OPS: One-shot Point prompted Semantics-aware Learning for Medical Image Registration*

Housheng Xie, Xiaoru Gao, and Guoyan Zheng (✉)

Institute of Medical Robotics, School of Biomedical Engineering, Shanghai Jiao Tong University, No. 800, Dongchuan Road, Shanghai 200240, China
Corresponding Author: Guoyan Zheng (Email: guoyan.zheng@sjtu.edu.cn)

Abstract. Unsupervised learning is the dominant paradigm in medical image registration, primarily due to its straightforward deployment and independence from labor-intensive manual annotations. However, its optimization objective fails to differentiate between semantically distinct anatomical structures and background regions, resulting in suboptimal registration performance. To introduce semantic priors into unsupervised registration with minimal annotation cost, we propose OPS, a One-shot Point prompt-driven Semantics-aware registration framework. Leveraging the powerful semantic clustering capabilities of the pre-trained vision foundation model DINOv3, OPS requires only extremely sparse point prompts (i.e., one point per organ) on a single reference image slice to extract global semantic anchors. These anchors are then propagated across the entire unlabeled dataset to compute feature similarities with the DINOv3 embeddings of each image volume, thereby localizing semantically consistent regions. This process enables the offline generation of dense semantic similarity maps, which serve as nearly annotation-free anatomical localization priors. Subsequently, we integrate these anatomical localization priors into the registration learning through a dual mechanism: (1) semantic feature embedding to modulate the feature representations, and (2) semantics-aware loss optimization to spatially re-weight the similarity loss, enabling semantic guidance for the registration learning. Comprehensive experiments on two datasets demonstrate that the proposed method enhances the registration performance of state-of-the-art methods. The code is available at https://github.com/xiehousheng/OPS.

Keywords: Medical image registration · Semantic prior · DINOv3 · One-shot · Unsupervised learning.

1 Introduction

Deep learning-based registration methods have become the dominant paradigm in medical image registration due to flexible architectural designs and superior inference efficiency. Among these, unsupervised registration approaches [2,

* This study was partially supported by the National Key R&D Program of China via project 2023YFB4706302 and by the National Natural Science Foundation of China via projects 62471293 and U20A20199.

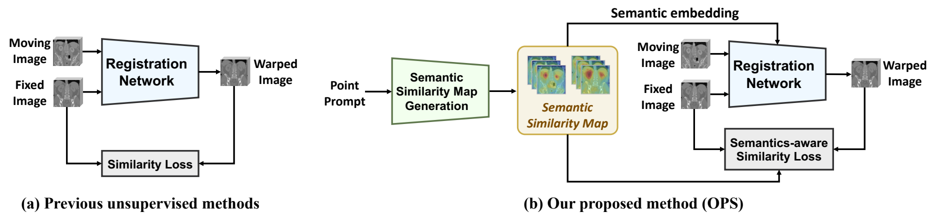


Fig. 1. Comparison between the proposed method and existing unsupervised learning methods. (a) A schematic illustration of existing unsupervised learning approaches which typically rely on global intensity similarity for learning optimization. (b) The proposed OPS leverages sparse point prompts to generate dense semantic similarity maps to guide registration learning.

4, 5, 17, 19, 3, 18, 20] directly exploit similarity metrics between image pairs to drive the network to predict deformation fields, achieving high deployment efficiency without the need for manual annotations. While effective, this optimization mechanism is inherently incapable of perceiving semantic information within images, resulting in suboptimal registration accuracy. In contrast, weakly-supervised registration methods [10, 7, 21] take advantage of manually annotated segmentation or landmarks to guide the network to focus on specific anatomical structures and regions, resulting in improved accuracy over their unsupervised counterparts. However, acquiring such medical image annotations requires significant expert efforts, making them difficult to directly apply to the vast number of unlabeled medical images in clinical practice. This naturally raises a critical challenge: how can we embed semantic guidance into unsupervised registration without relying on such costly manual supervision?

