Appendix: 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

A. Appendix

A.1. Comparison Between Diffeomorphic Methods and Inverse Consistent Methods
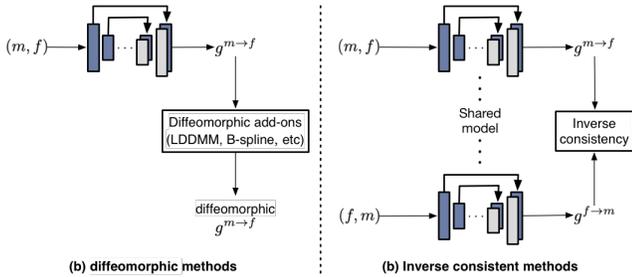
Figure A1. Illustration of two different types of methods. Technically, diffeomorphic methods [1, 4, 8, 5, 10, 9, 6, 7] are also inverse-consistent. But following the convention in [11, 2, 3], we restrict inverse consistency in the essence of operating on two different mappings predicted by the same model. Since the diffeomorphic methods only operate on one mapping, where the other mapping is calculated from that mapping, rather than predicted. Thus, they are called diffeomorphic methods instead.

A.2. Proof of Thm. 1

Theorem 1 (Relaxed ideal symmetric registration via cross-sanity check). *An ideal symmetric registration meets*

$$\varphi^{a\to b} \circ \varphi^{b\to a} = id,$$

in which id denotes the identity transformation. Then, a cross-sanity checked registration is a relaxed solution to the ideal registration, satisfying

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

Proof. A straightforward explanation for ideal symmetric registration $\varphi^{a\to b} \circ \varphi^{b\to a} = id$ would be that the coordinates of one pixel (p) for image a stay the same after two transformations: forward transformation (from a to b) and

---

[*]Corresponding author

backward transformation (from b back to a). Therefore, we have the chain of the coordinates changing

$$
\begin{aligned}
p & \\
\xrightarrow{a\to b}\ & g^{a\to b}(p) + p \\
\xrightarrow{b\to a}\ & g^{b\to a}(g^{a\to b}(p) + p) + g^{a\to b} + p \\
=\ & p.
\end{aligned}
\tag{A1}
$$

By simplification, we have the ideal inverse consistency:

$$
\underbrace{g^{b\to a}(g^{a\to b}(p) + p)}_{\tilde{g}^{b\to a}} + \underbrace{g^{a\to b}}_{g^{a\to b}} + p = p \\
\tilde{g}^{b\to a} \quad + \quad g^{a\to b} = 0.
\tag{A2}
$$

Different from the ideal inverse consistency, recall that our cross-sanity check in norm form is

$$||g^{a\to b} + \tilde{g}^{b\to a}||_2^2 < \alpha(||g^{a\to b}||_2^2 + ||\tilde{g}^{b\to a}||_2^2) + \beta N. \tag{A3}$$

Suppose $0 < \alpha < 1$ and $\beta > 0$. By expanding the cross-sanity check, we have

$$
\begin{aligned}
g^{a\to b\top}\tilde{g}^{b\to a} &< \frac{\beta N - (1-\alpha)(||g^{a\to b}||_2^2 + ||\tilde{g}^{b\to a}||_2^2)}{2} \\
&\le \frac{\beta N}{2}.
\end{aligned}
\tag{A4}
$$

So, we have

$$0 \le g^{a\to b\top}\tilde{g}^{b\to a} < \frac{\beta N}{2}, \tag{A5}$$

and

$$0 < ||g^{a\to b}||_2^2 + ||\tilde{g}^{b\to a}||_2^2 < \frac{\beta N}{1-\alpha}. \tag{A6}$$

Thus, we have

$$
\begin{aligned}
||g^{a\to b} + \tilde{g}^{b\to a}||_2^2 &= ||g^{a\to b}||_2^2 + ||\tilde{g}^{b\to a}||_2^2 + 2g^{a\to b\top}\tilde{g}^{b\to a} \\
&< \frac{\beta(2-\alpha)N}{1-\alpha}.
\end{aligned}
\tag{A7}
$$

Finally, we derive the lower/upper bound as

$$||g^{a \to b} + \tilde{g}^{b \to a}||_2^2 < \frac{\beta(2 - \alpha)N}{1 - \alpha}. \qquad \text{(A8)}$$

$\square$

It is obvious that our cross-sanity check formulation (Eq. (A8)) is a relaxed version of the strict symmetry in Eq. (A2). So far, we prove that under our cross-sanity check, the symmetry of $\tilde{g}^{b \to a}$ and $g^{a \to b}$ is bounded by α and β. When the strict symmetry is satisfied, our cross-sanity check is definitely satisfied. However, if our cross-sanity check is satisfied, it is not the other way around. We show in the experiments that our relaxed version of symmetry improves the overall results, quantitatively and qualitatively.

