

NR-Align: Non-Rigid Alignment for Non-simultaneous Two-View 3D Coronary Reconstruction in Complex Cardiac Interventions

Pengbo Wang¹, Zhu Meng^{1*}, Qiwan Yu¹, Fei Su¹, and Chunxia Di²

¹ Beijing University of Posts and Telecommunications, Beijing, China
bamboo@bupt.edu.cn

² Department of Cardiovascular Medicine, The Eighth Medical Center of Chinese PLA General Hospital, Beijing, China

Abstract. Accurate 3D coronary artery reconstruction from two non-simultaneous Digital Subtraction Angiography (DSA) projections is important for guidance in complex cardiac interventions, yet remains unreliable in practice. Cardiac pulsation and respiration introduce non-rigid motion between views, breaking rigid-geometry assumptions and causing cross-view misregistration, which often leads to vessel discontinuities, ghosting artifacts, and topology errors in reconstructed volumes. We propose NR-Align, a motion-aware non-rigid alignment module that reduces cross-view mismatch on dual-view backprojected volumes before downstream 3D reconstruction. To enable controlled training and analysis, we further build a physiology-driven simulation pipeline that synthesizes realistic cardiopulmonary deformations and provides paired 3D vascular ground truth from Coronary Computed Tomography Angiography (CCTA) volumes. Experiments on a synthesized dual-async benchmark constructed from the ImageCAS dataset and on unseen real clinical DSA cases show that NR-Align improves 3D Dice and reprojection agreement, and consistently reduces motion-induced artifacts, yielding more coherent coronary reconstructions. Our code is available at <https://github.com/wpbs/NR-Align>.

Keywords: Coronary artery reconstruction · Asynchronous dual-view DSA · Non-rigid motion compensation · Physiology-aware deformation synthesis.

1 Introduction

Accurate 3D coronary artery reconstruction is critical for guidance in complex cardiac interventions. In clinical Digital Subtraction Angiography (DSA), radiation dose and contrast-agent constraints typically limit acquisition to only a few 2D projections. More importantly, in single-plane systems, different views are predominantly acquired non-simultaneously. Consequently, cardiac pulsation

* Corresponding author.

and respiration introduce non-rigid motion between views. This violates rigid-geometry assumptions and yields severe cross-view structural mismatch, often manifesting as vessel discontinuities, ghosting artifacts, and topology errors in reconstructed volumes.

Recent work has demonstrated that two views can, in principle, contain sufficient information for vascular reconstruction [1]. Beyond two-view feasibility, related efforts also consider non-simultaneous angiography settings via integrated calibration and centerline reconstruction [11] or 3D/2D structure matching under imperfect segmentations [10, 9]. These studies highlight that handling non-simultaneity and imperfect 2D cues is already challenging, even before accounting for cardiopulmonary non-rigid deformation. However, most sparse-view coronary reconstruction approaches are developed and validated mainly on synthetic data generated from Coronary Computed Tomography Angiography (CCTA) volumes (e.g., ImageCAS [14]), including methods based on 3D Gaussian representations or implicit fields [2, 4, 5]. While these methods advance sparse-view reconstruction, they do not directly address the non-rigid cross-view mismatch induced by asynchronous clinical acquisition, and motion is often simplified as rigid or mild perturbations. Even improved angiography simulation efforts still face challenges in faithfully modeling clinical variability and motion [6]. In practice, robust 2D vessel masks and correspondence cues are also non-trivial in coronary angiography [8, 13, 7].

Among existing methods, DeepCA is most closely aligned with real interventional scenarios by targeting reconstruction from non-simultaneous projections and considering motion perturbations [3]. Nevertheless, the motion modeling in its training pipeline largely relies on rigid approximations, which cannot authentically capture cardiopulmonary non-rigid deformation in practice. More broadly, learning-based deformable registration has been extensively studied in medical imaging [16, 19, 20], suggesting that explicitly modeling non-rigid alignment with topology-preserving deformation is a principled direction for improving robustness under motion. Therefore, constructing physiologically consistent non-rigid training data and explicitly compensating cross-view structural mismatch prior to reconstruction remain urgent for robust clinical deployment.