A potential solution lies in leveraging pre-trained foundation models to automatically extract semantic context. For instance, annotations generated by segmentation models [13, 23] based on the Segment Anything Model (SAM) have been utilized as substitutes for manual segmentation labels to train registration models [22]. While this strategy alleviates the annotation burden, SAM typically still necessitates instance-specific prompts for each image and struggles to achieve reliable generalization on anatomical structures unseen during training. Recently, the self-supervised vision foundation model DINOv3 [15] has learned highly generalizable feature representations through large-scale pre-training, while simultaneously exhibiting remarkable semantic clustering properties [11, 12]. Due to the inherent anatomical consistency in medical imaging, these robust representations can implicitly capture structural correspondences across subjects, enabling the generation of dense semantic guidance for registration by relying solely on feature similarity.

Building on the above insight, we propose OPS, a One-shot Point prompt-driven Semantics-aware registration framework. As illustrated in Fig. 1, in contrast to existing unsupervised registration methods, our method introduces semantic awareness into the unsupervised learning process with minimal annotation cost. Specifically, as shown in Fig. 2, we first input extremely sparse point prompts (i.e., one point per organ) on a single reference image slice to extract feature representations from DINOv3 embeddings at the prompted locations, defining them as global semantic anchors. Subsequently, leveraging the cross-image semantic consistency of DINOv3 features, we utilize these anchors to retrieve corresponding regions across the entire unlabeled dataset based on feature similarity. This process propagates semantic priors from the single slice to the entire dataset, enabling the localization of similar anatomical structures across the full cohort with a one-shot image interaction, and generating dense semantic similarity map for each volume in the dataset. Serving as anatomical localization priors, these maps are integrated into the registration network via a dual mechanism comprising semantic feature embedding (SFE) and semantics-aware loss optimization (SLO), thereby injecting explicit semantic guidance into the unsupervised learning process. The contributions of this paper are summarized as follows:

- We propose a one-shot point prompt-driven registration paradigm. By leveraging the cross-image similarity of DINOv3 features, dense anatomical localization priors can be generated for the entire dataset through one-shot image interaction on a single image slice, and subsequently embedded into the registration network to enable semantics-aware registration learning.
- We employ a dual mechanism comprising semantic feature embedding and semantics-aware loss optimization to incorporate the anatomical localization priors into the registration learning process.

2 Methodology

2.1 Overview

Our framework (Fig. 2) leverages DINOv3 to propagate sparse point prompts from a single reference image slice into dense semantic similarity maps covering the entire dataset, followed by integrating them into the registration learning process as anatomical localization priors. This process comprises two key stages: semantic similarity map generation and semantics-aware registration learning.

2.2 Semantic Similarity Map Generation

To generate semantic similarity maps for the registration task, given a dataset of N volumetric images denoted as $\{V^i\}_{i=1}^{N}$, where $V^i \in \mathbb{R}^{H\times W\times D}$, we adopt the slice-wise encoding strategy from DINO-Reg [16] to adapt the 2D foundation model DINOv3. Each volume is sliced along the coronal plane and processed by the frozen DINOv3, yielding a DINOv3 feature embedding F_d^i.

Subsequently, let I_{ref} be a reference image slice selected from the training set. After encoding it to obtain the DINOv3 feature F_{ref}, for each target organ, we extract the feature vector at one manually specified point prompt location P_{ref} in F_{ref} as a global semantic anchor $A_g \in \mathbb{R}^C$. Due to the semantic

