Implicit Deformable Medical Image Registration with Learnable Kernels

Stefano Fogarollo[1], Gregor Laimer[2], Reto Bale[2], Matthias Harders[1]

[1] Interactive Graphics and Simulation Group (IGS), Department of Computer Science, University of Innsbruck, Technikerstraße 21 Innsbruck, Austria
[2] Interventional Oncology-Microinvasive Therapy (SIP), Department of Radiology, Medical University Innsbruck, Innsbruck, Austria

Corresponding authors: Reto.Bale@i-med.ac.at, Matthias.Harders@uibk.ac.at

Abstract. Deformable medical image registration is an essential task in computer-assisted interventions. This problem is particularly relevant to oncological treatments, where precise image alignment is necessary for tracking tumor growth, assessing treatment response, and ensuring accurate delivery of therapies. Recent AI methods can outperform traditional techniques in accuracy and speed, yet they often produce unreliable deformations that limit their clinical adoption. In this work, we address this challenge and introduce a novel implicit registration framework that can predict accurate and reliable deformations. Our insight is to reformulate image registration as a signal reconstruction problem: we learn a kernel function that can recover the dense displacement field from sparse keypoint correspondences. We integrate our method in a novel hierarchical architecture, and estimate the displacement field in a coarse-to-fine manner. Our formulation also allows for efficient refinement at test time, permitting clinicians to easily adjust registrations when needed. We validate our method on challenging intra-patient thoracic and abdominal zero-shot registration tasks, using public as well as internal datasets from the Innsbruck University Hospital. Our method not only shows competitive accuracy to state-of-the-art approaches, but also bridges the generalization gap between implicit and explicit registration techniques. In particular, our method generates deformations that better preserve anatomical relationships and matches the performance of specialized commercial systems, underscoring its potential for clinical adoption.

Keywords: Registration · Implicit · Liver · Interactive

1 Introduction

Accurate and reliable image registration is an essential step in computer-assisted interventions, with direct applications on intra-procedural navigation, treatment monitoring and evaluation. The task consists of finding the optimal transformation that aligns the two input images. In particular in the abdominal and thoracic

regions, deformable registration is needed to correctly model large nonlinear deformations resulting from the complex behavior and interaction of soft tissues [45]. Traditional image registration methods [2,23,32] can estimate accurate deformations for each image pair, but often struggle to balance accuracy, computational efficiency, and anatomical plausibility; modern techniques can leverage hardware acceleration (GPU) for faster inference [18, 36]. Recent advances in artificial intelligence (AI) have shown promising results in addressing these challenges, with learning-based methods achieving state-of-the-art performance in terms of both speed and accuracy. Building on the frameworks described in VoxelMorph [3] and SynthMorph [21], AI methods are trained to predict a tensor $\phi \in \mathbb{R}^{D\times H\times W\times 3}$ representing the displacement field that spatially aligns the two 3D inputs, namely the moving and fixed images, with D, H, W denoting the spatial dimensions. Backpropagation of the gradients is achieved by warping the image with ϕ using a Spatial Transformer Network layer [22]. Multi-stage incremental prediction has been shown to increase the registration accuracy with minimal computational overhead [8, 30, 42, 43], achieving comparable performance to pyramidal [33] and cascaded [47] architectures. Despite these improvements, current AI registration methods struggle with unseen anatomical variations or clinical scenarios not encountered during training [20]. Recent foundation models try to bridge this gap by pre-training on large-scale diverse datasets, but they still need test-time refinement in difficult zero-shot cases ("Type 2" out-of-distribution) [11, 39]. Implicit Neural Representations (INRs) have recently emerged as a powerful paradigm, offering high-resolution, and memory-efficient modeling of continuous spatial signals [28, 31, 37]. However, conventional INRs are inherently designed to overfit individual input signals and lack the capacity to generalize across diverse samples. For instance, in the context of medical imaging, the method introduced in [46] can model the spatial transformation between a pair of medical images, but requires retraining for each new pair, limiting clinical deployment. Several strategies have been explored to address this limitation, including hyper-networks [7, 14, 37], modulation of periodic activations [28], and conditioned multilayer perceptrons (MLP) [1]. Recently, the work of [48] proposed to condition learned image features for improved generalization. To our knowledge, this remains the only prior work addressing generalized INRs for medical image registration, highlighting a critical research gap. Therefore, a key question remains [6]: **how can we mitigate AI generalization issues to achieve reliable zero-shot registration?** This is the focus of our work, and to answer the question, we propose a fundamentally different approach that casts deformable registration as a continuous signal reconstruction task. Rather than fitting the dense displacement field directly from image intensities, we introduce a novel implicit model conditioned on sparse, automatically extracted keypoint correspondences. Our method leverages an implicit dual-stream attention mechanism to model the spatial and semantic dependencies of the neighboring displacements, allowing to learn complex deformations in a data-driven manner. Critically different from *explicit* sparse correspondence extrapolation methods such as [17], our framework learns a continuous *implicit* representa-

