UNDERSTANDING DOMAIN-SHIFT IMMUNITY IN DEEP DEFORMABLE REGISTRATION

Mingzhen Shao

Kahlert School of Computing
Scientific Computing and Imaging Institute
University of Utah

Sarang Joshi

Department of Biomedical Engineering
Scientific Computing and Imaging Institute
University of Utah

2026 IEEE International Conference on Image Processing (ICIP) | 979-8-3315-5151-3/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICIP61757.2026.11630227

ABSTRACT

Deep learning has achieved remarkable success in deformable image registration, yet the visual information that drives deformation estimation remains poorly understood. Rather than pursuing incremental performance improvements, this work investigates the fundamental source of robustness in deep registration models. Using diverse, domain-agnostic synthetic datasets, we decouple deformation learning from application-specific appearance and show that domain-shift immunity is an inherent, largely architecture-agnostic property of deep deformable registration when trained with a robust pipeline. To identify the mechanism underlying this immunity, we compare models trained directly on raw image intensities with models operating exclusively on local feature representations extracted by a fixed, pre-defined feature extractor. The comparable performance of these models provides strong empirical evidence that deformation estimation is governed primarily by local structural features, rather than global, domain-specific appearance cues. These findings offer a principled explanation for the cross-domain generalizability of deep registration networks and point toward feature-centric designs for domain-independent registration.

***Index Terms*—** Deformable registration, Domain-shift immunity, Local feature representation

1. INTRODUCTION

Recent advances in deep learning have substantially transformed image registration. A wide range of learning-based models have been proposed for both rigid and deformable registration tasks [1, 2, 3, 4, 5, 6, 7], demonstrating notable improvements in both accuracy and computational efficiency over traditional optimization-based approaches. Despite these successes, a fundamental question remains unresolved: **Do deep registration models remain robust under domain shift?**

Unlike recognition or segmentation tasks, which typically operate within a well-defined image distribution, image registration inherently demands robustness across domains. Classical optimization-based registration methods rarely faced this challenge, as their behavior was largely deterministic and independent of training data. In contrast, learning-based registration models rely on representations learned from specific data distributions, making their behavior under domain shift significantly harder to characterize.

Most prior efforts to improve generalization in deep registration focus on scaling training data diversity or pretraining on large datasets [8, 9, 10, 11]. While effective in practice, these approaches provide limited insight into why robustness emerges. SynthMorph [6], for example, demonstrates impressive cross-domain performance by training exclusively on synthetic data, yet the underlying mechanism that enables such robustness remains largely unexplained.

Recent work on deep rigid registration [12] provides an important clue, showing that domain-shift immunity can arise because deformation estimation is driven primarily by local structural features rather than global appearance statistics. This observation motivates the central question of this paper: **Does the same structure-inherent domain-shift immunity extend to deep *deformable* registration models, despite their increased spatial complexity and flexibility?**

Although several recent studies have explored feature-based representations for registration [13, 14], these methods primarily focus on optimization-based frameworks. The role of explicit feature representations in deep deformable registration—and their connection to domain-shift robustness—remains insufficiently understood.

In this work, we first demonstrate that deep learning–based deformable registration models exhibit strong cross-domain robustness, even when evaluated on unseen domains without fine-tuning. We show that this robustness is an inherent and largely architecture-agnostic property of deep registration models. To identify the source of this immunity, we adopt a controlled synthetic training framework inspired by SynthMorph and systematically replace raw image intensities with fixed local feature representations. The observed parity in performance reveals that deformation estimation is governed primarily by local structural features, rather than domain-specific appearance cues.

Before proceeding, we clarify our use of the term *domain-*

shift immunity. For deterministic, non-learning-based methods such as SyN [15], registration accuracy may vary across domains; however, such variation does not imply a loss of immunity. What matters is whether the method remains functionally effective—namely, capable of producing meaningful, smooth, and anatomically plausible deformation fields regardless of input domain. Under this definition, domain-shift immunity does not require identical accuracy across domains. Differences in accuracy can often be addressed through architectural choices, regularization strategies, or similarity metrics, and are not the focus of this study. Instead, our goal is to determine whether deep deformable registration models retain the ability to perform valid registration under domain shifts, in a manner comparable to classical methods.

2. METHOD

The goal of this work is to analyze the robustness of deep deformable registration models and to identify the source of such robustness. A key requirement for this analysis is a training and evaluation setup that produces consistent ground-truth deformations across different image domains. To this end, we adopt a synthetic data generation strategy inspired by SynthMorph [6], which allows deformation learning to be decoupled from domain-specific appearance.

