Test Time Optimized Generalized AI-based Medical Image Registration Method

Sneha Sree C[1] SNEHASREE.C@GEHEALTHCARE.COM
Dattesh Shanbhag[1] DATTESH.SHANBHAG@GEHEALTHCARE.COM
Sudhanya Chatterjee[1] SUDHANYA.CHATTERJEE@GEHEALTHCARE.COM
[1] *GE HealthCare, Banglore, India*

Abstract

Medical image registration is critical for aligning anatomical structures across imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound. Among existing techniques, non-rigid registration (NRR) is particularly challenging due to the need to capture complex anatomical deformations caused by physiological processes like respiration or contrast-induced signal variations. Traditional NRR methods, while theoretically robust, often require extensive parameter tuning and incur high computational costs, limiting their use in real-time clinical workflows. Recent deep learning (DL)-based approaches have shown promise; however, their dependence on task-specific retraining restricts scalability and adaptability in practice. These limitations underscore the need for efficient, generalizable registration frameworks capable of handling heterogeneous imaging contexts. In this work, we introduce a novel AI-driven framework for 3D non-rigid registration that generalizes across multiple imaging modalities and anatomical regions. Unlike conventional methods that rely on application-specific models, our approach eliminates anatomy- or modality-specific customization, enabling streamlined integration into diverse clinical environments.

Keywords: Non-Rigid Registration (NRR), Test Time Optimization (TTO), Deep Learning.

1. Introduction

Medical image registration is a fundamental technique used to align two or more images of the same anatomical region, acquired at different time points or through different imaging modalities (Hill et al., 2001; Maintz and Viergever, 1998). Its primary goal is to establish spatial correspondence between images, enabling the integration of information from multiple sources for improved analysis. This process plays a critical role in various clinical applications, including treatment planning, disease monitoring, surgical guidance, and multi-modal image fusion (Sotiras et al., 2013).

Image registration spatially aligns a moving image to a reference or fixed image. It uses (i) a transformation model (e.g., warp field), (ii) a similarity metric to assess alignment, and (iii) an optimization method that updates parameters to maximize similarity (Oliveira and Tavares, 2014). Transformation models can be rigid, affine, or deformable to capture complex anatomical variations, commonly referred to as non-rigid registration (NRR). While rigid and affine registration can be efficiently performed using analytical approaches (Avants et al., 2009), NRR is computationally challenging due to the complexity of deformation models and the iterative nature of optimization. Deep learning (DL)-based registration methods address these limitations by leveraging learned features to predict transformations directly, significantly reducing computation time for NRR (Balakrishnan et al., 2019; Reithmeir et al., 2026).

Recent advances in deep learning (DL)-based medical image registration have shown significant promise. Unsupervised frameworks such as *VoxelMorph* (Balakrishnan et al., 2019) and *DeepReg* (Fu et al., 2020) learn spatial transformations directly from data, enabling faster and more flexible registration. *GroupRegNet* (Zhang et al., 2021) performs groupwise one-shot 4D registration, removing fixed references and reducing inference time. *IIRP-Net* (Ma et al., 2024) uses residual pyramids and adaptive stopping without extra training. *multiGradICON* (Demir et al., 2024) and *uniGradICON* (Tian et al., 2024) offer zero-shot, cross-anatomy, cross-modality registration. Despite these innovations, generalization to unseen domains remains a major challenge, as most models rely heavily on training data distributions and struggle with domain shifts encountered in clinical practice (Reithmeir et al., 2026).

We present a 3D DL registration framework that generalizes across applications. We train on multi-modal-like synthetic data (Hoffmann et al., 2022). For each target, we use test-time optimization (TTO) with deformation-field regularization. To offset TTO latency, we apply knowledge distillation (Hinton et al., 2015) to reduce model size without loss. We evaluate on 4D DCE-MRI motion correction, MRI–CT alignment, pre/post-contrast CT with large deformations, and inter-subject brain MRI. Visual and quantitative metrics show high accuracy. All experiments use the same base model and TTO.

2. Method

We propose a three-stage framework for DL based medical image registration. The overall pipeline is shown in Figure 1. The three stages shown here are discussed in Section 2.1, 2.2, 2.3.

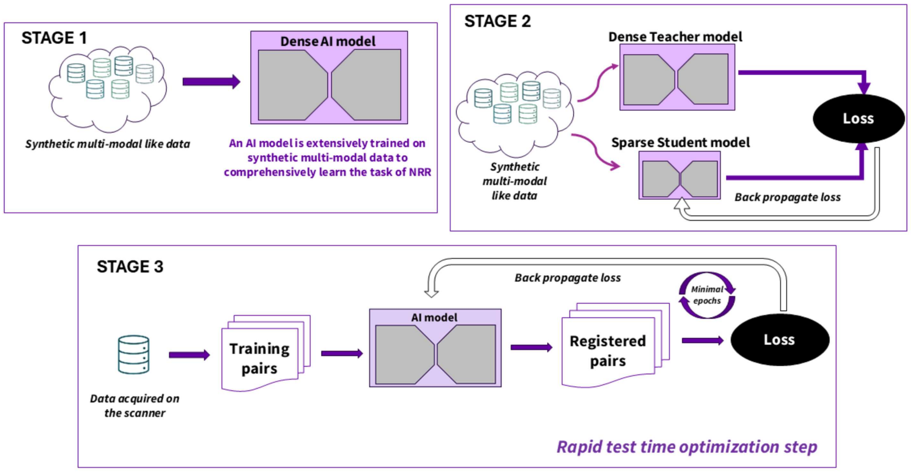

STAGE 1
Synthetic multi-modal like data
Dense AI model
An AI model is extensively trained on synthetic multi-modal data to comprehensively learn the task of NRR
STAGE 2
Dense Teacher model
Synthetic multi-modal like data
Sparse Student model
Loss
Back propagate loss
STAGE 3
Back propagate loss
Minimal epochs
Data acquired on the scanner
Training pairs
AI model
Registered pairs
Loss
Rapid test time optimization step

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

2.1. Stage-1: Model Pretraining

Let I_f and I_m denote the fixed and moving 3D images, respectively. The objective of image registration is to estimate a spatial transformation that aligns I_m to I_f. We adopt a learning-based approach where a 3D convolutional neural network (CNN), parameterized by θ, predicts a dense displacement field (DDF): $\mathbf{u} = \mathcal{F}_\theta(I_f, I_m)$, where $\mathbf{u} : \Omega \subset \mathbb{R}^3 \to \mathbb{R}^3$ represents voxel-wise displacements. The transformation ϕ applied to a spatial location $\mathbf{x}$ is defined as: $\phi(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$. This formulation enables non-rigid registration by allowing local deformations across the entire image domain.