tion of the dense displacement field. Distinct from related implicit methods, we condition the reconstruction of the displacement field on sparse cost-volume optimal displacements, not just on image features. Furthermore, our approach naturally supports interactive test-time refinement, enhancing its practicality for real-world clinical applications. We visualize the proposed method in Figure 1. In summary, the main contributions of this work are threefold: 1) We introduce a novel implicit framework for medical image registration leveraging learnable kernels. 2) We condition the representation on local keypoint correspondences. 3) We evaluate our method and compare it with several state-of-the-art approaches on zero-shot intra-patient registration tasks. In the following sections, we introduce our method, followed by empirical results and ablation studies supporting the design choices.

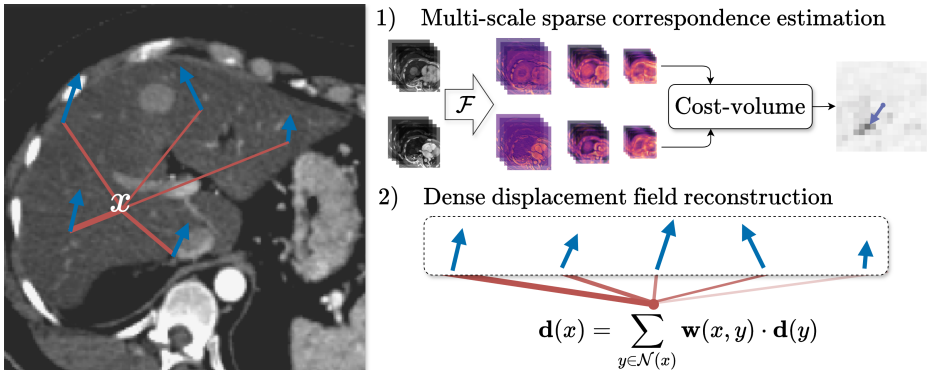


Fig. 1: In this work, we rephrase deformable image registration as signal reconstruction. First, we obtain sparse correspondences between the input images using cost-volume optimization on multi-scale features learned with $\mathcal{F}$. Then, we reconstruct the displacement field with a learnable kernel function $\mathbf{w}$, conditioned on the local displacements $\mathbf{d}(y)$ at locations in the neighborhood $y \in \mathcal{N}(x)$.

2 Method

In this work, we cast deformable medical image registration as a signal reconstruction problem. Our goal is to reconstruct a high-dimensional signal $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ from a sparse set of observations $O = \{\mathbf{p_i}, \mathbf{d_i}\}_{i \in \mathbb{N}: 1 \leq i \leq m}$, where $\mathbf{p_i}$ are spatial coordinates $\in \mathbb{R}^3$, and $\mathbf{d_i} \in \mathbb{R}^3$ denote the observed signal values. Under this formulation, we evaluate ρ on the regular Cartesian grid to represent the displacement field ϕ on the voxel domain.

