Learning Diffeomorphism for Medical Image Registration with Time-Embedded Architectures Using Semigroup Regularization

Supplementary Material

7. Proofs

In this section, we provide the proof of Theorem 1. We begin with some definitions and two lemmas establishing invertibility of ϕ_t and closure under composition. Since ϕ_t is implemented via differentiable neural networks, we assume it has mild smoothness and continuity properties.

Definition 1. *(Dyadic Composition) If the deformation ϕ_t satisfies the composition rule of Eq. (11), it satisfies the **dyadic composition**:*

$$\phi_s = \phi_{\frac{s+1}{2}} \circ \phi_{\frac{s-1}{2}} = \phi_{\frac{s-1}{2}} \circ \phi_{\frac{s+1}{2}}. \tag{13}$$

simply by the change of variable $s = 2t - 1$

Definition 2. *(Dyadic Expansion) Expanding each term of the dyadic composition of Eq. (13) yields:*

$$\begin{aligned}\phi_s &= \phi_{\frac{s+1}{2}} \circ \phi_{\frac{s-1}{2}} \\ &= (\phi_{\frac{s+3}{4}} \circ \phi_{\frac{s-1}{4}}) \circ (\phi_{\frac{s+1}{4}} \circ \phi_{\frac{s-3}{4}}) \\ &\vdots \\ &= (\phi_{\frac{s}{2^n}+\frac{2^n-1}{2^n}} \circ \phi_{\frac{s}{2^n}-\frac{1}{2^n}}) \circ \cdots \circ (\phi_{\frac{s}{2^n}+\frac{1}{2^n}} \circ \phi_{\frac{s}{2^n}-\frac{2^n-1}{2^n}}).\end{aligned} \tag{14}$$

*which forms a depth-n binary tree with 2^n leaves called a **n-dyadic expansion** of the initial dyadic composition. Specifically, $s = 0$ yields*

$$\phi_0 = (\phi_{\frac{2^n-1}{2^n}} \circ \phi_{\frac{-1}{2^n}}) \circ \cdots \circ (\phi_{\frac{1}{2^n}} \circ \phi_{\frac{1-2^n}{2^n}}), \tag{15}$$

Definition 3. *(Dyadic Grid) For any $n \in \mathbb{N}$, define the **n-dyadic grid** A_n as*

$$A_n := \left\{ \pm \frac{2m+1}{2^n} : 2m+1 < 2^n,\ m \in \mathbb{Z}^+ \right\}; \tag{16}$$

more precisely, A_n contains the indices appearing the n-dyadic expansions of ϕ_s for $s = 0$.

Lemma 1. *For any ϕ_t, $t \in A_n$, there exists a n-dyadic expansion of ϕ_0 starting from ϕ_t and a n-dyadic expansion ending with ϕ_t.*

Proof. Fix an arbitrary n-dyadic expansion of ϕ_0 and its n-dyadic grid A_n. This expansion corresponds to a full binary tree of depth n:

- The root represents ϕ_0
- A node ϕ_a, $a < n$ has two children $\phi_{\frac{a+1}{2}}$ and $\phi_{\frac{a-1}{2}}$.
- Dyadic composition of Eq. (13) allows us to compute each ϕ_a by either $\phi_{\frac{a+1}{2}} \circ \phi_{\frac{a-1}{2}}$ or $\phi_{\frac{a-1}{2}} \circ \phi_{\frac{a+1}{2}}$.
- The leaves of the tree are exactly $\{\phi_u : u \in A_n\}$.

Since we have a binary tree, for each leaf node ϕ_t, $t \in A_n$ there exists a unique path to the root:

$$\phi_t \to \pi(\phi_t) \to \pi(\pi(\phi_t)) \to \cdots \to \phi_0,$$

where $\pi(\phi_t)$ represents the parent node of ϕ_t.

We prove the lemma by construction. We now construct a valid dyadic expansion by following this path from the bottom up:

- Initialization: Set the current expression to ϕ_t.
- Climb one level: At each internal node on the path, ϕ_t is composed with the sibling node from left, producing the parent node.
- Iterate: Repeat this process at each higher ancestor until reaching the root node ϕ_0.

Since at each step (i) we use compositions already present in the dyadic tree and (ii) we only choose the local order (left vs right) at each internal node allowable by dyadic composition, the final expression is a valid n-dyadic expansion of ϕ_0 starting with ϕ_t.

To construct a n-dyadic expansion ending with ϕ_t we repeat exactly the same bottom-up construction but always place the current expression on the right side of each composition. □

Lemma 2. *The mapping $\phi_t(x)$ satisfying the composition rule of Eq. (11) and $\phi_0(x) = x$ is invertible for every $t \in (-1, 1)$:*

Proof. Fix $s \in (-1, 1)$. For an arbitrary $n \in \mathbb{N}$ we can obtain the n-dyadic expansions of ϕ_s as in Eq. (14).

Setting $s = 0$ in Eq. (14) and using $\phi_0 = \mathrm{Id}$, we obtain a composition of 2^n factors whose product equals the identity. According to Lemma 1, each factor in the expansion must be invertible (each factor can appear both in the beginning and end of the composition chain, leading to being left and right invertible). Considering the n-dyadic grid A_n corresponding to the n-dyadic expansion of $\phi_0 = \mathrm{Id}$, it is deduced that ϕ_t is invertible for each $t \in A_n$.

For any $s \in (-1, 1)$ and $\varepsilon > 0$, there exists a $n \in \mathbb{N}$ and $p_s \in A_n$ with $|s - p_s| < \varepsilon$. By continuity of $t \mapsto \phi_t$, we have

$$\lim_{p_s \to s} \phi_{p_s} = \phi_s.$$

Since the limit of invertible mappings is invertible, ϕ_s is invertible for every $s \in (-1, 1)$, and continuity extends the result to $s = \pm 1$. □

Lemma 3. *Take a n-dyadic grid A_n of ϕ_0 for an arbitrary n. For any distinct $p, q \in A_n$, the factors ϕ_p and ϕ_q appear adjacently in at least one of the $2^{\frac{n(n+1)}{2}}$ expansions in the n-dyadic expansions of Eq. (15).*

Proof. Removing parentheses using associativity of compositions, each expansion in Eq. (15) corresponds to a sequence of 2^n factors connected by $2^n - 1$ compositions. Note that there are a total of $2^{\frac{n(n+1)}{2}}$ possible expansions in Eq. (15) at level n.

The claim is trivial for $n = 1$. We can verify the claim for $n = 2$ with enumeration; there are 4 constituent factors $\phi_{3/4}, \phi_{-1/4}, \phi_{-3/4}$, and $\phi_{1/4}$, and all $2C_2^4 = 12$ compositions $\phi_{3/4} \circ \phi_{-3/4}, \phi_{3/4} \circ \phi_{-1/4}, \phi_{3/4} \circ \phi_{1/4}, \ldots$, appear in 8 possible expansions.

Assume the result for $n - 1$. Take $p, q \in A_n$ and $p \neq q$. If they have the same parent in A_{n-1}, they are adjacent. Otherwise, their parents must be adjacent by the inductive hypothesis, and one verifies adjacency among their children by direct construction (as in $n = 2$). Therefore, the claim follows by induction. □

Corollary 1. *The family $\{\phi_t\}$ is closed under composition.*

Proof. According to Lemma 2, ϕ_s is invertible for any $s \in (-1, 1)$. For any s, let ψ_s denote the inverse of ϕ_s. For any $s, t \in (-1, 1)$, both ϕ_s and ϕ_t are invertible. Thus, the composition $y = \phi_s \circ \phi_t$ must also be invertible with an inverse mapping $x = \psi_{f(s,t)}(y)$, where $f : (-1, 1)^2 \to \mathbb{R}$. Therefore, we must have $\phi_s \circ \phi_t = \phi_{f(s,t)}$, indicating that $\{\phi_t\}$ is closed under composition. □

Theorem 1. *The deformation $\phi_t(x)$ satisfying the composition rule of Eq. (11) and $\phi_0(x) = x$ is an exponential map, or equivalently, a one-parameter diffeomorphism solving the autonomous ODE of Eq. (1); see [7].*

Proof. Define $f(s, t)$ by $\phi_s \circ \phi_t = \phi_{f(s,t)}$. Applying f to the nested expansion Eq. (15) yields

$$0 = f\left(\tfrac{2^n-1}{2^n}, f(-\tfrac{1}{2^n}, \ldots, f(\tfrac{1}{2^n}, -\tfrac{2^n-1}{2^n}))\right). \tag{17}$$

