Synth-by-Reg (SbR): Contrastive learning for synthesis-based registration of paired images

Adrià Casamitjana[1], **Matteo Mancini**[2,3,4], **Juan Eugenio Iglesias**[1,5,6]

[1]Center for Medical Image Computing, University College London, UK

[2]Department of Neuroscience, University of Sussex, Brighton, UK

[3]NeuroPoly Lab, Polytechnique Montreal, Canada.

[4]CUBRIC, Cardiff University, UK

[5]Martinos Center for Biomedical Imaging, MGH and Harvard Medical School, USA

[6]Computer Science and AI Laboratory, Massachusetts Institute of Technology, USA

Abstract

Nonlinear inter-modality registration is often challenging due to the lack of objective functions that are good proxies for alignment. Here we propose a synthesis-by-registration method to convert this problem into an easier intra-modality task. We introduce a registration loss for weakly supervised image translation between domains that does not require perfectly aligned training data. This loss capitalises on a registration U-Net with frozen weights, to drive a synthesis CNN towards the desired translation. We complement this loss with a structure preserving constraint based on contrastive learning, which prevents blurring and content shifts due to overfitting. We apply this method to the registration of histological sections to MRI slices, a key step in 3D histology reconstruction. Results on two public datasets show improvements over registration based on mutual information (13% reduction in landmark error) and synthesis-based algorithms such as CycleGAN (11% reduction), and are comparable to registration with label supervision. Code and data are publicly available at https://github.com/acasamitjana/SynthByReg.

Keywords

Image synthesis; Inter-modality registration; Deformable registration; Contrastive estimation

1 Introduction

Image registration is a crucial step to spatially relate information from different medical images. Unpaired registration aligns images of different subjects into a common space to perform subsequent analysis (e.g., population studies [13], voxel-based morphometry [4], or multi-atlas segmentation [30,18]). On the other hand, paired registration aligns different images from the same anatomy and finds application in image guided intervention (e.g.,

a.casamitjana@ucl.ac.uk .

MR-CT in the prostate [15]); patient follow-up (e.g., pre- and post-operative scans [20]); or longitudinal [29] and multimodal studies (e.g., 3D histology reconstruction with MRI [27]).

Registration is often cast as an optimisation problem where a source image is deformed towards a target image such that it maximises a similarity metric of choice. Classical registration methods solve this problem independently for every pair of images with standard iterative optimisers [32]. Modern learning approaches predict a deformation directly from a pair of images using a convolutional neural network (CNN). Supervised learning methods use ground truth deformation fields in training, either synthetic [31] or derived from manual segmentations [7]. These have been superseded by unsupervised methods, in which CNNs are trained to optimise metrics like those used in classical registration, e.g., sum of squared differences (SSD) or local normalised cross-correlation (LNCC) [6,34], without wasting capacity in regions without salient features.

Widespread similarity functions like SSD or LNCC are well suited for intra-modality registration problems. However, the difficulty of designing accurate similarity functions across modalities hampers inter-modality registration. Mutual information (MI) is often used [21] but with unsatisfactory results in the nonlinear case, due to the excessive flexibility of the model [17]. Other metrics used in inter-modality registration are the Modality-Independent Neighbourhood Descriptor (MIND, [14], based on local patch similarities) or adversarial losses measuring whether two images are well aligned or not [12]. MIND is sensitive to initial alignment, bias field or rotations depending on the neighbourhood size, while adversarial losses are prone to missing local correspondences.

An alternative to inter-modality registrationis to convert the problem into an intra-modality task using a registration-by-synthesis framework: image-to-image (I2I) translation is first used to synthesise new source images with the target contrast, and then intra-modality registration (which is more accurate) is performed in the target domain. With accurate image synthesis, the errors introduced by the translation are outweighed by the improvement in registration [17]. In unsupervised synthesis, cycle-consistent generative adversarial networks (CycleGAN) can be used [33,36], but they lack structural consistency across views and may generate artefacts due to overfitting (e.g., flip contrast or even deform images). To mitigate this issue, additional losses between the original and synthetic images have been proposed, e.g., segmentation losses [16] or inter-modality similarities between the original and synthetic scans (e.g., MIND [37] or MI [35]).

