Towards Saner Deep Image Registration

Bin Duan   Ming Zhong   Yan Yan*

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

Illinois Tech, Chicago, IL 60616, USA

https://github.com/tuffr5/Saner-deep-registration

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 simultaneously. 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 registrations [42, 6, 13, 29, 23, 10, 41, 25, 9] focus on maximizing
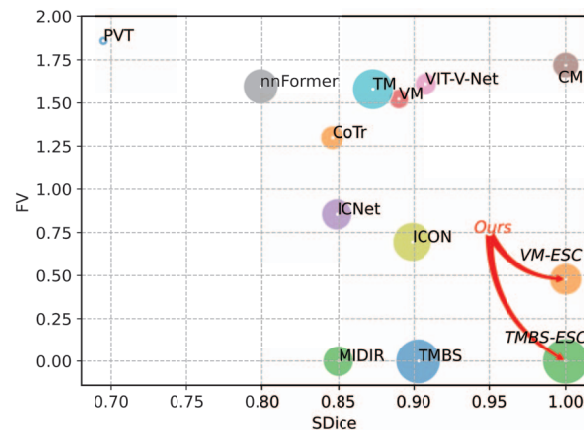
Figure 1. FV-SDice-Dice comparisons of deep registrations on IXI Brain dataset. The vertical axis is FV (% of folded voxels), the horizontal axis is SDice (Self-Dice), and the circle size is Dice. Both sanity-checked models (*VM-ESC* and *TMBS-ESC*) achieve better diffeomorphism, competitive registration performance, and significantly improved self-sanity, compared to other models, including models with inverse consistency (ICNet [50], ICON [16]).

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[1] Fig. A1.

---

*Corresponding author

[1]Appendix goes https://arxiv.org/pdf/2307.09696.pdf

Recent studies [50, 16, 31] have made great progress in modeling the relationship, i.e. inverse consistency, for different mappings. [50, 16] explicitly enforce the relationship in a strict form for ideal inverse consistency. However, for applications such as brain tumor registration [3], where there are regions with no valid correspondences, it is impractical to apply in such a strict manner. To select valid correspondences, [31] utilize mean errors of similar anatomical locations as thresholds. However, it is always tricky to determine similar anatomical locations, especially for unsupervised registrations in medical imaging [6, 13, 41]. Different from previous methods [50, 16, 31], we introduce two straightforward yet effective constraints, namely, self-sanity check to reduce error while registering identical pairs and cross-sanity check to ensure inverse consistency. Using our sanity checks, we test a wide range of registration models and find that despite better performance such as Dice, most models are suffering from over-optimization of image similarities, leading to folded transformations and low sanity awareness, as shown in Fig. 1. Moreover, our sanity checks can be applied to different registration models, where the extensive experiments certify that our method improves their sanity awareness, without sacrificing any performance.

Our findings are five-fold: (1) We find that despite better performance such as Dice, most models produce non-smooth transformations and are less aware of sanity errors due to over-optimization on image similarities; (2) We propose two novel sanity checks for the registration model training and derive corresponding theoretical foundations; (3) Our sanity checks not only help reduce the sanity errors of existing models but also assist them to produce more regular maps without diffeomorphic add-ons; (4) Our proposed sanity-checks are model-agnostic. It can be deployed to various models and is only needed during training so that it does genereate no side effects for inference; (5) Experiments on IXI [1], OASIS [26], and BraTSReg [3] datasets verify our findings and show on par or better performance.

2. Background and Related Work

Background. Consider a set of images defined on the domain $\Omega \subseteq \mathbb{R}^n$ (n-D spatial domain) of some Hilbert space H. Let f be the fixed, and m be the moving $\in \Omega$. Let $u : \Omega \to \mathbb{R}^n$ be a displacement vector field that maps from $H \times H \to H^\star$ (dual space of H) such that $\varphi = u + p$, where p is the image grid of coordinates. φ denotes the transformation from $m \to f$. The goal of registration is to compute such transformation φ, whose quality can be measured by some similarity functions in the form of $\text{Sim}(f, m \circ \varphi)$. $m \circ \varphi$ denotes that m is warped by φ. Without loss of generality, $\text{Sim}(\cdot)$ can be replaced by any differentiable similarity function, such as Normalized Cross Correlation (NCC) and Mutual Information (MI). Alternatively, we can use other