Defining f to be the addition operation (+), Eq. (17) becomes consistent. Moreover, note that the order of composition can start with any two adjacent factors in Eq. (15) (due to the associative property). According to Lemma 3, for any $p, q \in A_n$ and $p \neq q$, we know that ϕ_p and ϕ_q appear next to each other in at least one of the $2^{\frac{n(n+1)}{2}}$ possible expansions in Eq. (14). Consistent evaluations of these $2^{\frac{n(n+1)}{2}}$ expressions allow us to define $f(p, q) = p + q$ for any $p, q \in A_n$, for all n, such that $p \neq q$. Therefore, we will have $\phi_p \circ \phi_q = \phi_{p+q}$ for any $p, q \in A_n$ and $p \neq q$. Letting p_s and p_t denote the closest grid points to $s, t \in (-1, 1)$, where $s \neq t$, respectively, and taking the limits, we will get

$$\begin{aligned} \phi_s \circ \phi_t &= \lim_{p_s, p_t \in A_n, n \to \infty} \phi_{p_s} \circ \phi_{p_t} \\ &= \lim_{p_s, p_t \in A_n, n \to \infty} \phi_{p_s + p_t} \\ &= \phi_{s+t}. \end{aligned} \tag{18}$$

Additionally, taking the limit, we will get,

$$\phi_s \circ \phi_s = \lim_{t \to s, t \in (-1,1)} \phi_s \circ \phi_t = \lim_{t \to s, t \in (-1,1)} \phi_{s+t} = \phi_{2s}. \tag{19}$$

This proves that ϕ_s is a one-parameter family of transformations. Additionally, ϕ_s is differentiable, and therefore, it is an exponential map [7]. □

8. Datasets Preparation

OASIS Dataset The Open Access Series of Imaging Studies (OASIS) dataset[4] [30] contains 416 T1 weighted scans from subjects aging from 18 to 96 with 100 of them diagnosed with mild to moderate Alzheimer's disease. The segmentation masks of 35 subcortical regions available in the dataset serve as the ground truth for further evaluation of the registration. Out of 416 subjects, we use 256, 60, 100 subjects for training, validation, and testing, respectively. By random pairing we generate 1000 pairs for training, 100 pairs for validation and 1000 pairs for test.

CANDI Dataset The Child and Adolescent NeuroDevelopment Initiative (CANDI) dataset[5] [26] contains T1 weighted brain scans of 117 subjects divided into 4 different subgroups including Healthy Control (**HC**), Schizophrenia Spectrum (**SS**), Bipolar Disorder with Psychosis (**BPDwithPsy**), and Bipolar Disorder without Psychosis (**BPDwithoutPsy**). The segmentation masks of 32 subcortical regions available in the dataset are used as the ground truth for the evaluation. For training and evaluation, we mix all the groups, and from 117 subjects, we use 80, 11, 26 subjects for training, validation, and testing, respectively. We generate 400 pairs for training, 25 pairs for validation, and 325 pairs for test.

LPBA40 Dataset The LONI Probabilistic Brain Atlas (LPBA40) dataset[6] [41] comes with 40 scans and the segmentation maps of 56 cortical and subcortical regions. For

[4] https://github.com/adalca/medical-datasets/blob/master/neurite-oasis.md

[5] https://www.nitrc.org/projects/candi_share/

[6] https://www.loni.usc.edu/research/atlas_downloads

		Atlas-Based					Inter-Subject				
Type	**Method**	**Dice↑**	$\lvert J\rvert_{<0}\%\downarrow$	**HD95↓**	**SSIM↑**	**ASSD↓**	**Dice↑**	$\lvert J\rvert_{<0}\%\downarrow$	**HD95↓**	**SSIM↑**	**ASSD↓**
Diffeomorphic	SyN	76.88 ± 1.04	0.0015 ± 0.0009	2.65 ± 0.33	83.87 ± 1.34	0.75 ± 0.27	74.12 ± 1.13	0.0048 ± 0.0008	4.90 ± 0.87	82.26 ± 1.67	0.77 ± 0.38
	SYMNet	80.22 ± 1.47	0.0014 ± 0.0007	2.28 ± 0.57	85.14 ± 1.70	0.85 ± 0.36	78.08 ± 2.11	0.0071 ± 0.0023	4.65 ± 1.03	84.45 ± 1.45	0.99 ± 0.44
	LapIRN	80.36 ± 1.13	0.0011 ± 0.0007	2.21 ± 0.43	85.78 ± 1.76	0.93 ± 0.42	80.79 ± 1.61	0.0023 ± 0.0019	3.90 ± 1.01	87.60 ± 1.59	1.05 ± 0.47
	CycleMorph	82.07 ± 1.20	0.0084 ± 0.0123	2.79 ± 0.51	87.57 ± 1.49	0.75 ± 0.26	81.58 ± 2.03	0.0173 ± 0.0086	2.26 ± 0.79	88.10 ± 1.38	0.77 ± 0.28
	GradICON	84.53 ± 1.52	0.0022 ± 0.0019	2.23 ± 0.55	85.90 ± 1.65	**0.71 ± 0.31**	83.89 ± 1.82	0.0039 ± 0.0012	2.09 ± 0.36	90.17 ± 1.19	0.78 ± 0.36
	NePhi	80.51 ± 0.56	0.0005 ± 0.0006	2.37 ± 0.44	84.15 ± 0.87	0.88 ± 0.38	80.09 ± 1.17	0.0056 ± 0.0023	3.28 ± 0.44	86.15 ± 1.90	0.90 ± 0.32
	TransMorph-diff	84.63 ± 1.12	0.0091 ± 0.0115	2.25 ± 0.53	89.91 ± 1.13	0.76 ± 0.35	83.51 ± 1.52	0.0066 ± 0.0073	2.35 ± 0.76	91.25 ± 2.07	0.78 ± 0.27
	PULPo	78.94 ± 1.42	0.0103 ± 0.0057	2.63 ± 0.53	86.80 ± 1.67	0.79 ± 0.39	80.51 ± 0.99	0.0527 ± 0.0240	3.45 ± 0.98	86.71 ± 1.86	0.85 ± 0.34
	R2Net	80.78 ± 1.47	0.0036 ± 0.0028	2.63 ± 0.74	86.14 ± 1.52	0.87 ± 0.46	80.11 ± 1.53	0.0093 ± 0.0054	2.15 ± 0.49	87.44 ± 1.15	1.02 ± 1.02
	NODEO	81.87 ± 3.87	0.0016 ± 0.0020	2.38 ± 0.99	87.05 ± 2.69	0.77 ± 0.25	81.97 ± 1.39	0.0024 ± 0.0010	2.18 ± 0.71	87.68 ± 1.93	0.83 ± 0.24
	SGDIR DiT ($\lambda = 10^5$)	**86.53 ± 0.82**	**0.0 ± 0.0**	**1.90 ± 0.52**	**91.45 ± 1.01**	0.72 ± 0.22	**85.90 ± 1.94**	0.0003 ± 0.0005	**1.87 ± 0.59**	**92.52 ± 0.97**	**0.75 ± 0.19**
	SGDIR UNet ($\lambda = 10^5$)	86.16 ± 0.89	**0.0 ± 0.0**	1.96 ± 0.51	90.71 ± 1.01	0.74 ± 0.21	85.36 ± 1.67	**0.0001 ± 0.0002**	1.93 ± 0.47	92.12 ± 1.47	0.77 ± 0.20
Deformable	VoxelMorph	80.13 ± 1.15	0.0044 ± 0.0027	2.68 ± 0.52	87.35 ± 1.37	0.85 ± 0.33	77.34 ± 1.41	0.0173 ± 0.103	3.43 ± 0.93	85.26 ± 2.01	0.91 ± 0.37
	DiffuseMorph	80.06 ± 1.72	0.6344 ± 0.1023	2.57 ± 0.53	86.17 ± 1.61	0.76 ± 0.28	79.13 ± 2.03	1.1563 ± 0.2512	3.51 ± 0.77	87.79 ± 1.91	0.85 ± 0.29
	DiffuseReg	82.53 ± 1.04	0.1986 ± 0.0392	2.27 ± 0.53	88.02 ± 1.25	0.78 ± 0.30	80.12 ± 1.22	1.0877 ± 0.4312	3.73 ± 1.01	87.94 ± 2.15	0.90 ± 0.33
	TransMorph	85.26 ± 1.05	2.0155 ± 0.4260	2.39 ± 0.49	91.79 ± 1.90	0.72 ± 0.24	84.11 ± 1.30	1.0665 ± 0.5631	2.26 ± 0.68	91.14 ± 1.49	0.80 ± 0.26
	TransMatch	81.54 ± 1.45	0.0237 ± 0.0154	2.37 ± 0.57	88.22 ± 1.37	0.69 ± 0.23	83.48 ± 1.80	0.1493 ± 0.0151	2.23 ± 0.66	90.26 ± 2.11	0.75 ± 0.21
	HViT	85.38 ± 0.93	0.3566 ± 0.0832	2.13 ± 0.54	92.02 ± 1.07	0.77 ± 0.25	85.07 ± 1.05	0.4812 ± 0.0114	1.92 ± 0.46	92.13 ± 1.39	0.83 ± 0.31
	CorrMLP	83.07 ± 0.91	0.0125 ± 0.0059	2.13 ± 0.55	88.32 ± 1.88	0.69 ± 0.22	84.66 ± 1.24	0.4640 ± 0.1776	2.23 ± 0.62	90.29 ± 1.84	0.76 ± 0.22
	SACB-Net	81.60 ± 0.61	**0.0020 ± 0.0013**	2.19 ± 0.48	86.82 ± 0.94	0.68 ± 0.24	83.17 ± 1.18	**0.0184 ± 0.0093**	2.33 ± 0.48	88.66 ± 2.13	0.80 ± 0.21
	SGDIR DiT ($\lambda = 10^4$)	**88.09 ± 0.91**	0.4332 ± 0.1121	**1.73 ± 0.41**	**93.80 ± 1.15**	**0.65 ± 0.17**	**87.82 ± 0.77**	0.3982 ± 0.1213	1.83 ± 0.65	**95.32 ± 1.08**	**0.70 ± 0.16**
	SGDIR UNet ($\lambda = 10^4$)	87.97 ± 0.85	0.5991 ± 0.1043	1.96 ± 0.48	93.44 ± 0.80	0.68 ± 0.19	87.74 ± 1.13	0.5018 ± 0.0243	**1.63 ± 0.33**	94.60 ± 1.25	0.72 ± 0.26