2.1.1. Network Architecture

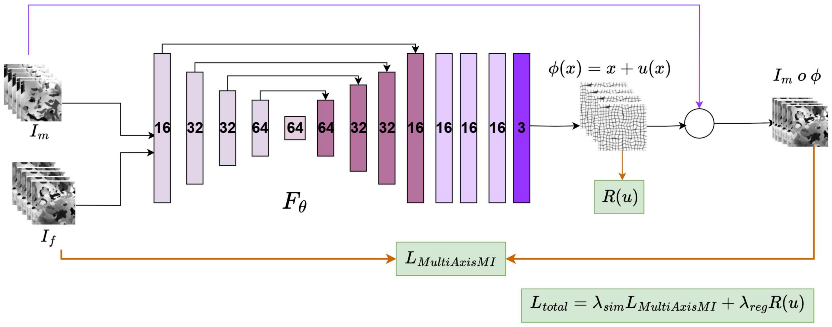


Figure 2: 3D U-Net based registration architecture.

Our registration framework is built upon a 3D U-Net architecture (Çiçek et al., 2016), optmized for learning dense deformation fields.

Encoder-Decoder Structure: The DL network is illustrated in Figure 2. The encoder consists of four hierarchical levels with feature channels of 16, 32, 32, and 64. Each level applies a 3 × 3 × 3 convolution (stride 1, padding 1) followed by LeakyReLU activation ($\alpha = 0.2$), with downsampling performed via 2 × 2 × 2 max-pooling.

Displacement Field Prediction: Following the decoder, the network employs a sequence of full-resolution refinement layers. Three convolutional blocks, each containing a 3 × 3 × 3 convolution with 16 output channels followed by LeakyReLU activation ($\alpha = 0.2$), progressively refine the feature representation while maintaining spatial resolution, yielding the displacement field as: $\mathbf{u} \in \mathbb{R}^{H \times W \times D \times 3}$.

Spatial Transformation: 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.

Loss Function: The network is trained on synthetically generated image pairs using a composite loss that combines image similarity with deformation regularization:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{sim}} \cdot \mathcal{L}_{\text{MultiAxisMI}}(I_f, I_m \circ \phi) + \lambda_{\text{smooth}} \cdot \mathcal{R}(\mathbf{u}), \tag{1}$$

where λ_{sim} and λ_{smooth} are weighting coefficients, $I_m \circ \phi$ denotes the warped moving image.The similarity term employs the Multi-Axis Mutual Information (MultiAxisMI) loss (see Appendix A, Eq. 3), which captures alignment across depth, height, and width. To ensure anatomically plausible transformations, the regularization term penalizes abrupt spatial variations and hence enforcing smoothness constraint on the deformation field, $\mathcal{R}(\mathbf{u}) = \|\nabla \mathbf{u}\|^2$, where $\nabla \mathbf{u}$ denotes spatial gradients of the displacement field.

2.2. Stage-2: Knowledge Distillation (KD) for Efficient Deployment

KD is a widely used technique for compressing deep learning models by transferring knowledge from a large, high-capacity teacher network to a smaller student network while maintaining accuracy of the target task (Hinton et al., 2015). To enable rapid test time optimization, this is a crucial step.

Student Network Architecture: The student network adopts a U-Net architecture similar to the teacher (refer Appendix H, Figure 11). It employs a reduced channel configuration for efficiency: encoder with 16, 24, 24, and 32 channels and a 32-channel bottleneck; decoder mirrors this with 32, 24, 24, and 16 channels. This compression lowers the parameter count from 0.6M to 0.23M, achieving nearly a three-fold reduction in model size.

Distillation Objective: For KD we employ output-based distillation. The student network learns to approximate the displacement fields predicted by the teacher, ensuring that the deformation patterns and spatial correspondences captured by the teacher are effectively transferred. Let $\mathbf{u}_T$ and $\mathbf{u}_S$ denote the displacement fields predicted by the teacher and student networks, respectively: $\mathbf{u}_T = \mathcal{F}_T(I_f, I_m), \quad \mathbf{u}_S = \mathcal{F}_S(I_f, I_m)$, with corresponding transformations: $\phi_T(\mathbf{x}) = \mathbf{x} + \mathbf{u}_T(\mathbf{x})$, $\phi_S(\mathbf{x}) = \mathbf{x} + \mathbf{u}_S(\mathbf{x})$. The total student training objective is formulated as:

$$\mathcal{L}_{\text{KD}} = \lambda_{\text{dist}} \cdot \mathcal{L}_{\text{MultiAxisMI}}(I_m \circ \phi_T, I_m \circ \phi_S) + \lambda_{\text{sim}} \cdot \mathcal{L}_{\text{MultiAxisMI}}(I_f, I_m \circ \phi_S) + \lambda_{\text{smooth}} \cdot \|\nabla \mathbf{u}_S\|^2$$

where, $\mathcal{L}_{\text{MultiAxisMI}}$ computes slice-wise mutual information across three directions (refer Appendix A equation 3), $\|\nabla \mathbf{u}_S\|^2$ enforces smoothness in the predicted displacement field.

2.3. Stage-3: Test-Time Optimization (TTO) at inference

During inference, the model undergoes a self-supervised fine-tuning phase to adapt to the domain-specific characteristics of the test data. This process, referred to as *Test-Time Optimization (TTO)*. It involves optimizing the student network parameters θ for a small number of epochs, the optimization objective is expressed as:

$$\theta^* = \arg\min_{\theta} \sum_{(I_f, I_m)} \mathcal{L}_{\text{TTO}}(I_f, I_m; \theta), \tag{2}$$

where $\mathcal{L}_{\text{TTO}}$ is the loss at TTO and has the following terms.