Method	Relaxation/Approximation	Error Bound
ICNet [50]	$\|g^{m \to f} + \tilde{g}^{f \to m}\|_F^2$	-
ICON [16]	$\|\varphi^{m \to f} \circ \varphi^{f \to m} - id\|_2^2 +$ $\epsilon^2 \|d\varphi^{m \to f} \sqrt{\text{Jac}(\varphi^{\text{m} \to \text{f}})}\|_2^2$	-
DIRAC [31]	$\|g^{m \to f} + \tilde{g}^{f \to m}\|_2 <$ $\frac{1}{N}\Sigma_N\|g^{m \to f} + \tilde{g}^{f \to m}\|_2 + \beta$	-
Ours	$\|g^{m \to f} + \tilde{g}^{f \to m}\|^2 <$ $\alpha(\|g^{m \to f}\|^2 + \|\tilde{g}^{f \to m}\|^2) + \beta$	$\lambda_c(1-\alpha)\beta N$

Table 1. Comparison between different inverse consistent registration methods. Due to the space limit, the formulations are shown only for $m \to f$, where $f \to m$ should follow the same way.

distance functions such as sum squared error (SSD) by replacing $-\text{Sim}(\cdot)$ with $\text{Dist}(\cdot)$. Hereafter, we will stick to the similarity notation for later derivations.

Let g be the mapping function, parameterized using the model, where $u = g^{m \to f} := g(m, f)$ stands for the displacement map from $m \to f$. In terms of such g learned by the model, we assume that the similarity operator is concave, making $-\text{Sim}$ convex when it is not a distance operator for a (m, f) pair. Therefore, the optimization in standard learning-based registrations can be formulated as

$$\min -\text{Sim}(f, m \circ (g^{m \to f} + p)) + \lambda_r \|\text{Reg}(g^{m \to f})\|_2^2, \quad (1)$$

trying to find such g for a (m, f) pair in the mapping search space. Here, $\text{Reg}(\cdot)$ term is added to smooth the transformation, penalizing sudden changes in the map. Most commonly, $\text{Reg}(\cdot)$ can be in the form of $\nabla(u)$ (first-order spatial derivative) or $\nabla^2(u)$ (second-order spatial derivative), where L^2 norm of the image gradient is generally adopted in medical registration, resulting in a H^1 regularization.

Definition 1 (Ideal/Strict inverse consistency). Given two different mappings: $g^{m \to f}$ and $g^{f \to m}$, if $\varphi^{m \to f} \circ \varphi^{f \to m} = id$, where id denotes the identity transformation, we called these two mappings are strictly inverse consistent. The strict inverse consistency is equivalently formulated as $g^{m \to f} + \tilde{g}^{f \to m} = 0$, where $\tilde{g}^{f \to m}$ is back-projected from $g^{f \to m}$.

Relation to other inverse consistent methods. We show the relationship between methods in Tab. 1. The formulation of ICNet [50] follows strict inverse consistency in Def. 1 with Frobenius norm. Besides strict inverse consistency, ICON [16] adds regularization using the determinant of the Jacobian matrix of the transformation. Instead of explicitly enforcing strict consistency as in [50, 16], DIRAC [31] and our approach both modulate the inverse problem with more relaxation. This inequality formulation allows us to tell whether a voxel has a valid correspondence, which is practically useful in registering image pairs with topological changes, e.g., brain tumor registration [31, 30]. Different from [31], where means of inverse errors are utilized, we closely associate our formulation with the data, allowing us to show that the errors are upper bounded without resorting to extra information to determine similar anatom-
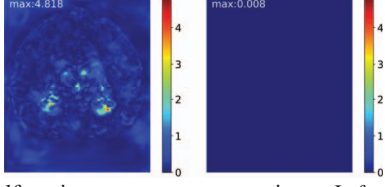
Figure 2. Self-sanity error maps comparison. *Left*: with no self-sanity check, *Right*: with self-sanity check. We unify the error maps' scale bars for a fair comparison.

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 $\varphi^{m\to f} \circ \varphi^{f\to m}$ 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.

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

$$g^{a\to a} = 0, \forall a \in \{m, f\}. \tag{2}$$

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.
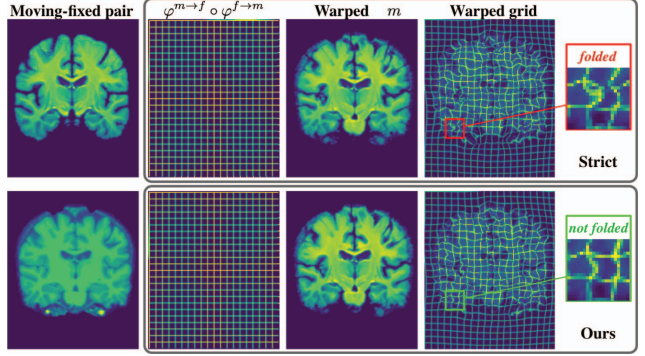
Next, we enforce the inverse consistency on different mappings for g by the model. We will search for correspondence in the fixed image for every point in the moving image such that the transformations between the two images are inconsistent. For example, suppose that a point in the space of image a 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 $g^{b\to a}$