Table 8. The complete quantitative registration results with atlas-based and inter-subject registration settings over the OASIS dataset. The best performing method in each category is shown in **bold** and the best competitor model is presented in blue.

the LPBA40 dataset 190 pairs from 20 subjects are used as the training set, 10 pairs from 5 subjects are used as the validation set and 105 pairs from 15 subjects are used for the test set.

IXI Dataset The Information eXtraction from Images (IXI) dataset [7] contains 581 brain scans of normal, healthy subjects. From all 581 scans, we use 450, 50, 81 subjects for training, validation, and testing, respectively, from which we create 1500, 100, 300 random pairs for training, validation, and testing.

Mindboggle101 Dataset The Mindboggle101 dataset[8] [29] contains manually labeled T1-weighted brain MRI scans from 101 healthy subjects, providing cortical and subcortical parcellations with high-quality annotations. From the 101 subjects, we use 70 subjects for training, 10 for validation, and 21 for testing. By random sampling, we generate 1000 image pairs for training, 45 pairs for validation, and 210 pairs for testing to evaluate registration performance.

AbdomenCTCT Dataset The AbdomenCTCT dataset is sourced from the Learn2Reg Challenge[9], consisting of paired abdominal CT scans with multi-organ annotations. The dataset includes intra-patient scans across different time points, enabling the evaluation of deformable registration in challenging abdominal regions. We follow the official Learn2Reg challenge protocol, using the predefined training, validation, and testing splits provided by the organizers, without altering the subject allocation.

LungCT Dataset The LungCT dataset, also originating from the Learn2Reg Challenge [10], contains thoracic CT scans with large anatomical variations due to respiration and pathology. The dataset is designed to benchmark non-rigid registration methods on lung structures. As per the challenge guidelines, we adopt the official partitioning of the dataset into training, validation, and test sets, preserving the predefined subject splits to ensure fair comparison with previous works.

ACDC Dataset The Automated Cardiac Diagnosis Challenge (ACDC) dataset[11] [40] provides cine-MRI scans of cardiac cycles with segmentations of the left ventricle, right ventricle, and myocardium across pathological subgroups. Out of 150 patients, we use 90 subjects for training, 15 subjects for validation, and 45 subjects for testing. Registration pairs are chosen based on the End Systolic (ES) and End Diastolic (ED) frames for each patient.

Data Preparation For all brain MRI datasets, we use skull-stripped images that are spatially normalized to the MNI152 1mm template and further center-cropped to a resolution of $160 \times 144 \times 192$. Across all datasets, we evaluate registration performance using the full set of avail-

[7] https://brain-development.org/ixi-dataset

[8] https://mindboggle.info

[9] https://learn2reg.grand-challenge.org/Datasets/

[10] https://learn2reg.grand-challenge.org/Datasets/

[11] https://www.creatis.insa-lyon.fr/Challenge/acdc/

able anatomical structures. Intensity values are scaled to the $[0, 1]$ range using min–max normalization. Since the IXI dataset does not provide segmentation labels, we generate them using FreeSurfer [18], obtaining masks for 32 subcortical anatomical regions. For the AbdomenCTCT and LungCT datasets, all CT volumes are resampled to an isotropic resolution of 1mm. In the case of the ACDC dataset, cardiac MRI slices are center-cropped to 128×128 to ensure consistency across samples. The pixel sizes are kept in the original format of 1.56×1.56 which are used for computing metrics such as HD95.

9. Implementation Details

To implement SGDIR, we use two backbone architectures common in diffusion models: a time-embedded U-Net and a U-Net augmented with a Diffusion Transformer (DiT). Both are optimized using normalized cross-correlation (NCC) loss with window size 11. We use sinusoidal positional embedding for the initial time embeddings for both architectures.

SGDIR U-Net (Time-Embedded U-Net). This model follows a U-Net encoder–decoder with temporal conditioning via diffusion-style time embeddings. The network uses symmetric downsampling and upsampling with channel widths of $[16, 32, 64, 128, 256]$. We condition only the decoder on a 128-dimensional time embedding, which we found sufficient for stable training and strong performance. To enhance feature propagation and contextual awareness, we incorporate attention gates on skip connections and squeeze-and-excitation (SE) blocks at each scale.

SGDIR DiT (U-Net + DiT Hybrid). The second model extends the U-Net with a Diffusion Transformer operating on volumetric tokens. The U-Net stem mirrors the same channel widths $[16, 32, 64, 128, 256]$, also with attention gates on the skip connections and Squeeze and Excitation (SE) modules. The latent feature map is patchified (patch size 1) and projected into a 789-dimensional token space. An 8-layer DiT with 16 heads and an MLP ratio of 4 processes these tokens. The DiT architecture is implemented according to [37], with AdaLN time modulation (Fig. 9). Decoder features are time-conditioned using a 128-dimensional temporal embedding and have the same dimensions as the encoder in reverse order.

Optimization. Both architectures are trained end-to-end using NCC loss on paired images. We use the Adam optimizer with a learning rate of $1e{-}4$, batch size of 1, on an NVIDIA RTX 3090 (24 GB VRAM). For the brain MR datasets and AbdomenCTCT dataset we train the model for 200 epochs. For the ACDC and LungCT datasets we train the model for 150 epochs.

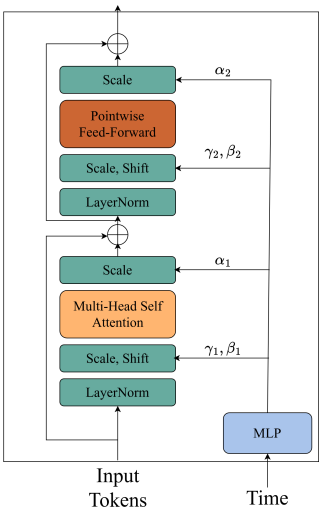

Figure 9. The architecture of DiT block [37] with AdaLN time modulation. The time is fed to a MLP module to produce six numbers used to scale and shift the outputs of LayerNorms, Multi-Head Self Attention, and FeedForward modules.

