameters from our test data. In particular, based on the cortex segmentation described in Section 2.3, we compute a Laplacian field between outer pial

and inner white matter surfaces using the *HighRes cortex* (Leprince et al., 2015) module included in *brainvisa* (Cointepas et al., 2001). We extract the following measures:

- *Equivolumetric cortical depth* (Bok, 1929) as the depth along cortical traverses following the gradient of the Laplacian field, compensating for the effect of cortical curvature through the divergence of the same field. It has values of 0 at the Pial and 1 at the gray–white matter surface.
- *White matter depth* defined as the shortest distance from each voxel within white matter to the interface between the cortical ribbon and white matter in millimeters.
- *Cortical curvature* as the divergence of the gradient of the Laplace field between pial and white matter surfaces (Goldman, 2005; equation 3.8 in Leprince et al., 2015)
- *Obliqueness* of the sectioning plane, computed as the absolute angle between the gradient of the Laplacian field and the sectioning plane with values in $[0°, 90°]$.

We follow the approach of Spitzer (2020) and quantify the extent to which these measures can be predicted from texture features by different methods using a linear model. Being able to predict a quantity with a simple linear model indicates a robust encoding of that quantity. We randomly select 10,000 voxels from the training set and compute both their feature representations using the trained models and the above-mentioned morphological measures. The features are standardized using Z-score normalization and used to fit a linear regression model via
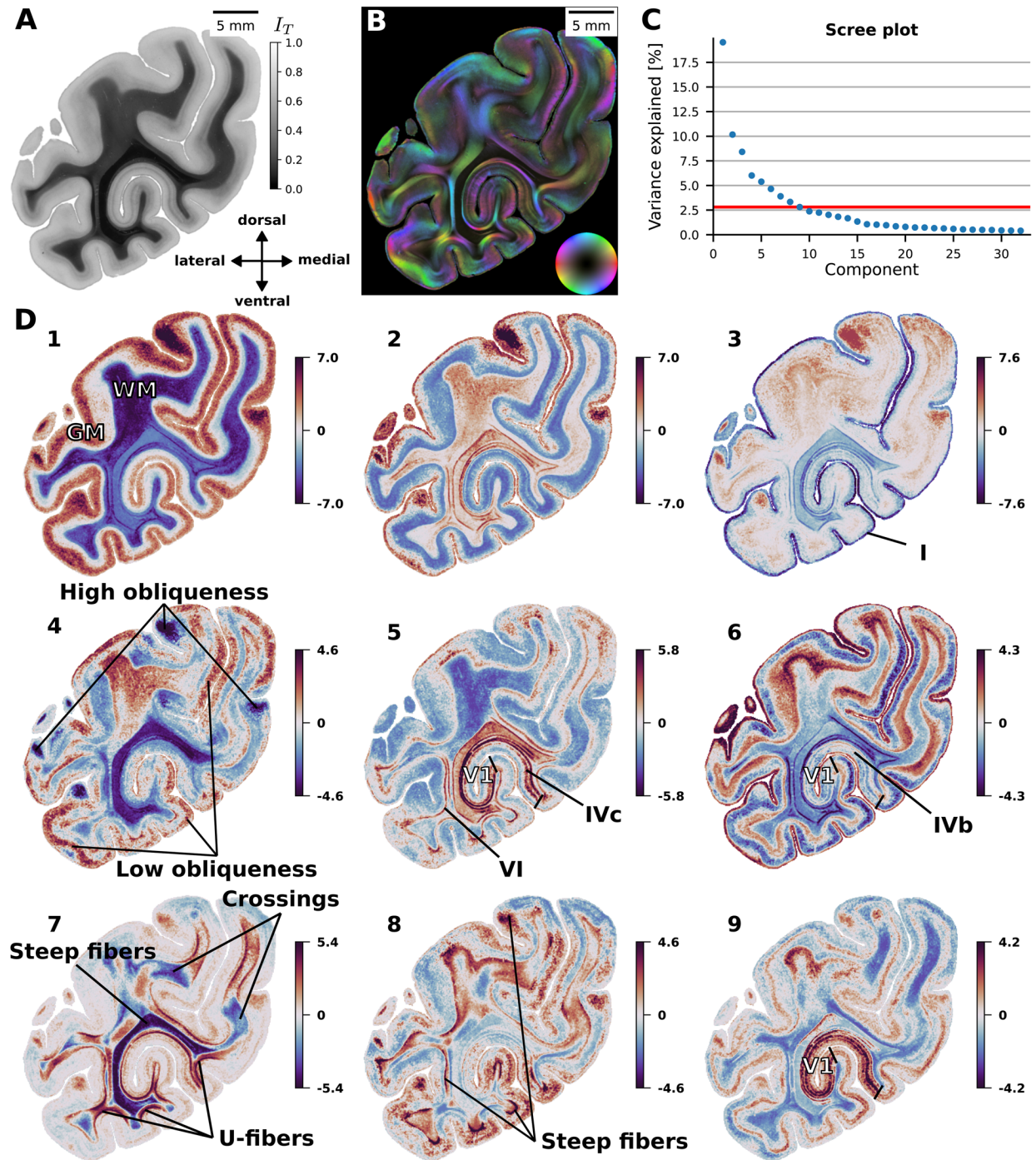
Fig. 7. Projection of CL-3D texture representations from section 898 onto the 9 PCA components with largest explained variance. (A) Transmittance and (B) fiber orientation map (FOM) for the section. (C) Scree plot showing the variance explained by the first 32 components. The horizontal red line indicates the variance explained by the 9th component of 2.8%. (D) Color-coded parameter maps of the selected PCA components, with background pixels masked as zero. The maps reveal anatomically plausible structures. GM: gray matter, WM: white matter, V1: primary visual cortex.

least squares. For the model, we use relatively high L2 regularization with a weight of 10^4 to reduce overfitting observed for the pretrained model (ImageNet) with high number of 2,048 features. The goodness of the fit is determined by calculating the coefficient of determination R^2, which denotes the proportion of variation in the measures that can be explained by the linear model from the feature representations. We compute R^2 for predicted values from 10,000 randomly selected voxels from the test set.

Our results in Table 1 show highest R^2 values for the prediction of all morphological measures from features

Table 1. Proportion of variance in morphological measures cortical depth, white matter (WM) depth, curvature, and obliqueness that can be explained by a linear model from extracted texture features.

Method	Dim.	r [μm]	Input	Cort. depth	WM depth	Curv.	Oblique.
GLCM	36	-	I_T, $\sin\delta$, $\hat{\varphi}$	0.53	0.26	0.01	0.05
Histogram	15			0.52	0.20	0.00	0.03
LBP	90			0.36	0.13	0.01	0.06
[Histo., LBP, GLCM]	141			0.63	0.27	0.02	0.08
Pretrained (ImageNet)	2048	-	I_T	0.60	0.22	0.00	0.18
			FOM	0.62	0.17	0.01	0.19
CL-2D	256	0	I_T, $\sin\delta$, φ	0.58	0.14	0.07	0.15
		118		0.79	0.34	0.13	0.43
		236		0.78	**0.36**	0.13	0.42
CL-3D	256	NN	I_T, $\sin\delta$, φ	0.79	0.35	**0.17**	0.49
		118		**0.82**	**0.36**	**0.17**	**0.52**
		236		**0.82**	0.34	0.14	0.51

To quantify the quality by which features encode each measure, we calculate the goodness of each fit using the coefficient of determination R^2. For CL-2D and CL-3D, r refers to the distance at which context sampling is performed, with nearest neighbor (NN) sampling being a special case of CL-3D. The methods use different input modalities (Input) and produce features with different dimensionalities (Dim.). Bold values indicate the highest R^2 value per column.

by our CL-3D method with medium sampling radius $r = 118$. It achieves R^2 values of 0.82 for cortical depth, 0.36 for white matter depth, 0.17 for curvature, and 0.52 for obliqueness, which are overall higher compared with CL-2D with in-plane context sampling. Using a larger or smaller radius for context sampling does not significantly increase R^2 values for CL-3D or CL-2D. While for $r = 0$, that is, using nearest neighbor (NN) context sampling for CL-3D, R^2 values are marginally smaller, they decrease considerably for CL-2D, when using no context sampling in model training. Classical texture descriptors (GLCM, Histogram, LBP) and the pretrained model show overall much smaller R^2 values and do not encode curvature at all, that is, showing values around 0.

Scatter plots shown in Figure 8 visualize predicted and true values at the example of CL-3D and combined classical texture features. As shown in Figure 8A, for obliqueness, especially larger angles can be predicted from CL-3D features, while the prediction of smaller angles in the scatter plot exhibits a deviation. Predictions of white matter depth from features by both methods show a clearer fit for smaller depths than for larger depths. As shown in Figure 8B, neither curvature nor obliqueness can be predicted from the selected classical texture features.

3.4. Clustering of learned features

3.4.1. Hierarchical clustering

We perform hierarchical cluster analysis in the embedding space of CL-3D features to evaluate the extent to which these features map characteristic nerve fiber configurations. Hierarchical clustering creates a tree-like structure of hierarchical relationships among data points

as a dendrogram based on their similarity in representational space. We choose the bottom-up approach for agglomerative clustering, which merges closest clusters based on our choice of Euclidean distance and Ward linkage. To reduce noise and computational effort, we cluster PCA-reduced feature maps from Section 3.2, where features are projected onto the first 20 components with 80.4% total explained variance. Additionally, using PCA-reduced feature maps mitigates potential negative effects of the curse of dimensionality when calculating distances in high-dimensional feature spaces. To increase the receptive field and reduce in-plane noise, each feature map is smoothed by an in-plane 2D Gaussian kernel with a standard deviation of σ = 1 which provides a good trade-off between noise reduction and keeping sensitivity to smaller structures.