To address these limitations, this paper improves both the data synthesis and the model design:

1. A physiologically consistent non-rigid deformation data generation pipeline that simulates combined cardiopulmonary motion is constructed, significantly bridging the gap between training assumptions and real clinical asynchronous deformations.
2. A reconstruction framework tailored for asynchronous dual-view DSA, featuring an explicit non-rigid mismatch compensation module (NR-Align), is proposed to directly address cross-view structural inconsistency prior to 3D reconstruction.
3. Extensive experiments on synthetic and real clinical datasets demonstrate that the proposed method significantly enhances the structural continuity

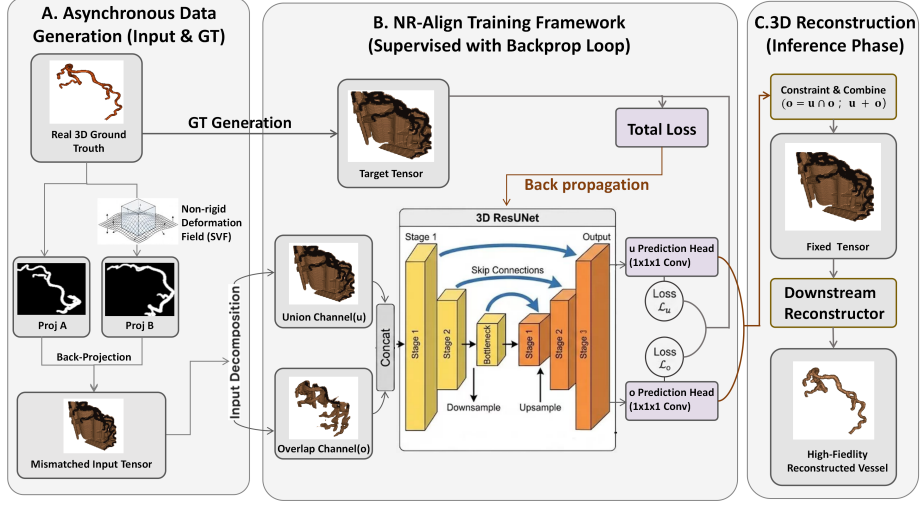


Fig. 1. Overview of the proposed framework. (A) We synthesize physiology-consistent non-rigid motion for supervision. (B) NR-Align is trained to compensate for cross-view mismatches. (C) The corrected tensors are fed into a downstream model for high-fidelity 3D reconstruction.

and topological rationality of the final reconstruction compared to existing state-of-the-art approaches.

2 Methodology

This section presents a framework for handling non-rigid vascular deformation in asynchronous dual-view DSA reconstruction [3]. The overall pipeline, illustrated in Fig. 1, is decomposed into three main components. First, a physiology-driven non-rigid deformation synthesis process (**Fig. 1A**) generates supervised training pairs with realistic cardiopulmonary motion [6, 12]. Second, the NR-Align training framework (**Fig. 1B**) is proposed to rectify cross-view spatial mismatches in the input tensor. Finally, during the inference phase (**Fig. 1C**), the aligned outputs are combined and fed into a downstream 3D reconstructor. This design improves cross-view structural consistency and helps preserve vessel topology in the reconstructed coronary arteries.

2.1 Physiology-aware Dual-Async Non-rigid Deformation Synthesis

A physiology-aware non-rigid deformation module is used to simulate cross-view mismatch under asynchronous dual-view DSA acquisition [6]. The discrepancy between two views is modeled by a shared time offset and two motion components: respiration-dominant translation and cardiac-dominant torsion. A station-

ary velocity field (SVF) formulation is adopted to ensure smooth and integrable deformations [19, 20, 16].

A shared physical time offset Δt (ms) is sampled and converted to respiratory and cardiac phase offsets:

$$\Delta\phi_r = \frac{\Delta t}{T_r}, \quad \Delta\phi_c = \frac{\Delta t}{T_c}, \quad (1)$$

where T_r and T_c denote respiration and cardiac periods. View-specific phases are then assigned (e.g., symmetric or same-side sampling) to mimic temporal misalignment.