2.1. Synthetic deformation generation

We generate a synthetic deformation field ϕ using the method detailed in SynthMorph and apply it consistently across domains. Given a fixed image I_f, the moving image I_m is obtained by warping:

$$I_m = \phi \circ I_f$$

Because the deformation field is generated independently of image appearance, this construction ensures that all training and testing pairs share the same underlying geometric correspondence, enabling controlled evaluation of cross-domain generalization.

2.2. Domain-independent label maps

We use generalized label maps rather than domain-specific images as training data. Each label map contains J distinct regions represented by randomized geometric patterns.

Specifically, we first generate J smoothly varying noise images $\{p_j\}_{j=1}^{J}$, where each p_j is sampled from a standard normal distribution $\mathcal{N}(0, 1)$ at a coarse resolution r_p and then upsampled to the target image size. The final label map s is constructed by assigning to each pixel k the label index corresponding to the noise image with the maximum intensity at that pixel:

$$s_k = \arg\max_j([\tilde{p}_j]_k),$$

where $\tilde{p}_j$ denotes the upsampled version of p_j.

This procedure yields randomized yet spatially coherent label maps that are independent of any specific domain, serving as a generalized training set for the registration network.

We randomly crop a 224x224 patch from the label map s as the fixed image I_f and applying a synthetic deformation field ϕ to I_f produces a transformed map I_m. The resulting pairs (I_f, I_m) serve as our training dataset LabelM.

2.3. Multi-modal data

To address multi-modal registration, we generate controlled and consistent image pairs using three strategies that span a wide range of content domains.

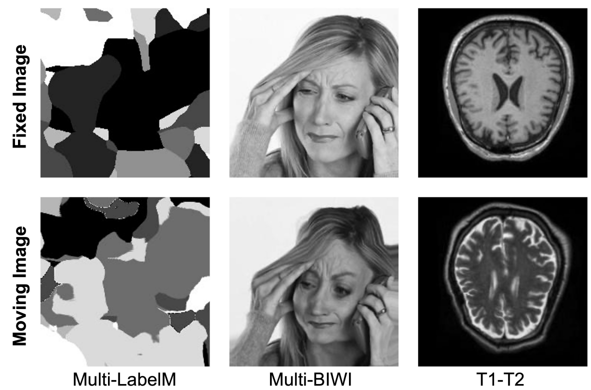

Fig. 1. Cross-domain multi-modal examples

Multi-modal LabelM: A random recoloring process is applied to the I_m in LabelM to create multi-modal pairs, preserving the underlying spatial correspondences.

Multi-modal BSD: We transform a normalized grayscale image $I \in [0, 1]^{H\times W}$ into a color image $I' \in [0, 1]^{H\times W\times 3}$ using a continuous, monotonic mapping

$$C : [0, 1] \to [0, 1]^3,$$

which assigns low-intensity values to dark colors and high-intensity values to bright colors. This transformation preserves the spatial structure of I while altering its photometric distribution, producing a modality-like variation that maintains geometric correspondence across images.

T1-T2 Image Pair: To align with standard multi-modal registration tasks, we also include T1–T2 image pairs from the IXI dataset [1]. We select T1 and T2 volumes acquired from the same scan session to ensure consistent anatomical representation and approximate alignment. From each volume pair, we extract 2D slices at random axial positions to generate multi-modal image pairs.

[1]https://brain-development.org/

2.4. Network architecture

Fig. 2 provides an overview of the proposed UniReg framework. The key design choice is to explicitly isolate local feature representations during training in order to analyze their role in deformation estimation.

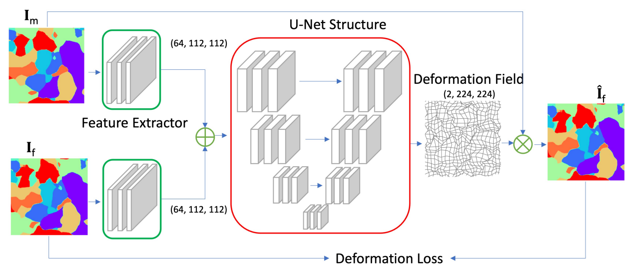

Fig. 2. Overview of the proposed UniReg.