Type	Method	Dice↑	$\lvert J\rvert_{<0}\%$↓	HD95↓	SSIM↑	ASSD↓
Diffeomorphic	SyN	69.96 ± 2.09	0.0006 ± 0.0012	3.46 ± 1.03	84.35 ± 2.15	1.12 ± 0.44
	SYMNet	70.93 ± 1.93	0.0012 ± 0.0018	4.15 ± 0.88	83.15 ± 1.93	1.08 ± 0.48
	LapIRN	72.15 ± 1.76	0.0022 ± 0.0018	3.20 ± 0.59	86.17 ± 1.81	1.44 ± 0.67
	CycleMorph	74.67 ± 1.33	0.0224 ± 0.0110	3.34 ± 0.87	84.75 ± 1.88	0.87 ± 0.38
	GradICON	76.43 ± 1.28	0.0018 ± 0.0022	3.19 ± 0.55	85.70 ± 1.73	0.87 ± 0.30
	NePhi	75.34 ± 1.78	0.0009 ± 0.0016	3.55 ± 0.55	85.05 ± 1.83	0.93 ± 0.34
	TransMorph-diff	75.98 ± 2.01	0.0142 ± 0.0098	3.81 ± 0.89	87.75 ± 1.92	0.94 ± 0.45
	PULPo	73.91 ± 2.11	0.0066 ± 0.0043	3.30 ± 0.39	84.35 ± 2.18	0.90 ± 0.41
	R2Net	73.39 ± 1.77	0.0011 ± 0.0016	3.69 ± 0.66	83.92 ± 1.97	1.20 ± 0.51
	NODEO	75.68 ± 1.82	0.0011 ± 0.0015	3.46 ± 0.84	86.46 ± 1.38	0.94 ± 0.42
	SGDIR DiT ($\lambda = 10^5$)	**80.18 ± 1.09**	**0.0 ± 0.0**	2.76 ± 0.22	87.47 ± 1.44	**0.77 ± 0.18**
	SGDIR UNet ($\lambda = 10^5$)	80.04 ± 1.02	0.0002 ± 0.0001	**2.56 ± 0.31**	**88.37 ± 1.21**	0.80 ± 0.21
Deformbale	VoxelMorph	71.71 ± 1.86	0.0389 ± 0.0112	3.73 ± 1.02	85.44 ± 1.99	0.83 ± 0.31
	DiffuseMorph	71.35 ± 2.05	1.6042 ± 0.7212	5.48 ± 1.45	83.19 ± 2.76	1.45 ± 0.47
	DiffuseReg	72.13 ± 1.88	0.5279 ± 0.2834	5.04 ± 1.12	84.13 ± 2.45	1.28 ± 0.43
	TransMorph	77.72 ± 1.76	1.2574 ± 0.7574	3.71 ± 0.80	89.04 ± 1.76	0.88 ± 0.32
	TransMatch	76.80 ± 1.85	0.0406 ± 0.0120	3.15 ± 0.41	87.58 ± 1.74	0.84 ± 0.34
	HViT	80.67 ± 1.67	0.5933 ± 0.1028	**2.95 ± 0.46**	89.48 ± 1.92	**0.78 ± 0.28**
	CorrMLP	77.54 ± 1.68	0.3675 ± 0.2168	3.15 ± 0.39	86.50 ± 1.64	0.81 ± 0.31
	SACB-Net	78.15 ± 1.87	**0.0201 ± 0.0099**	3.19 ± 0.51	87.61 ± 1.89	0.84 ± 0.24
	SGDIR DiT ($\lambda = 10^4$)	**83.88 ± 1.77**	0.6036 ± 0.0893	3.17 ± 0.26	**91.61 ± 1.49**	0.80 ± 0.22
	SGDIR UNet ($\lambda = 10^4$)	83.73 ± 1.52	0.5803 ± 0.0988	3.20 ± 0.61	91.57 ± 1.79	0.81 ± 0.26

Table 9. The complete quantitative registration results on IXI dataset. The best performing method in each category is shown in **bold** and the best competing model is presented in blue.

10. Complete Quantitative Results

In this section, we present the complete quantitative results for all datasets evaluated in this work. The full ACDC results are already reported in the main paper and are not repeated here. For all tables, the best-performing metric is highlighted in **bold**, while the second-best performing model is marked in blue. If the best performing model is SGDIR, we chose the second-best performing model from the comparison models to better empha-

size the improvement yielded by SGDIR. Each table includes results for both *diffeomorphic* and *deformable* registration models, evaluated across Dice score, Jacobian determinant percentage of folding voxels ($|J|_{<0}\%$), 95th percentile Hausdorff Distance (HD95), Structural Similarity Index (SSIM) and Average Symmetric Surface Distance (ASSD). Tab. 8 presents the complete results over the OASIS dataset both with atlas-based and inter-subject registration settings. Tab. 9 shows the performance comparisons on the IXI dataset. Similarly, Tab. 10, 11, 12 illustrate the results for the Mindboggle101, LPBA40, and CANDI datasets, respectively. The comprehensive results for the AbdomenCTCT dataset is reported in Tab. 13. Tab. 14 demonstrates the performance comparisons over the LungCT dataset for which the main evaluation metric is Target to Registration Error (TRE).

Type	Method	Dice↑	$\|J\|_{<0}\%$↓	HD95↓	SSIM↑	ASSD↓
Diffeomorphic	SyN	67.91 ± 1.85	0.0018 ± 0.0008	7.79 ± 2.41	84.78 ± 1.92	1.69 ± 0.48
	SYMNet	70.09 ± 1.67	0.0015 ± 0.0008	7.56 ± 1.03	86.30 ± 1.55	1.58 ± 0.47
	LapIRN	67.37 ± 1.94	0.0217 ± 0.0125	8.81 ± 1.21	87.10 ± 1.73	1.67 ± 0.63
	CycleMorph	70.64 ± 1.48	0.1282 ± 0.0873	7.75 ± 0.98	86.83 ± 1.64	1.56 ± 0.44
	GradICON	70.78 ± 1.52	0.0117 ± 0.0043	7.59 ± 1.12	87.51 ± 1.39	1.57 ± 0.55
	TransMorph-diff	69.45 ± 1.66	0.0364 ± 0.0157	7.89 ± 1.05	89.14 ± 1.88	**1.47 ± 0.30**
	NePhi	67.97 ± 1.91	0.0009 ± 0.0005	8.20 ± 1.23	84.47 ± 1.76	1.74 ± 0.52
	PULPo	68.69 ± 1.74	0.0287 ± 0.0104	7.92 ± 1.07	84.23 ± 1.69	1.61 ± 0.42
	R2Net	68.34 ± 1.69	0.0086 ± 0.0031	8.97 ± 1.44	83.74 ± 1.82	1.84 ± 0.64
	NODEO	70.18 ± 1.58	0.0023 ± 0.0011	7.70 ± 0.96	87.94 ± 1.45	1.66 ± 0.36
	SGDIR DiT ($\lambda = 10^5$)	**71.58 ± 1.12**	**0.0004 ± 0.0003**	7.60 ± 1.98	**89.52 ± 1.33**	1.55 ± 0.23
	SGDIR UNet ($\lambda = 10^5$)	71.20 ± 1.45	0.0005 ± 0.0002	**7.53 ± 1.83**	89.24 ± 1.72	1.59 ± 0.21
Deformable	VoxelMorph	69.23 ± 1.82	0.2562 ± 0.0944	8.12 ± 1.27	84.15 ± 1.71	1.55 ± 0.53
	DiffuseMorph	68.22 ± 1.77	1.7413 ± 0.6521	8.43 ± 1.30	84.62 ± 1.68	1.65 ± 0.47
	DiffuseReg	71.88 ± 1.54	1.0823 ± 0.4882	7.84 ± 1.03	89.20 ± 1.57	1.56 ± 0.38
	TransMorph	71.46 ± 1.63	1.4883 ± 0.5328	**7.27 ± 0.92**	93.71 ± 1.42	1.53 ± 0.35
	TransMatch	71.78 ± 1.49	0.2483 ± 0.1020	7.63 ± 1.08	89.73 ± 1.53	1.53 ± 0.49
	HViT	71.80 ± 1.72	1.4051 ± 0.6153	7.56 ± 1.11	92.04 ± 1.61	1.51 ± 0.46
	CorrMLP	71.92 ± 1.68	**0.1895 ± 0.0782**	7.65 ± 1.06	89.48 ± 1.49	1.51 ± 0.51
	SACB-Net	72.75 ± 1.57	0.3542 ± 0.1433	7.46 ± 0.97	91.83 ± 1.55	1.52 ± 0.37
	SGDIR DiT ($\lambda = 10^4$)	**73.59 ± 1.38**	0.6512 ± 0.0813	7.58 ± 0.94	**93.98 ± 1.49**	**1.51 ± 0.32**
	SGDIR UNet ($\lambda = 10^4$)	73.18 ± 1.63	0.6143 ± 0.2441	7.56 ± 0.93	93.92 ± 2.80	1.52 ± 0.45

Table 10. The complete quantitative registration results on Mindboggle101 dataset. The best performing method in each category is shown in **bold** and the best competitor is presented in blue.