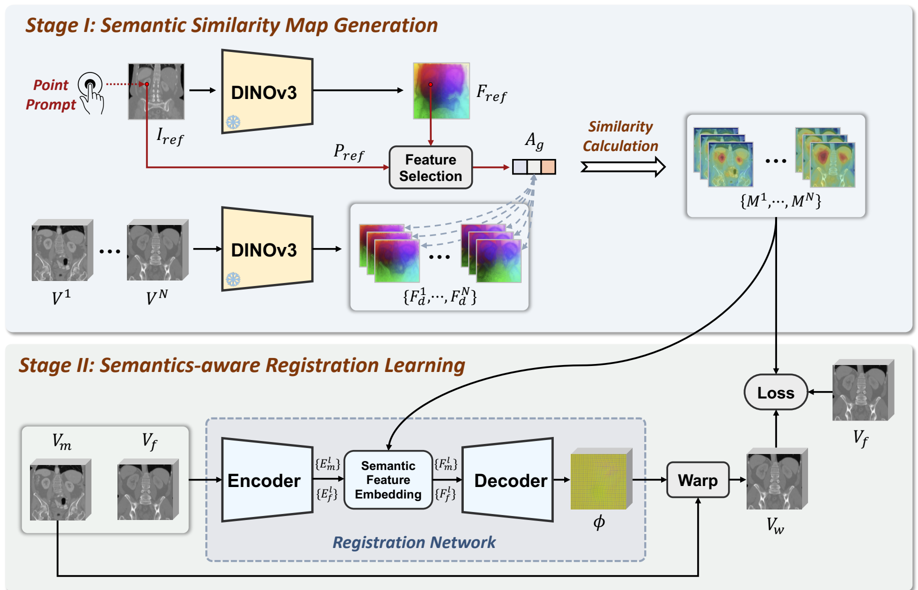


Fig. 2. Overview of our proposed framework. The method implements one-shot point prompt-driven semantics-aware registration via two stages: semantic similarity map generation and semantics-aware registration learning.

clustering property for the output features of DINOv3, features extracted from semantically similar regions across different subjects show high similarity to A_g. This observation motivates us to employ A_g as a query vector to perform similarity matching against the feature volumes F_d^i of all volumes in the dataset. Specifically, we compute the voxel-wise cosine similarity between A_g and each feature volume F_d^i, and then normalize the similarity values. The final semantic similarity map $\mathcal{M}^i$ is formulated as:

$$S^i(x) = \frac{A_g \cdot F_d^i(x)}{\|A_g\|_2 \|F_d^i(x)\|_2}, \quad \mathcal{M}^i(x) = \frac{S^i(x) - \min(S^i)}{\max(S^i) - \min(S^i)}, \tag{1}$$

where S^i denotes the cosine similarity volume, x is the spatial location of a voxel, and $\min(\cdot)$ and $\max(\cdot)$ represent the minimum and maximum values within the volume S^i, respectively.

2.3 Semantics-aware Registration Learning

Semantic feature embedding. To embed semantic information at the feature level, at each resolution level l of the decoder, the semantic similarity maps ($\mathcal{M}_m$ and $\mathcal{M}_f$) of the input moving and fixed images (V_m and V_f) are first spatially resampled (e.g., via trilinear interpolation) to align with the resolution of the current decoding level l. These aligned similarity maps, denoted as

$\{\mathcal{M}_m^l\}$ and $\{\mathcal{M}_f^l\}$, are then encoded into high-dimensional semantic embedding features through a level-specific convolutional layer. These semantic features are concatenated with the encoded features $\{E_m^l\}$ and $\{E_f^l\}$ at the corresponding levels in the registration network and fused through a convolutional layer. Finally, the fused features $\{F_m^l\}$ and $\{F_f^l\}$ replace the original encoded features and are fed into the decoder, thereby embedding semantic knowledge into the feature representation.

Semantics-aware loss optimization. During training, existing unsupervised registration learning typically relies on global image similarity measure, which treats all voxels equally. To achieve semantics-aware learning, we replace the uniform similarity loss optimization with a semantics-guided spatial weighting scheme. Specifically, the semantic similarity map $\mathcal{M}_f$ is first spatially upsampled to match the original image resolution, denoted as $\hat{\mathcal{M}}_f$. It is then utilized as a spatial weighting map to dynamically re-weight the similarity loss, emphasizing registration quality within semantically-relevant regions. The total loss function is formulated as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sim}}(V_m \circ \phi, V_f, \hat{\mathcal{M}}_f) + \lambda \mathcal{L}_{\text{smooth}}(\phi), \tag{2}$$

where the smoothness loss $\mathcal{L}_{\text{smooth}}$ is implemented using diffusion regularization [1], $\circ$ denotes the spatial warping operation that resamples the moving image V_m using the deformation field ϕ, and λ is a loss weighting hyperparameter.

