MemWarp: Discontinuity-Preserving Cardiac Registration with Memorized Anatomical Filters

Hang Zhang[1] ✉, Xiang Chen[2], Renjiu Hu[1], Dongdong Liu[3], Gaolei Li[4], and Rongguang Wang[5]

[1] Cornell University
[2] Hunan University
[3] New York University
[4] Shanghai Jiao Tong University
[5] University of Pennsylvania
hz459@cornell.edu

Abstract. Many existing learning-based deformable image registration methods impose constraints on deformation fields to ensure they are globally smooth and continuous. However, this assumption does not hold in cardiac image registration, where different anatomical regions exhibit asymmetric motions during respiration and movements due to sliding organs within the chest. Consequently, such global constraints fail to accommodate local discontinuities across organ boundaries, potentially resulting in erroneous and unrealistic displacement fields. In this paper, we address this issue with *MemWarp*, a learning framework that leverages a memory network to store prototypical information tailored to different anatomical regions. *MemWarp* is different from earlier approaches in two main aspects: firstly, by decoupling feature extraction from similarity matching in moving and fixed images, it facilitates more effective utilization of feature maps; secondly, despite its capability to preserve discontinuities, it eliminates the need for segmentation masks during model inference. In experiments on a publicly available cardiac dataset, our method achieves considerable improvements in registration accuracy and producing realistic deformations, outperforming state-of-the-art methods with a remarkable 7.1% Dice score improvement over the runner-up semi-supervised method. Source code will be available at https://github.com/tinymilky/Mem-Warp.

Keywords: Deformable image registration · Memory network · Discontinuity preserving

1 Introduction

Cardiovascular disease, a major cause of death worldwide [22], depends on medical imaging, especially cine-MRI [13], for diagnosis and treatment. Deformable image registration [3], a crucial step for cardiac analysis, has seen improvements through learning-based neural networks. These models vary from unsupervised to semi- and weakly-supervised frameworks. Unsupervised methods are favored

for their simplicity, requiring only raw images for training and inference. In contrast, semi-supervised methods need segmentation masks during training, while weakly-supervised models require them in both training and inference phases.

Learning-based registration models [3, 7, 16] outperform traditional iterative optimization approaches [1, 2, 4, 15, 23] in efficiency and precision. Yet, they often presuppose globally smooth deformations, a premise that doesn't align with the dynamic nature of cardiac motions influenced by heartbeat and respiratory-induced organ sliding. Additionally, volume shifts between end-diastole (ED) and end-systole (ES) phases, such as expansion of the left ventricular myocardium (LVM) and reductions in the left ventricular blood pool (LVBP) and right ventricle (RV), underscore the need for models that can handle local discontinuities to accurately depict cardiac motions.

Despite the clear need for discontinuity-preserving methods to capture the cardiac cycle's complexity, the field remains underexplored. Ng et al. [18] pioneered this area by integrating a discontinuous regularizer for local discontinuity without segmentation masks in an unsupervised manner, though accurately defining organ boundaries remains challenging. DDIR [9] and textCSF [8] address this by using segmentation masks to refine boundaries in a weakly-supervised manner, yet they require segmentation masks during both training and inference, making registration accuracy highly dependent on the quality of segmentation.

To tackle these challenges, we introduce MemWarp, a semi-supervised framework that balances local smoothness with the preservation of local discontinuities. MemWarp sets itself apart from existing approaches in two key ways. First, it decouples feature learning from similarity matching by utilizing Laplacian pyramids to create residual deformation fields at each level of a Unet [20], allowing it to capture deformations from coarse to fine. Second, unlike conventional learning-based methods that entangles features of moving and fixed images, MemWarp uses the fixed image's feature map to steer the creation of dynamic filters. These filters, tailored to specific anatomical regions, improve the model's ability to manage discontinuities across different areas. MemWarp's performance is validated on a public cardiac dataset [5], where it surpasses other state-of-the-art semi-supervised methods by a large margin. The main findings of this study are as follows:

- Decoupling feature extraction from similarity matching yields registration accuracies on par with intertwined methods in unsupervised contexts;
- This decoupling allows flexible use of fixed feature maps, leading to a memory network that retains dynamic filters specific to anatomical regions to promote local discontinuities;
- MemWarp excels beyond all leading semi-supervised methods in registration accuracy, achieving a significant 7.1% improvement in Dice score; it outperforms discontinuity-preserving models without needing the segmentation masks for inference that are typically required by these approaches.