We select features from foreground voxels across all test sections. Due to the huge number of 16 million data points, hierarchical clustering cannot be applied directly to the data points. Instead, we perform a two-step approach by first performing k-means clustering on a subset of 100,000 samples for 128 clusters. We use the 128 resulting cluster centers to assign all remaining 15.9 million data points to these clusters, allowing us to represent the whole test volume by superpixel-like clusters (Fig. 9A). Cluster assignments in the coronal plane appear visually smoother than the assignments across brain sections in the axial and sagittal planes. As a second step, we perform agglomerative hierarchical clustering to obtain a cluster dendrogram (Fig. 9B). From this result, we calculate silhouette scores for an increasing number of clusters to identify interesting candidate clusterings for visualization (Fig. 9C). Considering the overall
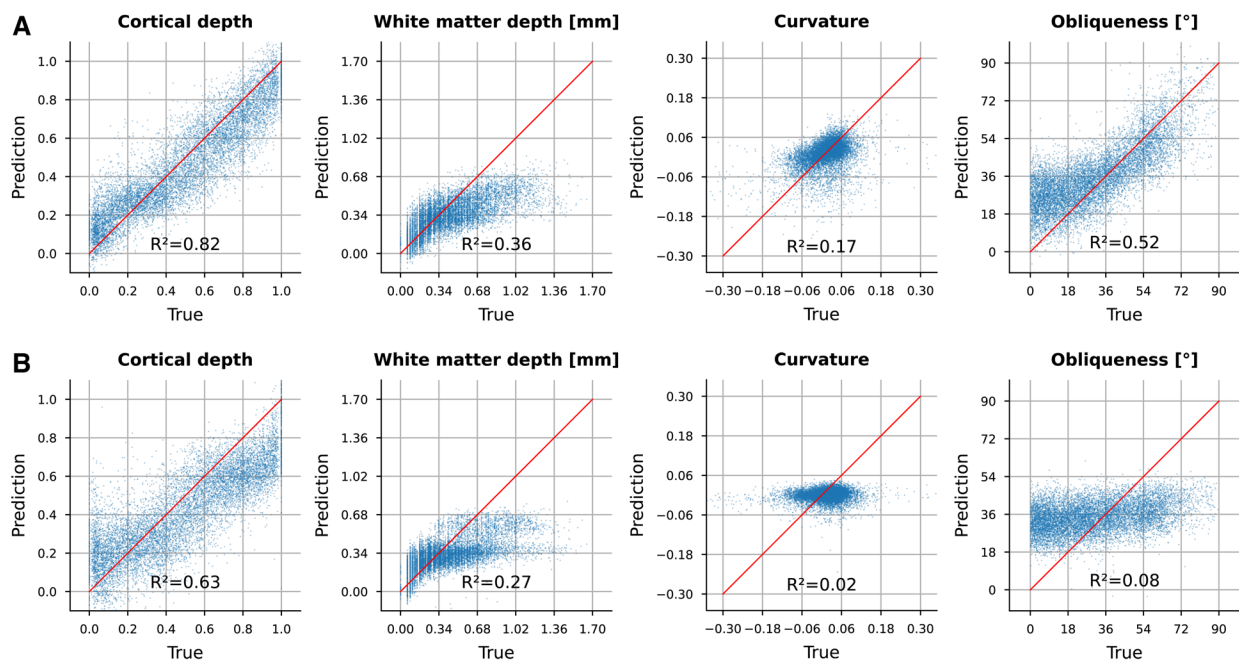
Fig. 8. Linear encoding of different morphological measures in (A) the proposed 256 dimensional CL-3D feature representations and (B) 141 dimensional combination of classical texture features (Histogram, LBP, GLCM). A linear model is fitted via least squares to predict cortical depth, white matter depth, curvature, and obliqueness. Predicted and true target measures are shown as blue scatter plots, where red lines indicate an optimal fit. Goodness of each fit is calculated by the coefficient of determination R^2.

decline in silhouette scores, we observe local maxima at 3, 7, and around 13–14 clusters. Notably, increasing from 13 to 14 clusters introduces a particularly interesting cluster, exclusively highlighting cortical layers in primary visual area (V1). Therefore, we select 14 clusters for visualization. We would like to point out that using fewer (32) or more (1,024) k-means centroids for subsequent hierarchical clustering resulted in overall lower silhouette scores and led to unspecific, over-simplifying clusters or more noisy clusters, respectively.

Results for selected candidate cluster configurations for visualization are shown in Figure 9E. We observe the same difference in spatial smoothness of cluster assignments between the coronal, axial, and sagittal planes as shown in Figure 9A, with the coronal plane appearing visually smoother. A range of characteristic aspects of fiber architecture are revealed, which are identified and confirmed by two neuroanatomists (N.P.-G. and M.N.). The descriptions are based on a comparison of each cluster with high-resolution 3D-PLI images across multiple sections and its overall distribution within the 3D geometry of the brain to ensure the consistency of descriptions.

3 clusters. Solutions for 3 clusters demonstrate a first global differentiation of the data into GM and WM. Due to its high fiber density, cortical layer VI is sometimes represented inside the WM cluster. We further observe a small cluster of tangentially cut cortex.

7 clusters. The configuration with seven clusters differentiates superficial and deep cortical layers. This segregation is shaped by the packing density of radial fibers in the deeper layers, and the tangentially running fibers close to the pial surface. For WM, voxels are split into two clusters: (1) the red cluster in Figure 9E highlights densely packed fibers with out-of-plane orientation of the sagittal stratum (SS), as well as surrounding densely packed in-plane fibers of the tapetum and (2) the green cluster in Figure 9E encompasses in-plane fibers or fibers with relatively low inclination, together with steep but less densely packed fiber bundles.

14 clusters. The configuration of 14 clusters reveals an increased sensitivity to specific WM fiber bundles and displays a cortical region delineation for the primary visual cortex (V1). Furthermore, a range of fiber architectural properties can be recognized in the maps corresponding to the clusters in Figure 9B:

- Cluster (1) shows the tapetum and stratum calcarinum (SC), characterized by approximate in-plane fibers with high packing density.
- Cluster (2) displays densely packed, highly inclined fibers and fibers of the sagittal stratum (SS) with out-of-plane orientation.
- Cluster (3) includes WM voxels from sections with an artifact-related increased light transmittance (not present in the section shown in Fig. 9E).
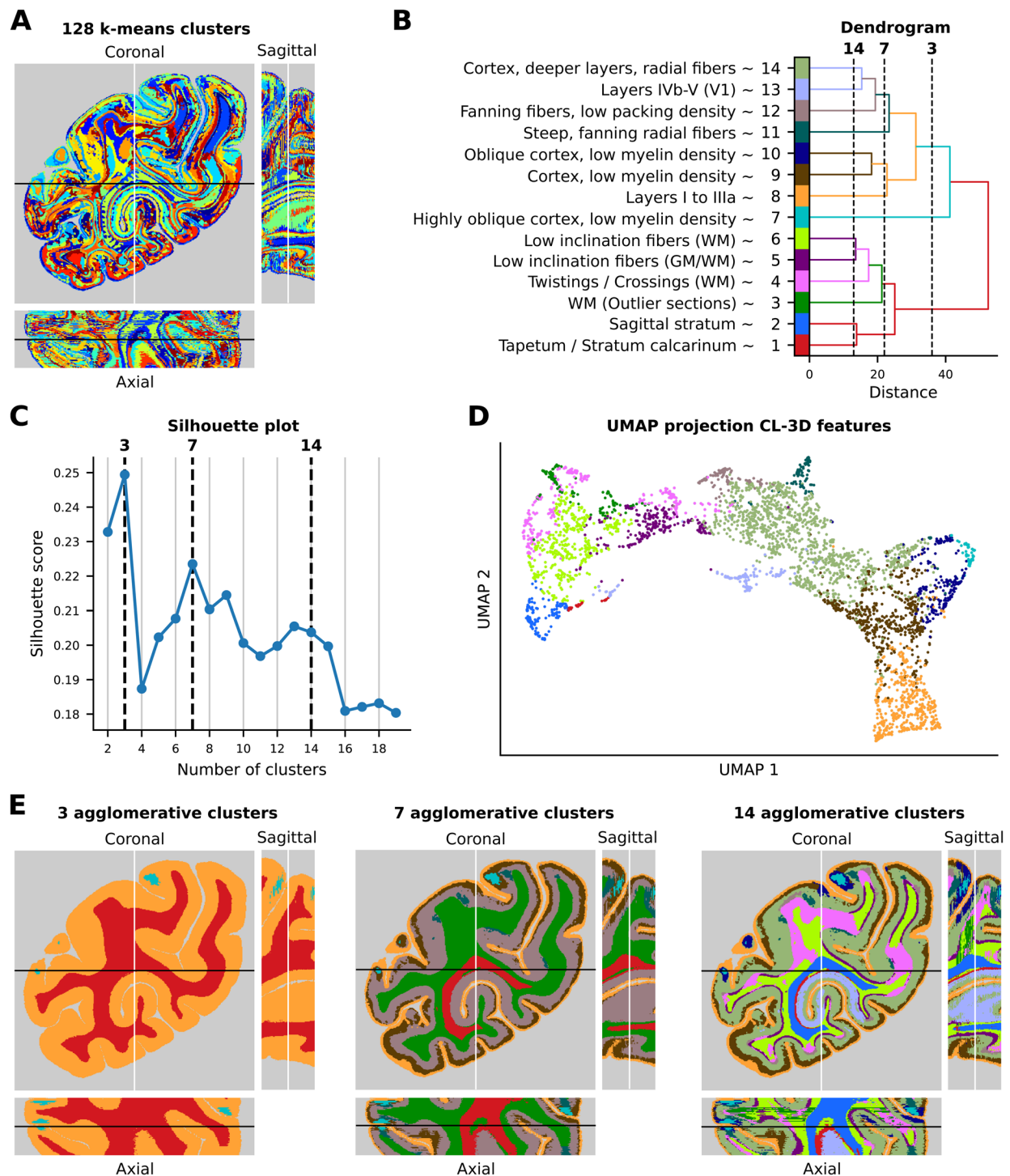
Fig. 9. Agglomerative hierarchical clustering of 128 k-means centroids shows a hierarchy of fiber architecture. (A) 128 k-means centroids forming superpixel-like clusters. (B) Dendrogram representing distances between identified clusters before merging and approximate labels naming the structures with highest overlap with each cluster. (C) Silhouette plot showing local maxima around 3, 7, and 14 clusters. (D) UMAP projection of the CL-3D features. The color of each point corresponds to the respective cluster assignment. (E) Clustering results for 3, 7, and 14 clusters.

