Efficient Test-Time Optimization via Knowledge Distillation for Modality-Agnostic Medical Image Registration

Sneha Sree C.[1], Dattesh Shanbhag[1], and Sudhanya Chatterjee[1,⋆]

[1] GE HealthCare, Bengaluru, India
{SnehaSree.C,dattesh.shanbhag,sudhanya.chatterjee}@gehealthcare.com

Abstract. Non-rigid medical image registration (NRR) is critical for aligning multi-modal anatomical structures but is challenged by complex physiological deformations (e.g., respiration, contrast variations). Traditional NRR incurs high computational costs and requires extensive parameter tuning, while deep learning (DL) approaches rely on restrictive task-specific retraining. Addressing these limitations, we introduce a novel AI-driven 3D NRR framework that generalizes across diverse imaging modalities and anatomies. By eliminating anatomy- and modality-specific customization, our approach provides an efficient, universal solution for streamlined integration into heterogeneous clinical workflows.

Keywords: Non-Rigid Registration (NRR) · Test Time Optimization (TTO)

1 Introduction

Medical image registration aligns multi-modal or temporal scans to establish spatial correspondence for critical clinical applications [4,5,6]. While rigid alignment is analytically efficient [15], non-rigid registration (NRR) relies on complex, computationally expensive iterative optimization [7]. Deep learning (DL) mitigates this bottleneck by directly predicting spatial transformations, drastically accelerating NRR [2,17]. Recent unsupervised frameworks such as VoxelMorph [2] and DeepReg [3] learn spatial transformations directly from data, enabling fast registration, but generalization to unseen domains remains challenging. Adaptive approaches have emerged to address this gap: LapIRN v2 [18] employs test-time optimization (TTO), ConvexAdam [19] uses dual-optimization, and Dino-Reg [20] leverages foundation model features for robust multimodal alignment.
The proposed framework differs from these paradigms by enabling efficient test-time optimization (TTO) through knowledge distillation (KD). While ConvexAdam often requires task-specific parameter tuning, and LapIRN v2, which incurs high latency due to dense instance optimization, our distilled approach learns the optimization behavior itself. This allows our lightweight student model

⋆ Work was done while at GE HealthCare, Bengaluru, India. Currently works at AI & Advanced Computing Lab, KLA Corporation, India.

($\sim$0.23M parameters) to perform rapid, instance-specific refinement without the heavy memory overhead of foundation backbones, making it uniquely suited for high-throughput clinical environments and 4D dynamic imaging.
A key contribution of this work is its clinical universality, demonstrated across a diverse range of imaging scenarios than contemporary benchmarks. While leading methods typically focus on specific anatomical regions or static modality pairs, we evaluate our unified framework across four fundamentally different registration challenges: (i) 4D DCE-MRI motion correction in the liver involving complex temporal dynamics; (ii) pelvic MRI–CT multi-modal registration with non-linear intensity relationships; (iii) pre/post-contrast CT with large deformations across multiple organs; and (iv) inter-subject cross-contrast brain MRI. Enabled by modality-agnostic synthetic pre-training [1] and distilled TTO, the proposed approach operates in a parameter-free setting (no task-specific tuning), providing a robust "one-model-fits-all" solution for heterogeneous clinical data. The use of a single model with fixed TTO parameters across all experiments highlights its strong adaptability.

2 Method

We propose a three-stage framework for DL based medical image registration.

