Robust joint registration of multiple stains and MRI for multimodal 3D histology reconstruction: Application to the Allen human brain atlas

Adrià Casamitjana[a,*], Marco Lorenzi[b], Sebastiano Ferraris[a], Loïc Peter[a], Marc Modat[c], Allison Stevens[d], Bruce Fischl[d,e,f], Tom Vercauteren[c], Juan Eugenio Iglesias[a,d,e]

[a]*Center for Medical Image Computing, University College London, UK*
[b]*Université Côte dÁzur, Inria, Epione Team, 06902 Sophia Antipolis, France*
[c]*School of Biomedical Engineering & Imaging Sciences, King's College London, UK*
[d]*Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, USA*
[e]*Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA*
[f]*Program in Health Sciences and Technology, Massachusetts Institute of Technology, USA*

Abstract

Joint registration of a stack of 2D histological sections to recover 3D structure ("3D histology reconstruction") finds application in areas such as atlas building and validation of *in vivo* imaging. Straightforward pairwise registration of neighbouring sections yields smooth reconstructions but has well-known problems such as "banana effect" (straightening of curved structures) and "z-shift" (drift). While these problems can be alleviated with an external, linearly aligned reference (e.g., Magnetic Resonance (MR) images), registration is often inaccurate due to contrast differences and the strong nonlinear distortion of the tissue, including artefacts such as folds and tears. In this paper, we present a probabilistic model of spatial deformation that yields reconstructions for multiple histological stains that that are jointly smooth, robust to outliers, and follow the reference shape. The model relies on a spanning tree of latent transforms connecting all the sections and slices of the reference volume, and assumes that the registration between any pair of images can be see as a noisy version of the composition of (possibly inverted) latent transforms connecting the two images. Bayesian inference is used to compute the most likely latent transforms given a set of pairwise registrations between image pairs within and across modalities. We consider two likelihood models: Gaussian (ℓ_2 norm, which can be minimised in closed form) and Laplacian (ℓ_1 norm, minimised with linear programming). Results on synthetic deformations on multiple MR modalities, show that our method can accurately and robustly register multiple contrasts even in the presence of outliers. The framework is used for accurate 3D reconstruction of two stains (Nissl and parvalbumin) from the Allen human brain atlas, showing its benefits on real data with severe distortions. Moreover, we also provide the registration of the reconstructed volume to MNI space, bridging the gaps between two of the most widely used atlases in histology and MRI. The 3D reconstructed volumes and atlas registration can be downloaded from `https://openneuro.org/datasets/ds003590`. The code is freely available at `https://github.com/acasamitjana/3dhirest`.

Keywords: histology, nonlinear registration, 3D reconstruction, linear programming, *ex vivo* MRI

1. Introduction

1.1. Motivation

Histology is the area of science concerned with microscopic exploration of tissue sections sampled from either a *post mortem* specimen or biopsy tissue. After a tissue processing pipeline (Bancroft & Gamble, 2008), thin sections can be inspected under the microscope and digitised with a scanner. The most common histology pipeline consists of fixation, processing and embedding with a hardening material (e.g., wax) for sectioning. Thin sections from the wax-embedded tissue are cut using a microtome and mounted in glass slides for staining.

---
*Corresponding author
Email address: `a.casamitjana@ucl.ac.uk` (Adrià Casamitjana)

Large specimens (e.g., a whole human brain) are typically first cut into several blocks, which are then processed independently.

Histological examination is the gold standard for many diagnostic protocols. Different staining procedures enable the visualisation of different microscopic structures. For example, the ubiquitous Haemotoxylin and Eosin (H&E) (Chan, 2014) stains cell nuclei purple and cytoplasm pink while immunohistochemistry techniques selectively identify antigens in cells (Ramos-Vara, 2005)). Stains are most often used in combination, e.g., to target different cell types in cancer diagnosis (Cooper et al., 2009) or to detect neuropathologies in neurodegenerative diseases (Montine et al., 2012).

On addition to clinical pathology, histology has many applications in medical imaging, often in combination with mm-scale modalities like Magnetic Resonance Imaging (MRI): the former yields excellent contrast at the microscopic scale, while the latter provides larger-scale context, 3D structure and minimal distortion. One such application is the validation of *in vivo* imaging techniques, like microstructure imaging (Bourne et al., 2017) or mass spectrometry imaging (Thiele et al., 2014). In these applications, histology provides a gold standard for the underlying anatomy.

Another successful application of combining μm-scale histology with mm-scale is the 3D modelling of anatomy at the microscopic level. In anatomy, histology is often used as the basis for fine anatomical delineation as it provides detailed information about the size, shape and cell density of structures.The mm-scale images are often used to inform the spatial registration of histological sections into a consistent 3D volume – a problem known as 3D histology reconstruction (Pichat et al., 2018). Sample applications include mammary glands (Shojaii et al., 2014) or lungs (Rusu et al., 2015) in mice; brain tissue (Malandain et al., 2004; Stille et al., 2013) or prostate (Gibson et al., 2013) in humans.

In the specific case of human neuroimaging, histology has been combined with MRI to build atlases. The most notable examples are the BigBrain project (Amunts et al., 2013) and the Allen atlas (Ding et al., 2016), which used special whole-brain microtomes to section through a whole human brain specimen each (a single hemisphere in the Allen atlas), without needing to block the tissue. At the single structure level, Yelnik et al. (2007) used immunohistochemistry and MRI to build an atlas of the basal ganglia; Adler et al. (2018) used a similar approach to build an atlas of the hippocampus; in previous work, we used Nissl staining with thionin to build a probabilistic atlas of the human thalamus (Iglesias et al., 2018a). Krauth et al. (2010) also built a thalamic atlas, but used histological sections in different orientations (rather than a reference MRI) to solve the 3D reconstruction problem (Song et al., 2013). Many of these works integrate multiple stains to capture a richer characterisation of the microstructural organisation of tissue. For example, the Allen atlas comprises three staining procedures (Nissl stain and two antibodies for immunohistochemistry) to gather structural evidence from both cyto- and chemoarchitectural features (Ding et al., 2016).

A central component of the works above is the 3D histology reconstruction. The invasive acquisition and the processing pipeline used in histology heavily distort the original shape of the tissue such that the 3D contextual information is lost and the spatial relationship between and within structures is broken. Therefore, image registration algorithms are required to recover the original 3D shape. Other processing artefacts, including but not limited to staining inconsistency, tears, folds, tissue loss and air bubbles, may be accounted for with dedicated preprocessing methods (Pichat et al., 2018).

In this work, we seek to achieve 3D reconstruction of multi-modality serial histology with a method that satisfies three main properties. First, producing spatially smooth reconstructions. Second, producing precise reconstructions, i.e., recovering shapes that accurately follow the underlying anatomy. And third, being robust against histological artefacts (e.g., folds or tears). We assume the availability of an external reference (e.g., an MRI scan, Annese 2012) that provides contextual information and enables unbiased reconstruction.

1.2. Related work

3D histology reconstruction without any additional shape information is an underconstrained problem. In this case, reconstruction is typically achieved by pairwise registration of adjacent slices. An important design choice in this approach is the reference slice, which is often chosen to be the one at the centre of the stack (Ourselin et al., 2001) – even though automatic selection methods (Bağci & Bai, 2010) have also been proposed. While these methods yield 3D reconstructions that are smooth (and thus visually pleasant), the lack of external

guidance tends to straighten curved shapes (incurring the so-called "banana effect") and also leads to accumulation of errors along the stack ("z-shift", Malandain et al. 2004).

These problems can be mitigated with a reference volume providing information on the true shape and thus constraining the original reconstruction problem. 3D histology reconstruction is then split into a linear 3D registration problem between the reference volume and the stack of histological images, and a set of nonlinear 2D registration problems between section in the stack and the corresponding resampled slice from the reference volume. The problem can be addressed in an iterative fashion (Malandain et al., 2004). Other existing approaches attempt to solve the linear 3D and nonlinear 2D problems simultaneously, i.e., jointly optimising the registration similarity metric with respect to all linear and nonlinear parameters (Alic et al., 2011; Yang et al., 2012). While this approach is potentially more accurate, it also requires dedicated registration algorithms.

Different methods have been used in the literature to initialise the stack, e.g., direct stacking of sections with alignment of their centres of mass (Goubran et al., 2013) or, most commonly, pairwise registration of consecutive slices starting from the bottom (Ceritoglu et al., 2010) or middle (Stille et al., 2013) of the stack. To decrease the z-shift effect and avoid large error due to badly distorted slices, some approaches consider larger neighbourhoods when registering the histological stack. For example, Yushkevich et al. (2006) consider a 5-neighbourhood centred on the reference slice and use a graph theoretical approach to find the shortest path from every slice to a selected reference slice, increasing the robustness against poorly registered slices. The use of an undistorted intermediate modality may also facilitate the 3D alignment. For example, blockface photographs are sometimes taken previous to sectioning and thus do not present many of the artefacts caused by the histology processing pipeline (Amunts et al., 2013).

If the linear 3D alignment is considered fixed (i.e., histological sections are linearly aligned to resampled slices of the reference volume), histology reconstruction reduces to a set of 2D in-plane registration problems. Naively, any intermodality registration model could be used to align histology and the resampled slices of the reference volume (henceforth "reference slices") one at the time, producing an unbiased 3D reconstruction. However, intermodal

nonlinear registration (typically using mutual information) is often difficult and inaccurate due to the artefacts discussed above, such as folding, tears, etc. (Jacobs et al., 1999). When treating each histological section independently, these inaccuracies yield jagged reconstructions in the orthogonal planes.

To improve reconstruction continuity, sequential approaches in the literature consider not only the reference slice but also the adjacent histological sections in each 2D registration step (Adler et al., 2014; Rusu et al., 2015; Wirtz et al., 2004). Hence, each histological slice is deformed to simultaneously match their reference counterpart and neighbouring slices in an iterative fashion. However, these approaches are prone to getting stuck in local minima, are biased towards the choice of the initial slice, and may propagate correlated errors at each step.

Instead, joint refinement of deformation fields provides a robuster alternative. For example, Feuerstein et al. (2011) define a 3D Markov random field between the control points of a set of precomputed 2D B-spline transforms and use discrete optimisation tools to globally minimise an energy function that encourages smoothness. Other *ad hoc* approaches involve low-pass filtering of the deformation fields along the direction of the stack, e.g., with a Gaussian filter that smooths rigid transforms (Yushkevich et al., 2006). Similarly, Malandain et al. (2004) apply a Gaussian filter directly on linear transform parameters. Finally, Casero et al. (2017) define a new framework for linear histology reconstruction and provide theoretical equivalence with Gaussian filter smoothing. In our prior work (Iglesias et al., 2018b), we defined a generative model over diffeomorphic deformation fields and used Bayesian inference to find a smooth solution along the stack direction. This framework yields 3D reconstructions that are simultaneously smooth and unbiased, i.e., without banana effect or z-shift.

1.3. Contribution

Here we present an extension of our previous conference article (Iglesias et al., 2018b) on 3D histology reconstruction with a reference volume. The contributions with respect to such article are:

- *Multiple stains:* we extend the framework to joint reconstruction of multiple histological stains, explicitly encouraging the spatial alignment of all modalities. To the best of our

knowledge, it is the first attempt at 3D histology reconstruction using multiple stains.

- *Robustness:* we greatly increase the robustness of our method by modelling the registration error with Laplacian distributions that penalise its ℓ_1 norm, and solving the optimisation problem with linear programming. Moreover we use histology and reference masks on the generative model to further mitigate the impact of artefacts on the optimisation.

