

Towards Saner Deep Image Registration

Bin Duan Ming Zhong Yan Yan*

Illinois Institute of Technology

bduan2@hawk.iit.edu, {mzhong3,yyan34}@iit.edu

Abstract

With recent advances in computing hardware and surges of deep-learning architectures, learning-based deep image registration methods have surpassed their traditional counterparts, in terms of metric performance and inference time. However, these methods focus on improving performance measurements such as Dice, resulting in less attention given to model behaviors that are equally desirable for registrations, especially for medical imaging. This paper investigates these behaviors for popular learning-based deep registrations under a sanity-checking microscope. We find that most existing registrations suffer from low inverse consistency and nondiscrimination of identical pairs due to overly optimized image similarities. To rectify these behaviors, we propose a novel regularization-based sanity-enforcer method that imposes two sanity checks on the deep model to reduce its inverse consistency errors and increase its discriminative power simultaneously¹. Moreover, we derive a set of theoretical guarantees for our sanity-checked image registration method, with experimental results supporting our theoretical findings and their effectiveness in increasing the sanity of models without sacrificing any performance.

1. Introduction

Learning maps between images or spaces, i.e. registration, is an important task, and has been widely studied in various fields, such as computer vision [15, 33], medical imaging [20, 52], and brain mapping [35, 47]. With recent advances in modern computing hardware and deep-learning techniques, learning-based deep image registration methods have surpassed their traditional counterparts, both in terms of metric performance and inference time. Different from the traditional style of optimizing on single image pair [12, 32, 7, 11, 2, 28, 19] using diffeomorphic formulations, such as elastic [4, 38], fluid mechanics [7, 18, 43] or B-spline [36], existing deep regis-

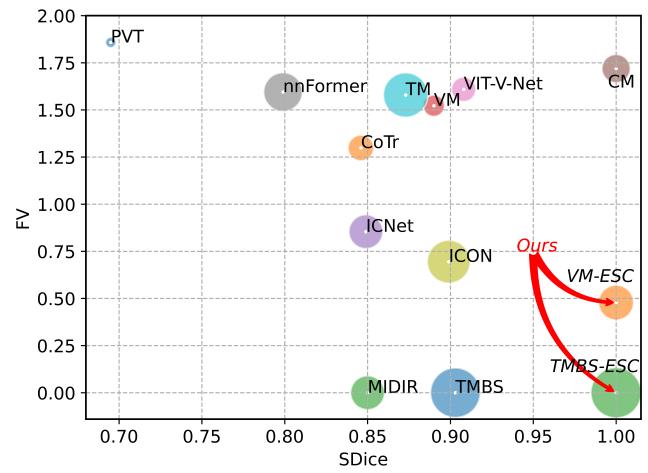


Figure 1. FV-SDice-Dice comparisons of deep registrations on IXI Brain dataset. The vertical axis is FV (% of folded voxels), the horizontal axis is SDice (Self-Dice), and the circle size is Dice. Both sanity-checked models (VM-ESC and TMBS-ESC) achieve better diffeomorphism, competitive registration performance, and significantly improved self-sanity, compared to other models, including models with inverse consistency (ICNet [50], ICON [16]).

tions [42, 6, 13, 29, 23, 10, 41, 25, 9] focus on maximizing image similarities between transformed moving images and fixed images. Despite the effectiveness of this approach, it inevitably leads to over-optimization of image similarities and thus introduces non-smooth mappings [6, 50, 16, 9], where smooth transformation maps are typically desirable, especially in the medical imaging domain.

To tackle the over-optimized issue, popular remedies [4, 36, 44, 37, 49, 45, 39, 40] utilize add-ons such as Large Deformations Diffeomorphic Metric Mapping (LDDMM) [7], vector Stationary Velocity Field (vSVF) [39], B-spline [36], Elastic [4] or Demons [43] to enforce diffeomorphism, requiring costly and iterative numerical optimizations [8, 49, 39]. Other methods [13, 17] seek probabilistic formulation for the registration but can lead to inferior performance [9]. Nonetheless, these methods operate only on one mapping in the direction from moving to fixed images, yet disregarding the relationship between different mappings from both directions, as shown in Appendix Fig. A1.