Stage-1: Model Pre-training To align fixed (I_f) and moving (I_m) 3D images, a 3D CNN (θ) predicts a dense displacement field (DDF) $\mathbf{u} = \mathcal{F}_\theta(I_f, I_m)$. Here, $\mathbf{u} : \Omega \subset \mathbb{R}^3 \to \mathbb{R}^3$ defines the voxel-wise non-rigid transformation ϕ at spatial location $\mathbf{x}$ as $\phi(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$, enabling domain-wide local deformations.
<u>Network Architecture</u> The proposed framework is built upon a 3D U-Net architecture [9]. The encoder comprises four levels (16, 32, 32, 64 channels), each applying a $3 \times 3 \times 3$ convolution (stride 1, padding 1), LeakyReLU ($\alpha = 0.2$), and $2 \times 2 \times 2$ max-pooling.
<u>Displacement Field Prediction</u>: Post-decoder, three full-resolution $3{\times}3{\times}3$ convolutional blocks (16 channels, LeakyReLU $\alpha = 0.2$) output the final displacement field $\mathbf{u} \in \mathbb{R}^{H\times W\times D\times 3}$.
<u>Spatial Transformation</u>: The moving image I_m is warped to align with the fixed image I_f using the predicted deformation field ϕ through trilinear interpolation: $I_m \circ \phi(\mathbf{x}) = I_m(\phi(\mathbf{x})) \quad \forall \mathbf{x} \in \Omega$ where Ω denotes the image domain.
<u>Loss Function</u>: The network is trained on synthetically generated image pairs using a composite loss function that combines image-similarity metric with a deformation regularization term: $\mathcal{L}_{\text{total}} = \lambda_{\text{sim}}\, \mathcal{L}_{\text{MultiAxisMI}}(I_f,\, I_m \circ \phi) + \lambda_{\text{smooth}}\, \mathcal{R}(\mathbf{u})$ where λ_{sim}, λ_{smooth} are regularization weights and $\mathbf{u}$ is the displacement field associated with ϕ. In $\mathcal{L}_{\text{total}}$, the term $\mathcal{L}_{\text{MultiAxisMI}}$ is a multi-axis mutual information loss, this multi-axis decomposition provides a computationally efficient approximation that significantly reduces memory and computational overhead. Let $\tilde{I}_m = I_m \circ \phi$. Then, $\mathcal{L}_{\text{MultiAxisMI}} = \frac{1}{D+H+W} \sum_{a\in\{d,h,w\}} \sum_{i=1}^{A_a} \mathcal{M}(I_f^{(i,a)}, \tilde{I}_m^{(i,a)})$. To ensure anatomically plausible transformations, we penalize abrupt spatial variations in the displacement field using a smoothness regularizer $\mathcal{R}(\mathbf{u}) = \|\nabla \mathbf{u}\|^2$, where $\nabla \mathbf{u}$ denotes the spatial gradients of the displacement field.

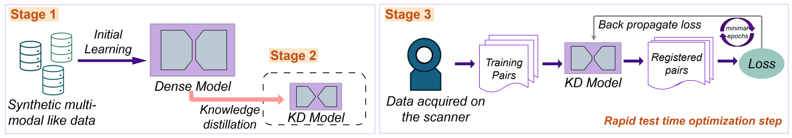


Fig. 1: Overview of the three stage pipeline of the proposed method.

Stage-2: Knowledge Distillation (KD) for Efficient Deployment To enable rapid TTO, Knowledge Distillation (KD) transfers representations from the dense teacher to a lightweight student, preserving target accuracy [16].
Student Architecture: The student U-Net employs a reduced channel configuration (encoder: 16, 24, 24, 32; bottleneck: 32; decoder: 32, 24, 24, 16). This compresses the parameter count nearly three-fold (from ∼0.6M to ∼0.23M).
Distillation Objective:We employ output-based KD: the student approximates the teacher's predicted displacement fields, effectively transferring learned deformation patterns and spatial correspondences. Let $\mathbf{u}_T, \mathbf{u}_S$ denote the teacher and student displacement fields, $\mathbf{u}_T = \mathcal{F}_T(I_f, I_m)$ and $\mathbf{u}_S = \mathcal{F}_S(I_f, I_m)$, with transformations $\phi_T(\mathbf{x}) = \mathbf{x} + \mathbf{u}_T(\mathbf{x})$ and $\phi_S(\mathbf{x}) = \mathbf{x} + \mathbf{u}_S(\mathbf{x})$. The total student training objective is formulated as: $\mathcal{L}_{\mathrm{KD}} = \lambda_{\mathrm{dist}}\,\mathcal{L}_{\mathrm{MultiAxisMI}}(I_m \circ \phi_T, I_m \circ \phi_S) + \lambda_{\mathrm{sim}}\,\mathcal{L}_{\mathrm{MultiAxisMI}}(I_f, I_m \circ \phi_S) + \lambda_{\mathrm{smooth}}\,\|\nabla \mathbf{u}_S\|^2$, $\|\nabla \mathbf{u}_S\|^2$ enforces smoothness in the predicted displacement field.

