Machine learning of dissection photographs and surface scanning for quantitative 3D neuropathology

Harshvardhan Gazula[1], Henry F. J. Tregidgo[2], Benjamin Billot[3], Yael Balbastre[1], Jonathan William-Ramirez[1], Rogeny Herisse[1], Lucas J Deden-Binder[1], Adrià Casamitjana[2,4], Erica J. Melief[5], Caitlin S. Latimer[5], Mitchell D. Kilgore[5], Mark Montine[5], Eleanor Robinson[2], Emily Blackburn[2], Michael S. Marshall[6], Theresa R. Connors[6], Derek H. Oakley[6], Matthew P. Frosch[6], Sean I. Young[1], Koen Van Leemput[1,7], Adrian V. Dalca[1,3], Bruce Fischl[1], Christine L. Mac Donald[8], C. Dirk Keene[5], Bradley T. Hyman[6], Juan Eugenio Iglesias[1,2,3]

***For correspondence:** jiglesiasgonzalez@mgh.harvard.edu (JEI)

[1]Martinos Center for Biomedical Imaging, MGH and Harvard Medical School, Charlestown, MA, USA; [2]Centre for Medical Image Computing, University College London, London, UK; [3]Computer Science and Artificial Intelligence Laboratory, MIT, MA, USA; [4]Biomedical Imaging Group, Universitat Politècnica de Catalunya, Spain; [5]BioRepository and Integrated Neuropathology (BRaIN) laboratory and Precision Neuropathology Core, UW School of Medicine, WA, USA; [6]Massachusetts Alzheimer Disease Research Center, MGH and Harvard Medical School, Charlestown, MA, USA; [7]Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland; [8]Department of Neurological Surgery, UW School of Medicine, WA, USA

Abstract

We present open-source tools for 3D analysis of photographs of dissected slices of human brains, which are routinely acquired in brain banks but seldom used for quantitative analysis. Our tools can: *(i)* 3D reconstruct a volume from the photographs and, optionally, a surface scan; and *(ii)* produce a high-resolution 3D segmentation into 11 brain regions per hemisphere (22 in total), independently of the slice thickness. Our tools can be used as a substitute for *ex vivo* magnetic resonance imaging (MRI), which requires access to an MRI scanner, *ex vivo* scanning expertise, and considerable financial resources. We tested our tools on synthetic and real data from two NIH Alzheimer's Disease Research Centers. The results show that our methodology yields accurate 3D reconstructions, segmentations, and volumetric measurements that are highly correlated to those from MRI. Our method also detects expected differences between *post mortem* confirmed Alzheimer's disease cases and controls. The tools are available in our widespread neuroimaging suite "FreeSurfer" (https://surfer.nmr.mgh.harvard.edu/fswiki/PhotoTools).

Introduction

Morphometric measurements such as cortical thickness and subcortical volumetry can be used as surrogate biomarkers of aging (***Salat et al., 2004***; ***Walhovd et al., 2005***; ***Coupé et al., 2017***) and disease (***Lerch et al., 2005***; ***Desikan et al., 2009***; ***Dickerson et al., 2009***; ***Blanken et al., 2017***) via confirmed histopathological changes. Integrating neuroimaging tools with neuropathological as-

sessments can enable neuropathological-neuroimaging correlations for studying neurodegenerative diseases, i.e., connecting macroscopic imaging with histological ground truth derived from microscopic imaging (***Webster et al., 2021***). While *ante mortem* MRI studies provide accurate and reliable morphometric data, they are often unavailable or occur too long before death, impeding reliable histopathological correlation. Cadaveric MRI can circumvent these challenges, but logistic and legal issues often complicate this procedure.

An alternative to cadaveric imaging is *ex vivo* MRI, which enables high-resolution image acquisition free of subject motion and other physiological noise (***Edlow et al., 2019***). However, *ex vivo* MRI also has disadvantages, including: tissue degradation due to bacteria and autolysis from death to initiation of tissue fixation; cross-linking of proteins due to fixation that dramatically changes MR properties of the tissue; and image artifacts caused by magnetic susceptibility interfaces that do not occur *in vivo* (***Shatil et al., 2016***).

While *ex vivo* MRI is relatively uncommon in brain banks that seek to establish neuropathological-neuroimaging correlations (***Ravid, 2009***; ***Love, 2005***), dissection photography is routine in nearly every brain bank. Collected specimens are typically dissected into coronal slices and photographed before further blocking and histological analysis. These photographs, often underutilized, are an invaluable information resource that, if leveraged appropriately, can play a vital role in advancing our understanding of various brain functions and disorders — mainly when *ex vivo* MRI is unavailable. To this end, we propose a novel software suite that, for the first time, enables 3D reconstruction and quantitative 3D morphometry of dissection photographs, powered by modern machine learning techniques (***Goodfellow et al., 2016***) and 3D surface scanning (***Salvi et al., 2004***). Provided that the photographs satisfy some basic requirements (presence of a ruler or fiducial markers to estimate pixel sizes), our tools enable volumetric analysis of brain structures from the photographs, computationally guided dissection via automated segmentation, and accurate spatial mapping for neuropathological-neuroimaging correlation.

Our suite is freely available and includes three modules. The first module is a set of preprocessing routines for the photographs that enables the correction of perspective and calibration of pixel sizes. The second module is a joint image registration algorithm that allows 3D reconstruction of the photographs using a 3D surface scan of the brain as a reference. This module can also use a probabilistic atlas as a reference for the reconstruction, thus circumventing the need for a surface scanner. This scan-free mode enables retrospective analysis of photographs without corresponding 3D scans, albeit with lower accuracy than if surface scanning was available. The third and final module uses machine learning to provide a high-resolution 3D image segmentation of the reconstructed stack. This module combines a state-of-the-art deep segmentation neural network (a U-Net, ***Ronneberger et al. 2015***) with a domain randomization approach (***Billot et al., 2023a***), thus enabling analysis of photographs with different intensity profiles, e.g., acquired under various illumination conditions, with different cameras or camera settings, or from fixed or fresh tissue. Moreover, the machine learning method enables the estimation of "smooth", isotropic segmentations that accurately interpolate across the gaps between the coronal planes captured in the photographs.

We note that this article extends our previous conference paper (***Tregidgo et al., 2020***) by *(i)* improving upon the 3D reconstruction methods; *(ii)* using machine learning (rather than Bayesian methods) to produce segmentations that are more accurate and also isotropic (i.e., provide labels in between slices); and *(iii)* providing extensive experiments on different synthetic and real datasets. The rest of this article is organized as follows. First, we present results on three different datasets, both synthetic (which enable fine-grained analysis with known ground truth) and real (which enables evaluation in real-world scenarios). Next, we discuss these results and their impact on quantitative *post mortem* neuroimaging without MRI. Finally, the Methods section elaborates on the technicalities of the preprocessing steps, the reconstruction algorithm, and the deep learning segmentation method.

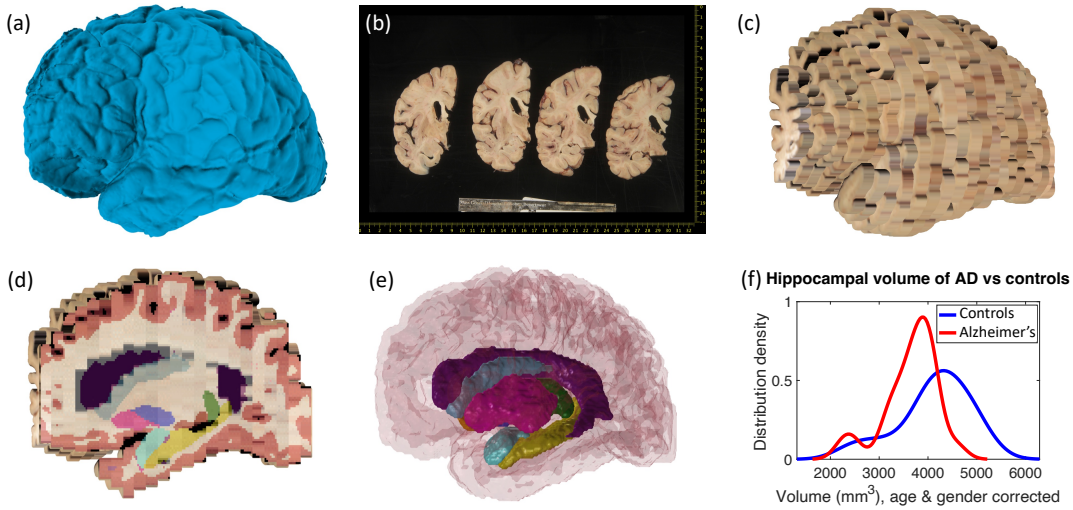


Figure 1. Examples of inputs and outputs from the MADRC dataset. **(a)** 3D surface scan of left human hemisphere, acquired prior to dissection. **(b)** Routine dissection photography of coronal slabs, after pixel calibration, with digital rulers overlaid. **(c)** 3D reconstruction of the photographs into an imaging volume. **(d)** Sagittal cross-section of the volume in (c) with the machine learning segmentation overlaid. The color code follows the FreeSurfer convention. Also, note that the input has low, anisotropic resolution due to the large thickness of the slices (i.e., rectangular pixels in sagittal view), whereas the 3D segmentation has high, isotropic resolution (squared pixels in any view). **(e)** 3D rendering of the 3D segmentation into the different brain regions, including hippocampus (yellow), amygdala (light blue), thalamus (green), putamen (pink), caudate (darker blue), lateral ventricle (purple), white matter (white, transparent), and cortex (red, transparent). **(f)** Distribution of hippocampal volumes in *post mortem* confirmed Alzheimer's disease vs controls in the MADRC dataset, corrected for age and gender.