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

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

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

$\forall(a, b) \in \{(m, f), (f, m)\}$. 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

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 $\varphi^{a\to b} \circ \varphi^{b\to a} = id$, defined in Def. 1. Then, a cross-sanity checked registration is a relaxed solution to this ideal registration, satisfying*

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

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

$$\min -\text{Sim}(f, m \circ (g^{m\to f} + p)) + \lambda_r ||\text{Reg}(g^{m\to f})||_2^2,$$

s.t. $||g^{a\to b} + \tilde{g}^{b\to a}||_2^2 < \alpha(||g^{a\to b}||_2^2 + ||\tilde{g}^{b\to a}||_2^2) + \beta N,$

$$||g^{a\to a}||_2^2 = 0, \quad \forall(a, b) \in \{(m, f), (f, m)\}. \tag{6}$$

Theorem 2 (Existence of the unique minimizer for our relaxed optimization). *Let m and f be two images defined on the same spatial domain Ω, which is connected, closed, and bounded in $\mathbb{R}^n$ with a Lipschitz boundary $\partial\Omega$. Let $g : H \times H \to H^\star$ be a displacement mapping from the Hilbert space H to its dual space $H^\star$. Then, there exists a unique minimizer $g^\star$ to the relaxed minimization problem.*

The detailed proof for Thm. 2 is shown in Appx. A.3. In summary, for a moving-fixed pair, since the relaxation by sanity checks and data term (e.g. SSD) are convex, together with the convex search space for g, our problem has a unique minimizer. Thus, the optimization problem is well-conditioned by properties of our regularity constraints such that g will not change dramatically.

3.3. Loyal Sanity-Checked Minimizer

In the next section, we further prove that the distance between the minimizer of the sanity-checked constrained optimization and the optimal minimizer for a single moving-fixed image pair can be represented and controlled by our proposed sanity checks. We first rewrite Eq. (6) as

$$\min - \text{Sim}(b, a \circ (g^{a\to b} + p)) + \lambda_r ||\text{Reg}(g^{a\to b})||_2^2,$$
$$\text{s.t.} \quad ||g^{a\to b} + \tilde{g}^{b\to a}||_2^2 < \alpha(||g^{a\to b}||_2^2 + ||\tilde{g}^{b\to a}||_2^2) + \beta N,$$
$$||g^{a\to a}||_2^2 = 0, \quad \forall (a,b) \in \{(m,f),(f,m)\}.$$
(7)

Remark 2.1. We denote this formulation as the bidirectional optimization since it operates in both directions. It is straightforward to show that this bidirectional optimization still satisfies Thm. 2 so that there exists a unique minimizer.

Then, with a slight abuse of notations, we define that

$$\mathbf{g} = [\underbrace{g^{m\to f}, g^{f\to m}}_{\mathbf{g}_c}, \underbrace{g^{m\to m}, g^{f\to f}}_{\mathbf{g}_s}]^\top. \quad (8)$$

We can rewrite the minimization problem as

$$\min - \text{Sim}(\mathbf{g}) + \lambda_r ||\text{Reg}(\mathbf{g})||_2^2,$$
$$\text{s.t.} \quad ||\mathbf{g}_s||_2^2 = 0,$$
$$||\mathbf{g}_c + \tilde{\mathbf{g}}_c||_2^2 < \alpha(||\mathbf{g}_c||_2^2 + ||\tilde{\mathbf{g}}_c||_2^2) + 2\beta N. \quad (9)$$

Here, $\text{Sim}(\mathbf{g}) := \text{Sim}(m, f \circ (g^{f\to m} + p)) + \text{Sim}(f, m \circ (g^{m\to f} + p))$, $||\text{Reg}(\mathbf{g})||_2^2 := ||\text{Reg}(g^{m\to f})||_2^2 + ||\text{Reg}(g^{f\to m})||_2^2$, $||\mathbf{g}_s||_2^2 := ||g^{m\to m}||_2^2 + ||g^{f\to f}||_2^2$, $||\mathbf{g}_c + \tilde{\mathbf{g}}_c||_2^2 := ||g^{m\to f} + \tilde{g}^{f\to m}||_2^2 + ||g^{f\to m} + \tilde{g}^{m\to f}||_2^2$, and $||\mathbf{g}_c||_2^2 + ||\tilde{\mathbf{g}}_c||_2^2 := ||g^{m\to f}||_2^2 + ||\tilde{g}^{m\to f}||_2^2 + ||g^{f\to m}||_2^2 + ||\tilde{g}^{f\to m}||_2^2$. Then, the optimal minimizer $\mathbf{g}_*$ can be written in the matrix equation as

$$\mathbf{g}_* = \arg\min_{\mathbf{g}\in H} - \text{Sim}(\mathbf{g}), \quad \text{s.t.} \quad \mathbf{A}\mathbf{g}_* = \mathbf{y}.$$

$$\mathbf{A} = \begin{bmatrix} \text{Reg}(\cdot) & 0 & 0 \\ \mathbf{CS}(\cdot) & 0 & 0 \\ 0 & 0 & \text{SS}(\cdot) \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0 \\ 0^\dagger \\ 0 \end{bmatrix}. \quad (10)$$