Both the fixed image I_f and moving image I_m are first processed by a shared feature extractor consisting of the first convolutional block of VGG16. We use VGG16 here for demonstration purposes; the supplementary material shows that other architectures can also serve as effective feature extractors. These layers are initialized with ImageNet-pretrained weights and remain frozen throughout training. As is well established, early convolutional layers predominantly capture low-level local structures such as edges and textures, making them well suited for extracting local features.

The extracted feature maps from I_f and I_m are concatenated and passed to a U-Net–style encoder–decoder network. This network predicts a dense deformation field ϕ. An additional upsampling layer ensures that the predicted deformation field matches the spatial resolution of the input images. The moving image is then warped to produce the registered image $\hat{I}_f = \phi \circ I_m$.

Except for the feature-level input, the remainder of UniReg follows the standard U-Net design commonly used in deep deformable registration.

2.5. Loss function

We adopt a standard deformation loss formulation commonly used in deep deformable registration. A smoothness regularization term penalizes spatial gradients of the predicted deformation field to encourage physically plausible transformations. We avoid specialized or task-specific loss designs, such as those used in UniGradICON, as our focus is on analyzing robustness and generalization mechanisms rather than optimizing performance through customized loss formulations.

$$\mathcal{L}_{\text{smooth}} = \frac{1}{N} \sum_{i,j} \Big(\|\nabla_x \phi_{i,j}\|_1 + \|\nabla_y \phi_{i,j}\|_1 \Big), \qquad (1)$$

where $\phi \in \mathbf{R}^{2\times H \times W}$ is the predicted flow field, N is the total number of the deformable field, and ∇_x, ∇_y denote the discrete spatial gradient (finite-difference) operators along the vertical and horizontal axes respectively.

$$\mathcal{L}_{\text{reg}} = \text{MSE}\Big(I_f, \hat{I}_f\Big) + \alpha\, \mathcal{L}_{\text{smooth}}, \qquad (2)$$

where MSE is the mean square error and $\alpha = 0.01$ is a hyperparameter to adjust the smooth strength.

2.6. Validation metrics

Traditional registration evaluation relies on anatomical correspondence measures such as Dice similarity and mean Target Registration Error (mTRE), which require segmentation masks or manually annotated landmarks. Although such annotations are available for the synthetic training data, they are unavailable for several evaluation domains (e.g.,BIWI and BSD). As a result, using annotation-dependent metrics would prevent consistent comparison across domains. Since our goal is to evaluate general-purpose registration performance under domain shift, we instead adopt annotation-free, intensity-based similarity metrics.

Correlation Coefficient (CC): For mono-modal registration, we measure similarity using the CC to capture linear intensity agreement between the fixed image I_f and the warped moving image $\hat{I}_f$:

$$\text{CC}(\mathrm{I_f}, \hat{\mathrm{I}}_\mathrm{f}) = \frac{\sum_{i=1}^{n} (I_f(i) - \bar{I}_f)(\hat{I}_f(i) - \overline{\hat{I}_f})}{\sqrt{\sum_{i=1}^{n} (I_f(i) - \bar{I}_f)^2 \sum_{i=1}^{n} (\hat{I}_f(i) - \overline{\hat{I}_f})^2}}, \qquad (3)$$

where $\bar{I}_f$ and $\overline{\hat{I}_f}$ denote the mean intensities of I_f and $\hat{I}_f$, respectively.

Mutual Information (MI): For multi-modal registration, we use MI to measure statistical dependence rather than direct intensity correspondence:

$$\text{MI}(I_f, \hat{I}_f) = H(I_f) + H(\hat{I}_f) - H(I_f, \hat{I}_f), \qquad (4)$$

where $H(\cdot)$ and $H(\cdot, \cdot)$ denote marginal and joint entropies.

Deformation Regularity: To assess physical plausibility, we examine the Jacobian matrix,

$$J_\phi(x) = \frac{\partial \phi(x)}{\partial x},$$

whose determinant $\det(J_\phi(x))$ measures the local change in area induced by the transformation. Regions where $\det(J_\phi(x)) \leq 0$ correspond to *folding*, indicating that the mapping is no longer locally invertible. We quantify deformation regularity using the percentage of non-positive Jacobians:

$$\%|J| = \frac{\big|\{x \mid \det(J_\phi(x)) \leq 0\}\big|}{|\Omega|} \times 100\%.$$

Lower $\%|J|$ indicates smoother and more anatomically consistent deformations.