Results

Volumetric group study of *post mortem* confirmed Alzheimer's disease

One of the main use cases of our tools is the volumetric analysis of different cerebral regions of interest (ROIs) from dissection photographs, without requiring cadaveric or *ex vivo* MRI. We used our tools to analyze 21 *post mortem* confirmed Alzheimer's disease (AD) cases from the Massachusetts Alzheimer's Disease Research Center (MADRC), as well as 12 age-matched controls (N=33 in total). We note that these cases comprise thick (~10 mm) slabs, sliced by hand, without cutting guides – as cutting on anatomic landmarks was prioritized over consistent slice thickness. Therefore, this dataset is representative of a challenging, real-world scenario.

Examples of the inputs and outputs of the pipeline can be found in Figure 1 and in Video 1 in the supplementary material (also available at https://youtu.be/wo5meYRaGUY). The 3D surface scan (Figure 1a) reveals the coarse shape of the specimen, in this case a left hemisphere. Our preprocessing tools correct for the perspective and pixel size of routine dissection photographs (b). Then, the 3D reconstruction tool uses information from the surface scan to produce a 3D reconstruction of the photographs into an imaging volume (c). This volume consists of highly anisotropic voxels, since the slice thickness is much larger than the pixel size of the photograph. The 3D reconstruction is fed to our machine learning segmentation method ("Photo-SynthSeg"), which produces a high-resolution, isotropic segmentation (d,e), independently of the slice thickness. The segmentations are then used to compare the volumes of brain ROIs between the two groups (e.g., the hippocampus, as in Figure 1f).

Table 1 shows the area under the receiver operating characteristic curve (AUROC) and the p-value for non-parametric Wilcoxon rank sum tests (***Mann and Whitney, 1947***) comparing the ROI volumes of AD vs controls; we leave the accumbens area and ventral diencephalon out of the analysis as their segmentations are not reliable due to poor contrast (***Fischl et al., 2002***). We note that the AUROC is the non-parametric equivalent of the effect size; a value of 0.5 represents chance, while 1.0 represents perfect separation. Age and gender were corrected with a general linear model,

Table 1. Area under the receiver operating characteristic curve (AUROC) and p-value of a non-parametric Wilcoxon rank sum test comparing the volumes of brain regions for Alzheimer's cases vs controls. The volumes were corrected by age and sex using a general linear model. We note that the AUROC is bounded between 0 and 1 (0.5 is chance) and is the non-parametric equivalent of the effect size (higher AUROC corresponds to larger differences). The sample size is $N = 33$.

Region	Wh Matter	Cortex	Vent	Thal	Caud	Putamen	Pallidum	Hippoc	Amyg
AUROC	.45	.52	.73	.48	.65	.64	.77	.75	. 77
p-value	.666	.418	.016	.596	.086	.092	.005	.009	.007

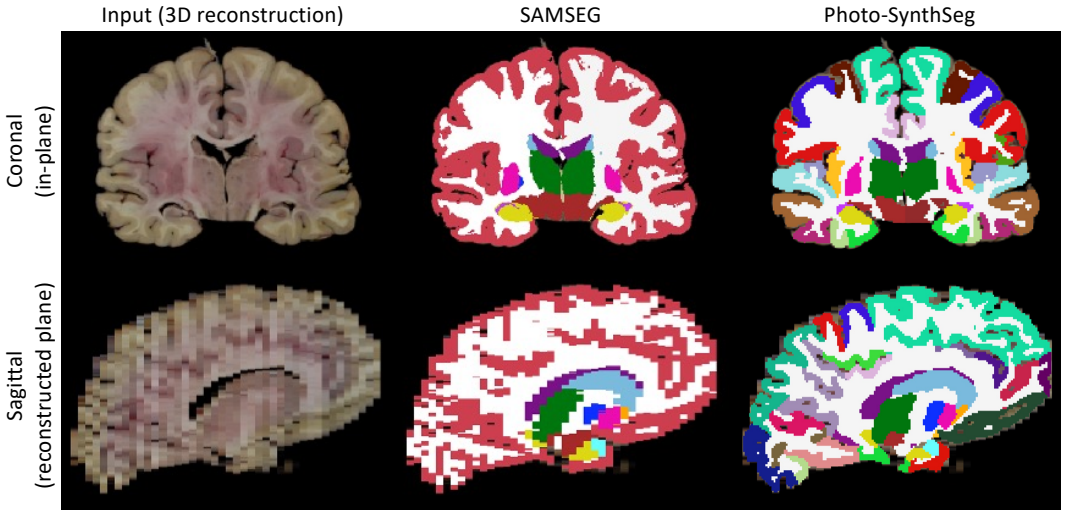


Figure 2. Qualitative comparison of SAMSEG vs Photo-SynthSeg: Coronal (top) and sagittal (bottom) views of the reconstruction and automated segmentation of a sample whole brain from the UW-ADRC dataset. Note that Photo-SynthSeg supports subdivision of the cortex with tools of the SynthSeg pipeline.

whereas volumes of contralateral ROIs were averaged when full brains (rather than hemispheres) were available. Our method successfully captures well-known atrophy patterns of AD, such as hippocampal atrophy and ventricle enlargement.

Quantitative evaluation of segmentation with Photo-SynthSeg

While the AD experiment illustrates the ability of our method to detect differences in real-world data, it is crucial to specifically assess the accuracy of our segmentation and 3D reconstruction methods. To evaluate Photo-SynthSeg, we 3D reconstructed the brain volumes from the photographs of 24 cases from the AD Research Center at the University of Washington (UW-ADRC), which were cut into uniform 4 mm thick slices. Crucially, isotropic FLAIR MRI scans were acquired for these specimens *ex vivo* prior to dissection, which enables the use of MRI as gold standard.

We compare Photo-SynthSeg against "SAMSEG", which is (to the best of our knowledge) the only available competing method. SAMSEG is a segmentation algorithm which we originally conceived for brain MRI (***Puonti et al., 2016***), and which employs Bayesian techniques to segment MRI scans irrespective of their contrast and pulse sequence. As we showed in ***Tregidgo et al.*** (***2020***), this technique can be adapted to 3D reconstructed photographs. Figure 2 shows the segmentation of an UW-ADRC case using both methods. As opposed to SAMSEG, Photo-SynthSeg effectively "interpolates" the segmentation in between slices, independently of their thickness, and is more robust against uneven intensities than the Bayesian method. Photo-SynthSeg also includes a volumetric parcellation of the cortex based on the original SynthSeg pipeline (***Billot et al., 2023b***), which relies on the Desikan-Killiany atlas (***Desikan et al., 2006***).