- *Computational efficiency:* we greatly increase the efficiency of our method by making approximations in the optimisation that have nearly no effect on the registration accuracy as well as by using modern deep learning registration techniques.

The rest of this paper is organised as follows. We introduce the proposed framework in Section 2. Experiments on two case studies are described in Section 3. A dataset with synthetic deformations is used to thoroughly test and compare different variants of the methodology in Section 3.4. A real-case scenario of 3D brain histology reconstruction is described in Section 3.5. Finally, Section 4 discusses the results and concludes the article.

2. Methods

2.1. Preliminaries

Consider a 3D histology reconstruction framework where different staining procedures are carried out and a reference volume is available (henceforth, we will assume this volume is an ex-vivo MRI scan). Let $\{I_n^c(\boldsymbol{x})\}_{n=1,\dots,N}$ be a stack of N histological sections of contrasts $c = 1, \dots, C$ (e.g., H&E staining), defined on pixel locations $\boldsymbol{x}$ over a discrete 2D image domain Ω. Paired sections of each contrast are cut a few microns apart and are thus assumed to be from the same tissue – but with independent deformations. We further assume that the reference MRI volume has been linearly aligned to the stack, and resampled into the planes of the histological sections: $\{I_n^0(\boldsymbol{x})\}_{n=1,\dots,N}$. Then, 3D histology reconstruction amounts to solving a set of interdependent 2D registration problems. We further assume that all images (MRI and histology) have associated binary masks $M_n^c(\boldsymbol{x})$ discriminating tissue vs. background and obtained with manual delineation or (semi-)automatic methods (e.g., Wang et al. 2016).

Furthermore, we consider all $(C + 1)N$ images vertices in a graph $\mathcal{G}$, connected through a spanning tree with $L = N(C+1)-1$ edges, such that any two images are connected by a unique path across the tree. The choice of spanning tree is irrelevant, as the algorithm presented below will guarantee convergence to the global optimum. Here we assume that the tree consists of $N - 1$ edges connecting the MRI slices to their neighbours (i.e., I_n^0 to I_{n+1}^0, $\forall n < N$), as well a $C \times N$ edges connecting each MRI slice I_n^0 to the C corresponding histological sections $I_n^1, \dots, I_n^C$ (Fig. 1).
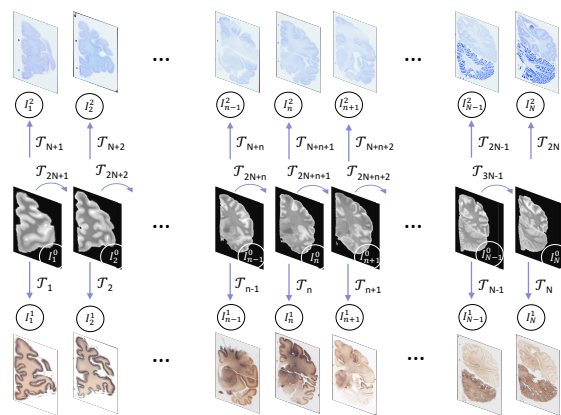

Figure 1: Our choice of spanning tree for C=2. Each MRI slice is connected to the corresponding histological sections, as well as the immediate neighbour in the MRI stack.

Associated with the edges of the spanning tree, we define a set of L latent, noise-free, nonlinear, diffeomorphic transforms $\{\mathcal{T}_l(\boldsymbol{x})\}_{l=1,\dots,L}$. These transforms introduce directionality in the graph $\mathcal{G}$. As for the choice of spanning tree, the chosen criterion for defining the direction of the transforms does not affect the results of the algorithm. Here, we assume that the transforms point from I_n^0 to I_n^c, $\forall n, c$, and from I_n^0 to I_{n+1}^0, $\forall n < N$ (Fig. 1). Since these transforms are assumed to be diffeomorphic and thus invertible, one can obtain the latent transform connecting any two images in $\mathcal{G}$ by composing a subset of (possibly inverted) transforms in $\{\mathcal{T}_l(\boldsymbol{x})\}$.

Finally, we consider a set of $K \geq L$ nonlinear diffeomorphic transforms between pairs of images in $\mathcal{G}$, estimated with a diffeomorphic registration algorithm (e.g., Avants et al. 2008; Modat et al. 2012; Dalca et al. 2018; Ashburner 2007; Vercauteren et al. 2008): $\{\mathcal{R}_k(\boldsymbol{x})\}_{k=1,\dots,K}$. Every $\mathcal{R}_k$ can be

seen as a noisy version of a composition of transforms in $\{\mathcal{T}_l\}$ and their inverses. We use a $K \times L$ matrix $\boldsymbol{W}$ to encode the transforms in $\{\mathcal{T}_l\}$ that each $\mathcal{R}_k$ traverses, as follows:

- $W_{kl} = 1$ if $\mathcal{T}_l$ is on the path of $\mathcal{R}_k$,

- $W_{kl} = -1$ if $\mathcal{T}_l^{-1}$ is on the path of $\mathcal{R}_k$, and

- $W_{kl} = 0$ otherwise.

Even though $\mathcal{R}_k$ can in principle connect any pair of nodes, one would normally only register images that are not too far in the graph. In practice, we compute registrations between images at the same level in the stack (MRI to histology, as well as between different histological stains), as well as between every image (MRI or histology) and its nearest neighbours in the stack (Fig. 2). It is the goal of our method to infer the most likely underlying $\{\mathcal{T}_l\}$ given the observed $\{\mathcal{R}_k\}$, as explained in the following section.
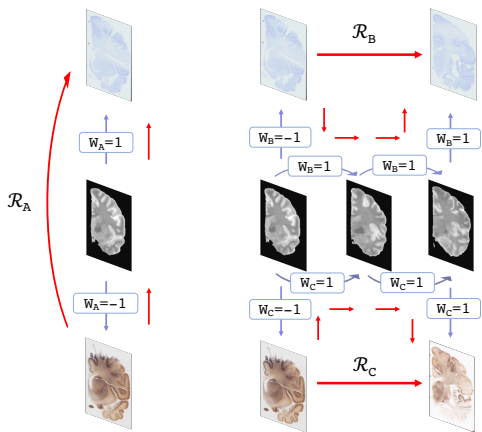

Figure 2: Example of three observations. $\mathcal{R}_A$: an intermodality registration between histological contrasts. $\mathcal{R}_B, \mathcal{R}_C$: intramodality registration within contrasts in the tree leaves. Blue and red arrows indicate the direction of the velocity fields from the spanning tree definition and observation path, respectively. Weight values are +1/-1 when blue and red arrows follow the same/opposite direction.

2.2. Probabilistic modelling and Bayesian inference

In this work, we assume that the latent transforms $\{\mathcal{T}_l\}$ and the observed registrations $\{\mathcal{R}_k\}$ are explained by a probabilistic generative model, such that 3D histology reconstruction can be posed as a Bayesian inference problem: given the $K \geq L$ observed noisy registrations, what is the most likely

set of L underlying transforms (i.e., those in the spanning tree) that gave rise to them?

The proposed probabilistic model relies on a number of key assumptions:

- The observed registrations $\{\mathcal{R}_k\}$ are conditionally independent, given the latent transforms $\{\mathcal{T}_l\}$.

- The likelihood of each registration $\mathcal{R}_k$ is parameterised by a set of parameters $\boldsymbol{\theta}$, for which we do not make any prior assumptions, i.e., $p(\boldsymbol{\theta}) \propto 1$.

- We do not make any prior assumptions on the distribution of the latent transforms $\{\mathcal{T}_l\}$, i.e., $\mathrm{p}(\{\mathcal{T}_l\}) \propto 1$.

Under these assumptions, the generative model describing the joint probability distribution of the latent transforms, the likelihood parameters and the observed registrations is:

$$
\begin{aligned}
p(\{\mathcal{T}_l\}, &\{\mathcal{R}_k\}, \boldsymbol{\theta}) \\
&= p(\{\mathcal{T}_l\})p(\boldsymbol{\theta}) \prod_{k=1}^{K} p(\mathcal{R}_k | \{\mathcal{T}_l\}, \boldsymbol{\theta}; W_{k,:}) \\
&\propto \prod_{k=1}^{K} p(\mathcal{R}_k | \{\mathcal{T}_l\}, \boldsymbol{\theta}; W_{k,:}), \quad (1)
\end{aligned}
$$

where we have dropped the dependency on $\boldsymbol{x}$ for simplicity.

Within this framework, one can compute the most likely 3D reconstruction by finding the most likely latent transforms that bring the histological sections into alignment; we note that this is a subset of $\{\mathcal{T}_l\}$, including only the transforms between histology and MRI (subset S_1); the transforms between reference MRI slices (subset S_2) are not needed for the 3D histology reconstruction. In a fully Bayesian formulation, solving this problem requires marginalising over all the variables we are not seeking to optimise, including the likelihood parameters and the subset of latent transforms connecting the reference MRI slices:

$$
\begin{aligned}
\{\hat{\mathcal{T}}_l\}_{l \in S_1} &= \underset{\{\mathcal{T}_l\}_{l \in S_1}}{\mathrm{argmax}}\, p(\{\mathcal{T}_l\}_{l \in S_1} | \{\mathcal{R}_k\}) \\
&= \underset{\{\mathcal{T}_l\}_{l \in S_1}}{\mathrm{argmax}} \int p(\{\mathcal{T}_l\}, \boldsymbol{\theta} | \{\mathcal{R}_k\}) d\boldsymbol{\theta} \prod_{l \in S_2} d\mathcal{T}_l. \quad (2)
\end{aligned}
$$

Equation 2 is often intractable due to the integral over transforms and likelihood parameters. Instead,

we optimise the joint probability of all transforms (both subsets) and likelihood parameters:

$$\{\hat{\mathcal{T}}_l\}, \hat{\boldsymbol{\theta}} = \underset{\{\mathcal{T}_l\}, \boldsymbol{\theta}}{\operatorname{argmax}}\, p(\{\mathcal{T}_l\}, \boldsymbol{\theta} | \{\mathcal{R}_k\})$$

$$= \underset{\{\mathcal{T}_l\}, \boldsymbol{\theta}}{\operatorname{argmax}}\, p(\{\mathcal{T}_l\}, \boldsymbol{\theta}, \{\mathcal{R}_k\})$$

$$= \underset{\{\mathcal{T}_l\}, \boldsymbol{\theta}}{\operatorname{argmax}} \prod_{k=1}^{K} p(\mathcal{R}_k | \{\mathcal{T}_l\}, \boldsymbol{\theta}; W)$$

$$= \underset{\{\mathcal{T}_l\}, \boldsymbol{\theta}}{\operatorname{argmax}} \sum_{k=1}^{K} \log p(\mathcal{R}_k | \{\mathcal{T}_l\}, \boldsymbol{\theta}; W). \quad (3)$$

2.3. Model instantiation

There are two main design choices in our model: the representation for the spatial transforms $\{\mathcal{T}_l\}, \{\mathcal{R}_k\}$, and the shape of the likelihood $p(\mathcal{R}_k | \{\mathcal{T}_l\}, \boldsymbol{\theta}; W)$.

