VoxelOpt: Voxel-Adaptive Message Passing for Discrete Optimization in Deformable Abdominal CT Registration

Hang Zhang[1], Yuxi Zhang[2], Jiazheng Wang[2], Xiang Chen[2], Renjiu Hu[1], Xin Tian[3], Gaolei Li[4], and Min Liu[2] (✉)

[1] Cornell University, Ithaca, USA
[2] Hunan University, Changsha, China
liu_min@hnu.edu.cn
[3] University of Oxford, Oxford, UK
[4] Shanghai Jiao Tong University, Shanghai, China

Abstract. Recent developments in neural networks have improved deformable image registration (DIR) by amortizing iterative optimization, enabling fast and accurate DIR results. However, learning-based methods often face challenges with limited training data, large deformations, and tend to underperform compared to iterative approaches when label supervision is unavailable. While iterative methods can achieve higher accuracy in such scenarios, they are considerably slower than learning-based methods. To address these limitations, we propose **VoxelOpt**, a discrete optimization-based DIR framework that combines the strengths of learning-based and iterative methods to achieve a better balance between registration accuracy and runtime. VoxelOpt uses displacement entropy from local cost volumes to measure displacement signal strength at each voxel, which differs from earlier approaches in three key aspects. First, it introduces voxel-wise adaptive message passing, where voxels with lower entropy receives less influence from their neighbors. Second, it employs a multi-level image pyramid with 27-neighbor cost volumes at each level, avoiding exponential complexity growth. Third, it replaces hand-crafted features or contrastive learning with a pretrained foundational segmentation model for feature extraction. In abdominal CT registration, these changes allow VoxelOpt to outperform leading iterative in both efficiency and accuracy, while matching state-of-the-art learning-based methods trained with label supervision. The source code will be available at https://github.com/tinymilky/VoxelOpt.

Keywords: Deformable image registration · Discrete optimization · Mean-field inference · Message passing · Abdominal

1 Introduction

Traditional deformable image registration (DIR) typically minimizes a Horn-Schunck-type energy function [14], combining a dissimilarity metric and a smoothness regularizer. While effective, these methods [4, 12] rely on computationally expensive iterative optimization, resulting in slow runtime for large volumetric data. In contrast, learning-based methods, pioneered by VoxelMorph [5], amortize iterative optimization by training neural networks to predict deformations in a single forward pass. This enables fast image registration and can potentially achieve higher accuracy when trained with label supervision. However, learning-based approaches struggle with limited training data, large deformations, and may underperform iterative methods in the absence of supervision.

To address these limitations, recent methods [21,24,27] have combined learning-based efficiency with iterative optimization. For example, ConvexAdam [21] performs iterative optimization within deep learning frameworks (e.g., PyTorch), leveraging parallel GPU processing to achieve considerable speed improvements over classical approaches. Additionally, lightweight neural networks can be integrated into the optimization loop [24, 27] to enhance accuracy. However, two critical challenges remain: 1) These hybrid methods **remain slower than pure learning-based approaches** despite their acceleration over classical methods; 2) They rely solely on photometric dissimilarity measures, which are empirically shown to provide **limited anatomical correspondence information** [16].

We hypothesize that message passing mechanisms [18] have been largely overlooked in existing registration methods. As shown in the original Horn-Schunck variational formulation [14], displacements in uniform or smooth regions with weak intensity gradients must be inferred from boundaries or gradient-rich regions. This ”filling” process, achieved through an isotropic diffusive regularizer, is essentially a form of message passing. To our knowledge, all leading registration methods, both learning-based and iterative, rely on this isotropic regularizer for message passing. However, this serves as one of the major factors limiting the efficiency of iterative methods. Propagating displacement signals requires numerous iterations, resulting in slow convergence, while stronger regularization, though accelerating convergence, inevitably oversmooths the displacement field.

To address these challenges, we propose **VoxelOpt**, a voxel-adaptive message passing framework for discrete optimization in deformable image registration. **First**, VoxelOpt quantifies displacement signal strength using the entropy of a probabilistic 27-neighbor cost volume to guide adaptive message passing: low-entropy (strong signal) voxels retain their values, while high-entropy (weak signal) voxels are more influenced by neighbors, resembling spatially adaptive filtering [33]. **Second**, while discrete optimization-based DIR methods [10, 12, 21, 23] converge faster than gradient descent, they suffer from exponential growth in displacement search space. VoxelOpt addresses this by using a multi-level image pyramid, where each level's cost volume is limited to a 27-neighborhood, avoiding exponential growth while remaining compatible with large-deformation diffeomorphic frameworks [6]. **Third**, inspired by recent work [8, 30, 32] showing that segmentation-derived features enhance registration, we use a pretrained CT segmentation foundation model for feature extraction. This leverages richer semantic context than raw images [16] without requiring contrastive learning.