To evaluate Photo-SynthSeg and SAMSEG quantitatively, we computed Dice scores against manual segmentations made on a single selected slice per subject. This slice is visually chosen to be close to the mid-coronal plane, while maximizing visibility of subcortical structures; an example of such slice, along with the manual and automated segmentations, is shown in Figure 3 of Ap-

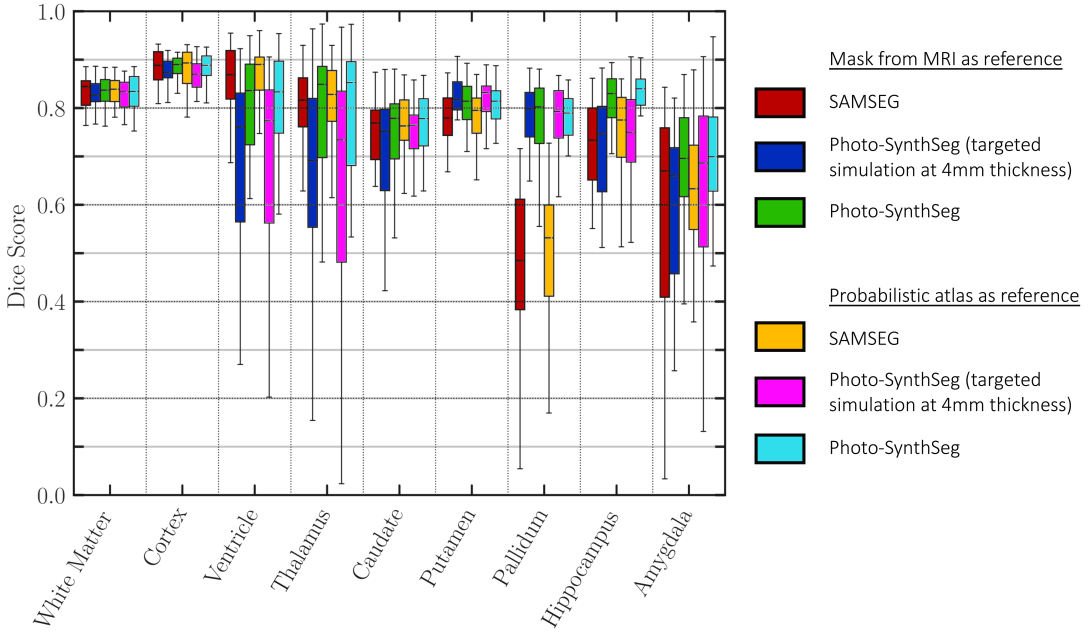


Figure 3. Dice scores of automated vs manual segmentations on select slices. Box plots are shown for SAMSEG, Photo-SynthSeg, and two ablations: use of probabilistic atlas, and targeted simulation with 4 mm slice spacing. Dice is computed in 2D, using manual segmentations on select slices. We also note that the absence of extracerebral tissue in the images contributes to high Dice for the cortex.

pendix 2. The Dice scores are displayed in Figure 3; as in the previous analysis, the accumbens area and ventral diencephalon are left out of the analysis. The figure also shows two ablations: using a probabilistic atlas instead of the case-specific reference; and using a version of Photo-SynthSeg dedicated to 4 mm slice thickness (i.e., the thickness of the UW-ADRC dataset, rather than using the general, thickness-agnostic model). The former assesses the impact of having to rely on a generic atlas when surface scans are not available, whereas the latter evaluates the ability of our neural network to adapt to thicknesses that are not known a priori (thanks to our domain randomization approach). The results show that Photo-SynthSeg generally outperforms SAMSEG, producing Dice scores over 0.8 for all structures except the amygdala – even though SAMSEG is better at segmenting the ventricles, thanks to their large size and strong contrast. The plots also show that using the probabilistic atlas produces realistic enough reconstructions, such that the 2D Dice scores remain high. Finally, the results also show the superiority of the thickness randomization strategy over the fixed spacing strategy, even if the latter had access to the ground truth spacing of the photographs. Even though the reader may initially find this result counterintuitive, it is consistent with our previous findings in MRI segmentation, where domain randomization strategies also outperformed targeted simulations (***Billot et al., 2020***).

This evaluation with Dice scores above is direct, but: *(i)* is based on a relatively small number of slices, and *(ii)* disregards the ultimate purpose of segmentation, which is downstream analysis in 3D (e.g., volumetry). For this purpose, we also indirectly evaluated the methods by analyzing the volumes of brain ROIs derived from the segmentations of the whole stack. Specifically, we correlated these volumes with silver standard values derived from the isotropic FLAIR MRI scans using our "standard" SynthSeg for MRI (***Billot et al., 2023a***). Table 2 shows the correlations and p-values for Steiger tests comparing the correlation coefficients achieved by SAMSEG and Photo-SynthSeg – while considering their dependency due to the common ground truth sample. The results show once more that Photo-SynthSeg outperforms SAMSEG for nearly every structure – and in the few cases in which it does not, the Steiger test does not yield statistically significant differences. We note that the correlations are above 0.8 for most brain structures, indicating that the 3D reconstructed photographs yield usable volumes in volumetric analysis. We also note that

Table 2. Correlations of volumes of brains regions estimated by SAMSEG and Photo-SynthSeg from the photographs against the ground truth values derived from the MRI. The p-values are for Steiger tests comparing the correlations achieved by the two methods (accounting for the common sample).

	Mask from MRI as reference			Probabilistic atlas as reference		
	SAMSEG	**Photo-SynthSeg**	***p*-value**	**SAMSEG**	**Photo-SynthSeg**	***p*-value**
White Matter	0.935	0.981	0.0011	0.886	0.935	0.0117
Cortex	0.930	0.979	0.0001	0.889	0.920	0.0366
Ventricle	0.968	0.988	0.0004	0.980	0.993	0.0006
Thalamus	0.812	0.824	0.4350	0.812	0.824	0.4252
Caudate	0.719	0.779	0.2525	0.733	0.792	0.2062
Putamen	0.904	0.779	0.9923	0.872	0.792	0.9598
Pallidum	0.727	0.694	0.6171	0.676	0.658	0.5698
Hippocampus	0.830	0.757	0.8873	0.764	0.776	0.4293
Amygdala	0.598	0.703	0.1663	0.576	0.763	0.0221

the correlations are slightly but consistently lower for the reconstructions with the probabilistic atlas, yielding correlations close to 0.8 for most brain regions.

Quantitative evaluation of reconstruction with digitally sliced MRI data

In addition to the segmentation, it is also desirable to evaluate the 3D reconstruction algorithm with the registration error (in mm). Measuring such error with real data would require manual annotation of pairs of matching landmarks, which is labor-intensive, error-prone, and not reproducible. Instead, we use simulated (digitally sliced) data created from MRI scans. While errors estimated this way may be optimistic compared with real sliced tissue, this approach enables us to analyze the error as a continuous function of slice thickness without manual annotation effort.

For this purpose, we used 500 subjects from the Human Connectome Project (HCP) dataset, which includes T1- and T2-weighted MRI scans acquired at 0.7 mm isotropic resolution. After skull stripping with FreeSurfer (***Fischl, 2012***), we simulated dissection photographs and matching surface scans by: *(i)* digitally slicing the T2 scans; and *(ii)* using the surface of the T1 as a 3D reference to reconstruct the T2 slices. Specifically, we simulated slices with $S \times 0.7$ mm thickness ($S = 2, 4, 8, 16$), including random affine transforms and illumination fields for every simulated slice (see example in Figure 1 of Appendix 2). While we could use a nonlinear model, the results would depend heavily on the strength of the simulated deformation. Instead, we keep the warps linear as we believe that the value of this experiment lies in the trends that the errors reflect, i.e., their relative rather than absolute value.

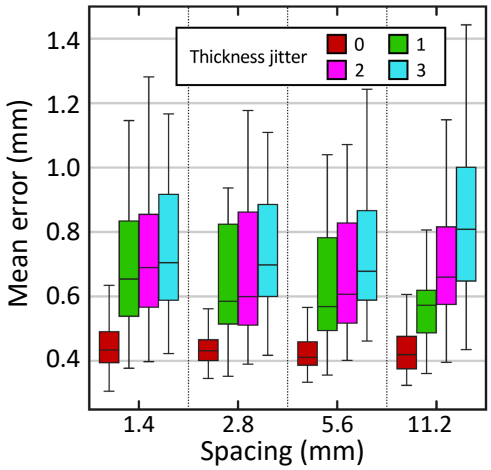


Figure 4. Reconstruction error (in mm) in synthetically sliced HCP data. The figure shows box plots for the mean reconstruction error as a function of spacing and thickness jitter. A jitter of j means that the nth slice is randomly extracted from the interval $[n-j, n+j]$ (rather than exactly n). The center of each box represents the median; the edges of the box represent the first and third quartile; and the whiskers extend to the most extreme data points not considered outliers (not shown, in order not to clutter the plot).

After digitally distorting the images, we used our method to 3D reconstruct the slices into their original shape. The registration error was calculated as the mean voxel displacement in mm between the reconstructed and ground truth T2 slices. Additionally, to test the robustness of the reconstruction algorithm to

uneven slice spacing during dissection, we also analyze the error when a random variation of the nominal thickness ("thickness jitter") is introduced for every slice.

Figure 4 shows the box plot for the mean reconstruction error as a function of the slice spacing and thickness jitter. The results show that the reconstruction error is reasonably robust to increased slice spacing. On the other hand, thickness jitter yields greater increases in reconstruction error, particularly at larger slice spacings. This result highlights the importance of keeping the slice thickness as constant as possible during acquisition.

