
From Fibers to Cells: Fourier-Based Registration Enables Virtual Cresyl Violet Staining From 3D Polarized Light Imaging

Alexander Oberstrass^{1,2,*} Esteban Vaca^{1,2,*} Eric Upschulte^{1,2,3} Meiqi Niu¹
 Nicola Palomero-Gallagher^{1,3} David Graessel¹ Christian Schiffer^{1,2}
 Markus Axer^{1,4} Katrin Amunts^{1,3} Timo Dickscheid^{1,2,5}

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

² Helmholtz AI, Research Centre Jülich, Germany

³ Cécile & Oskar Vogt Institute of Brain Research, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Germany

⁴ Department of Physics, University of Wuppertal, Germany

⁵ Institute of Computer Science, Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

May 19, 2025

ABSTRACT

Comprehensive assessment of the various aspects of the brain’s microstructure requires the use of complementary imaging techniques. This includes measuring the spatial distribution of cell bodies (cytoarchitecture) and nerve fibers (myeloarchitecture). The gold standard for cytoarchitectonic analysis is light microscopic imaging of cell-body stained tissue sections. To reveal the 3D orientations of nerve fibers, 3D Polarized Light Imaging (3D-PLI) has been introduced as a reliable technique providing a resolution in the micrometer range while allowing processing of series of complete brain sections. 3D-PLI acquisition is label-free and allows subsequent staining of sections after 3D-PLI measurement. By post-staining for cell bodies, a direct link between fiber- and cytoarchitecture can potentially be established within one and the same section. However, inevitable distortions introduced during the staining process make a costly nonlinear and cross-modal registration necessary in order to study the detailed relationships between cells and fibers in the images. In addition, the complexity of processing histological sections for post-staining only allows for a limited number of samples. In this work, we take advantage of deep learning methods for image-to-image translation to generate a virtual staining of 3D-PLI that is spatially aligned at the cellular level. We use a supervised setting, building on a unique dataset of brain sections, to which Cresyl violet staining has been applied after 3D-PLI measurement. To ensure a high correspondence between both modalities, we address the misalignment of training data using Fourier-based registration methods. In this way, registration can be efficiently calculated during training for local image patches of target and predicted staining. We demonstrate that the proposed method enables prediction of a Cresyl violet staining from 3D-PLI, matching individual cell instances.

Keywords deep learning, virtual staining, fiber architecture, cytoarchitecture, polarized light imaging, Cresyl violet, vervet monkey brain

*Equal contribution, co-corresponding authors ✉.

1 Introduction

To understand the organizational principles of the brain, complementary imaging techniques are used to highlight different aspects of brain architecture. Two important aspects of the microstructural organization are fiber- and cytoarchitecture (Nieuwenhuys, 2013; Amunts and Zilles, 2015). While cytoarchitecture encompasses the spatial distribution and shape of cell bodies in the cerebral cortex and subcortical nuclei, fiber architecture refers to the course and composition of nerve fibers. However, cytoarchitecture and fiber-architecture are usually studied using different staining protocols, applied in different sections. Only a few protocols are available to combine cyto- and fiber staining in a single protocol, e.g., Luxol fast blue (Klüver and Barrera, 1953), Bielschowsky (Bielschowsky, 1904) or the triple staining by Novotny (Novotny and Novotny, 1977) and Nowotny (2009). While they allow visualizing cell bodies and fibers in one and the same section, they do not allow the tracing of axons and fiber bundles over long distances.

3D-Polarized Light Imaging (3D-PLI) addresses this limitation. It is a microscopic imaging technique for evaluating the three-dimensional orientation of myelinated nerve fibers in entire, unstained histological brain sections (Axer and Amunts, 2022; Axer et al., 2011a,b). The technique can achieve an in-plane resolution of 1.3 μm , capturing structures at the level of individual fibers and small fiber bundles. 3D-PLI has been used to gain insights into the architecture of nerve fibers in different brain regions, such as the human hippocampus (Zeineh et al., 2017), the sagittal stratum (Caspers et al., 2022) and the vervet monkey visual system (Takemura et al., 2020). In addition, 3D-PLI has been used to validate fiber tractography algorithms and the interpretation of DW-MRI (Caspers and Axer, 2019). 3D-PLI potentially allows joint imaging of fiber tracts and neuronal cell bodies (Zeineh et al., 2017) due to scattering effects affecting light transmittance (Menzel et al., 2020). However, this possibility has not yet been validated.

Cytoarchitecture can be studied in histological sections of postmortem brains with Cresyl violet. The staining provides contrast due to staining of the rough endoplasmic reticulum. This allows to study cell shape, density and distribution, which vary between brain regions. Due to its high spatial resolution, microscopic analysis of histological sections is considered the gold standard to verify structural parcellations (Amunts and Zilles, 2015). Recent advances in high-throughput scanning, data processing algorithms, and computational capacities have enabled the creation of 3D human brain atlases based on cytoarchitecture, such as BigBrain (Amunts et al., 2013), the Allen Adult Human Brain Atlas (AAHA, Ding et al. 2016), and the Julich Brain probabilistic atlas (Amunts et al., 2020).

Since 3D-PLI relies solely on optical properties of the tissue, it is label-free and can be combined with a staining of the same tissue after its measurement. Post-staining, e.g. with Cresyl violet, enables a complementary visualization of neuronal cell bodies, potentially establishing a direct link between cytoarchitecture with 3D fiber-architecture. This requires, however, a complex histological processing, which limits the number of available samples, increases the risk of tissue damage and may lead to deformations of the section. To correct for deformation and artifacts in the images of the two modalities requires a nonlinear registration step. However, 3D-PLI and Cresyl violet stained tissue share only a few automatically identifiable cross-modal registration landmarks, such as blood vessels or morphological landmarks. I.e., post-staining of sections imaged with 3D-PLI is feasible, but technically challenging, and it hardly scales up.

Therefore, we aim to train a deep neural network model to perform image-to-image translation from 3D-PLI to a Cresyl violet staining. Such an approach is often denoted as *virtual histological staining*, which refers to computational methods that generate color-coded images of biological tissue without the need for traditional staining techniques. The methods instead utilize optical properties of the tissue, such as birefringence, autofluorescence, scattering, or absorption to create images that emulate the appearance of stained tissue. A virtual Cresyl violet staining spatially aligned with 3D-PLI would allow an application of established tools for cytoarchitectonic analysis, such as automatic cell instance segmentation (Upschulte et al., 2022), for 3D-PLI measurements. Furthermore, it would provide detailed registration landmarks for cross-modal registration (Ounkomol et al., 2018), thereby offering the opportunity to perform joint acquisition of aligned fiber and cytoarchitecture at a larger scale than possible today.

One of the earliest applications of label-free imaging for visual staining was Quantitative Phase Imaging (QPI, Curl et al. 2004). QPI measures the phase shift of light passing through a sample, producing high-resolution images that reveal the optical properties of the tissue. It was used to generate images of collagen fibers, red blood cells, and other tissue structures without staining (Curl et al., 2004; Park et al., 2018). Later, machine-learning algorithms, especially generative models, have been trained to recognize and virtually stain different tissue structures in unstained images by performing image-to-image translation. They have successfully generated color-coded images of tissue that replicate the appearance of histological stainings, such

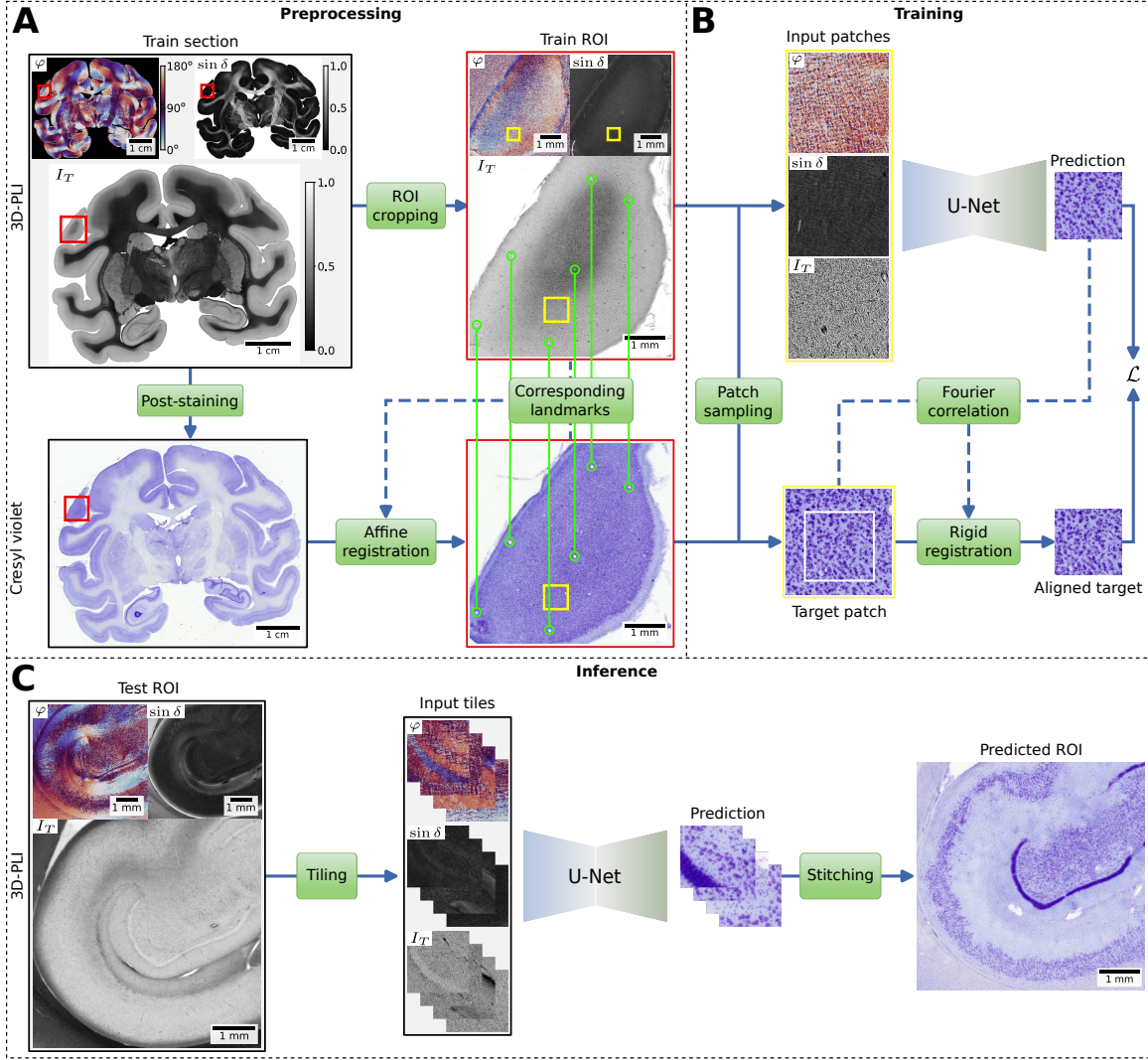


Figure 1: The proposed virtual staining workflow. **(A)** Preprocessing of 3D-PLI sections that were post-stained with Cresyl violet. As paired training data, regions of interest (ROIs) are manually cropped (red boxes) and affine registered using large blood vessels as landmarks (green marker). Background pixels in train sections are masked, and retardation values are scaled using gamma correction for visualization purposes. **(B)** Training of a U-Net model using patches, extracted from same random locations (yellow boxes) in 3D-PLI modalities direction φ , retardation $\sin \delta$, transmittance I_T and the Cresyl violet staining. 3D-PLI patches are used as input to the model to predict a virtual Cresyl violet staining. The Cresyl violet patch acts as target and is rigidly aligned with the prediction during the training procedure. The alignment is performed by our proposed online registration head using Fourier-based correlation of pixels. A loss \mathcal{L} is computed between aligned target and prediction. **(C)** Inference using the trained U-Net model to virtually stain unseen sections or ROIs. Inputs are divided into overlapping tiles, which are processed independently by the U-Net model. The predictions are then stitched back together to form the complete virtual staining.

as a virtual hematoxylin and eosin (H&E), Masson’s trichrome, and Jones’ stain from QPI of label-free tissue (Rivenson et al., 2019), a transformation of H&E stained tissue into Masson’s trichrome, periodic acid-Schiff (PAS), or Jones’ stain (de Haan et al., 2021; Yang et al., 2022). However, these stains are not very good at distinguishing the different components of the nervous tissue.

Machine learning algorithms were also used to predict fluorescence-labeled images from transmitted-light z-stacks (Ounkomol et al., 2018; Christiansen et al., 2018; Cross-Zamirski et al., 2022) or 3D fluorescence structures and a FluoroMyelin stain from bright-field and polarization images of brain slices (Guo et al., 2020). The methods typically use cross-entropy (Christiansen et al., 2018), mean absolute (L1) loss (Guo et al., 2020), or style-related losses such as conditional generative adversarial network (GAN) loss (Rivenson et al., 2019; de Haan et al., 2021; Cross-Zamirski et al., 2022). While including a GAN objective encourages prediction of realistic-looking images, it has no clear mechanism to preserve content when conditioned on a particular input image, and thus may introduce artificial structures (Cohen et al., 2018). A combination with a pixel-wise reconstruction loss (e.g. L1 loss) mitigates this problem of GAN training and improves accuracy of predictions (Isola et al., 2017).