Here, $\text{Reg}(\cdot) := ||\text{Reg}(\mathbf{g})||_2^2$, $\text{CS}(\cdot) := ||\mathbf{g}_c + \tilde{\mathbf{g}}_c||_2^2 - \alpha(||\mathbf{g}_c||_2^2 + ||\tilde{\mathbf{g}}_c||_2^2) - 2\beta N$, and $\text{SS}(\cdot) := ||\mathbf{g}_s||_2^2$, as defined previously. † here means that we can have < 0 solutions for the CS check as in Eq. (9), which we show will not interfere with the theoretical foundations by masking in the next section. In this way, we obtain our unique minimizer as

$$\mathbf{g}_{\text{sanity}} = \arg\min_{\mathbf{g}\in H} - \text{Sim}(\mathbf{g}) + \frac{\lambda}{2}||\mathbf{A}\mathbf{g} - \mathbf{y}||_2^2, \quad (11)$$

where $\lambda := [2\lambda_r, 2\lambda_c, 2\lambda_s]$ is the vector of parameters.

Theorem 3 (Loyalty of the sanity-checked minimizer). *Let $\mathbf{g}_*$ be the optimal minimizer to the bidirectional registration problem, defined in Eq. (10), and $\mathbf{g}_{\text{sanity}}$ as our sanity-checked minimizer, defined in Eq. (11). The distance between these two minimizers can be upper bounded as*

$$\text{Sim}(\mathbf{g}_{\text{sanity}}) - \text{Sim}(\mathbf{g}_*) \leq \frac{\lambda}{2}||A(\mathbf{g}_{\text{sanity}} - \mathbf{g}_*)||_2^2. \quad (12)$$

The proof is presented in Appx. A.4. We then expand the right-hand side of Eq. (12) as

$$\text{Sim}(\mathbf{g}_{\text{sanity}}) - \text{Sim}(\mathbf{g}_*) \leq \lambda_r ||\text{Reg}(\mathbf{g}_{\text{sanity}})||_2^2$$
$$+ \lambda_s ||\text{SS}(\mathbf{g}_{\text{sanity}})||_2^2) + \lambda_c ||\text{CS}(\mathbf{g}_{\text{sanity}})||_2^2,$$
$$\text{where} \quad \text{Reg}(\mathbf{g}_*) = 0, \text{SS}(\mathbf{g}_*) = 0, \text{CS}(\mathbf{g}_*) = 0. \quad (13)$$

Empirically, the first two terms contribute relatively less ($10\times$ smaller) than the cross-sanity error (See Experiments). Thus, we focus on the CS term here. Following Thm. 1, $||\text{CS}(\mathbf{g}_{\text{sanity}})||_2^2$ is upper bounded in the form of

$$||\text{CS}(\mathbf{g}_{\text{sanity}})||_2^2 < 2(1-\alpha)\beta N. \quad (14)$$

Here, $0 < \alpha < 1$ and $\beta > 0$ ensure the inequality's direction, where the multiplication of *two* accounts for cross-sanity errors from two directions, i.e., from moving image to fixed image and also fixed image to moving image. Thus, t. We put the full derivation of CS error bound in the Appx. A.5. In a nutshell, we prove that by satisfying the cross-sanity check, we have an upper bound for the CS error, constraining the relationship between two displacements from different directions via parameters α and β.

Lemma 4 (Upper-bound of distance between optimal minimizer and sanity-checked minimizer). *Let $\mathbf{g}_*$ and $\mathbf{g}_{\text{sanity}}$ respectively be the optimal minimizer and the constrained minimizer with cross-sanity check, then we have an upper bound for the distance between such two minimizers as*

$$\text{Sim}(\mathbf{g}_{\text{sanity}}) - \text{Sim}(\mathbf{g}_*) < 2\lambda_c(1-\alpha)\beta N. \quad (15)$$