2.3.1. Model for spatial transforms

We choose the Log-Euclidean framework to parameterise diffeomorphisms in the Lie group of stationary velocity fields (SVFs, Arsigny et al. 2006). Let $\{\boldsymbol{R}_k(\boldsymbol{x})\}$ and $\{\boldsymbol{T}_l(\boldsymbol{x})\}$ be the SVF infinitesimal generators whose integration using Lie exponentials result in the corresponding diffeomorphisms $\mathcal{R}_k(\boldsymbol{x}) = \exp[\boldsymbol{R}_k(\boldsymbol{x})]$ and $\mathcal{T}_l(\boldsymbol{x}) = \exp[\boldsymbol{T}_l(\boldsymbol{x})]$. For fast computation of exponentials we use the scaling-and-squaring approach (Arsigny et al., 2006). From Lie group manifolds, two relevant properties are derived. First, the inverse of a transform is equivalent to its negation in the log-space:

$$\mathcal{T}_l^{-1}(\boldsymbol{x}) = \exp[-\boldsymbol{T}_l(\boldsymbol{x})];$$

and second, in a scenario of small deformations, the composition of transforms can be approximated by truncating the Baker-Campbell-Hausdorff series at its first term (Vercauteren et al., 2008):

$$\mathcal{T}_l(\boldsymbol{x}) \circ \mathcal{T}_{l'}(\boldsymbol{x}) \approx \exp[\boldsymbol{T}_l(\boldsymbol{x}) + \boldsymbol{T}_{l'}(\boldsymbol{x})].$$

These two properties greatly simplify evaluation of the likelihood terms described below.

2.3.2. Likelihood models

We consider two different likelihood models, Gaussian and Laplacian. Both of them are based on two key assumptions. The first assumption is statistical independence across spatial locations and also between the horizontal and vertical components of

the transforms. We note that, in spite of such spatial independence, the smoothness of the solution will be guaranteed for two reasons: the observed registrations are often spatially smooth (leading to smooth solutions for $\{\mathcal{T}_l\}$), and the fact that we only solve the problem at a sparse set of control points, as explained in Section 2.7.2 below. The second assumption is that each observed registration is a noisy version of the "true" underlying transform, which is a composition of a subset of the hidden transforms, possibly inverted, as specified by the matrix $\boldsymbol{W}$ (as explained in Section 2.1 above).

Let $R_k^{\xi_1}(\boldsymbol{x})$ and $R_k^{\xi_2}(\boldsymbol{x})$ be the horizontal (ξ_1) and vertical (ξ_2) components of the SVF of registration k at $\boldsymbol{x}$, which we group into two $K \times 1$ vectors $\boldsymbol{R}^{\xi_j}(\boldsymbol{x}) = [R_1^{\xi_j}(\boldsymbol{x}), \ldots, R_K^{\xi_j}(\boldsymbol{x})]^T$, with $j \in \{1, 2\}$. In a similar fashion, let $T_l^{\xi_1}(\boldsymbol{x})$ and $T_l^{\xi_2}(\boldsymbol{x})$ be the two components of the SVF of the lth latent transform, grouped into two $L \times 1$ vectors $\boldsymbol{T}^{\xi_j}(\boldsymbol{x}) = [T_1^{\xi_j}(\boldsymbol{x}), \ldots, T_L^{\xi_j}(\boldsymbol{x})]^T$, with $j \in \{1, 2\}$. The model of spatial transforms in Section 2.3.1 enables us to write:

$$\boldsymbol{R}^{\xi_j}(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{T}^{\xi_j}(\boldsymbol{x}) + \boldsymbol{\zeta}^{\xi_j}(\boldsymbol{x}), \quad \text{with } j \in \{1, 2\},$$

where $\boldsymbol{\zeta}^{\xi_1}(\boldsymbol{x})$ and $\boldsymbol{\zeta}^{\xi_2}(\boldsymbol{x})$ are $K \times 1$ vectors with the horizontal and vertical components of errors in the SVFs of the K registrations at $\boldsymbol{x}$. The statistical distribution of the error $\boldsymbol{\zeta}$ (Gaussian or Laplacian) will shape the likelihood model, as described next.

Gaussian. The Gaussian model was presented in our previous conference article (Iglesias et al., 2018b). In short, it assumes that each error $\zeta_k^{\xi_j}(\boldsymbol{x})$ is independent from the others and follows a Gaussian distribution with zero mean and variance σ_k^2, such that the likelihood is:

$$\boldsymbol{R}^{\xi_j}(\boldsymbol{x}) \sim \mathcal{N}\left(\boldsymbol{W}\boldsymbol{T}^{\xi_j}(\boldsymbol{x}), \operatorname{diag}[\sigma_k^2]\right). \quad (4)$$

We use a model for the variances $\{\sigma_k^2\}$, which explicitly assumes that errors are larger when registering across modalities, or when registering slices or sections further apart. Specifically, the variance of a registration is assumed to be a linear combination of inter- and intra-modal variances:

$$\sigma_k^2 = c_k \sigma_{inter}^2 + \sum_{c=0}^{C} d_{k,c} \sigma_c^2,$$

where $c_k \in \{0, 1\}$ is an indicator variable which is equal to one if the registration is across modalities

(e.g., MRI to histology, or between different histological stains), and zero otherwise; $d_{k,c} \geq 0$ is the separation between the registered slices or sections along the stack for the intramodal registration of the c-th constrast, and zero otherwise; and σ_c^2, σ_c^2 are model parameters that need to be estimated, i.e., $\boldsymbol{\theta} = [\sigma_{\text{inter}}^2, \sigma_{c=0}^2, \cdots, \sigma_{c=C}^2]^T$.

Laplacian. In spite of the model for the variances, the Gaussian likelihood is sensitive to outliers in the registration. Such outliers occur frequently in histology due to common artifacts such as folding or tears. As an alternative, we propose a Laplacian model penalising the absolute value of the errors, i.e., the ℓ_1 norm:

$$R_k^{\xi_j}(\boldsymbol{x}) \sim \text{Laplace}\left(\boldsymbol{WT}^{\xi_j}(\boldsymbol{x}), b\right), \qquad (5)$$

where b is the scaling parameter of the Laplace distribution, which we consider constant across registrations; this assumption enables us to solve a single linear program per location during inference – rather than iteratively solving multiple linear programs. Moreover, modelling the dispersion of each registration separately as in the Gaussian case is not as important, due to the robustness of the ℓ_1 norm against outliers. Therefore, b is the only model parameter, i.e., $\boldsymbol{\theta} = [b]$.

2.4. Inference algorithms

Following the general inference framework in Section 2.2 and the design choices in Section 2.3, we now present two specific algorithms to solve the inference problem for the two proposed likelihood models.

2.4.1. Gaussian

Substituting the Gaussian likelihood from Equation 4 into the maximisation problem from Equation 3 and switching signs, we obtain the following cost function (Iglesias et al., 2018b):

$$C_{\ell_2}[\boldsymbol{T}^{\xi_1}(\boldsymbol{x}), \boldsymbol{T}^{\xi_2}(\boldsymbol{x}), \sigma_c^2, \sigma_d^2)] =$$
$$|\Omega| \sum_{k=1}^{K} \log[2\pi(c_k\sigma_c^2 + d_k\sigma_d^2)]$$
$$+ \sum_{j=1}^{2}\sum_{k=1}^{K}\sum_{\boldsymbol{x}\in\Omega} \frac{[R_k^{\xi_j}(\boldsymbol{x}) - \sum_{l=1}^{L} W_{kl}T_l^{\xi_j}(\boldsymbol{x})]^2}{2(c_k\sigma_c^2 + d_k\sigma_d^2)}. \qquad (6)$$

We use coordinate descent to solve this minimisation problem, alternately optimising for $\{\boldsymbol{T}^{\xi_j}\}_{j=1,2}$

and for $\boldsymbol{\theta} = [\sigma_c^2, \sigma_d^2]^T$, with the other fixed. For a constant $\boldsymbol{\theta}$, Equation 6 becomes a simple weighted least squares problem, with a closed-form solution given by:

$$T_l^{\xi_j}(\boldsymbol{x}) = \sum_{k=1}^{K} Z_{lk} R_k^{\xi_j}(\boldsymbol{x}), \qquad (7)$$

for the two spatial coordinates ξ_j, $j = 1, 2$. The regression matrix $\boldsymbol{Z}$ is by:

$$\boldsymbol{Z} = [\boldsymbol{W}^T \text{diag}(1/\sigma_k^2)\boldsymbol{W}]^{-1}\boldsymbol{W}^T\text{diag}(1/\sigma_k^2).$$

With the hidden transforms fixed, there is no closed-form expression for σ_c^2 and σ_d^2. However, Equation 6 becomes a smooth function of two variables that can be easily and quickly minimised with numerical methods, e.g., conjugate gradient (Shewchuk et al., 1994) or BFGS (Liu & Nocedal, 1989).

2.4.2. Laplacian

Substituting the Laplacian likelihood from Equation 4 into Equation 3, we obtain the following objective function:

$$\mathcal{O}_{\ell_1} = -2K|\Omega|\log(2b)$$
$$-\frac{1}{b}\sum_{j=1}^{2}\sum_{k=1}^{K}\sum_{\boldsymbol{x}\in\Omega}|R_k^{\xi_j}(\boldsymbol{x}) - \sum_{l=1}^{L} W_{kl}T_l^{\xi_j}(\boldsymbol{x})|. \qquad (8)$$

Since the values of the optimal latent transforms that minimise Equation 8 do not depend on the model parameter b, we can remove the terms related to b and switch signs to obtain the following cost function:

$$C_{\ell_1}[\boldsymbol{T}^{\xi_1}(\boldsymbol{x}), \boldsymbol{T}^{\xi_2}(\boldsymbol{x})] =$$
$$\sum_{j=1}^{2}\sum_{k=1}^{K}\sum_{\boldsymbol{x}\in\Omega}|R_k^{\xi_1}(\boldsymbol{x}) - \sum_{l=1}^{L} W_{kl}T_l^{\xi_1}(\boldsymbol{x})|, \quad (9)$$

which can be solved one spatial location $\boldsymbol{x}$ and direction (horizontal ξ_1 or vertical ξ_2) at the time. Crucially, the minimisation of Equation 9 can be rewritten as a linear program in standard form as follows:

$$\text{minimize } \boldsymbol{c}^T\boldsymbol{y}$$
$$\text{s. t. } \boldsymbol{A}_1^T\boldsymbol{y} \leq -\boldsymbol{R}^{\xi_j}(x),$$
$$\boldsymbol{A}_2^T\boldsymbol{y} \leq \boldsymbol{R}^{\xi_j}(x),$$

where:

- $\boldsymbol{y} = [D_1^{\xi_j}(\boldsymbol{x}), ..., D_K^{\xi_j}(\boldsymbol{x}), T_1^{\xi_j}(\boldsymbol{x}), ..., T_L^{\xi_j}(\boldsymbol{x})]^T$ is a $(K+L) \times 1$ vector concatenating the K absolute deviations , $D_k^{\xi_j}(\boldsymbol{x})$ (defined below), and the latent transforms to estimate, $T_l^{\xi_j}(\boldsymbol{x})$.

- $\boldsymbol{c} = [\mathbf{1}_K^T, \mathbf{0}_L^T]^T$, where $\mathbf{1}_K$ and $\mathbf{0}_L$ are the all-one and all-zero vectors with dimensions $K \times 1$ and $L \times 1$, respectively.

- $\boldsymbol{A}_1 = [-\boldsymbol{I}_K, -\boldsymbol{W}]$, where $\boldsymbol{I}_K$ is the $K \times K$ identity matrix.

- $\boldsymbol{A}_2 = [-\boldsymbol{I}_K, \boldsymbol{W}]$.

Since the vector $\boldsymbol{c}$ zeroes out the second part of $\boldsymbol{y}$, the objective function of this linear program and Equation 6 are identical. The inequality constraints effectively force the deviations $D_K^{\xi_j}(\boldsymbol{x})$ to be positive and equal to:

$$D_K^{\xi_j}(\boldsymbol{x}) = |R_k^{\xi_j}(\boldsymbol{x}) - \sum_{l=1}^{L} W_{kl} T_l^{\xi_j}(\boldsymbol{x})|.$$

Therefore, the linear program is equivalent to the problem of minimising $\mathcal{C}_{\ell_1}$ in Equation 9, with the difference that we can now use well-established linear programming algorithms to obtain the solution – which is simply the second part (last L elements) of the optimal $\boldsymbol{y}$.