Beyond CycleGAN, other approaches have attempted to enforce geometry consistency between the original and synthetic images via specific architectures or training schemes. An I2I translation model that explicitly learns to disentangle domain-invariant (i.e., content) from domain specific features (i.e., appearance) was proposed in [28]; the latent content features can then be used to train a registration network. More recently, a novel training scheme that forces the translation and registration steps to be commutative (thus discouraging deformation at synthesis) has been presented [2]. Nonetheless, GAN-based approaches are challenging to train, with well-known problems (e.g., vanishing gradients, instability [3]) and an increasing number of losses and hyperparameters. In this work, we turn the registration-by-synthesis framework around into a synthesis-by-registration (SbR)

approach, where a registration network trained on the target domain (and frozen weights) is used in the loss for training an I2I network. This allows us to greatly simplify the objective function and avoid potentially unstable adversarial training. Moreover, we use contrastive learning at the patch level to ensure geometric consistency. The SbR model outputs both the translated image and the deformation field. The contribution of this work is threefold: *(i)* we develop a novel registration loss for paired I2I translation; *(ii)* we adapt the contrastive PatchNCE loss [26] for image registration as a geometry-preserving constraint; and *(iii)* we combine *(i)* and *(ii)* into an unsupervised SbR framework for inter-modality registration that does not require multiple encoders / decoders and therefore has low GPU memory requirements.

2 Methods

2.1 Overview

Let us consider two misaligned 2D images of the same anatomy (e.g., a histological section and a corresponding MRI plane): the source $S(x)$ and target $T(x)$; x represents spatial location. We further assume the availability of an intra-modality registration CNN $\mathscr{R}$ with weights $\theta_\mathscr{R}$, which predicts a deformation field Φ from two images of the target modality: $\Phi = \mathscr{R}(T, T'; \theta_\mathscr{R})$, such that $T(x) \approx T'(\Phi(x))$. We also define an I2I translation CNN $\mathscr{G}$ (with weights $\theta_\mathscr{G}$) from contrast S to T that regresses the image intensities: $S^T = \mathscr{G}(S; \theta_\mathscr{G})$, such that S^T resembles the anatomy in S, had it been acquired with modality T. The crucial observation is that, if S^T is well synthesised, and $\Phi = \mathscr{R}(T, S^T; \theta_\mathscr{R})$, then $T(x) \approx S^T(\Phi(x))$. Specifically, we propose the following loss (Figure 1):

$$\mathscr{L}(\theta_\mathscr{G}) = \mathscr{L}_{\text{reg}}(S, T; \theta_\mathscr{G}, \theta_\mathscr{R}) + \lambda_{\text{geo}}\mathscr{L}_{\text{geo}}(S, T; \theta_\mathscr{G}, \theta_\mathscr{R}), \tag{1}$$

where $\mathscr{L}_{\text{reg}}$ is a "registration loss" measuring the similarity of T and (the deformed) S^T, $\mathscr{L}_{\text{geo}}$ is a geometric consistency loss that ensures that the contents of S and S^T are aligned, and λ_{geo} is a relative weight. A key implicit assumption of this framework is that, because the images are paired, there exists a spatial transform Φ that aligns S and T well, such that the synthesis does not need to shift or blur boundaries to minimise the error; the geometric consistency loss further discourages such mistakes. Crucially, the loss in Eq. 1 does not depend on $\theta_\mathscr{R}$: the registration CNN is trained on the target domain and its weights are frozen, such that gradients will backpropagate through these layers to improve the synthesis. This asymmetric scheme enables us to avoid using a distribution matching loss (e.g., CycleGAN) that may produce hallucination artefacts [9].

2.2 Intra-modality registration network

One of the key points in SbR is the differentiable registration method used to train the image synthesis model. We use a U-Net [8] model (as in [10]) that learns a diffeomorphic mapping between images from the same modality. The model is trained on pairs of images from the target domain and outputs a stationary velocity field (SVF), ψ, at half the input resolution. Then, a scaling and squaring approach is used to integrate ψ into a half-resolution deformation field, which is linearly upsampled to obtain the final deformation

$\Phi(\boldsymbol{x})$. Training uses LNCC as image similarity term, and the norm of the gradient of the SVF as regulariser:

$$\mathscr{L}_{\mathscr{R}}(T, T'; \boldsymbol{\theta}_{\mathscr{R}}) = \frac{1}{|\Omega|} \sum_{\boldsymbol{x} \in \Omega} LNCC[T(\boldsymbol{x}), T'(\Phi(\boldsymbol{x}; \boldsymbol{\theta}_{\mathscr{R}}))] + 2\lambda_{\mathscr{R}} \|\nabla \boldsymbol{\psi}(\boldsymbol{x}; \boldsymbol{\theta}_{\mathscr{R}})\|_2, \quad (2)$$

where Ω is the discrete image domain, and $\lambda_{\mathscr{R}}$ is a relative weight. Since the goal is to learn registration of images with approximately the same anatomy, we train the CNN with pairs of images that are similar to each other – specifically, within 3 neighbours in the image stack. In order to prevent overfitting, which may be problematic due to the relatively limited number of combinations of pairs, we use random spatial transformations for data augmentation at each iteration in the source and target images, including small random similarity transforms and smooth nonlinear deformations. Once this CNN has been trained, its weights $\boldsymbol{\theta}_{\mathscr{R}}$ are frozen during training of the rest of layers in our framework.