2 Methodology

2.1 Preliminaries

Deformable image registration aims to establish voxel-level correspondences between a moving image $\mathbf{I}_m$ and a fixed image $\mathbf{I}_f$. The spatial relationship is represented by $\phi(x) = x + \mathbf{u}(x)$, where x is a spatial location within the domain $\mathbf{\Omega} \subset \mathbf{R}^{H \times W \times D}$, and $\mathbf{u}(x)$ denotes the displacement vector at that location. In unsupervised learning, a network F_θ is trained to predict the deformation field ϕ, with its weights θ optimized by minimizing a composite loss function $\mathcal{L}$. This function combines metrics for dissimilarity between the warped moving image and the fixed image, and the smoothness of the deformation field: $\mathcal{L} = \mathcal{L}_{sim}(\mathbf{I}_f, \mathbf{I}_m \circ \phi) + \lambda \mathcal{L}_{reg}(\phi)$. Here, λ serves to balance the smoothness constraint on the deformation field, with methods like the discontinuous regularizer proposed by Ng et al. [18] falling under this strategy. Semi-supervised methods, including our MemWarp, introduce an additional Dice loss $\mathcal{L}_{dsc}(\mathbf{J}_f, \mathbf{J}_m \circ \phi)$ to assess the dissimilarity between the warped moving mask and the fixed mask. Weakly-supervised models need mask inputs for the network F_θ. For instance, DDIR [9] requires both moving and fixed masks, while textSCF [8] requires only the fixed mask.

2.2 Laplacian Pyramid Warping Network

To decouple feature extraction from similarity matching, we develop a Laplacian pyramid warping network (LapWarp) that leverages residual deformation fields across multiple scales, from coarse to fine. Contrary to previous method LapIRN [16,17], which applies image pyramids directly to raw images, LapWarp performs warping on feature maps and allows for interactions at all levels of the pyramid. This ensures stable training within its pyramid framework without requiring the warm starts or multi-stage coarse-to-fine training strategies.

Network Architecture: LapWarp deviates from classic Unet by stacking moving and fixed images across the batch dimension and employing a unique decoder structure. In each decoder level, moving image features are first warped using the previous level's field. A standard decoder layer then extracts features from both images as a batch, which a flow generator uses at each pyramid level to create the residual deformation field by re-stacking features along channels.

Given n pyramid levels, we obtain n residual deformation fields, labeled from $\Delta\tilde{\phi}_n$ to $\Delta\tilde{\phi}_1$, and $n+1$ total deformation fields, labeled from ϕ_{n+1} to ϕ_1, with both sets following the convention that a larger index indicates a coarser level. At level $i+1$, the feature maps $\hat{\mathbf{I}}_{m_{i+1}}$ and $\hat{\mathbf{I}}_{f_{i+1}}$ are generated by its decoder d_{i+1}. These feature maps, stacked along the channel dimension, are processed by the flow generator f_{i+1} to produce the residual deformation field $\Delta\phi_{i+1} = f_{i+1}(\hat{\mathbf{I}}_{m_{i+1}} \oplus \hat{\mathbf{I}}_{f_{i+1}})$. This residual field is then combined with the upsampled and scaled (by a factor of 2) deformation field $\tilde{\phi}_{i+2}$ from level $i+2$, resulting in the deformation field for level $i+1$, given by $\phi_{i+1} = \Delta\phi_{i+1} + \tilde{\phi}_{i+2}$. For the i_{th} level, the encoder feature maps $\mathbf{I}_{m_i}$ and $\mathbf{I}_{f_i}$, together with upsampled decoder