*Corresponding author

¹Our code and models are available at <https://github.com/tuffr5/Saner-deep-registration>

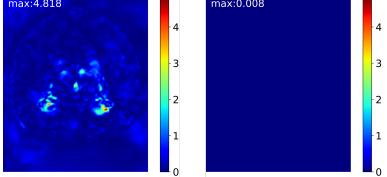


Figure 2. Self-sanity error maps comparison. *Left:* with no self-sanity check, *Right:* with self-sanity check. We unify the error maps’ scale bars for a fair comparison.

ical structures, which has not been studied in other related works. Experimentally, our relaxation shows better performance over various metrics. Besides, to the best of our knowledge, our work is the first to consider the self-sanity error directly on displacements rather than for image similarities as in [23] for medical image registration studies.

3. Methodology

3.1. Self-sanity and Cross-sanity Checks

To increase the discriminative power of identical pairs feeding into the model, we propose our self-sanity check as

$$g^{a \rightarrow a} = 0, \forall a \in \{m, f\}. \quad (2)$$

where the mapping function g learned using any models is restricted to output zero displacements for identical pairs. Such identical pairs can be filtered out using similarity measurements. However, users are unlikely to perform these filters, especially when they do not know that trained models would produce terrible predictions for identical pairs. Hence our self-sanity check is a natural remedy.

Next, we enforce the inverse consistency on different mappings for g by the model. We will search for correspondence in the fixed image for every point in the moving image such that the transformations between the two images are inconsistent. For example, suppose that a point in the space of image a is registered to the space of image b . If we register this point from image b back to image a , the point will arrive at the same location in image a (Def. 1). We first define the backward displacement map as in optical flow studies [22, 27, 24], back-projected from $g^{b \rightarrow a}$

$$\tilde{g}^{b \rightarrow a}(p) = g^{b \rightarrow a}(p + g^{a \rightarrow b}(p)), \quad (3)$$

making it convenient for calculation. We then introduce our cross-sanity check in the form of

$$|g^{a \rightarrow b} + \tilde{g}^{b \rightarrow a}|^2 < \alpha(|g^{a \rightarrow b}|^2 + |\tilde{g}^{b \rightarrow a}|^2) + \beta, \quad (4)$$

$\forall (a, b) \in \{(m, f), (f, m)\}$. Here, we allow the estimation errors to increase linearly with the displacement magnitude with slope α and intercept β . Instead of imposing zero-tolerance between forward and back-projected backward displacements [16, 31], we relax the inverse consistency with error tolerance, defined by α and β , to allow

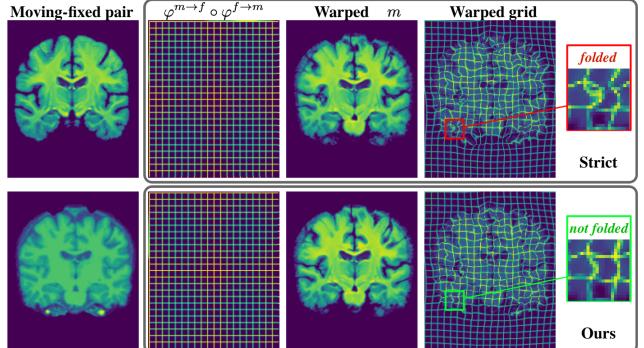


Figure 3. Comparisons between strict inverse consistency trained results (*Top*) and cross-sanity checked results (*Bottom*). Our relaxed sanity-checked result maintains a similar level of inverse consistency as $\varphi^{m \rightarrow f} \circ \varphi^{f \rightarrow m}$ is close to id transformation (second column). We can also observe that ours produces a more regular map, compared to the folded map from the model trained with strict inverse consistency. Best view zoomed.

occlusions which is more practical. I.e., This sanity check states that every point p in the moving image a should be able to map back from the fixed image b to its original place in image a with certain error tolerance. We then prove that this error tolerance is upper bounded.