2.3   Image-to-Image translation using a registration loss

The modality translation is performed by a generator network, $\mathscr{G}$, with a similar architecture to [26] and trained using a combination of two losses: $\mathscr{L}_{\text{reg}}$ and $\mathscr{L}_{\text{geo}}$. The first component $\mathscr{L}_{\text{reg}}$ is the registration loss between the target and the translated, deformed source. In section 2.2 above, we used the LNCC metric, which is known to work well in learning-based, intra-modality registration registration of most modalities, and can handle bias field in MRI [6]. However, in I2I we need to explicitly penalise absolute intensity differences, since encouraging local correlation is not enough to optimise the synthesis. For this purpose, we use the ℓ_1-norm, which has been widely used in the synthesis literature, and which is more robust than ℓ_2 against violations of the assumption that the anatomy is perfectly paired in the source and target images. The registration loss is:

$$\mathscr{L}_{\text{reg}}(S, T; \boldsymbol{\theta}_{\mathscr{G}}, \boldsymbol{\theta}_{\mathscr{R}}) = \frac{1}{|\Omega|} \sum_{\boldsymbol{x} \in \Omega} \left\| T(\boldsymbol{x}) - S^T(\Phi(\boldsymbol{x}; \boldsymbol{\theta}_{\mathscr{R}}); \boldsymbol{\theta}_{\mathscr{G}}) \right\|_1. \quad (3)$$

The second component of the loss $\mathscr{L}_{\text{geo}}$ seeks to enforce geometric consistency in the synthesis and is based on noise contrastive estimation (PatchNCE [26]). The idea behind PatchNCE is to maximise a lower bound on the MI between the pre- and post-synthesis images at the patch level. For this purpose, we define a "query" image q (e.g., S^T) and a "reference" image r (e.g., S), from which we extract patch descriptors from the stack of features computed by the encoding branch of the I2I CNN, $\mathscr{G}$. These descriptors are the output of L layers of interest, including: the input image, the downsampling convolutional layers and the first and last ResNet blocks. Specifically, we extract sets of features f_l at the layers of interest $l = 1, \ldots, L$ and N random locations $\boldsymbol{x}_{l,n}$ per layer, i.e., $\{f_l(\boldsymbol{x}_{l,n})\}_{l=1, \ldots, L; n=1, \ldots, N}$ (in practice, a tissue mask is used when drawing $\boldsymbol{x}_{l,n}$ in order not to sample the background). Each of these f_l encodes different image features (with different number of channels), from different neighbourhoods (patches), and at different resolution levels.

Given these descriptors, the contrastive loss builds on the principle that $f_l^q(x_{l,n})$ (for the query) and $f_l^r(x_{l,n'})$ (for the reference) should be similar for $n = n'$ and dissimilar for n

n'. Rather than using the descriptors f directly, we follow in [26] and run them through two-layer perceptrons (which are different for the descriptors in every layer l, since they have different resolutions), followed by unit-norm normalisation layers. This yields a new representation $\{z_{l,n}\}_{l=1,\ldots,L; n=1,\ldots,N}$, with:

$$z_{l,n} = \mathcal{Q}_l(f_l(x_{l,n}); \theta_z), \tag{4}$$

where θ_z groups the parameters of these representation layers. Given z, the contrastive PatchNCE loss is given by a softmax function of cosine similarities:

$$\mathscr{L}_{\text{PatchNCE}}(q, r; \theta_z, \tau) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{l=1}^{L} \log\left(\frac{\exp(z_{l,n}^q \cdot z_{l,n}^r / \tau)}{\sum_{n'=1}^{N} \exp(z_{l,n}^q \cdot z_{l,n'}^r / \tau)}\right), \tag{5}$$

where τ is a temperature parameter and $(\cdot)$ is the dot product. It can be shown that the lower bound on the MI becomes tighter with increasing N[25].