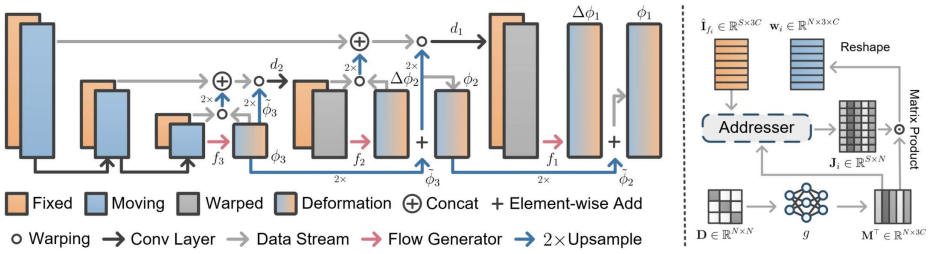

Fig. 1: Schematic representation of the MemWarp framework. The left panel depicts a 2-level LapWarp network employing Laplacian image pyramids; the right panel outlines the operation of the memory network.

feature maps $\tilde{\mathbf{I}}_{m_{i+1}}$ and $\tilde{\mathbf{I}}_{f_{i+1}}$, undergo processing by the decoder layer d_i and flow generator f_i to produce this level's deformation field:

$$\begin{aligned}
\hat{\mathbf{I}}_{f_i} &= d_i(\mathbf{I}_{f_i} \oplus \tilde{\mathbf{I}}_{f_{i+1}}),\\
\hat{\mathbf{I}}_{m_i} &= d_i((\mathbf{I}_{m_i} \circ \tilde{\phi}_{i+1}) \oplus (\tilde{\mathbf{I}}_{m_{i+1}} \circ \Delta\tilde{\phi}_{i+1})),\\
\Delta\phi_i &= f_i(\hat{\mathbf{I}}_{m_i} \oplus \hat{\mathbf{I}}_{f_i}),\\
\phi_i &= \Delta\phi_i + \tilde{\phi}_{i+1},
\end{aligned} \tag{1}$$

where $\Delta\tilde{\phi}_{i+1}$ and $\tilde{\phi}_{i+1}$ are the upsampled and scaled deformation fields from the previous level, $\Delta\phi_i$ denotes the residual deformation field at level i, and ϕ_i represents the cumulative deformation field at this level. It's worth noting that when $i+1$ is the coarsest level, we treat $\tilde{\phi}_{i+2}$ as an identity grid (with zero displacement field). Fig. 1 depicts a two-level LapWarp for visual illustration. Typically, the number of pyramid levels corresponds to the count of downsampling layers.

2.3 Discontinuity-Preserving Deformable Registration

DDIR [9] is the first neural network solution to generate high-quality, discontinuity-preserving deformation fields, but it requires segmentation masks for both training and inference, linking deformation field quality to segmentation accuracy. Additionally, DDIR's use of masks increases computational load by splitting image pairs per anatomical region. MemWarp tackles these challenges by incorporating a memory network [21] that adaptively learns prototypical feature representations for different anatomical regions. Empirical evidence suggests that learning such prototypical features is not feasible when features from moving and fixed images are entangled, which led to the development of LapWarp.

Anatomical Filters: Typically, the flow generator uses convolutional or self/cross-attention layers as in transformers, ending with a single convolutional filter of kernel size 1 to produce the deformation field. Our approach replaces this filter with dynamic filters [8, 26], adapting to the voxel context based on fixed feature maps. Given x as a location vector within $\mathbf{\Omega} \subset \mathbb{R}^{H_i \times W_i \times D_i}$, let $\hat{\mathbf{I}}_{m_i}$ and $\hat{\mathbf{I}}_{f_i}$ represent the moving and fixed feature maps from the decoder