2.1 Learnable kernels for image registration

Our key insight is that natural signals, such as displacement fields, exhibit strong local and global structure that can be encoded in a learnable basis representation. We adopt a kernelised formulation and compute $\mathbf{d}(x)$ as

$$\rho(x) = \sum_{y \in \mathcal{N}(x)} \mathbf{w}(x, y) \cdot \mathbf{d}(y), \tag{1}$$

where $\mathcal{N}(x)$ represents the neighborhood of x, and the kernel $\mathbf{w}$ determines the contribution of each neighbor to the final displacement at x. To construct $\mathcal{N}(x)$, we include the nearest $K = 30$ keypoints from x and associated correspondences. The signal observations can be obtained via point-to-point correspondences between the moving and fixed images, as described later on. While $\mathbf{w}$ is defined a priori in traditional techniques, in our method it is parameterized with a neural network. We design a dual-stream attention mechanism [41] to disentangle geometric and semantic information with two dedicated attention heads $\mathcal{H}_s, \mathcal{H}_f$, operating on the latent coordinate and feature space, respectively:

$$a(x, y) = \mathcal{H}_s(\mathcal{E}_s(x), \mathcal{E}_s(y))) + \mathcal{H}_f(\mathcal{E}_f(x_f) + \mathcal{E}_f(y_f))) + b(x, y), \quad y \in \mathcal{N}(x). \tag{2}$$

$a(x, y)$ encodes the attention score between two points in the image domain x, y, and $b(x, y) = \frac{1}{1+\|x-y\|^2}$ is the spatial bias component. $\mathcal{E}_s$ and $\mathcal{E}_f$ are the geometric and semantic feature encoders respectively, parameterized with a three-layer MLP with 128 hidden units and ReLU activations; overall our learnable kernel module consists of $\approx$ 150k trainable parameters. We extract dense semantic features from the inputs with a learnable encoder $\mathcal{F}$, which are sampled at x, y to obtain the feature vectors x_f, y_f respectively; $\mathcal{F}$ is parameterized with a multi-scale UNet encoder [35]. To calculate $\mathbf{w}$ we apply the softmax operation on the attention scores in the neighborhood, $\mathbf{w}(x, y) = \frac{e^{a(x,y)}}{\sum_{y \in \mathcal{N}(x)} e^{a(x,y)}}$.

2.2 Conditioning

To improve consistency and generalization, we propose to condition the reconstruction of the displacement field on the local neighborhood of displacements. To do so, we first detect a set of salient points on the image, and then estimate the optimal correspondences in the other image using a differentiable cost-volume layer on the learned multi-scale features [13,19]; during training, the optimal correspondences are identified within a $7 \times 7 \times 7$ voxel patch around each keypoint, whereas at inference this search window is expanded to $9 \times 9 \times 9$ voxels. These correspondences are the signal observations $\mathbf{d}(y), y \in \mathcal{N}(x)$, used in the reconstruction of the displacement field. In our experiments, we find that traditional detectors such as Förstner [18] and SIFT [27], as well as deep-learning ones such as DISK [40] and SuperPoint [12], provide a sufficient number of keypoints to achieve competitive performance, with minimal computational overhead. When using SIFT, DISK, and SuperPoint detectors we process each 2D slice in the input volume separately and retain only the points with confidence greater than a

pre-defined tolerance. During training, we limit the number of keypoints to 1024 by farthest point sampling, while at inference we allow maximum 3000 points.

Test-time Interactive Refinement. A unique feature of our method is the support for efficient test-time interactive refinement. We leverage the implicit kernelised representation and compute localized updates to $\mathbf{w}(x, u)$, where $u \in \mathcal{N}(x)$ and is a new point-to-point correspondence between images, without requiring an additional full registration.

2.3 Training

Following related work, we predict the final displacement field incrementally with a multi-scale pipeline [8, 30, 42]. We reconstruct the optimal dense deformation field at each scale, and use this estimate to warp the features at the next scale, iterating until full-resolution. In our experiments, we use five scales. We train our method with usual registration loss functions, imposing the normalized cross-correlation (NCC) loss to promote image similarity and the isotropic diffusion as regularizer [3]. If available, we also warp the segmentation masks and impose the Dice loss [5] on the warped and fixed segmentation masks. Similarly, if landmarks are available, we include the Euclidean distance between the fixed and warped landmarks in the loss function. We use the same weight for each loss function.