Since methods using paired training data for supervised image-to-image translation typically produce more accurate predictions than unpaired methods (Zhu et al., 2017), a pixel-accurate alignment of training data is necessary. This requires virtual staining methods to either perform a costly registration step or directly acquire paired images. A paired acquisition with 3D-PLI, however, is not feasible and a lack of structural overlap, such as a sufficient number of visible cell instances between the investigated modalities, makes pixel-accurate registration challenging. Therefore, to alleviate the need for perfectly paired training data, we propose a supervised learning objective performing local online registration of training pairs combined with a translational-invariant style comparison. This allows us to train the model on imperfectly registered image pairs with strong content preservation as in paired image-to-image translation, while enabling realistic prediction of subtle structures like cell bodies (Fig. 1). The main contributions are as follows:

- We apply the matching of Gram matrix representations as a *texture sensitive style loss* for the virtual staining, as previously used for texture synthesis (Gatys et al., 2015). Since the computation of Gram matrices is translation invariant, it allows a direct comparison of image statistics between coarsely aligned training examples. It therefore improves the accuracy of predicted cell instances over commonly used GAN style loss.
- An *online registration head* for improving registration accuracy of local image pairs after pre-alignment of larger tissue sections during training. We consider Fourier-based registration methods, which can be computed efficiently in real-time on modern GPU hardware.
- An *equivariance loss* to improve the accuracy of cell instance predictions by addressing the inherent agnosticism of loss computation to constant displacements through online registration.

2 Materials and Methods

2.1 Microscopic imaging of histological brain sections

We demonstrate the proposed virtual staining approach on a set of brain sections for which Cresyl violet staining has been performed after 3D-PLI measurement. The brain sample for this study was obtained from a healthy 2.4-year-old adult male vervet monkey (Wake Forest-ID 1818, Axer et al. 2020; Takemura et al. 2020; Menzel et al. 2022) 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. To obtain an undistorted volumetric reference, a T2 weighted MRI was acquired in-vivo one day prior to sacrifice. The brain was removed from the skull within 24 hours after flushing with phosphate-buffered saline and perfusion fixated with 4% buffered paraformaldehyde. It was stored for several weeks at -70°C in 20% glycerin solution for cryoprotection, and then sliced coronally with 60 µm section thickness using a large-scale cryostat microtome (Polycut CM 3500, Leica Microsystems, Germany). Blockface images of the frozen tissue block were taken with a CCD camera before cutting each brain section. The images were reconstructed into a 3D blockface volume to provide an undistorted reference for section realignment.

2.1.1 Image acquisition

For 3D-PLI acquisition, brain sections were scanned using a polarizing microscope (LMP-1, Taorad, Germany) with 1.3 µm resolution (Axer et al., 2011b; Axer and Amunts, 2022). The focus level of the LMP-1 was manually adjusted to the center of the tissue for each section. Inside the LMP-1 microscope, sections were

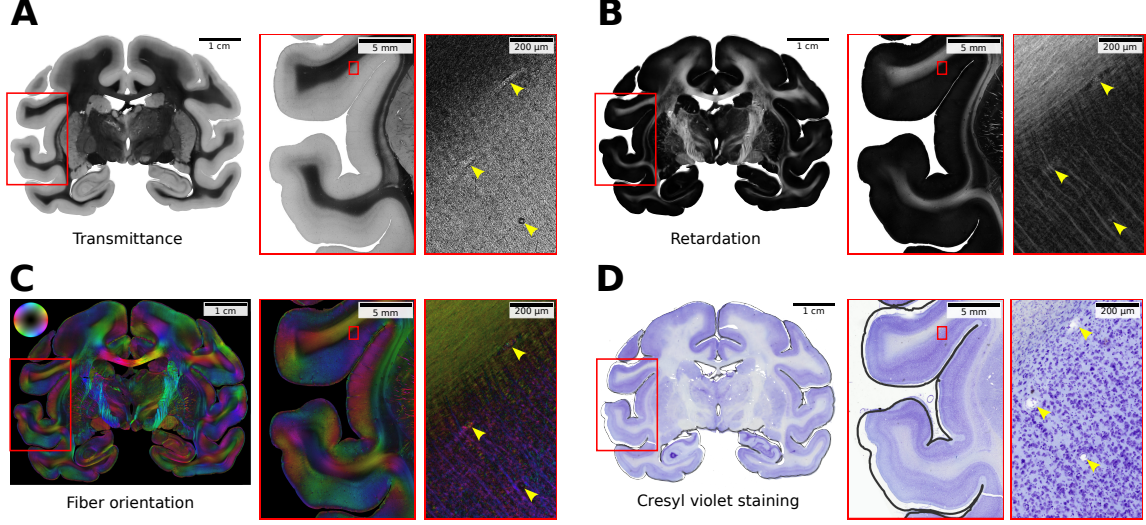


Figure 2: Data modalities and registration challenges for training section 544. (A-C) 3D-PLI parameter maps: Transmittance I_T , retardation $\sin \delta$ and fiber orientation in HSV color space (hue: direction φ ; saturation, brightness: retardation $\sin \delta$). Background pixels are masked for visualization purposes only. (D) Affine registered Cresyl violet staining. The pial surface of the 3D-PLI acquisition is shown as a contour plot in D for reference. Between both data acquisitions remains a nonlinear misalignment that cannot be resolved by a global affine transformation. At a local scale, the remaining misalignment is approximately linear. Yellow arrows indicate blood vessels that can be used as mutual registration landmarks for coarse alignment.

placed on a specimen stage between a rotating linear and a circular polarizer on top of an incoherent light source with a wavelength of 550 ± 5 nm. Images were taken by a CCD camera for nine equidistant rotation angles ρ of the rotating linear polarizer, covering 180° of rotation. At each pixel, the measured intensity of the images followed a sinusoidal profile as

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

Using harmonic Fourier analysis, parameter maps of transmittance (I_T), retardation ($\sin \delta$), and direction (φ) were obtained from the measurements, with a resolution of approximately $34,000 \times 44,000$ pixels per section, revealing their fine-grained nerve fiber architecture (Fig. 2A-C). After 3D-PLI acquisition, brain sections were washed, fixed and stained for cell bodies with Cresyl violet Nissl staining to reveal their cellular architecture. They were scanned using flatbed scanners at $1 \mu\text{m}$ in-plane resolution and saved as RGB color pictures with eight bit color depth, storing pixel values from 0 to 255 (Fig. 2D).

2.1.2 Tissue shrinkage correction

Deformations induced by histological processing include a volumetric shrinkage of brain tissue (Amunts et al., 2005). To correct this effect, the extent of shrinkage is usually estimated from the ratio between the histological processed and true brain volume represented by the fresh weight of the whole brain (Amunts et al., 2005, 2013) or an MRI reference (Wagstyl et al., 2020). To estimate the shrinkage of the 3D-PLI acquisition, a linear registration of the 3D reconstructed blockface volume and the postmortem MRI of the same brain was performed. The determinant of the transformation matrix showed a global volume change below 1%, which can be neglected. In a second step, we calculated 2D shrinkage factors for each brain section considered in this work as the quotient between the area occupied by tissue in the 3D-PLI measurement and its corresponding blockface image. For test section #559, we estimated a 2D shrinkage factor of 0.97 meaning a slight expansion of the 3D-PLI measurement relative to its original area within the MRI. This observation is consistent throughout all sections with 2D shrinkage factors between 0.95 and 0.99. We report in-plane cell sizes measured with 3D-PLI multiplied by an individual 2D shrinkage factor per section, assuming a uniform area change of cells and surrounding tissue.

2.2 Initial cross-modality alignment

After the subsequent processing of brain tissue, Cresyl violet images exhibit a deformation relative to the 3D-PLI acquisition. To align both modalities, an initial affine registration of whole brain sections is performed by manual identification of large blood vessels as landmarks visible in both modalities.

Performing the initial affine registration reveals remaining nonlinear deformations as shown in Fig. 2D. Since nonlinear deformations typically have low spatial frequencies, causing smooth, large-scale distortions, we expected near-linear deformations at smaller scales. Therefore, performing an additional more local linear registration would lead to a better fit. We subsequently crop square regions of interest (ROIs) with a size of 4,096 pixels (~ 5.3 mm) and without visible artifacts, covering distinct cellular architectures across the whole coronal plane. For all ROIs we perform additional affine registration and make sure that transformed landmarks have a maximum distance of 70 pixels (91 μm) from their matches. All ROIs are warped and resampled to 1.3 μm using linear interpolation to match the coordinate space of 3D-PLI. While this results in a loss of precision relative to the original resolution of Cresyl violet scans of 1 μm , matching the resolutions of both modalities facilitates subsequent processing and analysis steps.

2.3 Fourier-based online registration of image patches

To correct the remaining misalignment of 3D-PLI and Cresyl violet after affine registration of ROIs at a finer local scale, we introduce an online registration head that performs cross-modality alignment during training based on model predictions of small image patches. We assume that once the style transfer model has learned to reconstruct microscopic landmarks (e.g., individual cells or small blood vessels), such online registration will promote the learning of additional landmarks until the training can use pixel-aligned training examples.

The registration method performed during training needs to be computationally efficient, since training will require numerous registration iterations. Conventional feature-based image registration methods are accurate and can model nonlinear deformations but are computationally expensive and sensitive to image degradation. As we assume deformations to be approximately linear at a local scale, we take advantage of Fourier-based image correlations (Tong et al., 2019), which can efficiently recover a translation between images in the frequency domain.

2.3.1 Translational shift

Fourier-based image correlation methods are able to retrieve a translational shift $(\Delta u, \Delta v)$ between image functions $f(u, v)$ and $g(u, v)$ defined for integer pixel coordinates (u, v) , such that $f(u, v) = g(u + \Delta u, v + \Delta v)$. Both functions f and g represent images of equal height H and width W and are for now assumed to repeat periodically with a periodicity of H and W , respectively.

A common approach to retrieve the translational shift between f and g is to use circular cross-correlation (Tong et al., 2019), which can be efficiently computed in the frequency domain as

$$\begin{aligned} \text{CC}^\circ[a, b] &= (f \star g)[a, b] \\ &= \sum_{u, v} f(u - a, v - b) g(u, v) \\ &= \left(\mathcal{F}^{-1} \left\{ \overline{\mathcal{F}\{f\}} \mathcal{F}\{g\} \right\} \right) [a, b], \end{aligned} \quad (2)$$

for all integer shifts $[a, b]$, where \mathcal{F} denotes the Fourier transformation, \mathcal{F}^{-1} its inverse, $\overline{\mathcal{F}\{f\}}$ its complex-conjugate Fourier coefficients, and where we sum over all pixel coordinates (u, v) . The translational shift can subsequently be recovered by the location of the maximum value in CC° as

$$(\Delta u, \Delta v) = \arg \max_{(a, b)} \text{CC}^\circ[a, b]. \quad (3)$$

Eq. (2) can be extended to the mean over squared distances between pixel values as

$$\begin{aligned} \text{MSE}^\circ[a, b] &= \frac{1}{HW} \sum_{u, v} (f(u - a, v - b) - g(u, v))^2 \\ &= \frac{-2(f \star g)[a, b] + \sum_{u, v} (f^2(u, v) + g^2(u, v))}{HW}. \end{aligned} \quad (4)$$

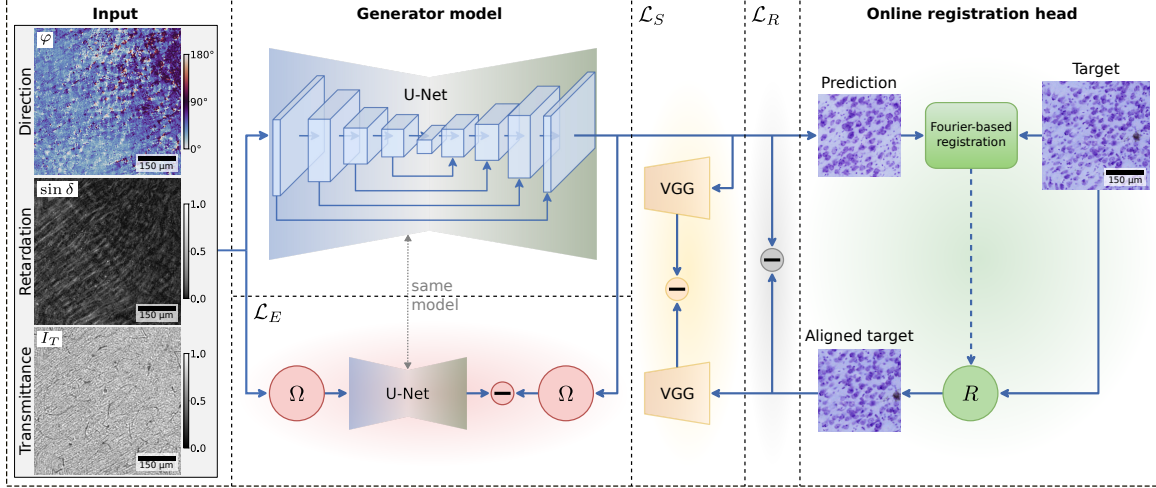


Figure 3: Illustration of the proposed virtual staining approach. Patches of 3D-PLI parameter maps transmittance I_T , retardation $\sin \delta$ (scaled using gamma correction for visualization), and direction φ are used as input to a 2D convolutional U-Net model as generator to predict a virtual Cresyl violet staining. An online registration head estimates a rigid transformation R between a coarsely aligned Cresyl violet target patch and the prediction via Fourier-based registration. Transformation R is used to align target and prediction at the patch level. We calculate three distinct loss components: \mathcal{L}_R , \mathcal{L}_S and \mathcal{L}_E . Reconstruction loss \mathcal{L}_R performs a pixel-wise comparison between prediction and aligned target. Style loss \mathcal{L}_S compares feature maps of a VGG network encoder using Gram matrices to mimic the style of the target image. Equivariance loss \mathcal{L}_E applies the same U-Net model a second time to a rotated version of the input by rotation Ω . The output is compared with the prediction rotated by same rotation Ω , which promotes stability and avoids learning a constant shift of pixels in the prediction.