Preliminary results on an abdominal CT dataset with limited training data and large deformations show that VoxelOpt is both effective and efficient. It outperforms the best unsupervised learning method by 14.7% Dice and the best iterative method by 9.2%, while significantly reducing runtime. VoxelOpt also matches the best semi-supervised method without using label supervision.

2 Methodology

2.1 Preliminaries

Deformable image registration (DIR) [9,22,25] is typically formulated as a variational optimization problem:

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\{ d_s(\boldsymbol{f}, \boldsymbol{m} \circ (\mathbf{u} + \mathbf{I}_d)) + \lambda r(\mathbf{u}) \right\}. \tag{1}$$

Here, $\boldsymbol{f}$, $\boldsymbol{m}$, $\mathbf{I}_d$, and $\mathbf{u}$ denote the fixed image, moving image, identity transformation grid, and displacement field, respectively. The terms $d_s(\cdot,\cdot)$ and $r(\cdot)$ represent the dissimilarity function and smoothness regularizer, while λ controls the regularization strength. For discrete optimization in image registration, we follow prior work [10,13,31] and construct the energy function using local cost aggregation with quadratic relaxation [7,29], formulated as follows:

$$\begin{aligned} \hat{\mathbf{u}}, \hat{\mathbf{v}} = \arg\min_{\mathbf{u},\mathbf{v}} \ & \sum_{x}^{\boldsymbol{\Omega}} \mathbf{C}^k(x) \circ \mathbf{v}(x) + \frac{1}{2\theta}\|\mathbf{v} - \mathbf{u}\|^2 + \lambda r(\mathbf{u}), \\ \text{s.t.} \ & \forall\, x \in \boldsymbol{\Omega}, \mathbf{v}(x) \in \mathcal{L}_k. \end{aligned} \tag{2}$$

Here, $\boldsymbol{\Omega} \subseteq \mathbb{Z}^3$ is the spatial space of the image, $\frac{1}{2\theta}$ is the interaction term coefficient, $n = h \times w \times d$ is the image spatial size, $\mathcal{L}_k = \{0, \pm1, \pm2, \ldots, \pm k\}^3$. The intensity at position x is $\boldsymbol{f}(x)$ and $\boldsymbol{m}(x)$, with displacement $\mathbf{v}(x) \in \mathbb{R}^3$ and $\mathbf{u}(x) \in \mathbb{R}^3$. The cost volume $\mathbf{C}^k$ is a 6D tensor: the first three dimensions represent spatial coordinates, and the last three define the local cost volume of kernel size k at each voxel. We denote the local 3D cost volume at position x as $\mathbf{C}^k(x) \in \mathbb{R}^{(2k+1)\times(2k+1)\times(2k+1)}$, pre-computed by measuring dissimilarities between each voxel and its neighbors:

$$\mathbf{C}^k(x, o) = d_s\big(\boldsymbol{f}(x), \boldsymbol{m}(x + o)\big). \tag{3}$$

Here, $o \in \mathcal{L}_k$ enumerates the offsets in a local neighborhood of kernel size k.

2.2 Voxel-wise Displacement Entropy

To enable voxel-adaptive discrete optimization, we quantify the signal strength at each voxel location. This strength reflects the confidence of the derived displacement vector and determines the extent of information exchange with neighboring voxels. Specifically, voxels with stronger signals should prevent their information from being diluted, while those with weak signals should put higher weights on receiving information from neighboring regions.

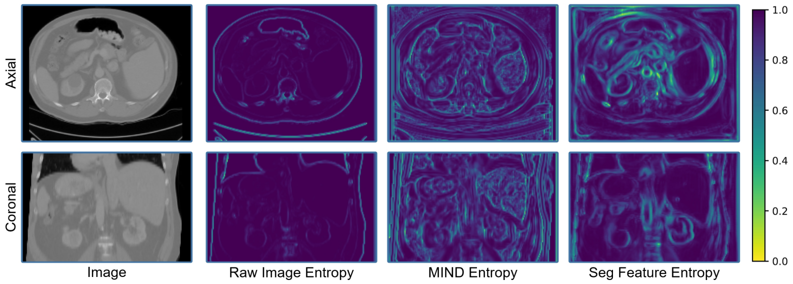

Fig. 1: Visual examples of how features affect entropy distribution. The fixed image is shown in the first column; the moving image is manually shifted by one voxel. Cost volumes are computed using a kernel size of 1, and per-voxel entropies are derived via Eq. (4), then normalized via Eq. (6).