In practice, we use two PatchNCE losses: one between the source and translated images; and another between the registered and target images:

$$\mathscr{L}_{\text{geo}}(S, T; \theta_{\mathcal{G}}, \theta_{\mathcal{R}}, \theta_z, \tau) = \mathscr{L}_{\text{PatchNCE}}\left(S^T(x; \theta_{\mathcal{G}}), S(x); \theta_z, \tau\right)$$
$$+ \mathscr{L}_{\text{PatchNCE}}\left(S^T(\Phi(x; \theta_{\mathcal{R}}); \theta_{\mathcal{G}}), T(x); \theta_z, \tau\right). \tag{6}$$

Combining the registration and geometric consistency losses in Equations 3 and 7 yields the final loss for our meta-architecture:

$$\mathscr{L}(\theta_{\mathcal{G}}, \theta_z) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \left\| T(x) - S^T(\Phi(x; \theta_{\mathcal{R}}); \theta_{\mathcal{G}}) \right\|_1$$
$$+ \lambda_{\text{geo}} \mathscr{L}_{\text{PatchNCE}}\left(S^T(x; \theta_{\mathcal{G}}), S(x); \theta_z, \tau\right)$$
$$+ \lambda_{\text{geo}} \mathscr{L}_{\text{PatchNCE}}\left(S^T(\Phi(x; \theta_{\mathcal{R}}); \theta_{\mathcal{G}}), T(x); \theta_z, \tau\right), \tag{7}$$

which we optimise with respect to $\theta_{\mathcal{G}}$ and θ_z – since τ is a fixed hyperparameter and $\theta_{\mathcal{R}}$ is frozen, as explained above.

3 Experiments and results

3.1 Data

We validate the presented methodology in the context of 3D histology reconstruction via registration to a reference MRI volume. We use two publicly available datasets with histological sections and an *ex vivo* 3D MRI of the same subject. A 3D similarity transform between the stack of histological sections and the MRI volume was used to align images from both domains [24]. The MRI volume was then resampled into the space of histological

stack, which yields a set of paired images to register: histological sections and corresponding MRI resampled planes. The two datasets are:

- **Allen Human Brain Atlas** [11]: this dataset includes 93 sections with manual delineations of hundreds of brain structures, which we grouped into four coarse tissue classes: cerebral white matter (WM), cerebral grey matter (GM), cerebellar white matter (WMc), and cerebellar grey matter (GMc). An *ex vivo* MRI is available, which was segmented into the same four tissue classes with SPM [5]. In addition, J.E.I. manually annotated 13.8 ± 4.4 pairs of matching landmarks in the histological sections and corresponding resampled MRI planes, uniformly distributed across all spatial locations.

- **BigBrain Initiative** [1]: we considered one every 20 sections, i.e., one section every 0.4 mm (344 sections in total). As in the previous dataset, an *ex vivo* MRI is available and J.E.I. manually annotated 11.6 ± 1.7 landmark pairs in the histological sections and corresponding MRI planes. No segmentations are available for this dataset.

3.2 Experimental setup

In our experiments, we register each histological section to the corresponding (resampled) MRI slice. For quantitative evaluation, we report the average root-mean-squared landmark error (both datasets) and the Dice score on brain tissue classes (only for the Allen dataset).

Our proposed method, **SbR**, was trained with the following hyper-parameters: $\lambda_{\mathrm{geo}} = 0.02$, $\tau = 0.05$ and $\lambda_{\mathscr{R}} = 1$, which were set from a subset of the Allen dataset and used elsewhere. We also tested three other configurations of our method: an ablated version without the structure preserving constraint, i.e., $\lambda_{\mathrm{geo}} = 0$ (**SbR-N**); fine-tuning the result of **SbR** by unfreezing the registration parameters (**SbR-R**); and an extension (**SbR-G**) that includes an LSGAN loss [22] with a PatchGAN discriminator [19] to discriminate between synthesised and target images (S^T and T). **SbR-G** enables us to assess the potential benefits of adding a distribution matching loss in training.

In addition, we compare our method against a number of other methods, to test differences against: standard registration metrics, other synthesis-based approaches without specific geometric constraints, and supervision with labels and Dice scores. Specifically, the competing methods are: (*i*) **Linear**, the initial affine registration with NiftyReg [24]; (*ii*) **NMI**, unsupervised training using normalised mutual information (NMI) with 20 bins on the image intensities; (*iii*) **NMIw**, weakly supervised training using NMI and an additional Dice loss [23] on the segmentations; (*iv*) **cGAN**, a CycleGAN [38] approach combined with our registration loss; and (*v*) **RoT**, the state-of-the-art method presented in [2] that consists of alternating the registration and translation steps. All learning-based methods above (including ours) use the same architecture for registration, and also the same nonlinear spatial augmentation scheme (sampling $9 \times 9 \times 2$ from zero-mean Gaussians and upsampling to full resolution)

3.3   Results

The quantitative results are summarised in Figure 2. The landmark errors show that our proposed method (SbR) outperforms all baseline approaches: 11%, 9 % and 7% error reduction with respect cGAN, RoT and NMIw in the Allen dataset and 23% and 33 % with respect cGAN and RoT in the BigBrain dataset; all improvements are statistically significant ($p < 0.001$) using a Wilcoxon signed-rank test. Interestingly, SbR is able to align tissue masks as well as NMIw, even though segmentations were not used in the training phase. The naive approach (SbR-N) suffers from synthetic artefacts in the generator, which degrades the results - thus highlighting the importance of including structure preserving constraints in the model. The other two extensions of the model, SbR-G and SbR-R, achieve similar performance to the initial configuration, without yielding any statistically significant additional benefits.