Type	Method	Dice↑	$\|J\|_{<0}\%$↓	HD95↓	SSIM↑	ASSD↓
Diffeomorphic	SyN	68.59 ± 1.34	0.0022 ± 0.0016	6.81 ± 0.92	90.53 ± 1.21	1.80 ± 0.77
	SYMNet	72.34 ± 1.28	0.0089 ± 0.0041	6.64 ± 0.87	88.46 ± 1.33	1.83 ± 0.82
	LapIRN	72.73 ± 1.22	0.0053 ± 0.0027	**4.21 ± 0.65**	88.12 ± 1.29	1.98 ± 0.82
	CycleMorph	71.90 ± 1.41	0.0064 ± 0.0032	6.76 ± 0.94	87.43 ± 1.37	1.88 ± 0.71
	GradICON	72.21 ± 1.26	0.0011 ± 0.0006	6.89 ± 0.89	91.02 ± 1.18	1.81 ± 0.62
	NePhi	72.39 ± 1.33	**0.0 ± 0.0**	6.75 ± 0.91	85.28 ± 1.42	1.84 ± 0.56
	TransMorph-diff	70.98 ± 1.52	0.0086 ± 0.0044	7.22 ± 1.03	**94.58 ± 1.09**	1.87 ± 0.64
	PULPo	74.82 ± 1.11	0.0009 ± 0.0005	6.22 ± 0.78	85.32 ± 1.36	1.68 ± 0.33
	R2Net	70.71 ± 1.49	0.0017 ± 0.0009	7.13 ± 1.01	84.40 ± 1.41	2.16 ± 1.01
	NODEO	72.73 ± 1.25	0.0008 ± 0.0004	7.25 ± 0.97	92.07 ± 1.17	1.88 ± 0.98
	SGDIR DiT ($\lambda = 10^5$)	**77.13 ± 1.10**	**0.0 ± 0.0**	5.72 ± 0.71	93.68 ± 1.14	**1.47 ± 0.25**
	SGDIR UNet($\lambda = 10^5$)	76.83 ± 1.15	**0.0 ± 0.0**	5.80 ± 0.74	93.11 ± 1.12	1.49 ± 0.37
Deformable	VoxelMorph	73.05 ± 1.39	0.0587 ± 0.0315	6.73 ± 0.92	92.11 ± 1.23	1.92 ± 0.74
	DiffuseMorph	73.11 ± 1.42	0.3610 ± 0.1429	7.21 ± 1.06	84.39 ± 1.48	2.08 ± 0.92
	DiffuseReg	72.98 ± 1.46	0.0723 ± 0.0289	7.27 ± 1.04	84.04 ± 1.51	2.08 ± 0.96
	TransMorph	74.42 ± 1.33	0.5275 ± 0.1835	6.88 ± 0.85	94.75 ± 1.12	1.77 ± 0.42
	TransMatch	74.13 ± 1.31	0.0251 ± 0.0116	6.53 ± 0.83	92.54 ± 1.25	1.68 ± 0.35
	HViT	76.57 ± 1.18	0.4428 ± 0.1655	6.74 ± 0.89	94.10 ± 1.09	1.71 ± 0.28
	CorrMLP	75.72 ± 1.27	0.0145 ± 0.0068	6.55 ± 0.87	90.39 ± 1.32	1.73 ± 0.43
	SACB-Net	73.91 ± 1.35	**0.0099 ± 0.0041**	6.70 ± 0.93	93.48 ± 1.20	1.79 ± 0.51
	SGDIR DiT ($\lambda = 10^4$)	**78.93 ± 1.02**	0.1821 ± 0.0731	**5.75 ± 0.69**	95.09 ± 1.05	**1.47 ± 0.28**
	SGDIR UNet($\lambda = 10^4$)	77.09 ± 1.08	0.1730 ± 0.0691	5.82 ± 0.72	**95.33 ± 1.03**	1.47 ± 0.39

Table 11. The complete quantitative registration results on LPBA40 dataset. The best performing method in each category is shown in **bold** and the best competitor is presented in blue.

Type	Method	Dice↑	$\|J\|_{<0}\%$↓	HD95↓	SSIM↑	ASSD↓
Diffeomorphic	SyN	75.43 ± 1.52	0.0005 ± 0.0003	2.41 ± 0.31	88.53 ± 1.46	0.75 ± 0.14
	SYMNet	80.46 ± 1.28	0.0011 ± 0.0007	2.31 ± 0.28	91.46 ± 1.15	0.78 ± 0.22
	LapIRN	79.73 ± 1.33	0.0042 ± 0.0018	2.34 ± 0.21	90.21 ± 1.27	0.87 ± 0.19
	CycleMorph	81.37 ± 1.25	0.0027 ± 0.0013	2.23 ± 0.29	92.43 ± 1.20	0.76 ± 0.11
	GradICON	83.39 ± 1.09	0.0015 ± 0.0008	2.14 ± 0.27	96.35 ± 1.04	**0.72 ± 0.08**
	NePhi	79.15 ± 1.46	0.0009 ± 0.0005	2.36 ± 0.30	90.07 ± 1.33	0.83 ± 0.25
	TransMorph-diff	83.20 ± 1.14	0.0026 ± 0.0012	**2.06 ± 0.26**	96.36 ± 1.09	0.79 ± 0.17
	PULPo	78.38 ± 1.57	0.0136 ± 0.0047	2.89 ± 0.33	89.59 ± 1.41	0.89 ± 0.14
	R2Net	78.79 ± 1.49	0.0032 ± 0.0016	2.78 ± 0.32	88.32 ± 1.39	0.87 ± 0.13
	NODEO	80.60 ± 1.26	0.0012 ± 0.0007	2.46 ± 0.29	92.56 ± 1.18	0.85 ± 0.10
	SGDIR DiT ($\lambda = 10^5$)	**84.50 ± 1.08**	0.0003 ± 0.0002	2.10 ± 0.20	**97.22 ± 1.02**	0.74 ± 0.09
	SGDIR UNet ($\lambda = 10^5$)	84.36 ± 1.11	**0.0001 ± 0.0001**	2.12 ± 0.21	97.12 ± 1.01	0.75 ± 0.11
Deformable	VoxelMorph	80.28 ± 1.34	0.0192 ± 0.0073	2.28 ± 0.27	91.84 ± 1.23	0.76 ± 0.15
	DiffuseMorph	77.22 ± 1.51	0.1277 ± 0.0412	2.53 ± 0.30	88.68 ± 1.44	0.89 ± 0.17
	DiffuseReg	79.31 ± 1.38	0.0322 ± 0.0125	2.31 ± 0.26	92.90 ± 1.19	0.86 ± 0.21
	TransMorph	84.48 ± 1.10	0.1059 ± 0.0388	2.13 ± 0.22	97.44 ± 1.03	0.75 ± 0.14
	TransMatch	82.60 ± 1.18	0.0160 ± 0.0061	2.28 ± 0.25	95.53 ± 1.09	0.81 ± 0.16
	HViT	84.67 ± 1.07	0.1368 ± 0.0435	2.09 ± 0.21	97.10 ± 0.99	0.77 ± 0.12
	CorrMLP	81.96 ± 1.25	0.0098 ± 0.0043	2.15 ± 0.26	95.07 ± 1.16	0.74 ± 0.12
	SACB-Net	82.75 ± 1.22	**0.0047 ± 0.0021**	2.19 ± 0.27	95.21 ± 1.12	0.79 ± 0.11
	SGDIR DiT ($\lambda = 10^4$)	**85.51 ± 0.96**	0.0473 ± 0.0157	2.10 ± 0.19	**97.96 ± 0.94**	**0.73 ± 0.07**
	SGDIR UNet ($\lambda = 10^4$)	85.28 ± 0.98	0.0331 ± 0.0132	**2.08 ± 0.20**	97.78 ± 0.93	0.73 ± 0.09

Table 12. The complete quantitative registration results on CANDI dataset. The best performing method in each category is shown in **bold** and the best competitor is presented in blue.