A.3. Proof of Thm. 2

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.*

Proof. In reality, a meaningful deformation field cannot be unbounded. We first restrict g to be a closed subset of $L^2(H^\star)$:

$$\mathcal{B} \triangleq \{g \in L^2(H^\star) :$$
$$||g||_{L^2(H^\star)}^2 \le B, B \in \mathbb{R}_+ \text{ only depends on } \Omega\}. \quad \text{(A9)}$$

We then seek solutions $g^\star$ to the minimization problem in the space $H^1(\Omega) \cap \mathcal{B}$, and meanwhile satisfying our proposed checks. For short notations, we denote the minimization problem as

$$\min_{g \in H^1 \cap \mathcal{B}} E(g),$$

s.t.

$$E(g) = -\text{Sim}(f, m \circ (g^{m \to f} + id)) + \lambda_r ||\text{Reg}(g^{m \to f})||_2^2, \quad \text{(A10)}$$

and λ_r is a positive constant. For $g \in H^1 \cap \mathcal{B}$, $E(g)$ is bounded below and there exists a minimizing sequence $\{g_k\}_{k=1}^\infty$ satisfying

$$E(g_{k+1}) \le E(g_k) \le \cdots \le E(g_1) \le \lim_{k \to \infty} E(g_k)$$
$$= \inf_{H^1 \cap B} E(g). \quad \text{(A11)}$$

For identical inputs, we have $||g||_2^2 = 0 < B \Rightarrow g \in H^1 \cap \mathcal{B}$, thus, the self-sanity is checked. For different image pairs, we have

$$||g + \tilde{g}||_2^2 < \alpha(||g||_2^2 + ||\tilde{g}||_2^2) + \beta N \le 2\alpha B + \beta N. \quad \text{(A12)}$$

By definition, g and its reversed displacement $\tilde{g}$ need to follow the constraint $g(x)\tilde{g}(x) \le 0, \forall x \in \Omega$, so we have

$$\max(||g + \tilde{g}||_2^2) < ||g||_2^2 \le B \Rightarrow 2\alpha B + \beta N < B. \quad \text{(A13)}$$

Here, we let $2\alpha B + \beta N < B$, since our cross-sanity check is considered a tighter bound than B. Thus, by choosing appropriate α and β, we can ensure that the cross-sanity is also checked, such that $g \in H^1 \cap \mathcal{B}$. Due to the fact that H^1 is precompact in L^2 space, there exists a convergent subsequence where we still denote as $\{g_k\}_{k=1}^\infty$, and $g^\star \in H^1$, such that $g_k \to g^\star$, which is strongly in L^2 and a.e. in Ω. Note that, either similarity functions (e.g. NCC) or distance functions (e.g. SSD) is naturally bounded in our image registration scenario, so that we can always have

$$-\text{Sim}(f, m \circ (g^\star + p)) \le \lim_{k \to \infty} -\text{Sim}(f, m \circ (g_k + p)). \quad \text{(A14)}$$

Besides, for the H^1 regularization, since $\{g_k\}_{k=1}^\infty$ is a bounded convergent sequence in $H^1 \cap \mathcal{B}$, and $g_k \to g^\star$ a.e. in Ω. By the dominant convergence theorem, we have

$$\lim_{k \to \infty} \text{Reg}(g_k) = \text{Reg}(g^\star). \quad \text{(A15)}$$

Combining Eq. (A14) with Eq. (A15), we obtain

$$E(g^\star) \le \lim_{k \to \infty} E(g_k) = \inf_{H^1 \cap B} E(g). \quad \text{(A16)}$$

Thus, $g^\star$ is indeed a solution to the minimization problem. So far, we prove that there exists a minimizer $g^\star$ of the modified optimization problem. We can then prove that $g^\star$ is unique. Note that, we assume that the similarity operator is concave (e.g., NCC, negative NCC is convex) when it is not a distance operator (e.g., SSD is convex) on the transformation with a pair of (m, f). I.e., $g^\star = g|m, f$, i.e., $g^\star$ learned by the model and conditioned on this specific (m, f) pair to satisfy the proposed sanity checks. Therefore, since the data term (e.g. SSD) and the regularization (H^1 regularization) are convex, together with the convex search space for g, the uniqueness of the minimizer $g^\star$ is proved. $\square$