Theorem 1 (Relaxed registration via cross-sanity check). *An ideal symmetric registration meets $\varphi^{a \rightarrow b} \circ \varphi^{b \rightarrow a} = id$, defined in Def. 1. Then, a cross-sanity checked registration is a relaxed solution to this ideal registration, satisfying*

$$\|g^{a \rightarrow b} + \tilde{g}^{b \rightarrow a}\|_2^2 < \frac{\beta(2 - \alpha)N}{1 - \alpha}. \quad (5)$$

Here, $0 < \alpha < 1$ and $\beta > 0$. N is a constant, representing the total pixel/voxel numbers. Theoretically, our proposed cross-sanity checks can be viewed as a relaxed version of the strict symmetric constraint, which is commonly used. We also derive the lower/upper bound for satisfying our forward/backward consistency check as shown in Eq. (5). The derivations’ details are shown in the Appx. A.2. To sum up, our cross-sanity check allows a series of solutions to the forward/backward consistency in an enclosed set with a radius of $\sqrt{\frac{\beta(2-\alpha)N}{1-\alpha}}$.

3.2. Unique Minimizer for Single Image Pair

Next, we show that there exists a unique solution for our sanity-checked optimization, in terms of a single image pair. We start with writing the standard optimization with our proposed sanity checks in the norm form as

$$\begin{aligned} \min & -\text{Sim}(f, m \circ (g^{m \rightarrow f} + p)) + \lambda_r \|\text{Reg}(g^{m \rightarrow f})\|_2^2, \\ \text{s.t. } & \|g^{a \rightarrow b} + \tilde{g}^{b \rightarrow a}\|_2^2 < \alpha(\|g^{a \rightarrow b}\|_2^2 + \|\tilde{g}^{b \rightarrow a}\|_2^2) + \beta N, \\ & \|g^{a \rightarrow a}\|_2^2 = 0, \quad \forall (a, b) \in \{(m, f), (f, m)\}. \end{aligned} \quad (6)$$

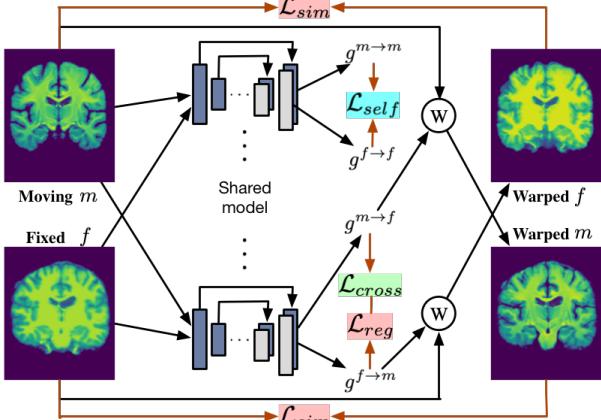


Figure 4. Training a sanity-checked model. \circledast denotes spatial warping, e.g., warped m is that we warp moving image m using the transformation map calculated from $g^{m \rightarrow f}$.

Interpretation of derived upper bound. By Lemma 4, we prove that if we average on the total number of voxels and also two directions, the similarity distance between the optimal minimizer and our sanity-checked minimizer per pixel/voxel is upper bounded by $\lambda_c(1 - \alpha)\beta$. That being said, to satisfy our sanity checks, and thus maintain the loyalty to the optimal minimizer, $\lambda_c(1 - \alpha)\beta$ should be small. Numerically speaking, for example, if $\alpha = 0.1$ and $\beta = 10$, we have $(1 - \alpha)\beta = 9$. This distance of 9 is extremely large for this delicate image registration task, causing the constrained minimizer to be untrustworthy. Therefore, we need to adopt a relatively small loss weight λ_c , e.g. $\lambda_c = 0.001$, to bound the distance between two minimizers tightly. This observation from our proven theoretical upper bound also coincides with our sanity loss weight ablation study.

3.4. Sanity-checked Registration Training

We show our sanity-checked training pipeline in Fig. 4. We introduce each loss in the subsection. The self-sanity check can be formulated into a loss function form as

$$\mathcal{L}_{self} = \frac{1}{2}(\|g^{m \rightarrow m}\|_2^2 + \|g^{f \rightarrow f}\|_2^2). \quad (16)$$

So that the self-sanity loss penalizes the squared differences between predicted displacement maps and the ideal ones. Next, we use $m \rightarrow f$ direction as an example ($f \rightarrow m$ direction follows the same principle) to formulate the proposed cross-sanity check loss. We calculate for every voxel and define a binary mask $\mathcal{M}^{m \rightarrow f}$ in the form of

$$\mathcal{M}^{m \rightarrow f} = \begin{cases} 0 & \text{if satisfies the cross-sanity check,} \\ 1 & \text{otherwise.} \end{cases} \quad (17)$$

