Global to Local Registration: Transferring Pretrained Registration Models to CT/MR–Ultrasound via Anatomical Supervision

Gabriella d'Albenzio[1], Mengting Liu[2†], Shixing Ma[2†], Chunna Yang[2], Yuhao Wei[3,14], Shrisharanyan Vasu[4], Aniketh Vijesh[4], Xihan Ma[5], Zachary P. Taylor[6], Tanmoy Sarkar Pias[7], Başar Demir[8], Miklós Gyöngy[9], Yasin Ceran[10], Marc Niethammer[11], Tina Kapur[12], Stephen Aylward[13], Gabor Fichtinger[1], Zhe Min[2,14,15], and Mirabela Rusu[16]

[1] School of Computing, Queen's University, Kingston, Canada
[2] School of Control Science and Engineering, Shandong University, Jinan, China
[3] School of Life Science and Technology, Xidian University, China
[4] Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, India
[5] Worcester Polytechnic Institute, Worcester MA 01609, USA
[6] Artificial Intelligence Imaging Research Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
[7] Stanford Medicine, Stanford University, Stanford, CA, USA
[8] University of California San Diego, San Diego, CA, USA
[9] Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary
[10] School of Information Systems and Technology, Lucas School of Business, San José State University, San José, CA, USA
[11] Department of Computer Science and Engineering and Department of Neurological Surgery, University of California San Diego, La Jolla, CA, USA
[12] Department of Radiology, Harvard Medical School, Boston, MA, USA
[13] Medical Devices, Healthcare and Life Sciences, NVIDIA, USA
[14] Shenzhen Loop Area Institute, Shenzhen, China
[15] UCL Hawkes Institute, Department of Medical Physics and Biomedical Engineering, University College London, London, UK
[16] Department of Radiology, Department of Urology and Department of Biomedical Data Science, Stanford University, Stanford, CA, USA

[†]These authors contributed equally.

Abstract. Transferring pretrained registration models to ultrasound is challenging because of weak cross-modal correspondence, limited field of view, and large initial pose differences. We propose US-multiGradICON, a global-to-local framework for CT–ultrasound and MR–ultrasound registration. Surface-based PCA–ICP initialization provides robust global alignment, after which pretrained multiGradICON is finetuned with image-similarity, gradient inverse-consistency, and Dice supervision to predict the final deformation field in a single feed-forward pass. We evaluate the method on kidney CT–ultrasound and prostate MRI–ultrasound registration. US-multiGradICON outperformed all competing methods across both applications. Compared with the strongest baseline, it significantly

reduces target registration error from 4.92 to 2.07 mm for kidney registration and from 4.46 to 4.40 mm for prostate registration. These results demonstrate that robust surface initialization and anatomical supervision enable effective transfer of pretrained multimodal registration models to distinct CT/MR–ultrasound applications.

Keywords: Ultrasound · Medical image registration · Multimodal registration · Deep learning · Domain adaptation

1 Introduction

Ultrasound (US) is widely used for diagnostic purposes and to guide interventions as it provides real-time, non-invasive, cost-effective, and non-ionizing imaging at the bedside [1]. Despite its advantages, ultrasound images are noisy, capture a limited field of view, suffer from acoustic shadowing, speckle and may have low soft-tissue contrast [1]. Often, ultrasound-guided interventions require registering with high resolution imaging modalities, like CT or MRI, to visualize and project complementary information. For example, kidney or prostate biopsies are performed under the real-time guidance of ultrasound, with lesions projected from computed tomography (CT) or Magnetic Resonance Imaging (MRI) onto ultrasound to enable targeting. Yet, CT/MRI–US registration remains challenging because these modalities have different physical contrast mechanisms and because reliable cross-modal landmarks are difficult to define consistently.