Discussion

Neuroimaging to neuropathology correlation studies explore the relationship between gold-standard pathological diagnoses and imaging phenotypes by transferring the microscopic signatures of pathology to *in vivo* MRI. A significant impediment to this effort is the lack of quantitative tools for *post mortem* tissue analysis. For example, quantitative measurements such as cortical thickness and specific regional atrophy are often estimated qualitatively from 2D coronal slices during gross examination. A solution to this problem is leveraging dissection photography of these coronal slices routinely acquired before histology. By designing algorithms to reconstruct 3D volumes from 2D photographs and subsequently segment them, we have enabled a cost-effective and time-saving link between morphometric phenotypes and neuropathological diagnosis.

Our new tools are publicly available and utilize modern deep learning techniques and surface scanning. Being available as part of Freesurfer makes the tools easy to use by anyone with little or no training. Furthermore, the absence of tunable parameters in Photo-SynthSeg makes the results produced by our methods highly reproducible, and the robustness of the tools enables widespread application to heterogeneous datasets. This robustness has been demonstrated by applying the tools to images from two different biobanks, with different slice thickness, tissue processing, and photographic setup. On the UW-ADRC dataset, for which MRI scans were available, we achieved correlations above 0.8 between the volumes derived from the photographs and the ground truth obtained from the MRI.

Retrospective reconstruction of datasets with no accompanying surface scan available can be achieved with a probabilistic atlas. However, such a reconstruction is laden with ambiguities because the probabilistic atlas does not have access to the true shape of the tissue – which the surface scan directly measures. Whether the increased reconstruction error is tolerable depends on the downstream task, e.g., shape analysis vs volumetry.

While deep learning segmentation tools are increasingly common in medical imaging research, their practical applicability in modalities with highly varying appearance (like dissection photography) has been hindered by their limited generalization ability. Photo-SynthSeg circumvents this problem by building on our recent work on domain randomization (***Billot et al., 2023a***), and can segment 3D reconstructed stacks of photographs irrespective of the thickness of the slices and of the contrast properties of the tissue. Compared with SAMSEG, Photo-Synthseg also has the advantage of estimating the segmentation in between slices (Figure 2). Moreover, Photo-Synthseg inherits the computational efficiency of neural networks, and can segment a whole case in a few seconds (or tens of seconds if no GPU is available) without any special requirements on machine learning expertise – thus enabling broad applicability.

While registration and volumetric segmentation enable morphometry and neuropathology-neuroimaging correlation, precise white matter and pial surface placement on 3D reconstructed photographs are crucial for accurate cortical analyses – e.g., producing topologically correct segmentations, as opposed to volumetric segmentations that can, for example, leak across gyri. In the future, we will extend Photo-SynthSeg to enable surface analysis for cortical placement and parcellation. While the convoluted nature of cortical surfaces makes this task difficult for photographic volumes, integrating the triangular mesh provided by the surface scanner could enable accurate surface placement.

Another direction of future work will be extending the tools to axial and sagittal slices of the cerebellum and brainstem. While adapting the 3D reconstruction (Equation 1) is straightforward, the U-net will need additional image synthesis and manual labeling efforts – particularly if one wishes to include new regions, such as brainstem nuclei. Additional future analyses will include: correlating the segmentation-derived volumes with clinical scores, disease subtypes, and disease duration; using techniques like SynthSR (***Iglesias et al., 2023***) to improve the resolution of the reconstructed volumes; exploring nonlinear deformation models for the 3D reconstruction; fully automatizing tissue segmentation from the background using neural networks; and extending the tools to 3D analysis of histological sections.

Leveraging the vast amounts of dissection photographs available at brain banks worldwide to perform morphometry is a promising avenue for enhancing our understanding of various neurodegenerative diseases. Our new tools will allow extraction of quantitative phenotypical information from these photographs – and thus augmentation of histopathological analysis. We expect this new methodology to play a crucial role in the discovery of new imaging markers to study neurodegenerative diseases.

Materials and Methods

Datasets

MADRC: Dissection photography of fixed tissue and companion surface scans for 76 cases from the Massachusetts Alzheimer's Research Center (18 whole cerebrums and 58 hemispheres). Ruling out cases with frontotemporal dementia and other comorbidities, as well as subjects that did not pass manual quality control (by JWR, RH, and LJD), led to a final sample size of $N = 33$ (21 *post mortem* confirmed Alzheimer's and 12 controls). The surface scans were acquired using a turntable with an Einscan Pro HD scanner (Shining 3D, Hangzhou, China, 0.05 mm point accuracy). Slices with variable thickness were cut with a dissecting knife on a predefined set of landmarks and photographed with a 15.1 MP Canon EOS 50D Digital SLR camera. Further details on the dissection and processing of the specimens can be found in Appendix 1.

UW-ADRC: Dissection photography of fixed tissue and companion *ex vivo* MRI scans for 24 cases (all of them with both hemispheres) from the Alzheimer's Disease Research Center at the University of Washington. The MRI scans were acquired at 0.8 mm isotropic resolution using a FLAIR sequence. Coronal slices were cut with 4 mm thickness using a modified deli slicer and photographed with a 35 Megapixel (MP) camera. While no 3D surface scanning was available for this dataset, we obtained surfaces by skull stripping the MRI scans and meshing the brain surface. This dataset enables us to compute volumetric measurements from the 3D reconstructed photographs and compare them with reference values obtained from the corresponding MRI scans. Furthermore, two experienced labelers manually traced the contour of 9 brain regions (white matter, cortex, lateral ventricle, thalamus, caudate, putamen, pallidum, hippocampus, and amygdala) in one slice per case, which enables computation of 2D Dice scores. Further details on the dissection and processing of the specimens can be found in Appendix 1 and ***Latimer et al.*** (***2023***).

HCP: T1- and T2-weighted MRI scans of 500 subjects from the Human Connectome Project, acquired at 0.7 mm isotropic resolution. The scans were skull stripped using Freesurfer (***Fischl, 2012***). We use these scans to simulate dissection photographs and matching surface scans by digitally slicing the T2 scans and using the T1 as a 3D reference to reconstruct the T2 slices. Further details on the MRI acquisition can be found in ***Van Essen et al.*** (***2012***).

T1-39: 39 T1-weighted MRI scans at 1 mm isotropic resolution, with manual volumetric segmentations for 36 brain structures, which we used to train Photo-SynthSeg. We note that this is the labeled dataset that was used to build the probabilistic atlas in FreeSurfer (***Fischl et al., 2002***). The 36 structures include 22 that are segmented by Photo-Synthseg: left and right white matter, cortex, ventricle, thalamus, caudate, putamen, pallidum, hippocampus, amygdala, accumbens area, and ventral diencephalon. The remaining 14 labels include: four labels for the cerebellum (left and

right cortex and white matter); the brainstem; five labels for cerebrospinal fluid regions that we do not consider; the left and right choroid plexus; and two labels for white matter hypointensities in the left and right hemisphere.

Surface scanning

Surface scanning (also known as "profilometry") is a technology that is becoming increasingly inexpensive (a few thousand dollars), mainly via structured light technology (***Salvi et al., 2004***). The preferred version of our proposed pipeline relies on a surface scan of the specimen (Figure 1a) acquired before slicing. This surface scan, represented by a triangular mesh, is used as an external reference to guide the 3D reconstruction of the photographs (details below). While there are no specific technical requirements for the surface scan (e.g., minimum resolution), minimizing geometric distortion (i.e., deformation) between scanning and subsequent slicing is crucial. The surface scan may be acquired with a handheld scanner with the specimen placed on a table, or with the scanner on a tripod and the sample on a turntable. As mentioned earlier, our pipeline does not strictly require the surface scan, as it can be replaced with a probabilistic atlas with a slight loss of accuracy (as shown in the Results section above).

Tissue slicing and photography

In preparation for the downstream 3D reconstruction, some requirements exist for the dissection and photography of the brain slabs. Specifically, our pipeline assumes that approximately parallel slices are cut in coronal direction – which is the standard practice in most brain banks. We further assume that the thickness of these slices is approximately constant, as "thickness jitter" has a detrimental effect on the results (Figure 4).

Moreover, we require the presence of fiducials to enable pixel size calibration and perspective correction. Ideally, four fiducials are placed on the corners of a rectangle of known dimensions (Figure 5a). In the absence of such fiducials, we require the presence of at least a ruler or two orthogonal rulers; the former enables pixel size calibration via image scaling, whereas the latter enables approximate perspective correction by fitting an affine transform.

Our pipeline allows for multiple slices to be present in one photograph but requires that all photographs are of the same side of the slices (either anterior or posterior) and that the inferior-superior direction of the anatomy is approximately aligned with the vertical axis of the image (as in Figure 5a). Ideally, the slices should be photographed on a flat board with a solid background color that stands out from the brain tissue, as stronger contrast between tissue and background greatly facilitates image segmentation when preprocessing the photographs (further details below).