An interpretation of this binary mask $\mathcal{M}^{m \rightarrow f}$ is that it records violations of the cross-sanity check (Eq. (4)) for

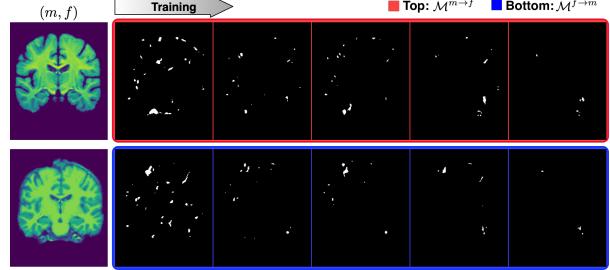


Figure 5. Mask evolution during training. Overall, as training proceeds, violators of the cross-sanity check are decreasing.

each individual point. In this way, < 0 solution in Eq. (10) will not challenge the theoretical formulations since these points are masked out. Thus, we can formulate the proposed cross-sanity check in the form of a loss function of

$$\begin{aligned} \mathcal{L}_{cross}^{m \rightarrow f} = & ||\mathcal{M}^{m \rightarrow f} \odot (g^{m \rightarrow f} + \tilde{g}^{f \rightarrow m})||_2^2 - \beta ||\mathcal{M}^{m \rightarrow f}||_2^2 \\ & - \alpha(||\mathcal{M}^{m \rightarrow f} \odot g^{m \rightarrow f}||_2^2 + ||\mathcal{M}^{m \rightarrow f} \odot \tilde{g}^{f \rightarrow m}||_2^2). \end{aligned} \quad (18)$$

The final $\mathcal{L}_{cross} = \frac{1}{2}(\mathcal{L}_{cross}^{m \rightarrow f} + \mathcal{L}_{cross}^{f \rightarrow m})$. Here, \odot denotes element-wise multiplication so that we only retain points violating the cross-sanity check. In this case, the loss value is only calculated for those violators, visualized as occlusion masks, shown in Fig. 5. Finally, the total loss is

$$\mathcal{L}_{total} = \mathcal{L}_{sim} + \lambda_r \mathcal{L}_{reg} + \lambda_s \mathcal{L}_{self} + \lambda_c \mathcal{L}_{cross}. \quad (19)$$

Here, \mathcal{L}_{sim} as NCC loss, \mathcal{L}_{reg} as $||\nabla(u)||_2^2$, and $\lambda_r = 1$, following standard deep registrations [6, 9]. If not specified otherwise, $\lambda_s = 0.1$ and $\lambda_c = 0.001$. While training, the model optimizes the total loss \mathcal{L}_{total} on different moving-fixed image pairs (m, f) in the training set \mathcal{D} as

$$\min_{(m, f) \in \mathcal{D}} \min_{g \in H} \mathcal{L}_{total}. \quad (20)$$

In this way, we fulfill our novel regularization-based sanity-enforcer formulation for training a registration model.

4. Experiments

Evaluation metrics. We study model behaviors in a wide range of metrics. For the main metric, we use dice to measure how fit is the transformed segmentation to its ground truth, following previous studies. To study the model taking an identical pair as inputs, we also report self dice (SDice), i.e., the dice when registering the moving image to itself. For pre-operative and post-recurrence registration, we measure the mean target registration error (TRE) of the paired landmarks with Euclidean distance in millimeters and also self mean registration error (STRE) to study the self-sanity of the models. Besides, we report 95% Hausdorff Distance (HD95) as in [9], which measures the 95th percentile of the distances between boundary points

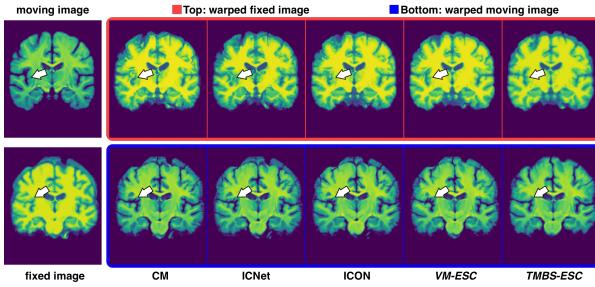


Figure 6. Comparisons between different models on IXI dataset.

of the transformed subject and the actual subject. We follow [3, 31] to report robustness (ROB) for a pair of scans as the relative number of successfully registered landmarks.