A.4. Proof of Thm. 3

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*

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

Proof. Since $\mathbf{g}_*$ is optimal, thus we have that $-\text{Sim}(\mathbf{g}_*) + \frac{\lambda}{2}||\mathbf{A}\mathbf{g}_* - \mathbf{y}||_2^2 \le -\text{Sim}(\mathbf{g}_{\text{sanity}}) + \frac{\lambda}{2}||\mathbf{A}\mathbf{g}_{\text{sanity}} - \mathbf{y}||_2^2$. Since $||\mathbf{A}\mathbf{g}_* - \mathbf{y}||_2^2 = 0$, with elimination we have that $-\text{Sim}(\mathbf{g}_*) \le -\text{Sim}(\mathbf{g}_{\text{sanity}}) + \frac{\lambda}{2}||\mathbf{A}\mathbf{g}_{\text{sanity}} - \mathbf{y}||_2^2$. By
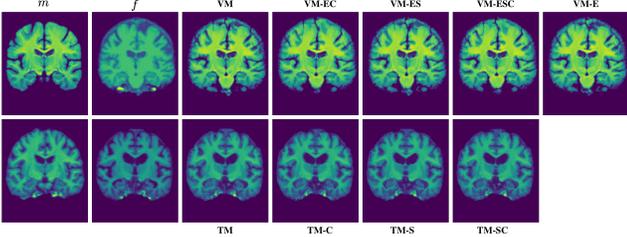
Figure A2. Qualitative comparisons of different model variants on: (*Top*) IXI dataset, (*Bottom*) OASIS validation dataset.

substituting $\mathbf{y}$ with $\mathbf{Ag}_*$ and combining like terms, we have Eq. (12). Therefore, we prove the loyalty of the unique minimizer $\mathbf{g}_{\text{sanity}}$ to the optimal minimizer $\mathbf{g}_*$, controlled by weight λ. $\qquad\square$

A.5. Proof of CS Error Upper Bound

Recall we want to prove that the CS error is upper bounded in the form of

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

Proof. Recall CS error in Eq. (9), and by Thm. 1, we have

$$\|\text{CS}(\mathbf{g}_{\text{sanity}})\|_2^2$$
$$= \|\mathbf{g}_{\text{sanity}} + \tilde{\mathbf{g}}_{\text{sanity}}\|_2^2 - \alpha(\|\mathbf{g}_{\text{sanity}}\|_2^2 + \|\tilde{\mathbf{g}}_{\text{sanity}}\|_2^2) - 2\beta N$$
$$= (1-\alpha)(\|\mathbf{g}_{\text{sanity}}\|_2^2 + \|\tilde{\mathbf{g}}_{\text{sanity}}\|_2^2$$
$$\qquad\qquad + \frac{2}{1-\alpha}\mathbf{g}_{\text{sanity}}^\top\tilde{\mathbf{g}}_{\text{sanity}}) - 2\beta N$$
$$\leq (1-\alpha)(\|\mathbf{g}_{\text{sanity}}\|_2^2 + \|\tilde{\mathbf{g}}_{\text{sanity}}\|_2^2$$
$$\qquad\qquad + 2\mathbf{g}_{\text{sanity}}^\top\tilde{\mathbf{g}}_{\text{sanity}}) - 2\beta N$$
$$= (1-\alpha)\|\mathbf{g}_{\text{sanity}} + \tilde{\mathbf{g}}_{\text{sanity}}\|_2^2 - 2\beta N$$
$$< (1-\alpha)\frac{2\beta(2-\alpha)}{1-\alpha}N - 2\beta N$$
$$= 2(1-\alpha)\beta N, \quad \text{where} \quad 0 < \alpha < 1 \quad \text{and} \quad \beta > 0.$$
$$\text{(A17)}$$

The first inequality holds since $\mathbf{g}_{\text{sanity}}^\top\tilde{\mathbf{g}}_{\text{sanity}} \leq 0$ for valid inverse consistent displacements. The second inequality holds for two directions ($m{\to}f$ and $f{\to}m$) by Thm. 1. Thus, the proof for CS error upper bound is completed. $\quad\square$

A.6. Cross-sanity Check Numerical Study

We present our numerical study for α and β parameters in Tab. A1. This ablation study is conducted on the same subset of IXI training dataset described in the ablation study section but validate/test in the entire validation/test set. Our thought is that compared to the problem of enforcing strict inverse consistency over two different displacements, the relaxed version might be easier to solve, mathematically.