Type	Method	Dice↑	$\|J\|_{<0}\%$↓	HD95↓	SSIM↑	ASSD↓
Diffeomorphic	SyN	46.07 ± 2.31	0.0032 ± 0.0014	11.51 ± 1.12	56.06 ± 2.85	3.53 ± 0.71
	SYMNet	30.06 ± 2.78	0.0011 ± 0.0006	13.15 ± 1.37	40.74 ± 3.14	4.89 ± 0.98
	LapIRN	38.82 ± 2.94	0.0008 ± 0.0003	13.12 ± 0.92	52.57 ± 3.45	4.93 ± 1.23
	CycleMorph	42.56 ± 2.65	0.0053 ± 0.0025	13.04 ± 1.28	52.68 ± 3.02	4.77 ± 1.08
	GradICON	41.02 ± 2.81	0.0009 ± 0.0004	13.21 ± 1.33	51.86 ± 3.27	4.86 ± 1.19
	NePhi	45.32 ± 2.24	0.0008 ± 0.0003	12.48 ± 1.19	66.34 ± 2.54	3.90 ± 0.68
	TransMorph-diff	41.41 ± 2.72	0.0034 ± 0.0017	12.39 ± 1.25	62.61 ± 2.98	4.71 ± 0.76
	PULPo	44.43 ± 2.13	0.0766 ± 0.0214	10.93 ± 1.04	53.80 ± 2.67	3.51 ± 0.62
	R2Net	39.99 ± 2.89	0.0010 ± 0.0005	13.05 ± 1.32	51.52 ± 3.36	4.93 ± 0.91
	NODEO	39.37 ± 2.96	0.0007 ± 0.0004	13.16 ± 1.38	52.45 ± 3.42	3.74 ± 0.56
	SGDIR DiT ($\lambda = 10^5$)	52.23 ± 2.04	0.0001 ± 0.0001	10.27 ± 0.94	66.08 ± 2.27	**2.89 ± 0.38**
	SGDIR UNet ($\lambda = 10^5$)	**53.64 ± 2.01**	**0.0 ± 0.0**	**10.07 ± 0.85**	**67.76 ± 2.19**	2.97 ± 0.45
Deformable	VoxelMorph	46.92 ± 2.27	0.0203 ± 0.0084	11.94 ± 1.21	56.91 ± 2.68	4.75 ± 0.99
	DiffuseMorph	31.50 ± 2.81	0.0292 ± 0.0113	12.88 ± 1.34	43.02 ± 3.09	4.85 ± 1.33
	DiffuseReg	35.91 ± 2.56	2.8106 ± 0.7348	12.55 ± 1.25	46.87 ± 2.88	4.32 ± 1.02
	TransMorph	46.66 ± 2.48	3.1310 ± 0.8113	12.61 ± 1.28	**71.84 ± 2.54**	4.32 ± 1.12
	TransMatch	43.64 ± 2.35	2.6378 ± 0.6922	12.50 ± 1.22	71.47 ± 2.33	3.87 ± 0.94
	HViT	47.66 ± 2.41	0.9601 ± 0.4175	10.11 ± 1.10	66.90 ± 2.74	3.88 ± 0.83
	CorrMLP	50.28 ± 2.18	**0.1656 ± 0.0643**	11.47 ± 1.16	62.66 ± 2.59	3.00 ± 0.53
	SACB-Net	53.38 ± 2.63	0.9348 ± 0.1135	13.09 ± 1.41	66.73 ± 2.96	3.67 ± 0.49
	SGDIR DiT ($\lambda = 10^4$)	54.15 ± 1.91	0.6244 ± 0.2184	10.19 ± 0.89	68.10 ± 2.11	**2.45 ± 0.27**
	SGDIR UNet ($\lambda = 10^4$)	**56.57 ± 1.83**	0.2683 ± 0.0945	**9.19 ± 0.78**	71.66 ± 1.96	2.45 ± 0.31

Table 13. The complete quantitative registration results on AbdomenCTCT dataset. The best performing method in each category is shown in **bold** and the best competitor is presented in blue.

11. Visual Results

In this section, we present visual results highlighting the advantages of SGDIR. Throughout all figures, the suffix *-diff* denotes the diffeomorphic variant of SGDIR, trained with $\lambda = 10^5$, while the unsuffixed SGDIR refers to the deformable variant trained with $\lambda = 10^4$.

Fig. 15 illustrates the qualitative performance of the SGDIR variants compared to the best competing methods on the OASIS dataset, sliced along the y-axis. In this figure, the moving and fixed images are shown in the top-left corner. The first row displays the warped moving images produced by SGDIR, and the second row presents the corresponding deformation grids. For other methods, we follow the same visualization scheme, stacking each warped image above its deformation grid for consistency. Similarly, Fig. 16 presents qualitative results on the IXI dataset, sliced along the z-axis.

Type	Method	TRE↓	$\lvert J\rvert_{<0}\%$↓	SSIM↑
Diffeomorphic	SyN	3.59 ± 0.22	0.0062 ± 0.0018	44.56 ± 2.91
	SYMNet	3.31 ± 0.20	0.0037 ± 0.0012	48.58 ± 3.12
	LapIRN	2.93 ± 0.18	0.0043 ± 0.0014	54.15 ± 3.05
	CycleMorph	3.04 ± 0.19	0.7180 ± 0.211	56.99 ± 3.44
	GradICON	2.64 ± 0.17	0.0009 ± 0.0004	54.06 ± 2.86
	NePhi	3.12 ± 0.21	0.0093 ± 0.0027	56.66 ± 3.33
	TransMorph-diff	2.89 ± 0.18	0.0042 ± 0.0015	66.85 ± 2.95
	R2Net	2.99 ± 0.19	0.0211 ± 0.0068	68.06 ± 2.81
	NODEO	2.87 ± 0.18	0.0038 ± 0.0013	**70.80 ± 2.74**
	SGDIR DiT($\lambda = 10^5$)	2.77 ± 0.77	**0.0 ± 0.0**	66.06 ± 2.46
	SGDIR UNet ($\lambda = 10^5$)	**2.37 ± 0.15**	**0.0 ± 0.0**	66.55 ± 2.61
Deformable	VoxelMorph	3.19 ± 0.20	0.0221 ± 0.0072	51.87 ± 3.08
	DiffuseMorph	3.38 ± 0.21	0.0794 ± 0.0195	62.98 ± 2.92
	DiffuseReg	3.26 ± 0.20	0.7765 ± 0.2337	65.75 ± 2.71
	TransMorph	2.85 ± 0.17	0.4767 ± 0.1546	68.82 ± 2.65
	TransMatch	3.02 ± 0.19	0.1343 ± 0.0482	67.47 ± 2.89
	HViT	3.22 ± 0.20	0.1791 ± 0.0554	46.72 ± 3.41
	CorrMLP	2.48 ± 0.16	0.2673 ± 0.0711	55.57 ± 3.08
	SACB-Net	3.01 ± 0.19	0.9824 ± 0.2761	70.06 ± 2.77
	SGDIR DiT($\lambda = 10^4$)	2.57 ± 0.65	**0.0543 ± 0.0187**	70.02 ± 2.66
	SGDIR UNet ($\lambda = 10^4$)	**2.23 ± 0.14**	0.0615 ± 0.0223	**70.48 ± 2.44**

Table 14. The complete quantitative registration results on LungCT dataset. The best performing method in each category is shown in **bold** and the best competitor is presented in blue.

Among all comparison methods, DiffuseMorph [28] and DiffuseReg [58] are capable of generating time-resolved deformations. Fig. 10 visualizes the temporal evolution of their warped outputs. Following the standard inference protocol of DiffuseMorph, we use $\Delta t = 1/7$ time steps. Notably, SGDIR supports deformation generation at arbitrary temporal resolutions, owing to its continuous-time training formulation, whereas DiffuseMorph and DiffuseReg are constrained to a fixed number of time steps.

Finally, to analyze how each anatomical region contributes to the overall Dice score, Fig. 11, 12, 13, 14 present the distribution of regional Dice scores over the OASIS, IXI, CANDI, and ACDC datasets, respectively. For clarity, on brain MRI datasets, left and right anatomical regions are merged and averaged into a single score per region.