Traditional registration approaches estimate rigid, affine, or more general transforms through iterative pairwise optimization of handcrafted similarity objectives with deformation regularization. Even when using multimodal scoring functions, like mutual information, and advanced parameterizations of deformations, e.g., via B-splines or SyN [2–4], these approaches remain sensitive to initialization, optimized metrics, regularization, and are often computationally intensive. Learning-based methods replace iterative optimization with neural networks that directly predict deformation fields. QuickSilver and VoxelMorph introduced supervised and unsupervised registration frameworks respectively, while TransMorph used transformers to capture long-range spatialcorrespondence [5–7]. SynthMorph improved contrast-invariant registration through training on syntheticimages [8]. However, many learning-based methods remain specific to a particular anatomy or modality. Their performance may therefore degrade when applied to heterogeneous CT/MR–US registration tasks [9]. More recently, uniGRADICON and multiGradICON introduced general-purpose learning-based registration approaches through large-scale training across multiple anatomies and imaging modalities [9, 10]. These pretrained models provide a strong basis for transfer to new registration tasks. However, ultrasound was not included in the training data of multiGradICON. Direct application to CT/MR–US registration is therefore difficult because ultrasound has a limited field of view, modality-specific artifacts, weak intensity correspondence with CT or MRI, and potentially large initial pose differences. Existing US–CT and US–MR methods

often address these challenges using landmarks, vascular structures, segmentations, or task-specific registration pipelines [1, 11].

Here, we present US-multiGradICON, a framework for adapting multiGradICON to CT/MR–US registration. We adapt a pretrained multimodal registration model to CT/MR–US registration without training a separate network from scratch for each anatomy. We hypothesize that the pretrained multiGradICON model can be effectively adapted to CT/MR–US registration when combined with robust organ-surface initialization and anatomy-guided finetuning. The surface initialization first establishes global anatomical correspondence despite large pose differences and limited overlap. Finetuning then adapts the pretrained deformation model to ultrasound using modality-robust image descriptors, inverse-consistency regularization, and segmentation-based anatomical supervision. Our main contributions are:

1. We introduced US-multiGradICON, an adaptation of pretrained multiGradICON for CT–US and MR–US registration, enabling accurate ultrasound registration without training anatomy-specific models from scratch.
2. We developed an ultrasound-specific registration pipeline that combines robust organ-surface initialization with anatomy-guided finetuning, allowing the model to handle large pose differences, limited overlap, and weak multimodal image correspondence.
3. We provide a systematic evaluation against previous registration methods and establish a MONAI-based benchmark for evaluating pretrained and future CT/MR–US registration models. The benchmark implementation and source code are publicly available at `https://github.com/Project-MONAI/wg-ultrasound/tree/main/registrator`.

2 Method

Let $I_S : \Omega_S \to \mathbb{R}$ denote the source CT or MR image and $I_T : \Omega_T \to \mathbb{R}$ the target ultrasound image, with corresponding anatomical masks S_S and S_T. The proposed global-to-local framework consists of two components: rigid surface-based alignment followed by deformable refinement with multiGradICON (Fig. 1). The rigid transformation T_{rigid} follows a forward point-transform convention and maps source coordinates $x_S \in \Omega_S$ into the ultrasound coordinate system:

$$x_T = T_{\text{rigid}}(x_S) = Rx_S + t, \tag{1}$$

where $R \in SO(3)$ is a rotation matrix and $t \in \mathbb{R}^3$ is a translation vector. The rigidly aligned source image and target ultrasound are subsequently provided to multiGradICON, which estimates the remaining local deformation.

2.1 Rigid Surface-Based Alignment

Direct deformable registration is unreliable when the source and target anatomies have large differences in position or orientation. Such misalignment can place