For diffeomorphism measurements, we report the percentage of folded voxels (FV) whose Jacobian determinant < 0 , the Absolute Value of Negative Jacobian (AJ) where we sum up all negative Jacobian determinants, and the Standard Deviation of the logarithm of the Jacobian determinant (SDlogJ). All these three Jacobian determinant-related metrics reflect how regular are the transformation maps.

For quantifying sanity errors, we present the mean of self-sanity error (SSE) and the mean of cross-sanity error (CSE), defined in Eq. (16) and Eq. (18), respectively. These two metrics are designed to study model behaviors per the level of each single image pair and are essential for our sanity analysis of different learning-based deep models.

Implemented models. We denote the bidirectional optimization as Enclosed (E) Image registration, operating to maximize the similarity score and minimize the spatial gradient regularization. We also have Self-sanity checked (S) and Cross-sanity checked (C) image registrations. *Moreover, as proof of concept that our sanity checks can be applied to different learning-based models*, we implement our proposed techniques on models such as VoxelMorph [6] (VM), TransMorph-Large [9] (TM), TransMorph-B-spline [9] (TMBS), and DIRAC [31].

4.1. Results

Atlas-to-subject registration. We split in total 576 T1-weighted brain MRI images from the Information eXtraction from Images (IXI) [1] database into 403, 58, and 115 volumes for training, validation, and test sets. The moving image is an atlas brain MRI obtained from [23]. FreeSurfer [14] was used to pre-process the MRI volumes. The pre-processed image volumes are all cropped to the size of $160 \times 192 \times 224$. Label maps, including 30 anatomical structures, are obtained using FreeSurfer for evaluating registration. Besides CNN-based models [6, 23, 10], we also include other transformer-based deep methods [10, 51, 48, 46] as baselines following [9]. The comparison results are presented in Tab. 2, and qualitative comparisons in Fig. 6. For cross-sanity check, we set $\alpha = 0.1$ and

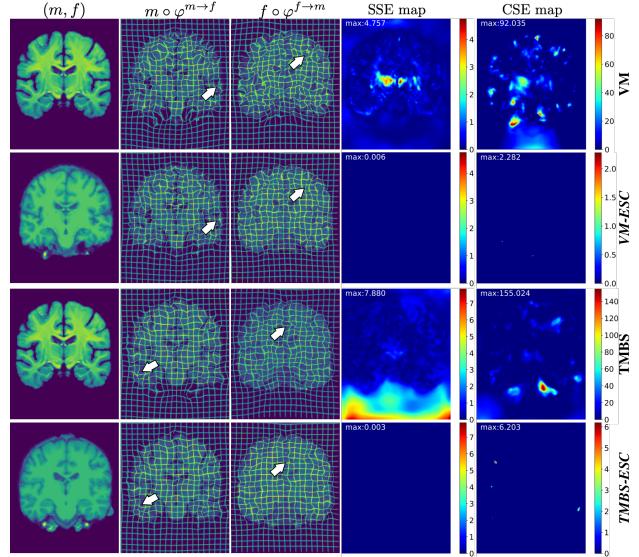


Figure 7. Qualitative comparisons on IXI, where we mark maximum values of error maps on each top left. Best view zoomed.

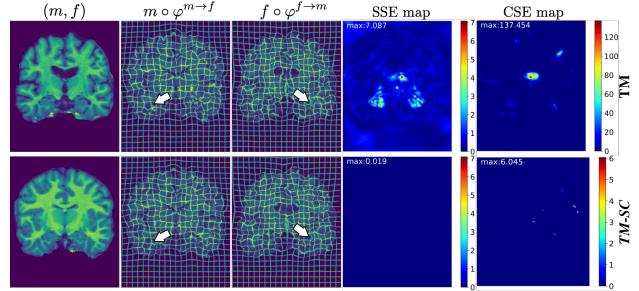


Figure 8. Comparisons on OASIS validation set. Our method produces more regular maps for all input image pairs.