2.5. Spatially-varying subgraphs

So far we have assumed any location in the image domain ($\boldsymbol{x} \in \Omega$) is modelled with the same static graph structure, encoded in the matrix $\boldsymbol{W}$. However, it is desirable to model in the graph image features and artefacts that may influence registration. For example, one may want to remove edges when modelling pixels far away from the masks $\{M_n^c\}$, which cannot be mapped reliably across images, or edges modelling pixels with artefacts such as folds and tears.

To tackle this problem, we propose to build spatially-varying subgraphs of $\mathcal{G}$ for each spatial location $\boldsymbol{x} \in \Omega$, which we represent as $\mathcal{G}(\boldsymbol{x})$. We also build companion matrices $\boldsymbol{W} = \boldsymbol{W}(\boldsymbol{x})$ that encode the relationship between the registrations and the hidden transforms at each location, considering the structures of $\mathcal{G}(\boldsymbol{x})$. These subgraphs are built by simply removing from $\mathcal{G}$ all edges going from $I_n^c(\boldsymbol{x})$ to $I_{n'}^{c'}(\boldsymbol{x})$, for which no tissue is present in the mask or the source image at location $\boldsymbol{x}$, i.e., when $M_n^c(\boldsymbol{x}) = 0$.

This approach easily accommodates the algorithm to cases for which we have irregular graphs; e.g., histological contrasts with different number of sections or missing correspondence between contrasts.

2.6. Registration networks

A learning-based strategy inspired by (Balakrishnan et al., 2019) is used to compute the SVF maps $\{\boldsymbol{R}_k(\boldsymbol{x})\}$. It consists of a backbone network (U-Net type, Çiçek et al. 2016) that produces a low-resolution SVF at 1/8 of the original resolution. Then, a rescaling layer with linear interpolation is used to get a full resolution velocity field. "Scaling and squaring" (Arsigny et al., 2006) is used to integrate the SVFs and compute the deformation fields $\{\mathcal{R}_k(\boldsymbol{x})\}$ used for spatial alignment. The inverse deformation field is computed by integrating the negated SVFs and used to induce symmetry in the training by evaluating the cost function at both reference and target image spaces. A local normalised cross-correlation and a smoothness regularisation term constraining the spatial gradient of the deformation are used as the composite loss function.

The learning approach is used to train a total of $C \times (C+1)/2$ registration networks on the entire dataset, one for every possible pair of contrasts. Intermodality networks are trained with randomly selected pairs of matching images I_n^c, $I_n^{c'}$ at every minibatch (i.e., with fixed n, and $c \neq c'$). Intramodality networks use randomly selected pairs of images, constrained to be within a maximum number of slices – in practice, we use 4, i.e., I_n^c, $I_{n'}^c$, with fixed c and $|n-n'| \leq 4$. In order to increase the generalisation ability of the models, all networks use an augmentation scheme whereby images are spatially deformed at every minibatch with a random, smooth, nonlinear displacement field, obtained by linear interpolation of a low resolution grid of control points (9×9) whose strength depends on the image resolution.

2.7. Implementation details and summary of the algorithm

2.7.1. Structure of the observational graph

As explained in Section 2.1, we compute K registrations between pairs of images in the graph. First, we compute intermodality registrations between all pairs of corresponding slices from different contrasts, i.e., between I_n^c and $I_n^{c'}$, with $c \neq c'$. And

second, we compute intramodality registrations between every pair of images of the same contrast, that are not more than P slices apart, i.e., between I_n^c and $I_{n'}^c$, with $|n - n'| \leq P$. The maximum separation P controls the smoothness of the solution (higher values of P yield smoother reconstruction but also more banana effect). An example of observational graph is shown in Figure 3.
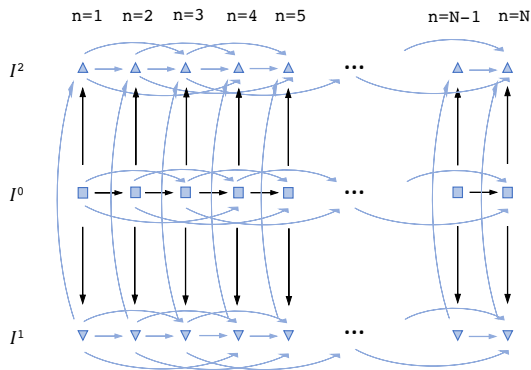

Figure 3: Example of the observational graph used with $C = 2$ and P=3. Arrows show all computed registrations in the graph with their corresponding direction and the spanning tree is highlighted in black.

2.7.2. Optimisation details

Since registration yields smoothly varying SVF maps (upscaled from 1/8 resolution, as explained in Section 2.6), we greatly reduce the computational requirements of our method by running the proposed inference algorithm on low-resolution SVFs. As in Section 2.6, we use linear interpolation to upsample the refined SVFs to the full resolution and use scaling and squaring to compute the final deformation fields.

In terms of optimisation approaches, we use different methods for the two likelihood models. For the Gaussian likelihood, we use an interative coordinate descent approach, where Equation 7 is used to update the latent transforms, and a quasi-Newton method (L-BFGS) is used to update the model parameters. In the *Laplacian* likelihood, we solve the linear program in Section 2.4.2 with the dual simplex method (Lemke, 1954).

2.7.3. Summary

The presented algorithm goes through a sequential number of steps. First, we build the graph and

its associated spanning tree by finding correspondent slices between contrasts (Section 2.1). Then, we compute the specified registrations above (Section 2.7.1) using the learning based approach introduced in Section 2.6. With this information and the likelihood model of choice in hand, we use Equation 3 to solve the inference problem at each voxel in the space of the low-resolution velocity fields. Finally, we upsample the solution with linear interpolation and integrate it with the scaling and squaring technique (Arsigny et al., 2006) to yield the final estimate of the latent deformation fields.

The package is written in Python and made publicly available in `https://github.com/acasamitjana/3dhirest`. Registration networks are optimised using PyTorch and the linear program is solved using the implementations found in the Gurobi package `https://www.gurobi.com/downloads/`.

3. Experiments and results

3.1. Data

To test and evaluate our framework, we use two datasets: one synthetic, which enables evaluation of registration errors with dense ground truth, and one real, to test the performance of the model on images with real distortions due to histological processing.

3.1.1. Synthetic dataset

The synthetic dataset consists of T1, T2 and FLAIR MRI scans from the training subset ($N = 285$) of the Brain Tumor Segmentation (BraTS) dataset (Menze et al., 2014). The T1 images with contrast ("T1c") are used as reference volumes, and the T2 and FLAIR images are used as a proxy for two histological contrasts. The BraTS scans are resampled (by the challenge organisers) to 1mm^3 isotropic resolution, and we define the z coordinate along the inferior-superior (I-S) direction. We generate synthetic 2D non-linear deformation fields using a grid of control points and B-spline interpolation (Prautzsch et al., 2002), independently for each axial slice of the T2 and FLAIR images, in order to mimic geometric distortion due histological processing (further details in Section 3.3 and Figure 7 below).

3.1.2. Real dataset

The real dataset is a set of publicly available images from the left hemisphere from a 34-year-old donor, distributed by the Allen institute (Ding

et al., 2016) at `http://atlas.brain-map.org/`. The dataset includes a multi-echo flash MRI scan acquired on a 7T scanner at 200 μm resolution, which we use as reference for the 3D histology reconstruction. Two different histological contrasts are also available at sub-μm in-plane resolution (which we resampled to 250 μm for convenience): 641 Nissl stained coronal sections with 200 μm spacing, and 287 coronal sections with 400 μm spacing and immunostained using parvalbumin. The correspondence between histological sections is found by matching the closest pair, with errors no superior than 0.1mm. For each section, we generate an associated tissue mask by thresholding and morphological operations. The final stacks of histological images do not have regular spacing due to missing sections, and there are a number of gaps of several mm without any sections at all, which divide the hemisphere into five different slabs.

In order to estimate the linear registration between the reference volume and histological stacks, we first used our previously presented algorithm (Tregidgo et al., 2020) to co-register the Nissl sections and the MRI. Once the MRI was linearly aligned to the stack, we used a block-matching algorithm (Ourselin et al., 2001) (as implemented in NiftyReg, Modat et al. 2010) to independently compute an affine transform between each histological section and its corresponding resampled MRI slice. Moreover, using the left hemisphere from the ICBM nonlinear 2009b symmetric atlas (Fonov et al., 2009), we compute the projection from the subject space to MNI coordinates.

Quantitative evaluation for this dataset is carried out using landmarks generated as follows. First, one salient point for each Nissl section (details below) is automatically sampled. Next, a first observer (JEI) marked the equivalent locations on the parvalbumin sections (where available) and on the resampled MRI. This generates a set of 641 pairs of landmarks for Nissl/MRI, and 287 for Nissl/parvalbumin, which can be used for quantitative evaluation. The same observer (JEI) reannotated the landmarks on a different day, for estimation of intra-observer variability. A second observer (AC) also annotated the same landmarks, for estimation of inter-observer variability.

The reference landmarks on the Nissl sections were sampled in a manner that ensured both salience and uniform spatial distribution. First, we applied a Harris corner detector (Harris et al., 1988) with a low threshold on the quality of the corner

(0.001) to every section. Next, we randomly sampled a location on every section with a uniform distribution, and centred on it a Gaussian kernel with standard deviation σ=20 pixels (i.e., 5 mm) in both x and y. We used this spatial Gaussian distribution to modulate (multiply) the scores from the Harris detector, and picked the landmark with the highest score.

3.2. Metrics

The synthetic dataset enables dense quantitative evaluation of the 3D reconstruction methods at the pixel level. Let $\boldsymbol{\phi}_n^c(\boldsymbol{x})$ represent the ground truth 2D deformation field between the reference slice n and histological contrast c, and let $\hat{\boldsymbol{\phi}}_n^c(\boldsymbol{x})$ be the deformation field estimated by an algorithm. To assess the performance of the presented framework, we define different performance metrics.

We first define the *pixel-wise error* as the bivariate deviation of the estimate from the ground truth deformation field at each pixel. Hence, for slice n, contrast c, and location x, we have:

$$\boldsymbol{e}_n^c(\boldsymbol{x}) = \boldsymbol{\phi}_n^c(\boldsymbol{x}) - \hat{\boldsymbol{\phi}}_n^c(\boldsymbol{x}). \qquad (10)$$

Based on this error, we define two performance metrics. First, the *intra-slice error*, defined as a global average over all pixels and slices of the module of the pixel-wise error. The metric is computed only within the tissue mask that results from the intersection of the reference and registered slices, resulting in a valid domain Ω_n for each slice:

$$E_W^c = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{|\Omega_n|} \sum_{x \in \Omega_n} \|\boldsymbol{e}_n^c(\boldsymbol{x})\| \qquad (11)$$

We also want to measure the consistency across slices, something that the *intra-slice error* doesn't capture. For this purpose, we use a second metric referred to as *inter-slice error*, which measures the error consistency across the direction of the stack. Intuitively, the *inter-slice error* measures the smoothness of the reconstruction, by comparing the consistency of the errors across neighbouring slices – if errors are consistent, the reconstruction is smooth. Its specific definition is:

$$E_B^c =$$
$$\frac{1}{N-1} \sum_{n=1}^{N-1} \frac{1}{|\Omega_n|} \sum_{x \in \Omega_n} \|\boldsymbol{e}_n^c(\boldsymbol{x}) - \boldsymbol{e}_{n+1}^c(\boldsymbol{x})\|. \qquad (12)$$

These two metrics complement each other in measuring the trade-of between accuracy and smoothness of the recovered 3D volume.