In Figure 3, we show an example of the synthesised and registered images using SbR for each dataset. The method displays robustness against common artefacts, such as: cracks, missing tissue and inhomogeneous staining (in histology), or intensity inhomogeneity (in MRI). Our method is able to accurately register convoluted structures such as the cortex, as seen in Figure 4

4   Discussion and conclusion

We have presented Synth-by-Reg, a synthesis-by-registration framework for inter-modality registration, which we have validated on a histology-to-MRI registration task. The method uses a single I2I translation network trained with a robust registration loss (based on the ℓ_1-norm) and a geometric consistency term (based on contrastive learning). In histology-MRI registration, Synth-by-Reg enables us to avoid using a CycleGAN approach, which often falters in presence of histological artefacts – since it needs to learn to simulate them and subsequently recover from them. Future work will focus on adapting our method to the unpaired scenario, as well as to other imaging modalities. We believe that synthesis-by-registration can be a very useful alternative in difficult inter-modality registration problems when weakly paired data are available, e.g., MRI and histology.

References

1. Amunts K, Lepage C, Borgeat L, Mohlberg H, Dickscheid T, Rousseau MÉ, Bludau S, Bazin PL, Lewis LB, et al. : BigBrain: an ultrahigh-resolution 3D human brain model. Science 340(6139), 1472–1475 (2013) [PubMed: 23788795]

2. Arar M, Ginger Y, Danon D, Bermano AH, Cohen-Or D: Unsupervised multi-modal image registration via geometry preserving image-to-image translation. In: CVPR. pp. 13410–13419. IEEE (2020)

3. Arjovsky M, Bottou L: Towards principled methods for training generative adversarial networks. arXiv preprint arXiv:1701.04862 (2017)

4. Ashburner J, Friston K: Voxel-based morphometry-the methods. Neuroimage 11(6), 805–821 (2000) [PubMed: 10860804]

5. Ashburner J, Friston K: Unified segmentation. Neuroimage 26, 839–851 (2005) [PubMed: 15955494]

6. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV: Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging 38(8), 1788–1800 (2019)

7. Cao X, Yang J, Zhang J, Nie D, Kim M, Wang Q, Shen D: Deformable image registration based on similarity-steered CNN regression. In: MICCAI. pp. 300–308. Springer (2017)

8. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: MICCAI. pp. 424–432. Springer (2016)

9. Cohen JP, Luck M, Honari S: Distribution matching losses can hallucinate features in medical image translation. In: MICCAI. pp. 529–536. Springer (2018)

10. Dalca AV, Balakrishnan G, Guttag J, Sabuncu MR: Unsupervised learning for fast probabilistic diffeomorphic registration. In: International Conference on MICCAI. pp. 729–738. Springer (2018)

11. Ding SL, Royall JJ, Sunkin SM, Ng L, Facer BA, Lesnar P, Guillozet-Bongaarts A, McMurray B, et al. : Comprehensive cellular-resolution atlas of the adult human brain. Journal of Comparative Neurology 524(16), 3127–3481 (2016)

12. Fan J, Cao X, Wang Q, Yap PT, Shen D: Adversarial learning for mono-or multi-modal registration. Medical image analysis 58, 101545 (2019) [PubMed: 31557633]

13. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL: Unbiased average age-appropriate atlases for pediatric studies. NeuroImage 54(1), 313–327 (2011) [PubMed: 20656036]

14. Heinrich MP, Jenkinson M, Bhushan M, Matin T, Gleeson FV, Brady M, Schnabel JA: MIND: Modality independent neighbourhood descriptor for multimodal deformable registration. Medical image analysis 16(7), 1423–1435 (2012) [PubMed: 22722056]

15. Hu Y, Ahmed HU, Taylor Z, Allen C, Emberton M, Hawkes D, Barratt D: MR to ultrasound registration for image-guided prostate interventions. Medical image analysis 16(3), 687–703 (2012) [PubMed: 21216180]

16. Huo Y, Xu Z, Bao S, Assad A, Abramson RG, Landman BA: Adversarial synthesis learning enables segmentation without target modality ground truth. In: ISBI. pp. 1217–1220. IEEE (2018)

17. Iglesias JE, Konukoglu E, Zikic D, Glocker B, Van Leemput K, Fischl B: Is synthesizing mri contrast useful for inter-modality analysis? In: MICCAI. pp. 631–638. Springer (2013)