The similarity loss $\mathcal{L}_{\text{sim}}$ is derived using the Modality Independent Neighborhood Descriptor with Self-Similarity Context (MIND-SSC) [9, 8]. We formulate the semantics-aware similarity loss as:

$$\mathcal{L}_{\text{sim}} = \sum_{x \in \Omega} \hat{\mathcal{M}}_f(x) \cdot \|\mathcal{D}(V_m \circ \phi(x)) - \mathcal{D}(V_f(x))\|_2^2, \tag{3}$$

where $\mathcal{D}(\cdot)$ represents the dense feature descriptors extracted via the MIND-SSC.

3 Experiments

3.1 Datasets

We evaluate our method on two public datasets from the Learn2Reg challenge [14], including the abdominal CT dataset and the hippocampus MRI dataset.

Abdominal dataset. This dataset contains 54 CT scans with segmentation annotations for four major abdominal organs: liver, spleen, right kidney, and left kidney. All volumes are normalized to $192 \times 160 \times 192$ with $2 \times 2 \times 2$ mm^3 voxel spacing. We randomly split 38 CT scans for training and the remaining 16 CT scans for testing, respectively generating 1,406 training pairs and 240 test pairs.

Hippocampus dataset. This dataset comprises 260 T1-weighted MR images with manual segmentation labels for anterior and posterior hippocampus. All volumes are normalized to $64 \times 64 \times 64$ with $0.5 \times 0.5 \times 0.5$ mm^3 voxel spacing. We use 208 subjects for training (43,056 training pairs) and pair each of the remaining 52 test subjects with four others, yielding 208 test pairs.

3.2 Preprocessing and Evaluation Metrics

In the semantic similarity map generation stage, we preprocessed all slices from both datasets following the official implementation of DINOv3 prior to feeding them into the DINOv3 model. During the semantics-aware registration learning stage, for the abdominal dataset, the voxel intensities of CT volumes were clipped to the Hounsfield Unit (HU) range of $[-1000, 1000]$ and subsequently normalized to $[0, 1]$ using min-max scaling. For the hippocampus dataset, the intensities of MRI volumes were clipped to the $[0.5, 99.5]$ percentile range and normalized to the $[0, 1]$ interval. We employed three metrics to quantify registration accuracy and diffeomorphic properties: the Dice Similarity Coefficient (DSC), the 95% Hausdorff Distance (HD95), and the percentage of voxels with negative Jacobian determinants ($|J_\phi| \leq 0$).

3.3 Implementation Details

We integrate our method into two representative backbone registration architectures, i.e., RDP [19] and CGNet [3], using three frozen DINOv3 variants (ViT-S, ViT-B, ViT-L) for semantic similarity map generation, resulting in six configurations: OPS-RDP-{S, B, L} and OPS-CGNet-{S, B, L}. We compare against state-of-the-art methods, including VoxelMorph [2], TransMorph [4], SACB-Net [6], GroupMorph [17], DINO-Reg [16], RDP [19] and CGNet [3], which employ the unweighted version of Eq. (3) as the similarity loss. For semantic similarity map generation, we select a single reference slice containing all structures of interest (SOI) from each dataset and provide one point prompt per organ (liver, spleen, left kidney, and right kidney for abdominal data; anterior hippocampus and posterior hippocampus for hippocampus data). Dense similarity maps are pre-computed offline for all SOIs, introducing only minimal registration training overhead. During training, the networks are trained with Adam optimizer for 50,000 iterations on both datasets, using a learning rate of 1e-4 and a batch size of 1. The weighting parameter λ in Eq. (2) was empirically set to 0.5. To facilitate organ-specific registration learning during the semantics-aware loss optimization, we feed the pre-computed similarity map of a single point prompt into the network per iteration, cyclically alternating the similarity maps from different prompt points across successive iterations. During inference, we compute the pixel-wise maximum across similarity maps of different organ prompts to incorporate multi-organ semantic information as the input prior.