We compute them not only between the reference volume and the two (surrogate) histological stains (i.e., T2, FLAIR), but also between these two contrasts - which is easily achieved by defining:

$$\phi_n^{c,c'} = (\phi_n^c)^{-1} \circ \phi_n^{c'}$$

$$\boldsymbol{e}_{n+1}^{c,c'}(\boldsymbol{x}) = \phi_n^{c,c'}(\boldsymbol{x}) - \hat{\phi}_n^{c,c'}(\boldsymbol{x})$$

and use the estimation error to compute $E_B^{c,c'}$ and $E_W^{c,c'}$

3.3. Experimental setup

To generate the synthetic deformations in the BraTS dataset, we independently deform each axial slice of the T2 and FLAIR with 2D deformation fields generated as follows. First, normally distributed, low-resolution deformation fields of size $9 \times 9 \times 2$ are generated independently for each slice, subject and modality (T2, FLAIR). Each element in the low-resolution field is an independent Gaussian variable with zero mean and a standard deviation which is constant for each slice, and which is sampled from a uniform distribution $\mathcal{U}[3,7]$ (in pixels). Each deformation field is then resized to the original image size using B-Spline interpolation. We explicitly avoid using velocity fields to prevent imitating the deformation model used in the algorithm. The final volume is built by applying the deformations and resampling with bilinear interpolation.

To measure the robustness of our framework against large registration errors, we introduce outliers in the synthetic dataset. Specifically, we further distort subsets of the T2 and FLAIR slices (2%, 5%, 10% and 20%, to test scenarios with increasing number of outliers) by applying large random rotations (90°, 180°, or 270°), which lead to large errors in the nonlinear registration algorithm. Each modality is distorted independently. Evaluation is carried out only on the undistorted slices.

In order to analyse the impact of the deep learning registration techniques, we consider two additional diffeomorphic registration algorithms:

- *NiftyReg (NR)* (Modat et al. 2012): we use NiftyReg with SVF parameterisation ("–vel" option), control point spacing of 8 pixels, local normalised cross correlation similarity metric and bending energy penalty term. This setup makes the NR model as similar to our deep learning registration framework as possible.

- *Registration Networks* (RegNet): we use learning-based registration networks, trained as reported in Section. 2.6.

In our experiments, we compare slice-wise registration using these two algorithms, with different versions of our Spanning Tree framework (henceforth, ST):

- ST2-L2: single-contrast framework presented in Iglesias et al. (2018b) using masks in the graph and a Gaussian likelihood for the registrations. We run the algorithm independently for each histological contrast.

- ST2-L1: same as ST2-L2 but with a Laplacian likelihood.

- ST3-L2: multi-contrast framework presented in this work (joint registration of contrasts) using masks and a Gaussian distribution for the registrations.

- ST3-L1: same as ST3-L2 but with a Laplacian likelihood.

3.4. Results on synthetic dataset

In Fig. 4 we compare different configurations of the algorithm using NR and RegNet as baseline registration methods. RegNet, which is trained specifically for the problem at hand (as opposed to used a generic optimiser) provides better initial alignment between modalities than NR. Over the initial linear alignment, RegNet improves the *intra-slice* error by 49% and 42% for T2 and FLAIR modalities, respectively, compared to the 37% and 8% using NR. Moreover, estimation error variability in RegNet is consistently $\sim$ 5x times lower than in NR. According to these results, the dependence of standard intermodality registration algorithms (e.g., NiftyReg) on the contrast and image appearance can be partly mitigated using learning-based approaches. Further refinement can be achieved using the presented framework with different modelling options. In our previous work, we used a single-contrast approach with the ℓ_2-norm as a cost function (ST2-L2), yielding around 10% extra improvement in each modality.

The extensions introduced in this work improve upon that result in two ways: *(i)* the ℓ_1 norm
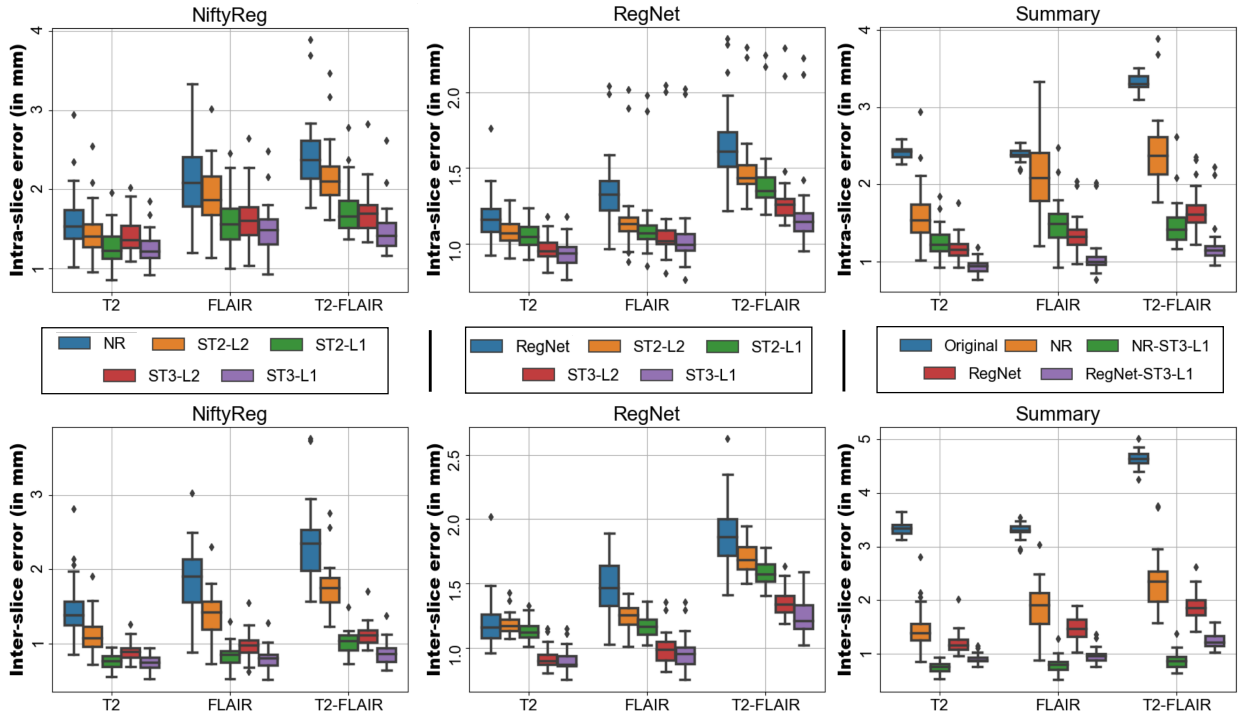
Figure 4: Top row: the first two columns show the *intra-slice* the *inter-slice* error for NiftyReg and RegNet when used in isolation (i.e., for each slice and contrast independently), and when refined with the different variants of the algorithms proposed in this paper. The last column shows a summary of previous plots comparing the errors for the distorted images, the registration approaches used in isolation, and the proposed approach ST3-L1. Each metric is computed for T2 and FLAIR modalities as well as the consistency of the error between them.

can correct larger registration errors with respect to direct registrations (no refinement); and *(ii)* a multi-contrast framework can correct registration errors in one modality using redundant measurements from other modalities. We perform statistical significance analysis between all methods compared, with the multi-contrast framework using the ℓ_1 norm (ST3-L1) significantly outperforming all other algorithms ($p < 0.01$), both with NR and RegNet. Over the initial registrations, ST3-L1 yield improvements of 19% and 22% (RegNet) and 23% and 29% (NR) on the *intra-slice* error, for T2 and FLAIR, respectively. NR yields smoother reconstructions when combined with our proposed refinement method, while RegNet achieves better inter-modality alignment (Fig. 4, summary). Unless otherwise specified, we use RegNet with ST3-L1 as default configuration for the proposed method throughout the rest of this paper.

Figure 5 shows the results of varying the number of neighbours in the observational graph, P. This parameter represent a trade-off between sev-

eral factors: smoothness, banana effect, z-shift error accumulation, accuracy and computational requirements. For example, the larger the redundancy in the graph (larger P) the more robust and smooth the method is at the cost of straighten curved structures. The baseline, $P = 0$, builds a disconnected graph using the raw intermodality observations from RegNet directly. The results in this figure show that the optimal number of neighbours depends on the likelihood function. In the ℓ_2-norm case, $P = 1$ is the optimal number of neighbours for the trade-off between smoothness and banana effect. However, the ℓ_1-norm is more robust to errors in intramodality registration with increased number of neighbours in the observational graph. We use a paired t-test to quantify statistically significant mean differences and found that $P = 2$ is the optimal number of neighbours for both the *inter-slice* and *intra-slice* error when using the Laplacian likelihood; from $P \geq 3$ the z-shift error start to accumulate. Therefore, we will use $P = 2$ (combined with RegNet and ST3-L1) throughout the rest of
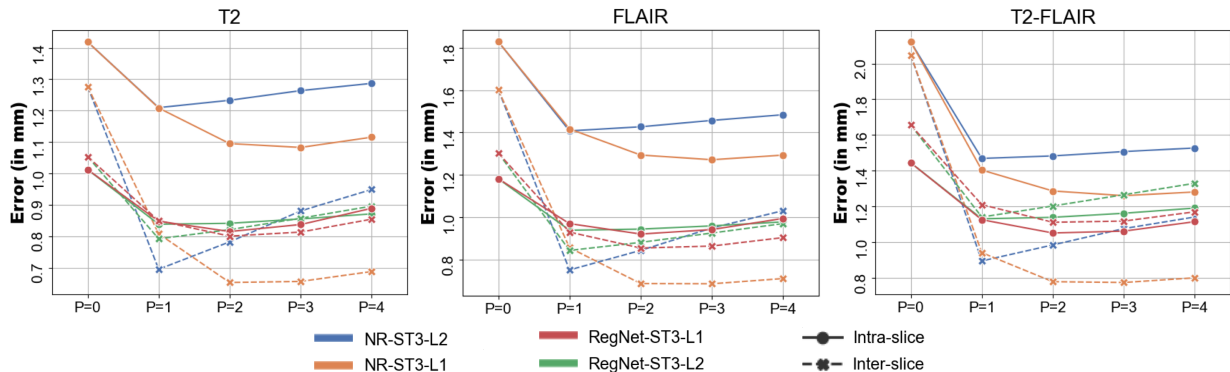
Figure 5: *Inter-* (dashed) and *intra-slice* (solid) errors as a function of neighbours in the observational graph (P), for the different versions of our algorithm. Setting $P = 0$ is equivalent to running the algorithm independently for each slice.

this manuscript, unless explicitly specified.

In order to test the robustness of the framework against outliers, we used RegNet for the initial registrations and compared different algorithm configurations against increasing rates of outliers. Fig. 6 shows the *inter-* and *intra-slice* error for each contrast as well as the error consistency across contrasts. Errors grow with the proportion of outliers, as expected, but the rate at which performance decreases is different for the different configurations. The ℓ_1-norm appears to be robust to a considerable number of outliers (up to 10-20%). The ℓ_2-norm is much more sensitive to outliers and its performance decreases much faster. We note that the higher error of the ST3 configurations is partly due to the fact that the higher number of registrations K means that more outliers are present in the observational graph. Nonetheless, ST3 appears to improve the error consistency and, when combined with the robust ℓ_1-norm, it boosts the performance of the algorithm.