18. Iglesias JE, Sabuncu MR: Multi-atlas segmentation of biomedical images: a survey. Medical image analysis 24(1), 205–219 (2015) [PubMed: 26201875]

19. Isola P, Zhu JY, Zhou T, Efros AA: Image-to-image translation with conditional adversarial networks. In: CVPR. pp. 1125–1134. IEEE (2017)

20. Kwon D, Niethammer M, Akbari H, Bilello M, Davatzikos C, Pohl KM: PORTR: Pre-operative and post-recurrence brain tumor registration. IEEE transactions on medical imaging 33(3), 651–667 (2013)

21. Maes F, Vandermeulen D, Suetens P: Medical image registration using mutual information. Proceedings of the IEEE 91(10), 1699–1722 (2003)

22. Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S: Least squares generative adversarial networks. In: CVPR. pp. 2794–2802. IEEE (2017)

23. Milletari F, Navab N, Ahmadi SA: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3DV Conf. pp. 565–571 (2016)

24. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, Fox NC, Ourselin S: Fast free-form deformation using graphics processing units. Computer methods and programs in biomedicine 98(3), 278–284 (2010) [PubMed: 19818524]

25. Oord A.v.d., Li Y, Vinyals O: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)

26. Park T, Efros AA, Zhang R, Zhu JY: Contrastive learning for unpaired image-to-image translation. In: ECCV. pp. 319–345. Springer (2020)

27. Pichat J, Iglesias JE, Yousry T, Ourselin S, Modat M: A survey of methods for 3D histology reconstruction. Medical image analysis 46, 73–105 (2018) [PubMed: 29502034]

28. Qin C, Shi B, Liao R, Mansi T, Rueckert D, Kamen A: Unsupervised deformable registration for multi-modal images via disentangled representations. In: IPMI. pp. 249–261. Springer (2019)

29. Reuter M, Schmansky NJ, Rosas HD, Fischl B: Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402–18 (2012) [PubMed: 22430496]

30. Rohlfing T, Brandt R, Menzel R, Maurer CR Jr: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 21(4), 1428–1442 (2004) [PubMed: 15050568]

31. Sokooti H, De Vos B, Berendsen F, Lelieveldt BP, Išgum I, Staring M: Nonrigid image registration using multi-scale 3D convolutional neural networks. In: MICCAI. pp. 232–239. Springer (2017)

32. Sotiras A, Davatzikos C, Paragios N: Deformable medical image registration: A survey. IEEE transactions on medical imaging 32(7), 1153–1190 (2013) [PubMed: 23739795]

33. Tanner C, Ozdemir F, Profanter R, Vishnevsky V, Konukoglu E, Goksel O: Generative adversarial networks for MR-CT deformable image registration. arXiv preprint arXiv:1807.07349 (2018)

34. de Vos BD, Berendsen FF, Viergever MA, Staring M, Išgum I: End-to-end unsupervised deformable image registration with a convolutional neural network. In: International Workshop DLMIA, pp. 204–212. Springer (2017)

35. Wang C, Yang G, Papanastasiou G, Tsaftaris SA, Newby DE, Gray C, et al. : DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis. Information Fusion 67, 147–160 (2021) [PubMed: 33658909]

36. Wei D, Ahmad S, Huo J, Peng W, Ge Y, Xue Z, Yap PT, Li W, Shen D, Wang Q: Synthesis and inpainting-based MR-CT registration for image-guided thermal ablation of liver tumors. In: MICCAI. pp. 512–520. Springer (2019)

37. Xu Z, Luo J, Yan J, Pulya R, Li X, Wells W, Jagadeesan J: Adversarial uni-and multi-modal stream networks for multimodal image registration. In: MICCAI. pp. 222–232. Springer (2020)