α	β	Dice	SDice	FV	AJ$\times 10^4$	SSE$\times 10^{-1}$	CSE
*0.1	5	0.7263	0.8850	1.5220	3.00	5.156	10.72
†0.1	5	0.7223	0.8745	1.6130	3.30	4.337	11.05
0	0	0.7178	0.9486	1.4041	2.74	0.693	$7.58\to1.81_{76.12\%}$
	0.1	0.7199	0.9533	1.2460	2.30	0.721	$6.39\to2.21_{65.41\%}$
	1	0.7223	0.9654	1.2070	2.02	0.643	$6.81\to4.60_{32.45\%}$
	3	0.7233	0.9685	1.0160	1.51	0.596	$8.70\to5.19_{40.34\%}$
	5	0.7247	0.9588	0.8216	1.09	0.723	$10.72\to5.03_{53.07\%}$
	7	0.7226	0.9689	0.6632	0.77	0.642	$12.53\to4.52_{63.92\%}$
0.1	8	0.7232	0.9600	0.6152	0.68	0.668	$13.39\to4.27_{68.11\%}$
	9	0.7226	0.9699	0.5534	0.58	0.572	$14.25\to4.03_{71.72\%}$
	11	0.7206	0.9665	0.4472	0.41	0.564	$15.83\to3.53_{77.70\%}$
	12	0.7211	0.9668	0.4219	0.38	0.572	$16.55\to3.43_{79.27\%}$
	13	0.7186	0.9742	0.4278	0.39	0.543	$17.22\to3.39_{80.31\%}$
	14	0.7189	0.9678	0.4182	0.37	0.543	$17.80\to3.24_{81.80\%}$
	16	0.7179	0.9681	0.4345	0.40	0.561	$18.79\to2.89_{84.62\%}$
	20	0.7124	0.9693	0.4668	0.44	0.577	$20.36\to2.82_{86.15\%}$
0.01		0.7229	0.9677	0.9842	1.46	0.623	$11.30\to5.43_{51.94\%}$
0.10		0.7247	0.9588	0.8216	1.09	0.723	$10.72\to5.03_{53.07\%}$
0.15	5	0.7230	0.9624	0.7891	0.99	0.591	$10.42\to4.77_{54.22\%}$
0.20		0.7222	0.9656	0.7479	0.90	0.586	$10.15\to4.53_{55.37\%}$
0.50		0.7218	0.9570	0.7276	0.86	0.642	$8.64\to3.47_{59.83\%}$
1		0.7165	0.9353	0.9622	1.45	1.277	$7.41\to1.93_{73.95\%}$

Table A1. Ablation study for α, β, where the loss weights for both sanity check losses are set to 0.001. $*$ denotes the checkpoint model used in this ablation (resulting in initial CSE errors on the left-side of arrows), and † represents where we follow the standard protocol to finetune the model for the same epochs as the rest ablative experiments. In this case, α and β are only used to calculate the CSE. Along with SSE, they are not part of the training loss.

Again, practically speaking, it is rather difficult to find one-to-one correspondence for every point in the moving-fixed image pair, which is the strict inverse consistency saying, relaxing or setting an error threshold can be effective from this perspective. Our error-bound formulation implicitly presents guidance for training such sanity-checked models.

Settings of α and β. A uniform estimate of α and β is possible, however, such a bound is not sharp, and it will lead to over-estimation of λ_c (the regularization parameter) for different applications. Hence, we prefer to derive the bounds on α and β on particular sets of applications, where we can easily find such bounds from the formulations. As stated in **Thm. 1**, α and β control the relaxation. We can also directly derive an upper bound from the check that constrains the ratio of two displacements. Since $g^{a\to b}\tilde{g}^{b\to a}<0$, then $\frac{g^{a\to b}}{\tilde{g}^{b\to a}} + \frac{\tilde{g}^{b\to a}}{g^{a\to b}} < \frac{2}{1-\alpha}$, β is neglected for simplicity. These two bounds estimate ranges of α and β for our relaxation. E.g., we use existing models (e.g., VM) to predict ten samples randomly, and set β to 0.15× maximum displacement; set α to 0.1 for models outputting absolute displacements (e.g., VM and TM), or α to 0.01 for models outputting relative displacements (e.g., DIRAC). Note that this only needs to be done once, while the previous experiment shows that the registrations are pretty robust among a range of α and β. Thus, it is safe to choose α and β within the range.