Under sufficiently downsampled input images, we can assume the largest displacement magnitude is within one voxel. Thus, we compute the displacement entropy $\mathbf{E} \in \mathbb{R}^{h \times w \times d}$ using the 6D probabilistic cost volume tensor $\mathbf{P}^1$ as follows:

$$\mathbf{P}^1(x, o) = \frac{\exp(-\mathbf{C}^1(x, o)/\beta)}{\sum_p^{\mathcal{L}_1} \exp(-\mathbf{C}^1(x, p)/\beta)}, \quad \mathbf{E}(x) = \sum_o^{\mathcal{L}_1} -\mathbf{P}^1(x, o) \log(\mathbf{P}^1(x, o)), \tag{4}$$

where $\exp(\cdot)$ is the exponential operator, $\log(\cdot)$ is the logarithm operator, and β controls the temperature. Here $\mathbf{E}(x)$ is the per-voxel displacement entropy.

Displacement Entropy Interpretation: The per-voxel entropy $\mathbf{E}(x)$ quantifies the uncertainty of the displacement derived from the cost volume (by taking the argmin). For voxels in uniform or smooth regions, entropy tends to be high due to similar neighboring intensities, resulting in a uniform probability distribution. In contrast, voxels near boundaries or in texture-rich regions exhibit sparse probability distributions, leading to lower entropy. Thus, smaller $\mathbf{E}(x)$ indicates stronger displacement signals with less uncertainty, and vice versa.

2.3 Feature Impact on Entropy Distribution

Lower entropy values across more voxels simplify the overall displacement field extraction, as displacement signals propagate faster and more reliably. Thus, image features should promote a wider distribution of voxels with lower entropies. However, while certain features yield better entropy distributions, low entropy can also arise from local noise. To evaluate feature impact on entropy, we study raw image intensity, MIND feature maps [11], and feature maps from a pretrained foundational segmentation model [19] on abdominal CT scans.

As shown in Fig. 1, both MIND and segmentation (seg) feature maps produce more widely distributed low-entropy voxels compared to raw images. While raw images exhibit high entropy even at boundaries of large objects like the liver, seg features accurately capture boundary signals without introducing local noise in uniform regions. In contrast, MIND's self-similarity introduces considerable amount of local noise in uniform areas such as the liver.

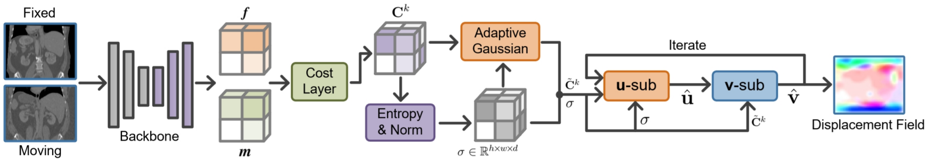

Fig. 2: Schematic of the VoxelOpt framework.

2.4 Voxel-wise Adaptive Message Passing

To solve Eq. (2), we employ coordinate descent, alternating between $\mathbf{v}$ and $\mathbf{u}$ while progressively reducing θ to finally achieve $\hat{\mathbf{v}} \approx \hat{\mathbf{u}}$. Then the optimization of Eq. (2) can be decomposed into two subproblems as follows:

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}} \frac{1}{2\theta}\|\mathbf{v}-\hat{\mathbf{u}}\|^2 + \sum_{x}^{\mathbf{\Omega}} \mathbf{C}^k(x) \circ \mathbf{v}(x), \text{ s.t. } \forall\, x \in \mathbf{\Omega}, \mathbf{v}(x) \in \mathcal{L}_k;$$
$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \frac{1}{2\theta}\|\hat{\mathbf{v}}-\mathbf{u}\|^2 + \lambda r(\mathbf{u}). \tag{5}$$