3 Experiments and Results

We compare our method with state-of-the-art registration approaches, including six incremental multi-scale registration techniques (corrMLP [29], H-ViT [15], ModeT [42], NICE-Trans [30], RDP [43], WiNet [8]), the implicit method AM SIREN [48], and the recent foundation model uniGradICON [39]. To ensure a fair comparison, we fine-tuned uniGradICON with 50 test-time iterations ($\approx$ 1 minute). We implemented a multi-scale version of the "3l-512" AM SIREN architecture, totaling 7.8m of trainable parameters. The remaining architectures were configured to have 3.5 ± 0.3m of trainable parameters by adjusting the number of feature channels. We train all the learning methods until convergence on the validation set (for a maximum of 100 epochs), and test them using the checkpoint with the best validation metrics; each training run never exceeds 7 hours. During training we adopt common data augmentation techniques such as Gaussian noise and blurring. We set the same randomization seed in each training run so that each model is trained on exactly the same data.

We use two challenging intra-patient registration datasets, namely the public NLST data[3] [10], and a large-scale dataset from the Innsbruck University Hospital containing 96 colorectal cancer hepatic interventions. In the following tables, we report the mean and standard deviation across measurements, as well as the 5% and 95% confidence intervals in squared brackets [9]. Arrows indicate

[3] `https://www.cancerimagingarchive.net/collection/nlst`

whether higher (↑) or lower values (↓) are better. * represents statistically significant differences with respect to our results, after Bonferroni correction.

NLST data. This dataset, featured in the MICCAI Learn2Reg 2023 challenge [20], is extensively used to benchmark intra-patient registration methods. The registration task involves estimating lung deformations between inhale and exhale CT scans, a challenging problem due to the presence of large non-linear displacements. In our experiments, we observed that using a mean-squared error (MSE) loss yielded superior results compared to using the NCC loss. To ensure a fair comparison with memory-intensive methods, we resized the input volumes to half of their original resolution ($112 \times 96 \times 112$), and trained all models at this resolution; for the additional pre-processing steps we follow the challenge instructions [20]. We compare the methods on the salient target registration error (TRE, measured in mm) using the provided ground-truth landmarks and on the regularity of the deformations, in terms of the standard deviation of the logarithm of the Jacobian determinant (SDlogJ). In order to evaluate robustness, we also determine the 30th percentile of largest landmark distances (TRE30) [20]. The results are compiled in Table 1.

Table 1: Quantitative results on the Learn2Reg NLST dataset.

Method	TRE (mm) ↓	TRE30 (mm) ↓	SDlogJ ↓
ours	$1.72_{\pm 0.43}$ $[1.40, 2.04]$	$1.89_{\pm 0.38}$ $[1.60, 2.17]$	0.02
AM SIREN [48]	$3.51_{\pm 1.06}$ $[2.71, 4.31]^*$	$4.11_{\pm 1.23}$ $[3.19, 5.04]^*$	0.08^*
corrMLP [29]	$3.30_{\pm 1.33}$ $[2.29, 4.30]^*$	$3.80_{\pm 1.42}$ $[2.73, 4.87]^*$	0.05^*
H-ViT [15]	$3.77_{\pm 1.39}$ $[2.72, 4.82]^*$	$4.45_{\pm 1.57}$ $[3.26, 5.63]^*$	0.05^*
ModeT [42]	$2.33_{\pm 0.76}$ $[1.75, 2.90]$	$2.51_{\pm 0.72}$ $[1.97, 3.05]$	0.06^*
NICE-Trans [30]	$3.27_{\pm 1.32}$ $[2.28, 4.27]^*$	$3.80_{\pm 1.67}$ $[2.54, 5.06]^*$	0.07^*
RDP [43]	$2.42_{\pm 1.04}$ $[1.63, 3.20]$	$2.56_{\pm 1.06}$ $[1.77, 3.36]$	0.05^*
uniGradICON [39]	$1.77_{\pm 0.29}$ $[1.55, 1.98]$	$1.87_{\pm 0.31}$ $[1.63, 2.10]$	0.04^*
WiNet [8]	$3.60_{\pm 1.31}$ $[2.61, 4.59]^*$	$4.17_{\pm 1.33}$ $[3.16, 5.17]^*$	0.03^*