Stage-3: Test-Time Optimization (TTO) at inference During inference, the model undergoes a self-supervised fine-tuning phase to adapt to each test image pair individually. This process, referred to as *Test-Time Optimization (TTO)*, performs instance-specific refinement by optimizing the student network parameters θ for a small number of epochs for a given pair, i.e., $\theta^* = \arg\min_\theta \mathcal{L}_{\mathrm{TTO}}(I_f, I_m; \theta)$. where $\mathcal{L}_{\mathrm{TTO}}$ is the test-time optimization loss comprising: (i) an image-similarity term that enforces alignment between the fixed image I_f and the warped moving image $I_m \circ \phi$ using Multi-Axis MI and NCC, i.e., $\mathcal{L}_{\mathrm{sim}} = \mathcal{L}_{\mathrm{MultiAxisMI}}(I_f, I_m \circ \phi) + \mathcal{L}_{\mathrm{NCC}}(I_f, I_m \circ \phi)$; (ii) a smoothness regularizer that penalizes spatial gradients of the displacement field, $\|\nabla \mathbf{u}\|^2$, to promote coherent deformations; and (iii) a divergence penalty, $\mathcal{R}_{\mathrm{div}}(\mathbf{u}) = \alpha_{\mathrm{div}}\|\nabla \cdot \mathbf{u}\|^2$, which encourages near-incompressibility by discouraging local volume changes. The complete objective is $\mathcal{L}_{\mathrm{TTO}} = \lambda_{\mathrm{sim}}\mathcal{L}_{\mathrm{sim}} + \lambda_{\mathrm{smooth}}\|\nabla \mathbf{u}\|^2 + \alpha_{\mathrm{div}}\|\nabla \cdot \mathbf{u}\|^2$. This design reflects the distinct roles of pre-training and TTO: Stage-1 uses a simpler objective to learn a robust, modality-agnostic deformation prior, while Stage-3 augments the loss with NCC and divergence constraints to enable precise, subject-specific refinement and improved physical plausibility during test-time optimization.

Implementation Details All models were implemented in PyTorch (v2.4.1+) and trained on NVIDIA RTX 6000 GPUs (48 GB VRAM). The stage-1 and stage-2 training was done using synthetic data which was generated following the exact protocol established in SynthMorph [1]. The teacher model was trained for

300 epochs on synthetic image pairs generated on-the-fly (500 pairs per epoch). Adam optimization. Learning rate of $1e{-}4$. After teacher convergence, the student model was trained using the knowledge distillation framework on the synthetic dataset for 500 epochs. *TTO at inference:* The pre-trained model was fine-tuned on target data. Stopping Criteria for TTO: 100 epochs or one minute of training, whichever is reached earlier. During the TTO step, data is resampled as needed to the nearest resolution of 2^4 for $H \times W \times D$, ensuring compatibility with the U-Net-based model architecture.

2.1 Experiments

We evaluate the proposed method across diverse clinical scenarios spanning multiple modalities and contrasts. As part of the preprocessing pipeline, affine registration and resampling to a uniform spatial resolution are performed where required to correct for global misalignment and ensure consistency prior to non-rigid registration. The registered outputs are mapped back to their native resolution for evaluation.

Dynamic Contrast-Enhanced MRI (DCE-MRI)4D (3D+t) DCE-MRI captures contrast dynamics, but respiratory liver motion misaligns frames. Registering all temporal phases to a reference restores alignment for accurate contrast-uptake curves. Due to dynamic intensity changes, this inherently constitutes a multi-contrast registration task. *Dataset and Experiment*: Data was obtained on a 1.5T GE SIGNA EXCITE system, 3D EFGRE, TE/TR=1.12/4.8ms, matrix is $256 \times 256 \times 32$, 30 bolus phase volumes (temporal frames), FA=15, FOV=$450mm^2$. The middle time point is selected as the fixed reference volume, and the remaining 29 frames are treated as moving volumes. TTO is performed using the dense teacher model and the KD student model. *Analysis*: To understand the alignment of structures over the temporal frames, the contrast uptake curves are plotted in a region of interest at liver dome (region highly susceptible to motion). Additionally, we trained a comparison model, *Application Specific AI model*, using the same dense teacher network and loss function as the TTO-based approach exclusively for liver DCE-MRI (500 epochs, 261 paired volumes). Achieving comparable performance between this model and the TTO-based method would suggest that TTO maintains structural alignment accuracy while providing enhanced generalizability.