Here, each subproblem is convex and can be solved globally [23]. The $\mathbf{v}$ subproblem is solved optimally and pointwise across $x \in \mathbf{\Omega}$, as each voxel's displacement is independent of others. Notably, since $\mathbf{v}(x)$ is restricted to a discrete space, we approximate the optimizer by propagating the interaction term's effects to the cost volume and performing an argmin search. When $r(\mathbf{u}) = \|\nabla\mathbf{u}\|$, typically the $\mathbf{u}$ subproblem can be solved through the fixed-point iteration [7], yielding $\mathbf{u}$ as a smoothed version of $\hat{\mathbf{v}}$. This can be approximated by applying Gaussian filtering to $\hat{\mathbf{v}}$ as $\hat{\mathbf{u}} = \mathcal{K} * \hat{\mathbf{v}}$ across spatial locations.

Adaptive Message Passing: Solving the $\mathbf{u}$ subproblem resembles mean-field inference, where displacement signals act as messages passed between neighboring voxels. While widely used in prior approaches [13,21], isotropic Gaussian filtering can dilute strong signals with weak ones, leading to oversmoothing. This may require larger offset search spaces, increasing complexity exponentially.

To address this issue, we propose voxel-wise adaptive message passing. Using voxel-wise displacement entropy (see §2.2), we determine signal strength, which controls the extent of information each voxel receives from its neighbors: stronger signals receive less input, while weaker signals receive more. Specifically, we use normalized voxel entropies in $(0, 1)$ to compute a per-voxel $\sigma(x)$, adaptively controlling the information received during each message passing phase:

$$\sigma(x) = \alpha \log\left(\frac{\mathbf{E}(x)}{\max(\mathbf{E})} + 1\right), \tag{6}$$

where $\max(\cdot)$ computes the tensor's maximum value, $\log(\cdot)$ is the logarithm operator, and α controls the overall blurring strength.

2.5 Overall Framework of VoxelOpt

Unlike iterative methods requiring many iterations to converge, VoxelOpt uses discrete optimization [13, 21] and converges in just 6 iterations. Crucially, VoxelOpt operates like a pre-trained learning-based method without training, as it also computes the displacement field in a single forward pass.

As shown in Fig. 2, fixed and moving images are processed by a pretrained foundational network [19] to extract feature maps $\boldsymbol{f}$ and $\boldsymbol{m}$, which are fed into a cost layer (Eq. (3)) to compute the 6D cost volume $\mathbf{C}^k$. Using Eq. (4), $\mathbf{C}^k$ computes voxel-wise displacement entropy, normalized and scaled via Eq. (6) to obtain $\sigma \in \mathbb{R}^{h \times w \times d}$, guiding adaptive message passing. The voxel-adaptive σ filters $\mathbf{C}^k$ spatially to produce $\tilde{\mathbf{C}}^k$, which, along with σ, supports iterative optimization of the **v**- and **u**-subproblems, yielding the final displacement field.

With kernel size $k = 1$ (capturing displacements within one voxel using 27 neighbors), we employ an N-level Laplacian image pyramid for large-deformation diffeomorphic transformation. After extracting feature maps $\boldsymbol{f}$ and $\boldsymbol{m}$, we trilinearly downsample them 2× at each level to construct the N-level pyramid. At each level, the fixed feature map and the moving feature map (warped by the previous level's deformation field) undergo discrete optimization to compute the residual field. Each residual field is processed via scaling and squaring [3], and the final field is obtained by composing all levels' fields.

3 Experiments & Results

3.1 Datasets

We evaluated our method on a public dataset of 30 abdominal CT scans [28], each with segmentation masks for 13 organs. We generated 380 training pairs (20×19), 6 validation pairs (3×2), and 42 testing pairs (7×6) for comparison with iterative and learning-based methods. All images were resampled to 2 mm voxel spacing, resized to 192×160×256, clipped to $[-800, 500]$, and normalized to $[0, 1]$. In the unsupervised setting, no segmentation masks were used. In the semi-supervised setting, masks were used only during training to compute Dice loss for anatomical alignment. This dataset was previously collected and published in accordance with institutional ethical standards and the Declaration of Helsinki; no new human data collection was conducted in this study.

3.2 Baselines, Implementations & Evaluation Metrics

Baseline Methods: We compare VoxelOpt against three categories: iterative optimization, unsupervised learning-based, and semi-supervised learning-based methods. Iterative methods include Ants SyN [4], Deeds [12], ConvexAdam [21], NODEO [27], ccINR [24], and PRIVATE [15]. Learning-based methods include VoxelMorph [5], FourierNet [17], RDP [26], and CorrMLP [20].