The deformation is localized around the coronary region using a Gaussian window. Two basis velocity fields are defined within this region: a respiratory translation field v_{resp} and a cardiac torsion field v_{card} (torsion-like motion around a local axis, proportional to $n \times (x - c)$). For each view i , the SVF is constructed as

$$v^{(i)} = a_r^{(i)} v_{\text{resp}} + a_c^{(i)} v_{\text{card}}, \quad (2)$$

with phase-dependent coefficients $a_r^{(i)} = \sin(2\pi t_r^{(i)})$ and $a_c^{(i)} = \sin(2\pi t_c^{(i)} + \frac{\pi}{2})$. The view-specific transformation is obtained via the exponential map [19, 20]:

$$\phi^{(i)} = \exp\left(v^{(i)}\right). \quad (3)$$

The deformation is applied in signed distance field (SDF) space and converted back to a binary vessel volume by thresholding. Other steps such as sample gating and backprojection construction are implementation details and omitted for brevity.

2.2 NR-Align Network

Given a dual-view backprojected volume, the objective is to compensate cross-view mismatch before downstream 3D reconstruction. The input is represented as a two-channel 3D volume $X = [X_u, X_o]$, where X_u is the union occupancy map (voxels supported by at least one view) and X_o is the overlap occupancy map (voxels jointly supported by both views) [3]. This representation preserves both global vessel support and cross-view consistency cues.

Network Design A 3D encoder-decoder architecture is adopted to predict compensated structures from $X = [X_u, X_o]$ [17, 18]. The network uses a dual-head design: a union head predicting the corrected global vessel support, and an overlap head predicting the cross-view consistent region. This explicitly separates “where vessels may exist” from “where the two views agree”, which is beneficial under asynchronous non-rigid motion.

Training Objective For each head, a combined Binary Cross-Entropy (BCE) and Dice loss is used:

$$\mathcal{L}_{\text{seg}}(p, y) = \mathcal{L}_{\text{BCE}}(p, y) + \mathcal{L}_{\text{Dice}}(p, y). \quad (4)$$

Let $\mathcal{L}_u$ and $\mathcal{L}_o$ denote the segmentation losses of the union and overlap heads, respectively. To enforce structural consistency, a constraint encourages the overlap prediction to be a subset of the union prediction:

$$\mathcal{L}_{\text{cons}} = \text{mean}[\text{ReLU}(\sigma(O) - \sigma(U))], \quad (5)$$

where U and O denote the logits of the union and overlap heads, and $\sigma(\cdot)$ is the sigmoid function. The final objective is:

$$\mathcal{L} = \mathcal{L}_u + \lambda_o \mathcal{L}_o + \lambda_c \mathcal{L}_{\text{cons}}, \quad (6)$$

where λ_o and λ_c are weighting coefficients.

3 Experiments and Results

3.1 Datasets

(1) **Synthesized Dataset:** Synthesized training and evaluation samples are generated from the ImageCAS dataset [14] using the pipeline described in Sec. 2.1. Right Coronary Artery (RCA) extraction follows the same preprocessing and RCA selection protocol as DeepCA [3], which is used as the instantiation in the experiments. (2) **Clinical Dataset:** A clinical dataset comprising 10 dual-view asynchronous coronary angiography sequences was collected from the Department of Cardiovascular Medicine, The Eighth Medical Center of Chinese PLA General Hospital. All clinical data were de-identified before analysis. This dataset is used exclusively for testing to evaluate the sim-to-clinical generalization of NR-Align.

3.2 Implementation Details

A patient-level split of 8:1:1 is used for training, validation, and testing to prevent subject overlap across subsets. For each patient in the synthesized set, $N = 9$ dual-async motion configurations are generated by combining three deformation bins ($amp \in \{5, 10, 15\}$ mm) and three phase-pattern groups ($k \in \{0, 1, 2\}$). The deformation magnitude is controlled by amp , defined as the 95th-percentile (P95) displacement magnitude within the coronary region-of-interest in the simulator, which avoids sensitivity to rare extreme warps.

The model is implemented in PyTorch and trained on a single NVIDIA GeForce RTX 5090 GPU. Unless otherwise stated, input volumes are at 128^3 resolution, with two channels formed by the union and overlap occupancy maps. Projection and back-projection follow the same geometry setting as DeepCA [3, 15]. Optimization uses AdamW with a learning rate of 2×10^{-4} and weight decay of 10^{-4} . We employ a batch size of 1 with gradient accumulation of 4 (effective batch size 4) and a binarization threshold of 0.5.