MRI–CT Multi-Modal Registration MRI and CT exhibit significantly different tissue contrasts due to their distinct physical principles (e.g., inverted bone intensities), make MRI-CT alignment a challenging multi-modal registration problem. *Dataset and Experiment*: MRI and CT volumes are acquired from the same subject, each volume originally has dimensions of $256 \times 256 \times 120$ voxels. The MRI and CT images are treated as the fixed and moving images, respectively. The proposed non-rigid registration (NRR) using the TTO approach is applied to capture complex local anatomical variations and spatially varying deformations across modalities. *Analysis*: A visual assessment and Patch-wise Mutual Information Map (PMM) analysis, where PMM quantifies local registration accuracy using mutual information computed over overlapping patches, were used to demonstrate the alignment post registration using the proposed method.

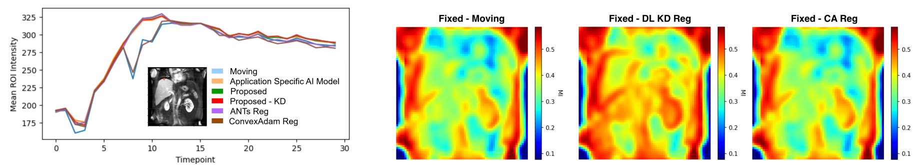
Pre- and Post-Contrast CT Registration Pre-contrast and post-contrast CT scans are routinely acquired for diagnostic evaluation across anatomical regions. For comparative analysis, it is important for these images to be anatomically aligned. *Dataset and Experiment*: TCIA datasets are used for this experiment: *Data-1 (Thorax)* [10] (Case ID: AMC-015), *Data-2 (Abdomen)* [11] (Case ID: TCGA-DI-A2QT), *Data-3 (Cardiac)* [12] (Case ID: AP-26JK), and *Data-4 (Liver)* [13] (Case ID: HCC_023). For all datasets, pre-contrast and post-contrast scans are treated as the fixed and moving images, respectively. The proposed DL-based non-rigid registration (NRR) method is applied to align the image pairs. *Analysis*: TotalSegmentator [14] is used to segment the primary organ in the pre-contrast image, post-contrast image before registration, and post-contrast image after registration. Alignment between the segmented organs is quantified using Dice scores.

Cross-Contrast Brain MRI Registration Brain MRI studies often include multiple contrast sequences (e.g., T1-weighted, T2-weighted, Fluid-Attenuated Inversion Recovery (FLAIR)), requiring accurate cross-contrast alignment.
Dataset and Experiment: We consider three experimental settings: inter-subject same-contrast ($n = 24$ cases), inter-subject cross-contrast ($n = 24$ cases), and same-subject cross-contrast ($n = 8$ cases), covering variations in both anatomical structure and image contrast. T2-weighted FLAIR (providing contrast variation via fluid suppression) and T2-weighted FSE MRI volumes are used. Both share native dimensions ($512 \times 512 \times 22$ voxels) and spacing ($0.4297 \times 0.4297 \times 6$ mm). The proposed non-rigid registration (NRR) method is applied to align the image pairs. *Analysis*: In addition to visual assessment, Patch-wise Mutual Information Maps (PMMs) are computed, and mean PMM values within the brain mask are used to quantify registration performance across the population.

3 Results and Discussion

We present TTO outcomes (Section 2.1) comparing the dense teacher (**DL Reg**), distilled student (**DL KD Reg**), ANTs (**ANTs Reg**) and ConvexAdam [19] (**ConvexAdam Reg**). To explicitly isolate the architectural contribution of knowledge distillation, we introduce a native lightweight baseline that shares the student's exact parameter dimensions but trained without KD (inferred with TTO to ensure fairness in comparison).

Dynamic Contrast Enhanced MRI (DCE-MRI) Fig. 2a shows liver dome contrast uptake curves. The moving image (blue) shows a drop in signal, which is not observed in the DL Reg (green) and DL KD Reg (red) results. Both closely match the application-specific model (orange), proving generalizability without performance loss, while the overlapping green and red curves confirm the miniaturized KD model matches dense model performance. Alignment was quantified via mean PMMs, computed on $16 \times 16 \times 16$ patches (stride 4) and averaged across timepoints (Fig. 2b). Mean PMMs improved from 0.3801 (fixed–moving) to 0.4291 (Fixed–DL Reg) and 0.4195 (Fixed–DL KD Reg). ANTs achieved 0.4192, while ConvexAdam (Fixed-CA Reg) lagged at 0.3798. These results place our proposed method at or above the performance of established classical registration baselines, particularly around challenging regions such as air pockets near the liver, kidney, and spleen. Additional visual results are provided in the supplementary material.