Baseline Implementation Details: All learning methods were implemented using the CorrMLP framework and trained under identical conditions: 100 epochs on an A6000 GPU with PyTorch, Adam optimizer (learning rate 1×10^{-4}), batch size 1, regularization $\lambda = 1$, and NCC dissimilarity (window size 9). Scaling and squaring [2] with 7 integration steps ensured diffeomorphic transformations, as

Table 1: Quantitative results on the abdominal dataset, with metrics averaged over the 42 testing pairs. "Initial" refers to baseline results before registration, and a down arrow (↓) indicates that lower values are better. The best-performing metric in each column is bolded, and the second-best is underlined. Each method within the respective category is sorted by Dice score, from low to high.

Model	Type	Remarks	Dice (%)	HD95 ↓	SDLogJ ↓	Runtime (s) ↓
Initial	-	-	30.86	29.77	-	-
Ants SyN [4]	Iterative Methods	Affine+Deform	37.02	25.48	**0.07**	22.6
ccINR [24]		2500 iterations	39.07	24.09	0.21	77.1
PRIVATE [15]		500 iterations	45.49	24.11	0.13	7.1
ConvexAdam [21]		80 iterations	50.23	22.60	0.13	7.0
NODEO [27]		300 iterations	51.75	22.60	0.15	18.5
Deeds [12]		MIND+NCC	53.57	20.08	<u>0.12</u>	110.1
FourierNet [17]	Learning-based Methods Unsupervised	Start channels 32	41.83	25.25	**0.11**	< 1.0
VoxelMorph [5]		Start channels 32	41.90	25.97	0.12	< 1.0
RDP [26]		Channels 16	50.91	22.88	0.14	< 1.0
CorrMLP [20]		Enc 8, Dec 16	51.01	22.80	0.13	< 1.0
FourierNet [17]	Learning-based Methods Semi-supervised	Start channels 32	42.80	22.95	0.13	< 1.0
VoxelMorph [5]		Start channels 32	47.05	23.08	0.13	< 1.0
CorrMLP [20]		Enc 8, Dec 16	56.58	20.40	0.16	< 1.0
RDP [26]		Channels 16	**58.77**	<u>20.07</u>	0.22	< 1.0
VoxelOpt (ours)	Iterative Methods	6 iterations, $k = 1$	<u>58.51</u>	**18.54**	0.21	< 1.0

in [5]. Iterative methods were evaluated using their public code repositories, with configurations tuned for optimal performance (see remarks in results).

VoxelOpt Implementation Details: To handle large deformations, we employ a 5-level image pyramid with a local cost volume kernel size of $k = 1$ (unless specified otherwise). The blurring strength cap in Eq. (6) is set to $\alpha = 1.5$. Following prior work [13], we use 6 iterations with gradually decaying $\theta = \{150, 50, 15, 5, 1.5, 0.5\}$ in solving Eq. (5). VoxelOpt uses L_1 dissimilarity and scaling-and-squaring with 7 steps for diffeomorphic transformations. Adaptive 3D Gaussian filtering is performed via three separable 1D filters along x-, y-, and z-axes. The backbone uses the foundational segmentation model [19], with pre-softmax features as input. All experiments were run on the same machine and environment as the baselines.

Evaluation Metrics: Anatomical alignment is evaluated using Dice Similarity Coefficient (Dice) and 95% Hausdorff Distance (HD95), while smoothness is assessed via the standard deviation of the log Jacobian determinant (SDlogJ). Runtime, averaged over 42 testing pairs, is denoted as < 1 for sub-second durations, as sub-second differences for volumetric image registration are negligible.

3.3 Results & Analysis

Quantitative results of VoxelOpt compared to baseline methods are shown in Table 1. Deeds [12], using Markov random fields and tree-DP for discrete optimization, achieves the highest registration accuracy among iterative methods but requires the most runtime. It also outperforms all unsupervised learning-based methods in Dice (%). However, when trained with label supervision, multi-scale learning-based methods like CorrMLP and RDP surpass all iterative methods in registration accuracy. The proposed VoxelOpt achieves Dice (%) on par with

Table 2: Ablation study of VoxelOpt on feature type, kernel size k, adaptive message passing, and pre-filtering of the cost volume before optimization. Models are numbered for clarity. "Foundation" refers to features extracted using a foundational segmentation model.