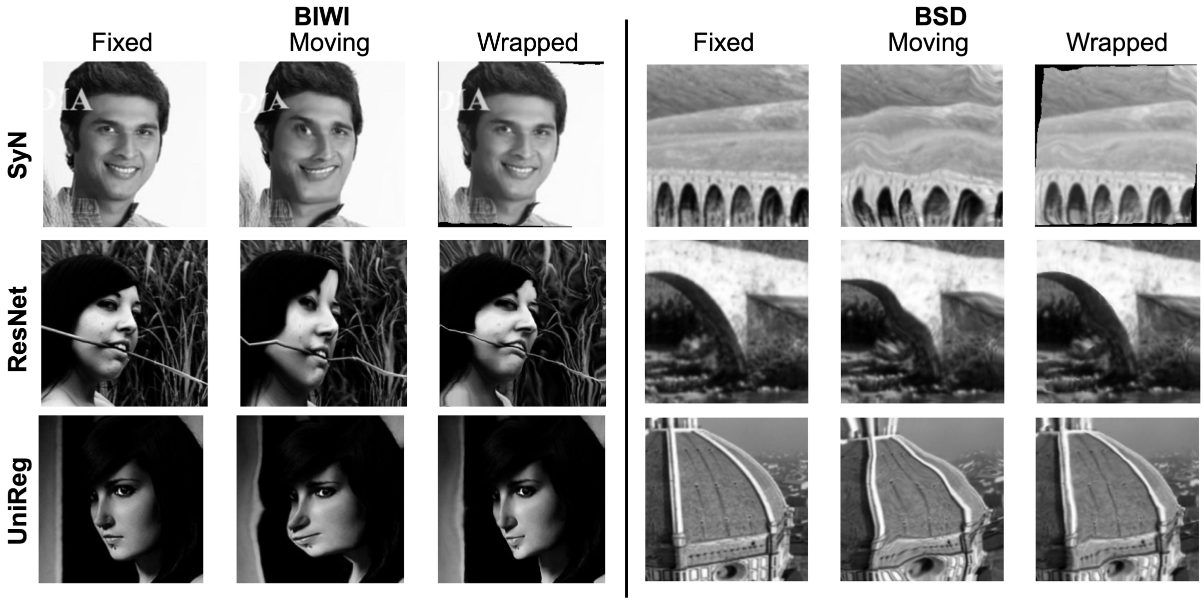

Fig. 3. Example mono-modal registration from different models.

3. EXPERIMENT

In this section, we first evaluate mono-modal registration across diverse image domains and then extend the analysis to multi-modal settings. In both cases, we show that deep registration models retain strong cross-domain performance and use the proposed UniReg framework to demonstrate that this domain-shift immunity arises from a reliance on local feature representations.

Our experiments are designed to explain the generality and mechanism of this robustness rather than to optimize task-specific accuracy. Accordingly, we do not focus on specialized architectures or training strategies developed for maximal performance in particular medical imaging applications, such as VoxelMorph or GradICON. These methods largely share a common U-Net–based deformation backbone and differ mainly in regularization or optimization details, offering limited additional insight into the source of domain-shift immunity.

For the same reason, we do not evaluate 3D volumetric registration on specific modalities such as MRI or CT. While such experiments are important for clinical deployment, they do not fundamentally affect the questions addressed in this work. Instead, we focus on isolating the core factors governing cross-domain generalization in deep deformable registration under controlled and interpretable settings.

The shaded column in tables indicates that all models were trained on this domain; the remaining columns evaluate cross-domain generalization.

3.1. Mono-model registration

To verify the existence of domain-shift immunity, we follow the same evaluation strategy as in [12], where a model is trained on one domain and directly tested on unseen domains without any additional fine-tuning or modification. We include a non-learning registration method (SyN) as a baseline. As a well-established non-learning registration algorithm, SyN does not rely on domain-specific training and is thus inherently domain-independent. If a deep learning model outperforms this baseline on an unseen domain, it indicates that the model remains functional under domain shifts.

To ensure that our findings are not architecture-specific, we evaluate multiple backbone networks, including VGG16, ResNet50, and U-Net. To achieve diverse dataset coverage, we employ BSD300 (natural images), BIWI (facial images), and NIH chest X-ray [16] (X-ray images), which collectively span a broad range of visual characteristics and application domains. For architectures such as VGG16 and ResNet50, we append a series of upsampling layers after the final convolutional block to ensure that the output deformation field matches the resolution of the input image.

Table 1. Registration results across domains. Values are reported as CC ($\%|J|$), where CC ($\uparrow$) and $\%|J|$ ($\downarrow$).