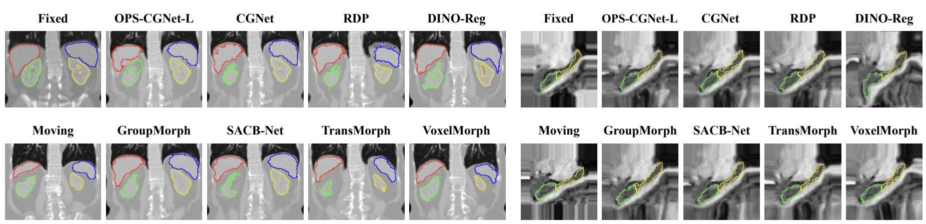
4 Results and Discussions

4.1 Comparison Results

Table 1 presents the quantitative comparison of our OPS framework against other registration methods. Integrating OPS with two backbone architectures (RDP and CGNet) consistently improves performance on both abdominal and hippocampus datasets. On the abdominal dataset, OPS-CGNet-L achieves the

Table 1. Quantitative comparison of registration performance. The best results are highlighted in **bold**.

Methods	Abdominal Dataset			Hippocampus Dataset		
	DSC (%) ↑	HD95 (mm) ↓	$\lvert J_\phi \rvert \leq 0$ ↓	DSC (%) ↑	HD95 (mm) ↓	$\lvert J_\phi \rvert \leq 0$ ↓
Initial	35.15 ± 12.43	22.58 ± 7.24	-	55.95 ± 13.48	3.82 ± 1.32	-
Baseline Methods						
VoxelMorph [2]	41.73 ± 14.48	21.69 ± 7.70	3.03e-3	67.38 ± 11.92	3.10 ± 1.26	**4.33e-4**
TransMorph [4]	46.09 ± 13.40	19.89 ± 6.01	**1.43e-3**	70.29 ± 10.67	2.86 ± 1.18	5.84e-4
SACB-Net [6]	53.60 ± 15.63	19.59 ± 7.99	1.89e-3	71.77 ± 9.40	2.89 ± 1.20	5.15e-4
GroupMorph [17]	55.07 ± 16.60	19.92 ± 8.04	2.16e-2	72.20 ± 9.53	2.90 ± 1.24	4.74e-3
DINO-Reg [16]	64.05 ± 14.57	16.71 ± 1.26	2.89e-2	67.79 ± 10.05	4.02 ± 2.12	1.60e-2
RDP [19]	50.32 ± 16.45	20.01 ± 7.46	1.98e-3	73.52 ± 9.82	2.70 ± 1.20	5.03e-4
CGNet [3]	58.52 ± 16.82	18.92 ± 8.36	9.29e-3	74.52 ± 9.27	2.63 ± 1.24	1.38e-3
Ours (RDP architecture)						
OPS-RDP-S	53.16 ± 16.01	19.50 ± 7.90	2.71e-3	74.40 ± 10.20	2.61 ± 1.25	1.25e-3
OPS-RDP-B	52.89 ± 16.10	19.41 ± 7.65	2.04e-3	74.25 ± 9.95	2.60 ± 1.18	1.30e-3
OPS-RDP-L	52.29 ± 17.57	19.52 ± 8.02	3.26e-3	74.36 ± 10.15	2.62 ± 1.25	1.37e-3
Ours (CGNet architecture)						
OPS-CGNet-S	64.12 ± 15.89	16.93 ± 7.11	1.20e-2	75.80 ± 7.97	2.53 ± 1.16	2.06e-3
OPS-CGNet-B	63.62 ± 15.83	16.67 ± 6.57	1.31e-2	75.84 ± 7.99	2.55 ± 1.20	2.46e-3
OPS-CGNet-L	**66.54 ± 13.96**	**15.31 ± 5.68**	1.15e-2	**76.07 ± 7.69**	**2.52 ± 1.17**	2.39e-3



(a) Results on the abdominal dataset **(b) Results on the hippocampus dataset**

Fig. 3. Qualitative comparison of registration performance on the abdominal dataset and hippocampus dataset, showing warped labels overlaid on the warped images.

highest mean DSC of 66.54% and the lowest HD95 of 15.31 mm, representing a substantial improvement of 8.02% in terms of DSC and 3.61 mm reduction in terms of HD95 over the baseline CGNet. For DINOv3 variants (S/B/L), the large-capacity ViT-L excels at abdominal tasks while the lightweight ViT-S suffices for hippocampus registration (e.g., OPS-CGNet-S achieves 75.80% mean DSC). Although semantics-aware registration learning increases folding voxels marginally, values remain within reasonable bounds, indicating OPS preserves topological plausibility while enhancing accuracy. In Fig. 3, we further visualize the qualitative comparison of different methods on representative samples from both datasets.