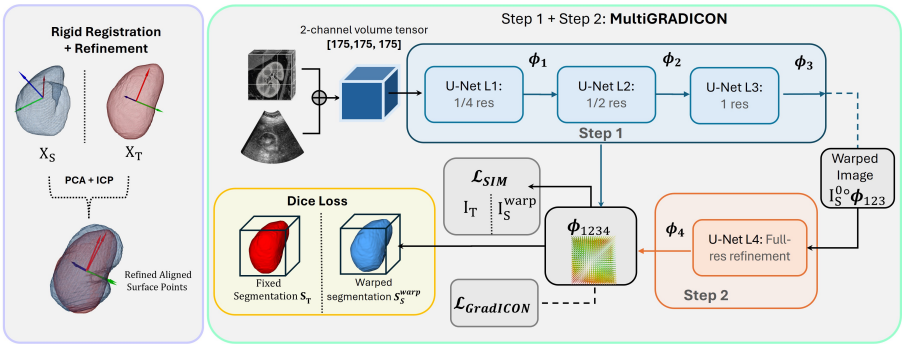

Fig. 1. Overview of the proposed global-to-local registration framework. Source and target surface point sets, X_S and X_T, are first aligned using PCA and ICP to estimate T_{rigid}. The rigidly aligned source image I_S^0 and target ultrasound image I_T are then passed to multiGradICON as a two-channel input. The network estimates transformation maps ϕ_1–ϕ_4 at progressively finer resolutions, which are composed to produce the final deformable registration. Dice supervision is applied between the warped source segmentation S_S^{warp} and the target segmentation S_T.

corresponding structures outside the effective capture range of the deformable model, leading to incorrect correspondences. We therefore estimate a rigid transformation before deformable refinement. Surface point sets $X_S = \{\mathbf{x}_i \in \mathbb{R}^3\}_{i=1}^N$ and $X_T = \{\mathbf{y}_j \in \mathbb{R}^3\}_{j=1}^M$ are extracted from the source and target surfaces S_S and S_T, respectively, and are represented in physical coordinates. We first align their centroids and principal axes using Principal Component Analysis (PCA) [12]. Because the direction of each PCA axis is ambiguous, the source and target bases may differ by axis sign even when they represent the same anatomical orientation. After expressing both PCA bases as right-handed coordinate systems, we evaluate the four valid sign configurations that preserve orientation.

For each configuration k, the candidate rotation and translation are $R_k = U_T D_k U_S^\top$, $t_k = \mu_T - R_k \mu_S$ where U_S and U_T are the source and target PCA bases, μ_S and μ_T are their centroids, and D_k is a diagonal sign matrix satisfying $\det(D_k) = 1$. The candidate producing the lowest symmetric surface distance is retained as T_{PCA}.

The PCA alignment is then refined using Iterative Closest Point (ICP), which alternates between establishing closest-point correspondences and updating the rigid transformation [13]. The final rigid $T_{\text{rigid}} = T_{\text{ICP}} \circ T_{\text{PCA}}$ provides a consistent global alignment while leaving local anatomical differences to the deformable model.

2.2 Deformable Refinement with multiGradICON

Because T_{rigid} maps source points into the target coordinate system, image resampling uses its inverse. The rigidly aligned source image and mask are defined

on the ultrasound grid as

$$I_S^0(x) = I_S\left(T_{\text{rigid}}^{-1}(x)\right), \qquad S_S^0(x) = S_S\left(T_{\text{rigid}}^{-1}(x)\right), \quad x \in \Omega_T. \tag{2}$$

The input to multiGradICON consists of the resampled volumes ($175 \times 175 \times 175$), I_S^0 and I_T:

$$[I_S^0, I_T] \in \mathbb{R}^{2\times 175\times 175\times 175}. \tag{3}$$

We initialized the network from the pretrained second-stage multiGradICON configuration and finetuned it on CT–US and MR–US image pairs [9,10]. multiGradICON estimates transformation maps ϕ_1, ϕ_2, and ϕ_3 at one-quarter, one-half, and full resolution, respectively (Fig. 1). The second stage estimates an additional full-resolution refinement transformation, ϕ_4. These transformations are composed sequentially to obtain the final backward sampling transformation $\phi_\theta : \Omega_T \to \Omega_S$. The deformable registration model is optimized using

$$\mathcal{L} = \mathcal{L}_{\text{sim}}\left(I_S^{\text{warp}}, I_T\right) + \lambda \mathcal{L}_{\text{GradICON}} + \gamma\left[1 - \text{Dice}\left(S_S^{\text{warp}}, S_T\right)\right], \tag{4}$$

where $\mathcal{L}_{\text{sim}}$ denotes the image-similarity loss and is instantiated as either MIND-SSC or -LNCC^2. MIND-SSC compares modality-robust local self-similarity descriptors between I_S^{warp} and I_T [14], whereas LNCC^2 measures local intensity correspondence while accounting for both positive and negative correlations. The GradICON term enforces gradient inverse consistency between the forward and backward transformations, while the Dice term directly supervises anatomical alignment. For experiments without segmentation supervision, γ is set to zero. All loss terms are evaluated within the target region of interest to limit the contribution of background voxels. Following multiGradICON, we evaluated both MIND-SSC and LNCC^2 as image-similarity losses during finetuning [9].