- Cluster (4) mainly displays WM voxels, but in some cortical segments also encompasses layer VI. WM covered by these voxels is characterized by fibers with high packing density, small inclination angles, and twisting or crossing patterns.
- Cluster (5) mainly highlights layer VI of cortical segments complementing cluster 4. WM voxels encompass fibers with low packing density and low inclination at the border between GM/WM.
- Cluster (6) reveals only WM voxels with very small fiber inclination mostly parallel configurations, with only a few crossings.
- Cluster (7) highlights voxels located in highly oblique cortex with layers characterized by low myelination. When not tangentially sectioned, this portion of the cortex is encompassed by cluster (9).
- Cluster (8) highlights the most superficially located bands of tangential fibers, namely those of the zonal layer and the Kaes–Bechterew stripe, which are located in cytoarchitectonic layers I and IIIa, respectively (Zilles et al., 2015), and found throughout the whole cortex.
- Cluster (9) covers cortical layers with low density of myelin. The width of this cluster varies along the cortical ribbon. Some areas, where radial fibers reach almost up to layer II, are characterized by a narrow band of cluster (9), while in other areas it is broad because their radial fibers only reach into layer IIIb. The cluster disappears in the cortex of highly compressed sulci.
- Voxels in cluster (10) are also located in obliquely sectioned layers with a low myelination while being less oblique than those of cluster (7).
- Cluster (11) encompasses steep radial fibers fanning out at the apex of the gyrus.
- Cluster (12) highlights approximate in-plane fibers with low packing density, fanning out at the apex of the gyrus.
- Voxels of cluster (13) are restricted to the primary visual area (V1). They are mainly found in layers IVb–V, but, depending on packing density, sometimes also those of layer (VI). This cluster could reflect local cortical connectivity within V1.
- Cluster (14) highlights deeper layers mainly due to their density of radial fibers. Its width varies along the cortical ribbon and is inversely related to the width of cluster (9). In V1 cluster (14) is restricted to layers IIIb and IVa. In the remaining cortex, depending on the area it can reach from layer IIIa or IIIb to layer V or VI.

To visualize the organization of data points in the learned representational space, we additionally perform UMAP (McInnes et al., 2018) projection of the PCA-reduced features used for clustering on two dimensions for 4,000 example data points. Projections in Figure 9D show the organization of features in a continuous band along cortical depth starting from superficial layers (clusters 8 and 9) to deeper layers (clusters 13, 14, 5, and partially 4) until reaching WM clusters (3, 6, and partially 4). Branches to the sides highlight different degrees of obliqueness of cortical layers (i.e., clusters 7, 9, and 10). Clusters for structures such as the SS (cluster 2), tapetum/SC (cluster 1), or layers of V1 (cluster 13) form separate branches in the projected space. We observe splits into multiple fragments by clusters 3, 4, and 13 in UMAP space. These splits can be partially explained by the features providing a more fine-grained distinction between patches from different cortical and white matter depths than captured by the clustering results. While increasing the number of clusters in hierarchical clustering reveals these additional details of fiber architecture, we limit the number of clusters to 14 in our qualitative analysis to maintain interpretability. Beyond this point, the complexity of clusterings gradually increases, and the content of individual clusters becomes more difficult to describe.

3.4.2.　Consistency of cluster assignments across sections

In addition to 2D cluster maps, we assemble 3D renderings of the configuration with 14 clusters (Fig. 10). While CL-3D features were generated based only on in-plane texture information without cross-section constraints, the volume renderings reveal consistent cluster boundaries across sections, as can be observed by the smooth cluster shapes in the cross-sections. Cluster 13 stands out in being more noisy than other clusters.

To better quantify the cross-section consistency of feature clusters, we compute the intersection over union (IoU) of cluster assignments between adjacent sections for different number of k-means clusters (2, 8, 32, and 128). Since absolute coordinates are not included in the features, this analysis provides insights into the robustness of representations to intersection variations arising from histological processing.

IoU scores reported in Table 2 decrease overall as the number of clusters increases. Cluster assignments based on CL-3D features are significantly more consistent across all number of clusters compared with the other methods. When using nearest neighbor (NN) context sampling for CL-3D, scores slightly decrease. IoU scores of other methods (CL-2D, GLCM, pretrained encoder on ImageNet) are relatively close to each other. Setting $r = 0$ for CL-2D, that is, not performing context sampling, stands out due to particularly low IoU values, which fall
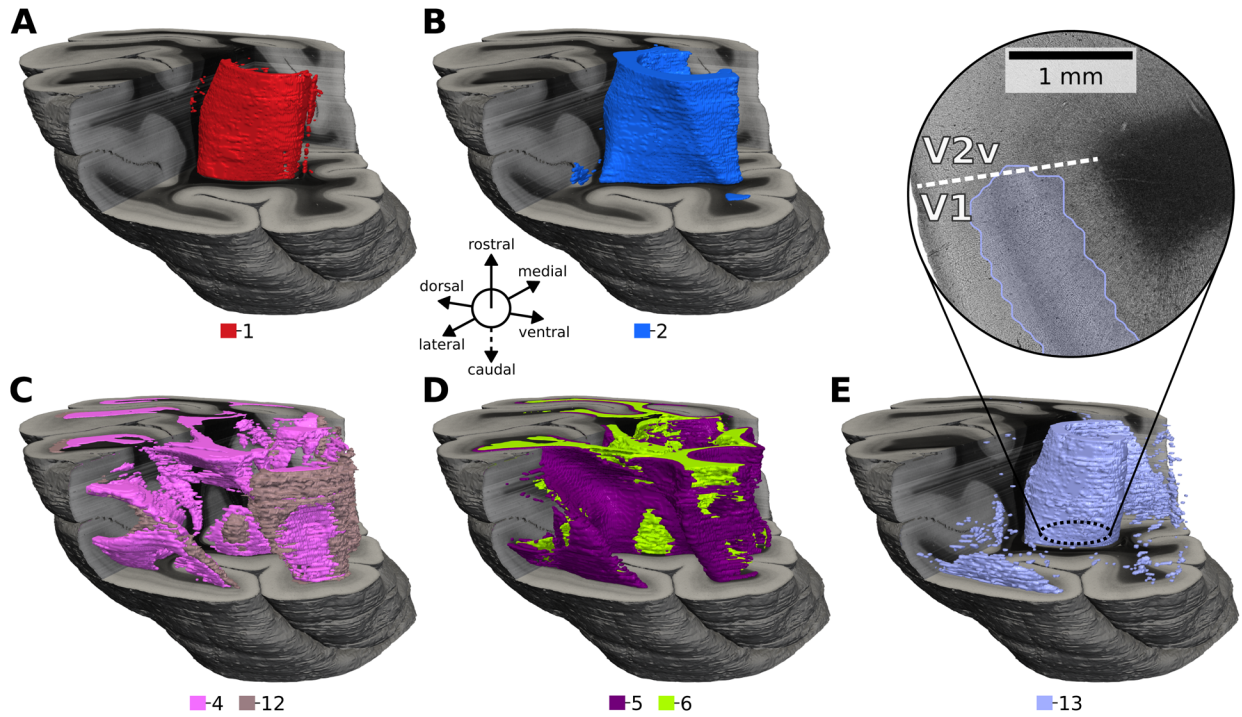
Fig. 10. Hierarchical agglomerative clustering of representations extracted by the CL-3D approach produces consistent 3D segments with anatomical relevance. (A) Approximately the tapetum and stratum calcarinum (1). (B) Contains steep fibers of the sagittal stratum (2). (C) Close to in-plane twisting or crossing fibers (4) and flat fanning fibers (12). (D) Flat fibers for layer VI (5) and WM (6). (E) Mainly layers IVb–V of primary visual area (V1), sometimes including layer VI (13). It is less clearly defined than other clusters.

Table 2. Cross-section consistency as mean IoU of k-means cluster assignments between neighboring sections based on different texture features.