4.2 Ablation Study

We conducted ablation studies on both datasets to evaluate the contribution of each component in the proposed framework to the overall performance, where all

Table 2. Ablation study results on the effectiveness of key components. The best results are highlighted in **bold**.

Components		Abdominal Dataset			Hippocampus Dataset		
SFE	SLO	DSC (%) ↑	HD95 (mm) ↓	$\lvert J_\phi \rvert \leq 0$ ↓	DSC (%) ↑	HD95 (mm) ↓	$\lvert J_\phi \rvert \leq 0$ ↓
✗	✗	58.52 ± 16.82	18.92 ± 8.36	9.29e-3	74.52 ± 9.27	2.63 ± 1.24	1.38e-3
✓	✗	59.39 ± 17.07	18.02 ± 6.99	**9.17e-3**	74.70 ± 8.31	2.62 ± 1.18	**9.11e-4**
✗	✓	65.42 ± 14.96	16.19 ± 6.01	1.24e-2	75.95 ± 8.04	2.60 ± 1.24	2.66e-3
✓	✓	**66.54 ± 13.96**	**15.31 ± 5.68**	1.15e-2	**76.07 ± 7.69**	**2.52 ± 1.17**	2.39e-3

Table 3. Robustness analysis of our method to point prompt selection variations.

Prompt Selection	Abdominal Dataset			Hippocampus Dataset		
	DSC (%) ↑	HD95 (mm) ↓	$\lvert J_\phi \rvert \leq 0$ ↓	DSC (%) ↑	HD95 (mm) ↓	$\lvert J_\phi \rvert \leq 0$ ↓
Trial 1 (Ours)	66.54 ± 13.96	15.31 ± 5.68	1.15e-2	76.07 ± 7.69	2.52 ± 1.17	2.39e-3
Trial 2	66.73 ± 13.74	15.22 ± 5.57	1.36e-2	76.04 ± 8.33	2.51 ± 1.20	2.29e-3
Trial 3	66.27 ± 14.05	15.49 ± 5.78	1.28e-2	76.21 ± 8.01	2.51 ± 1.17	2.17e-3
Trial 4	66.82 ± 14.12	15.28 ± 5.68	1.29e-2	75.95 ± 8.04	2.60 ± 1.24	2.66e-3

ablation experiments were based on the OPS-CGNet-L configuration. As shown in Table 2, integrating only SFE into CGNet (without loss re-weighting) improved average DSC from 58.52% to 59.39% on the abdominal dataset and from 74.52% to 74.70% on the hippocampus dataset, while achieving the lowest deformation field folding ratio. Incorporating solely SLO substantially outperformed the baseline, raising average DSC to 65.42% on the abdominal dataset. The full OPS framework (combining SFE and SLO) achieved the best performance, attaining respectively 66.54% and 76.07% average DSC when evaluated on the two datasets. This confirms the effectiveness of each individual component of the proposed method. We further conducted an ablation study to evaluate the robustness of the proposed method to input point prompt selection. Specifically, we performed three additional independent trials by randomly selecting different reference training sample slices and new point prompts. As shown in Table 3, the registration performance remains relatively stable across different point prompt selections, demonstrating the strong robustness of our method to prompt selection.

5 Conclusion

In this paper, we propose OPS, a novel framework that advances unsupervised medical image registration by introducing semantic awareness with minimal annotation cost. By leveraging the powerful semantic clustering capabilities of features encoded by the DINOv3, OPS leverages one-shot geometric prompts to acquire dense anatomical priors across the entire dataset, thereby guiding the registration learning process. Comprehensive experiments on abdominal and hippocampus datasets demonstrate that OPS consistently enhances the performance of state-of-the-art baselines, offering an efficient, nearly label-free solution for embedding semantic priors into the unsupervised registration paradigm.