The Lemma 4 follows Thm. 3 and Eq. (14) (upper bound of CS error). That said, the loyalty w.r.t the distance of the constrained minimizer is controlled by the combination of α, β, and the weight parameter λ_c.
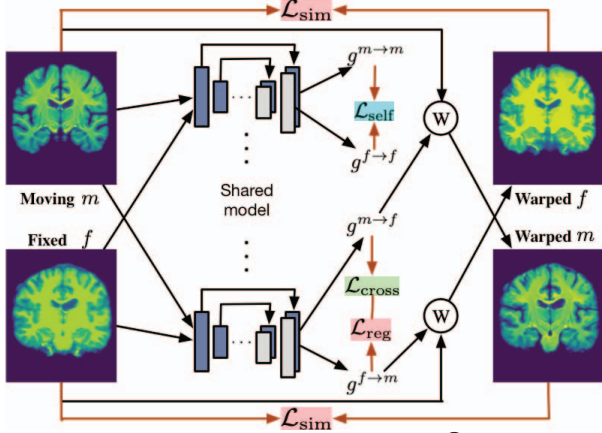
Figure 4. Training a sanity-checked model. $\textcircled{W}$ denotes spatial warping, e.g., warped m is that we warp moving image m using the transformation map calculated from $g^{m \to f}$.

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

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

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

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

An interpretation of this binary mask $\mathcal{M}^{m \to f}$ is that it records violations of the cross-sanity check (Eq. (4)) for
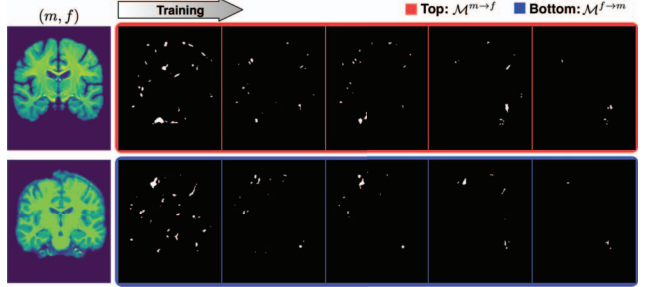
Figure 5. Mask evolution during training. Overall, as training proceeds, violators of the cross-sanity check are decreasing.

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

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

The final $\mathcal{L}_{\text{cross}} = \frac{1}{2}(\mathcal{L}_{\text{cross}}^{m \to f} + \mathcal{L}_{\text{cross}}^{f \to m})$. 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

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

Here, $\mathcal{L}_{\text{sim}}$ as NCC loss, $\mathcal{L}_{\text{reg}}$ 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 $\mathcal{L}_{\text{total}}$ on different moving-fixed image pairs (m, f) in the training set $\mathcal{D}$ as

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

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
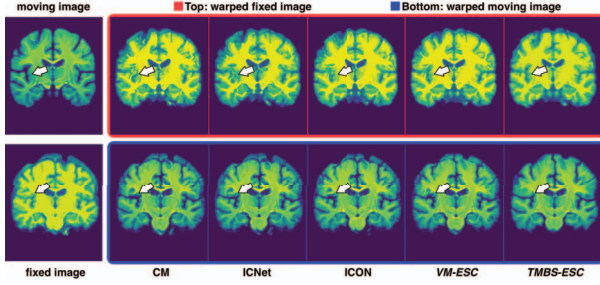
Figure 6. Comparisons between different models on IXI dataset.

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.

95th percentile of the distances between boundary points 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 160×192×224. 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
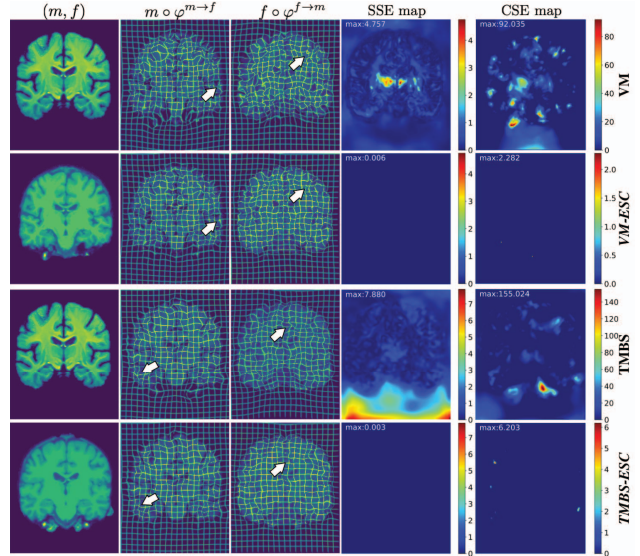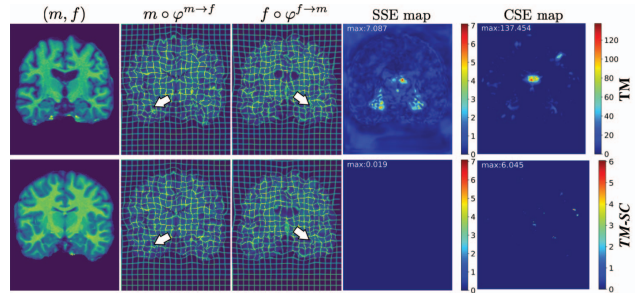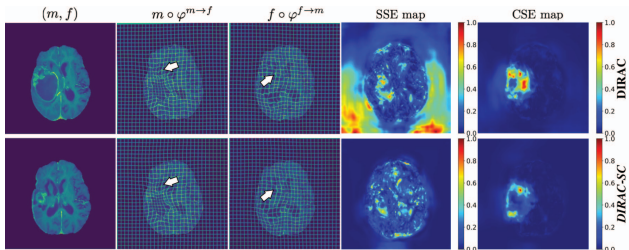
in Fig. 6. For cross-sanity check, we set $\alpha = 0.1$ and $\beta = 12$. Overall, we can observe models with sanity checks achieve better diffeomorphisms without impairing any Dice performance and even improve the performance for the VM model. We show qualitative comparisons in Fig. 7. Compared to 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 160×192×224. 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, indi-