Colorectal liver cancer data. This dataset includes 96 CT scans from different patients before and immediately after radio-frequency ablation [4], acquired in arterial and venous phase respectively. This minimally invasive intervention induces highly non-linear deformations in the liver, primarily due to respiratory motion and tissue shrinkage [26, 45], as well as significant intensity changes near the tumor region. These challenges make the dataset particularly valuable for evaluating the robustness and accuracy of registration methods under complex, real-world conditions. Ground-truth segmentation masks for liver, tumor, and treatment area have been semi-automatically obtained, checked, and corrected by two clinicians at the Innsbruck University Hospital. Pre-processing steps involve resampling the volumes to the same voxel spacing ($3.0 \times 1.4 \times 1.4$ mm^3), cropping a region of $[80 \times 192 \times 192]$ voxels around the liver mask obtained with TotalSegmentator [44], masking the image intensities using the 5th and 95th

percentiles, and normalizing them to $[0, 1]$. We used a three cross-fold training scheme, with 48 cases for training, 16 for validation, and 32 for testing. We compare with related work on the following two tasks: liver registration accuracy and safety margin assessment (SMA). For the liver registration, we compute the average symmetric surface distance (ASSD), and the 95th percentile of the Hausdorff distance (HD95) on the liver mask, both measured in mm. To calculate SMA, we measure the distance between the warped treatment area and the pre-operative tumor, using a 5 mm margin as the threshold to determine treatment success [24]: this is widely recognized as an independent predictor of local disease progression [25]. We calculate the receiver operating characteristic (ROC) curve for this classification task [34], and report the results in Table 2. In the last row of the table, we show the results from a state-of-the-art commercial software, specifically designed for this task [38]. For detailed visual results, we refer to the supplementary material.

Table 2: Quantitative results on the dataset from the Innsbruck University Hospital.

Method	Liver ASSD (mm) ↓	Liver HD95 (mm) ↓	SMA (%) ↑
ours	$1.10_{\pm 0.96}$ [0.90, 1.29]	$4.71_{\pm 3.78}$ [3.94, 5.48]	70.59
AM SIREN [48]	$0.99_{\pm 0.90}$ [0.81, 1.17]	$4.86_{\pm 3.76}$ [4.10, 5.63]	52.94
corrMLP [29]	$2.53_{\pm 2.46}$ [2.03, 3.04]*	$8.51_{\pm 6.98}$ [7.09, 9.93]*	53.27
H-ViT [15]	$1.63_{\pm 1.24}$ [1.37, 1.88]*	$6.20_{\pm 4.48}$ [5.29, 7.11]	50.33
ModeT [42]	$0.98_{\pm 0.87}$ [0.80, 1.15]	$4.77_{\pm 3.72}$ [4.02, 5.53]	56.21
NICE-Trans [30]	$1.26_{\pm 1.13}$ [1.03, 1.49]	$5.58_{\pm 4.18}$ [4.73, 6.43]	59.15
RDP [43]	$0.82_{\pm 0.88}$ [0.64, 1.00]	$4.13_{\pm 3.45}$ [3.42, 4.83]	58.82
uniGradICON [39]	$1.01_{\pm 0.88}$ [0.84, 1.19]	$4.29_{\pm 3.15}$ [3.65, 4.93]	64.71
WiNet [8]	$1.68_{\pm 1.27}$ [1.42, 1.94]*	$6.33_{\pm 4.61}$ [5.39, 7.26]	50.65
Ablation-fit [38]	-	-	71.24

4 Discussion, Limitations, and Conclusion

The empirical results on the public NLST dataset demonstrate that our method either outperforms or pars state-of-the-art registration accuracy on large deformation modeling, while generating smoother deformations. In the case of the internal colorectal cancer dataset, while the liver surface is accurately aligned by most approaches, safety margin assessment remains a significant challenge. Notably, ours is the only AI method matching the performance of specialized commercial systems for this task, with dramatically reduced inference time ($1.6_{\pm 0.3}$ seconds including keypoint extraction, compared to over two minutes), underscoring its potential for clinical translation.