The translational shift $(\Delta u, \Delta v)$ can be recovered analog to Eq. (3) by computing the arg min. For periodic image functions f and g , solutions of CC° and MSE° are identical as the sum over squared functions f^2 and g^2 is constant (Fienup, 1997).

Since histological images are not periodic, zero padding is applied to fill both images up to a shape of $(H_f + H_g - 1, W_f + W_g - 1)$, in order to break periodicity and allow processing of images with different heights H_f, H_g and widths W_f, W_g . The zero-padded images are denoted by new image functions f_0 and g_0 . Furthermore, additional masks M_f and M_g are introduced, which have a value of one at all pixel coordinates within original height and width and zero elsewhere. We reformulate Eq. (4) to a non-circular form as

$$\text{MSE}[a, b] = \left(\frac{(f_0^2 \star M_g) - 2(f_0 \star g_0) + (M_f \star g_0^2)}{M_f \star M_g} \right) [a, b], \quad (5)$$

which can be efficiently computed by multiple applications of Eq. (2). Here, the cross-correlations of f_0^2 and g_0^2 with masks M_g and M_f , respectively, ensure that pixel values of the original unpadded images are not compared with zero padding values. Furthermore, the score is divided by the correlation between M_f and M_g to account for the number of overlapping pixels between original unpadded images. We apply the same division by $M_f \star M_g$ also for CC° to retrieve a non-circular variant called CC. While the solutions for circular CC° and MSE° are identical, the solutions for non-circular CC and MSE differ, resulting in distinct registration metrics.

2.3.2 Rotation and scale

While relative scale and rotation between images can also be retrieved in the frequency domain (Sheng, 1989; Reddy and Chatterji, 1996), we expect only small relative rotation angles and minor scale variations due to the initial affine registration by matching blood vessels. Therefore, we leverage the parallel processing capability of GPUs to perform an exhaustive search over a fixed set of rotation angles without scaling adjustments. We calculate registration metrics for every translational shift and rotation and select the rotation angle and shift combination that yields a global optimum.

2.4 Conditional generation of Cresyl violet staining

Image-to-image translation refers to the process of generating images using a generator model G conditioned on an input image x . Predictions $G(x)$ of the model are compared with actual target images y .

Our style transfer algorithm consists of three main elements:

1. **Reconstruction Loss** \mathcal{L}_R : Similar to (Isola et al., 2017), we use an L_1 loss between target y and prediction $G(x)$ to encourage pixel correspondence as

$$\mathcal{L}_R = \mathbb{E}_{x,y} \|R(y) - G(x)\|_1, \quad (6)$$

where R performs the proposed online registration to spatially align y and $G(x)$ before loss calculation. This enables the utilization of imperfectly aligned training data to penalize any discrepancies in corresponding pixel values between $G(x)$ and y .

2. **Style Loss** \mathcal{L}_S : We apply a texture-sensitive style loss proposed by Gatys et al. (2015) based on squared distances between Gram matrix representations of neural network features. Given a pre-trained VGG encoder (Simonyan and Zisserman, 2015), Gram matrix representations are computed from feature activations of its layers to characterize the texture of images at different complexities. For each layer l , the encoder produces a different number of N_l feature maps, each storing K_l spatial entries (i.e., height \times width). Elements of the Gram matrix G_{ij}^l at layer l are computed as the inner product between the i -th and j -th feature map F_i^l and F_j^l , where each map is flattened to a K_l -dimensional vector:

$$G_{ij}^l = \sum_{k=1}^{K_l} F_{ik}^l F_{jk}^l. \quad (7)$$

Since the Gram matrix computation captures global feature correlations rather than spatial locations, this allows a translation-invariant comparison of image statistics. The style loss is computed over all L layers of the VGG encoder, using Gram matrix representations G^l for the online registered target and \hat{G}^l for the prediction:

$$\mathcal{L}_S = \mathbb{E}_{x,y} \left[\sum_{l=1}^L \frac{1}{K_l^2 N_l^2} \sum_{i=1}^{N_l} \sum_{j=1}^{N_l} (G_{ij}^l - \hat{G}_{ij}^l)^2 \right]. \quad (8)$$

We refer to this style loss as *Gram loss*.

3. **Equivariance Loss** \mathcal{L}_E : By registering target y to generator prediction $G(x)$ before loss calculation, displacing objects in $G(x)$ relative to its input x is not penalized by reconstruction loss \mathcal{L}_R . To prevent pixel shifts in $G(x)$, we introduce a third loss component enforcing equivariance with respect to rotations as

$$\mathcal{L}_E = \mathbb{E}_x \|\Omega(G(x)) - G(\Omega(x))\|_2, \quad (9)$$

where operator Ω represents an image rotation of 180° . Minimizing Eq. (9) ensures that pixels in $G(x)$ correspond to pixels at the same pixel coordinates in x as any discrepancy would cause a mismatch of $\Omega(G(x))$ and $G(\Omega(x))$.

All components are aggregated into total loss

$$\mathcal{L} = \lambda \mathcal{L}_R + (1 - \lambda) \mathcal{L}_S + \eta \mathcal{L}_E, \quad (10)$$

with relative weightings $\lambda \in [0, 1]$ and $\eta \geq 0$ as hyperparameters. An overview of the framework is illustrated in Fig. 3.

As an alternative for style loss \mathcal{L}_S , we consider adversarial training (Goodfellow et al., 2014) using a *GAN loss* to compare with previous work in virtual staining (Rivenson et al., 2019; Cross-Zamirski et al., 2022). We implement the GAN loss in the form of a Wasserstein GAN (Arjovsky et al., 2017). In contrast to conditional GAN training (Mirza and Osindero, 2014; Isola et al., 2017), we do not condition the discriminator on input images x , as this would cause the model to reproduce any misalignment in the training data. We refer to models trained with *Gram loss* for \mathcal{L}_S as Gram and models trained with *GAN loss* as GAN. In combination with the online registration head, we refer to the models as Gram+Reg and GAN+Reg.

2.5 Model training

For the generator G , we use the same 5-layer U-Net (Ronneberger et al., 2015) with numbers of features [32, 64, 128, 256, 512] in all experiments, and adjust the input and output channels to 3 according to

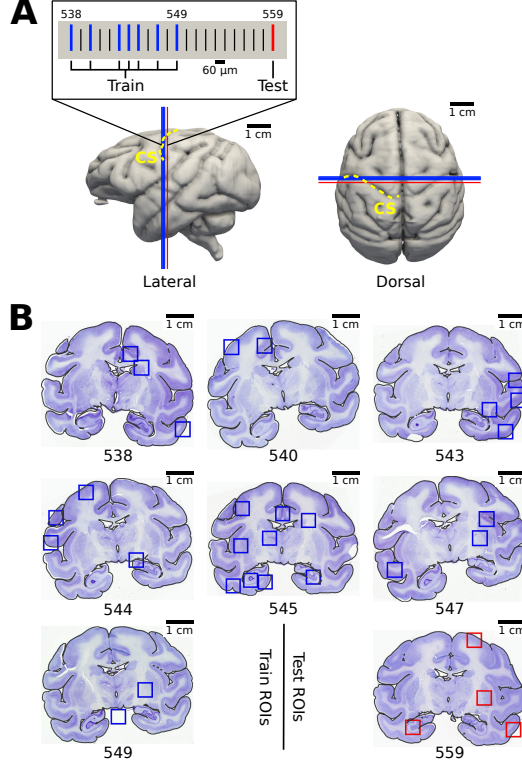


Figure 4: Localization of train and test data. **(A)** Seven sections used for training (blue stripes) and one section used for testing (red stripe) were taken at the level of the central sulcus (CS; yellow dashed lines). Locations are shown on top of the 3D reconstructed blockface of the brain for reference. Train and test data are 0.6 mm apart from each other. **(B)** Selected locations of train and test regions of interest (ROIs), which are used for training and testing the models. The images show ROIs from each of the train and test sections on top of globally affine registered Cresyl violet images. Black contour plots outline the pial surface of corresponding 3D-PLI sections for reference.

our setup. To train G , we use square 3D-PLI patches of 444 pixels size, represented by parameter maps transmittance (I_T), retardation ($\sin \delta$) and direction (φ). We reformulate the 3D-PLI parameters as triplets ($I_T, \sin \delta \cos(2\varphi), \sin \delta \sin(2\varphi)$) to resolve the circular behavior of direction φ , standardize the channels and stack them to the input of generator G . Due to the fully convolutional approach of the U-Net model without padding, the generated output predictions have a reduced size of 260 pixels. Unless specified otherwise, we use a patch size of 360 pixels for the target Cresyl violet images, centered at the input patch position. They are chosen to be larger than the model predictions to allow the online registration to correct translational shifts of up to 50 pixels in any direction, while keeping the predictions fully contained within the target images. We normalize image pixel values of the Cresyl violet staining to the range of $[0, 1]$.

For computing style loss \mathcal{L}_S , we extract features from a VGG19 model (Simonyan and Zisserman, 2015) to compute the *Gram loss*. The VGG feature encoder network has a depth of four layers and three input channels with pre-trained weights on ImageNet (Deng et al., 2009). We multiply the style loss \mathcal{L}_S by a constant factor of 10^4 to bring it to the same order of magnitude as reconstruction loss \mathcal{L}_R . For the training, we use Adam optimizer (Kingma and Ba, 2017) with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate of 10^{-3} . If not stated otherwise, we use $\eta = 0.1$ and $\lambda = 0.5$ as default in the training objective in Eq. (10).

In the case of GAN style loss \mathcal{L}_S , we use a 4-layer convolutional network as discriminator with kernel size 4, stride 2, padding 1, and feature size of $[32, 64, 128, 256]$, followed by batch normalization (Ioffe and Szegedy, 2015) and Leaky ReLU after each convolution. For training, we use the Wasserstein GAN (Arjovsky et al., 2017) objective and a separate Adam optimizer for the discriminator and the generator, using $\beta_1 = 0.5$, $\beta_2 = 0.999$ and a learning rate of 10^{-4} . We perform five updates for the discriminator for one update of the generator and clamp discriminator weights at 0.03 after each step.

Training data For model training and evaluation, we use eight coronal sections at the level of the central sulcus. Seven sections are used for training and one section is kept for testing with a gap of 0.6 mm between train and test sections (Fig. 4 A). From the training sections, 27 affine-aligned ROIs are extracted (Fig. 4 B), where one ROI is held out for validation to identify possible overfitting. For each ROI, we extract joint target Cresyl violet images and 3D-PLI modalities at the same center location. To maximize the diversity of the training examples, we do not pre-compute training patches but sample them randomly during the training process. We use a batch size of 128 and draw 32,768 paired random patches per epoch evenly distributed across the training ROIs, resulting in approximately 1,260 random samples per ROI per epoch. All training is performed for 150 epochs or until model convergence if validation loss did not decrease for at least 50 epochs.

Data augmentation To enhance the robustness of our trained models we employ 3D-PLI-specific data augmentations, which were carefully modified to keep physically plausible signal parameters (Oberstrass et al., 2024). Specifically, we perform random rotation by angles between -180° and 180° with mirror padding and horizontal and vertical random flipping. In both cases, direction parameter maps φ are corrected accordingly. Additionally, we perform Gaussian blurring of 3D-PLI parameter maps for random standard deviations up to $\sigma = 1.5$ and kernel sizes of 3 or 5. We scale thickness and attenuation coefficients for 3D-PLI parameter maps by random values between 0.5 and 2.