Table 1. Quantitative performance on the synthesized dual-async benchmark in terms of 3D Dice. All values are reported as mean ($\pm$ standard deviation). Upper bound denotes an oracle setting with perfectly aligned views.

amp (mm)	Raw	NR-Align (ours)	Upper bound
5.0	0.4645 ± 0.1153	0.5492 ± 0.0854	0.6265 ± 0.0728
10.0	0.3897 ± 0.1036	0.5356 ± 0.0862	0.6265 ± 0.0728
15.0	0.3074 ± 0.1210	0.5036 ± 0.0999	0.6265 ± 0.0728
Overall	0.3872 ± 0.1304	0.5295 ± 0.0927	0.6265 ± 0.0728

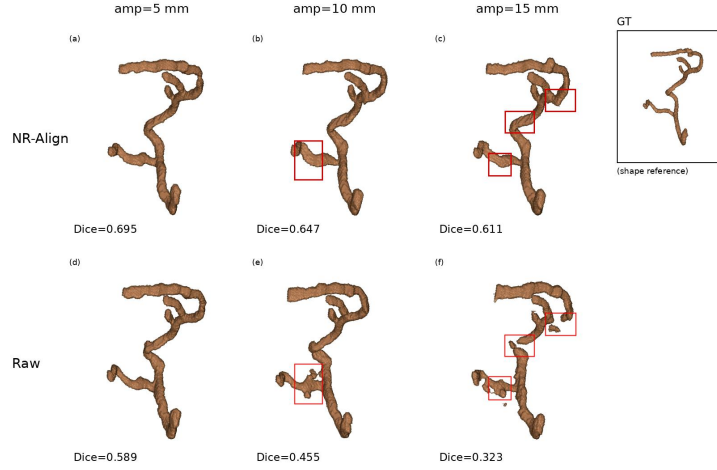


Fig. 2. Qualitative 3D reconstructions on synthesized data under increasing deformation ($amp = 5, 10, 15$ mm). NR-Align substantially reduces discontinuity and ghosting artifacts (red boxes) present in the Raw inputs.

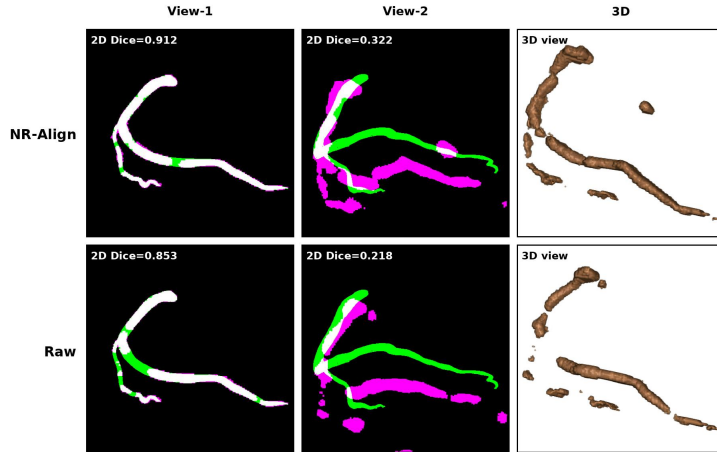
3.3 Results on Synthesized Dual-Async Data

Table 1 reports quantitative performance on the synthesized dual-async benchmark in terms of 3D Dice. NR-Align consistently outperforms the uncorrected input (Raw) across all deformation bins, and the relative improvement increases with stronger deformation: **+18.2%** at 5 mm, **+37.4%** at 10 mm, and **+63.8%** at 15 mm. Overall, NR-Align improves Dice by **+36.8%**. Across all bins, NR-Align is consistently closer to the ground-truth-projection (upper bound), where both views are generated by forward-projecting the same ground-truth volume, indicating that the improvement is mainly driven by reducing motion-induced cross-view mismatch rather than simply smoothing the volume.

Qualitatively (Fig. 2), Raw reconstructions exhibit typical asynchronous failure modes, such as vessel discontinuities and ghosting artifacts around complex bifurcations. NR-Align effectively mitigates these artifacts, yielding coherent topologies that closely match the ground truth.

Table 2. Quantitative performance on real clinical DSA cases. Results are reported as mean $\pm$ standard deviation.