$\beta = 12$. Overall, we can observe models with sanity checks achieve better diffeomorphisms without impairing any Dice performance and even improve the performance for the VM model. We show qualitative comparisons in Fig. 7. Compared to the naive counterparts, our sanity-checked model produces more regular maps from both directions and reduces huge sanity errors, comparably using error heatmaps for both self-sanity and cross-sanity errors.

Subject-to-subject registration. OASIS dataset [26, 21] contains a total of 451 brain T1 weighted MRI images, with 394/19/38 images used for training/validation/testing purposes, respectively. The pre-processed image volumes are all cropped to the size of $160 \times 192 \times 224$. Label maps for 35 anatomical structures are provided using FreeSurfer [14] for evaluation. For the cross-sanity check, we set $\alpha = 0.1$ and $\beta = 10$. The results are shown in Tab. 3. In terms of main metrics, we achieve on-par performance with all the state-of-the-art methods, further certifying our sanity checks will not contaminate the performance of models. We also show our qualitative comparison in Fig. 8, indicating that our sanity-enforced image registration results in

Acknowledgements

Authors thank anonymous reviewers for their constructive comments and Yuzhang Shang for the fruitful discussions. Authors would also want to thank Junyu Chen, Tony C. W. Mok for their generous help with their papers and codes, and LEARN2REG challenge organizer for evaluating our models. This work was supported by NSF SCH-2123521. This article solely reflects the opinions and conclusions of its authors and not the funding agencies.

References

- [1] Ixi brain, 2022. <https://brain-development.org/ixi-dataset>. 2, 6
- [2] Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *MIA*, 12(1):26–41, 2008. 1, 7
- [3] Bhakti Baheti, Diana Waldmannstetter, Satrajit Chakrabarty, Hamed Akbari, Michel Bilello, Benedikt Wiestler, Julian Schwarting, Evan Calabrese, Jeffrey Rudie, Syed Abidi, et al. The brain tumor sequence registration challenge: Establishing correspondence between pre-operative and follow-up mri scans of diffuse glioma patients. *arXiv:2112.06979*, 2021. 2, 6, 7
- [4] Ruzena Bajcsy and Stane Kovacić. Multiresolution elastic matching. *CVGIP*, 46(1):1–21, 1989. 1
- [5] Spyridon Bakas. Brain tumor sequence registration challenge, 2022. <https://www.med.upenn.edu/cbica/brats-reg-challenge>. 7
- [6] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *TMI*, 38(8):1788–1800, 2019. 1, 2, 5, 6, 7, 8
- [7] M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV*, 61(2):139–157, 2005. 1, 7
- [8] Xiaohuan Cao, Jianhua Yang, Jun Zhang, Qian Wang, Pew-Thian Yap, and Dinggang Shen. Deformable image registration using a cue-aware deep regression network. *TBE*, 65(9):1900–1911, 2018. 1
- [9] Junyu Chen, Eric C Frey, Yufan He, William P Segars, Ye Li, and Yong Du. Transmorph: Transformer for unsupervised medical image registration. *MIA*, page 102615, 2022. 1, 5, 6, 7, 8
- [10] Junyu Chen, Yufan He, Eric C Frey, Ye Li, and Yong Du. Vit-v-net: Vision transformer for unsupervised volumetric medical image registration. *arXiv:2104.06468*, 2021. 1, 6, 7, 8
- [11] Gary E Christensen, Xiujuan Geng, Jon G Kuhl, Joel Bruss, Thomas J Grabowski, Imran A Pirwani, Michael W Vannier, John S Allen, and Hanna Damasio. Introduction to the non-rigid image registration evaluation project (nirep). In *WBIR*, 2006. 1, 8
- [12] Gary E Christensen, Sarang C Joshi, and Michael I Miller. Volumetric transformation of brain anatomy. *TMI*, 16(6):864–877, 1997. 1
- [13] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *MIA*, 57:226–236, 2019. 1, 2, 7, 8
- [14] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012. 6
- [15] Kexue Fu, Shaolei Liu, Xiaoyuan Luo, and Manning Wang. Robust point cloud registration framework based on deep graph matching. In *CVPR*, 2021. 1
- [16] Hastings Greer, Roland Kwitt, Francois-Xavier Vialard, and Marc Niethammer. Icon: Learning regular maps through inverse consistency. In *ICCV*, 2021. 1, 2, 3, 7, 8
- [17] Daniel Grzech, Mohammad Farid Azampour, Ben Glocker, Julia Schnabel, Nassir Navab, Bernhard Kainz, and Loïc Le Folgoc. A variational bayesian method for similarity learning in non-rigid image registration. In *CVPR*, 2022. 1
- [18] Gabriel L Hart, Christopher Zach, and Marc Niethammer. An optimal control approach for deformable registration. In *CVPR*, 2009. 1
- [19] Mattias P Heinrich, Oskar Maier, and Heinz Handels. Multi-modal multi-atlas segmentation using discrete optimisation and self-similarities. *VISCERAL Challenge@ ISBI*, 1390:27, 2015. 1, 7
- [20] Alessa Hering, Lasse Hansen, Tony C. W. Mok, Albert CS Chung, Hanna Siebert, Stephanie Häger, Annkristin Lange, Sven Kuckertz, Stefan Heldmann, Wei Shao, et al. Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *TMI*, 2022. 1, 7
- [21] Andrew Hoopes, Malte Hoffmann, Douglas N Greve, Bruce Fischl, John Guttag, and Adrian V Dalca. Learning the effect of registration hyperparameters with hypermorph. *MELBA*, 3:1–30, 2022. 6
- [22] Junhwa Hur and Stefan Roth. Mirrorflow: Exploiting symmetries in joint optical flow and occlusion estimation. In *ICCV*, 2017. 3
- [23] Boah Kim, Dong Hwan Kim, Seong Ho Park, Jieun Kim, June-Goo Lee, and Jong Chul Ye. Cyclemorph: cycle consistent unsupervised deformable image registration. *MIA*, 71:102036, 2021. 1, 3, 6, 7, 8
- [24] Pengpeng Liu, Michael Lyu, Irwin King, and Jia Xu. Selflow: Self-supervised learning of optical flow. In *CVPR*, 2019. 3
- [25] Jinxin Lv, Zhiwei Wang, Hongkuan Shi, Haobo Zhang, Sheng Wang, Yilang Wang, and Qiang Li. Joint progressive and coarse-to-fine registration of brain mri via deformation field integration and non-rigid feature fusion. *TMI*, 41(10):2788–2802, 2022. 1, 7
- [26] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498–1507, 2007. 2, 6