Preprocessing of photographs

Image preprocessing starts with geometric correction, i.e., pixel size calibration and perspective correction. In the ideal scenario with four fiducials, the widespread widespread scale-invariant feature transform (SIFT, ***Lowe 1999***) is used to detect such fiducials (Figure 5b) and compute a spatial transform that corrects for perspective distortion and calibrates the pixel size – see Figure 5c, where the pixel size calibration enables superimposition of a digital ruler.

Our software also supports a manual mode where the user clicks on two, three, or four landmarks: two points with a known distance in between (enables approximate pixel size calibration); three points at the ends of two rulers plus their intersection (enables approximate perspective correction with an affine transform); or four points on the corners of a rectangle of known dimensions (enables full perspective correction). This manual mode is useful when no fiducials are present, but the user can still identify features with known dimensions in the photograph, e.g., on an imprinted grid pattern, or along rulers.

After geometric correction, our methods require a binary segmentation of the brain tissue, separating it from the background. While our tools do not have any specific requisites in terms of background, using a solid background with a distinct color (e.g., a "green screen") can greatly

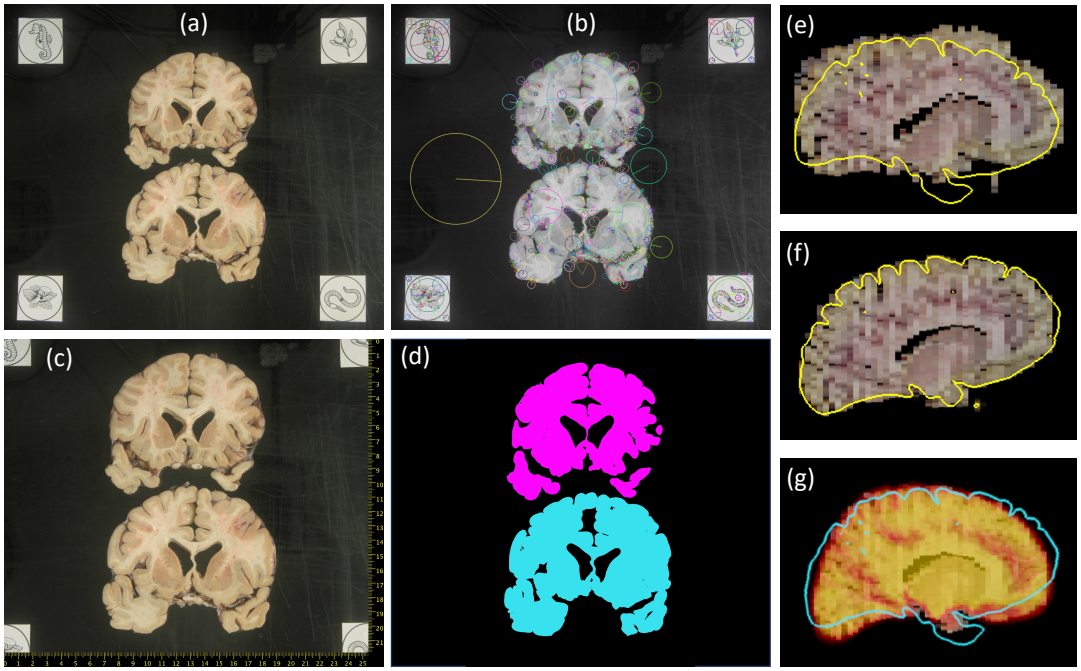


Figure 5. (a) Dissection photograph with brain slices on black board with fiducials. **(b)** SIFT features for fiducial detection. **(c)** Photograph from (a) corrected for pixel size and perspective, with digital ruler overlaid. **(d)** Segmentation against the background, grouping pieces of tissue from the same slice. **(e)** Sagittal slice of the initialization of a 3D reconstruction. **(f)** Corresponding slice of the final 3D reconstruction, obtained with a surface as reference (overlaid in yellow). **(g)** Corresponding slice of the 3D reconstruction provided by a probabilistic atlas (overlaid as a heat map); the real surface is overlaid in light blue for comparison.

facilitate segmentation; otherwise, more extensive manual intervention may be needed. In our experiments, using a flat black background enabled us to automatically segment the tissue with a combination of thresholding and morphological operations, keeping manual edits to a minimum (a couple of minutes per photograph, mostly to erase bits of cortical surface that are sometimes visible around the edge of the face of the slice).

Given this binary mask, the final preprocessing step requires the user to specify the order of the slices within the photograph – which may be anterior to posterior or vice versa, but must be consistent within and across photographs of the same case. Moreover, the user also needs to specify whether two connected components belong to the same slice, which often happens around the temporal pole. This can be quickly accomplished with a dedicated graphical user interface that we distribute with our tools (Figure 5d).

3D volumetric reconstruction from photographs

Recovering a 3D volume from a stack of 2D images entails a consistent 3D reconstruction of the stack via joint image alignment – known as "registration" (***Maintz and Viergever, 1998***; ***Pluim et al., 2003***; ***Zitova and Flusser, 2003***; ***Sotiras et al., 2013***). We pose 3D reconstruction as a joint optimization problem (***Pichat et al., 2018***; ***Mancini et al., 2019***) and use a 3D reference volume for the registration. This volume is ideally a binary 3D mask obtained by rasterizing ("filling") the triangular mesh from the surface scan. If a surface scan is not available, one can instead use a probabilistic atlas (***Shattuck et al., 2008***) of brain shape, which provides a rough reference for the reconstruction. However, this reference cannot drive the reconstruction toward the actual shape of the specimen nor correct deviations from the user-provided slice thickness.

Given N photographed slices and their corresponding masks, the goal of this optimization framework is to simultaneously identify: *(i)* a set of N 2D affine geometric transforms $\{\phi_n\}_{n=1,\ldots,N}$ (rotation, translation, shear, and scaling for each slice); *(ii)* a scaling factor s in the anterior-posterior direction shared by all slices; and *(iii)* a rigid 3D transform Ψ for the reference (rotation and trans-

lation). These transforms seek to align the slices with each other and with the reference volume.

We note that affine transforms (rather than rigid) are required for the photographs due to imperfections in the image preprocessing of the previous section; nevertheless, we expect the shear and scaling of these transforms to be small. Furthermore, we use affine rather than nonlinear transforms because the latter compromise the robustness of the registration, as they introduce huge ambiguity in the space of solutions (e.g., one could add an identical small nonlinear deformation to every slice almost without changing the feature of the 3D reconstruction). This affine model makes it particularly important to place connected components of the same slice in a correct relative position when there is more than one, e.g., in the temporal pole. We further note that the scaling s in the anterior-posterior direction is required to correct deviations from the slice thickness specified by the user, which in practice is never completely exact.

The optimal set of transforms is obtained by maximizing the objective function $\mathcal{F}$ in Equation 1. This objective encodes four desired attributes of the reconstructed data:

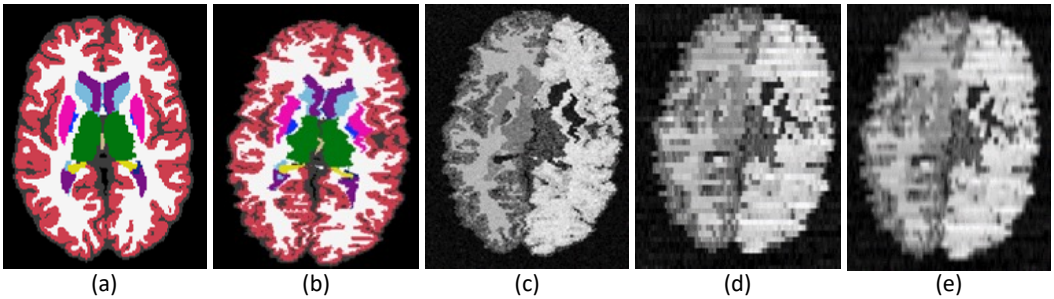
1. A high overlap between the stack of N 3D reconstructed slice masks $M\left[\boldsymbol{x};\left\{\Phi_n\right\},s\right]$ and the aligned reference volume $R[\boldsymbol{x};\Psi]$; we note that images are a function of spatial location $\boldsymbol{x}$.
2. A high similarity between the image intensities of successive (reconstructed) slices $S_n\left[\boldsymbol{x};\Phi_n,s\right]$ and $S_{n+1}\left[\boldsymbol{x};\Phi_{n+1},s\right]$, for $n=1,\ldots,N-1$.
3. A high overlap between successive (reconstructed) slice masks $M_n\left[\boldsymbol{x};\Phi_n,s\right]$ and $M_{n+1}\left[\boldsymbol{x};\Phi_{n+1},s\right]$, for $n=1,\ldots,N-1$.
4. Minimal scaling and shear in the 2D affine transforms; the function f is a regularizer that prevents excessive deformation of the slices – particularly those showing little tissue, e.g., the first and last slices of the stack.