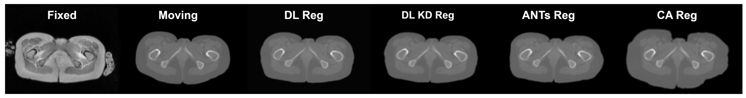
(a) (b)

Fig. 2: (a) ROI contrast uptake curves demonstrate improved temporal alignment for DL Reg, DL KD Reg and ANTs Reg while ConvexAdam shows suboptimal alignment. (b) PMM maps indicate higher alignment for DL KD Reg, with lower values for ConvexAdam, reflecting poorer non-rigid alignment.

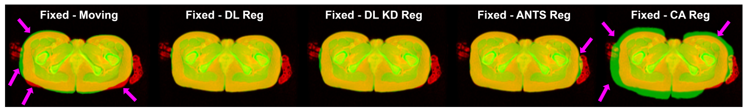
MRI-CT Multi-Modal Registration TTO results (Fig. 3a) demonstrate accurate cross-modal alignment, maintaining each modality's appearance while preserving anatomical consistency. The false-color overlay (Fig. 3b; MR: red, registered CT: green) shows good correspondence of body contours and internal structures, with yellow indicating overlap. Quantitatively, mean PMMs ($16 \times 16 \times 16$ patches, stride 4) increase from 0.0986 (Fixed–Moving) to 0.1797 (Fixed–DL Reg) and 0.1680 (Fixed–KD Reg), comparable to ANTs (0.1742). In contrast, ConvexAdam shows weaker cross-modal consistency, reflected in its lower PMM of 0.0862 (Fixed–CA Reg).

Pre- and Post-Contrast CT Registration To evaluate performance in this experiment, we measured organ-specific segmentation overlap (lung for Data-1; liver for Data-2–4). Table 1 reports Dice scores for all methods: DL Reg (with/without KD), ConvexAdam, ANTs, and a lightweight non-KD baseline. Overall, our framework achieves performance comparable to ANTs. On an NVIDIA RTX 6000, the distilled student model converges in 46 seconds, outperforming both the dense teacher (56 seconds) and the clinical gold standard ANTs (1.25 minutes), with runtime evaluated within optimization-based methods. This efficiency is achieved without the significant memory or latency overheads seen in dual-optimization or foundation-model approaches.

Brain MRI Registration Fig. 4 shows an inter-subject cross-contrast example where our method visually outperforms ANTs, restoring post-affine misaligned ventricles (red box). Table 2 compares quantitative in-mask PMMs for F–M, F–R, F–R–KD, F–R–S, F–AR, and F–CA across three scenarios: *Same-subject, Cross-contrast* (n = 8), *Inter-subject, Same-contrast* (n = 24), and *Inter-subject, Cross-contrast* (n = 24). All proposed models consistently im-



(a)



(b)

Fig. 3: MRI/CT registration. Fig. 3a Representative single-slice visualizations compare alignment quality across methods. Fig. 3b False-color overlay (fixed MR: red, registered CT: green) demonstrates precise spatial and body contour correspondence. Visually, the proposed method outperforms ANTs.

proved PMMs. DL Reg scored highest, closely followed by DL Reg KD. Expectedly, *Same-subject, Cross-contrast* yielded the highest, most stable PMMs, while *Inter-subject, Cross-contrast* scored lower due to combined anatomical/modality complexities (same-contrast pairs inherently inflate PMMs).

Comparative Analysis We compare the distilled TTO (F-R-KD) against the dense teacher (F-R) and ConvexAdam (F-CA). Across 4D DCE-MRI, MRI–CT, and cross-contrast brain MRI, F-CA consistently underperformed (e.g., PMM 0.0862 vs. 0.1680 in MRI–CT). In pre/post-contrast CT, both F-R and F-R-KD achieved higher Dice scores than F-CA, highlighting the benefit of TTO refinement. While F-R slightly outperformed F-R-KD in isolated cases (e.g., PMM 0.2078 vs. 0.2032 in same-subject brain MRI), this is offset by the efficiency of F-R-KD, which achieves $\sim 3\times$ parameter reduction ($\sim$0.23M vs. $\sim$0.6M) and faster inference (20–25%), providing a compact yet accurate solution.