Method	r [µm]	Input	2 clusters	8 clusters	32 clusters	128 clusters
GLCM	-	I_T, $\sin\delta$, $\hat{\varphi}$	95.2	58.4	34.1	19.4
Histogram			95.4	49.8	27.7	14.3
LBP			49.4	30.9	17.3	9.2
[Histo., LBP, GLCM]			92.9	42.6	24.2	14.1
Pretrained (ImageNet)	-	I_T	86.6	50.0	29.9	17.0
		FOM	77.8	50.9	35.5	23.6
CL-2D	0	I_T, $\sin\delta$, φ	88.9	47.1	25.0	12.8
	118		95.4	61.2	35.9	20.0
	236		95.0	55.8	36.7	21.0
CL-3D	NN	I_T, $\sin\delta$, φ	89.2	70.8	50.2	30.3
	118		**96.1**	70.5	**51.0**	**32.3**
	236		95.4	**71.9**	50.7	32.0

For CL-2D and CL-3D, r refers to the distance at which context sampling is performed, with nearest neighbor (NN) sampling being a special case of CL-3D. Column "Input" denotes modalities used as input for each method. Using a cross-section sampling strategy in CL-3D achieves the overall highest consistency. Bold values indicate the highest mean IoU score per column.

below those of GLCM and the pretrained encoder on ImageNet for more than two clusters.

3.5. Using CL-3D features for retrieval of common fiber orientation patterns

The proposed CL-3D method embeds image patches with similar fiber architectural properties as close points in the learned feature space. Therefore, we expect that the resulting feature representations can be used for retrieval of similar structures, given a known "prototype" image patch. To evaluate the suitability for such an application, we investigate how far U-fiber structures can be found in 3D-PLI image data from a few image examples. Although U-fibers are represented by one of the principal axes (Fig. 7D(7)), they do not appear as individual clusters

in Figure 9E. To demonstrate that the representations can still be used to identify specific nerve fiber architectures, we provide a few positive examples of U-fibers to perform a query for similar fiber configurations (Fig. 11).

Unfortunately, access to detailed, annotated data in combination with 3D-PLI measurements is limited, as reliable identification of positive and negative examples is not always possible. However, we can still search for U-fibers by selecting only a few positive examples that could be identified with confidence from 3D-PLI images by Takemura et al. (2020). For this purpose, we select up to six query points as positive examples (Fig. 11A) and compute affinity maps that show the similarity of all voxels in the feature maps to the averaged query features in
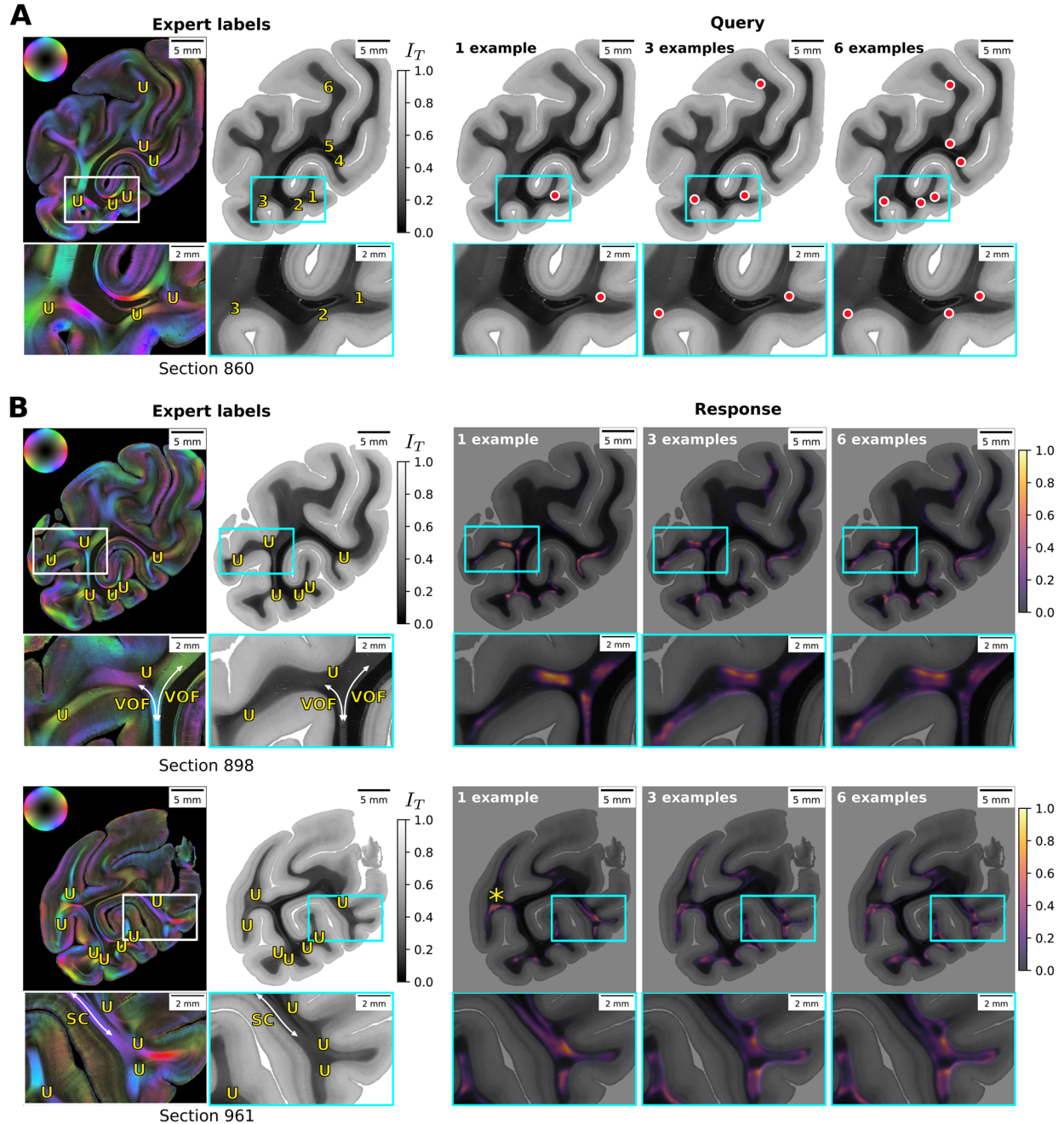
Fig. 11. CL-3D features allow retrieval of distinct nerve fiber patterns by using prototypical patches as queries and searching for nearby patches in feature space. (A) A section from the test set with query points that identify U-fibers. Yellow "U" labels in fiber orientation maps (left) and numbers in transmittance images (right) show identified U-fibers. Red dots mark query points for U-fibers with increasing examples (1, 3, and 6). (B) shows responses to queries in two other sections from the test set, displaying similarities of voxels to the mean representation of example points as affinity maps. The asterisk highlights the position of a U-fiber not detected by this procedure. U: U-fibers, VOF: vertical occipital fascicle, SC: stratum calcarinum.

representational space as responses. For this experiment, we smooth PCA-reduced feature maps with 20 components (80.4% total explained variance) by a 2D Gaussian kernel analog to Section 3.4, but set σ = 2 to increase the receptive field of texture features. This allows to represent texture for larger structures such as U-fibers and improves retrieval results. We calculate the affinity between feature points using a Gaussian radial basis function (RBF) kernel with σ = 3.5.

The affinity maps reveal peak regions at locations displaying U-fiber structures (Fig. 11B). They activate to all U-fibers identified by an expert, except for one missing activation for a U-fiber, highlighted by the asterisk. In addition, the activations highlight some fiber bundles that are not labeled as U-fibers, such as the vertical occipital fascicle (VOF) or the stratum calcarinum (SC). Although the false positive activations for VOF and SC could be suppressed through more query points, they do not disappear completely.

4. DISCUSSION

4.1. CL-3D features encode fundamental aspects of fiber architecture

Feature representations extracted by the proposed CL-3D method from 3D-PLI image patches encode distinct aspects of fiber architecture. Our experiments in Section 3.4 showed that the features form hierarchical clusters that represent gray and white matter, myeloarchitectonic layer structures, fiber bundles, fiber crossings, and fiber fannings. These clusters are spatially consistent and often highlight specific characteristics of fiber architecture as locally connected structures (Fig. 9). Even without explicit clustering, the main PCA components of the feature embedding space produce maps that highlight fundamental principles of fiber architecture (Fig. 7). This indicates that proximity of features in the embedding space, which is efficient and easy to compute, serves as a suitable proxy measure for similarity of fiber architecture as captured in the corresponding image patches. This is in contrast to directly measuring similarity of 3D-PLI image patches, where strong differences can occur despite very similar fiber configurations. Thus, CL-3D features are well suited to facilitate downstream applications for 3D-PLI image analysis.

4.2. CL-3D features are robust to variations in histological processing

While encoding relevant aspects of fiber architecture, the proposed CL-3D features proved to be robust against many other sources of texture variation. We were able to observe this robustness in the 3D stacking of consecutive images with derived cluster segments from CL-3D features. The segments showed a high overlap across brain sections (Table 2), in particular, compared with the clustering of classical texture features and a pretrained encoder on ImageNet as baselines, but also with respect to CL-2D features. Clustering results by these methods were overall not consistent enough (see Appendix Fig. A3) to perform the same in-depth qualitative evaluation as performed for CL-3D in Section 3.4.1. Volume renderings in Figure 10 illustrated the consistency of clustering CL-3D features as spatially consistent 3D segments. This suggests that CL-3D promotes the learning of representations that are more robust to discrepancies between independently processed sections. CL-3D features were also found to be robust to the absolute in-plane orientation of texture, as shown by consistent laminar patterns of cluster assignments regardless of their absolute orientation in 2D (Fig. 9).

Some of the robustness of CL-3D features can be attributed to the introduced context sampling across brain sections, as shown in Table 2. CL-3D features demonstrated significantly higher overlap in cluster assignments between sections compared with models trained with in-plane (CL-2D) or without context sampling (CL-2D with $r = 0$).