at pyramid level i, respectively. The function f_{i_c} denotes the operation of the convolutional layer. With $\mathbf{w}_i(x) \in \mathbb{R}^{C\times 3}$ as the designated filter at position x in the flow generator, the residual displacement vector at x is defined as $\Delta\phi_i(x) = \mathbf{w}_i(x)^\top f_{i_c}(\hat{\mathbf{I}}_{m_i} \oplus \hat{\mathbf{I}}_{f_i})(x)$, where $f_{i_c}(\hat{\mathbf{I}}_{m_i} \oplus \hat{\mathbf{I}}_{f_i})(x) \in \mathbb{R}^C$ yields the context vector at x. Unlike the conventional approach that applies a uniform $\mathbf{w}_i(x)$ across all locations, our method allows for dynamic filter generation.

Filter generation involves a memory query, addressing, and reconstruction process. The fixed feature vector $\hat{\mathbf{I}}_{f_i}(x) \in \mathbb{R}^{3C}$ acts as the query, with $\mathbf{M} \in \mathbb{R}^{3C\times N}$ representing the memory matrix containing N slots, where N denotes the number of anatomical regions. Instead of storing $\mathbf{M}$ directly as learnable parameters, it is produced by a multi-layer perceptron (MLP).

Memory Addressing & Filter Generation: Define $\mathbf{D} \in \mathbb{R}^{N\times N}$ as a diagonal matrix filled with ones and g as the MLP operation. The memory matrix $\mathbf{M}$ is derived as $\mathbf{M} \in \mathbb{R}^{3C\times N} = g(\mathbf{D})$. Utilizing the fixed feature map $\hat{\mathbf{I}}_{f_i} \in \mathbb{R}^{S_i\times 3C}$ ($S_i = H_i \times W_i \times D_i$) as the query, memory addressing and filter generation proceed as follows:

$$\mathbf{J}_i = \text{softmax}\left(\frac{\hat{\mathbf{I}}_{f_i}\mathbf{M}}{\|\mathbf{M}\|_{2,a_1}}\right), \tag{2}$$

$$\mathbf{w}_i = \text{reshape}(\mathbf{J}_i\mathbf{M}^\top), \tag{3}$$

where the division by $\|\mathbf{M}\|_{2,a_1}$ applies L_2 normalization along the 1_{st} axis of the tensor $\mathbf{M}$, and the softmax is then applied along the 2_{nd} axis of the tensor. With $\mathbf{w}_i$ obtained, the deformation field is generated in accordance with Eq. (1). The reshape function transforms $\mathbf{w}_i \in \mathbb{R}^{S\times 3C}$ into $\mathbf{w}_i \in \mathbb{R}^{S\times 3\times C}$.

Anatomical Region Loss: Anatomical Region Loss: The feature representation $\hat{\mathbf{I}}_{f_i}(x)$at pyramid level i of the fixed image produces the memory-addressed $\mathbf{J}_i \in \mathbb{R}^{S\times N}$, which acts as a segmentation probabilities across N regions. Across all pyramid levels, we apply Dice loss: $\mathcal{L}_{rgn} = \sum_i^n \text{DSC}(\text{up}(\mathbf{J}_i), \mathbf{J}_f)\times \frac{1}{2^{i-1}}$, where $\mathbf{J}_i$ is the network output, $\mathbf{J}_f$ is the fixed segmentation mask from the dataset, and the up function upsamples $\mathbf{J}_i$ to match $\mathbf{J}_f$'s resolution.

2.4 Loss function & Overall Framework

The composite loss function for MemWarp is formulated as:

$$\mathcal{L} = \mathcal{L}_{sim}(\mathbf{I}_f, \mathbf{I}_m \circ \phi) + \mathcal{L}_{dsc}(\mathbf{J}_f, \mathbf{J}_m \circ \phi) + \lambda_1\mathcal{L}_{reg} + \mathcal{L}_{rgn}, \tag{4}$$

with $\mathcal{L}_{reg} = \sum_{x\in\Omega} ||\nabla \mathbf{u}_i(x)||^2$ ($\mathbf{u}_i(x) = \phi_i(x) - x$) and λ adjusting the smoothness regularization strength. The framework of MemWarp aligns with traditional registration frameworks like VoxelMorph but introduces three critical adjustments: 1) Moving and fixed images are combined along the batch dimension; 2) Flow generators, enhanced by memory networks, supplement a conventional Unet, yielding a gradually warped moving image for each decoder level; 3) Deep supervision [14] is employed on the memory-addressed tensors to encourage discontinuities across regions.