Assimilation of Knowledge Distillation: KD maximizes the lightweight model's spatial reasoning. While the native small baseline (F-R-S) is robust, KD (F-R-KD) consistently outperforms. In lung (Data-1) and liver (Data-2)

Table 1: Dice scores for pre-/post-contrast CT organ segmentation across six settings: Fixed–Moving (F-M), Fixed–DL Reg (F-R), Fixed–DL Reg with KD (F-R-KD), Fixed–ANTs (F-AR), small DL Reg without KD (F-R-S), and Fixed–ConvexAdam Reg (F-CA).

Data	**Dice Score**					
	F-M	F-R	F-R-KD	F-AR	F-R-S	F-CA
Data-1	0.9834	0.9838	0.9928	0.9844	0.9836	0.9956
Data-2	0.9393	0.9737	0.9744	0.9733	0.9662	0.9703
Data-3	0.9576	0.9755	0.9691	0.9526	0.9655	0.9503
Data-4	0.8096	0.9675	0.9519	0.9724	0.9655	0.9140

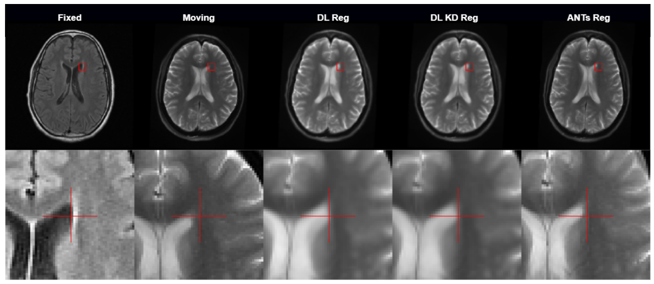


Fig. 4: Qualitative comparison of inter-subject cross-contrast MRI registration. The zoomed-in region within the red bounding box highlights accurate alignment of the frontal horn of the lateral ventricles post registration.

Table 2: Mean and standard deviation of PMMs across categories (values computed inside brain mask).

	F-M		F-R		F-R-KD		F-AR		F-R-S		F-CA	
Category	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Same-subject, Cross-contrast	0.1677	0.0108	0.2078	0.0135	0.2032	0.0138	0.1861	0.0209	0.2029	0.0136	0.1654	0.0224
Inter-subject, Same-contrast	0.0887	0.0114	0.1707	0.0192	0.1618	0.0190	0.1389	0.0151	0.1616	0.0189	0.1054	0.0088
Inter-subject, Cross-contrast	0.0827	0.0077	0.1372	0.0096	0.1291	0.0092	0.1120	0.0092	0.1289	0.0091	0.0950	0.0050

CTs, F-R-KD yields Dice scores of 0.9928 and 0.9744, outperforming F-R-S (0.9836, 0.9662) and rivaling the dense teacher. Mirrored by higher PMMs in cross-contrast brain MRIs, these results confirm KD empowers miniaturized networks to achieve peak accuracy without compromising speed.

4 Conclusion

We propose a generalizable DL-based registration framework that combines pretraining, knowledge distillation, and test-time optimization (TTO) for robust cross-modal alignment. Unlike inference-time optimization approaches, our distilled model enables efficient, parameter-free adaptation while maintaining comparable accuracy. TTO further provides instance-specific refinement, achieving a practical balance between accuracy, adaptability, and computational efficiency across diverse modalities and anatomies.

Acknowledgments. Authors acknowledge use of data from the following sources:

- The Cancer Genome Atlas Uterine Corpus Endometrial Carcinoma (TCGA–UCEC) collection [11], generated by the TCGA Research Network (`http://cancergenome.nih.gov/`).
- The VA Research Precision Oncology Program – APOLLO (VAREPOP–APOLLO) [12], produced by the Veterans Health Administration's Research for Precision Oncology Program and the APOLLO Research Network.