- **Image Similarity Term:** The similarity between the fixed image I_f and the warped moving image $I_m \circ \phi$ is enforced using a combination of 3D Multi-Axis MI (Appendix A Equation 3) and 3D Normalized Cross-Correlation (NCC) (Appendix B Equation 4): $\mathcal{L}_{\text{sim}} = \mathcal{L}_{\text{MultiAxisMI}}(I_f, I_m \circ \phi) + \mathcal{L}_{\text{NCC}}(I_f, I_m \circ \phi)$.
- **Smoothness Term:** The gradient-based regularization $\|\nabla \mathbf{u}\|^2$ is used to ensure smooth deformation field.

- **Divergence Regularization:** To further improve anatomical plausibility and prevent non-physical deformations, we introduce penalties on the divergence of the displacement field $\mathbf{u}$: $\mathcal{R}_{\text{div}}(\mathbf{u}) = \alpha_{\text{div}} \|\nabla \cdot \mathbf{u}\|^2$, where $\nabla \cdot \mathbf{u}$ measures local volume changes (encouraging near-incompressibility). The coefficient α_{div} controls the strength of this constraint. This loss prevents tearing-like unrealistic effects in the registered image.

Hence, the TTO loss is given by: $\mathcal{L}_{\text{TTO}} = \lambda_{\text{sim}} \cdot \mathcal{L}_{\text{sim}} + \lambda_{\text{smooth}} \cdot \|\nabla \mathbf{u}\|^2 + \alpha_{\text{div}} \|\nabla \cdot \mathbf{u}\|^2$.

2.3.1. Implementation Details

Framework and Hardware: All models were implemented in PyTorch (v2.4.1+) and trained on NVIDIA RTX 6000 GPUs (48 GB VRAM).
Teacher Network: 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 $1e{-}4$.
Student Network: After teacher convergence, the student model was trained using the knowledge distillation framework (Section 2.2) 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.

2.4. Experiments

We evaluate the proposed method across multiple clinically relevant scenarios involving different imaging modalities, imaging contrasts (dynamic imaging included). In addition to visual confirmation of alignment, efforts are made to assess the performance using quantitative metrics. We introduce Patch-wise MI Map (PMM) for registration evaluation. At each voxel, we extract a 3D patch and compute normalized mutual information (NMI) between the corresponding reference and target patches. Computational details are provided in Appendix C. Note that, in multi-modal and multi-contrast applications, the values will not reach a value of 1. Across most experiments, we compare the proposed method against well established ANTs registration (Avants et al., 2009).

2.4.1. Dynamic Contrast-Enhanced MRI (DCE-MRI)

DCE-MRI acquires multiple 3D volumes over time (4D, 3D+t) to capture exogenous contrast dynamics. Respiratory motion–dominated by diaphragm motion in liver scans–misaligns frames and hinders reliable pharmaco-kinetic uptake quantification. We correct motion by registering all temporal phases to a reference phase using the proposed method. This restores alignment and enables accurate contrast-uptake curves. Because image contrast changes markedly across time, the task is effectively a multi-contrast registration problem.
Dataset and Experiment description: The MRI data was obtained on a 1.5T GE SIGNA EXCITE system, 3D EFGRE, TE/TR=1.12/4.8ms, matrix=$256 \times 256 \times 32$, 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 (ROI) 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.

2.4.2. MRI–CT Multi-Modal Registration

MRI and CT are different imaging modalities based on different principles. Consequently, tissue contrast varies significantly between modalities, for example, in MRI signal in bone is negligible but exhibits high intensity in CT. Due to these non-trivial intensity relationships, MRI–CT alignment is truly a multi-modal registration problem.

Dataset and Experiment description: We evaluate the proposed method on paired pelvic MRI and CT volumes acquired from the same subject. Each volume originally has dimensions of $256 \times 256 \times 120$ voxels. The MRI and CT image are treated as the fixed and moving image respectively. Prior to this, the images are affine registered (Avants et al., 2009). The volumes are resampled to $256 \times 256 \times 128$ and NRR is then performed using the proposed TTO approach. After registration, results are resampled back to the native resolution for evaluation.

Analysis: A visual comparison and PMM analysis were performed to demonstrate the alignment post registration using proposed method.

2.4.3. Pre- and Post-Contrast CT Registration

Pre-contrast and post-contrast CT scans are routinely acquired for diagnostic evaluation across anatomical regions. For a comparative analysis, it is important for the images to be aligned to each other.

Dataset and Experiment Description: TCIA datasets are used for this experiment. *Data-1 (Thorax)* (Bakr et al., 2017) (Case ID: AMC-015), *Data-2 (Abdomen)* (Erickson et al., 2016) (Case ID: TCGA-DI-A2QT), *Data-3 (Cardiac)* (for Precision Oncology Program et al., 2024) (Case ID: AP-26JK), and *Data-4 (Liver)* (Moawad et al., 2021) (Case ID: HCC_023). For all datasets pre- and post-contrast scans were considered as fixed and moving respectively. Before applying proposed DL based NRR, an affine alignment was performed. All the volumes were resampled to size of $256 \times 256 \times 128$. After inference, the registered outputs were resampled back to their original resolution for evaluation.

Analysis: TotalSegmentator (Wasserthal et al., 2023) was used to segment the primary organ for pre-contrast image, post-contrast image before registration and post-contrast image after registration. Alignment between the segmented organs were quantified using Dice and Intersection-over-Union (IoU) scores.

2.4.4. Cross-Contrast Brain MRI Registration

Brain MRI studies often include multiple contrast sequences (e.g., T1-weighted, T2-weighted, Fluid-Attenuated Inversion Recovery (FLAIR)), and accurate registration across these contrasts can be important in comparative tasks.

Dataset and Experiment Description: In this experiment, we consider three experiments: Inter-subject and same contrast ($n = 24$ cases), Inter-subject and cross-contrast ($n = 24$ cases), and Same-subject but cross-contrasts ($n = 8$ cases). These experiments cover all variations: registration across shape and contrast variations. The two contrasts considered here are T2-w (FLAIR) and T2-w Fast Spin Echo (FSE) MRI data. The fluid suppression in FLAIR images provide sufficient contrast variations between the images. The T2 FLAIR volume has dimensions $512 \times 512 \times 22$ voxels with spacing $0.4297 \times 0.4297 \times 6$ mm, and the T2-w FSE volume has similar dimensions and spacing. As a pre-processing step, both volumes are resampled to 1mm isotropic spacing. After registration, the outputs are resampled back to their native spacing and dimensions for evaluation.
Analysis: In addition to visual confirmations, PMM were computed and mean PMM values were obtained inside brain mask to quantify registration performance over entire population.