Train / Test	Model	Dice	SDice	HD95	SDlogJ	FV	AJ×10⁴	SSE×10⁻¹	CSE
IXI / OASIS	VM	0.714	0.848	3.039	0.118	0.831	1.52	6.24	7.49
	VM-ESC	0.759	1.000	2.563	0.080	0.173	0.16	0.00	2.86
	TMBS	0.775	0.901	2.201	0.072	0.00	0.00	24.4	15.16
	TMBS-ESC	0.783	1.000	2.144	0.061	0.00	0.00	0.00	3.73
OASIS / IXI	TM	0.705	0.918	4.083	0.132	1.390	2.67	3.18	13.78
	TM-SC	0.718	1.000	3.903	0.097	0.674	0.84	0.01	5.58

Table A2. Generalization ability study on sanity-checked registers to study cross-dataset registration performance. Here, we set $\alpha=0.1$, and $\beta=12$ to calculate cross-sanity errors in both settings.
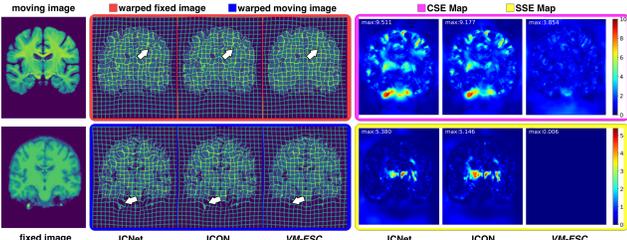
Figure A3. Qualitative comparisons between ICNet [11], ICON [2] and our method on IXI dataset.

A.7. Ablative Qualitative Comparisons

Qualitative comparisons between our ablative model variants are shown in Fig. A2.

A.8. Sanity-awareness Preservation Study on Cross-Dataset Scenario

We test whether sanity awareness is preserved in cross-dataset scenarios. We train our sanity-checked register in one dataset and then test it on different dataset so no overlapping between training and testing datasets. The results are shown in Tab. A2. Compared to methods without sanity checks, our sanity-checked models improve in every metric, certifying that our sanity checks do not harm the model training. Besides, the sanity-checked registers still preserve good sanity for preventing corresponding errors.

A.9. Experimental Results Statistical Significance

We specifically study whether our results are statistically significant, compared to the other strong baselines, e.g., ICON [2] and DIRAC [3]. We calculate `p value` using `scipy` package. Compared to ICON, our *VM-ESC* (`p value: 0.0174`) and *TMBS-ESC* (`p value: 0.0280`), while for DIRAC, our *DIRAC-SC* (`p value: 0.0204`), considering all metrics shown in the corresponding tables. All the `p values` are < 0.05, indicating that our results are statistically significant.

A.10. Error Map Comparisons between Inverse Consistent Methods

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

Method	TRE↓	STRE↓	ROB↑	FV↓	AJ×10²↓
DIRAC	2.760±0.247	0.274±0.027	0.776±0.055	0.025±0.009	4.242±2.954
DIRAC-C	2.721±0.262	0.268±0.039	0.791±0.044	0.022±0.008	3.012±1.442
DIRAC-SC	2.719±0.259	0.218±0.046	0.795±0.034	0.022±0.005	3.001±1.314

Table A3. Performance of replacing the inverse consistent error.

A.11. Performance of Replacing DIRAC's Inverse Error

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

A.12. 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 $\mathcal{L}_{\text{sim}}$. 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.

References

[1] Ruzena Bajcsy and Stane Kovačič. Multiresolution elastic matching. *CVGIP*, 46(1):1–21, 1989. 1

[2] Hastings Greer, Roland Kwitt, Francois-Xavier Vialard, and Marc Niethammer. Icon: Learning regular maps through inverse consistency. In *ICCV*, 2021. 1, 4

[3] 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. 1, 4

[4] 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

[5] Tanya Schmah, Laurent Risser, and François-Xavier Vialard. Left-invariant metrics for diffeomorphic image registration with spatially-varying regularisation. In *MICCAI*, 2013. 1

[6] Zhengyang Shen, Xu Han, Zhenlin Xu, and Marc Niethammer. Networks for joint affine and non-parametric image registration. In *CVPR*, 2019. 1

[7] Zhengyang Shen, François-Xavier Vialard, and Marc Niethammer. Region-specific diffeomorphic metric mapping. *NeurIPS*, 2019. 1

[8] 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

[9] Jian Wang and Miaomiao Zhang. Deepflash: An efficient network for learning-based medical image registration. In *CVPR*, 2020. 1

[10] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, 158:378–396, 2017. 1

[11] Jun Zhang. Inverse-consistent deep networks for unsupervised deformable image registration. *arXiv:1809.03443*, 2018. 1, 4