3 Experiments

We conducted three sets of experiments to evaluate registration accuracy, analyze the contribution of each design choice, and assess deformation quality.

1. **Comparison to prior art.** We compared our approach with ConvexAdam [15], TransMorph [7], VoxelMorph [6], LocalNet [16], and MIDIR [17]. We selected these methods on the basis of (i) representation of different registration paradigms, including optimization-based, CNN-based, Transformer-based, and diffeomorphic approaches; (ii) suitability for our problem setting; and (iii) reproducibility using the original manuscripts and released code where available. We made necessary adaptations to account for differences in the dataset, and we finetuned the hyperparameters to allow fair comparison with our approach. For fairness, all non-rigid registration methods used the same rigid surface-based initialization before the non-rigid registration stage.
2. **Ablation studies.** We assessed the contribution of the training losses and multimodal training strategy. We first finetuned multiGradICON with each

similarity loss, with and without Dice segmentation supervision, and varied the Dice loss weight to assess its effect. The architecture and gradient inverse-consistency regularization were kept fixed, allowing the effects of the similarity loss, segmentation supervision, and Dice weighting to be isolated. We then compared task-specific training, using separate CT–US and MR–US models, with joint multimodal training, using a single model trained on both modality pairs.

3. **Qualitative analysis**. We examined warped images, deformation grid overlays, and difference maps before and after registration to assess anatomical alignment and transformation plausibility. This analysis complemented the quantitative metrics by revealing under-registration, residual misalignment, excessive local deformation, and anatomically inconsistent warping.

Datasets and Preprocessing. We included the TRUSTED kidney dataset [18] and the μRegPro prostate dataset [19]. TRUSTED provides paired transabdominal 3D US and CT kidney images with kidney masks and anatomical landmarks, while μRegPro provides paired prostate MRI and TRUS with prostate masks. Volumes were resampled to $175 \times 175 \times 175$ voxels, using trilinear interpolation for images and nearest-neighbor interpolation for masks. Left and right kidneys were treated as separate registration pairs and split into training and test sets at a 9:1 ratio, while the prostate MRI–TRUS pairs were split at a 7:3 ratio. For the kidney landmarks we only used the superior pole, inferior pole, and lateral renal cortical border, as they are more consistently identifiable across CT and US, and excluded other landmarks due to ambiguous correspondences.

Implementation Details. We initialized our model from the second-stage multiGradICON configuration and finetuned it with a batch size of 1 on a single GPU on the two datasets for 300 epochs, using an Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 0) with a learning rate of 5×10^{-5}. All network parameters were updated during fine-tuning; no layers were frozen. The gradient inverse-consistency regularization weight is set to $\lambda = 1.5$, following the default GradICON setting [10]. All variants were evaluated in a single feed-forward pass, without per-instance optimization.

Evaluation Metrics. Registration accuracy was evaluated using TRE, Dice Coefficient, Chamfer distance, and HD95. TRE measured landmark alignment error, and Dice measured the overlap between the warped moving mask and the fixed ultrasound mask. Chamfer distance evaluated global surface correspondence by averaging nearest-neighbor distances between boundary point clouds sampled from the warped moving and fixed organ masks (up to 5000 points/mask). Metrics were computed after warping CT/MRI into the ultrasound space.

4 Results and Discussion

Baseline Comparison. US-multiGradICON achieved the best performance across most metrics and both anatomies (Table 1). For kidney registration, the largest gain was observed in TRE, which was significantly reduced by approximately 58% relative to the strongest baseline. This improvement is important be-

cause TRE directly measures the accuracy of local landmark alignment, suggesting that the proposed method improves fine-scale anatomical correspondence, as well as surface overlap and boundary agreement. For prostate registration, US-multiGradICON achieved the highest performance across all metrics. The most substantial improvement was observed in surface alignment, with a significant 38% reduction in Chamfer distance compared to the strongest baseline, while also retaining the best landmark accuracy, overlap, and boundary agreement. This comprehensive analysis demonstrates consistent performance, indicates that the proposed surface initialization and anatomy-guided finetuning yield accurate and anatomically coherent registration across both CT–US and MR–US tasks.