Implementation All models were trained on the supercomputer JURECA-DC at the Jülich Supercomputing Centre (JSC, Thörnig 2021) on a single node by splitting each batch equally onto 4 NVIDIA A100 GPUs using distributed data-parallel strategy. For data pre-processing 128 worker processes were spawned on 128 CPU cores. For reference, training for 100 epochs took 8 hours with online registration and 4 hours without on this hardware. The implementations are based on the Quicksetup-ai template by the HelmholtzAI Consultants Munich², using the frameworks: *PyTorch* (Paszke et al., 2019), *PyTorch Lightning* (Borovec et al., 2022) and *Hydra* (Yadan, 2019).

Online registration For the online registration head, we restrict accepted solutions of Eq. (5) to translation + rotation pairs that cause the registered target to have full overlap with the prediction, avoiding loss calculation over zero-padded values. We check the translation correction for 31 rotation angles from -7.5° to 7.5° with steps of 0.5° and take the translation + rotation pair with the best registration score.

3 Experiments and Results

We compare a selection of performance scores to identify optimal hyperparameters of the proposed method and assess the overall potential of the best performing model. The main hyperparameters are the choice of the online registration metric, the type of style loss \mathcal{L}_S , its relative weighting λ to reconstruction loss \mathcal{L}_R , and whether to use the additional equivariance loss \mathcal{L}_E . To reduce the massive computational demands by a rigorous grid search across all hyperparameters, we choose to identify a suitable choice for the online registration metric and model variants using different loss components independently before determining an optimal relative weighting λ .

3.1 Experiment setup

3.1.1 Test data

We manually select four ROIs for model evaluation, ensuring a diverse representation of different cytoarchitectonic characteristics from the held-out test section (Fig. 4B). The ROIs contain primary motor cortex 4a, temporal cortical area TE, the hippocampal cornu Ammonis (CA) region, and parts of the putamen and globus pallidus (Pars interna and Pars externa) as subcortical structures (Fig. 5).

As the computation of image metrics requires a precise alignment of test data, we perform elastic registration of test ROIs based on landmarks and image intensity using the *bUnwarpJ* (Arganda-Carreras et al., 2006) algorithm. We use predictions of an independently trained Gram+Reg model as target. The predictions are used to manually identify 15-25 characteristic cell clusters as landmarks per ROI, which are confirmed by the location and arrangement of faint shadows of cells visible in the 3D-PLI transmittance. For registration, we use an image weight of 1.0, a landmark weight of 10.0, and a consistency weight of 10.0. The strong weights

²<https://github.com/HelmholtzAI-Consultants-Munich/Quicksetup-ai>

for landmarks and consistency are chosen to prevent the transformation field from overly conforming to the predictions and to overcome local optima.

3.1.2 Evaluation scores

We evaluate the impact of different model parameter choices on the quality of the predicted virtual staining by applying structural similarity index measure (SSIM, Wang et al. 2004), mutual information (MI), and root-mean-square error (RMSE). For each metric, we report the mean over all test ROIs. To evaluate the ability to localize cell bodies, we compute F1 scores for instance segmentations of predicted and target cells using a contour proposal network (CPN, Upschulte et al. 2022, 2023). We only consider cells with a minimum in-plane size of $100 \mu\text{m}^2$ and require a minimum intersection over union (IoU) of 30%. Smaller cells could not be reconstructed accurately and would introduce noise into the evaluation. For the cell detection model, we fine-tuned a pre-trained³ CPN for cell body segmentation in cell-stained microscopy images using the *celldetection*⁴ Python package. Fine-tuning was performed on a diverse mix of manually annotated images, as well as synthetic data.

We restrict the computation of evaluation metrics to gray matter, where most of the neuronal cell bodies are located, excluding white matter and background pixels. In test ROIs showing cortical regions, we further exclude molecular layer I as it consists of only a few individual neurons and exhibits a nonlinear deformation due to tissue shrinkage, which could not be reliably corrected in the registration. For the ROI showing the hippocampus, we exclude the fascia dentata, as its granular layer consists of very densely packed neurons, which cannot reliably be distinguished and restrict the analysis to hippocampal CA1-CA4 regions.

3.2 Online registration metric

To compute a pixel-aligned reconstruction loss we apply the online registration head (Sec. 2.3). We consider correlation-based registration metrics CC and MSE with phase correlation (PC, Kuglin 1975) and blur-invariant phase correlation (BIPC, Ojansivu and Heikkila 2007). We apply a Hann window to PC and BIPC before their calculation to mitigate their bias towards the sharp image edges (Gonzales and Woods, 2008).

To understand how image degradation effects can influence the success of the online registration head, we compare the robustness of the metrics against noise and blur. A square target image with 460 pixels size is extracted from a random location in the Cresyl violet staining. Next, a smaller moving tile with a size of 260 pixels is extracted from a random location within the target and distorted by gradually increasing noise or blur on the image. We add noise from a zero-centered Gaussian distribution with standard deviation σ increasing from 0 to 25 and Gaussian blur with increasing kernel sizes with standard deviations σ growing from 0 to 50. The registration head is applied with each metric to realign the degraded moving tile with its location in the target image. We compare the hit rate over 100 examples per degradation step. The registration is considered successful if the determined translational displacement remains within 5 pixels of the actual displacement.

Fig. 6 shows that as blur increases, BIPC with a Hann window achieves the best performance. This result is expected, as the metric remains invariant to blur. We also observe that CC and MSE perform best in mitigating the effects of noise, while the other metrics fail already with a minor amount of noise. Overall, MSE provides a balanced tradeoff between blur and noise response.

To make an optimal choice of a registration metric for online registration, we train Gram+Reg models with CC, BIPC, PC, and MSE compared to a Gram model without online registration. We use a Cresyl violet target patch size of 360 pixels and predictions of 260 pixels, centered within the target patch. This allows the online registration to correct for a maximum of 50 pixels translation in each dimension. For Gram+Reg models trained with CC, BIPC and the Gram model, $\lambda = 0.1$ was chosen as it improved results compared to $\lambda = 0.5$ for PC and MSE.

Tab. 1 shows a quantitative comparison of the models. Independent of the metric, the best results are observed when using Gram+Reg with online registration. Among all metrics considered, the model with MSE performs best.

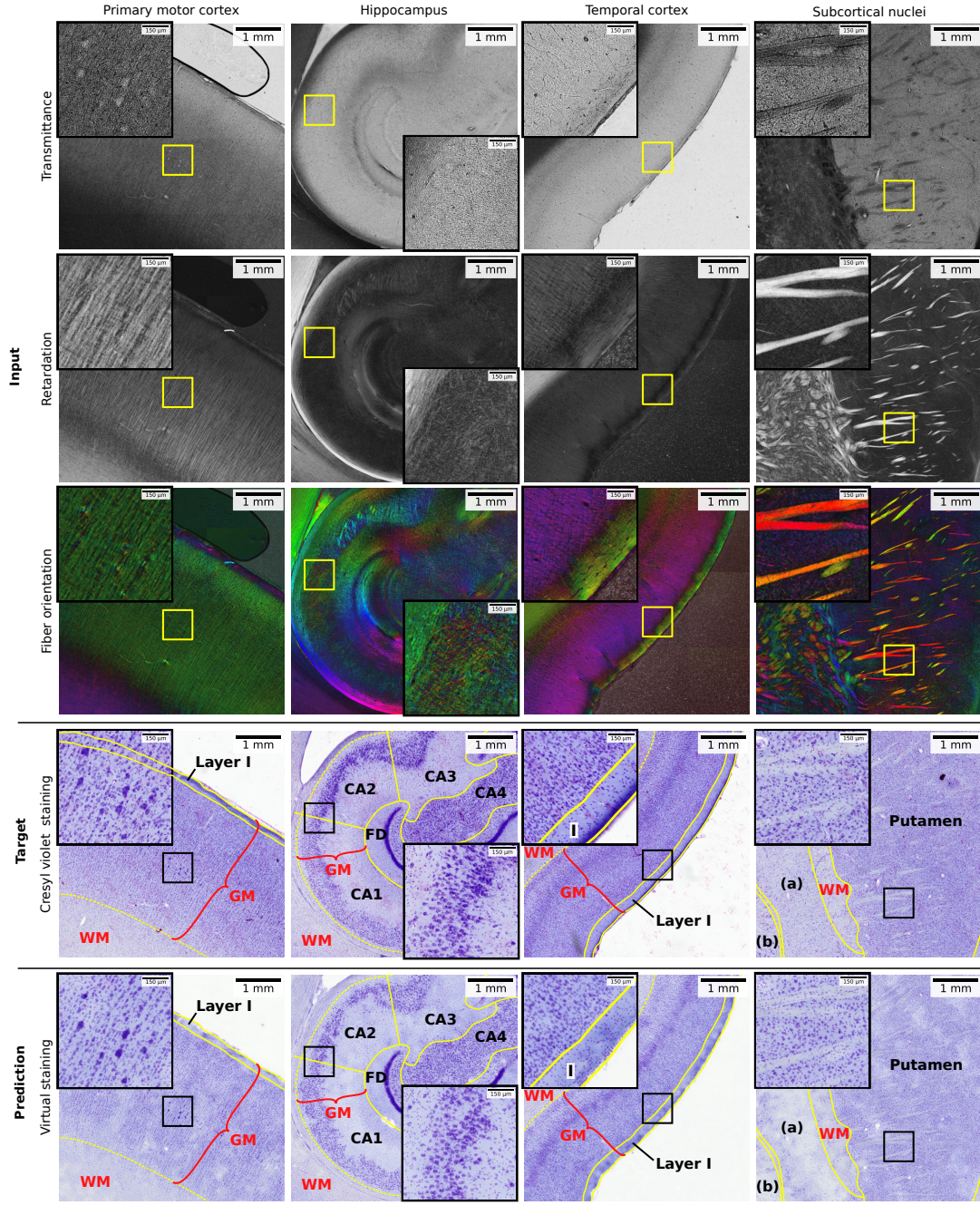


Figure 5: Overview of ROIs used for the evaluation, which represent distinct cellular architectures. They were extracted from section 559, located 0.6 mm apart from the training sections. Embedded windows show magnified details inside each ROI. The locations of the magnified views are indicated by yellow frames. Columns each show one of the four test ROIs taken from the anterior subdivision of the primary motor cortex (4a), the hippocampal cornu Ammonis (CA) region, temporal cortical area TE, and parts of the putamen and globus pallidus (GP; a: GP Pars interna; b: GP Pars externa) as subcortical nuclei. The first three rows demonstrate 3D-PLI modalities transmittance I_T , retardation $\sin \delta$ (scaled using gamma correction for visualization) and fiber orientation in HSV color space (hue: direction φ ; saturation: retardation $\sin \delta$). The 3D-PLI modalities are compared to the registered target Cresyl violet and predicted virtual staining.

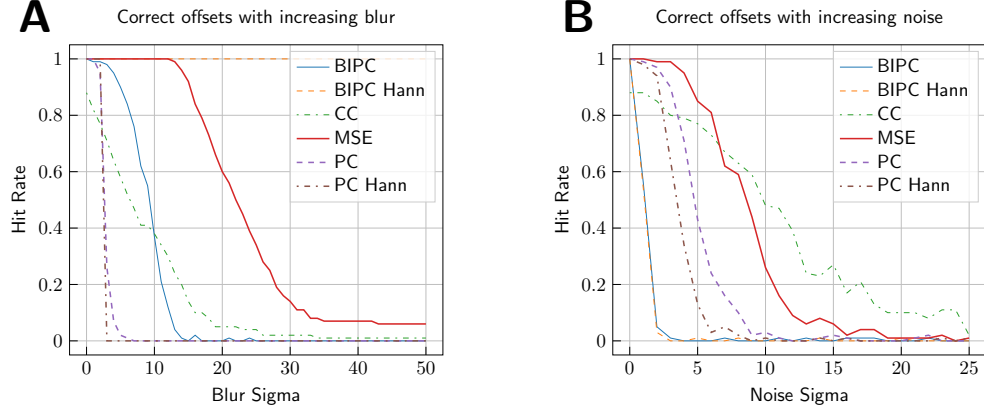


Figure 6: Robustness of different online registration metrics against synthetic image distortions. We report the proportion of correct rigid alignments for different blur (A) and noise (B) levels by registration metrics MSE, CC, PC, and BIPC. The registration is considered successful (a hit) if the translational displacement obtained remains within 5 pixels of the actual displacement.

Table 1: Performance of models trained with different online registration metrics in terms of similarity of generated images with ground truth. We compare cross-correlation (CC), phase-correlation (PC), blur-invariant phase-correlation(BIPC), and mean-squared error (MSE) as online registration metrics against using no online registration (-).