- [27] Simon Meister, Junhwa Hur, and Stefan Roth. Unflow: Unsupervised learning of optical flow with a bidirectional census loss. In *AAAI*, 2018. 3
- [28] Marc Modat, Gerard R Ridgway, Zeike A Taylor, Manja Lehmann, Josephine Barnes, David J Hawkes, Nick C Fox, and Sébastien Ourselin. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, 98(3):278–284, 2010. 1, 7
- [29] Tony C. W. Mok and Albert Chung. Conditional deformable image registration with convolutional neural network. In *MICCAI*, 2021. 1, 7
- [30] Tony C. W. Mok and Albert Chung. Robust image registration with absent correspondences in pre-operative and follow-up brain mri scans of diffuse glioma patients. *arXiv:2210.11045*, 2022. 2, 7
- [31] Tony C. W. Mok and Albert Chung. Unsupervised deformable image registration with absent correspondences in pre-operative and post-recurrence brain tumor mri scans. In *MICCAI*, 2022. 2, 3, 6, 7, 1
- [32] Xavier Pennec, Pascal Cachier, and Nicholas Ayache. Understanding the “demon’s algorithm”: 3d non-rigid registration by gradient descent. In *MICCAI*, 1999. 1
- [33] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *CVPR*, 2022. 1
- [34] Huaiqi Qiu, Chen Qin, Andreas Schuh, Kerstin Hammernik, and Daniel Rueckert. Learning diffeomorphic and modality-invariant registration using b-splines. In *MIDL*, 2021. 7, 8
- [35] Lei Qu, Yuanyuan Li, Peng Xie, Lijuan Liu, Yimin Wang, Jun Wu, Yu Liu, Tao Wang, Longfei Li, Kaixuan Guo, et al. Cross-modal coherent registration of whole mouse brains. *Nature Methods*, 19(1):111–118, 2022. 1
- [36] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *TMI*, 18(8):712–721, 1999. 1
- [37] Tanya Schmah, Laurent Risser, and François-Xavier Vialard. Left-invariant metrics for diffeomorphic image registration with spatially-varying regularisation. In *MICCAI*, 2013. 1
- [38] Dinggang Shen and Christos Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration. *TMI*, 21(11):1421–1439, 2002. 1
- [39] Zhengyang Shen, Xu Han, Zhenlin Xu, and Marc Niethammer. Networks for joint affine and non-parametric image registration. In *CVPR*, 2019. 1
- [40] Zhengyang Shen, François-Xavier Vialard, and Marc Niethammer. Region-specific diffeomorphic metric mapping. *NeurIPS*, 2019. 1
- [41] Hanna Siebert, Lasse Hansen, and Matthias P Heinrich. Fast 3d registration with accurate optimisation and little learning for learn2reg 2021. In *MICCAI*, 2021. 1, 2, 7
- [42] Hessam Sokoooti, Bob De Vos, Floris Berendsen, Boudeijn PF Lelieveldt, Ivana Isgum, and Marius Staring. Nonrigid image registration using multi-scale 3d convolutional neural networks. In *MICCAI*, 2017. 1
- [43] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2009. 1
- [44] François-Xavier Vialard, Laurent Risser, Daniel Rueckert, and Colin J Cotter. Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. *IJCV*, 97:229–241, 2012. 1
- [45] Jian Wang and Miaomiao Zhang. Deepflash: An efficient network for learning-based medical image registration. In *CVPR*, 2020. 1
- [46] Wenhui Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *ICCV*, 2021. 6, 7, 8
- [47] Xuechun Wang, Weilin Zeng, Xiaodan Yang, Yongsheng Zhang, Chunyu Fang, Shaoqun Zeng, Yunyun Han, and Peng Fei. Bi-channel image registration and deep-learning segmentation (birds) for efficient, versatile 3d mapping of mouse brain. *Elife*, 10:63455, 2021. 1
- [48] Yutong Xie, Jianpeng Zhang, Chunhua Shen, and Yong Xia. Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. In *MICCAI*, 2021. 6, 7, 8
- [49] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration—a deep learning approach. *NeuroImage*, 158:378–396, 2017. 1
- [50] Jun Zhang. Inverse-consistent deep networks for unsupervised deformable image registration. *arXiv:1809.03443*, 2018. 1, 2, 7, 8, 3
- [51] Hong-Yu Zhou, Jiansen Guo, Yinghao Zhang, Lequan Yu, Liansheng Wang, and Yizhou Yu. nnformer: Interleaved transformer for volumetric segmentation. *arXiv:2109.03201*, 2021. 6, 7, 8
- [52] S Kevin Zhou, Hoang Ngan Le, Khoa Luu, Hien V Nguyen, and Nicholas Ayache. Deep reinforcement learning in medical imaging: A literature review. *MIA*, 73:102193, 2021. 1