Table 1. Comparison with previous registration methods on the prostate and kidney datasets.

	Prostate				Kidney			
Method	TRE ↓	Chamfer ↓	Dice ↑	HD95 ↓	TRE ↓	Chamfer ↓	Dice ↑	HD95 ↓
ConvexAdam [15]	6.14 ± 3.03*	4.09 ± 1.11*	0.72 ± 0.07*	10.69 ± 2.38*	6.16 ± 0.58*	5.53 ± 0.84*	0.85 ± 0.04	7.00 ± 0.99
ConvexAdam+nnU-Net [15]	4.73 ± 2.08	2.31 ± 0.69*	0.877 ± 0.02*	3.91 ± 1.28	7.22 ± 1.70*	5.27 ± 1.13*	0.86 ± 0.03	7.12 ± 2.43
VoxelMorph [6]	4.51 ± 2.06*	2.27 ± 0.46*	0.88 ± 0.03*	4.30 ± 1.03*	5.87 ± 1.54*	**1.99 ± 0.73**	0.82 ± 0.04	6.25 ± 2.63
MIDIR [17]	4.46 ± 2.03*	2.48 ± 0.58*	0.86 ± 0.03*	4.39 ± 1.08*	7.44 ± 2.21*	7.15 ± 2.08*	0.82 ± 0.04	12.72 ± 5.05*
TransMorph [7]	4.65 ± 2.05*	2.29 ± 0.54*	0.88 ± 0.03*	4.26 ± 1.17*	7.69 ± 3.59*	4.97 ± 1.44*	0.86 ± 0.04	7.13 ± 3.73
LocalNet [16]	4.71 ± 2.01*	2.46 ± 0.49*	0.87 ± 0.03*	4.05 ± 1.03*	4.92 ± 1.78*	5.62 ± 1.81*	0.86 ± 0.04	6.06 ± 1.84
US-multiGradICON	**4.40 ± 3.76**	**1.40 ± 0.36**	**0.89 ± 0.03**	**3.77 ± 1.15**	**2.07 ± 1.54**	2.03 ± 0.63	**0.89 ± 0.03**	**5.25 ± 1.82**

Note. Bold and underlined values indicate the best and second-best performance, respectively. * denotes a statistically significant difference from US-multiGradICON according to a paired two-sided t-test (raw $p < 0.05$).

Ablation Study. Table 2 evaluates the effects of deformable registration, Dice supervision, and finetuning strategy. The PCA–ICP row reports performance after rigid alignment only, before deformable registration. All other experiments use the same multiGradICON architecture and inverse-consistency regularization, while varying the initialization, similarity loss, and Dice weight. Task-specific models were finetuned separately for kidney and prostate from the pretrained multiGradICON checkpoint, whereas joint finetuning used both organs in a single training set. This joint setting also enabled a controlled comparison between pretrained initialization and training from scratch with randomly initialized weights.

Effect of Dice supervision. Adding Dice supervision improved anatomical overlap and boundary alignment across both organs and both similarity losses (Table 2). The effect was most pronounced for prostate registration with LNCC^2, where Dice supervision more than doubled the Dice score (from 42% to 89%) and substantially reduced TRE, Chamfer distance, and HD95; gains for the kidney and for MIND-SSC were more modest but consistently positive. Across Dice weights $\gamma \in \{0.3, 0.5, 1.0\}$, prostate registration was consistently best or near-best at $\gamma = 1$, whereas kidney results varied more across weights with no single value dominating. We therefore fixed $\gamma = 1$ for all subsequent experiments, favoring the setting that performed best on the prostate task while remaining competitive on the kidney task, and avoiding an additional tuning step.

Table 2. Ablation of task-specific and joint multimodal fine-tuning.