3. Results and Discussion

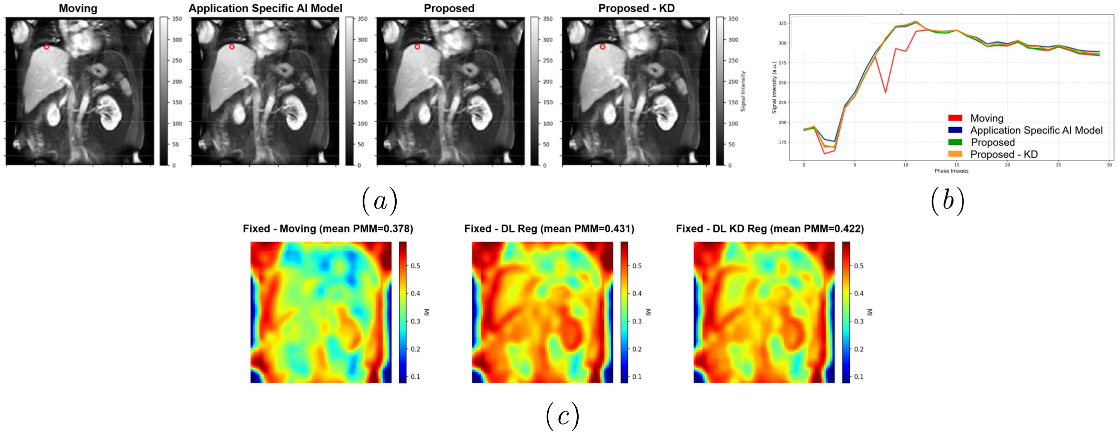
We present the outcomes of the proposed TTO framework across the experimental scenarios described in Section 2.4. The registered output obtained using the dense teacher model is referred to as **DL Reg**, the output from the knowledge-distilled student model is denoted as **DL KD Reg**, registration output using ANTs is referred to as **ANTs Reg**.

3.1. Dynamic Contrast Enhanced MRI (DCE-MRI)

The contrast uptake curves for red ROI in Figure 3(a) are shown in Figure 3(b). The sudden drop and spike in the contrast uptake curve (red) for moving image render them physiologically implausible. Post registration using the proposed method, both DL Reg (green curve) and DL KD Reg (orange curve), have corrected such abberations in the contrast uptake curve. The contrast uptake curves using proposed method matches the contrast uptake curve post correction using the application specific model (blue curve). This shows that the proposed generalizeable method performs as good as an AI model specifically trained for the application (hence, no compromise in performance). The green and orange curve matching each other indicates that the miniaturized KD model performance is equivalent to the dense model. Image alignment post registration was evaluated using PMMs (see Appendix C). Patch-wise normalized MI was computed for the image pairs: (*fixed, moving*), (*fixed, DL Reg*), and (*fixed, DL KD Reg*) on 3D patches of size $16 \times 16 \times 16$ voxels with a stride of 4. The PMMs were subsequently averaged across timepoints for each slice to obtain a mean PMM. The mean PMMs for the target slice is shown in Figure 3(c). The aggregated mean PMM values were: *fixed–moving* = 0.3801, *fixed–DL Reg* = 0.4291, and *fixed–DL KD Reg* = 0.4195. The increased values indicate higher alignment post registration. This is evident from increased values in air pockets around the liver, kidney and spleen regions. Visual assessments were also performed by overlaying the image pairs pre- and post registration using the proposed method (refer Appendix D Figure 6).

3.2. MRI-CT Multi-Modal Registration

Figure 4 illustrates qualitative results of registration using the proposed TTO framework. The comparison in Figure 4(a) highlights accurate alignment of anatomical structures (red



(a) (b)

(c)

Figure 3: Left to right for Figure 3(a) show moving, registered using application specific AI model, DL Reg, and DL KD Reg. The contrast uptake curves in Figure 3(b) indicate expected correction post registration. Higher PMM values observed post-registration (Figure 3(c)) indicate improved structural alignment.

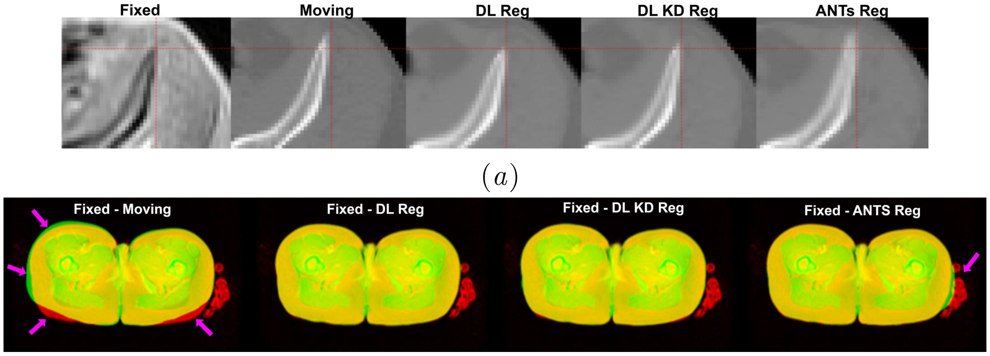
crosshair at bone tip) across modalities while preserving modality-specific characteristics. The MRI-CT image overlay for another slice is shown in Figure 4(b). The fixed image is shown in red and the registered image in green (blue channel suppressed). Regions of accurate correspondence appear yellow (red + green), while red-only and green-only regions indicate anatomical structures present exclusively in the fixed or registered image, respectively. It demonstrates fine alignment post proposed registration method (both body contour and internal tissues). PMMs for few slices from this data are shown in Appendix E (Fig. 7). These maps were computed using 3D patches of size 16 × 16 × 16 voxels with a stride of 4. Mean PMM values across all slices are: Fixed–Moving = 0.0986, Fixed–DL Reg = 0.1797, Fixed–KD Reg = 0.1680, and Fixed–ANTs Reg = 0.1742.