38. Zhu JY, Park T, Isola P, Efros AA: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: CVPR. pp. 2223–2232. IEEE (2017)
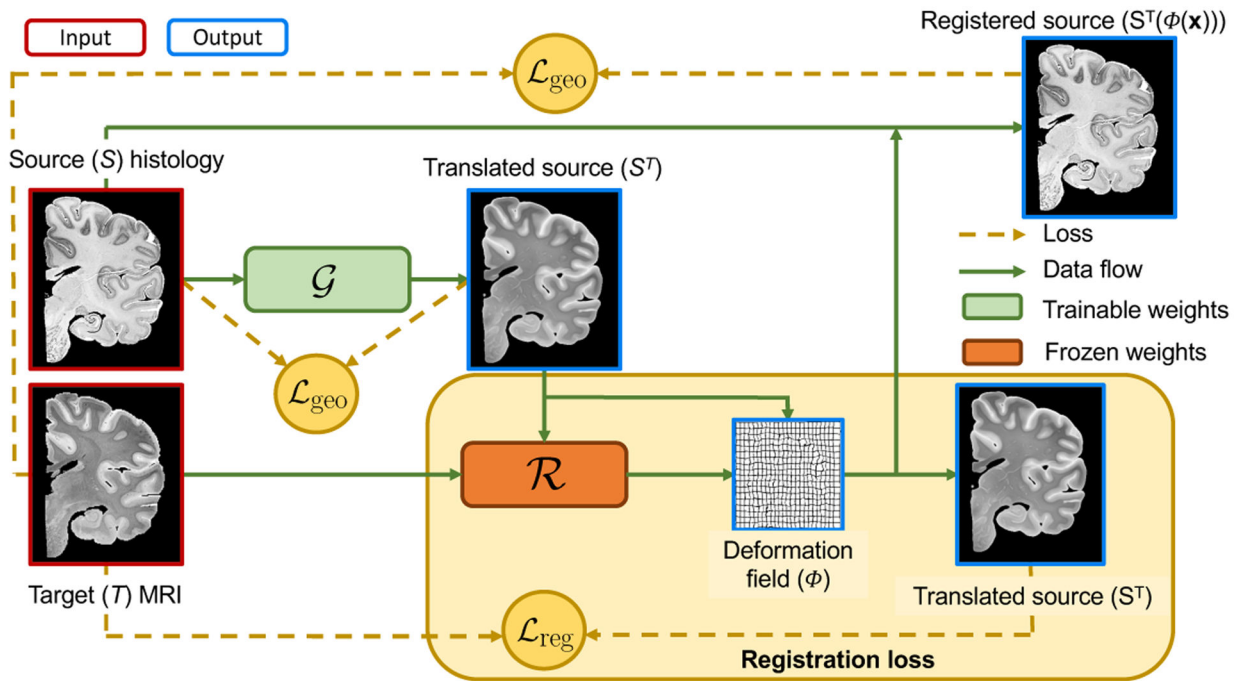
Fig. 1.
Overview of proposed pipeline, using histology and MRI as source and target contrasts, respectively.
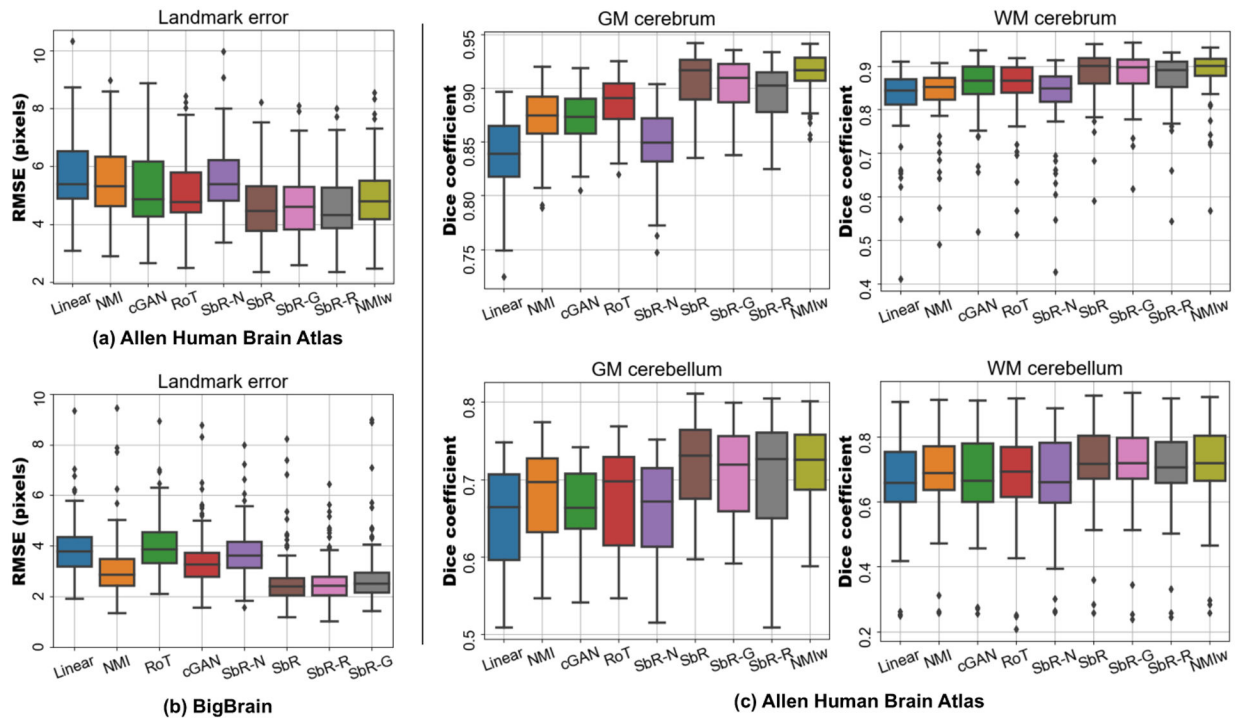
Fig. 2.
Landmark mean squared error on the Allen human brain atlas dataset (a) and the BigBrain dataset (b). Dice score coefficient for the Allen dataset is shown in (c).