Method	TRE \downarrow	STRE \downarrow	ROB \uparrow	FV \downarrow	AJ $\times 10^2 \downarrow$
DIRAC	2.760 \pm 0.247	0.274 \pm 0.027	0.776 \pm 0.055	0.025 \pm 0.009	4.242 \pm 2.954
<i>DIRAC-C</i>	2.721 \pm 0.262	0.268 \pm 0.039	0.791 \pm 0.044	0.022 \pm 0.008	3.012 \pm 1.442
<i>DIRAC-SC</i>	2.719 \pm 0.259	0.218 \pm 0.046	0.795 \pm 0.034	0.022 \pm 0.005	3.001 \pm 1.314

Table A3. Performance of replacing the inverse consistent error.

A.9. Error Map Comparisons between Inverse Consistent Methods

Qualitative comparisons between inverse consistent methods on IXI dataset are shown in Fig. A3.

A.10. Performance of Replacing DIRAC’s Inverse Error

We denote it as *DIRAC-C*, and report in Tab. A3.

A.11. Role of Image Similarity Loss

The image similarity loss still plays a very important role during training. The reason is that $\mathcal{L}_{\text{self}}$ and $\mathcal{L}_{\text{cross}}$ are defined on displacements, to calculate such losses, we need to ensure that those displacements are meaningful, which is guaranteed via \mathcal{L}_{sim} . Compared to the value of NCC (<1), the cross-sanity error is relatively large (Tab. 2), and using small λ_c will not interfere with the optimizations.