3.3. Pre- and Post-Contrast CT Registration

The results of the experiments discussed in Section 2.4.3 are discussed here. The corresponding organ segmentation overlaps (performed using (Wasserthal et al., 2023)) are considered for comparison: lung segmentation for *Data-1* and liver segmentation for *Data-2*, *Data-3*, and *Data-4*. Dice and IoU scores pre and post registration using proposed DL registration method (with and without KD), and ANTs are shown in Table 1. Post registration, the dice and IoU scores have improved across cases. The performance of the proposed method is similar to that of ANTs. Refer to Appendix G Figure 10 for additional qualitative assessments.

3.4. Brain MRI Registration

The results of the experiments discussed in Section 2.4.4 are discussed here. Results of an inter-subject cross-contrast registration is shown in Figure 5. The red square zooms into



(a)

(b)

Figure 4: MRI/CT registration. The zoomed-in region in Figure 4(a) shows good alignment of bone tip between MRI and registered CT images using proposed method. The False-color overlays shown in Figure 4(b) demonstrates spatial correspondence between fixed (MR in red) and registered (CT in green) volumes for a representative slice. Post registration there is good alignment of body contour. For both examples, proposed methods seems to perform slightly better than ANTs.

Data	**Dice Score**				**IoU Score**			
	F-M	F-R	F-R-KD	F-AR	F-M	F-R	F-R-KD	F-AR
Data-1	0.9834	0.9838	0.9928	0.9844	0.9673	0.9681	0.9857	0.9694
Data-2	0.9393	0.9737	0.9744	0.9733	0.8856	0.9488	0.9500	0.9480
Data-3	0.9576	0.9755	0.9691	0.9526	0.9186	0.9523	0.9401	0.9095
Data-4	0.8096	0.9675	0.9519	0.9724	0.6801	0.9371	0.9082	0.9463

Table 1: Comparison of Dice and IoU scores for primary organ segmentation from pre- and post-contrast CT datasets are shown here. Evaluations are performed across four settings: Fixed-Moving (F-M), Fixed-DL Reg (F-R), Fixed-DL Reg KD (F-R-KD), and Fixed-ANTs registered (F-AR).

region where ventricle ends are present in the 2D slice, and have sufficient misalignment post affine registration. The structural alignment is restored post registration using the proposed method. Based on the qualitative analysis for this example, proposed method performs slightly better than ANTs registration. For quantitative evaluation for all cases, PMM was computed for each fixed–moving/registered pair (see Appendix C for visual results). Mean PMM values were computed for each case inside the brain mask. Table 2 summarizes results for three scenarios: *Same-subject, Cross-contrast* ($n = 8$), *Inter-subject, Same-contrast* ($n = 24$), and *Inter-subject, Cross-contrast* ($n = 24$). Comparisons include the unregistered moving image (F–M), DL-registered outputs using the dense model (F–R) and KD-based model (F–R–KD), and ANTs registration (F–AR). Across all evaluation combinations, proposed registration method improved PMM values relative to the unregistered baseline. The DL Reg model achieved the highest PMM. The DL Reg KD model closely follows its performance. *Same-subject, Cross-contrast* yielded the highest post-registration PMM with low variability, while *Inter-subject, Cross-contrast* showed a relatively lower PMM, reflecting the increased complexity of inter-subject and cross-contrast alignment. Otherwise, one must note that same-contrast image pairs will have higher PMM values by virtue of contrast matching. Additional qualitative comparisons are shown in Appendix F.

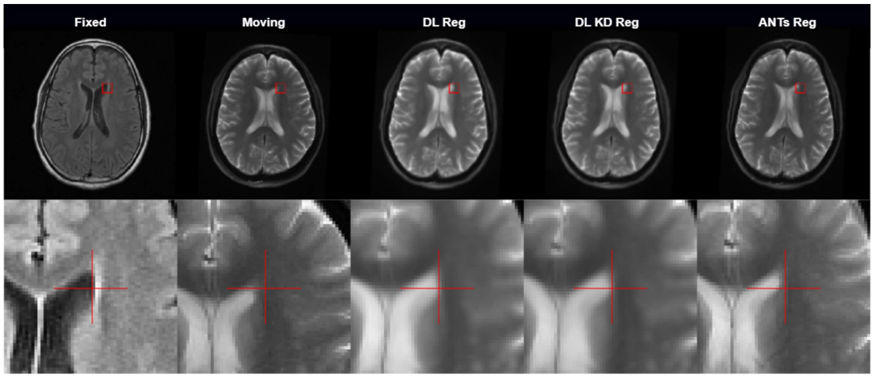


Figure 5: Qualitative comparison of inter-subject cross-contrast MRI registration.The zoomed-in region within the red bounding box highlights accurate alignment of the frontal horn of the lateral ventricles post registration.

	F-M		F-R		F-R-KD		F-AR	
Category	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
Inter-subject, Same-contrast	0.0887	0.0114	0.1707	0.0192	0.1618	0.0190	0.1389	0.0151
Inter-subject, Cross-contrast	0.0827	0.0077	0.1372	0.0096	0.1291	0.0092	0.1120	0.0092

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

4. Conclusion

In this work we proposed a DL-based registration framework which generalizes well across applications. Even though test time optimization concepts are used, by using model miniaturization approaches, the proposed method on an average takes half the time (or lesser) taken by popular analytical registration libraries. The ability of the proposed method to align structures across modalities and anatomies have been demonstrated using qualitative and quantitative measures.

Acknowledgments

The authors acknowledge the use of data from the following sources:

- The Cancer Genome Atlas Uterine Corpus Endometrial Carcinoma Collection (TCGA-UCEC) (Erickson et al., 2016). The results published here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.
- VA Research Precision Oncology Program – APOLLO (VAREPOP-APOLLO) (for Precision Oncology Program et al., 2024). Data used in this publication were generated by the Veterans Health Administration's Research for Precision Oncology Program and the Applied Proteogenomics Organizational Learning and Outcomes (APOLLO) Research Network.