Method	Metric	MI \uparrow	RMSE \downarrow	SSIM \uparrow	F1 \uparrow
Gram	-	0.126	33.6	0.354	48.8
Gram+Reg	CC	0.136	33.0	0.379	55.2
	BIPC	0.149	31.5	0.371	54.0
	PC	0.185	31.9	0.415	59.4
	MSE	0.226	29.8	0.444	63.7

3.3 Performance analysis of different model variants

To justify including equivariance loss \mathcal{L}_E (Eq. (9)) in the total loss, Fig. 7 illustrates the effect of training Gram+Reg models with this loss component enabled and disabled. With \mathcal{L}_E disabled ($\eta = 0$), cell instances in the model predictions exhibit a relative displacement. Enabling \mathcal{L}_E ($\eta = 0.1$) improves the overlap of cell instances with the original Cresyl violet staining.

To quantify the impact of the loss component \mathcal{L}_E on the predictions, we compare the models with \mathcal{L}_E enabled and disabled in Tab. 2. We consider Gram and Gram+Reg, as well as GAN and GAN+Reg as model variants. All models use $\lambda = 0.5$ except for the Gram model without online registration, where $\lambda = 0.1$ showed better performance. Since displacements of cell instances in the model predictions occur randomly with varying strength, we report mean and standard errors across four independent trainings. Models trained with \mathcal{L}_E enabled show the highest correspondence with the target staining. Without online registration, the advantage of including \mathcal{L}_E becomes negligible. Furthermore, results in Tab. 2 demonstrate that Gram models outperform their corresponding GAN models in all evaluation metrics. Regardless of the variant used for style loss \mathcal{L}_S , all models benefit from the online registration.

A comparison of model predictions is shown in Fig. 8C for the whole cortical depth of an Isocortex sample from temporal cortical area TE. GAN+Reg and Gram+Reg provide the most realistic-looking reconstructions in terms of relative size and shape of cell bodies, along with the differences between layers concerning cell packing density. While Gram+Reg seems to be influenced by technically-related inhomogeneities in staining intensity, GAN+Reg produces a clearer contrast between stained cells and surrounding tissue. Gram+Reg

³ginoro_CpnResNeXt101UNet-fbe875f1a3e5ce2c

⁴<https://github.com/FZJ-INM1-BDA/cellldetection>

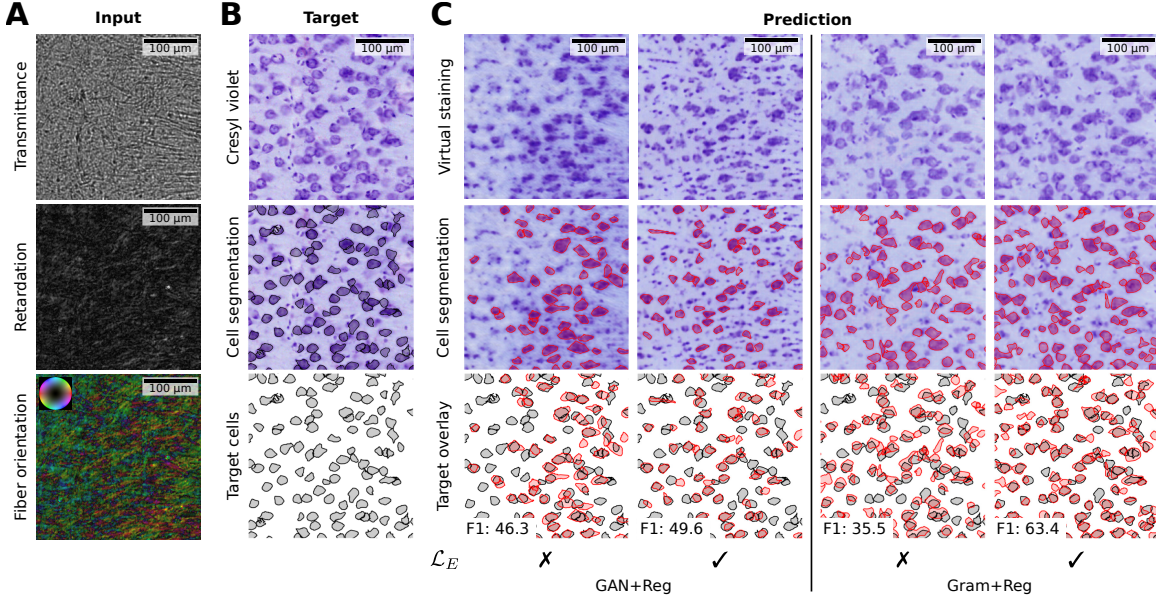


Figure 7: Including equivariance loss \mathcal{L}_E in the training improves cell instance overlap of Gram+Reg and GAN+Reg models. (A) 3D-PLI input as transmittance I_T , retardation $\sin \delta$ (scaled using gamma correction) and fiber orientation map in HSV color space (hue: direction φ ; saturation, brightness: retardation $\sin \delta$) for an example patch from the pyramidal layer of the hippocampal CA1 region. (B) Registered target Cresyl violet staining, a cell segmentation by a CPN model, and the identified target cell instances. (C) Predicted virtual staining of models trained with equivariance loss \mathcal{L}_E enabled (✓) or disabled (✗). Cell instances are overlaid with their target cells. Models trained with equivariance loss show less displacement of cells and increased F1 scores for this specific image patch. Since the displacement occurs randomly, a patch and random seeds for the models are selected that show a particularly strong effect.

Table 2: Effect of the proposed online registration on models trained with different choices of style loss \mathcal{L}_S . Each variant is trained with online registration head (Reg.) and equivariance loss (\mathcal{L}_E) enabled or disabled. MSE metric is used for online registration and $\eta = 0.1$ for weighting of \mathcal{L}_E . The deviation is reported as standard error over four independent trainings with different random seeds.

\mathcal{L}_S	Reg.	\mathcal{L}_E	MI \uparrow	RMSE \downarrow	SSIM \uparrow	F1 \uparrow
Gram	✓	✓	0.224 \pm 0.003	29.6 \pm 0.6	0.445 \pm 0.003	63.1 \pm 0.5
		✗	0.211 \pm 0.004	30.2 \pm 0.4	0.432 \pm 0.004	61.6 \pm 0.4
	✗	✓	0.111 \pm 0.012	33.6 \pm 0.8	0.314 \pm 0.023	38.9 \pm 7.6
		✗	0.107 \pm 0.007	34.2 \pm 0.3	0.304 \pm 0.015	40.6 \pm 5.7
GAN	✓	✓	0.160 \pm 0.008	30.5 \pm 0.5	0.389 \pm 0.012	50.3 \pm 2.1
		✗	0.126 \pm 0.008	33.0 \pm 0.4	0.347 \pm 0.010	43.0 \pm 1.4
	✗	✓	0.085 \pm 0.008	38.1 \pm 0.5	0.249 \pm 0.001	21.1 \pm 0.3
		✗	0.096 \pm 0.008	38.8 \pm 1.7	0.253 \pm 0.005	22.2 \pm 1.3

detects the transition from the cortex to white matter better than GAN+Reg. Both methods introduce an artificial arrangement of cells into cortical columns not present in the original Cresyl violet staining.

To further analyze the ability of the methods to reconstruct the laminar cell organization, we compute grey level index (GLI) values (Schleicher et al., 1978; Zilles et al., 1978). GLI values provide an established proxy for volume density of stained cell bodies in gray matter regions and are used to characterize the laminar architecture of cortical areas (Schleicher et al., 2000). To compute GLI values, adaptive thresholding is applied to create a mask of pixels occupied by cell bodies for each image. The masks are subsequently down scaled by a factor of 16 to a spatial resolution of 20.8 μm , with each value representing the fraction of segmented pixels. Resulting GLI images, cropped to the area between pial surface and the cortex/white matter transition, are displayed for the Cresyl violet target and the predictions in Fig. 8D.

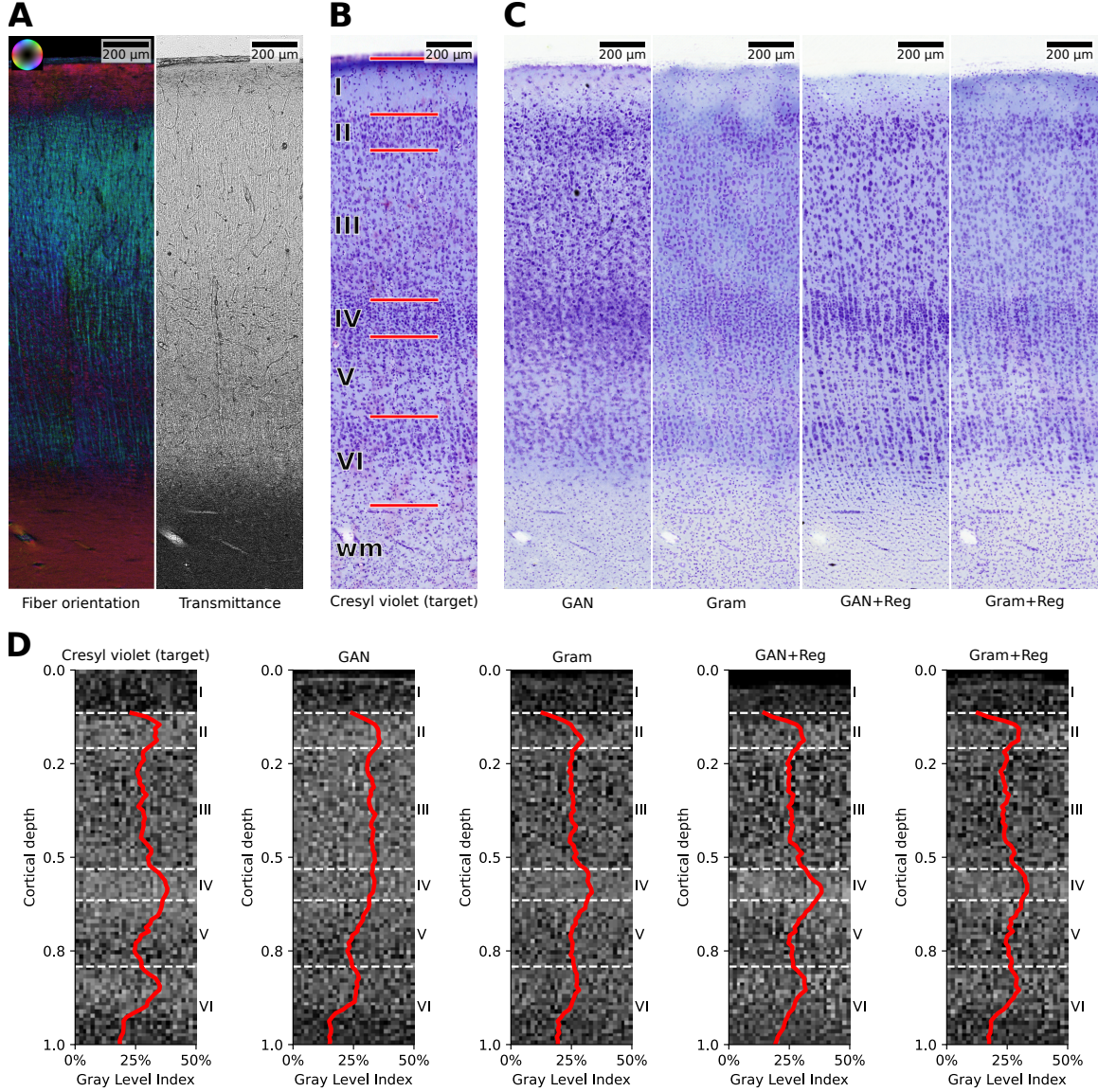


Figure 8: Comparison of virtual stainings for a patch of temporal area TE. (A) 3D-PLI input visualized as fiber orientation in HSV color space (hue: direction φ ; saturation, brightness: retardation $\sin \delta$) and transmittance I_T . (B) Registered target Cresyl violet staining for reference with annotations of cortical layers I-VI and white matter (wm). (C) Predicted virtual stainings of GAN and Gram models and extended variants using MSE online registration (GAN+Reg, Gram+Reg). (D) GLI images ranging from the pial surface to the cortex/white matter transition and average profiles (red lines) for layers II - VI.

For each of the GLI images, 31 intensity profiles vertically oriented to the cortical layers are extracted to represent the columnar distribution of GLI values. The profiles are restricted to layers II - VI since staining inhomogeneities make GLI values for layer I unreliable. To obtain one representative profile for each image, the profiles are averaged and smoothed using a mean filter with kernel size 3 to improve the signal-to-noise ratio. The average profiles are shown on top of each GLI image in Fig. 8D.

The average GLI profile for the GAN model shows no clear distinction between cortical layers II-IV and misses a peak for layer IV. The profile by Gram replicates peaks for layers II and IV but does not provide a clear representation of the cortical layers. GAN+Reg and Gram+Reg both replicate peaks for the higher cell densities in layers II, IV and VI. Gram+Reg shows the clearest distinction between layers and even replicates slightly nuanced peaks of the profile within layers.