3 Experiments & Results

We evaluate MemWarp's effectiveness using the ACDC dataset [5], which includes 150 subjects. Each subject is provided with images from both End-diastole (ED) and End-systole (ES) phases alongside segmentation masks. For intra-subject registration, images from both ED to ES and ES to ED phases are required to be aligned, resulting in a total of 300 pairs ($[100+50]\times 2$). Of these, 170 pairs are allocated for training, 30 for validation, and the remaining 100 for testing. The distribution is stratified to ensure subjects with various diseases are evenly represented across training, validation, and testing phases, with no overlap of subjects between training or validation and testing. All images undergo a min-max normalization to (0,1), are resampled to a voxel size of $1.8\times 1.8\times 10$ mm and adjusted to a size of $128\times 128\times 16$.

3.1 Implementation Details & Comparator Methods

Experiments use Python 3.7 and PyTorch 1.9.0 [19] on a machine equipped with an A100 GPU, and a 16-core CPU with 32GB RAM. Training employs the Adam optimizer with a learning rate of 4e-4, a batch size of 4, and cosine decay, running for 400 epochs. The Mean Square Error (MSE) serves as the similarity loss, complemented by L2 regularization on the spatial gradients of the deformation field ($\lambda = 0.01$ in Eq. (4)), following [3, 10], with seven integration steps in the diffeomorphic layer. For MemWarp, a diffeomorphic layer is used at all pyramid levels except the first. Other models apply only MSE loss, Dice Loss, and regularization as outlined in Eq. (4)'s initial three terms.

Comparator Methods: MemWarp is benchmarked against top learning-based models such as VoxelMorph [3], TransMorph [7], LKU-Net [12], and Slicer Network [25], as well as DDIR [9] which is recognized for its discontinuity-preserving capabilities in cardiac registration. For DDIR, we employ the leading model nnFormer [27] for segmentation, achieving a Dice score of 90.15% on the test set. Slicer Network is assessed with an added guidance loss per its original configuration, while MemWarp and the other models are tested under a consistent experimental framework. We also include traditional methods like ANTs [2] and Demons [24]. While MemWarp is model-agnostic, we utilize the backbone of LKU-Net in this implementation.

Evaluation Metrics: Aligned with standard practices [3, 7], our evaluation employs the Dice coefficient and the 95th percentile Hausdorff Distance (HD95) for anatomical alignment evaluation. HD95 values are averaged across all anatomical structures for individual subjects. Additionally, the standard deviation of the logarithm of the Jacobian determinant (SDlogJ) is utilized to evaluate the quality of diffeomorphism.

3.2 Results & Analysis

Registration Accuracy: Table 1 illustrates that all methods produce smooth displacement fields with low SDlogJ values; however, increased SDlogJ alongside higher Dice scores indicates inherent discontinuities in cardiac alignments.

Table 1: Comparative analysis of MemWarp (LapWarp denotes a MemWarp variant wihtout the memory module) and other models on the test set of the ACDC dataset, with top performing metric in bold. Metrics include Average Dice (%), RV Dice (%), LVM Dice (%), LVBP Dice (%), HD95 (mm), and SDlogJ, with lower values preferred for HD95 and SDlogJ. For clarity, models are categorized as unsupervised (trained solely on raw images), semi-supervised (using segmentation masks in training), and weakly-supervised (requiring masks during both training and inference).