Ablation studies. To analyze the factors influencing the performance of our method, we conduct ablation studies on the core architectural elements using the

public NLST dataset for reproducibility, and summarize the results in Table 3. Although not statistically significant, there is a clear performance drop when removing each component. In particular, the proposed learnable interpolation offers a key advantage in terms of deformation accuracy. Further, we observe that keypoint detections from deep-learning methods (DISK [40], SuperPoint [12]) remain effective despite significant domain shift, and keypoint location outside of relevant anatomical areas.

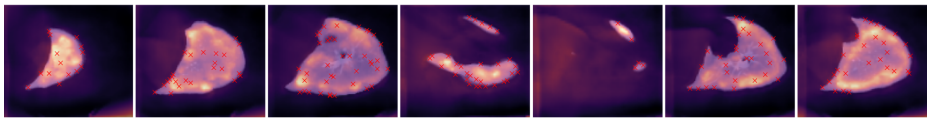
Limitations. We note that our current implementation allocates $17.8_{\pm 1.4}$ GB of GPU memory during training due to the dense voxel-wise sampling and cost-volume computations. We are actively developing more efficient implementations to enhance scalability, especially for resource-constrained clinical environments. In particular, the method could be optimized further to have sub-second response times, and register non-contrast CT scans.

Table 3: Ablation studies on the NLST data. * represents statistically significant differences with respect to our method with Förstner keypoints. We use the following abbreviations: *kpts.* for keypoints, and *extrap.* for extrapolation.

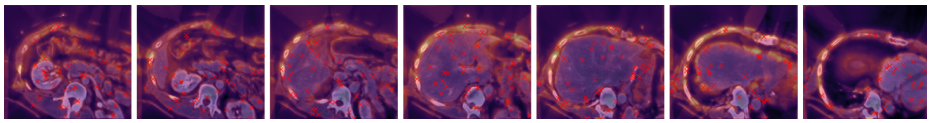
Method	TRE (mm) ↓	TRE30 (mm) ↓	SDlogJ ↓
ours	$1.72_{\pm 0.43}$ [1.40, 2.04]	$1.89_{\pm 0.38}$ [1.60, 2.17]	0.02
only $\mathcal{H}_s$	$2.00_{\pm 0.54}$ [1.59, 2.40]	$2.24_{\pm 0.60}$ [1.79, 2.69]	0.01
only $\mathcal{H}_f$	$1.79_{\pm 0.45}$ [1.45, 2.14]	$1.98_{\pm 0.41}$ [1.67, 2.29]	0.01
TPS extrap.	$2.23_{\pm 0.62}$ [1.76, 2.69]	$2.53_{\pm 0.67}$ [2.02, 3.04]*	0.02
DISK kpts.	$1.75_{\pm 0.35}$ [1.48, 2.01]	$1.89_{\pm 0.36}$ [1.62, 2.17]	0.02
SIFT kpts.	$1.78_{\pm 0.45}$ [1.43, 2.12]	$1.95_{\pm 0.52}$ [1.55, 2.34]	0.02*
SuperPoint kpts.	$2.15_{\pm 0.64}$ [1.66, 2.63]	$2.42_{\pm 0.66}$ [1.92, 2.91]	0.02*

Test-time behavior. We highlight the practicality of our approach by computing the standard deviation of the attention scores as a proxy for the "confidence" of the displacements [16], and visualize it in Figure 2. Finally, we also examine the value of interactive refinement by re-computing the displacements based on the provided ground-truth landmarks: we achieve a TRE improvement of $5, 9, 13\%$ adding $10, 20, 30$ uniformly randomly sampled landmarks, respectively.