Method	Dice↑	FV↓
Affine	0.386±0.195	-
SyN [2]	0.645±0.152	<0.01
NiftyReg [28]	0.645±0.167	0.020±0.046
LDDMM [7]	0.680±0.135	<0.01
deedsBCV [19]	0.733±0.126	0.147±0.050
MIDIR [34]	0.742±0.128	<0.01
PVT [46]	0.727±0.128	1.858±0.314
CoTr [48]	0.735±0.135	1.298±0.342
VM [6]	0.732±0.123	1.522±0.336
VM-diff [13]	0.580±0.165	<0.01
ICNet [50]	0.742±0.223	0.854±0.316
CM [23]	0.737±0.123	1.719±0.382
ViT-V-Net [10]	0.734±0.124	1.609±0.319
nnFormer [51]	0.747±0.135	1.595±0.358
ICON [16]	0.752±0.155	0.695±0.248
TM [9]	0.754±0.124	1.579±0.328
TM-diff [9]	0.594±0.163	<0.01
TMBS [9]	0.761±0.122	<0.01
VM-ESC	0.743±0.025	0.478±0.101
TMBS-ESC	0.762±0.023	<0.01

Table 2. IXI dataset results.

Method	Dice↑	HD95↓	SDlogJ↓
VTN [25]	0.827±0.013	1.722±0.318	0.121±0.015
ConAdam [41]	0.846±0.016	1.500±0.304	0.067±0.005
VM [6]	0.847±0.014	1.546±0.306	0.133±0.021
ClapIRN [29]	0.861±0.015	1.514±0.337	0.072±0.007
TM [9]	0.862±0.014	1.431±0.282	0.128±0.021
TM-SC	0.862±0.014	1.449±0.310	0.084±0.005

(a) OASIS validation set results.

Method	Dice↑	HD95↓	SDlogJ↓
Initial	0.56	3.86	1.50
VTN [25]	0.80	1.77	0.08
ConAdam [41]	0.81	1.63	0.07
ClapIRN [29]	0.82	1.67	0.07
TM [9]	0.82	1.66	0.12
TM-SC	0.820	1.666	0.085

(b) OASIS test set results.

Table 3. OASIS dataset results, obtained from the LEARN2REG challenge organizers [20]. Since the listed top-ranking methods are already enclosed, we mainly study the impact of adding sanity checks, where the impact is proved to be negligible on these main metrics.

Method	TRE↓	STRE↓	ROB↑	FV↓	AJ×10^2↓
Initial	6.864±0.996	-	-	-	-
*DIRAC [31]	2.760±0.247	0.274±0.027	0.776±0.055	0.025±0.009	4.242±2.954
†DIRAC-SC	2.719±0.259	0.218±0.046	0.795±0.034	0.022±0.005	3.001±1.314

Table 4. BraTSReg dataset results, reported as means±deviations out of five-fold validations. *: We use their code and train models until convergence. †: We implement the cross-sanity check to replace the inverse consistent error part of DIRAC, and add our self-sanity check.

Figure 9. Example axial T1-weighted MR slices comparisons on BraTSReg dataset. The error maps are normalized to incorporate displacements in unit space predicted by DIRAC.

cating that our sanity-enforced image registration results in less self-sanity error from the self-sanity error map comparison and more regular maps in the deformed grid. All these comparisons prove the efficacy of our regularization-based sanity-enforcer to prevent both sanity errors.