$$\mathcal{F}\left(\Psi,\left\{\Phi_n\right\}\right)=\underbrace{\alpha\mathcal{D}\left(M\left[\boldsymbol{x};\left\{\Phi_n\right\},s\right],R[\boldsymbol{x};\Psi]\right)}_{\text{1) volume-to-volume mask overlap}}+\underbrace{\beta\frac{1}{N-1}\sum_{n=1}^{N-1}\mathcal{C}\left(S_n\left[\boldsymbol{x};\Phi_n,s\right],S_{n+1}\left[\boldsymbol{x};\Phi_{n+1},s\right]\right)}_{\text{2) slice-to-slice cross-correlation}}$$
$$+\underbrace{\gamma\frac{1}{N-1}\sum_{n=1}^{N-1}\mathcal{D}\left(M_n\left[\boldsymbol{x};\Phi_n,s\right],M_{n+1}\left[\boldsymbol{x};\Phi_{n+1},s\right]\right)}_{\text{3) slice-to-slice mask overlap}}-\underbrace{\nu\frac{1}{N}\sum_{n=1}^{N}f\left(\Phi_n\right)}_{\text{4) regularizer}} \tag{1}$$

Mathematically, overlap in Equation 1 is measured with the soft Dice coefficient $\mathcal{D}$ (***Dice, 1945***; ***Sorensen, 1948***; ***Milletari et al., 2016***); image similarity is measured with the normalized cross-correlation $\mathcal{C}$; and the scaling and shearing are measured with the absolute value of the logarithm of the determinant of the affine matrices, i.e., $f\left(\Phi_n\right)$ is the absolute log-determinant of the 3×3 matrix encoding Φ_n. The constants α,β,γ,ν are relative weights, which we set via visual inspection of the output on a small pilot dataset. For the surface reference, we used: $\alpha=.95,\beta=\gamma=.025,\nu=\gamma/100$. For the probabilistic atlas, we trust the reference less and also use more regularization to prevent excessive deformation of slices: $\alpha=.8,\beta=\gamma=\nu=.1$. Either way, the 3D reconstruction is in practice not very sensitive to the exact value of these parameters. We also note that, as opposed to the preprocessing described in the previous section, SIFT is not a good candidate for matching consecutive slices: while it is resilient against changes in pose (e.g., object rotation), perspective, and lightning, it is not robust against changes in the object itself – such as changes between one slice to the next.

The objective function $\mathcal{F}$ is minimized with standard numerical methods – specifically the LBFGS algorithm (***Fletcher, 1987***). The LBFGS optimizer is initialized by stacking the photographs with their centers of gravity on coordinate $(0,0)$, and then matching the center of gravity of the whole stack with the center of gravity of the 3D reference (as illustrated in Video 1 in the supplement, see also https://youtu.be/wo5meYRaGUY).

If a probabilistic atlas is used instead of the surface scan, the same objective function is used but with a slightly different search space. Since we can no longer trust the external reference to correct



(a) (b) (c) (d) (e)

Figure 6. Intermediate steps in the generative process: **(a)** Randomly sampled input label map from the training set. **(b)** Spatially augmented input label map; imperfect 3D reconstruction is simulated with a deformation jitter across the coronal plane. **(c)** Synthetic image obtained by sampling from a Gaussian Mixture Model conditioned on the segmentation, with randomized means and variances. **(d)** Slice spacing is simulated by downsampling to low resolution. This imaging volume is further augmented with a bias field and intensity transforms (brightness, contrast, gamma). **(e)** The final training image is obtained by resampling (d) to high resolution. The neural network is trained with pairs of images like (e) (input) and (b) (target).

for fine shape correction: *(i)* we keep the scaling factor of the anterior-posterior direction fixed to $s = 1$ (i.e., we trust the slice thickness specified by the user); *(ii)* we use rigid rather than affine transforms $\{\phi_n\}$ for the slices (which also has the effect of making the regularizer equal to zero); and *(iii)* we use a full 3D affine transform Ψ for the reference. Sample reconstructions of a case with a 3D surface and a probabilistic atlas are shown in Figure 5e-g. The probabilistic atlas produces a plausible reconstruction, which is however far from the real shape of the specimen given by the surface (Figure 5g). An additional example from the MADRC dataset is shown in Figure 2 of Appendix 2. We note that 3D reconstruction is implemented in PyTorch, and runs efficiently on a graphics processing unit (GPU) – less then five minutes on a Nvidia Quadro RTX 6000.

Segmentation

The 3D reconstructed photographs have various applications (e.g., volumetry, computationally guided dissection) that require image segmentation, i.e., assigning neuroanatomical labels to every spatial location (***Pham et al., 2000***; ***Despotović et al., 2015***; ***Akkus et al., 2017***). There are two main challenges when segmenting the 3D reconstructed photographs. First, the appearance of the images varies widely across cases due to differences in camera hardware (sensor, lens), camera settings, illumination conditions, and tissue preparation (e.g., fixed versus fresh). And second, accurate volumetry requires estimating the segmentation not only on the planes of the photographs but also in between slices.

In our previous work (***Tregidgo et al., 2020***), we adopted a Bayesian segmentation strategy that handled the first issue but not the second. Here, we extend a machine learning approach based on domain randomization (***Tobin et al., 2017***) that we have successfully applied to segment clinical brain MRI scans with large slice spacing (***Billot et al., 2023a***,b). Specifically, our newly proposed approach "Photo-SynthSeg" trains a convolutional neural network for image segmentation (a 3D U-Net, ***Ronneberger et al. 2015***; ***Çiçek et al. 2016***) with synthetic data as follows.

"Photo-SynthSeg" starts from a training dataset (the T1-39 dataset) that comprises a pool of 3D segmentations of brain images at isotropic 3D resolution (Figure 6a). At every iteration during training, one of these 3D segmentations is randomly selected and geometrically deformed with a random nonlinear transform, which simulates imperfect 3D reconstruction by including a "deformation jitter" in the coronal direction, i.e., small but abrupt deformations from one coronal slice to the next (Figure 6b). This deformed isotropic segmentation is used to generate a synthetic 3D image (also at isotropic resolution) using a Gaussian mixture model conditioned on the segmentation (Figure 6c). Crucially, we randomize the Gaussian parameters (means and variances) to make the neural network robust against variations in image appearance. This synthetic isotropic image is then "digitally sliced" into a 3D stack of synthetic coronal photographs, each of which is independently corrupted by a random 2D smooth illumination field that simulates inhomogeneities

in the illumination of the photographs (Figure 6d). Importantly, we also randomize the thickness of the simulated slices in every iteration and introduce small stochastic variations in thickness around the central value within each simulation. This strategy makes the network agnostic to the slice thickness of the reconstructed 3D stack at test time. Finally, the simulated stack, which has anisotropic resolution (e.g., 1mm in-plane and several mm in slice thickness), is resampled to the original isotropic resolution of the (deformed) 3D segmentation (Figure 6e) – typically 1 mm isotropic.

This generative process mimics the image formation process in the real world and has two crucial aspects. First, the super-resolution component: Photo-SynthSeg is a U-Net that will produce a high-resolution (1 mm) isotropic 3D segmentation for every input at test time, independently of the slice spacing of the 3D reconstructed stack of photographs (Figure 2). And second, domain randomization: sampling different Gaussian parameters, illumination fields, and slice thicknesses at every iteration beyond realistic limits (as in Figure 6, where even contralateral regions have different appearance) forces the U-Net to learn features that are independent of the intensity profiles of the photographs, as well as of the spacing between slices. This process makes the U-Net robust against changes in the acquisition. We note that the Gaussian distributions are univariate rather than trivariate, i.e., they model grayscale rather than red-green-blue triplets; we tried training a U-Net with the latter, but the performance on color images was worse than when converting them to grayscale and using the U-Net trained with univariate Gaussians.

During training, resampled stacks and corresponding segmentations (Figure 6b-e) are fed to a U-Net. The U-Net architecture is the same as in our previous works with synthetic scans (***Billot et al., 2020***, ***2023a***,a): it consists of five levels, each separated by a batch normalization layer (***Ioffe and Szegedy, 2015***) along with a max-pooling (contracting path) or an upsampling operation (expanding path). All levels comprise two convolution layers with $3 \times 3 \times 3$ kernels. Every convolutional layer is associated with an Exponential Linear Unit activation (***Clevert et al., 2016***), except for the last one, which uses a softmax. While the first layer comprises 24 feature maps, this number is doubled after each max-pooling, and halved after each upsampling. Following the original U-Net architecture, skip connections are used across the contracting and expanding paths. The network is trained with a soft Dice loss (***Milletari et al., 2016***) and the Adam optimizer (***Kingma and Ba, 2014***). The deep learning model is implemented in Keras (***Chollet, 2015***) with a Tensorflow backend (***Abadi et al., 2016***) and runs on a few seconds on a Nvidia Quadro RTX 6000 GPU. Training takes around seven days on the same GPU.