Another factor contributing to the robustness may result from the data augmentations specifically designed for 3D-PLI parameter maps, introduced in Section 2.2. CL-3D and CL-2D models trained with the introduced augmentations showed increased feature quality and robustness in Section 3.1.2. A comparison of individual augmentations identified the highest benefit from using color distortions (modulation of section thickness and the attenuation coefficient). For CL-2D, using the blur augmentation alone performed worse than training a model without augmentations. This is in line with T. Chen et al. (2020), who found the highest benefit from color distortions and the least benefit from blur. For CL-3D, geometric affine and flip augmentations each demonstrated a negative effect when used individually as the only augmentation. This is surprising at first glance but could be explained by natural geometric distortions in the training pairs, which were sampled from unregistered, neighboring tissue sections. Including geometric transformations to the full set of augmentations, however, did not negatively impact CL-3D training.

Remaining inconsistencies in cluster assignments between sections as shown in the sagittal and axial planes in Figure 9A and E can be attributed to some variations between brain sections still captured in CL-3D features. For example, the "WM (Outlier sections)" cluster highlights white matter in sections with degraded

transmittance, which we regard as a histological artifact. Imprecisions in the 3D reconstruction of adjacent brain sections could also contribute to this effect. The cross-section discontinuity of cluster assignments could be mitigated by postprocessing with the same spatial smoothing of features in the axial and sagittal planes as performed for feature maps in the coronal plane.

CL-3D and CL-2D features showed sensitivity to the relative cutting angle of cortical voxels, as observed by their ability to predict measured obliqueness using a linear model (Table 1). Obliqueness is a local feature of histological images that is consistent across adjacent sections and could be exploited by the contrastive learning objective to identify nearby positive pairs. For the majority of cortical voxels, however, this effect on the features seems to be small. Only for very high obliqueness, CL-3D features formed some smaller branches in the UMAP projection in Figure 9D or isolated clusters in Figure 9E. This aligns with observations from the scatter plots (Fig. 8A) indicating that CL-3D features primarily encode obliqueness for larger cutting angles, showing limited ability to predict smaller angles. If an encoding of obliqueness in downstream applications is nevertheless not intended, a supervisory signal that combines texture patches with different cutting angles into the same label could be helpful. For unsupervised learning, treating obliqueness as a confounding variable (Dinga et al., 2020; Snoek et al., 2019) could also help to reduce its effect.

4.3.   Retrieval and mapping as possible downstream applications

Fiber architecture is expressed in highly complex textures when measured at microscopic resolution. This makes it extremely challenging to navigate and explore larger stacks of 3D-PLI images. An obvious, albeit simple, application is the search for similar local configurations of nerve fibers, given a template image patch used as a prototype for such a query. We took a search for U-fiber structures as an example (Fig. 11), where independent expert annotations could be obtained from a previous study, and were able to demonstrate the feasibility of such a retrieval task with the proposed CL-3D features.

We showed that features cluster into groups that follow certain fiber bundles such as the sagittal stratum or the tapetum (Fig. 9). While this might at first suggest the use of features for automated brain mapping tasks, the clusters did not lend themselves to a sufficiently accurate delineation of anatomical structures. This could be due to partial volume effects of patches used to represent texture, or to the smoothing performed before clustering to denoise the features. As contrastive learning focuses on

the most characteristic properties of texture to identify positive pairs, some aspects of fiber architecture might overshadow others. The clustering shown in Figure 9E, for example, did not fall into accurate GM/WM segments. This could be due to features reflecting the density of myelinated fibers more than other aspects of fiber architecture, thus including the deepest cortical layers in the WM segment, where the density of myelinated fibers is still very high. Since the degree of sharpness of the boundary between cortex and white matter constitutes a criterium for the identification of architectonically distinct areas (Niu et al., 2020), another explanation would, therefore, be that brain areas with a blurry boundary are those in which voxels of layer VI merge into the WM segment. It should be noted though that the proposed feature extraction method is not specifically designed for performing automatic brain mapping. A supervised approach for brain mapping as a downstream application can nevertheless be promising with the proposed features. Linear evaluation, as shown in Figure 5, demonstrated that a linear classifier on top of CL-3D or CL-2D features required only 30 examples per tissue class (gray matter, white matter, or background) to perform convincing classification, significantly reducing the amount of manual annotation required. A more systematic investigation into fiber architectonic mapping based on CL-3D features will be an important follow-up work of this study.

4.4.   Relationship of feature representations with brain morphology

Cortical layers are arranged along the cortical depth and have distinct characteristic architectures. Being able to regress cortical depth from texture features using a simple linear model, therefore, suggests that features robustly encode information about the layering structure of the cortex, which is important for downstream applications such as brain mapping. The high amount of variance in cortical depth that could be explained by a linear model from CL-3D and CL-2D features (Table 1) indicates that these models indeed encode layer-related textures. For CL-3D, this claim is also supported by the observation that the main PCA components highlight individual cortical layers (Fig. 7), while clustering of the features shows several clusters that group more superficial and deeper layers (Fig. 9). Furthermore, being able to regress the depth of patches within white matter (Table 1) suggests that CL-3D and CL-2D features robustly separate deep from superficial fiber bundles such as U-fibers (Decramer et al., 2018; Shinohara et al., 2020). Classical texture features, as well as a pretrained model on ImageNet, however, demonstrated significantly lower capability in predicting cortical and white matter depth. A

CL-2D model with sampling radius $r = 0$, that is, not performing context sampling, performed much worse. This highlights a positive effect of the introduced context sampling in learning expressive representations for fiber architecture in 3D-PLI.

Methods for analyzing the laminar structure of the cortex typically require representations that are robust to cortical folding (Leprince et al., 2015; Schleicher et al., 1999; Waehnert et al., 2014). While CL-3D and CL-2D learned to represent some curvature-related patterns such as for fanning radial fibers at gyral crowns (Fig. 9), predicting curvature from the features did not work well (Table 1), indicating moderate robustness of the proposed feature representations to cortical folding. For CL-3D and CL-2D, the weak encoding might be attributed to the affine transformation applied in the contrastive learning setup, which performs scaling and shearing operations on the texture examples. It should be noted that in addition to the curvature definition used in Section 3.3, other established definitions (Goldman, 2005) that have not been considered in this study might lead to different results.

5.   CONCLUSION AND OUTLOOK

Aiming to improve automatic mapping and analysis of fiber orientation distributions in the brain, we introduced a self-supervised contrastive learning scheme for extracting "deep" feature representations for 3D-PLI image patches at micrometer resolution. We specifically proposed *3D context Contrastive Learning* (CL-3D), introducing a *context sampling* strategy to sample positive pairs based on their spatial proximity across nearby brain sections. Without any anatomical prior information given during training, the feature representations extracted by CL-3D were shown to highlight fundamental patterns of fiber architecture in both gray and white matter, such as myelinated radial and tangential fibers within the cortex, fiber bundles, crossings, and fannings. At the same time, feature representations by CL-3D proved to be more robust to variations between independently measured sections, such as artifacts arising from histological processing, compared with statistical methods, an encoder pretrained on natural images, and representations by in-plane sampling of positive pairs in contrastive learning (CL-2D).

The present study opens new perspectives for automated analysis of fiber architecture in 3D-PLI. Due to the low-dimensional embedding space, CL-3D feature representations can aid in interpretation of 3D-PLI textures and improve computational efficiency of downstream 3D-PLI analysis. For example, the learned feature representations can be used to develop spatial maps of specific aspects of fiber orientation distributions, such as U-fibers, which allow comparison of fiber architecture

with other modalities linked to brain atlases. They can also be used to train discriminative models for downstream tasks such as segmenting tissue classes, cortical layers, fiber bundles, or even brain areas with minimal number of positive and negative labeled examples.

An important direction for future research will be to extend the trained models to larger training datasets. We intend to extend the approach to whole-brain datasets, possibly including multiple species. Since the main challenge for establishing training data is the precise 3D reconstruction from individual brain sections, it will be helpful to investigate how far approximate registrations can be sufficient for CL-3D. Furthermore, we plan to integrate deep 3D-PLI features into brain atlases to provide easy accessibility. For the human brain, the BigBrain (Amunts et al., 2013) model would be an ideal reference model for integration, which is already used for multimodal data integration from other imaging modalities, such as cyto- and receptor architecture.

DATA AND CODE AVAILABILITY

The software used to implement and train self-supervised CL-3D and CL-2D models is available on GitHub (https://github.com/FZJ-INM1-BDA/cl-3d). An implementation of the introduced data augmentations for 3D-PLI images (https://jugit.fz-juelich.de/inm-1/bda/software/data_processing/pli-transforms) as well as additional dependencies (https://jugit.fz-juelich.de/inm-1/bda/software) are available on GitLab.

Volumetric clustering results, PCA projections of extracted CL-3D features, measures of cortex morphology, and 3D-PLI fiber orientation and transmittance maps for reference are available on the EBRAINS data sharing platform (Oberstrass et al., 2024).

The complete set of high-resolution 3D-PLI images for the vervet monkey occipital lobe is hosted on the Jülich supercomputing facility. A subset of selected 3D-PLI images is available on EBRAINS (Axer, Gräßel, et al., 2020), with a future publication of the whole stack of images planned.

ETHICS STATEMENT

Vervet monkeys used were part of the Vervet Research Colony housed at the Wake Forest School of Medicine. Our study did not include experimental procedures with living animals. Brains were obtained when animals were sacrificed to reduce the size of the colony, where they were maintained and sacrificed in accordance with the guidelines of the Wake Forest Institutional Animal Care and Use Committee IACUC #A11-219 and the AVMA Guidelines for the Euthanasia of Animals.