	LabelM	BIWI	BSD	X-Ray
SyN	0.92(0)	0.89(0)	0.82(0)	0.86(0)
VGG16	0.97(0.82)	0.95(0.47)	0.93(0.37)	0.96(0.11)
Res50	0.97(1.13)	0.95(0.76)	0.94(0.42)	0.97(0.19)
U-Net	0.98(1.24)	0.96(0.92)	0.93(0.54)	0.96(0.13)
UniReg	0.98(1.38)	0.96(0.96)	0.94(0.48)	0.97(0.19)

Table 1 reports the correlation coefficients and $\%|J|$ across the four evaluation domains. All deep learning–based methods achieve low folding rates while consistently outperforming SyN in appearance similarity, indicating that learned registration models generalize effectively to unseen data. Moreover, the consistent performance trends observed across different network architectures demonstrate that domain-shift immunity is a general property of learning-based deformable registration.

We notice that the UniReg achieves accuracy comparable to that of U-Net, and UniReg is trained **purely using local structural features** from the extractor. In contrast, the original U-Net has direct access to the full intensity information from raw input images. Despite this limitation, UniReg still provide a similar accuracy, indicating that local features alone play a critical role in accurate deformation field estimation. Fig. 3 shows example registration results for several models.

3.2. Multi-modal registration

For multi-modal registration, we do not use SyN as a baseline to assess whether a model is functional. SyN is a strong classical method for multi-modal registration, and even recent state-of-the-art deep learning approaches [17] only marginally outperform it. As a result, the simplified models used in this study—designed for analysis rather than peak performance—are not expected to surpass SyN.

Instead, we report the initial mutual information (MI) between each fixed–moving image pair prior to registration and evaluate performance based on the improvement in MI after registration. This allows us to compare robustness and generalization behavior across models without relying on a strong external baseline.

All models are trained using two multi-modal datasets with different levels of appearance diversity: Multi-LabelM (high diversity) and Multi-BIWI (low diversity). Training protocols are kept identical across models, and the same loss function is used throughout. During optimization, the predicted deformation field is applied to the mono-modal image.

Retraining on multi-modal datasets is necessary because, although the models exhibit domain-shift immunity, the feature distributions presented to the U-Net differ substantially between mono-modal and multi-modal inputs. Without retraining, this input-level feature perturbation can obscure the analysis of robustness under modality change.

We report multi-modal registration results for models trained on Multi-LabelM in Tab. 2. Similar to the mono-modal setting, all models achieve robust improvements in MI across domains, largely independent of architectural choice. Notably, the comparable performance of U-Net and UniReg indicates that local structural information is still the primary driver of deformation estimation in the multi-modal setting.

Fig. 4 shows example multi-modal registration results trained on Multi-LabelM for several models. A complete set of results for all models is provided in the supplementary material.

Table 2. Multi-modal registration results (trained on Multi-LabelM). Values are reported as MI ($\%|J|$).

Models	Multi-LabelM	Multi-BIWI	T1-T2
Init MI	1.17	0.80	0.51
VGG16	1.64(0.15)	1.10(0.11)	0.66(0.05)
Res50	1.62(0.44)	1.17(0.10)	0.65(0.11)
U-Net	1.74(0.64)	1.10(0.11)	0.65(0.18)
UniReg	1.69(0.50)	1.05 (0.07)	0.66(0.04)

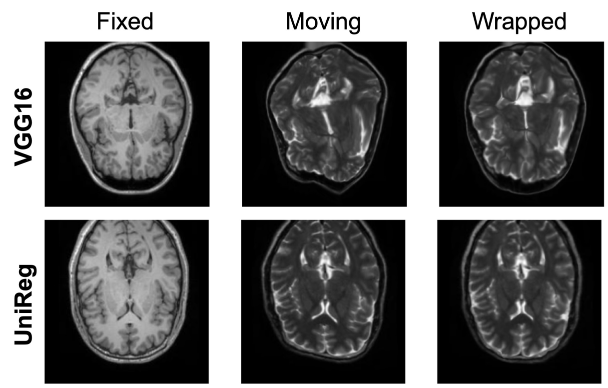

Fig. 4. Multi-modal registration from different models.

3.3. Discussion

Although U-Net remains the dominant architecture for deformable registration, our results suggest that architectures leveraging invariant feature encoders (e.g., VGG or other pretrained extractors) can achieve stronger robustness under unseen domains. Thus, UniReg serves as a proof-of-concept for a broader design principle: decoupling feature representation from deformation estimation can yield domain-stable registration models regardless of the specific backbone or feature extractor used.