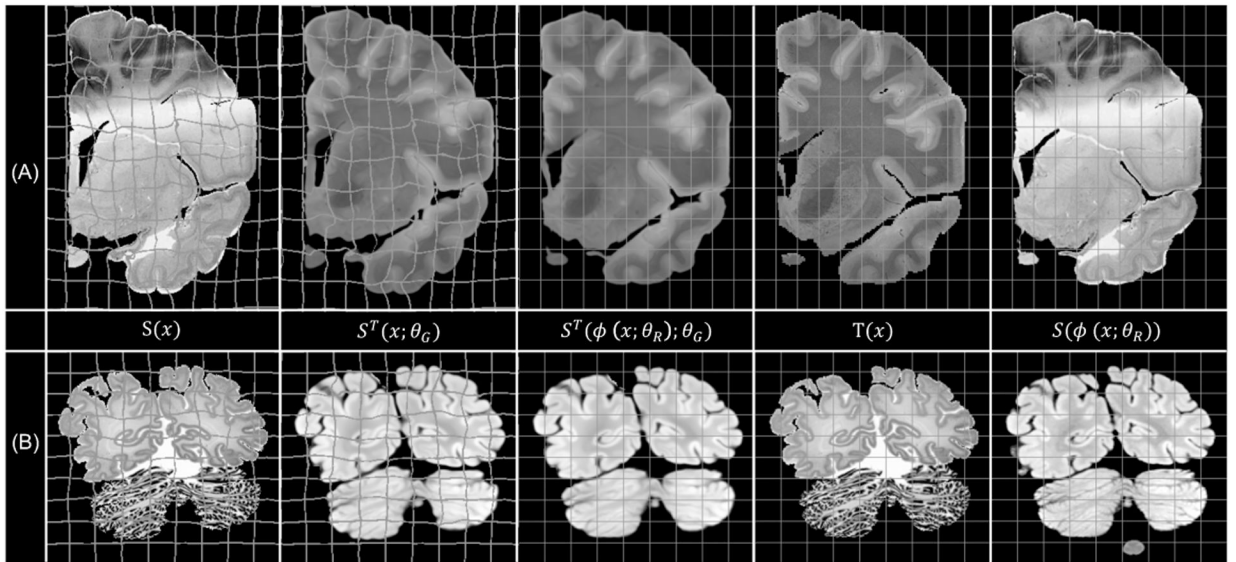
Fig. 3.
Image examples from (a) the Allen human brain atlas, and (b) the BigBrain project, with the deformed and rectangular grid overlaid on the source and target spaces, respectively.
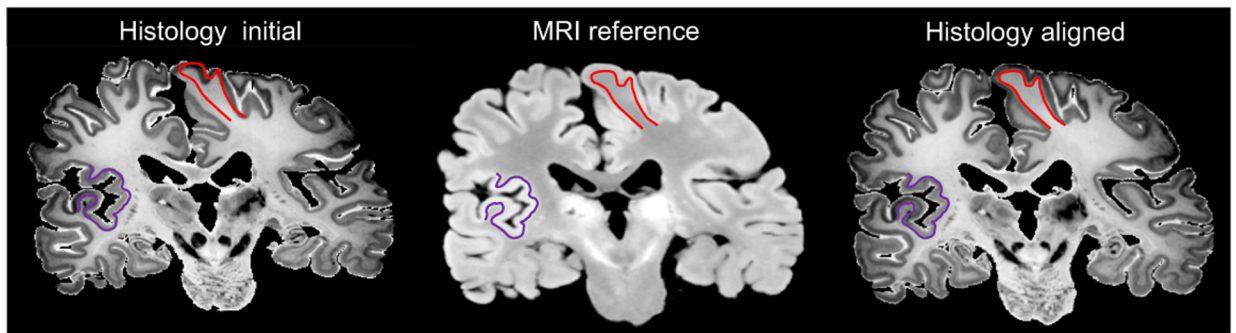
Fig. 4.
Section 170 from BigBrain, with cortical boundaries manually traced on the target domain (MRI) and overlaid on the histology, before and after registration.