Conclusion and Future Work. In summary, this work introduces a novel implicit framework that achieves a unique balance of accuracy, reliability, and clinical usability by conditioning the signal reconstruction on sparse keypoint correspondences. Our approach not only mitigates generalization issues in existing AI-based registration methods but it also provides a robust and practical solution for real-world clinical applications, achieving performance comparable to specialized commercial systems. The code is available at `https://git.uibk.ac.at/informatik/igs/open/msm`. In the future, we will conduct further evaluation on different anatomical areas and modalities.



(a) Case 101 from the NLST data, landmarks TRE = 1.10 mm.



(b) Case 130 from the Innsbruck University Hospital, liver ASSD = 0.87 mm.

Fig. 2: Overlay of the prediction "confidence" on the warped volume, with detected keypoints marked in red.

Acknowledgements. This project has been supported by the Austrian Science Foundation (10.55776/DOC110). The computational results presented have been achieved (in part) using the Vienna Scientific Cluster (VSC).
Disclosure of Interests. The authors declare no conflict of interest.

References

1. Amiranashvili, T., et al.: Learning shape reconstruction from sparse measurements with neural implicit functions. In: MIDL. pp. 22–34 (2022)
2. Avants, B., et al.: Symmetric diffeomorphic image registration with cross-correlation. Medical Image Analysis **12**, 26–41 (2008)
3. Balakrishnan, G., et al.: Voxelmorph: A learning framework for deformable medical image registration. IEEE TMI **38**, 1788–1800 (2019)
4. Bale, R., Widmann, G., Haidu, M.: Stereotactic radiofrequency ablation. Cardio-Vascular and Interventional Radiology **34**, 852–856 (2011)
5. Chen, J., et al.: Transmorph: Transformer for unsupervised medical image registration. Medical Image Analysis **82**, 102615 (2022)
6. Chen, J., et al.: A survey on deep learning in medical image registration. Medical Image Analysis **100**, 103385 (2025)
7. Chen, Y., et al.: Transformers as meta-learners for implicit neural representations. In: ECCV. pp. 170–187 (2022)
8. Cheng, X., et al.: Winet: Wavelet-based incremental learning for efficient medical image registration. In: MICCAI. pp. 761–771 (2024)
9. Christodoulou, E., et al.: Confidence intervals uncovered: Are we ready for real-world medical imaging AI? In: MICCAI. pp. 124 – 132 (2024)
10. Clark, K., et al.: The cancer imaging archive (TCIA). Journal of Digital Imaging **26**(6), 1045–1057 (2013)
11. Demir, B., et al.: Multigradicon: A foundation model for multimodal medical image registration. In: Biomedical Image Registration. pp. 3–18 (2024)
12. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: CVPRW. pp. 337–33712 (2018)
13. Dosovitskiy, A., et al.: Flownet: Learning optical flow with convolutional networks. In: ICCV (2015)