To sum up, the results on the synthetically deformed data show that ST3-L1 with RegNet achieves the highest consistency across contrasts and presents the best trade-off between the different factors such as accuracy and smoothness. It outperforms our previous method from Iglesias et al. (2018b), which is based on NiftyReg and ST2-L2. Qualitative results comparing both algorithms on a case from BraTS are shown in Fig. 7

3.5. Results on Allen atlas

For 3D reconstruction of the histology of the Allen atlas, we use RegNet as the baseline registration algorithm and ST3-L1 configuration with $P = 4$ neighbours in each of the three stacks of images; we increase P with respect to the BraTS dataset to compensate for the lower spacing between sections. Due to a number of larger gaps between sections in this dataset, this process yields a disconnected graph with 5 separate slabs. Therefore, each slab can be processed independently, which reduces the memory footprint of the algorithm.

Qualitative results in the sagittal and axial planes are shown in Figures 8 and 9, where the reconstructed volumes are resampled at 0.25mm isotropic resolution. 3D histology stacks of both Nissl and parvalbumin contrast appear to be aligned with the reference volume at the same time that provide smooth reconstructions. A close-up in the temporal lobe shows that, even though RegNet provides good initial alignment in more homogeneous areas (e.g., cerebral cortex), more jagged reconstructions are recovered in heterogeneous areas such as the hippocampus and the cerebellum. The framework presented here is able to smooth out some of these effects

The algorithm not only recovers smooth reconstructions, but also produces registrations that are accurate and robust. Figure 10 shows the registration of a heavily damaged coronal section in plane, i.e., in coronal view. Despite the presence of strong artefacts, particularly for the parvalbumin stain (inhomogeneous staining, torn tissue, big cracks), our method is able to produce smooth deformations that are robust against these artefacts and yield accurate 3D reconstructions.

Finally, Figure 11 shows boxplots for the intra-slice errors computed from the manually placed landmarks, and compares them to the inter- and intra-observer variabilities. Even though the me-
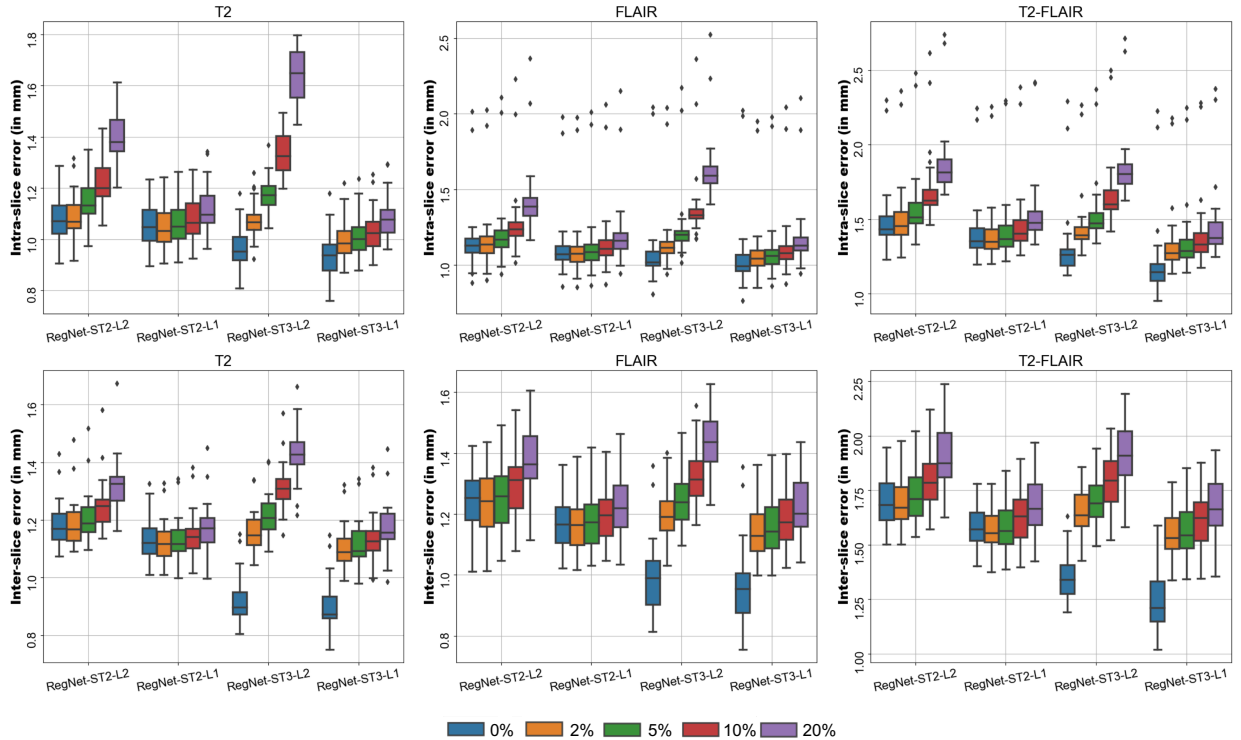
13

Figure 6: Comparison of the *intra-slice* error (top row) and *inter-slice* error (bottom row) for different level of outliers per modality (0-20%). Different configurations of the algorithm using RegNet as base algorithm are tested (ST2-L2, ST2-L1, ST3-L2, ST3-L1) and the errors are independently reported for each modality (T2, FLAIR) as well as for the consistency of the error between them.

dian error does not improve much with respect to RegNet (as it did on the synthetic dataset, Figure 7), there is a small decrease in outliers after refinement of the registrations with our proposed approach. This is despite the fact that the additional smoothness imposed by our approach (apparent from Figures 8 and 9) necessarily represents a trade-off with accuracy. Compared with the initial affine alignment, our method achieves reductions of 17% in Nissl/MR registration, 17% in parvalbumin/MRI registration, and 25% in Nissl/parvalbumin in the median error. While these differences may not seem large at first, they have a very noticeable impact on the quality of the output (again, see Figures 8 and 9).

In absolute terms, our approach achieves median registration errors of approximately 1 mm or below: 0.89 mm for Nissl/MRI, 1.09 mm for parvalbumin/MRI, and 0.89 mm for Nissl/parvalbumin. These are approximately within 0.5 mm of the intra- and inter-observer variability (0.71/0.79 mm for Nissl/MRI, and 0.59/0.73 mm for

Nissl/parvalbumin). Using NiftyReg as a base algorithm, landmark misalignment is increased mostly due to errors in the initial observations, with the exception of the Nissl/parvalbumin registration that achieves an inter-observer level of error.

4. Discussion and conclusion

In this manuscript, we have presented a probabilistic model for joint registration of image stacks in histology and MRI. Volumetric histology reconstruction is posed as a Bayesian inference problem, which can be solved analytically (ℓ_2-norm) or with standard linear programming techniques (ℓ_1-norm). The algorithm effectively balances banana effect, z-shift effect and registration accuracy to produce 3D reconstruction that are booth smooth and precise – and also robust, if the version with the ℓ_1-norm is used.

Our algorithm builds on standard registration techniques, with the only requirement that they parameterise the deformations with stationary ve-
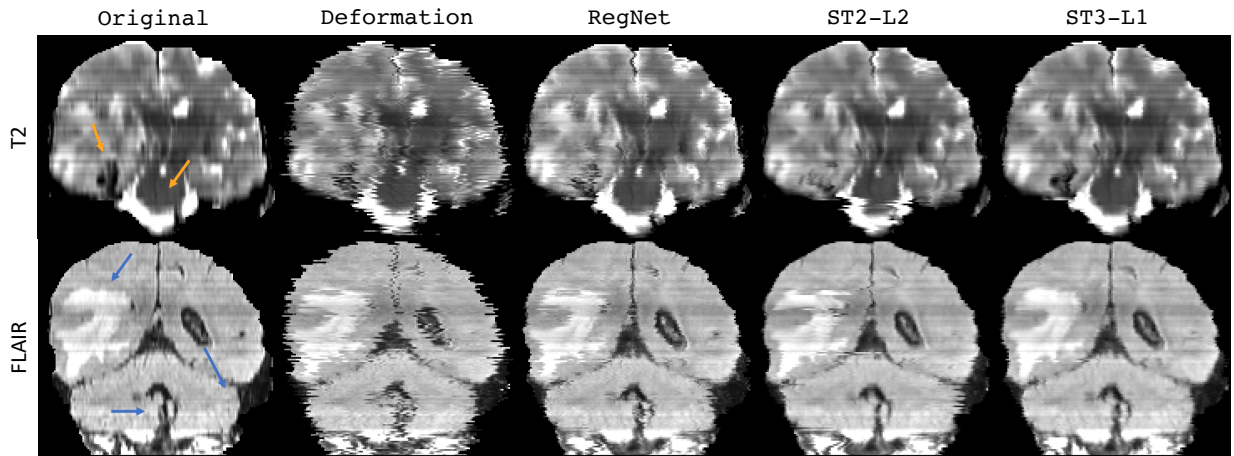
Figure 7: Reconstructed coronal view for the T2 and FLAIR scans of a sample subject from BraTS (Menze et al., 2014). From left to right we show the original image, the original image with synthetic deformations using B-Splines, the initial registrations using RegNet, the output of the ST2-L2 / NiftyReg algorithm presented in Iglesias et al. (2018b), the output of the ST3-L1 (RegNet) algorithm presented here, and the ground truth for each contrast. The arrows point at tumorous areas and other regions (e.g., brain stem, cerebellum) which are shown to be hard to register.
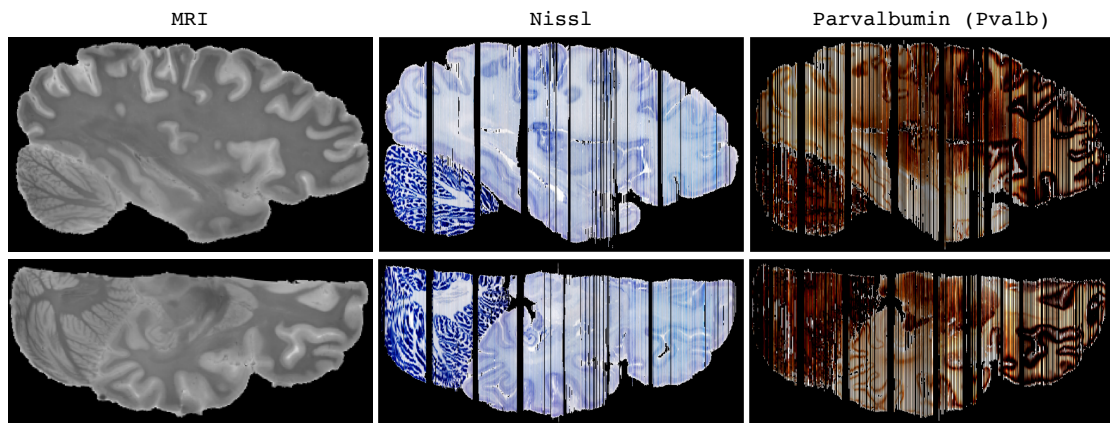
Figure 8: Sagittal and axial views of reference MRI and the reconstruction of the available contrasts (Nissl and parvalbumin), using RegNet combined with ST3-L1.

locity fields. In our experiments, we have compared a classical algorithm building on explicit optimisation (NiftyReg) and an unsupervised machine learning approach building on modern neural networks (RegNet). The latter learns a global, data-dependent deformation model that, in practice, is shown to be more accurate and superior to generic optimisation algorithms such as the conjugate gradient strategy used by NiftyReg.