Model	Type	Avg. (%)	RV (%)	LVM (%)	LVBP (%)	HD95 (mm) ↓	SDlogJ ↓
Initial	-	58.14	64.50	48.33	61.60	11.95	-
ANTs [2]	Traditional	71.04	68.61	67.53	76.96	13.15	0.056
Demons [24]		72.37	70.85	69.34	76.93	11.46	0.031
Bspline [15]		74.36	72.18	71.68	79.22	11.18	0.030
TransMorph [7]	Unsupervised	74.97	73.08	71.49	80.34	9.44	0.045
VoxelMorph [3]		75.26	73.10	71.80	80.88	9.33	0.044
LKU-Net [7]		76.53	74.25	73.23	82.12	9.13	0.049
Slicer Network [25]		79.52	77.83	76.80	83.93	8.21	0.044
LapWarp (ours)		77.25	75.86	73.92	81.99	9.23	0.074
TransMorph [7]	Semi-supervised	81.08	81.73	75.27	86.23	7.51	0.091
VoxelMorph [3]		81.34	82.03	75.35	86.64	6.87	0.082
LKU-Net [7]		83.08	84.59	77.24	87.39	6.43	0.099
Slicer Network [25]		83.68	84.94	77.97	88.12	6.10	0.099
MemWarp (ours)		**89.61**	89.30	**86.49**	**93.04**	**3.93**	0.149
DDIR [9]	Weakly-supervised	88.03	**90.02**	85.47	87.61	9.91	0.121

Among unsupervised learning-based models, all outperform traditional methods, with Slicer Network at the forefront due to its large effective receptive field (ERF) and TransMorph lagging, likely hindered by insufficient training data for its transformer architecture. In semi-/weakly-supervised contexts, MemWarp and DDIR, which focus on preserving discontinuities, lead the pack. Despite Slicer Network's strong unsupervised performance, its limited handling of local discontinuities relegates it behind MemWarp. Notably, MemWarp surpasses all semi-supervised methods with a significant 7.1% Dice score gain. DDIR, while competitive, shows potential drawbacks from segmentation inaccuracies, indicated by a higher HD95 value.

Ablation Analysis: Table 2 details our ablation study, examining the impact of the Laplacian pyramid, the memory network, and the inclusion of Dice loss. The base model, labeled as #1, functions as the backbone network in the unsupervised setting, with enhancements observed in #2 upon integrating the Laplacian pyramid. In the semi-supervised scenario, the memory network generates segmentation masks comparable in accuracy to top-tier models like nnFormer [27], utilized in DDIR's mask generation (89.68% vs 90.15%). Yet, we observe that excluding $\mathcal{L}_{dsc}$ can tilt the network's focus towards segmentation, which consequently degrades registration accuracy and the smoothness of the displacement field, as evidenced by #4. Moreover, comparing #4 and #5, registration accuracy improves even in the absence of the Laplacian pyramid when $\mathcal{L}_{dsc}$ is included. The optimal performance in both registration and segmentation is achieved when all three components are included, as with #6.

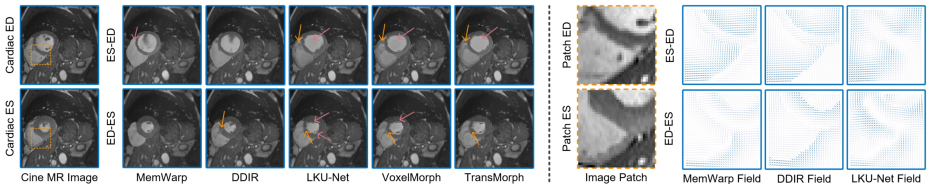

Fig. 2: Comparative visualization of MemWarp against other methods on cardiac MR images, highlighting deformable registration across ED ⇔ ES phases. Pink arrows show omitted trabeculations; orange arrows identify artifacts. The right panel focuses on deformation fields outlined by the left panel's yellow dash, with arrow darkness indicating displacement magnitude.