Table 3: Different values for weighting parameter λ show that a balance between style loss \mathcal{L}_S and reconstruction loss \mathcal{L}_R is required in the training of a Gram+Reg model. Setting $\lambda = 0$ means only \mathcal{L}_S is used. Setting $\lambda = 1$ means only \mathcal{L}_R is used.

Method	λ	MI \uparrow	RMSE \downarrow	SSIM \uparrow	F1 \uparrow
Gram+Reg	0	0.124	33.8	0.340	48.7
	0.03	0.162	31.9	0.394	57.7
	0.25	0.210	30.7	0.436	62.8
	0.5	0.226	29.8	0.444	63.7
	0.75	0.238	29.8	0.448	63.2
	0.97	0.230	30.7	0.446	59.6
	1.0	0.090	38.0	0.347	0.0

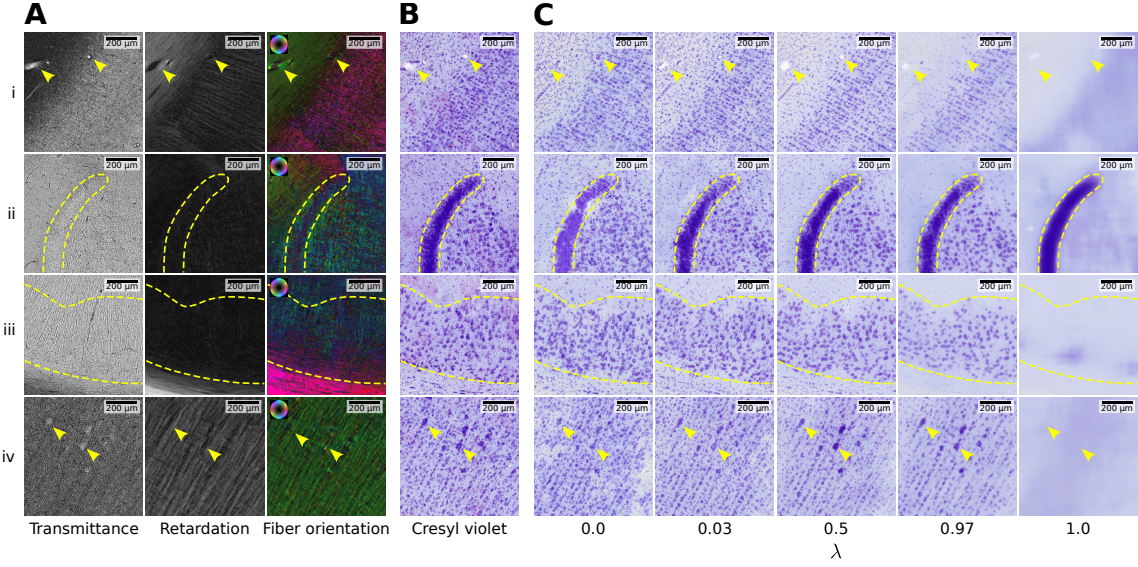


Figure 9: Predicted virtual Cresyl violet stainings for different weightings between reconstruction and style loss components. (A) 3D-PLI inputs as transmittance I_T , retardation $\sin \delta$ and fiber orientation in HSV color space (hue: direction φ ; saturation, brightness: retardation $\sin \delta$). The retardation has been scaled using gamma correction for visualization. (B) Registered target Cresyl violet staining. (C) Predicted virtual stainings of models trained with different weightings λ between reconstruction loss \mathcal{L}_R and style loss \mathcal{L}_S . Setting $\lambda = 0$ focuses on \mathcal{L}_S and setting $\lambda = 1$ on \mathcal{L}_R exclusively. Gram loss is used for \mathcal{L}_S . Within each row, yellow markers indicate the same structures across all columns. (i) Temporal area TE and underlying white matter with arrowheads highlighting blood vessels within white matter and cortical layer VI. (ii) Dentate gyrus of the hippocampus with a dashed line delineating the proximal end of the granular layer of the fascia dentata (FD). (iii) Pyramidal layer of the cornu Ammonis (CA1) region of the hippocampus highlighted by dashed lines. (iv) Layer V of the primary motor cortex with arrowheads highlighting two Betz cells.

3.4 Optimal balancing of style and content

To identify an optimal weighting λ of the style loss in Eq. (10), we train Gram+Reg models with different values for λ and MSE online registration metric. A quantitative evaluation of the trained models in Tab. 3 shows that larger values for λ up to $\lambda = 0.75$ achieve the best MI, RMSE, and SSIM scores. This indicates that the reconstruction loss should be given a stronger weight but should not be used alone ($\lambda = 1$). The highest overlap of predicted cell instances measured by the F1 score is obtained for a balanced weighting with $\lambda = 0.5$.

A visual comparison of predictions of Gram+Reg models trained with different choices for λ is shown in Fig. 9 for selected crops from the test data. By using \mathcal{L}_R only (i.e. $\lambda = 1$), the model is unable to reconstruct details from the Cresyl violet staining. It is only able to locate strongly pronounced structures such as the granular layer delineated in (ii). Independent of the style loss weighting, the models do not resolve individual cell instances present within the granular layer. Instead, they capture the overall cell density, represented by a

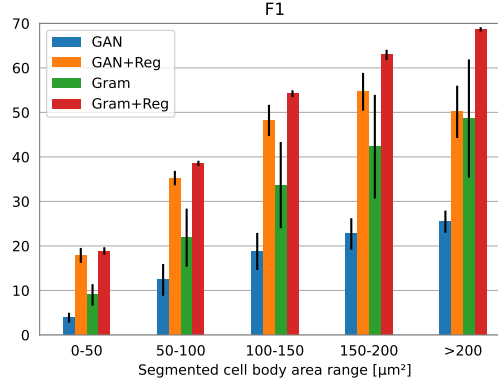


Figure 10: Reliability of predicted cell instances increases with their size. We report F1 scores between detected cell instances by a CPN model in the predicted and target stainings. F1 scores are reported for intervals of the segmented in-plane cell body area. Each bar represents the average over four independently trained models. Error bars show standard deviation.

continuous dark purple color. By increasing emphasis on \mathcal{L}_S (i.e. with decreasing values for λ), the generator is able to reconstruct details such as neuronal cell bodies in less dense regions in (i)-(iv) and blood vessels in (i). It is also able to generate Glial cells to match the appearance of white matter in (i) and (iii). However, generators trained with an overweighting of style loss also tend to miss some strongly pronounced structures in the predictions, such as the blood vessels shown in (i) or the Betz cells in (iv).

3.5 Reliability of predicted cell instances across cell sizes

Larger cells are expected to be more pronounced in 3D-PLI parameter maps compared to smaller cells that may be overshadowed by other tissue components such as nerve fibers. To quantify the reliability of predicted cells in the virtual staining in relation to their size, we compute F1 scores for multiple bins of 2D in-plane cell sizes. Obtaining F1 scores for each bin requires computation of true and false positives, as well as false negatives through matching of predicted and target cells with a minimum IoU threshold of 0.3. We modify the computation of true and false positives by matching only predicted cells within that range with all cells in the target image. To count false negatives, target cells within that range are matched with all cells in the prediction. Cell sizes are corrected with the 2D shrinkage factor of 0.97 estimated in Sec. 2.1.2, assuming homogeneous shrinkage of cells and surrounding tissue.

Results in Fig. 10 show that smaller cells $< 50 \mu\text{m}$ have much lower F1 scores, below 20.0, with all methods. With increasing 2D cell size, they can be identified more accurately. The Gram+Reg model shows highest F1 scores across all cell sizes and smallest variation between four independently trained models, measured as standard deviation.

4 Discussion

We introduced an online registration head capable of approximating smooth nonlinear deformations, given sufficiently small patch sizes for training. Its accuracy is inherently influenced by the choice of the registration metric. We identified MSE in the Fourier domain through multiple applications of cross-correlation as superior to other typically used metrics in Fourier-based image registration, including PC. This is a notable observation, as PC is a popular choice due to its robustness to intensity variations and frequency-dependent noise (Tong et al., 2019). However, during early training stages, model predictions are often blurred. The network tends to learn coarse, low-frequency structures first. Additionally, random weight initialization introduces noise in the first steps. MSE appears more resilient to such blur and frequency-independent noise, leading to more accurate predictions (Fig. 6). This behavior can be explained by PC’s reliance on strong high-frequency spectral peaks, which are diminished by blur (Pedone et al., 2013). MSE, in contrast, maintains sensitivity across the frequency spectrum and can better handle smooth, low-detail inputs that dominate during initial training.

The applied pixel-wise reconstruction loss assumes a precise alignment to provide an informative feedback signal. However, the online registration head will typically only succeed if a good contrast between structures is being predicted. In other words, there is a cross-dependency of accurate pixel reconstructions and producing a high contrast over training iterations, which may result in a conflict during training. This can be seen in Fig. 9, where a Gram+Reg model focussing on pixel reconstruction only made heavily blurred predictions that do not provide structural details for cell-precise registration. Adding a style loss component leads to more pronounced contours while simultaneously achieving higher pixel reconstruction accuracy (Tab. 3) through online registration. This underpins the importance of the additional style loss, which does not depend on perfect alignment and guides the training to produce structures.

Apparently, the choice for Gram or GAN as style loss should be considered in the light of the application. Training with Gram loss led to more accurate cell localization and pixel values (Tab. 2). However, it was affected by staining intensity inhomogeneities, resulting in blurry cells (Fig. 8). GAN loss, on the other hand, produced higher contrast in cells but with lower accuracy. This could be a consequence of the Gram loss referring to the specific target image, while the GAN loss refers to the average scoring of the discriminator across the distribution, less than the precision with respect to an individual image.

The choice of style loss function also impacts the stability and complexity of the training process. When using GAN loss, the training process has to be carefully monitored. If training collapses, restarting is often necessary. It requires careful balancing of discriminator and generator capacities and tuning of style and reconstruction loss weighting. Training with Gram loss, on the other hand, is generally more convenient to train. It requires no balancing of model capacities, is robust to loss balancing, and does not collapse during training.

Throughout all models, we observed occasional staining inhomogeneities in the predicted virtual stainings. This effect was particularly pronounced in the form of stripe patterns in the primary motor cortex ROI (Fig. 5). Cell body membranes and blood vessel walls pose hard edges in the cortex, resulting in diffraction patterns even brighter than the background field. Varying of the focus level of the LMP-1 microscope allows optical scanning of the Poisson spots induced by diffraction. Cell bodies are highlighted with higher contrast by shifting and lowering the focal plane to the top and the bottom of the tissue by approximately $\pm 30 \mu\text{m}$ (Fig. 11). This means that the ability to predict cells from 3D-PLI measurements is sensitive to the focus level of the microscope, and it is difficult to generalize to unseen focus levels that were not part of the model training.

During the mounting process of the $60 \mu\text{m}$ thick sections in 20% glycerol solution on the object slide the gray matter is more prone to swelling than the more compact white matter, leading to tissue expansion, mechanical stress and tape flutter especially in the cortex due to its higher flexibility. The free-floating tissue between object slide and coverslip shapes waves of up to $30 \mu\text{m}$ in amplitude which affects the focus level and the light incidence on the tissue, hence the effective birefringence, modifying the polarization state of the transmitted light. Therefore, the stripe-like inhomogeneities in the virtual staining can be explained by such variations in focus levels within the same section.

An effective countermeasure against waves in the tissue is to weight the open coverslips after embedding overnight. In addition, the acquisition of multiple focus levels for each section could potentially provide missing information. For the given dataset, however, these potential solutions were not applicable, and measurements could not be easily repeated. In addition, the acquisition of multiple focus levels would significantly increase the measurement effort. Nevertheless, it will be an important consideration for future investigations to include multiple focus levels in each measurement to improve cell extraction.

While most of the larger cells ($>100 \mu\text{m}^2$ in-plane size) could be localized in 3D-PLI parameter maps, smaller cells were often missed or misplaced in the predictions. This is expected as smaller cell structures are often dominated by a stronger signal of intersecting nerve fibers or overlay each other due to the high section thickness of $60 \mu\text{m}$ required for 3D-PLI. The predictions also showed a stronger arrangement of cells into cortical columns than in the target images. The presence or absence of this arrangement is one of the criteria used to identify cortical areas, which might impair downstream interpretation of the predicted architecture. It may be attributed to the pronounced signal of radial fibers in 3D-PLI, which are used by the virtual staining model to infer potential cell locations. Further investigations are necessary to understand this effect.

Our ablation experiments in Sec. 3.2, Sec. 3.3 and Sec. 3.4 support the design decisions of the proposed virtual staining method. They showed an overall good reconstruction and localization of individual cell bodies. But, of course, the virtual stainings can not yet fully replace real measurements. They occasionally include staining inhomogeneities, miss smaller cells, or introduce implausible cytoarchitectonic features, such as artificial columnar arrangements of cells. In contrast, extracting GLI profiles from the predictions demonstrated that the

laminar cell organization can be reconstructed quite well. However, further investigations with larger datasets spanning multiple brain areas are necessary to thoroughly assess their reliability.