AUTHOR CONTRIBUTIONS

FUNDING

DECLARATION OF COMPETING INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

Alexander, D., Pierpaoli, C., Basser, P., & Gee, J. C. (2001a). An algorithm for preservation of orientation during non-rigid warps of diffusion tensor magnetic resonance (DT-MR) images. *Proceedings of the International Society for Magnetic Resonance in Medicine*, *9*, 791. https://doi.org/10.1109/42.963816

Alexander, D., Pierpaoli, C., Basser, P., & Gee, J. (2001b). Spatial transformations of diffusion tensor magnetic resonance images. *IEEE Transactions on Medical Imaging*, *20*(11), 1131–1139. https://doi.org/10.1109/42.963816

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. (2013). BigBrain: An ultrahigh-resolution 3D human brain model. *Science*, *340*(6139), 1472–1475. https://doi.org/10.1126/science.1235381

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

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025

Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., & Gee, J. C. (2010). The optimal template effect in hippocampus studies of diseased populations. *NeuroImage*, *49*(3), 2457–2466. https://doi.org/10.1016/j.neuroimage.2009.09.062

Axer, M., & Amunts, K. (2022). Scale matters: The nested human connectome. *Science*, *378*(6619), 500–504. https://doi.org/10.1126/science.abq2599

Axer, M., Amunts, K., Grässel, D., Palm, C., Dammers, J., Axer, H., Pietrzyk, U., & Zilles, K. (2011). A novel approach to the human connectome: Ultra-high resolution mapping of fiber tracts in the brain. *NeuroImage*, *54*(2), 1091–1101. https://doi.org/10.1016/j.neuroimage.2010.08.075

Axer, M., Gräßel, D., Palomero-Gallagher, N., Takemura, H., Jorgensen, M. J., Woods, R., & Amunts, K. (2020). Images of the nerve fiber architecture at micrometer-resolution in the vervet monkey visual system [Data set]. *EBRAINS*. https://doi.org/10.25493/AFR3-KDK

Axer, M., Graessel, D., Kleiner, M., Dammers, J., Dickscheid, T., Reckfort, J., Huetz, T., Eiben, B., Pietrzyk, U., Zilles, K., & Amunts, K. (2011). High-resolution fiber tract reconstruction in the human brain by means of three-dimensional polarized light imaging. *Frontiers in Neuroinformatics*, *5*, 34. https://doi.org/10.3389/fninf.2011.00034

Axer, M., Poupon, C., & Costantini, I. (2020). Fiber structures of a human hippocampus based on joint DMRI, 3D-PLI, and TPFM acquisitions [Data set]. *Human Brain Project Neuroinformatics Platform*. https://doi.org/10.25493/JQ30-E08

Bardes, A., Ponce, J., & LeCun, Y. (2022). VICReg: Variance-invariance-covariance regularization for self-supervised learning. *International Conference on Learning Representations*. https://openreview.net/forum?id=xm6YD62D1Ub

Bok, S. T. (1929). Der Einflu\der in den Furchen und Windungen auftretenden Krümmungen der Gro\hirnrinde auf die Rindenarchitektur. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, *121*(1), 682–750. https://doi.org/10.1007/BF02864437

Borovec, J., Falcon, W., Nitta, A., Jha, A. H., otaj, Brundyn, A., Byrne, D., Raw, N., Matsumoto, S.,

Koker, T., Ko, B., Oke, A., Sundrani, S., Baruch, Clement, C., POIRET, C., Gupta, R., Aekula, H., Wälchli, A., … zlapp. (2022). Lightning-ai/lightning-bolts: Minor patch release (0.6.0.post1). *Zenodo*. https://doi.org/10 .5281/zenodo.7447212

Breen, J., Ravikumar, N., Allen, K., Zucker, K., & Orsi, N. M. (2024). Reducing histopathology slide magnification improves the accuracy and speed of ovarian cancer subtyping. *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. 1–5. https:// doi.org/10.1109/ISBI56570.2024.10635499

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information*, *11*(2), 125. https://doi.org/10.3390/info11020125

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A. (2020). Unsupervised learning of visual features by contrasting cluster assignments. *Advances in Neural Information Processing Systems*, *33*. https:// proceedings.neurips.cc/paper/2020/hash/70feb62b69f16 e0238f741fab228fec2-Abstract.html

Caspers, S., & Axer, M. (2019). Decoding the microstructural correlate of diffusion MRI. *NMR in Biomedicine*, *32*(4), e3779. https://doi.org/10.1002/nbm.3779

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *International Conference on Machine Learning*, *119*, 1597–1607. http://proceedings.mlr.press /v119/chen20j.html

Chen, X., & He, K. (2021). Exploring simple Siamese representation learning. In *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15750–15758). IEEE. https://doi.org/10 .1109/CVPR46437.2021.01549

Chen, X., Wang, X., Zhang, K., Fung, K.-M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, *79*, 102444. https://doi.org/10.1016/j.media .2022.102444

Cointepas, Y., Mangin, J.-F., Garnero, L., Poline, J.-B., & Benali, H. (2001). BrainVISA: Software platform for visualization and analysis of multi-modality brain data. *Neuroimage*, *13*(6), 98. https://doi.org/10.1016/S1053 -8119(01)91441-7

Decramer, T., Swinnen, S., van Loon, J., Janssen, P., & Theys, T. (2018). White matter tract anatomy in the rhesus monkey: A fiber dissection study. *Brain Structure and Function*, *223*(8), 3681–3688. https://doi.org/10.1007 /s00429-018-1718-x

de Matos, J., Ataky, S. T. M., de Souza Britto, A., Soares de Oliveira, L. E., & Lameiras Koerich, A. (2021). Machine learning methods for histopathological image analysis: A review. *Electronics*, *10*(5), 562. https://doi.org/10.3390 /electronics10050562

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). IEEE. https://doi .org/10.1109/CVPR.2009.5206848

Dinga, R., Schmaal, L., Penninx, B. W. J. H., Veltman, D. J., & Marquand, A. F. (2020). Controlling for effects of confounding variables on machine learning predictions. *bioRxiv*. https://doi.org/10.1101/2020.08.17.255034

Doersch, C., Gupta, A., & Efros, A. A. (2015). Unsupervised visual representation learning by context prediction. In *Proceedings of the 2015 IEEE International Conference on Computer Vision* (pp. 1422–1430). IEEE. https://doi .org/10.1109/ICCV.2015.167

Gildenblat, J., & Klaiman, E. (2019). Self-supervised similarity learning for digital pathology. *MICCAI 2019 Workshop COMPAY*. https://openreview.net/forum?id =BkxeSkO7ZB

Goldman, R. (2005). Curvature formulas for implicit curves and surfaces. *Computer Aided Geometric Design*, *22*(7), 632–658. https://doi.org/10.1016/j.cagd.2005 .06.005

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., Piot, B., Kavukcuoglu, K., Munos, R., & Valko, M. (2020). Bootstrap your own latent—A new approach to self-supervised learning. *Advances in Neural Information Processing Systems*, *33*. https://proceedings .neurips.cc/paper/2020/hash/f3ada80d5c4ee70142b 17b8192b2958e-Abstract.html

Hadsell, R., Chopra, S., & LeCun, Y. (2006). Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (pp. 1735–1742). IEEE. https://doi.org/10.1109/CVPR.2006.100

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-3*(6), 610–621. https://doi.org/10.1109/TSMC.1973 .4309314

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9729–9738). IEEE. https://doi.org/10 .1109/CVPR42600.2020.00975

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR .2016.90

Howard, A. F. D., Huszar, I. N., Smart, A., Cottaar, M., Daubney, G., Hanayik, T., Khrapitchev, A. A., Mars, R. B., Mollink, J., Scott, C., Sibson, N. R., Sallet, J., Jbabdi, S., & Miller, K. L. (2023). An open resource combining multi-contrast MRI and microscopy in the macaque brain. *Nature Communications*, *14*(1), 4320. https://doi.org/10 .1038/s41467-023-39916-1

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on International Conference on Machine Learning, PMLR*, *37*, 448–456. https://proceedings.mlr .press/v37/ioffe15.html

Ji, X., Henriques, J. F., & Vedaldi, A. (2019). Invariant information clustering for unsupervised image classification and segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9865–9874. https://openaccess.thecvf.com /content_ICCV_2019/html/Ji_Invariant_Information _Clustering_for_Unsupervised_Image_Classification_and _Segmentation_ICCV_2019_paper.html

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., & Krishnan, D. (2020). Supervised contrastive learning. *Advances in Neural Information Processing Systems*, *33*, 18661–18673. https://proceedings.neurips.cc/paper_files/paper/2020 /hash/d89a66c7c80a29b1bdbab0f2a1a94af8-Abstract .html

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR*. https://doi.org/10.48550 /arXiv.1412.6980

Klein, S., Staring, M., Murphy, K., Viergever, M. A., & Pluim, J. P. W. (2010). Elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, *29*(1), 196–205. https://doi.org/10.1109/TMI.2009.2035616

Krishnan, R., Rajpurkar, P., & Topol, E. J. (2022). Self-supervised learning in medicine and healthcare. *Nature Biomedical Engineering*, *6*(12), 1346–1352. https://doi.org/10.1038/s41551-022-00914-1

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). IEEE. https://doi.org/10.1109/ISBI.2015.7163940

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2980–2988. https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html

Liu, M., Liu, Y., Xu, P., & Ma, J. (2024). Unleashing the infinity power of geometry: A Novel Geometry-Aware Transformer (GOAT) for whole slide histopathology image analysis. *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, 1–5. https://doi.org/10.1109/ISBI56570.2024.10635269