Method				Kidney				Prostate			
				TRE ↓	Chamfer ↓	Dice ↑	HD95 ↓	TRE ↓	Chamfer ↓	Dice ↑	HD95 ↓
PCA–ICP rigid alignment				4.85 ± 1.71	4.24 ± 0.68	0.85 ± 0.04	6.05 ± 1.81	4.33 ± 3.78	1.77 ± 0.39	0.86 ± 0.04	4.50 ± 1.11
Task-specific fine-tuning (Dice-weight ablation)											
Training configuration				Kidney				Prostate			
Pretr.	Rand.	Similarity	γ	TRE ↓	Chamfer ↓	Dice ↑	HD95 ↓	TRE ↓	Chamfer ↓	Dice ↑	HD95 ↓
✓	✗	LNCC^2	0	6.49 ± 1.55	4.60 ± 0.61	0.80 ± 0.07	8.22 ± 1.63	9.49 ± 5.00	12.62 ± 2.53	0.42 ± 0.08	24.28 ± 3.83
✓	✗	LNCC^2	0.3	5.12 ± 1.06	4.10 ± 0.39	0.85 ± 0.06	6.20 ± 2.15	**4.49 ± 3.75**	1.47 ± 0.32	0.88 ± 0.02	3.84 ± 1.05
✓	✗	LNCC^2	0.5	5.39 ± 1.63	4.14 ± 0.38	0.86 ± 0.04	5.83 ± 1.43	4.57 ± 3.81	1.43 ± 0.34	0.89 ± 0.02	3.80 ± 1.11
✓	✗	LNCC^2	1.0	5.13 ± 1.42	4.12 ± 0.25	0.86 ± 0.04	6.34 ± 1.76	4.60 ± 3.86	**1.39 ± 0.33**	**0.89 ± 0.02**	**3.71 ± 1.09**
✓	✗	MIND-SSC	0	4.80 ± 0.81	4.07 ± 0.20	0.85 ± 0.05	6.54 ± 1.41	5.69 ± 3.66	2.62 ± 0.48	0.77 ± 0.05	7.00 ± 1.61
✓	✗	MIND-SSC	0.3	4.87 ± 1.43	4.09 ± 0.33	**0.87 ± 0.03**	**5.27 ± 1.30**	4.94 ± 3.75	1.49 ± 0.41	0.89 ± 0.03	3.90 ± 1.05
✓	✗	MIND-SSC	0.5	**4.55 ± 1.23**	**4.05 ± 0.41**	**0.87 ± 0.03**	5.34 ± 1.21	4.89 ± 3.77	1.45 ± 0.39	0.89 ± 0.02	3.84 ± 1.09
✓	✗	MIND-SSC	1.0	4.85 ± 1.64	4.12 ± 0.37	**0.87 ± 0.04**	5.48 ± 1.54	4.87 ± 3.77	1.44 ± 0.37	0.89 ± 0.03	3.81 ± 1.09
Joint multimodal fine-tuning											
✓	✗	LNCC^2	1	2.07 ± 1.54	**2.03 ± 0.63**	**0.89 ± 0.03**	**5.25 ± 1.82**	**4.40 ± 3.76**	1.40 ± 0.36	**0.89 ± 0.03**	3.77 ± 1.15
✓	✗	MIND-SSC	1	2.23 ± 1.62	2.04 ± 0.53	0.88 ± 0.03	5.32 ± 1.65	4.48 ± 3.75	**1.39 ± 0.38**	**0.89 ± 0.03**	**3.72 ± 1.15**
✗	✓	LNCC^2	1	2.19 ± 1.45	2.31 ± 0.60	0.87 ± 0.03	5.68 ± 1.58	4.42 ± 3.77	1.55 ± 0.29	0.88 ± 0.02	4.00 ± 0.95
✗	✓	MIND-SSC	1	**1.70 ± 1.28**	2.29 ± 0.48	0.87 ± 0.02	5.41 ± 1.34	**4.40 ± 3.82**	1.50 ± 0.31	0.88 ± 0.02	3.93 ± 0.97

Note. Pretr. and *Rand.* denote pretrained and randomly initialized networks. PCA–ICP denotes the rigid-alignment baseline prior to deformable registration. γ denotes the Dice-loss weight. Bold and underlined values denote the best and second-best results within each fine-tuning setting, respectively.