Model	Feature Type	Kernel Size k	Adaptive	$\mathbf{C}^k$ Filtering	Dice (%)	HD95 ↓	SDLogJ ↓	Runtime (s) ↓
Initial	-	-	-	-	30.86	29.77	-	-
#1	Raw Image	1	Yes	Yes	45.67	26.14	0.18	< 1.0
#2	MIND	1	Yes	Yes	49.98	24.97	0.16	< 1.0
#3	Foundation	1	Yes	Yes	58.51	18.54	0.21	< 1.0
#4	Foundation	1	No	Yes	56.40	18.85	0.20	< 1.0
#5	Foundation	2	Yes	Yes	57.95	19.21	0.25	1.0
#6	Foundation	3	Yes	Yes	56.78	20.58	0.27	2.5
#7	Foundation	1	Yes	No	56.93	19.47	0.19	< 1.0
#8	Foundation	1	No	No	54.75	19.44	0.17	< 1.0

the state-of-the-art (SOTA) learning-based method RDP (semi-supervised), with similar runtime (< 1s), while reducing HD95 by 7.6%. Moreover, VoxelOpt outperforms the best unsupervised learning method CorrMLP by 14.7% in Dice (%) and the best iterative method Deeds by 9.2%, with a substantial runtime reduction. This demonstrates that, with readily available foundational models, VoxelOpt offers competitive performance without hand-crafted features, complex contrastive learning, or label supervision.

3.4 Ablation Studies

Effects of Image Features: Table 2 shows that performance improves progressively when using raw images (#1), MIND features [11] (#2), and pre-softmax features from a foundational segmentation model (#3). This aligns with Fig. 1, where raw images struggle to provide strong displacement signals, and MIND features introduce local noise in uniform regions.

Effects of The Kernel Size k: Comparing models #3, #5, and #6, increasing the kernel size does not improve registration accuracy but degrades deformation smoothness and exponentially increases computational complexity, as the cost volume space grows from $(2\times1+1)^3$ to $(2\times2+1)^3$ and $(3\times1+1)^3$, with runtime increasing from < 1.0s to 1.0s and 2.5s. This demonstrates the superiority of the multi-level image pyramid and 26-neighborhood approach.

Effects of Voxel-Adaptive Message Passing: Comparing models #3 and #4, voxel-adaptive message passing improves Dice by 3.7%. While the foundational segmentation model contributes significantly (comparing #1 and #3), the entropy measure of displacement signal strength is crucial for identifying pre-softmax segmentation features as naturally suitable for image registration.

Effects of # of Iterations: We vary the number of optimization iterations from 1 to 9 and observe that while Dice slightly increases with more iterations, the change remains $< 0.5\%$, likely because adaptive cost volume filtering already propagates key displacement signals. As shown in Table 2, comparing #3 vs. #7 and #4 vs. #8, removing pre-optimization cost volume filtering consistently degrades performance, with or without adaptive filtering.

4 Discussions & Conclusions

ConvexAdam (CA) [21] is perhaps the most comparable method to VoxelOpt, but we highlight four key differences that enable a better accuracy-efficiency trade-off and expansibility: 1) VoxelOpt uses a fixed 27-neighbor cost volume, avoiding CA's exponential complexity and enabling full-resolution processing. 2) It extracts displacement in one pass, unlike CA's two-stage pipeline that requires downsampling. With the same MIND feature, VoxelOpt achieves similar Dice but runs significantly faster (7s vs. $< 1s$, see #2 in Table 2 vs. "CA" in Table 1). 3) VoxelOpt uses displacement entropy to assess feature effectiveness, facilitating the use of various pre-trained foundational models beyond the one used here. 4) While unrolling is common in reconstruction [1,34], it is costly [10] or architecture-sensitive [31] in registration. In contrast, VoxelOpt integrates seamlessly for end-to-end training. In conclusion, **VoxelOpt** bridges the gap between learning-based and iterative methods for deformable image registration via a discrete optimization framework. It achieves SOTA accuracy with sub-second runtime, outperforming iterative methods in efficiency and matching semi-supervised approaches (with label supervision) in accuracy.

Disclosure of Interests: The authors have no competing interests to declare that are relevant to the content of this article. **Acknowledgments:** This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFE0134700, and in part by the National Natural Science Foundation of China under Grant 62425305 and U22B2050.