McCormick, M., Liu, X., Ibanez, L., Jomier, J., & Marion, C. (2014). ITK: Enabling reproducible research and open science. *Frontiers in Neuroinformatics*, *8*, 13. https://doi.org/10.3389/fninf.2014.00013

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform manifold approximation and projection. *Journal of Open Source Software*, *3*(29), 861. https://doi.org/10.21105/joss.00861

Mekki, I., Vivar, G., Subramanian, H., & Merdivan, E. (2022). Quicksetup-ai. *Zenodo*. https://doi.org/10.5281/zenodo.10044608

Menzel, M., Reuter, J. A., Gräßel, D., Costantini, I., Amunts, K., & Axer, M. (2022). Automated computation of nerve fibre inclinations from 3D polarised light imaging measurements of brain tissue. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-08140-0

Niu, M., Impieri, D., Rapan, L., Funck, T., Palomero-Gallagher, N., & Zilles, K. (2020). Receptor-driven, multimodal mapping of cortical areas in the macaque monkey intraparietal sulcus (T. E. Behrens & W. Vanduffel, Eds.). *eLife*, *9*, e55979. https://doi.org/10.7554/eLife.55979

Noroozi, M., & Favaro, P. (2016). Unsupervised learning of visual representations by Solving Jigsaw Puzzles. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision—ECCV 2016* (pp. 69–84). Springer International Publishing. https://doi.org/10.1007/978-3-319-46466-4_5

Oberstrass, A., Muenzing, S. E. A., Palomero-Gallagher, N., Schiffer, C., Jorgensen, M. J., Woods, R., Axer, M., Amunts, K., & Dickscheid, T. (2024). Deep texture features characterizing fiber architecture in the vervet monkey occipital lobe (v1) [Data set]. *EBRAINS*. https://doi.org/10.25493/78A4-KTU

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987. https://doi.org/10.1109/TPAMI.2002.1017623

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: feature learning by inpainting. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2536–2544). IEEE. https://doi.org/10.1109/CVPR.2016.278

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Scalco, E., & Rizzo, G. (2017). Texture analysis of medical images for radiotherapy applications. *The British Journal of Radiology*, *90*(1070), 20160642. https://doi.org/10.1259/bjr.20160642

Schiffer, C., Amunts, K., Harmeling, S., & Dickscheid, T. (2021). Contrastive representation learning for whole brain cytoarchitectonic mapping in histological human brain sections. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)* (pp. 603–606). IEEE. https://doi.org/10.1109/ISBI48211.2021.9433986

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. https://doi.org/10.1006/nimg.1998.0385

Schober, M., Schlömer, P., Cremer, M., Mohlberg, H., Huynh, A.-M., Schubert, N., Kirlangic, M. E., & Amunts, K. (2015). Reference volume generation for subsequent 3D reconstruction of histological sections. In H. Handels, T. M. Deserno, H.-P. Meinzer, & T. Tolxdorff (Eds.), *Bildverarbeitung für die Medizin 2015* (pp. 143–148). Springer. https://doi.org/10.1007/978-3-662-46224-9_26

Shamonin, D., Bron, E., Lelieveldt, B., Smits, M., Klein, S., & Staring, M. (2014). Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Frontiers in Neuroinformatics*, *7*, 50. https://doi.org/10.3389/fninf.2013.00050

Shinohara, H., Liu, X., Nakajima, R., Kinoshita, M., Ozaki, N., Hori, O., & Nakada, M. (2020). Pyramid-shape crossings and intercrossing fibers are key elements for construction of the neural network in the superficial white matter of the human cerebrum. *Cerebral Cortex*, *30*(10), 5218–5228. https://doi.org/10.1093/cercor/bhaa080

Snoek, L., Miletić, S., & Scholte, H. S. (2019). How to control for confounds in decoding analyses of neuroimaging data. *NeuroImage*, *184*, 741–760. https://doi.org/10.1016/j.neuroimage.2018.09.074

Spitzer, H. (2020). *Automatic Analysis of Cortical Areas in Whole Brain Histological Sections using Convolutional Neural Networks* [Thesis]. HHU Düsseldorf. https://nbn-resolving.org/urn/resolver.pl?urn=urn:nbn:de:hbz:061-20200706-083633-7

Spitzer, H., Kiwitz, K., Amunts, K., Harmeling, S., & Dickscheid, T. (2018). Improving cytoarchitectonic segmentation of human brain areas with self-supervised Siamese networks. In A. F. Frangi, J. A. Schnabel, C.

Davatzikos, C. Alberola-López, & G. Fichtinger (Eds.), *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018* (pp. 663–671). Springer International Publishing. https://doi.org/10.1007/978-3-030-00931-1_76

Srinidhi, C. L., Kim, S. W., Chen, F.-D., & Martel, A. L. (2022). Self-supervised driven consistency training for annotation efficient histopathology image analysis. *Medical Image Analysis*, *75*, 102256. https://doi.org/10.1016/j.media.2021.102256

Striedter, G. F., Srinivasan, S., & Monuki, E. S. (2015). Cortical folding: When, where, how, and why? *Annual Review of Neuroscience*, *38*, 291–307. https://doi.org/10.1146/annurev-neuro-071714-034128

Takemura, H., Palomero-Gallagher, N., Axer, M., Gräßel, D., Jorgensen, M. J., Woods, R., & Zilles, K. (2020). Anatomy of nerve fiber bundles at micrometer-resolution in the vervet monkey visual system (T. Verstynen, T. E. Behrens, & T. Verstynen, Eds.). *eLife*, *9*, e55444. https://doi.org/10.7554/eLife.55444

Thörnig, P. (2021). JURECA: Data centric and booster modules implementing the modular supercomputing architecture at Jülich Supercomputing Centre. *Journal of Large-Scale Research Facilities (JLSRF)*, *7*, A182–A182. https://doi.org/10.17815/jlsrf-7-182

van den Oord, A., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* arXiv. https://doi.org/10.48550/arXiv.1807.03748

Van Essen, D. C. (1997). A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature*, *385*(6614), 313–318. https://doi.org/10.1038/385313a0

Van Gansbeke, W., Vandenhende, S., Georgoulis, S., & Van Gool, L. (2021). Revisiting contrastive methods for unsupervised learning of visual representations. *Advances in Neural Information Processing Systems*, *34*, 16238–16250. https://proceedings.neurips.cc/paper_files/paper/2021/hash/8757150decbd89b0f5442ca3db4d0e0e-Abstract.html

Waehnert, M. D., Dinse, J., Weiss, M., Streicher, M. N., Waehnert, P., Geyer, S., Turner, R., & Bazin, P.-L. (2014). Anatomically motivated modeling of cortical laminae. *NeuroImage*, *93*, 210–220. https://doi.org/10.1016/j.neuroimage.2013.03.078

Wagner, D., & Schmalstieg, D. (2007). ARToolKitPlus for pose tracking on mobile devices. *Computer Vision Winter Workshop: CVWW 2007*, 139–146.

Wu, H., Brémond-Martin, C., Bouaou, K., & Clouchoux, C. (2024). Tumor segmentation on whole slide images: Training or prompting? *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, 1–5. https://doi.org/10.1109/ISBI56570.2024.10635239

Yadan, O. (2019). *Hydra—A framework for elegantly configuring complex applications*. https://github.com/facebookresearch/hydra

Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S. (2021). Barlow twins: Self-supervised learning via redundancy reduction. *Proceedings of the 38th International Conference on Machine Learning*, 12310–12320. https://proceedings.mlr.press/v139/zbontar21a.html

Zeineh, M. M., Palomero-Gallagher, N., Axer, M., Gräßel, D., Goubran, M., Wree, A., Woods, R., Amunts, K., & Zilles, K. (2017). Direct visualization and mapping of the spatial course of fiber tracts at microscopic resolution in the human hippocampus. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(3), 1779–1794. https://doi.org/10.1093/cercor/bhw010

Zilles, K., Palomero-Gallagher, N., & Amunts, K. (2015). Myeloarchitecture and maps of the cerebral cortex. In A. Toga (Ed.), *Brain mapping: An encyclopedic reference* (pp. 137–156). Academic Press. https://doi.org/10.1016/B978-0-12-397025-1.00209-8