The proposed methodology might be further improved by extending the amount of training sections, brain regions and focus levels in 3D-PLI acquisition. We also expect that the expressiveness of the Gram matrix representation could be enhanced by replacing the VGG encoder, pre-trained on ImageNet, with a domain-specific model, which is powerful on microscopy data. This is technically motivated by the observation that higher-level features in ImageNet-pretrained networks, such as VGG, are optimized for object semantics in natural images (e.g., faces, animals, vehicles), which are irrelevant in the context of fine-grained texture characteristic of microscopy images (Stuckner et al., 2022). In contrast, microscopy-specific encoders are better suited to capture such low- to mid-level textural features and have been shown to improve clustering (Oberstrass et al., 2024), segmentation (Stuckner et al., 2022), and the evaluation of generative models (Kropp et al., 2024) over those trained on ImageNet. Therefore, computation of Gram matrices using domain-specific encoders (Spitzer et al., 2017; Schiffer et al., 2021b,a) or emerging foundation models for histology (Tran et al., 2024) might further enhance representation of cytoarchitectural characteristics.

5 Conclusions

Motivated by previous observations that larger cells are encoded in 3D-PLI parameter maps alongside fiber orientations (Zeineh et al., 2017), we introduced a deep learning model for transforming 3D-PLI maps into virtual Cresyl violet cell body stainings. This approach enables joint visualization of fiber tracts and cell bodies in the same tissue. Compared to real post-staining, the model offers a scalable alternative that avoids manual labor, limited only by its generalization to unseen architectural patterns. Validation against post-stained test sections confirms that our virtual stainings reliably localize most of the larger cell bodies ($>100\text{ }\mu\text{m}^2$ in-plane size).

A central contribution of our approach is the integration of an online registration head during training. This component eliminates the need for explicit multimodal registration, which is commonly required in virtual staining pipelines (Rivenson et al., 2020; de Haan et al., 2021; Yang et al., 2022). It is a simple but highly efficient add-on that can be combined with various loss formulations, leveraging model-estimated landmarks, to continuously refine the alignment over time.

The method enables accurate prediction of cell bodies from 3D-PLI and successful adaptation to the appearance of real Cresyl violet stainings. As such, it expands the usability of 3D-PLI in large-scale data settings, allowing virtual staining at scale. While such synthetic data cannot and should not replace real histological measurements, it offers promising opportunities in downstream analysis. Potential applications include cross-modal image registration to align real Cresyl violet and 3D-PLI images, disentangling signal sources to identify voxels dominated by cell bodies (e.g., for improved tractography), and missing data imputation in serial section stacks. Especially in interleaved modalities, this could enable the reconstruction of complete datasets. Of course, such applications require careful quality control and distinction between synthetic and real measurements.

The outcomes of this study lay the groundwork for prospective investigations focused on enhancing 3D-PLI analysis, particularly through the exploration of dedicated cell detection techniques to directly extract cell body instances from 3D-PLI data. Furthermore, the findings serve as a motivation for gathering additional training data, aiming to refine and extend the application of the virtual Cresyl violet staining to a broader range of sections, brains, and species. While the current model is trained specifically to replicate Cresyl violet stainings, the same methodology can in principle be adapted to other staining types, provided appropriate retraining is performed. Future research should focus on further investigating its transferability to other datasets, staining protocols, and brains.

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.

Data and Code Availability

The training pipeline for presented Gram+Reg and GAN+Reg models is available on GitHub⁵.

Code for the online registration⁶, data augmentations for 3D-PLI images⁷, visualization methods for 3D-PLI modalities⁸, as well as additional dependencies⁹ are hosted on our external GitLab server.

The ROIs employed for the training and testing of the models in this study, along with a selection of model predictions, are available in our central institutional repository (Oberstrass et al., 2025).

Author Contributions

Alexander Oberstrass: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Esteban Vaca:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Eric Upschulte:** Software, Formal analysis, Writing - review & editing, Visualization. **Meiqi Niu:** Validation, Data curation, Investigation, Writing - review & editing. **Nicola Palomero-Gallagher:** Validation, Investigation, Writing - review & editing. **David Graessel:** Validation, Investigation, Writing - review & editing. **Christian Schiffer:** Conceptualization, Methodology, Software, Writing - review & editing. **Markus Axer:** Conceptualization, Validation, Investigation, Supervision, Writing - review & editing, Project administration, Funding acquisition. **Katrin Amunts:** Conceptualization, Validation, Investigation, Supervision, Writing - Review & Editing, Resources, Project administration, Funding acquisition. **Timo Dickscheid:** Conceptualization, Methodology, Supervision, Writing - original draft, Writing - Review & Editing, Project administration, Funding acquisition.

Conflict of Interest Statement

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.

Funding

This project received funding from the Helmholtz Association’s Initiative and Networking Fund through the Helmholtz International BigBrain Analytics and Learning Laboratory (HIBALL) under the Helmholtz International Lab grant agreement InterLabs-0015, the Helmholtz Association portfolio theme “Supercomputing and Modeling for the Human Brain”, the Priority Program 2041 (SPP 2041) "Computational Connectomics" of the German Research Foundation (DFG), the Federal Ministry of Education and Research (BMBF) under project number 01GQ1902, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; PA 1815/1-1), and the European Union’s Horizon 2020 Research and Innovation Programme, grant agreement 945539 (HBP SGA3), which is now continued in the European Union’s Horizon Europe Programme, grant agreement 101147319 (EBRAINS 2.0 Project). Computing time was granted through JARA on the supercomputer JURECA at Jülich Supercomputing Centre (JSC). Vervet monkey research was supported by the National Institutes of Health under grant agreements R01MH092311 and P40OD010965.

Acknowledgments

The authors gratefully thank Roger Woods (David Geffen School of Medicine, UCLA, USA) and Karl Zilles (INM-1, Forschungszentrum Jülich GmbH, Germany) for their collaboration in the vervet brain project, and the lab team of the INM-1 for preparing the brain sections. We also sincerely thank the members of the Fiber Architecture and Big Data Analytics groups of the INM-1 for their constructive discussions and thoughtful insights throughout the development of this work.

⁵<https://github.com/FZJ-INM1-BDA/pli2cells>

⁶https://jugit.fz-juelich.de/inm-1/bda/software/data_processing/ffreg

⁷https://jugit.fz-juelich.de/inm-1/bda/software/data_processing/pli-transforms

⁸https://jugit.fz-juelich.de/inm-1/bda/software/data_processing/pli-styles

⁹<https://jugit.fz-juelich.de/inm-1/bda/software>

References

- Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N., Habel, U., Schneider, F., and Zilles, K. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. *Anatomy and Embryology*, 210(5):343–352.
- Amunts, K., Lepage, C., Borgeat, L., Mohlberg, H., Dickscheid, T., Rousseau, M.-É., Bludau, S., Bazin, P.-L., Lewis, L. B., Oros-Peusquens, A.-M., Shah, N. J., Lippert, T., Zilles, K., and Evans, A. C. (2013). BigBrain: An Ultrahigh-Resolution 3D Human Brain Model. *Science*, 340(6139):1472–1475.
- Amunts, K., Mohlberg, H., Bludau, S., and Zilles, K. (2020). Julich-Brain: A 3D probabilistic atlas of the human brain’s cytoarchitecture. *Science*, 369(6506):988–992.
- Amunts, K. and Zilles, K. (2015). Architectonic Mapping of the Human Brain beyond Brodmann. *Neuron*, 88(6):1086–1107.
- Arganda-Carreras, I., Sorzano, C. O. S., Marabini, R., Carazo, J. M., Ortiz-de-Solorzano, C., and Kybic, J. (2006). Consistent and Elastic Registration of Histological Sections Using Vector-Spline Regularization. In Beichel, R. R. and Sonka, M., editors, *Computer Vision Approaches to Medical Image Analysis*, Lecture Notes in Computer Science, pages 85–95, Berlin, Heidelberg. Springer.
- Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 214–223. PMLR.
- Axer, M. and Amunts, K. (2022). Scale matters: The nested human connectome. *Science*, 378(6619):500–504.
- Axer, M., Amunts, K., Grässel, D., Palm, C., Dammers, J., Axer, H., Pietrzyk, U., and Zilles, K. (2011a). A novel approach to the human connectome: Ultra-high resolution mapping of fiber tracts in the brain. *NeuroImage*, 54(2):1091–1101.
- Axer, M., Graessel, D., Kleiner, M., Dammers, J., Dickscheid, T., Reckfort, J., Huetz, T., Eiben, B., Pietrzyk, U., Zilles, K., and Amunts, K. (2011b). High-Resolution Fiber Tract Reconstruction in the Human Brain by Means of Three-Dimensional Polarized Light Imaging. *Frontiers in Neuroinformatics*, 5.
- Axer, M., Gräbel, D., Palomero-Gallagher, N., Takemura, H., Jorgensen, M. J., Woods, R., and Amunts, K. (2020). Images of the nerve fiber architecture at micrometer-resolution in the vervet monkey visual system.
- Bielschowsky, M. (1904). Die Silberimpregnations der Neurofibrillen. *J Psychol Neurol*, 3:169.
- 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., Phatak, A., Kessler, I., Wang, J., Lee, J., Mehta, S., Yang, Z., O’Donnell, G., and zlapp (2022). Lightning-ai/lightning-bolts: Minor patch release. Zenodo.
- Caspers, S. and Axer, M. (2019). Decoding the microstructural correlate of diffusion MRI. *NMR in Biomedicine*, 32(4):e3779.
- Caspers, S., Axer, M., Gräbel, D., and Amunts, K. (2022). Additional fiber orientations in the sagittal stratum—noise or anatomical fine structure? *Brain Structure and Function*, 227(4):1331–1345.
- Christiansen, E. M., Yang, S. J., Ando, D. M., Javaherian, A., Skibinski, G., Lipnick, S., Mount, E., O’Neil, A., Shah, K., Lee, A. K., Goyal, P., Fedus, W., Poplin, R., Esteva, A., Berndl, M., Rubin, L. L., Nelson, P., and Finkbeiner, S. (2018). In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. *Cell*, 173(3):792–803.e19.
- Cohen, J. P., Luck, M., and Honari, S. (2018). Distribution Matching Losses Can Hallucinate Features in Medical Image Translation. In Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C., and Fichtinger, G., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Lecture Notes in Computer Science, pages 529–536, Cham. Springer International Publishing.
- Cross-Zamirski, J. O., Mouchet, E., Williams, G., Schönlieb, C.-B., Turkki, R., and Wang, Y. (2022). Label-free prediction of cell painting from brightfield images. *Scientific Reports*, 12(1):10001.
- Curl, C. L., Bellair, C. J., Harris, P. J., Allman, B. E., Roberts, A., Nugent, K. A., and Delbridge, L. M. (2004). Quantitative phase microscopy: A new tool for investigating the structure and function of unstained live cells. *Clinical and Experimental Pharmacology and Physiology*, 31(12):896–901.
- de Haan, K., Zhang, Y., Zuckerman, J. E., Liu, T., Sisk, A. E., Diaz, M. F. P., Jen, K.-Y., Nobori, A., Liou, S., Zhang, S., Riahi, R., Rivenson, Y., Wallace, W. D., and Ozcan, A. (2021). Deep learning-based transformation of H&E stained tissues into special stains. *Nature Communications*, 12(1):4884.