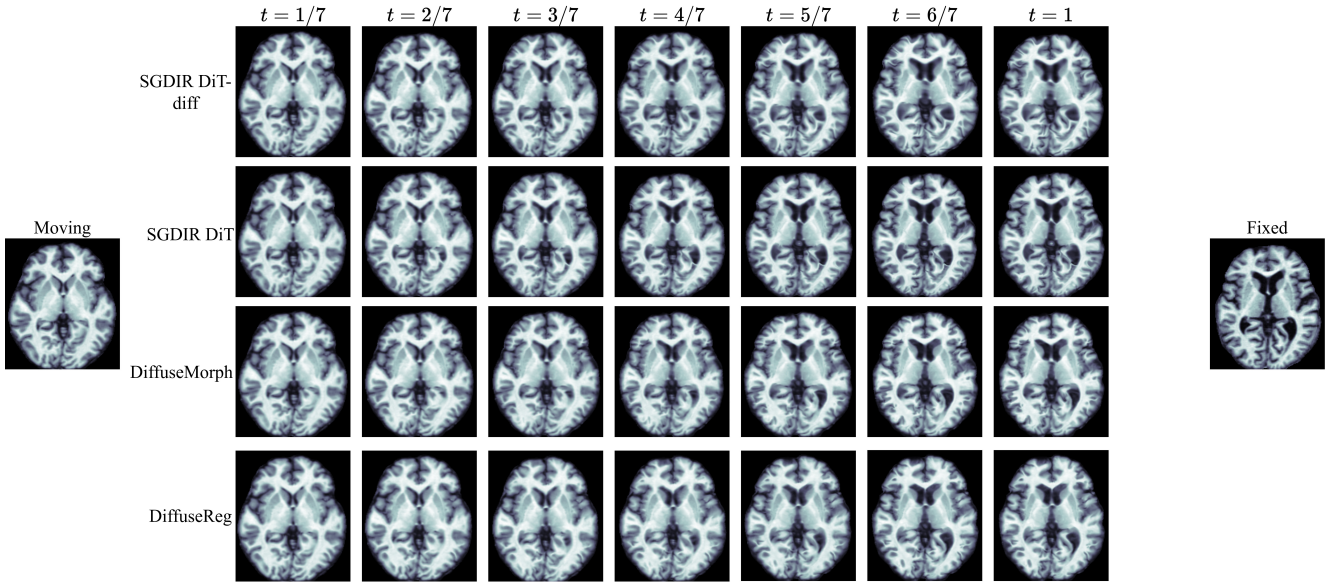

Figure 10. Visual results and comparisons over the IXI dataset. The deformation grid predicted by each method is attached below.

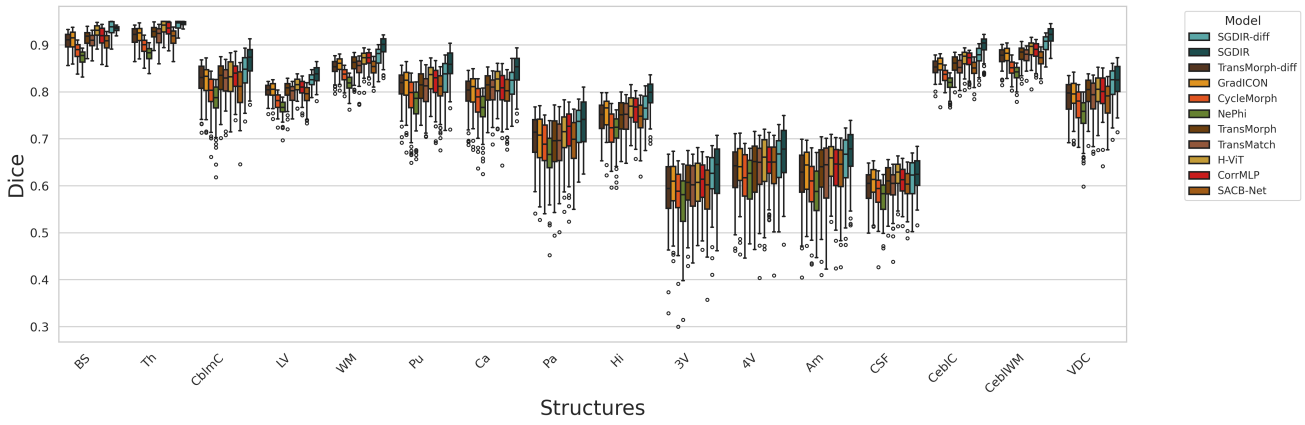

Figure 11. The boxplot of average Dice score across different anatomical regions over the OASIS dataset.

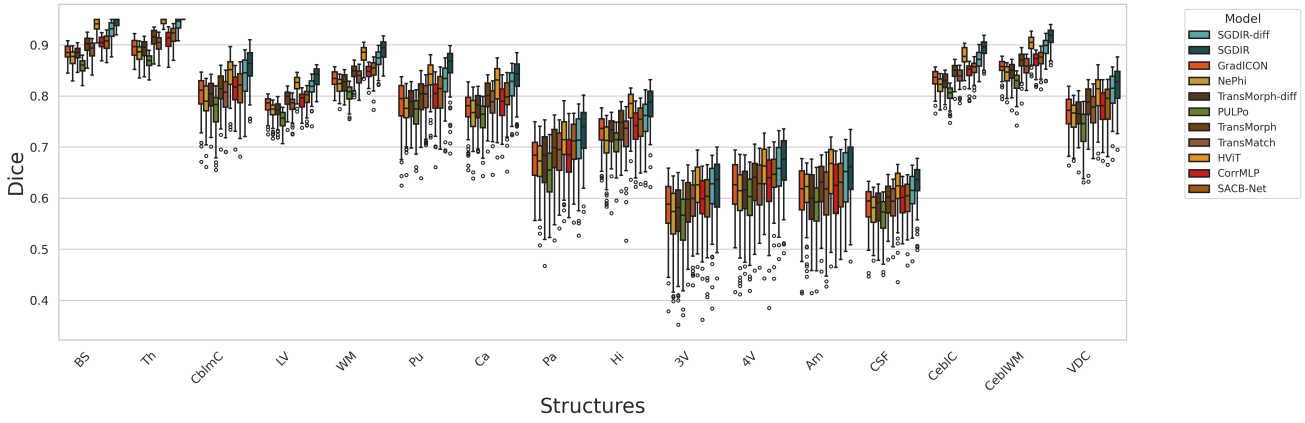

Figure 12. The boxplot of average Dice score across different anatomical regions over the IXI dataset.

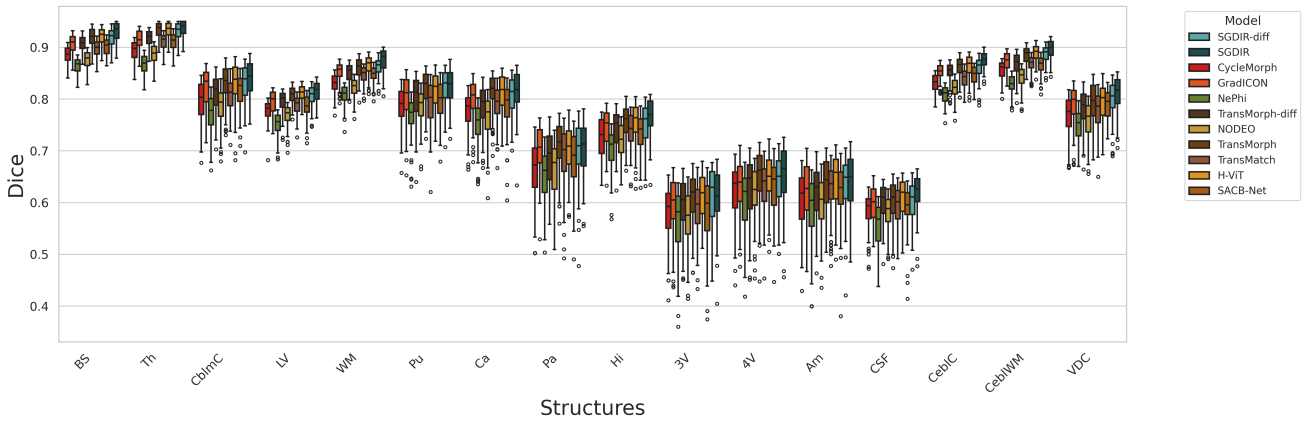

Figure 13. The boxplot of average Dice score across different anatomical regions over the CANDI dataset.

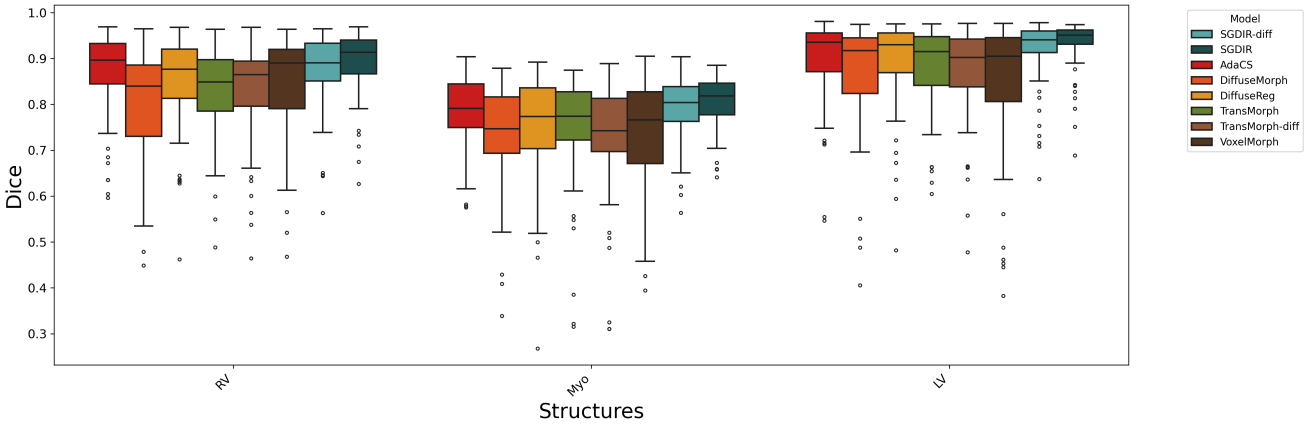

Figure 14. The boxplot of average Dice score across different anatomical regions over the ACDC dataset.