Pre-operative and post-recurrence scans registration. BraTSReg [3] dataset contains 160 pairs of pre-operative and follow-up brain MR scans of glioma patients taken from different time points. Each time point contains native T1, contrast-enhanced T1-weighted, T2-weighted, and FLAIR MRI. Within the dataset, 140 pairs of scans are associated with 6 to 50 manual landmarks in both scans. The other 20 pairs of scans only have landmarks in the follow-up scan. Following [31], we create a five-fold schema where each fold has 122/10/28 for training/validation/test, respectively. We set $\alpha = 0.01$ and $\beta = 0.03$ in the cross-sanity check since DIRAC predicts normalized displacement (range 0 to 1). We take an average of five folds to report our results in Tab. 4. All baselines and our models use the same training setup. We specifically focus on the model training part and directly train all models on data downloaded from the offi-

cial website [5] without pre-affine and post-processing steps mentioned in their paper [30]. The best model is selected on the validation set, which has the lowest TREs. Despite the high performance of the baseline, we can observe that the substitution by cross-sanity check and the addition of self-sanity check enable all metrics' improvements to be claimed as state-of-the-art. We also show qualitative comparisons in Fig. 9, where the comparison between SSE error maps demonstrates the effectiveness of our self-sanity check.

4.2. Sanity Analysis

We conduct a broad sanity analysis on a wide range of models on the IXI dataset, including popular models (VM [6] and TM [9]), model with cycle consistency (CM [23]), models with inverse consistency (ICNet [50] and ICON [16]), probabilistically-formulated models (VM-diff [13] and TM-diff [9]), and models with diffeomorphic add-ons (MIDIR [34] and TMBS [9]). The results are shown in Tab. 5. We find that models with probabilistic formulation or cycle consistency have higher self-sanity, whereas the other models are insensitive to this scenario and have an inferior SDice. However, these probabilistic formulated models suffer from an relatively low Dice score, refraining from practical usage. For CM, albeit low self-sanity error, the correspondence error between image pairs is not negligible. We also find it interesting that models with diffeomorphic add-ons still suffer from large sanity errors, despite their ability to produce a diffeomorphic/regular map, caused by not modeling different mappings from both directions. Overall, our behavior studies show that most models suffer from inferior Dice/SDice performance, lower diffeomorphism, or significant sanity errors. Our proposed sanity checks improve performance from every aspect, producing a more regular map and preventing sanity errors.

Method	Main Metric↑		Diffeomorphism↓		Sanity Error↓	
	Dice	SDice	FV	AJ×10^4	SSE×10^{-1}	CSE
PVT [46]	0.727	0.695	1.858	3.64	52.95	19.43
CoTr [48]	0.735	0.846	1.298	2.10	6.28	21.37
MIDIR [34]	0.742	0.850	<0.1	<0.1	6.75	15.75
VM [6]	0.732	0.890	1.522	3.00	5.16	16.55
VM-diff [13]	0.580	1.000	<0.1	<0.1	0.13	<0.1
ICNet [50]	0.742	0.849	0.854	1.90	13.23	7.48
CM [23]	0.737	1.000	1.719	3.44	1.18	19.47
VIT-V-Net [10]	0.734	0.908	1.609	2.97	14.52	12.98
nnFormer [51]	0.747	0.799	1.595	2.90	13.31	20.68
ICON [16]	0.752	0.899	0.695	0.90	7.31	5.68
TM [9]	0.754	0.873	1.579	2.75	85.63	19.62
TM-diff [9]	0.594	1.000	<0.1	<0.1	0.16	<0.1
TMBS [9]	0.761	0.903	<0.1	0.02	23.84	14.08
VM-ESC	0.743	1.000	0.478	0.43	<0.1	2.55
TMBS-ESC	0.762	1.000	<0.1	<0.1	<0.1	2.99

Table 5. Sanity analysis of various models on IXI dataset.

[46]	[48]	[6]	[23]	[10]	[51]	[50]	[16]	[9]	[6]	$\frac{TMBS}{ESC}$	$\frac{VM}{ESC}$
19.24	10.02	5.63	7.58	11.27	10.44	3.46	2.78	22.61	7.31	1.96	**0.92**

Table 6. Cross-sanity Error using strict inverse consistency on IXI dataset, i.e., $\tilde{g}^{b \to a} + g^{a \to b} = 0$. We choose VM for [6] and TMBS for [9] as our baselines. We show our results in the last two columns. Still, we achieve state-of-the-art performance even under strict inverse consistency for calculating cross-sanity error.