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.
- Ding, S.-L., Royall, J. J., Sunkin, S. M., Ng, L., Facer, B. A., Lesnar, P., Guillozet-Bongaarts, A., McMurray, B., Szafer, A., Dolbeare, T. A., Stevens, A., Tirrell, L., Benner, T., Caldejon, S., Dalley, R. A., Dee, N., Lau, C., Nyhus, J., Reding, M., Riley, Z. L., Sandman, D., Shen, E., van der Kouwe, A., Varjabedian, A., Write, M., Zollei, L., Dang, C., Knowles, J. A., Koch, C., Phillips, J. W., Sestan, N., Wahnoutka, P., Zielke, H. R., Hohmann, J. G., Jones, A. R., Bernard, A., Hawrylycz, M. J., Hof, P. R., Fischl, B., and Lein, E. S. (2016). Comprehensive cellular-resolution atlas of the adult human brain. *Journal of Comparative Neurology*, 524(16):3127–3481.
- Fienup, J. R. (1997). Invariant error metrics for image reconstruction. *Applied Optics*, 36(32):8352–8357.
- Gatys, L., Ecker, A. S., and Bethge, M. (2015). Texture Synthesis Using Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.
- Gonzales, R. C. and Woods, R. E. (2008). *Digital Image Processing*. Pearson Education.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.
- Guo, S.-M., Yeh, L.-H., Folkesson, J., Ivanov, I. E., Krishnan, A. P., Keefe, M. G., Hashemi, E., Shin, D., Chhun, B. B., Cho, N. H., Leonetti, M. D., Han, M. H., Nowakowski, T. J., and Mehta, S. B. (2020). Revealing architectural order with quantitative label-free imaging and deep learning. *eLife*, 9:e55502.
- Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML’15*, pages 448–456, Lille, France. JMLR.org.
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-To-Image Translation With Conditional Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1125–1134.
- Kingma, D. P. and Ba, J. (2017). Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR*.
- Klüver, H. and Barrera, E. (1953). A Method for the Combined Staining of Cells and Fibers in the Nervous System*. *Journal of Neuropathology & Experimental Neurology*, 12(4):400–403.
- Kropp, J.-O., Schiffer, C., Amunts, K., and Dickscheid, T. (2024). Denoising Diffusion Probabilistic Models for Image Inpainting of Cell Distributions in The Human Brain. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–5.
- Kuglin, C. D. (1975). The phase correlation image alignment method. In *Proc. International Conference on Cybernetics and Society, 1975*, pages 163–165.
- Menzel, M., Axer, M., De Raedt, H., Costantini, I., Silvestri, L., Pavone, F. S., Amunts, K., and Michielsen, K. (2020). Toward a High-Resolution Reconstruction of 3D Nerve Fiber Architectures and Crossings in the Brain Using Light Scattering Measurements and Finite-Difference Time-Domain Simulations. *Physical Review X*, 10(2):021002.
- Menzel, M., Reuter, J. A., Gräbel, D., Costantini, I., Amunts, K., and Axer, M. (2022). Automated computation of nerve fibre inclinations from 3D polarised light imaging measurements of brain tissue. *Scientific Reports*, 12(1):1–14.
- Mirza, M. and Osindero, S. (2014). Conditional Generative Adversarial Nets. *arXiv preprint*.
- Nieuwenhuys, R. (2013). The myeloarchitectonic studies on the human cerebral cortex of the Vogt-Vogt school, and their significance for the interpretation of functional neuroimaging data. *Brain Structure & Function*, 218(2):303–352.
- Novotny, G. E. K. and Novotny, E. (1977). Triple Staining of Normal and Degenerating Nervous Tissue. *Stain Technology*, 52(2):97–99.
- Oberstrass, A., Muenzing, S. E., Niu, M., Palomero-Gallagher, N., Schiffer, C., Axer, M., Amunts, K., and Dickscheid, T. (2024). Self-supervised representation learning for nerve fiber distribution patterns in 3D-PLI. *Imaging Neuroscience*, 2:1–29.
- Oberstrass, A., Vaca-Cerda, E., Upschulte, E., Niu, M., Palomero-Gallagher, N., Graessel, D., Schiffer, C., Axer, M., Amunts, K., and Dickscheid, T. (2025). Replication Data for: From Fibers to Cells: Fourier-Based Registration Enables Virtual Cresyl Violet Staining From 3D Polarized Light Imaging. Jülich DATA.

- Ojansivu, V. and Heikkilä, J. (2007). Image Registration Using Blur-Invariant Phase Correlation. *IEEE Signal Processing Letters*, 14(7):449–452.
- Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F., and Johnson, G. R. (2018). Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nature Methods*, 15(11):917–920.
- Park, Y., Depeursinge, C., and Popescu, G. (2018). Quantitative phase imaging in biomedicine. *Nature Photonics*, 12(10):578–589.
- 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., Bai, J., and Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.
- Pedone, M., Flusser, J., and Heikkilä, J. (2013). Blur invariant translational image registration for n -fold symmetric blurs. *IEEE Transactions on Image Processing*, 22(9):3676–3689.
- Reddy, B. and Chatterji, B. (1996). An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5(8):1266–1271.
- Rivenson, Y., de Haan, K., Wallace, W. D., and Ozcan, A. (2020). Emerging Advances to Transform Histopathology Using Virtual Staining. *BME Frontiers*, 2020:9647163.
- Rivenson, Y., Wang, H., Wei, Z., de Haan, K., Zhang, Y., Wu, Y., Günaydin, H., Zuckerman, J. E., Chong, T., Sisk, A. E., Westbrook, L. M., Wallace, W. D., and Ozcan, A. (2019). Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nature Biomedical Engineering*, 3(6):466–477.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham. Springer International Publishing.
- Schiffer, C., Amunts, K., Harmeling, S., and Dickscheid, T. (2021a). Contrastive Representation Learning For Whole Brain Cytoarchitectonic Mapping In Histological Human Brain Sections. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 603–606.
- Schiffer, C., Spitzer, H., Kiwitz, K., Unger, N., Wagstyl, K., Evans, A. C., Harmeling, S., Amunts, K., and Dickscheid, T. (2021b). Convolutional neural networks for cytoarchitectonic brain mapping at large scale. *NeuroImage*, 240:118327.
- Schleicher, A., Amunts, K., Geyer, S., Kowalski, T., Schormann, T., Palomero-Gallagher, N., and Zilles, K. (2000). A stereological approach to human cortical architecture: Identification and delineation of cortical areas. *Journal of Chemical Neuroanatomy*, 20(1):31–47.
- Schleicher, A., Zilles, K., and Kretschmann, H. J. (1978). Automatische Registrierung und Auswertung eines Grauwertindex in histologischen Schnitten. *Verh Anat Ges*, 72:413–415.
- Sheng, Y. (1989). Fourier-Mellin Spatial Filters For Invariant Pattern Recognition. *Optical Engineering*, 28(5):494–500.
- Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint*.
- Spitzer, H., Amunts, K., Harmeling, S., and Dickscheid, T. (2017). Parcellation of visual cortex on high-resolution histological brain sections using convolutional neural networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 920–923.
- Stuckner, J., Harder, B., and Smith, T. M. (2022). Microstructure segmentation with deep learning encoders pre-trained on a large microscopy dataset. *npj Computational Materials*, 8(1):1–12.
- Takemura, H., Palomero-Gallagher, N., Axer, M., Gräbel, D., Jorgensen, M. J., Woods, R., and Zilles, K. (2020). Anatomy of nerve fiber bundles at micrometer-resolution in the vervet monkey visual system. *eLife*, 9:e55444.
- 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.

- Tong, X., Ye, Z., Xu, Y., Gao, S., Xie, H., Du, Q., Liu, S., Xu, X., Liu, S., Luan, K., and Stilla, U. (2019). Image Registration With Fourier-Based Image Correlation: A Comprehensive Review of Developments and Applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(10):4062–4081.
- Tran, M., Schmidle, P., Wagner, S. J., Koch, V., Novotny, B., Lupperger, V., Feuchtinger, A., Böhner, A., Kaczmarczyk, R., Biedermann, T., Comfere, N. I., Guo, R. R., Wang, C., Eyerich, K., Braun, S. A., Peng, T., and Marr, C. (2024). Generating clinical-grade pathology reports from gigapixel whole slide images with HistoGPT.
- Upschulte, E., Harmeling, S., Amunts, K., and Dickscheid, T. (2022). Contour proposal networks for biomedical instance segmentation. *Medical Image Analysis*, 77:102371.
- Upschulte, E., Harmeling, S., Amunts, K., and Dickscheid, T. (2023). Uncertainty-Aware Contour Proposal Networks for Cell Segmentation in Multi-Modality High-Resolution Microscopy Images. In *Proceedings of The Cell Segmentation Challenge in Multi-modality High-Resolution Microscopy Images*, pages 1–12. PMLR.
- Wagstyl, K., Larocque, S., Cucurull, G., Lepage, C., Cohen, J. P., Bludau, S., Palomero-Gallagher, N., Lewis, L. B., Funck, T., Spitzer, H., Dickscheid, T., Fletcher, P. C., Romero, A., Zilles, K., Amunts, K., Bengio, Y., and Evans, A. C. (2020). BigBrain 3D atlas of cortical layers: Cortical and laminar thickness gradients diverge in sensory and motor cortices. *PLOS Biology*, 18(4):e3000678.
- Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- Yadan, O. (2019). Hydra - a framework for elegantly configuring complex applications. Github.
- Yang, X., Bai, B., Zhang, Y., Li, Y., de Haan, K., Liu, T., and Ozcan, A. (2022). Virtual Stain Transfer in Histology via Cascaded Deep Neural Networks. *ACS Photonics*, 9(9):3134–3143.
- Zeineh, M. M., Palomero-Gallagher, N., Axer, M., Gräbel, D., Goubran, M., Wree, A., Woods, R., Amunts, K., and 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.
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232.
- Zilles, K., Schleicher, A., and Kretschmann, H.-J. (1978). Automatic morphometric analysis of retrograde changes in the nucleus n. facialis at different ontogenetic stages in the rat. *Cell and Tissue Research*, 190(2):285–299.

A Appendix

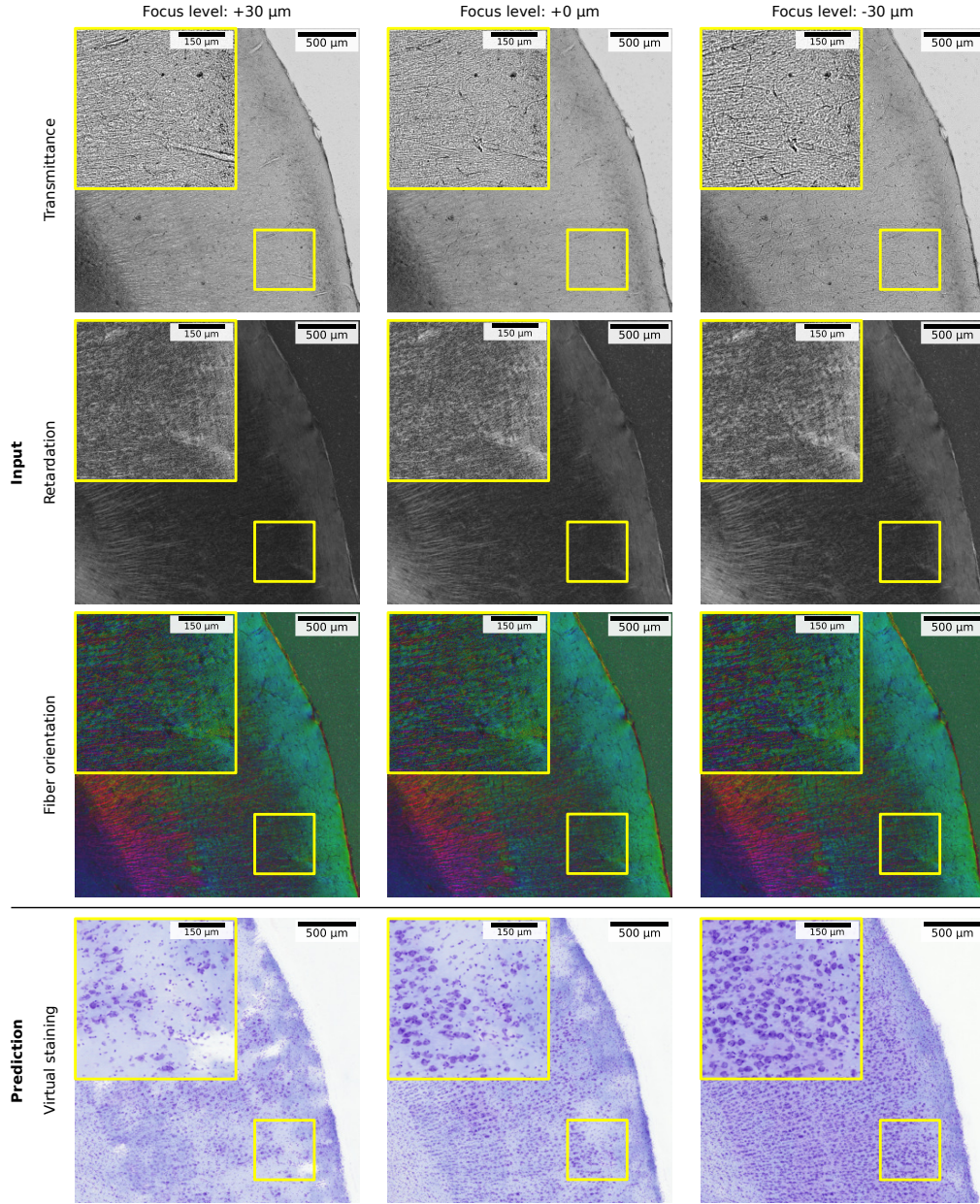


Figure 11: Virtual staining quality varies with the focus level of the LMP-1 due to differences in the captured diffraction patterns. Example images show the vervet entorhinal cortex with zoom-ins to the architecture of cortical layers II-III. 3D-PLI inputs are represented by the transmittance I_T , retardation $\sin \delta$ (scaled with gamma correction) and fiber orientation in HSV color space (hue: direction φ ; saturation, brightness: retardation $\sin \delta$). The same Gram+Reg model is used to predict virtual stainings from a focus level manually adjusted to the tissue center (+0 μm) and 30 μm above and below. A lower focus level of -30 μm achieves the highest consistency in staining intensity and produces the most cells. Higher focus levels introduce an increasing amount of staining artifacts.