While the neural network requires (unsupervised) training, it provides quick predictions for the measurements $\{\mathcal{R}_k(\boldsymbol{x})\}$ (i.e., the pairwise registrations), which can be computed in the order of seconds to a few minutes, depending on the size of the problem (number of images, number of contrasts, size of the images, etc.). NiftyReg is between one and two orders of magnitude slower, as it iteratively optimises the deformation for every $\mathcal{R}_k$, such that computing the whole set of registrations take between minutes and hours. Such differences are reduced when accounting for network training times, being more efficient for larger datasets e.g., several minuts per subjects in the BraTS dataset. Once the registrations have been computed, solving the linear program at the control points takes approximately 20 seconds for a BraTS case on an Intel
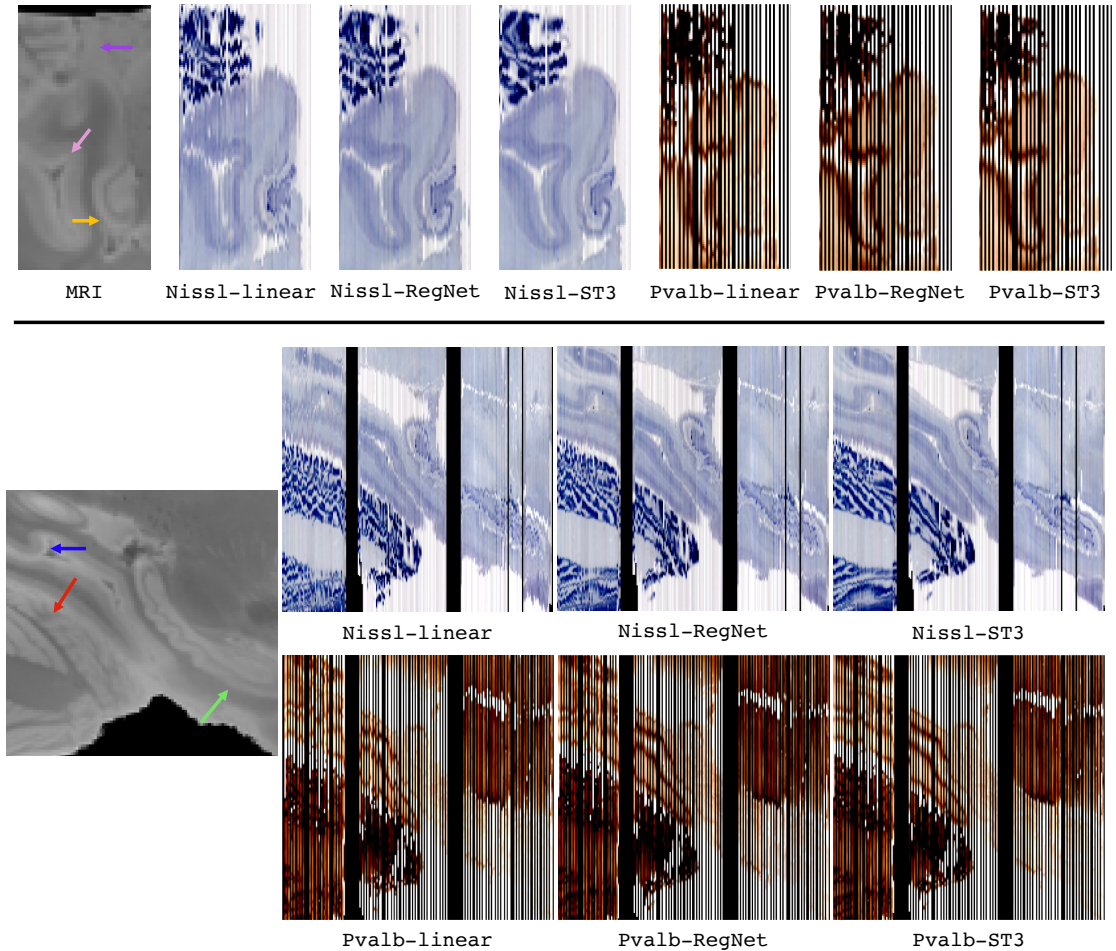
Figure 9: Close-up of 3D reconstructions produced by linear alignment, RegNet alone and its combination with the proposed approach (ST3-L1). Heterogeneous regions such as the cerebellum (red and magenta arrow) or the hippocampus (green and yellow arrows) as well as tissue boundaries (blue and pink arrows) are corrected by the algorithm.

Core i7-9800X processor with 8 cores (150 minutes if the problem is solved at every pixel instead).

The extension to multiple contrasts has two main implications. First, there is an increasing demand of computation and memory, as the number of observations increase quadratically with the number of contrasts. And second, there is an increase of the number of cycles in the observation graph, making it more redundant and robust. In other words: since the number of latent variables increase linearly (rather than quadratically), the inference model becomes progressively more overdetermined. Moreover, a multi-contrast framework is shown to improve the consistency between reconstructed volumes for different histological contrasts as seen in Fig. 4. The use of an ℓ_1-norm further increases the robustness of the framework in the presence of low

to moderate levels of outliers. No running time difference has been found in practice between using an ST3 approach or solving the inference problem independently for each contrast using ST2.

The usefulness of the framework we have presented has been shown with a publicly available real test case from the Allen human brain atlas, which has two available contrasts (stains). Despite the fact that the two stains are sampled at different frequencies (yielding an irregular graph structure) and that typical histology artefacts are found in both contrasts (e.g., tears, folding, cracks, inhomogeneous staining), our method is able to recover the original 3D shape with smooth transitions between slices. We have shared our results through the OpenNeuro repository (https://openneuro.org/datasets/ds003590) along
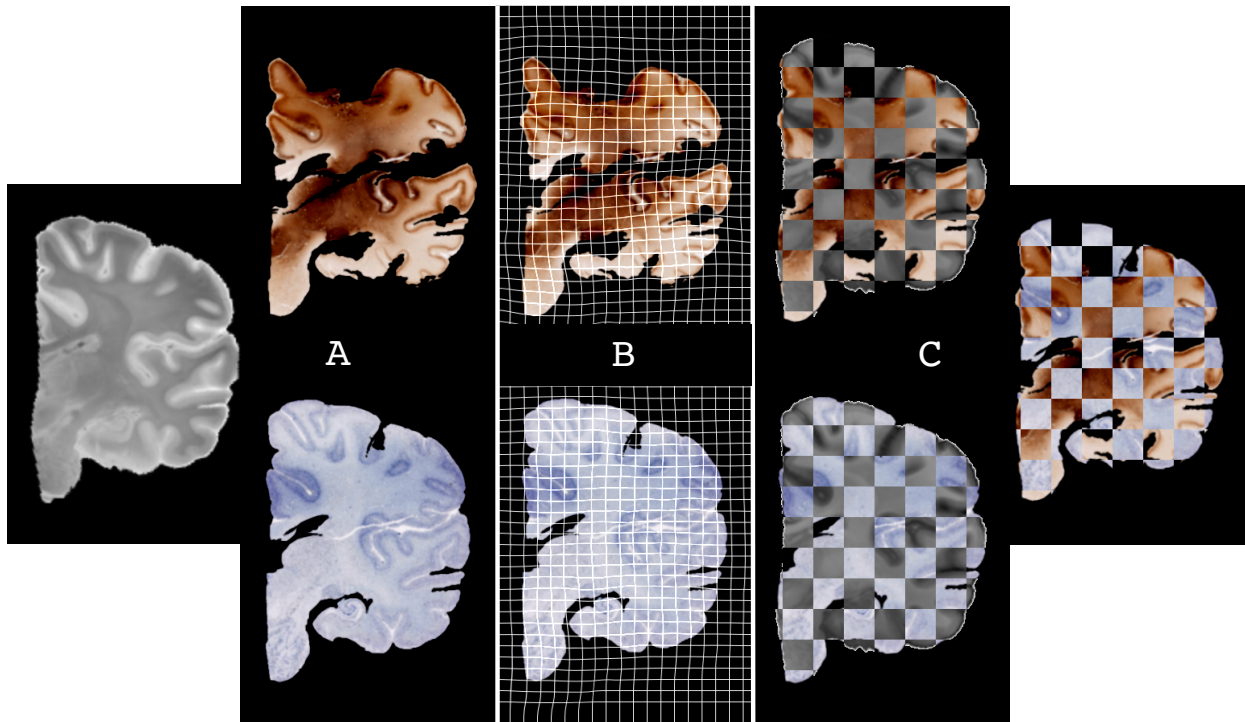
16

Figure 10: Coronal plane view of (A) section 1624 of the Allen human brain atlas for the Parvalbumin (above) and Nissl (below) stains and the correspondent MRI slice (middle), (B) the deformation field superimposed on the original section image and (C) checkerboard visualisation of the aligned histology section and MRI slice and the cross-alignment between contrasts.
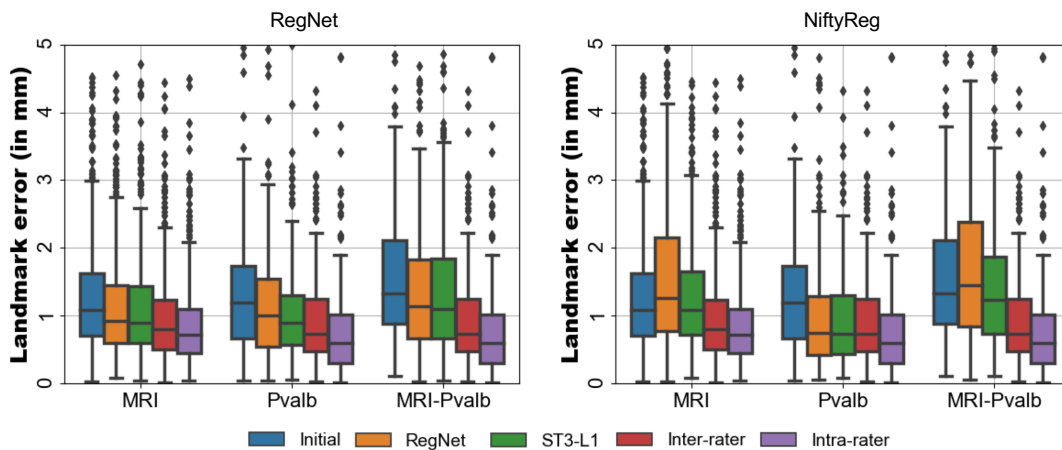

Figure 11: Landmark error using RegNet (left) and NiftyReg (right) as base algorithms for alignment between the Nissl/MRI, Nissl/parvalbumin and MRI/parvalbumin. Inter-slice and intra-slice errors are computed as stated in Section 3.1.2.

with the mapping to MNI space, providing the neuroimaging community with a cross-scale link between the two atlases.

The framework builds upon a set of assumptions, such as the independence between registration noise and spatial location, or the conditional independence of the observations given the latent variables. While these assumptions are violated to different extents in different scenarios (application, base registration algorithm), the results on the two datasets have shown that our proposed method works well in practice. We speculate that the different degrees

of departure from the assumptions may be behind some of our empirical results.

Extensions of this work can follow several directions, being the first one evaluating learning-based registration-by-synthesis approaches to improve initial intermodality alignment (Qin et al., 2019; Xu et al., 2020). A second direction is to consider a more realistic approach to model the registration errors that accounts for spatial correlations. A third step would be to integrate intensity homogenisation techniques in the framework to jointly correct for uneven staining. Moreover, imputation methods to fill the gaps and improve continuity between sections may be very well explored.

We plan to use this framework on the 3D histology reconstruction pipeline introduced by Mancini et al. (2020) to build a μm-resolution probabilistic atlas of the human brain. The accurate 3D reconstruction of histological atlases with the proposed method will enable volumetric studies of the whole human brain at the subregion level, with much higher specificity than current approaches.