Qualitative Analysis: Fig. 2 showcases MemWarp's qualitative performance. Notably, MemWarp minimizes artifacts and consistently captures cardiac structures like trabeculations. DDIR's artifacts, particularly between the ventricles, may stem from segmentation inaccuracies. MemWarp and DDIR both display clear organ boundary discontinuities, in contrast to LKU-Net's blending of these regions. MemWarp also manages background deformations adeptly, avoiding DDIR's tendency to reduce displacement magnitude. Within organs, MemWarp finely tunes deformation with respect to the underlying texture instead of overly smoothing the field.

3.3 Discussions

Let $\mathbf{I}_{m_i}$ and $\mathbf{I}_{f_i}$ be the feature maps of moving and fixed images at pyramid level i. MemWarp operates under the assumption that the 'brightness' at any given location $p \in \Omega$ in $\mathbf{I}_{f_i}$ remains constant compared to moving image [11], which is formulated as:

$$\nabla \mathbf{I}_{f_i}(p) \cdot \mathbf{u}(p) = \mathbf{I}_{m_i}(p) - \mathbf{I}_{f_i}(p), \tag{5}$$

where $\nabla \mathbf{I}_{f_i}(p) = \left[\frac{\partial \mathbf{I}_{f_i}}{\partial x}(p), \frac{\partial \mathbf{I}_{f_i}}{\partial y}(p), \frac{\partial \mathbf{I}_{f_i}}{\partial z}(p)\right]^T$. Eq. (5) holds provided that the magnitude of $\mathbf{u}(x)$ is less than one voxel. In the MemWarp framework, we employ an n-level Laplacian image pyramid to ensure $2^{(n-1)} > d_{max}$, where d_{max} is the maximum possible displacement magnitude. This setup ensures that the coarsest level meets the conditions of Eq. (5), with each finer level processing a pre-warped moving image, thus maintaining the model's assumption throughout all levels.

Based on the assumption, we've implemented two major modifications in neural network architecture to enhance registration performance. First, we decouple feature learning from flow estimation. Unlike traditional registration networks that combine moving and fixed images at the network's input, MemWarp employs a U-net for feature extraction and adds a simple convolution layer at each pyramid level to compute flow and performs warping, ensuring each level satisfies Eq. (5). Second, the smoothness requirement of Eq. (5) aligns well with features derived from segmentation networks, as segmentation can be regarded as the ultimate form of image harmonization [6]. This reinforces that effective segmentation features are equally beneficial for registration. Consequently, MemWarp

Table 2: Ablation results outlining the individual and combined contributions of the Laplacian pyramid, memory network, and Dice loss to the performance of our model, achieving optimal outcomes when all three modules are employed.

Model ID	Pyramid	$\mathcal{L}_{dsc}$	Memory	Type	Avg. (%)	HD95 (mm) ↓	SDlogJ ↓	Seg Dice (%)
# 1	×	×	×	Unsupervised	76.53	9.13	0.049	-
# 2	✓	×	×		77.25	9.23	0.074	-
# 3	×	✓	×	Semi-supervised	83.08	6.43	0.099	-
# 4	✓	×	✓		74.81	9.26	0.950	89.06
# 5	×	✓	✓		85.87	5.32	0.085	87.76
# 6	✓	✓	✓		89.61	3.93	0.149	89.68

uses fixed feature maps to steer dynamic filter creation, enhancing feature map smoothness within organs and preserving local discontinuities across boundaries.

4 Conclusions

In conclusion, MemWarp establishes a new benchmark for cardiac registration, outperforming existing methods by effectively preserving essential anatomical details and reducing artifacts. Its success hinges on two pivotal elements: the decoupling of moving and fixed feature maps via LapWarp, and the memory network's use of region loss for maintaining discontinuities across boundaries. MemWarp's effectiveness is validated by a significant 7.1% Dice score enhancement over the nearest semi-supervised competitors. Moreover, MemWarp uniquely addresses discontinuities without needing segmentation masks at inference, yet it can still generate segmentation masks comparable to top segmentation methods.