References

1. Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V.: SynthMorph: Learning Contrast-Invariant Registration Without Acquired Images. IEEE Trans. Med. Imaging 41(3), 543–558 (2022)
2. Balakrishnan, G., Zhao, A., Sabuncu, M., Guttag, J., Dalca, A.V.: VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Trans. Med. Imaging 38(8), 1788–1800 (2019)
3. Fu, Y., Brown, N.M., Saeed, S.U., Casamitjana, A., Baum, Z.M.C., Delaunay, R., Yang, Q., Grimwood, A., Min, Z., Blumberg, S.B., Iglesias, J.E., Barratt, D.C., Bonmati, E., Alexander, D.C., Clarkson, M.J., Vercauteren, T., Hu, Y.: DeepReg: A deep learning toolkit for medical image registration. J. Open Source Softw. 5(55), 2705 (2020)
4. Hill, D.L.G., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical image registration. Phys. Med. Biol. 46(3), R1 (2001)
5. Maintz, J.B.A., Viergever, M.A.: A survey of medical image registration. Med. Image Anal. 2(1), 1–36 (1998)
6. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. IEEE Trans. Med. Imaging 32(7), 1153–1190 (2013)
7. Oliveira, F.P.M., Tavares, J.M.R.S.: Medical image registration: a review. Comput. Methods Biomech. Biomed. Eng. 17(2), 73–93 (2014)
8. Haskins, G., Kruger, U., Yan, P.: Deep learning in medical image registration: a survey. Mach. Vis. Appl. 31(8), 1–18 (2020)
9. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: MICCAI, pp. 424–432. Springer (2016)
10. Bakr, S., Gevaert, O., Echegaray, S., Ayers, K., Zhou, M., Shafiq, M., Zheng, H., Zhang, W., Leung, A., Kadoch, M., Shrager, J., Quon, A., Rubin, D., Plevritis, S., Napel, S.: Data for NSCLC Radiogenomics (Version 4) [Data set]. The Cancer Imaging Archive (2017)
11. Erickson, B.J., Mutch, D., Lippmann, L., Jarosz, R.: The Cancer Genome Atlas Uterine Corpus Endometrial Carcinoma Collection (TCGA-UCEC) (Version 4) [Data set]. The Cancer Imaging Archive (2016)
12. The Research for Precision Oncology Program, APOLLO Research Network: VA Research Precision Oncology Program – APOLLO (VAREPOP-APOLLO) (Version 3) [Data set]. The Cancer Imaging Archive (2024)
13. Moawad, A.W., Fuentes, D., Morshid, A., Khalaf, A.M., Elmohr, M.M., Abusaif, A., Hazle, J.D., Kaseb, A.O., Hassan, M., Mahvash, A., Szklaruk, J., Qayyom, A., Elsayes, K.: Multimodality annotated HCC cases with and without advanced imaging segmentation [Data set]. The Cancer Imaging Archive (2021)
14. Wasserthal, J., Breit, H., Meyer, M., et al.: TotalSegmentator: Robust Segmentation of 104 Anatomical Structures in CT Images. Radiol. Artif. Intell. 5(5), e230024 (2023)
15. Avants, B.B., Tustison, N.J., Song, G.: Advanced Normalization Tools (ANTS). Insight J., 1–35 (2009)
16. Hinton, G., Vinyals, O., Dean, J.: Distilling the Knowledge in a Neural Network. arXiv:1503.02531 (2015)
17. Reithmeir, A., Spieker, V., Sideri-Lampretsa, V., Rueckert, D., Schnabel, J.A., Zimmer, V.A.: From model based to learned regularization in medical image registration: A comprehensive review. Med. Image Anal. 108, 103854 (2026)

18. Mok, T.C., Li, Z., Xia, Y., Yao, J., Zhang, L., Zhou, J., Lu, L.: Deformable medical image registration under distribution shifts with neural instance optimization. In: International Workshop on Machine Learning in Medical Imaging, pp. 126–136. Springer (2023)
19. Siebert, H., Großbröhmer, C., Hansen, L., Heinrich, M.P.: ConvexAdam: Self-configuring dual-optimisation-based 3D multitask medical image registration. IEEE Trans. Med. Imaging (2024)
20. Jia, H., et al.: Dino-Reg: Efficient Multimodal Image Registration with Distilled Features. In: Proc. MICCAI (2024)