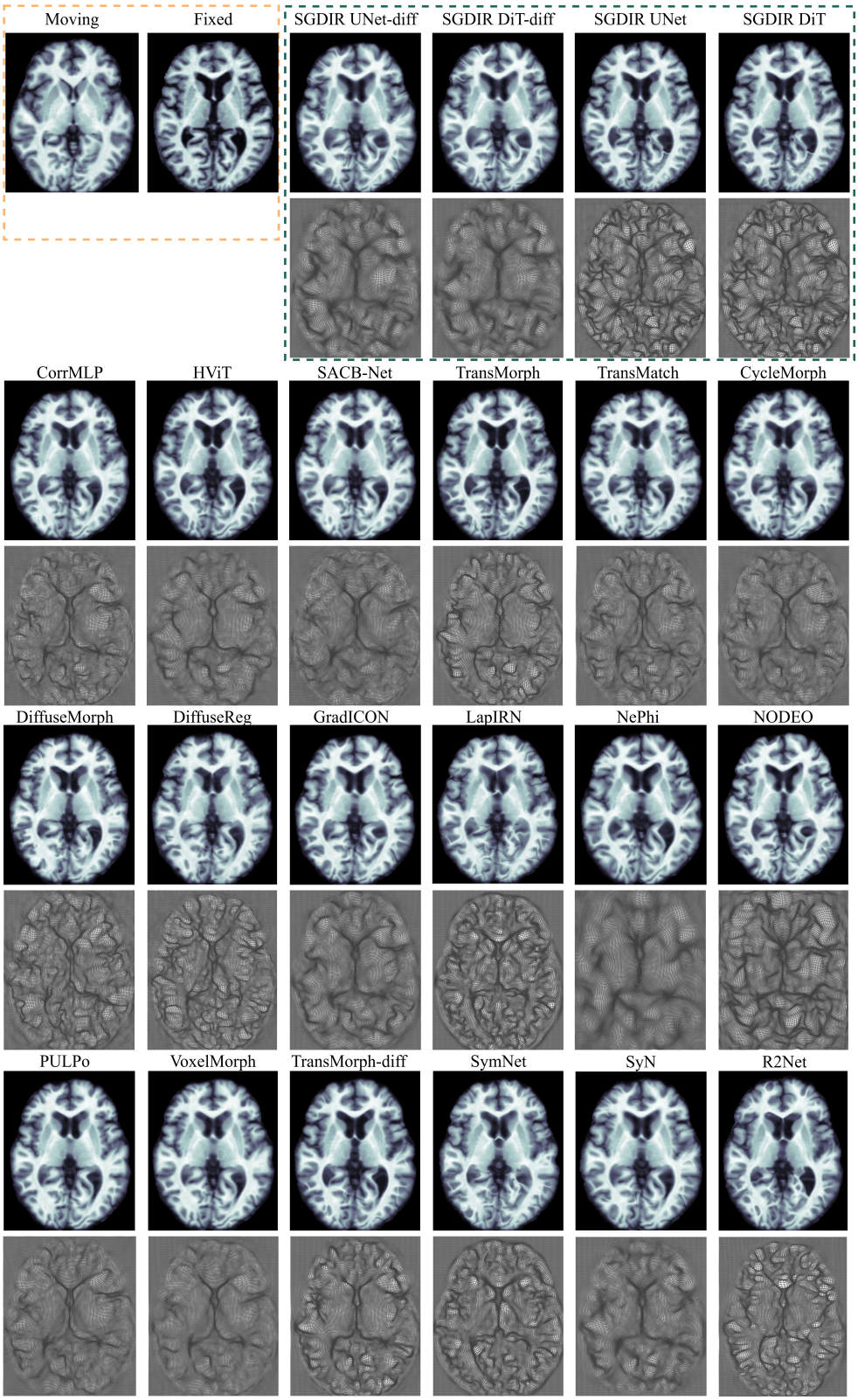

Figure 15. Visual results and comparisons over the OASIS dataset. The deformation grid predicted by each method is attached below its warped image.

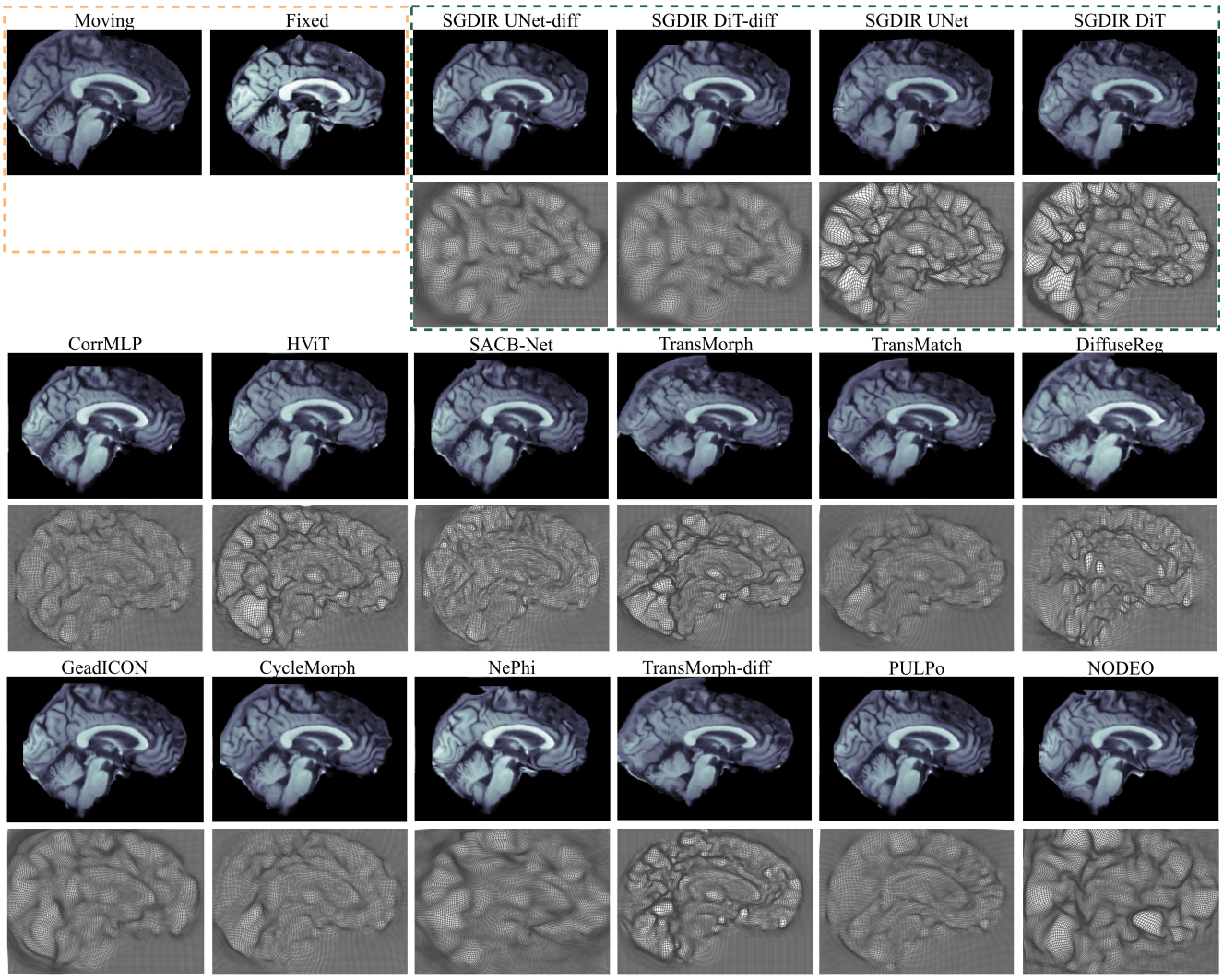

Figure 16. Visual results and comparisons over the IXI dataset. The deformation grid predicted by each method is attached below its warped image.