14. Fehrentz, M., et al.: Intraoperative Registration by Cross-Modal Inverse Neural Rendering. In: MICCAI. pp. 317 – 327 (2024)
15. Ghahremani, M., et al.: H-vit: A hierarchical vision transformer for deformable image registration. In: CVPR. pp. 11513–11523 (2024)
16. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: ICML. p. 1321–1330 (2017)
17. Hansen, L., Heinrich, M.P.: Graphregnet: Deep graph regularisation networks on sparse keypoints for dense registration of 3D lung CTs. IEEE TMI **40** (2021)
18. Hansen, L., Heinrich, M.P.: Revisiting iterative highly efficient optimisation schemes in medical image registration. In: MICCAI. pp. 203–212 (2021)
19. Heinrich, M.P.: Closing the gap between deep and conventional image registration using probabilistic dense displacement networks. In: MICCAI. pp. 50–58 (2019)
20. Hering, A., et al.: Learn2reg: Comprehensive Multi-Task Medical Image Registration Challenge, Dataset and Evaluation in the Era of Deep Learning. IEEE TMI **42**, 697–712 (2023)
21. Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V.: Synthmorph: Learning contrast-invariant registration without acquired images. IEEE TMI **41**(3), 543–558 (2022)
22. Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K.: Spatial transformer networks. In: NIPS. p. 2017–2025 (2015)
23. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: elastix: A toolbox for intensity-based medical image registration. IEEE TMI **29**, 196–205 (2010)
24. Laimer, G., et al.: Minimal ablative margin (MAM) assessment with image fusion. European Radiology **30**(5), 2463–2472 (2020)
25. Laimer, G., et al.: Multicenter and inter-software evaluation of ablative margins after thermal ablation of colorectal liver metastases. European Radiology **35**(2), 1046–1056 (2025)
26. Liu, D., et al.: Local contractive registration with biomechanical model. IEEE Journal of Biomedical and Health Informatics **28**(1), 415–426 (2024)
27. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**, 91–110 (2004)
28. Mehta, I., et al.: Modulated periodic activations for generalizable local functional representations. In: ICCV. pp. 14194–14203 (2021)
29. Meng, M., Feng, D., Bi, L., Kim, J.: Correlation-aware coarse-to-fine mlps for deformable medical image registration. In: CVPR. pp. 9645–9654 (2024)
30. Meng, M., et al.: Non-iterative coarse-to-fine transformer networks for joint affine and deformable image registration. In: MICCAI. pp. 750–760 (2023)
31. Mildenhall, B., et al.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: ECCV (2020)
32. Modat, M., et al.: Fast free-form deformation using graphics processing units. Computer Methods and Programs in Biomedicine **98**(3), 278–284 (2010)
33. Mok, T.C.W., Chung, A.C.S.: Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks. In: MICCAI. pp. 211–221 (2020)
34. Paolucci, I., et al.: Identification of a0 minimum ablative margins for colorectal liver metastases. British Journal of Surgery **111**(9) (2024)
35. Ronneberger, O., et al.: U-net. In: MICCAI. pp. 234–241 (2015)
36. Siebert, H., et al.: Convexadam. IEEE TMI **44**(2), 738–748 (2025)
37. Sitzmann, V., et al.: Implicit neural representations with periodic activation functions. In: NIPS. pp. 7462–7473 (2020)
38. Solbiati, M., et al.: A novel software platform for volumetric assessment of ablation completeness. International Journal of Hyperthermia **36**, 336–342 (2019)

39. Tian, L., et al.: uniGradICON: A foundation model for medical image registration. In: MICCAI. pp. 749–760 (2024)
40. Tyszkiewicz, M., Fua, P., Trulls, E.: Disk: Learning local features with policy gradient. In: NIPS. vol. 33, pp. 14254–14265 (2020)
41. Vaswani, A., et al.: Attention is all you need. In: NIPS. vol. 30 (2017)
42. Wang, H., Ni, D., Wang, Y.: Modet: Learning deformable image registration via motion decomposition transformer. In: MICCAI. pp. 740–749 (2023)
43. Wang, H., et al.: Recursive deformable pyramid network for unsupervised medical image registration. IEEE TMI **43**(6), 2229–2240 (2024)
44. Wasserthal, J., et al.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)
45. Wise, P.A., et al.: Intraoperative liver deformation and organ motion caused by ventilation, laparotomy, and pneumoperitoneum in a porcine model for image-guided liver surgery. Surgical Endoscopy **38**(3), 1379–1389 (2024)
46. Wolterink, J.M., et al.: Implicit neural representations for deformable image registration. In: PMLR. pp. 1349–1359 (2022)
47. Zhao, S., et al.: Recursive cascaded networks for unsupervised medical image registration. In: ICCV. pp. 10599–10609 (2019)
48. Zimmer, V.A., et al.: Towards generalised neural implicit representations for image registration. In: Deep Generative Models. pp. 45–55 (2024)