Task-specific versus joint multimodal finetuning. Joint multimodal finetuning improved over task-specific finetuning across nearly all metrics, with the largest gains observed for the kidney (Table 2). Among the joint models, pretrained initialization generally gave better Dice, Chamfer, and HD95 than random initialization for both organs, supporting the use of pretrained multiGradICON for multimodal adaptation. TRE showed a different trend. For the kidney, training from scratch with MIND-SSC achieved the lowest TRE among all configurations (1.70 mm). For the prostate, however, the rigid PCA–ICP initialization itself achieved the lowest TRE (4.33 mm), slightly better than the best deformable configuration (4.40 mm). Importantly, the deformable models still substantially improved Chamfer distance, Dice, and HD95 over rigid alignment, indicating better overall anatomical and surface correspondence. The prostate TRE result therefore suggests that the additional deformation does not necessarily improve sparse landmark alignment, even when it improves the organ-level registration. Together, these results motivate further investigation of how landmark accuracy can be better preserved or optimized during multimodal deformable registration.

Figure 2 shows representative prostate MRI–TRUS and kidney CT–US results, illustrating improved anatomical alignment together with smooth deformation fields. Beyond registration accuracy, US-multiGradICON may also help reduce the annotation burden in future ultrasound studies. Although Dice supervision requires ultrasound segmentations during fine-tuning, a model adapted using a limited annotated dataset could subsequently be used to propagate CT or MRI labels to ultrasound images, providing pseudo-labels for weak supervision or assisting quality control. The reliability of such label propagation, however, will depend on the quality of both the source segmentations and the estimated registration, and may be affected by variations in scanner, operator, probe position, field of view, and image artifacts. Future studies should therefore investigate how registration and label propagation performance change as the amount of

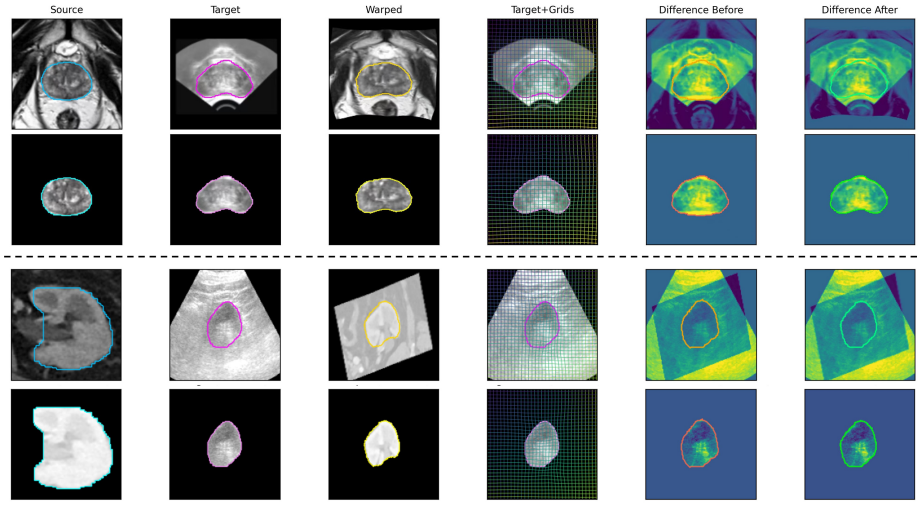

Fig. 2. Qualitative registration results for a prostate case (top) and a kidney case (bottom). For each case, the bottom row shows a cropped, organ-masked view of the row above. Masking is applied for visualization only; the network receives full, unmasked images as input.

annotated ultrasound data is reduced, and assess robustness across a broader range of acquisition conditions.

Several limitations should also be considered. First, the proposed approach relies on organ segmentations during fine-tuning, which may be difficult or costly to obtain for many clinical applications. Second, the experiments were conducted on relatively small datasets, reflecting the practical difficulty of acquiring paired multimodal images with reliable spatial correspondence. Finally, we did not re-evaluate the fine-tuned model on the original multiGradICON benchmark tasks. Consequently, whether adaptation to ultrasound registration leads to performance degradation on previously learned registration tasks, including potential catastrophic forgetting, remains to be investigated.