Model	View-1 Dice	View-2 $O_t(1)$	View-2 $O_t(2)$
Raw	0.7638 ± 0.0952	0.2944 ± 0.1042	0.3920 ± 0.1257
NR-Align (ours)	0.8151 ± 0.0862	0.3526 ± 0.0954	0.4322 ± 0.1083

**Fig. 3.** Qualitative comparison on a real clinical DSA case. Columns show View-1 (higher-overlap projection), View-2, and the resulting 3D reconstruction. For 2D overlays, white indicates agreement with the manual mask, green indicates false negatives, and magenta indicates false positives.

3.4 Results on Real Clinical DSA Data

To assess clinical applicability and domain generalization, we evaluated our method on unseen, real-world asynchronous dual-view coronary DSAs. Lacking 3D ground truth in clinical data, we measure the agreement between reprojected 3D reconstructions and manual 2D vessel masks using the overlap metric $O_t(d)$ with a distance tolerance d (mm), noting that $O_t(0)$ is equivalent to the Dice score. We report the Dice score on the better-aligned projection (View-1) and $O_t(1)/O_t(2)$ on the secondary projection (View-2) to quantify robustness under relaxed spatial tolerances.

As shown in Table 2, our motion compensation model (NR-Align) significantly improves cross-view consistency compared to the uncorrected input (Raw). This demonstrates the successful transfer of the learned compensation from synthetic environments to real clinical domains.

Fig. 3 illustrates a representative clinical case. While the Raw input suffers from severe cross-view mismatch causing fragmented and spurious 3D structures, NR-Align notably increases the 2D reprojection overlap (View-1: $0.853 \rightarrow 0.912$; View-2 Dice: $0.218 \rightarrow 0.322$) and yields a highly coherent 3D reconstruction.

Table 3. Ablation on NR-Align architecture on synthesized dual-async data (3D Dice, mean over the test set).

Variant	3D Dice
Single-head (union only)	0.4292
Dual-head (union + overlap)	0.5295

Table 4. Ablation on deformation synthesis strategies, evaluated on unseen clinical DSA data. Results are reported as mean $\pm$ standard deviation.

Training synthesis	View-1 Dice ($O_t(0)$)	View-2 $O_t(1)$	View-2 $O_t(2)$
SVF-based baseline	0.7412 $\pm$ 0.0821	0.2735 $\pm$ 0.0886	0.3718 $\pm$ 0.0975
Physiological (ours)	0.8151 $\pm$ 0.0862	0.3526 $\pm$ 0.0954	0.4322 $\pm$ 0.1083

3.5 Ablation Studies

Network design We ablate the prediction heads of NR-Align while keeping the training protocol and synthesized data fixed. As shown in Table 3, the dual-head variant (union + overlap) substantially outperforms the single-head variant (union only) in 3D Dice. This suggests that explicitly modeling cross-view overlap provides a useful structural constraint for motion compensation, especially around complex vessel topology.

Deformation synthesis strategy The physiological deformation synthesis is evaluated against a general non-rigid baseline using unconstrained SVF. As shown in Table 4, the SVF-based baseline fails to generalize clinically, performing slightly worse than the uncorrected Raw input. This suggests that unconstrained deformations introduce arbitrary vascular distortions, causing the network to learn non-anatomical mappings. In contrast, explicitly modeling cardiopulmonary motions significantly improves cross-view consistency, highlighting the critical role of physiological synthesis in effective sim-to-real transfer.

4 Conclusion

This paper addressed the cross-view non-rigid mismatch in dual-async DSA coronary 3D reconstruction caused by cardiopulmonary motion. Our contributions are twofold: (1) a physiology-aware SVF-based deformation synthesis pipeline that models realistic respiratory and cardiac motions, significantly bridging the sim-to-real gap; and (2) NR-Align, a motion compensation network that explicitly corrects cross-view structural mismatch on backprojected volumes using dual-head union/overlap predictions. Extensive experiments validate that NR-Align substantially improves reconstruction continuity and topological accuracy on both synthesized data and unseen real clinical cases.