References

Brian B. Avants, Nicholas J. Tustison, and Gang Song. Advanced normalization tools (ants). *Insight Journal*, pages 1–35, 2009. URL https://github.com/ANTsX/ANTs. https://www.insight-journal.org/browse/publication/681.

S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, W. Zhang, A. Leung, M. Kadoch, J. Shrager, A. Quon, D. Rubin, S. Plevritis, and S. Napel. Data for nsclc radiogenomics (version 4) [data set], 2017. URL https://doi.org/10.7937/K9/TCIA.2017.7hs46erv. Acknowledgment: Publications using data from this program are requested to include the following statement: "The results ¡published or shown¿ here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.".

Guha Balakrishnan, Amy Zhao, Mert Sabuncu, John Guttag, and Adrian V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *IEEE TMI: Transactions on Medical Imaging*, 38:1788–1800, 2019.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.

Basar Demir, Lin Tian, Thomas Hastings Greer, Roland Kwitt, Francois-Xavier Vialard, Raul San Jose Estepar, Sylvain Bouix, Richard Jarrett Rushmore, Ebrahim Ebrahim, and Marc Niethammer. multigradicon: A foundation model for multimodal medical image registration. *arXiv preprint arXiv:2408.00221*, 2024.

B. J. Erickson, D. Mutch, L. Lippmann, and R. Jarosz. The cancer genome atlas uterine corpus endometrial carcinoma collection (tcga-ucec) (version 4) [data set], 2016. URL https://doi.org/10.7937/K9/TCIA.2016.GKJ0ZWAC. Acknowledgment: Publications using data from this program are requested to include the following statement: "The results ¡published or shown¿ here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.".

The Research for Precision Oncology Program, the Applied Proteogenomics Organizational Learning, and Outcomes (APOLLO) Research Network. Va research precision oncology program – apollo (varepop-apollo) (version 3) [data set], 2024. URL https://doi.org/10.7937/ghkn-md15. Acknowledgment: Data used in this publication were generated by the Veterans Health Administration's Research for Precision Oncology Program and the Applied Proteogenomics Organizational Learning and Outcomes (APOLLO) Research Network.

Yunguan Fu, Nina Montaña Brown, Shaheer U. Saeed, Adrià Casamitjana, Zachary M. C. Baum, Rémi Delaunay, Qianye Yang, Alexander Grimwood, Zhe Min, Stefano B. Blumberg, Juan Eugenio Iglesias, Dean C. Barratt, Ester Bonmati, Daniel C. Alexander, Matthew J. Clarkson, Tom Vercauteren, and Yipeng Hu. Deepreg: a deep learning toolkit for medical image registration. *Journal of Open Source Software*, 5(55):2705, 2020. doi: 10.21105/joss.02705. URL https://doi.org/10.21105/joss.02705.

Derek LG Hill, Philipp G Batchelor, Mark Holden, and David J Hawkes. Medical image registration. *Physics in Medicine & Biology*, 46(3):R1, 2001.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. URL https://arxiv.org/abs/1503.02531.

Malte Hoffmann, Benjamin Billot, Douglas N. Greve, Juan Eugenio Iglesias, Bruce Fischl, and Adrian V. Dalca. Synthmorph: Learning contrast-invariant registration without acquired images. *IEEE Transactions on Medical Imaging*, 41(3):543–558, March 2022. ISSN 1558-254X. doi: 10.1109/tmi.2021.3116879. URL http://dx.doi.org/10.1109/TMI.2021.3116879.

Tai Ma, Suwei Zhang, Jiafeng Li, and Ying Wen. Iirp-net: Iterative inference residual pyramid network for enhanced image registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11546–11555, 2024.

JB Antoine Maintz and Max A Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

A. W. Moawad, D. Fuentes, A. Morshid, A. M. Khalaf, M. M. Elmohr, A. Abusaif, J. D. Hazle, A. O. Kaseb, M. Hassan, A. Mahvash, J. Szklaruk, A. Qayyom, and K. Elsayes. Multimodality annotated hcc cases with and without advanced imaging segmentation [data set], 2021. URL https://doi.org/10.7937/TCIA.5FNA-0924.

Francisco PM Oliveira and João Manuel RS Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2):73–93, 2014.

Anna Reithmeir, Veronika Spieker, Vasiliki Sideri-Lampretsa, Daniel Rueckert, Julia A. Schnabel, and Veronika A. Zimmer. From model based to learned regularization in medical image registration: A comprehensive review. *Medical Image Analysis*, 108: 103854, 2026. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2025.103854. URL https://www.sciencedirect.com/science/article/pii/S1361841525004001.

Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: a survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013.

Lin Tian, Hastings Greer, Roland Kwitt, Francois-Xavier Vialard, Raul San Jose Estepar, Sylvain Bouix, Richard Rushmore, and Marc Niethammer. unigradicon: A foundation model for medical image registration. *arXiv preprint arXiv:2403.05780*, 2024.

Jakob Wasserthal, Holger Breit, Martin Meyer, and et al. Totalsegmentator: Robust segmentation of 104 anatomical structures in ct images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023. doi: 10.1148/ryai.230024. URL https://github.com/wasserth/TotalSegmentator.

Yunlu Zhang, Xue Wu, H Michael Gach, Harold Li, and Deshan Yang. Groupregnet: A groupwise one-shot deep learning-based 4d image registration method. *Physics in Medicine & Biology*, 66(4):045030, 2021.

Appendix A. Multi-Axis Mutual Information Similarity Loss

We adopt the Multi-Axis Mutual Information (MultiAxisMI) approach instead of a full 3D Mutual Information (MI) loss. The similarity loss is defined as:

$$\mathcal{L}_{\text{MultiAxisMI}} = \frac{1}{D+H+W}\left(\sum_{d=1}^{D}\mathcal{M}(I_f^{(d)}, I_m^{(d)}\circ\phi)+\sum_{h=1}^{H}\mathcal{M}(I_f^{(h)}, I_m^{(h)}\circ\phi)+\sum_{w=1}^{W}\mathcal{M}(I_f^{(w)}, I_m^{(w)}\circ\phi)\right), \tag{3}$$

where:

- $\mathcal{M}(\cdot,\cdot)$ denotes the mutual information between corresponding slices.
- $I_f^{(d)}, I_m^{(d)}$ represent slices along the depth axis, and similarly for height (h) and width (w).
- D, H, W are the number of slices along each axis.