λ_c	λ_s	Dice↑	SDice↑	FV↓	AJ×10^4↓	SSE×10^{-1}↓	CSE↓
5e-4	1e-3	0.720	0.957	0.913	1.24	0.658	6.96
	1e-2	0.720	0.983	0.920	1.27	0.299	6.77
	1e-1	0.719	1.000	0.896	1.22	<0.1	6.73
1e-3	1e-3	0.721	0.967	0.422	0.38	0.572	3.43
	1e-2	0.720	0.989	0.426	0.40	0.301	3.46
	1e-1	0.721	1.000	0.443	0.40	<0.1	3.43
1e-2	1e-3	0.473	0.912	<0.1	<0.1	1.966	<0.1
	1e-2	0.420	0.980	<0.1	<0.1	1.210	<0.1
	1e-1	0.406	1.000	<0.1	<0.1	0.209	<0.1

Table 7. Sanity loss weight study of VM on sub-IXI dataset.

Method	Dice↑	SDice↑	FV↓	AJ×10^4↓	SSE×10^{-1}↓	CSE↓
VM	0.732	0.890	1.522	3.00	5.156	16.55
VM-E	0.742	0.919	1.574	3.04	2.647	26.97
VM-ES	0.740	1.000	1.442	2.72	0.050	25.32
VM-EC	0.743	0.950	0.447	0.39	1.178	2.61
VM-ESC	0.743	1.000	0.478	0.43	<0.1	2.55
TM	0.862	0.925	0.752	1.52	3.069	10.72
TM-S	0.861	1.000	0.777	1.66	0.018	11.43
TM-C	0.862	0.948	0.246	0.26	0.991	2.93
TM-SC	0.862	1.000	0.307	0.35	<0.1	3.17

Table 8. Model ablations on IXI dataset (**top rows**) and OASIS validation dataset (**bottom rows**). Since TM is already enclosed (E), we focus on studying the impacts of adding our sanity checks.

4.3. Ablation Study

Cross-sanity error under strict inverse consistency. We testify the performance under strict inverse consistency as in Def. 1, the same as CICE in [11], shown in Tab. 6. The reductions of strict inverse consistency error compared using relaxed inverse consistency are caused by many small errors taken into account for calculating the average, where our sanity-checked methods still show better performance, compared to all models, including the model trained using explicitly strict inverse consistency, e.g. [16].

Sanity checks loss weight study. We derive an upper bound for our saner registration in Lemma 4. We show in the study that we can utilize this upper bound to guide us to set appropriate loss weight λ_c for the cross-sanity check. For fast verification, we randomly sample 50 subjects from the training set of the IXI dataset and validate/test in the full validation/test set. We also include the self-sanity check loss weight λ_s here for comparing purposes. The experimental results are shown in Tab. 7. Compared to λ_c, λ_s is insensitive to weight change. As we discussed earlier, if we set higher loss weight to cross-sanity check, for example, 0.01/0.001, approximately having upper bounds as 0.1/0.01, respectively. This creates a huge difference when we train our models. This high tolerance for error results in significantly degenerated performance for Dice, where our derived upper bound gives useful guidance for the setting of loss weights. More discussions can be found in Appendix.

Model ablation study. We study our proposed sanity checks on IXI and OASIS validation datasets and report the results in Tab. 8. We show that each sanity check reduces the corresponding errors without compromising other metrics' performance compared to their naive counterparts. We also find that the bidirectional registration and cross-sanity check can also mitigate self-sanity errors to a certain level, but cannot eliminate such errors completely, proving that our self-sanity check is necessary to regulate the models' behavior for mapping identical image pairs. More experiments such as parameters α and β numerical study, sanity preservation study, statistical significance of results, ablative qualitative results, etc, can be seen in the Appendix.

5. Conclusion

This paper focuses on correcting learning-based deep models' behaviors on single moving-fixed image pairs. In our model sanity analysis, we find that most existing models suffer from significant sanity errors, with no exceptions for models equipped with diffeomorphic add-ons. We show that this sanity-checked model can prevent such sanity errors without contaminating any registration performance. While the experimental results certify the effectiveness of our proposed sanity checks, our sanity checks are supported by a set of theoretical guarantees derived in this paper. We first show that there is an error upper bound for our sanity-checked formulation to the optimal condition. Then, we show that this upper bound can give significant guidance to train a sanity-checked registration model, where we believe it is beneficial for preventing overly optimization on image similarities when training deep registration models.

Acknowledgements

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 Kovačič. 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, 4

[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, 4

[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] Huaqi 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 Mattias P Heinrich. Fast 3d registration with accurate optimisation and little learning for learn2reg 2021. In *MICCAI*, 2021. 1, 2, 7

[42] Hessam Sokooti, Bob De Vos, Floris Berendsen, Boudewijn PF Lelieveldt, Ivana Išgum, 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] Wenhai 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