Finally, we note that there are two different versions of Photo-SynthSeg: one for full cerebra and one for single hemispheres. The later is trained with left hemispheres and flipped right hemispheres, and can thus be used as is for left hemispheres. To process a right hemisphere, we simply left-right flip it, segment it, and flip the results back. While the default Photo-SynthSeg pipeline segments the cerebral cortex as a whole (i.e., as in Figure 6a), our tool also offers the option of subdividing it into parcels as defined by the Desikan-Killiany atlas (***Desikan et al., 2009***), like in Figure 2. This is achieved with the cortical parcellation module ("Segmenter S3") of our tool "SynthSeg" (***Billot et al., 2023b***), which is also distributed with FreeSurfer.

Acknowledgments

First and foremost, we thank the research participants who donated their brains to science and their families, without whom this work would be impossible.

This research was supported by primarily supported by the National Institute of Aging (R01AG070988 and P30AG062421). Other support was provided by NIH grants RF1MH123195, R01EB031114, UM1MH130981, P30AG066509 (UW ADRC), U19AG066567, U19AG060909, K08AG065426, R01NS112161; the European Union (ERC Starting Grant 677697); Alzheimer's Research UK (ARUK-IRG2019A-003); and the Massachusetts Life Sciences Center. AC was further funded by the European Union, the Ministry of Universities and Recovery, and the Transformation and Resilience Plan, through a call from Universitat Politècnica de Catalunya (Grant Ref. 2021UPC-MS-67573). Additional support was

provided by NIH grants U01MH117023, 1R01EB023281, R01EB006758, R21EB018907, R01EB019956, P41EB030006, 1R56AG064027, 1R01AG064027, 5R01AG008122, R01AG016495, 1R01AG070988, UM1MH130 R01 MH123195, R01 MH121885, 1RF1MH123195, R01NS0525851, R21NS072652, R01NS070963, R01NS083534, 5U01NS086625,5U24NS10059103, R01NS105820, 1S10RR023401, 1S10RR019307, 1S10RR023043, 5U01MH093765, and by the Buster Alvord Endowment.

BF has a financial interest in CorticoMetrics, a company developing brain MRI measurement technology. His interests are reviewed and managed by Massachusetts General Hospital.

References

Abadi M, Barham P, Chen J, Chen Z, Davis A. Tensorflow: A system for large-scale machine learning. In: *Symposium on Operating Systems Design and Implementation*; 2016. p. 265–283.

Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. Journal of digital imaging. 2017; 30:449–459.

Billot B, Greve D, Van Leemput K, Fischl B, Iglesias JE, Dalca A. A Learning Strategy for Contrast-agnostic MRI Segmentation. In: *Medical Imaging with Deep Learning*; 2020. p. 75–93.

Billot B, Greve DN, Puonti O, Thielscher A, Van Leemput K, Fischl B, Dalca AV, Iglesias JE, et al. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. Medical Image Analysis. 2023; p. 102789.

Billot B, Magdamo C, Cheng Y, Arnold SE, Das S, Iglesias JE. Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets. Proceedings of the National Academy of Sciences. 2023; 120(9):e2216399120.

Blanken AE, Hurtz S, Zarow C, Biado K, Honarpisheh H, Somme J, Brook J, Tung S, Kraft E, Lo D, et al. Associations between hippocampal morphometry and neuropathologic markers of Alzheimer's disease using 7T MRI. NeuroImage: Clinical. 2017; 15:56–61.

Chollet F, Keras; 2015. Https://keras.io.

Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19* Springer; 2016. p. 424–432.

Clevert DA, Unterthiner T, Hochreiter S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). arXiv:151107289 [cs]. 2016; .

Coupé P, Catheline G, Lanuza E, Manjón JV, Initiative ADN. Towards a unified analysis of brain maturation and aging across the entire lifespan: A MRI analysis. Human brain mapping. 2017; 38(11):5501–5518.

Desikan RS, Cabral HJ, Hess CP, Dillon WP, Glastonbury CM, Weiner MW, Schmansky NJ, Greve DN, Salat DH, Buckner RL, et al. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain. 2009; 132(8):2048–2057.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006; 31(3):968–980.

Despotović I, Goossens B, Philips W. MRI segmentation of the human brain: challenges, methods, and applications. Computational and mathematical methods in medicine. 2015; 2015.

Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945; 26(3):297–302.

Dickerson BC, Feczko E, Augustinack JC, Pacheco J, Morris JC, Fischl B, Buckner RL. Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. Neurobiology of aging. 2009; 30(3):432–440.

Edlow BL, Mareyam A, Horn A, Polimeni JR, Witzel T, Tisdall MD, Augustinack JC, Stockmann JP, Diamond BR, Stevens A, et al. 7 Tesla MRI of the ex vivo human brain at 100 micron resolution. Scientific data. 2019; 6(1):1–10.

Fischl B. FreeSurfer. Neuroimage. 2012; 62(2):774–781.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Van Der Kouwe A, Killiany R, Kennedy D, Klaveness S, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002; 33(3):341–355.

Fletcher R. Practical methods of optimization john wiley & sons. New York. 1987; 80(4).

Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016.

Iglesias JE, Billot B, Balbastre Y, Magdamo C, Arnold SE, Das S, Edlow BL, Alexander DC, Golland P, Fischl B. SynthSR: A public AI tool to turn heterogeneous clinical brain scans into high-resolution T1-weighted images for 3D morphometry. Science advances. 2023; 9(5):eadd3607.

Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: *International Conference on Machine Learning*; 2015. p. 448–456.

Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014; .

Latimer CS, Melief EJ, Ariza-Torres J, Howard K, Keen AR, Keene LM, Schantz AM, Sytsma TM, Wilson AM, Grabowski TJ, et al. Protocol for the Systematic Fixation, Circuit-Based Sampling, and Qualitative and Quantitative Neuropathological Analysis of Human Brain Tissue. In: *Alzheimer's Disease* Springer; 2023.p. 3–30.

Lerch JP, Pruessner JC, Zijdenbos A, Hampel H, Teipel SJ, Evans AC. Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. Cerebral cortex. 2005; 15(7):995–1001.

Love S. Neuropathological investigation of dementia: a guide for neurologists. Journal of Neurology, Neurosurgery & Psychiatry. 2005; 76(suppl 5):v8–v14.

Lowe DG. Object recognition from local scale-invariant features. In: *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2 Ieee; 1999. p. 1150–1157.

Maintz JA, Viergever MA. A survey of medical image registration. Medical image analysis. 1998; 2(1):1–36.

Mancini M, Crampsie S, Thomas DL, Jaunmuktane Z, Holton JL, Iglesias JE. Hierarchical joint registration of tissue blocks with soft shape constraints for large-scale histology of the human brain. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* IEEE; 2019. p. 666–669.

Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics. 1947; p. 50–60.

Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 fourth international conference on 3D vision (3DV)* Ieee; 2016. p. 565–571.

Pham DL, Xu C, Prince JL. Current methods in medical image segmentation. Annual review of biomedical engineering. 2000; 2(1):315–337.

Pichat J, Iglesias JE, Yousry T, Ourselin S, Modat M. A survey of methods for 3D histology reconstruction. Medical image analysis. 2018; 46:73–105.

Pluim JP, Maintz JA, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE transactions on medical imaging. 2003; 22(8):986–1004.

Puonti O, Iglesias JE, Van Leemput K. Fast and sequence-adaptive whole-brain segmentation using parametric Bayesian modeling. NeuroImage. 2016; 143:235–249.

Ravid R. Biobanks for biomarkers in neurological disorders: the Da Vinci bridge for optimal clinicopathological connection. Journal of the Neurological Sciences. 2009 Aug; 283(1-2):119–126. doi: 10.1016/j.jns.2009.02.364.

Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* Lecture Notes in Computer Science, Cham: Springer International Publishing; 2015. p. 234–241. doi: 10.1007/978-3-319-24574-4_28.

Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RS, Busa E, Morris JC, Dale AM, Fischl B. Thinning of the cerebral cortex in aging. Cerebral cortex. 2004; 14(7):721–730.

Salvi J, Pages J, Batlle J. Pattern codification strategies in structured light systems. Pattern recognition. 2004; 37(4):827–849.

Shatil AS, Matsuda KM, Figley CR. A method for whole brain ex vivo magnetic resonance imaging with minimal susceptibility artifacts. Frontiers in neurology. 2016; 7:208.

Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW. Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage. 2008; 39(3):1064–1080.

Sorensen TA. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Biol Skar. 1948; 5:1–34.

Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: A survey. IEEE transactions on medical imaging. 2013; 32(7):1153–1190.