This formulation ensures robustness to local intensity variations and captures anatomical alignment across all three dimensions.

Appendix B. Normalized Cross-Correlation (NCC) Loss

The Normalized Cross-Correlation (NCC) loss used in our experiments is defined as:

$$\mathcal{L}_{\text{NCC}}(I_f, I_m\circ\phi) = -\frac{\sum_x (I_f(x)-\bar{I}_f)(I_m(\phi(x))-\bar{I}_m)}{\sqrt{\sum_x (I_f(x)-\bar{I}_f)^2}\sqrt{\sum_x (I_m(\phi(x))-\bar{I}_m)^2}}, \tag{4}$$

where $\bar{I}_f$ and $\bar{I}_m$ denote the mean intensities of the fixed and warped moving images, respectively.

Appendix C. Patchwise Mutual Information over ROI or Full Volume

To evaluate local registration quality, we compute a *patchwise mutual information (MI) map* over a specified spatial domain, which can be either a predefined anatomical region of interest (ROI) or the entire image volume.

Let F and M denote the fixed and moving (or registered) 3D images with voxel intensities $F(\mathbf{r}), M(\mathbf{r})$ at spatial location $\mathbf{r} \in \Omega \subset \mathbb{Z}^3$. Define a binary mask $\mathcal{D} \subseteq \Omega$ representing the domain of interest (e.g., ROI or full volume). A voxel belongs to the domain iff $\mathbf{r} \in \mathcal{D}$.

We define a family of overlapping cubic patches $\{P_k\}_{k=1}^{K}$ covering Ω. For each patch P_k, we consider its intersection with the domain:

$$P_k^{\mathcal{D}} = P_k \cap \mathcal{D}.$$

If $|P_k^{\mathcal{D}}|$ exceeds a minimal sample threshold, we estimate the empirical joint intensity distribution $p_{F,M}^{(k)}(f,m)$ over voxels in $P_k^{\mathcal{D}}$, and compute the local mutual information:

$$I_k = I(F; M \mid P_k^{\mathcal{D}}) = \sum_{f,m} p_{F,M}^{(k)}(f,m) \, \log \frac{p_{F,M}^{(k)}(f,m)}{p_F^{(k)}(f) \, p_M^{(k)}(m)}.$$

Assigning I_k to all voxels in $P_k^{\mathcal{D}}$ and averaging over overlapping patches produces a continuous MI field:

$$\mathrm{MI}_{\mathcal{D}}(\mathbf{r}) = \frac{1}{N(\mathbf{r})} \sum_{k:\, \mathbf{r} \in P_k^{\mathcal{D}}} I_k, \quad \mathbf{r} \in \mathcal{D},$$

where $N(\mathbf{r})$ is the number of domain-containing patches covering $\mathbf{r}$.

This spatial MI map localizes registration quality: higher values indicate stronger statistical dependence and better alignment. Patchwise aggregation reveals heterogeneous performance (e.g., cortical vs. deep structures), which a single global MI would conceal. Mean statistics of $\mathrm{MI}_{\mathcal{D}}$ before and after registration (or across methods) provide interpretable improvement scores (e.g., $\Delta\mathrm{MI}_{\mathcal{D}}$).

Appendix D. Qualitative analysis of DCE MRI registration

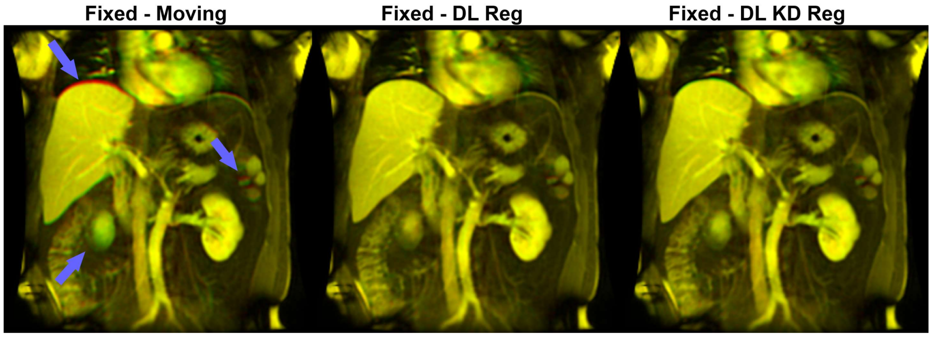


Figure 6: Qualitative overlays for liver DCE-MRI registration.

Appendix E. Qualitative analysis of MRI-CT Multi-Modal Registration

Figure 7 illustrates patch-wise MI maps for several slices, computed using a patch size of $16 \times 16 \times 16$ voxels and a stride of 4, following the approach described in Section C.

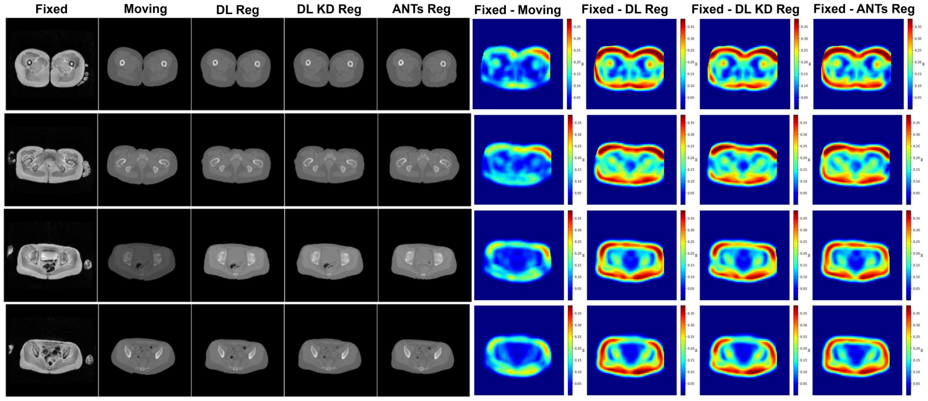


Figure 7: MRI/CT Registration - patchwise MI Maps of several slices

Appendix F. Qualitative analysis of Cross-Contrast Brain-MR Registration

Figure 8 illustrates patch-wise MI maps for several slices, computed using a patch size of $16 \times 16 \times 16$ voxels and a stride of 4, (refer Appendix C. The computation is performed within the brain mask to ensure anatomical relevance.