5 Conclusion

We presented a global-to-local framework for CT–US and MR–US registration that combined PCA–ICP initialization with multiGradICON-based deformable refinement. The results showed that finetuning and segmentation-guided supervision improved anatomical alignment without per-instance optimization. We also established a MONAI-based benchmark to support reproducible evaluation of pretrained and future foundation registration models for ultrasound-guided applications. Future work will focus on reducing annotation dependence and validating the method on larger, more diverse clinical datasets.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grants U25A20145, 22IAA01849, and 62303275, and in part by the Jinan Science and Technology Bureau under Grant 202333011. Gabor Fichtinger is a Canada Research Chair in Computer Integrated Surgery, Tier 1.

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

References

1. Ndzimbong, W., Thome, N., Fourniol, C., Keeza, Y., Sauer, B., Marescaux, J., George, D., Hostettler, A., Collins, T.: Global registration of kidneys in 3d ultrasound and ct images. International Journal of Computer Assisted Radiology and Surgery **20**(1), 65–75 (2025)
2. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE transactions on Medical Imaging **16**(2), 187–198 (1997)
3. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: application to breast mr images. IEEE transactions on medical imaging **18**(8), 712–721 (1999)
4. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis **12**(1), 26–41 (2008)
5. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: Fast predictive image registration–a deep learning approach. NeuroImage **158**, 378–396 (2017)
6. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging **38**(8), 1788–1800 (2019)
7. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. Medical Image Analysis **82**, 102615 (2022)
8. Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V.: Synthmorph: learning contrast-invariant registration without acquired images. IEEE transactions on medical imaging **41**(3), 543–558 (2021)
9. Demir, B., Tian, L., Greer, H., Kwitt, R., Vialard, F.X., Estépar, R.S.J., Bouix, S., Rushmore, R., Ebrahim, E., Niethammer, M.: Multigradicon: A foundation model for multimodal medical image registration. In: Biomedical Image Registration. pp. 3–18. Springer Nature Switzerland, Cham (2024)
10. Tian, L., Greer, H., Vialard, F.X., Kwitt, R., Estépar, R.S.J., Rushmore, R.J., Makris, N., Bouix, S., Niethammer, M.: Gradicon: Approximate diffeomorphisms via gradient inverse consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18084–18094 (2023)
11. Miura, S., Nakayama, M., Xu, K., Bo, Z., Kuromatsu, R., Nakano, M., Noda, Y., Kawaguchi, T.: Point cloud registration algorithm using liver vascular skeleton feature with computed tomography and ultrasonography image fusion. International Journal of Computer Assisted Radiology and Surgery **20**(12), 2469–2478 (2025)

12. Greenacre, M., Groenen, P.J., Hastie, T., d'Enza, A.I., Markos, A., Tuzhilina, E.: Principal component analysis. Nature Reviews Methods Primers **2**(1), 100 (2022)
13. Besl, P.J., McKay, N.D.: Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures. vol. 1611, pp. 586–606. Spie (1992)
14. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, M., Schnabel, J.A.: Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. Medical image analysis **16**(7), 1423–1435 (2012)
15. Siebert, H., Großbröhmer, C., Hansen, L., Heinrich, M.P.: Convexadam: Self-configuring dual-optimization-based 3d multitask medical image registration. IEEE Transactions on Medical Imaging **44**(2), 738–748 (2024)
16. Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C.M., Emberton, M., Noble, J.A., Barratt, D.C., Vercauteren, T.: Label-driven weakly-supervised learning for multimodal deformable image registration. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 1070–1074 (2018)
17. Qiu, H., Qin, C., Schuh, A., Hammernik, K., Rueckert, D.: Learning diffeomorphic and modality-invariant registration using b-splines. In: Medical Imaging with Deep Learning. pp. 645–664. PMLR (2021)
18. Ndzimbong, W., Fourniol, C., Themyr, L., Thome, N., Keeza, Y., Sauer, B., Piéchaud, P.T., Méjean, A., Marescaux, J., George, D., et al.: Trusted: The paired 3d transabdominal ultrasound and ct human data for kidney segmentation and registration research. Scientific Data **12**(1), 615 (2025)
19. Baum, Z., Saeed, S., Min, Z., Hu, Y., Barratt, D.: Mr to ultrasound registration for prostate challenge-dataset. In: Medical Image Computing and Computer Assisted Intervention–MICCAI. vol. 2023 (2023)