Tobin J, Fong R, Ray A, Schneider J, Zaremba W, Abbeel P. Domain randomization for transferring deep neural networks from simulation to the real world. In: *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)* IEEE; 2017. p. 23–30.

Tregidgo HF, Casamitjana A, Latimer CS, Kilgore MD, Robinson E, Blackburn E, Van Leemput K, Fischl B, Dalca AV, Donald CLM, et al. 3D reconstruction and segmentation of dissection photographs for MRI-free neuropathology. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23* Springer; 2020. p. 204–214.

Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TE, Bucholz R, Chang A, Chen L, Corbetta M, Curtiss SW, et al. The Human Connectome Project: a data acquisition perspective. Neuroimage. 2012; 62(4):2222–2231.

Walhovd KB, Fjell AM, Reinvang I, Lundervold A, Dale AM, Eilertsen DE, Quinn BT, Salat D, Makris N, Fischl B. Effects of age on volumes of cortex, white matter and subcortical structures. Neurobiology of aging. 2005; 26(9):1261–1270.

Webster JM, Grabowski TJ, Madhyastha TM, Gibbons LE, Keene CD, Latimer CS. Leveraging Neuroimaging Tools to Assess Precision and Accuracy in an Alzheimer's Disease Neuropathologic Sampling Protocol. Frontiers in Neuroscience. 2021; 15. https://www.frontiersin.org/article/10.3389/fnins.2021.693242.

Zitova B, Flusser J. Image registration methods: a survey. Image and vision computing. 2003; 21(11):977–1000.

Appendix 1

Data acquisition at the Massachusetts Alzheimer's Disease Research Center (MADRC)

The whole brain is removed using standard autopsy procedures. After removal of the dura matter, the brain is weighed, photo-documented, and then fixed in 10% neutral-buffered formalin (NBF). Some specimens are cut along the midline sagittal plane with one hemibrain frozen for future biochemical studies. After at least one week of fixation, the specimens are photo-documented again. The posterior fossa is then dissected off the fixed full or hemibrain through the midbrain at the level of cranial nerve III, and the supratentorial portion of the brain is 3D surface scanned.

After scanning, the brain or hemibrain is sectioned at defined anatomical landmarks (anterior temporal tips, optic chiasm, infundibulum, mamillary bodies, cerebral peduncles, red nuclei, and colliculi) with the frontal and occipital lobes additionally sectioned at approximately 10 mm intervals. The front and back of all sections are then photographed with a metric ruler included for size reference. The photographs are then deidentified for further analysis.

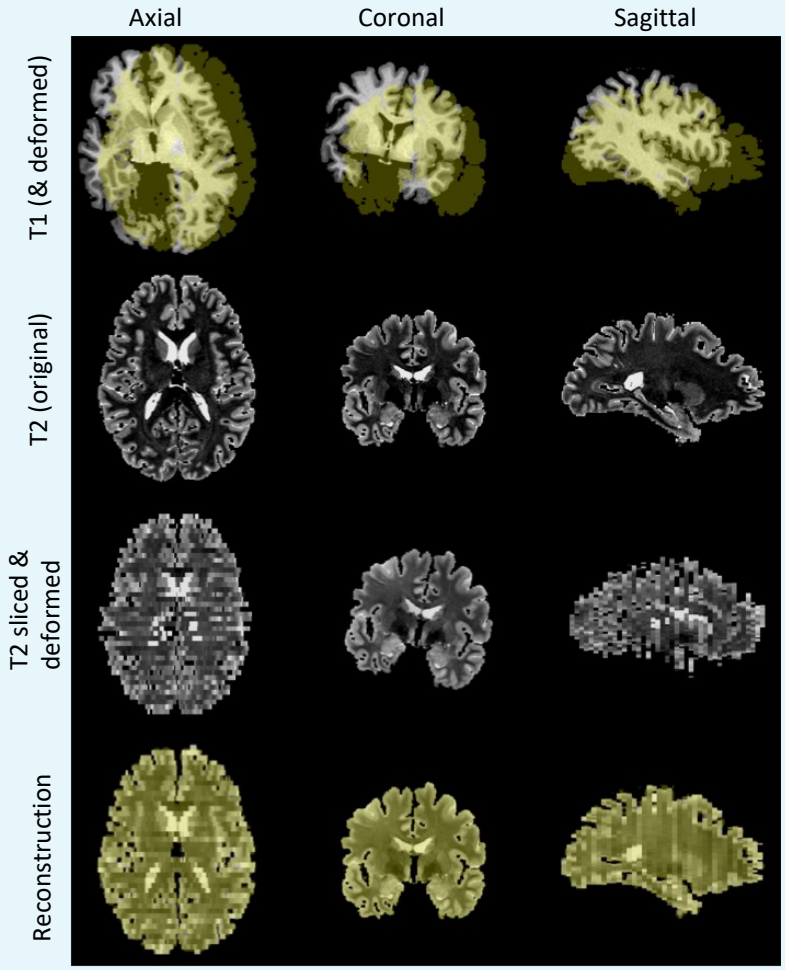
Data acquisition at the Alzheimer's Disease Research Center of the University of Washington (UW-ADRC)

The whole brain is removed using standard autopsy procedures. The brain is weighed and photo-documented, and then the meninges are removed. For donors with a postmortem interval (PMI: the time interval between death and procurement) longer than 12 hours, the whole brain is fixed in 10% NBF with no tissue frozen. A rapid slicing protocol is performed for donors with a post-mortem interval of fewer than 12 hours. For a rapid protocol, the brain is bisected along the midline, and one hemibrain is placed in a bucket with 10% NBF for two weeks. The other hemibrain is 3D surface scanned, then placed in a scaffolding box with the vermis of the cerebellum flush against the posterior wall. The hemibrain is embedded in freshly mixed dental alginate (540g powder blended in 4L of water) and submerged until the alginate is set. The alginate block containing the hemibrain is placed in a slicing sled with the frontal pole toward the front. The hemibrain is sliced anterior to posterior in 4mm slices. All slices are set on Teflon-coated aluminum plates and photographed. Alternating slices are set to be flash-frozen, with the remaining slices fixed in 10% NBF. The hemibrain (rapid) or whole (non-rapid) brain is fixed in 10% NBF for two weeks, embedded in agarose, and scanned in a 3T MRI scanner.

Following the MRI, the agarose-embedded brain is sliced on a modified deli slicer for precise 4 mm tissue slabs aligned with the ex vivo MRI images for image-guided tissue sampling to standard tissue sampling following the current National Institute of Aging-Alzheimer's Association (NIA-AA) consensus guidelines. After sampling, all blocks are processed and embedded in paraffin according to standard techniques. Following current NIA-AA guidelines, the resulting Fast-Frozen Paraffin-Embedded blocks are cut and stained for diagnostic analysis. Histologically stained slides are scanned into the HALO (https://indicalab.com/halo/) imaging workstation using an Aperio AT2 slide scanner and are analyzed by a trained neuropathologist.

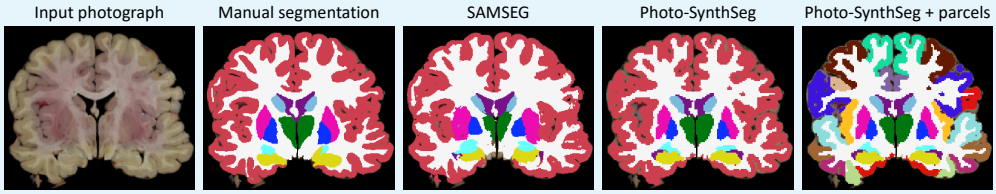
Appendix 2

Supplementary figures



Appendix 2—figure 1. Simulation and reconstruction of synthetic data. Top row: skull stripped T1 scan and (randomly translated and rotated) binary mask of the cerebrum, in yellow. Second row: original T2 scan. Third row: randomly sliced and linearly deformed T2 images. Bottom row: output of the 3D reconstruction algorithm, i.e., reconstructed T2 slices and registered reference mask overlaid in yellow.

Appendix 2—figure 2. Reconstruction with surface scan vs probabilistic atlas. **(a)** Initialization, with contour of 3D surface scan superimposed. **(b)** Reconstruction with 3D surface scan.
(c) Reconstruction with probabilistic atlas (overlaid as heat map with transparency); the contour of the surface scan is overlaid in light blue, for comparison. Even though the shape of the reconstruction in (c) is plausible, it is clearly inaccurate in light of the surface scan.



Appendix 2—figure 3. Example of mid-coronal slice selected for manual segmentation and computation of Dice scores. Compared with the FreeSurFer protocol, we merge the ventral diencephalon (which has almost no visible contrast in the photographs) with the cerebral white matter in our manual delineations. We also merged this structures in the automated segmentations from SAMSEG and Photo-SynthSeg in this figure, for a more consistent comparison.