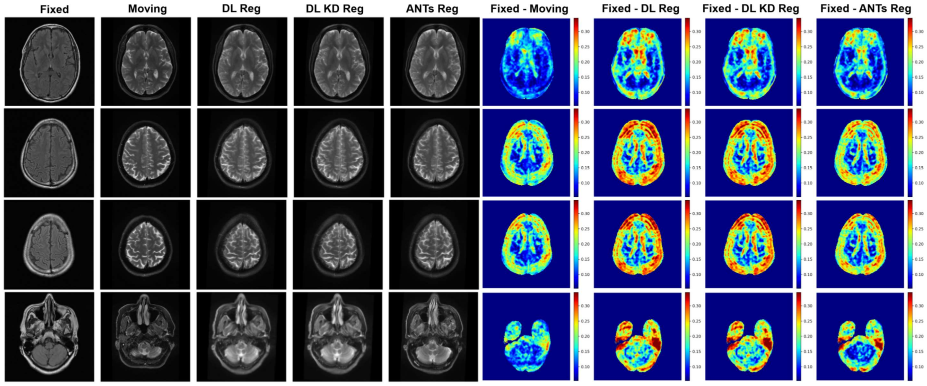


Figure 8: Inter Subject Cross Contrast Brain MRI Registration - patchwise MI Maps of several slices

Figure 9 shows checkerboard overlays of inter-subject cross-contrast brain MR images before and after registration, visually emphasizing the improved alignment achieved through the proposed method.

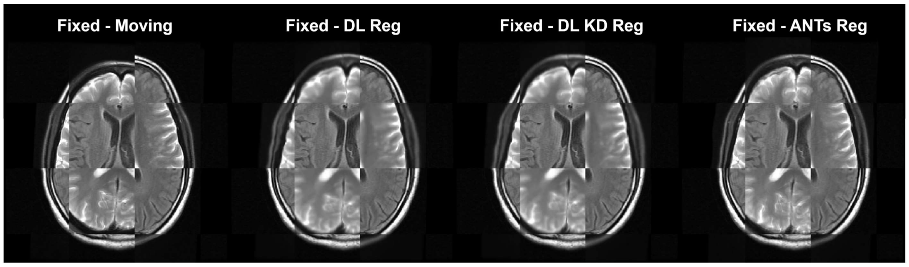
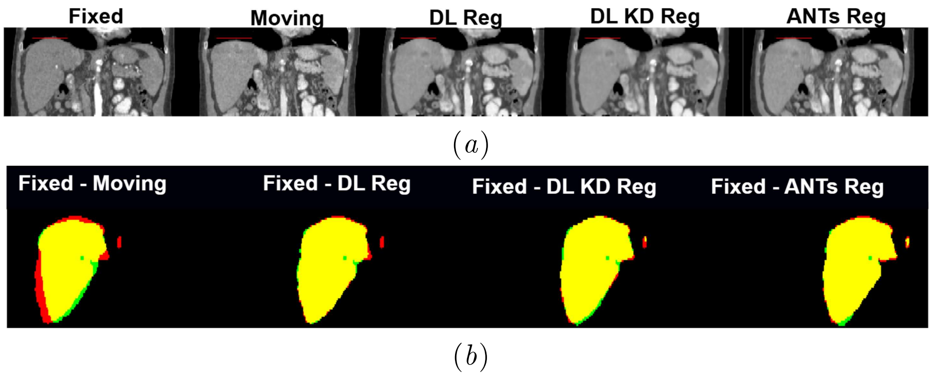


Figure 9: Checkerboard overlays illustrating local alignment between fixed and registered volumes for a slice. Alternating tiles display either the fixed image or the corresponding registered image, facilitating localized visual assessment of registration accuracy. Smooth transitions across tile boundaries indicate accurate local registration, whereas visible seams or discontinuities reveal residual misalignment.

Appendix G. Pre and Post Contrast CT Qualitative evaluation

For Data-4 (Section 2.4.3), Figure 10 presents qualitative results of pre- and post-contrast CT registration, illustrating improved liver alignment and segmentation consistency after registration.



(a)

(b)

Figure 10: Pre- and post-contrast CT registration. Figure 10(a) illustrates accurate alignment of the liver dome between the pre-contrast CT and the registered post-contrast CT using the proposed method. The segmentation overlays in Figure 10(b) further confirm spatial correspondence between the fixed and registered volumes for a representative slice, demonstrating improved alignment of liver segmentation after registration.

Appendix H. Knowledge Distillation Framework

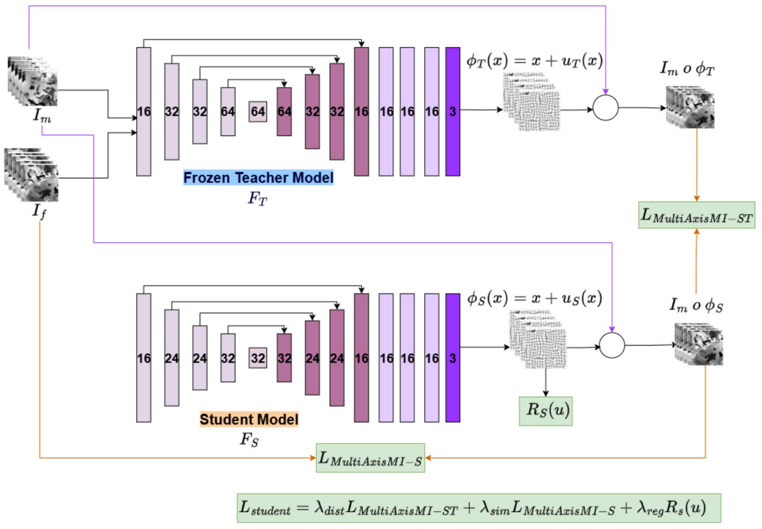


Figure 11: Knowledge Distillation framework - The trained and frozen teacher model guides the student using output based distillation, image similarity and smoothness losses.