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

References

1. Ashburner, J.: A fast diffeomorphic image registration algorithm. Neuroimage **38**(1), 95–113 (2007)
2. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration. Neuroimage **54**(3), 2033–2044 (2011)
3. 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)
4. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International journal of computer vision **61**, 139–157 (2005)
5. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep Learning Techniques for Automatic MRI Cardiac Multi-structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Transactions on Medical Imaging **37**(11), 2514–2525 (2018)

6. Blake, A., Zisserman, A.: Visual reconstruction. MIT press (1987)
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. Chen, X., Liu, M., Wang, R., Hu, R., Liu, D., Li, G., Zhang, H.: Spatially covariant image registration with text prompts (2024)
9. Chen, X., Xia, Y., Ravikumar, N., Frangi, A.F.: A deep discontinuity-preserving image registration network. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24. pp. 46–55. Springer (2021)
10. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised Learning of Probabilistic Diffeomorphic Registration for Images and Surfaces. Medical Image Analysis **57**, 226–236 (2019)
11. Horn, B.K., Schunck, B.G.: Determining optical flow. Artificial intelligence **17**(1-3), 185–203 (1981)
12. Jia, X., Bartlett, J., Zhang, T., Lu, W., Qiu, Z., Duan, J.: U-net vs transformer: Is u-net outdated in medical image registration? In: International Workshop on Machine Learning in Medical Imaging. pp. 151–160. Springer (2022)
13. Khalil, A., Ng, S.C., Liew, Y.M., Lai, K.W.: An overview on image registration techniques for cardiac diagnosis and treatment. Cardiology research and practice **2018** (2018)
14. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets. In: Artificial intelligence and statistics. pp. 562–570. Pmlr (2015)
15. Marstal, K., Berendsen, F., Staring, M., Klein, S.: SimpleElastix: A User-friendly, Multi-lingual Library for Medical Image Registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 134–142 (2016)
16. Mok, T.C., Chung, A.C.: Large deformation diffeomorphic image registration with laplacian pyramid networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23. pp. 211–221. Springer (2020)
17. Mok, T.C., Chung, A.C.: Large deformation image registration with anatomy-aware laplacian pyramid networks. In: Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data: MICCAI 2020 Challenges, ABCs 2020, L2R 2020, TN-SCUI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 23. pp. 61–67. Springer (2021)
18. Ng, E., Ebrahimi, M.: An unsupervised learning approach to discontinuity-preserving image registration. In: Biomedical Image Registration: 9th International Workshop, WBIR 2020, Portorož, Slovenia, December 1–2, 2020, Proceedings 9. pp. 153–162. Springer (2020)
19. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
21. Sukhbaatar, S., Weston, J., Fergus, R., et al.: End-to-end memory networks. Advances in neural information processing systems **28** (2015)

22. Timmis, A., Vardas, P., Townsend, N., Torbica, A., Katus, H., De Smedt, D., Gale, C.P., Maggioni, A.P., Petersen, S.E., Huculeci, R., et al.: European society of cardiology: cardiovascular disease statistics 2021. European Heart Journal **43**(8), 716–799 (2022)
23. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage **45**(1), S61–S72 (2009)
24. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., et al.: Diffeomorphic Demons Using ITK's Finite Difference Solver Hierarchy. The Insight Journal **1** (2007)
25. Zhang, H., Chen, X., Wang, R., Hu, R., Liu, D., Li, G.: Slicer networks (2024)
26. Zhang, H., Wang, R., Zhang, J., Liu, D., Li, C., Li, J.: Spatially covariant lesion segmentation. In: Elkind, E. (ed.) Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23. pp. 1713–1721. International Joint Conferences on Artificial Intelligence Organization (8 2023). https://doi.org/10.24963/ijcai.2023/190, https://doi.org/10.24963/ijcai.2023/190, main Track
27. Zhou, H.Y., Guo, J., Zhang, Y., Han, X., Yu, L., Wang, L., Yu, Y.: nnformer: Volumetric medical image segmentation via a 3d transformer. IEEE Transactions on Image Processing (2023)