Zilles, K., Palomero-Gallagher, N., Gräßel, D., Schlömer, P., Cremer, M., Woods, R., Amunts, K., & Axer, M. (2016). Chapter 18—High-resolution fiber and fiber tract imaging using polarized light microscopy in the human, monkey, rat, and mouse brain. In K. S. Rockland (Ed.), *Axons and brain architecture* (pp. 369–389). Academic Press. https://doi.org/10.1016/B978-0-12-801393-9.00018-9
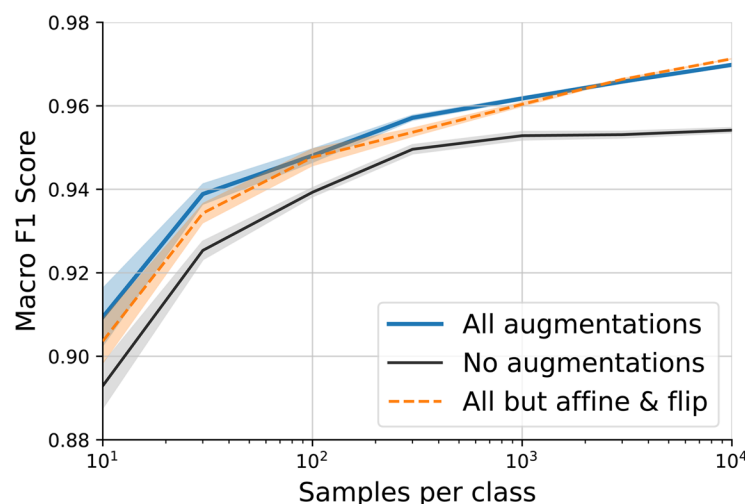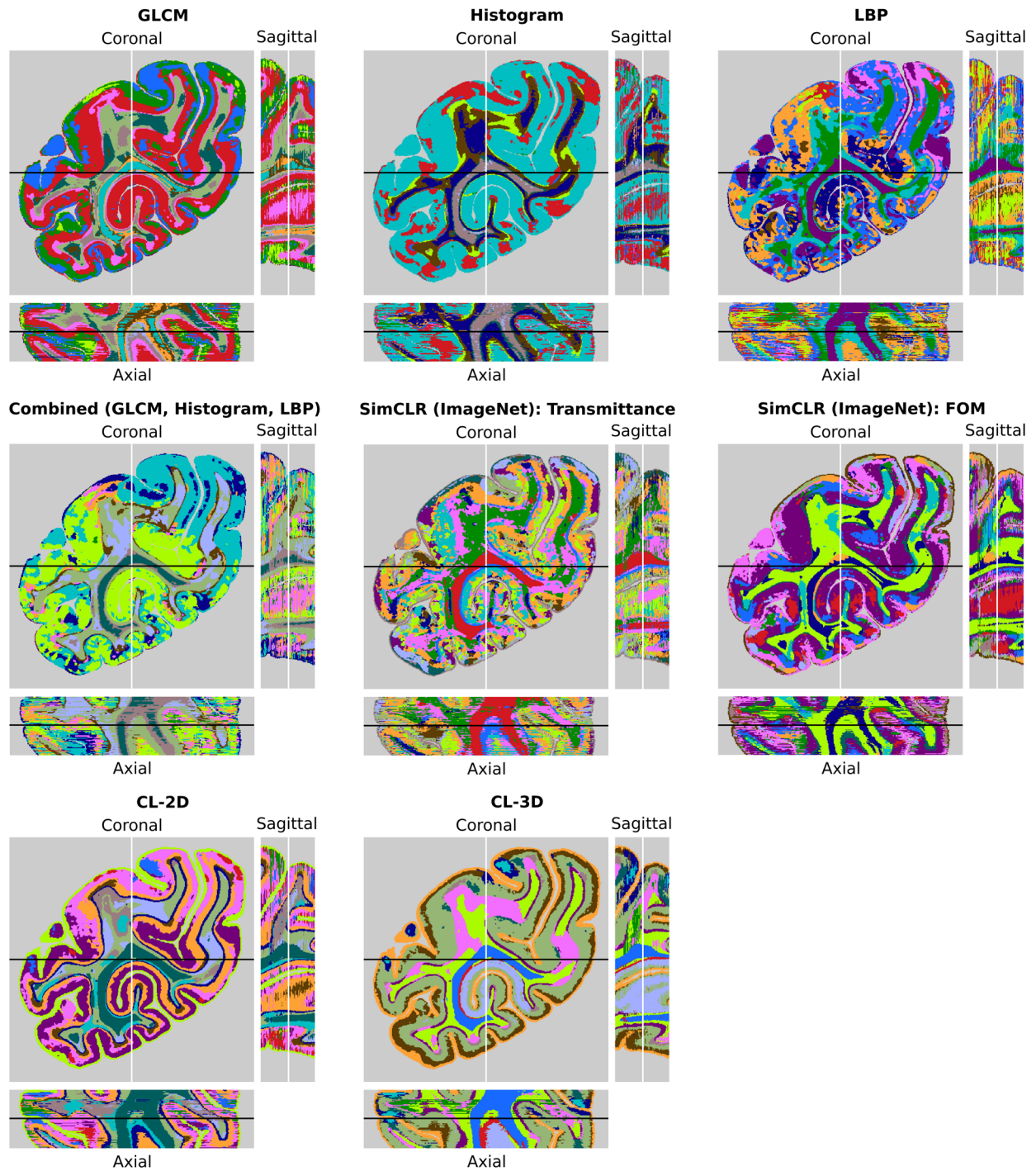
APPENDIX

Appendix Fig. A1. Excluding geometric affine and flip data augmentations from CL-3D training does not improve model performance. Quality and robustness of features are evaluated using the linear evaluation protocol under an increasing number of labeled samples per class. A simple linear classifier is fitted on extracted features to differentiate texture patches as gray matter, white matter, or background. Macro F1 scores are presented for each model across different augmentation sets. Shaded areas indicate the standard error over 50 independent runs with random training samples.
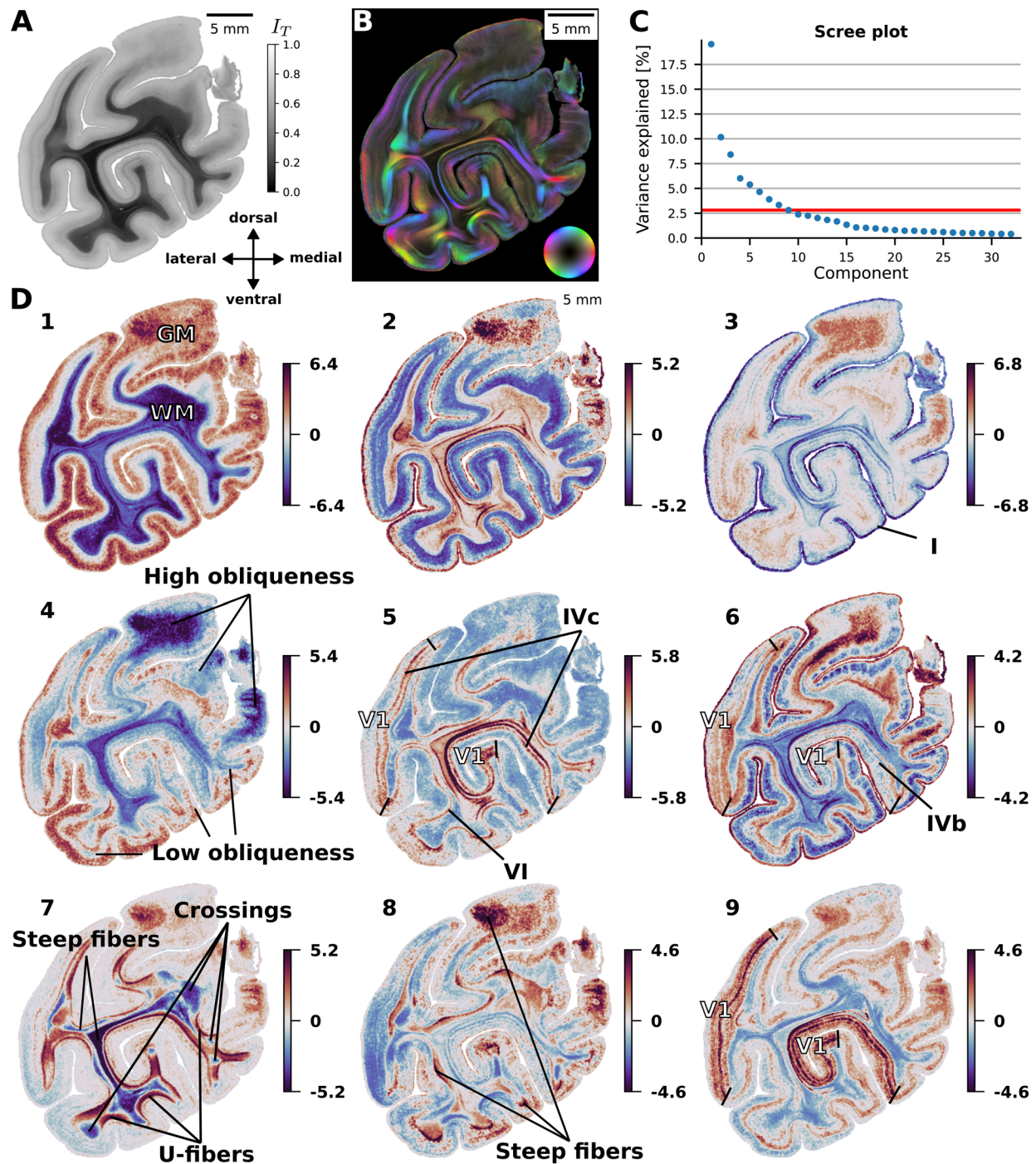
Appendix Fig. A2.   Projection of CL-3D features from section 961 onto the 9 PCA components with largest explained variance. (A) Transmittance and (B) fiber orientation maps (FOM). (C) Scree plot showing the variance explained by the first 32 components. The horizontal red line indicates the variance explained by the 9th component of 2.8%. (D) Color-coded parameter maps of the selected PCA components, with background pixels masked as zero. The maps reveal anatomically plausible structures. Structures identified by each PCA component in this section are consistent with section 898 (Fig. 7). GM: gray matter, WM: white matter, V1: primary visual cortex.

Appendix Fig. A3. Agglomerative hierarchical clustering results for 14 clusters across all feature extraction methods show highest cluster quality by CL-3D. Baseline GLCM, histogram, LBP, and combined features, as well as a pretrained encoder on ImageNet using transmittance and FOM images, produce fragmented cluster assignments in the coronal plane and inconsistent assignments in the sagittal and axial planes. In contrast, CL-2D and CL-3D demonstrate more organized cluster assignments in the coronal plane, with CL-3D showing most consistent cluster assignments in the axial and sagittal planes.