Disclosure of Interests. The authors declare no conflict of interest.

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9252–9260 (2018)
2. 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)
3. Chang, Y., Li, Z., Xu, W.: Cgnet: A correlation-guided registration network for unsupervised deformable image registration. IEEE Transactions on Medical Imaging (2024)
4. 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)
5. Chen, J., He, Y., Frey, E.C., Li, Y., Du, Y.: Vit-v-net: Vision transformer for unsupervised volumetric medical image registration. arXiv preprint arXiv:2104.06468 (2021)
6. Cheng, X., Zhang, T., Lu, W., Meng, Q., Frangi, A.F., Duan, J.: Sacb-net: spatial-awareness convolutions for medical image registration. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5227–5237 (2025)
7. Heinrich, M.P., Hansen, L.: Voxelmorph++ going beyond the cranial vault with keypoint supervision and multi-channel instance optimisation. In: International workshop on biomedical image registration. pp. 85–95. Springer (2022)
8. Heinrich, M.P., et al.: Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. Medical image analysis **16**(7), 1423–1435 (2012)
9. Heinrich, M.P., Jenkinson, M., Papież, B.W., Brady, S.M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: International conference on medical image computing and computer-assisted intervention. pp. 187–194. Springer (2013)
10. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., et al.: Weakly-supervised convolutional neural networks for multimodal image registration. Medical image analysis **49**, 1–13 (2018)
11. Li, Y., Wu, Y., Lai, Y., Hu, M., Yang, X.: Meddinov3: How to adapt vision foundation models for medical image segmentation? arXiv preprint arXiv:2509.02379 (2025)
12. Liu, C., Chen, Y., Shi, H., Lu, J., Jian, B., Pan, J., Cai, L., Wang, J., Yu, J., Gao, Z., et al.: Does dinov3 set a new medical vision standard? benchmarking 2d and 3d classification, segmentation, and registration. arXiv preprint arXiv:2509.06467 (2025)
13. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature communications **15**(1), 654 (2024)
14. Shusharina, N., Heinrich, M.P., Huang, R.: 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. Springer Nature (2021)

15. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)
16. Song, X., Xu, X., Zhang, J., Reyes, D.M., Yan, P.: Dino-reg: Efficient multimodal image registration with distilled features. IEEE Transactions on Medical Imaging (2025)
17. Tan, Z., Zhang, L., Lv, Y., Ma, Y., Lu, H.: Groupmorph: medical image registration via grouping network with contextual fusion. IEEE Transactions on Medical Imaging (2024)
18. Tian, L., Greer, H., Kwitt, R., Vialard, F.X., San José Estépar, R., Bouix, S., Rushmore, R., Niethammer, M.: unigradicon: A foundation model for medical image registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 749–760. Springer (2024)
19. Wang, H., Ni, D., Wang, Y.: Recursive deformable pyramid network for unsupervised medical image registration. IEEE Transactions on Medical Imaging **43**(6), 2229–2240 (2024)
20. Xie, H., Gao, X., Zheng, G.: Promptreg: Universal medical image registration via task prompt learning and domain knowledge transfer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 498–507. Springer (2025)
21. Xie, H., Gao, X., Zheng, G.: Biasnet: A bidirectional feature alignment and semantics-guided network for weakly-supervised medical image registration. Medical Image Analysis **109**, 103913 (2026)
22. Xu, H., Xue, T., Fan, J., Liu, D., Chen, Y., Zhang, F., Westin, C.F., Kikinis, R., O'Donnell, L.J., Cai, W.: Medical image registration meets vision foundation model: Prototype learning and contour awareness. In: International Conference on Information Processing in Medical Imaging. pp. 79–93. Springer (2025)
23. Zhao, Z., Zhang, Y., Wu, C., Zhang, X., Zhou, X., Zhang, Y., Wang, Y., Xie, W.: Large-vocabulary segmentation for medical images with text prompts. NPJ Digital Medicine **8**(1), 566 (2025)