References

Adler, D. H., Pluta, J., Kadivar, S., Craige, C., Gee, J. C., Avants, B. B., & Yushkevich, P. A. (2014). Histology-derived volumetric annotation of the human hippocampal subfields in postmortem MRI. *Neuroimage*, *84*, 505–523.

Adler, D. H., Wisse, L. E., Ittyerah, R., Pluta, J. B., Ding, S.-L., Xie, L., Wang, J., Kadivar, S., Robinson, J. L., Schuck, T. et al. (2018). Characterizing the human hippocampus in aging and Alzheimer's disease using a computational atlas derived from ex vivo MRI and histology. *Proceedings of the National Academy of Sciences*, *115*, 4252–4257.

Alic, L., Haeck, J. C., Bol, K., Klein, S., van Tiel, S. T., Wielepolski, P. A., de Jong, M., Niessen, W. J., Bernsen, M., & Veenland, J. F. (2011). Facilitating tumor functional assessment by spatially relating 3D tumor histology and in vivo MRI: image registration approach. *PLoS One*, *6*, e22835.

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. et al. (2013). Big-Brain: an ultrahigh-resolution 3D human brain model. *Science*, *340*, 1472–1475.

Annese, J. (2012). The importance of combining MRI and large-scale digital histology in neuroimaging studies of brain connectivity and disease. *Frontiers in Neuroinformatics*, *6*, 13.

Arsigny, V., Commowick, O., Pennec, X., & Ayache, N. (2006). A log-euclidean framework for statistics on diffeomorphisms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 924–931). Springer.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, *38*, 95–113.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, *12*, 26–41.

Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., & Dalca, A. V. (2019). Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, *38*, 1788–1800.

Bancroft, J. D., & Gamble, M. (2008). *Theory and practice of histological techniques*. Elsevier health sciences.

Bağci, U., & Bai, L. (2010). Automatic best reference slice selection for smooth volume reconstruction of a mouse

brain from histological images. *IEEE Transactions on Medical Imaging*, *29*, 1688–1696.

Bourne, R. M., Bailey, C., Johnston, E. W., Pye, H., Heavey, S., Whitaker, H., Siow, B., Freeman, A., Shaw, G. L., Sridhar, A. et al. (2017). Apparatus for histological validation of in vivo and ex vivo magnetic resonance imaging of the human prostate. *Frontiers in oncology*, *7*, 47.

Casero, R., Siedlecka, U., Jones, E. S., Gruscheski, L., Gibb, M., Schneider, J. E., Kohl, P., & Grau, V. (2017). Transformation diffusion reconstruction of three-dimensional histology volumes from two-dimensional image stacks. *Medical image analysis*, *38*, 184–204.

Ceritoglu, C., Wang, L., Selemon, L. D., Csernansky, J. G., Miller, M. I., & Ratnanather, J. T. (2010). Large deformation diffeomorphic metric mapping registration of reconstructed 3D histological section images and in vivo MR images. *Frontiers in human neuroscience*, *4*, 43.

Chan, J. K. (2014). The wonderful colors of the hematoxylin–eosin stain in diagnostic surgical pathology. *International journal of surgical pathology*, *22*, 12–32.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention* (pp. 424–432). Springer.

Cooper, L., Sertel, O., Kong, J., Lozanski, G., Huang, K., & Gurcan, M. (2009). Feature-based registration of histopathology images with different stains: An application for computerized follicular lymphoma prognosis. *Computer methods and programs in biomedicine*, *96*, 182–192.

Dalca, A. V., Balakrishnan, G., Guttag, J., & Sabuncu, M. R. (2018). Unsupervised learning for fast probabilistic diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 729–738). Springer.

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. et al. (2016). Comprehensive cellular-resolution atlas of the adult human brain. *Journal of Comparative Neurology*, *524*, 3127–3481.

Feuerstein, M., Heibel, H., Gardiazabal, J., Navab, N., & Groher, M. (2011). Reconstruction of 3-D histology images by simultaneous deformable registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 582–589). Springer.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, (p. S102).

Gibson, E., Gaed, M., Gómez, J. A., Moussa, M., Pautler, S., Chin, J. L., Crukley, C., Bauman, G. S., Fenster, A., & Ward, A. D. (2013). 3D prostate histology image reconstruction: quantifying the impact of tissue deformation and histology section location. *Journal of pathology informatics*, *4*.

Goubran, M., Crukley, C., de Ribaupierre, S., Peters, T. M., & Khan, A. R. (2013). Image registration of ex-vivo MRI to sparsely sectioned histology of hippocampal and neocortical temporal lobe specimens. *Neuroimage*, *83*, 770–781.

Harris, C. G., Stephens, M. et al. (1988). A combined corner and edge detector. In *Alvey vision conference* (pp. 10–5244). Citeseer volume 15.

Iglesias, J. E., Insausti, R., Lerma-Usabiaga, G., Bocchetta, M., Van Leemput, K., Greve, D. N., Van der Kouwe, A., Fischl, B., Caballero-Gaudes, C., Paz-Alonso, P. M. et al. (2018a). A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology. *Neuroimage*, *183*, 314–326.

Iglesias, J. E., Lorenzi, M., Ferraris, S., Peter, L., Modat, M., Stevens, A., Fischl, B., & Vercauteren, T. (2018b). Model-based refinement of nonlinear registrations in 3D histology reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 147–155). Springer.

Jacobs, M. A., Windham, J. P., Soltanian-Zadeh, H., Peck, D. J., & Knight, R. A. (1999). Registration and warping of magnetic resonance images to histological sections. *Medical physics*, *26*, 1568–1578.

Krauth, A., Blanc, R., Poveda, A., Jeanmonod, D., Morel, A., & Székely, G. (2010). A mean three-dimensional atlas of the human thalamus: generation from multiple histological data. *Neuroimage*, *49*, 2053–2062.

Lemke, C. E. (1954). The dual method of solving the linear programming problem. *Naval Research Logistics Quarterly*, *1*, 36–47.

Liu, D. C., & Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical programming*, *45*, 503–528.

Malandain, G., Bardinet, E., Nelissen, K., & Vanduffel, W. (2004). Fusion of autoradiographs with an MR volume using 2-D and 3-D linear transformations. *NeuroImage*, *23*, 111–127.

Mancini, M., Casamitjana, A., Peter, L., Robinson, E., Crampsie, S., Thomas, D. L., Holton, J. L., Jaunmuktane, Z., & Iglesias, J. E. (2020). A multimodal computational pipeline for 3d histology of the human brain. *Scientific reports*, *10*, 1–21.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R. et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, *34*, 1993–2024.

Modat, M., Daga, P., Cardoso, M. J., Ourselin, S., Ridgway, G. R., & Ashburner, J. (2012). Parametric non-rigid registration using a stationary velocity field. In *2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis* (pp. 145–150). IEEE.

Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C., & Ourselin, S. (2010). Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, *98*, 278–284.

Montine, T. J., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Dickson, D. W., Duyckaerts, C., Frosch, M. P., Masliah, E., Mirra, S. S. et al. (2012). National institute on aging–Alzheimer's association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta neuropathologica*, *123*, 1–11.

Ourselin, S., Roche, A., Subsol, G., Pennec, X., & Ayache, N. (2001). Reconstructing a 3D structure from serial histological sections. *Image and vision computing*, *19*, 25–31.

Pichat, J., Iglesias, J. E., Yousry, T., Ourselin, S., & Modat, M. (2018). A survey of methods for 3D histology reconstruction. *Medical image analysis*, *46*, 73–105.

Prautzsch, H., Boehm, W., & Paluszny, M. (2002). *Bézier and B-spline techniques*. Springer Science & Business Media.

Qin, C., Shi, B., Liao, R., Mansi, T., Rueckert, D., & Kamen, A. (2019). Unsupervised deformable registration for multi-modal images via disentangled representations. In *International Conference on Information Processing in Medical Imaging* (pp. 249–261). Springer.

Ramos-Vara, J. (2005). Technical aspects of immunohistochemistry. *Veterinary pathology*, *42*, 405–426.

Rusu, M., Golden, T., Wang, H., Gow, A., & Madabhushi, A. (2015). Framework for 3D histologic reconstruction and fusion with in vivo MRI: Preliminary results of characterizing pulmonary inflammation in a mouse model. *Medical physics*, *42*, 4822–4832.

Shewchuk, J. R. et al. (1994). An introduction to the conjugate gradient method without the agonizing pain.

Shojaii, R., Bacopulos, S., Yang, W., Karavardanyan, T., Spyropoulos, D., Raouf, A., Martel, A., & Seth, A. (2014). Reconstruction of 3-dimensional histology volume and its application to study mouse mammary glands. *JoVE (Journal of Visualized Experiments)*, (p. e51325).

Song, Y., Treanor, D., Bulpitt, A. J., & Magee, D. R. (2013). 3D reconstruction of multiple stained histology images. *Journal of pathology informatics*, *4*.

Stille, M., Smith, E. J., Crum, W. R., & Modo, M. (2013). 3D reconstruction of 2D fluorescence histology images and registration with in vivo MR images: application in a rodent stroke model. *Journal of neuroscience methods*, *219*, 27–40.

Thiele, H., Heldmann, S., Trede, D., Strehlow, J., Wirtz, S., Dreher, W., Berger, J., Oetjen, J., Kobarg, J. H., Fischer, B. et al. (2014). 2D and 3D MALDI-imaging: conceptual strategies for visualization and data mining. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1844*, 117–137.

Tregidgo, H. F., Casamitjana, A., Latimer, C. S., Kilgore, M. D., Robinson, E., Blackburn, E., Van Leemput, K., Fischl, B., Dalca, A. V., Mac Donald, C. L. et al. (2020). 3D reconstruction and segmentation of dissection photographs for MRI-free neuropathology. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 204–214). Springer.

Vercauteren, T., Pennec, X., Perchant, A., & Ayache, N. (2008). Symmetric log-domain diffeomorphic registration: A demons-based approach. In *International conference on medical image computing and computer-assisted intervention* (pp. 754–761). Springer.

Wang, J., MacKenzie, J. D., Ramachandran, R., & Chen, D. Z. (2016). A deep learning approach for semantic segmentation in histology tissue images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 176–184). Springer.

Wirtz, S., Fischer, B., Modersitzki, J., & Schmitt, O. (2004). Superfast elastic registration of histologic images of a whole rat brain for 3D reconstruction. In *Medical Imaging 2004: Image Processing* (pp. 328–334). International Society for Optics and Photonics volume 5370.

Xu, Z., Luo, J., Yan, J., Pulya, R., Li, X., Wells, W., & Jagadeesan, J. (2020). Adversarial uni-and multi-modal stream networks for multimodal image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 222–232). Springer.

Yang, Z., Richards, K., Kurniawan, N. D., Petrou, S., & Reutens, D. C. (2012). MRI-guided volume reconstruction of mouse brain from histological sections. *Journal of neuroscience methods*, *211*, 210–217.

Yelnik, J., Bardinet, E., Dormont, D., Malandain, G., Ourselin, S., Tandé, D., Karachi, C., Ayache, N., Cornu, P., & Agid, Y. (2007). A three-dimensional, histological and deformable atlas of the human basal ganglia. i. atlas construction based on immunohistochemical and MRI data. *Neuroimage*, *34*, 618–638.

Yushkevich, P. A., Avants, B. B., Ng, L., Hawrylycz, M., Burstein, P. D., Zhang, H., & Gee, J. C. (2006). 3D mouse brain reconstruction from histology using a coarse-to-fine approach. In *International Workshop on Biomedical Image Registration* (pp. 230–237). Springer.