Acknowledgements. This work was supported by the Beijing Natural Science Foundation under Grant QY25324 and the Chinese National Natural Science Foundation (62401069).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cafaro, A., Dorent, R., Haouchine, N., et al.: Two projections suffice for cerebral vascular reconstruction. In: MICCAI 2024, LNCS, vol. 15007, pp. 722–731. Springer (2024)
2. Fu, X., Li, Y., Tang, F., et al.: 3DGR-CAR: Coronary artery reconstruction from ultra-sparse 2D X-ray views with a 3D Gaussians representation. In: MICCAI 2024, LNCS, vol. 15007, pp. 14–24. Springer (2024)
3. Wang, Y., Banerjee, A., Choudhury, R.P., et al.: DeepCA: Deep Learning-Based 3D Coronary Artery Tree Reconstruction from Two 2D Non-Simultaneous X-Ray Angiography Projections. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 337–346. IEEE (2025)
4. Wang, Y., Banerjee, A., Grau, V.: NeCA: 3D coronary artery tree reconstruction from two 2D projections via neural implicit representation. *Bioengineering* **11**(12), 1227 (2024)
5. Maas, K.W.H., Ruijters, D., Vilanova, A., et al.: NeRF-CA: Dynamic reconstruction of X-ray coronary angiography with extremely sparse-views. *IEEE Transactions on Visualization and Computer Graphics* **31**(10), 8782–8795 (2025)
6. Zhang, B., Zhang, Z., Liu, S., et al.: XA-Sim2Real: Adaptive representation learning for vessel segmentation in X-ray angiography. In: MICCAI 2024, LNCS, vol. 15006, pp. 747–756. Springer (2024)
7. Zhao, C., Esposito, M., Xu, Z., et al.: HAGMN-UQ: Hyper association graph matching network with uncertainty quantification for coronary artery semantic labeling. *Medical Image Analysis* **99**, 103374 (2025)
8. Wang, G., Zhou, P., Gao, H., et al.: Coronary vessel segmentation in coronary angiography with a multi-scale U-shaped transformer incorporating boundary aggregation and topology preservation. *Physics in Medicine & Biology* **69**(2), 025012 (2024)
9. Kim, H.W., Noh, S.C., Kim, S.H., et al.: Effective descriptor extraction strategies for correspondence matching in coronary angiography images. *Scientific Reports* **14**(1), 18630 (2024)
10. Wu, W., Xie, H., Zhang, S., et al.: Exhaustive matching of 3D/2D coronary artery structure based on imperfect segmentations. *International Journal of Computer Assisted Radiology and Surgery* **19**(1), 109–117 (2024)
11. Lin, H., Li, H., Xie, S., et al.: An Integrated Approach for Simultaneous Calibration and 3-D Coronary Artery Centerline Reconstruction From Two Non-Simultaneous Angiographic Images. *IEEE Transactions on Medical Imaging* **44**(12), 4800–4810 (2025)
12. Wang, F., Luo, Y., Wen, K., et al.: Groupwise deformable registration of diffusion tensor cardiovascular magnetic resonance: Disentangling diffusion contrast, respiratory and cardiac motions. In: MICCAI 2024, LNCS, vol. 15002, pp. 640–650. Springer (2024)

13. Ruan, Y., Gu, L., Kurose, Y., et al.: Physiology-Aware PolySnake for Coronary Vessel Segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 8873–8882. IEEE (2025)
14. Zeng, A., Wu, C., Lin, G., et al.: ImageCAS: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images. *Computerized Medical Imaging and Graphics* **109**, 102287 (2023)
15. Biguri, A., Dosanjh, M., Hancock, S., et al.: TIGRE: A MATLAB-GPU toolbox for CBCT image reconstruction. *Biomedical Physics & Engineering Express* **2**(5), 055010 (2016)
16. Chen, J., Liu, Y., Wei, S., et al.: A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond. *Medical Image Analysis* **100**, 103385 (2025)
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI 2015, LNCS, vol. 9351, pp. 234–241. Springer (2015)
18. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., et al.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: MICCAI 2016, LNCS, vol. 9901, pp. 424–432. Springer (2016)
19. Balakrishnan, G., Zhao, A., Sabuncu, M.R., et al.: VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging* **38**(8), 1788–1800 (2019)
20. Dalca, A.V., Balakrishnan, G., Guttag, J., et al.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical Image Analysis* **57**, 226–236 (2019)