4. CONCLUSION

Our study reveals that deep deformable registration models possess structure-inherent domain-shift immunity, driven by their reliance on local feature representations. The analysis with UniReg confirms that maintaining consistent local features is sufficient for robust cross-domain and multi-modal registration. These findings shift the focus from deformation architecture to feature extraction as the key determinant of generalization, suggesting that future research should reexamine backbone design and develop feature-extractor architectures explicitly optimized for cross-domain robustness.

5. ACKNOWLEDGEMENT

This work is supported by NIH Grant R01CA259686.

6. REFERENCES

[1] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich, "Deep image homography estimation," *arXiv*, vol. preprint arXiv: 1606.03798, 2016.

[2] Xiaohuan Cao, Jianhua Yang, Jun Zhang, Dong Nie, Minjeong Kim, Qian Wang, and Dinggang Shen, "Deformable image registration based on similarity-steered cnn regression," in *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*. Springer, 2017, pp. 300–308.

[3] Ty Nguyen, Steven W Chen, Shreyas S Shivakumar, Camillo J Taylor, and Vijay Kumar, "Unsupervised deep homography: A fast and robust homography estimation model," in *Robotics and Automation Letters 3(3)*, 2018, pp. 2346–2353.

[4] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[5] Jiahui Zhang, Dawei Sun, Zixin Luo, Anbang Yao, Lei Zhou, Tianwei Shen, Yurong Chen, Long Quan, and Hongen Liao, "Learning two-view correspondences and geometry using order-aware network," *International Conference on Computer Vision (ICCV)*, 2019.

[6] Malte Hoffmann, Benjamin Billot, Douglas N. Greve, Juan Eugenio Iglesias, Bruce Fischl, and Adrian V. Dalca, "Synthmorph: Learning contrast-invariant registration without acquired images," *IEEE Transactions on Medical Imaging*, vol. 41, no. 3, pp. 543–558, 2022.

[7] Lin Tian, Hastings Greer, François-Xavier Vialard, Roland Kwitt, Raúl San José Estépar, and Marc Niethammer, "Gradicon: Approximate diffeomorphisms via gradient inverse consistency," *arXiv preprint arXiv:2206.05897*, 2022.

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

[9] Zi Li, Jianpeng Zhang, Tai Ma, Tony CW Mok, Yan-Jie Zhou, Zeli Chen, Xianghua Ye, Le Lu, and Dakai Jin, "Unireg: Foundation model for controllable medical image registration," *arXiv preprint arXiv:2503.12868*, 2025.

[10] Junyu Chen, Shuwen Wei, Yihao Liu, Aaron Carass, and Yong Du, "Pretraining deformable image registration networks with random images," *arXiv preprint arXiv:2505.24167*, 2025.

[11] Neel Dey, Benjamin Billot, Hallee E Wong, Clinton J Wang, Mengwei Ren, P Ellen Grant, Adrian V Dalca, and Polina Golland, "Learning general-purpose biomedical volume representations using randomized synthesis," *arXiv preprint arXiv:2411.02372*, 2024.

[12] Mingzhen Shao, Tolga Tasdizen, and Sarang Joshi, "Analyzing the domain shift immunity of deep homography estimation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 4800–4808.

[13] Zi Li, Lin Tian, Tony CW Mok, Xiaoyu Bai, Puyang Wang, Jia Ge, Jingren Zhou, Le Lu, Xianghua Ye, Ke Yan, et al., "Samconvex: Fast discrete optimization for ct registration using self-supervised anatomical embedding and correlation pyramid," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 559–569.

[14] Hanna Siebert, Christoph Großbröhmer, Lasse Hansen, and Mattias P Heinrich, "Convexadam: Self-configuring dual-optimisation-based 3d multitask medical image registration," *IEEE Transactions on Medical Imaging*, 2024.

[15] B.B. Avants, C.L. Epstein, M. Grossman, and J.C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[16] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers, "ChestX-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, jul 2017.

[17] Başar Demir, Lin Tian, Hastings Greer, Roland Kwitt, François-Xavier Vialard, Raúl San José Estépar, Sylvain Bouix, Richard Rushmore, Ebrahim Ebrahim, and Marc Niethammer, "Multigradicon: A foundation model for multimodal medical image registration," in *International Workshop on Biomedical Image Registration*. Springer, 2024, pp. 3–18.