References

1. Aggarwal, H.K., Mani, M.P., Jacob, M.: Modl: Model-based deep learning architecture for inverse problems. IEEE transactions on medical imaging **38**(2), 394–405 (2018)
2. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 924–931. Springer (2006)
3. Ashburner, J.: A fast diffeomorphic image registration algorithm. Neuroimage **38**(1), 95–113 (2007)
4. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis **12**(1), 26–41 (2008)
5. 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)
6. 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)
7. Chambolle, A.: An algorithm for total variation minimization and applications. Journal of Mathematical imaging and vision **20**, 89–97 (2004)
8. Chen, X., Liu, M., Wang, R., Hu, R., Liu, D., Li, G., Zhang, H.: Spatially covariant image registration with text prompts. IEEE Transactions on Neural Networks and Learning Systems pp. 1–11 (2024)

9. Haskins, G., Kruger, U., Yan, P.: Deep learning in medical image registration: a survey. Machine Vision and Applications **31**, 1–18 (2020)
10. Heinrich, M.P.: Closing the gap between deep and conventional image registration using probabilistic dense displacement networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 50–58. Springer (2019)
11. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, M., Schnabel, J.A.: Mind: Modality independent neighbourhood descriptor for multimodal deformable registration. Medical image analysis **16**(7), 1423–1435 (2012)
12. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: Mrf-based deformable registration and ventilation estimation of lung ct. IEEE transactions on medical imaging **32**(7), 1239–1248 (2013)
13. Heinrich, M.P., Papież, B.W., Schnabel, J.A., Handels, H.: Non-parametric discrete registration with convex optimisation. In: Biomedical Image Registration: 6th International Workshop, WBIR 2014, London, UK, July 7-8, 2014. Proceedings 6. pp. 51–61. Springer (2014)
14. Horn, B.K., Schunck, B.G.: Determining optical flow. Artificial intelligence **17**(1-3), 185–203 (1981)
15. Hu, J., Gan, W., Sun, Z., An, H., Kamilov, U.: A plug-and-play image registration network. In: International Conference on Learning Representations (2024)
16. Jena, R., Sethi, D., Chaudhari, P., Gee, J.: Deep learning in medical image registration: Magic or mirage? In: The Thirty-eighth Annual Conference on Neural Information Processing Systems (2024)
17. Jia, X., Bartlett, J., Chen, W., Song, S., Zhang, T., Cheng, X., Lu, W., Qiu, Z., Duan, J.: Fourier-net: Fast image registration with band-limited deformation. In: Proceedings of the AAAI Conference on Artificial Intelligence (2023)
18. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. Advances in neural information processing systems **24** (2011)
19. Liu, J., Zhang, Y., Chen, J.N., Xiao, J., Lu, Y., A Landman, B., Yuan, Y., Yuille, A., Tang, Y., Zhou, Z.: Clip-driven universal model for organ segmentation and tumor detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21152–21164 (2023)
20. Meng, M., Feng, D., Bi, L., Kim, J.: Correlation-aware coarse-to-fine mlps for deformable medical image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9645–9654 (2024)
21. Siebert, H., Großbröhmer, C., Hansen, L., Heinrich, M.P.: Convexadam: Self-configuring dual-optimisation-based 3d multitask medical image registration. IEEE Transactions on Medical Imaging (2024)
22. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. IEEE transactions on medical imaging **32**(7), 1153–1190 (2013)
23. Steinbrücker, F., Pock, T., Cremers, D.: Large displacement optical flow computation without warping. In: 2009 IEEE 12th International Conference on Computer Vision. pp. 1609–1614. IEEE (2009)
24. Van Harten, L.D., Stoker, J., Išgum, I.: Robust deformable image registration using cycle-consistent implicit representations. IEEE Transactions on Medical Imaging (2023)
25. Viergever, M.A., Maintz, J.A., Klein, S., Murphy, K., Staring, M., Pluim, J.P.: A survey of medical image registration–under review (2016)
26. Wang, H., Ni, D., Wang, Y.: Recursive deformable pyramid network for unsupervised medical image registration. IEEE Transactions on Medical Imaging (2024)

27. Wu, Y., Jiahao, T.Z., Wang, J., Yushkevich, P.A., Hsieh, M.A., Gee, J.C.: Nodeo: A neural ordinary differential equation based optimization framework for deformable image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20804–20813 (2022)
28. Xu, Z., Lee, C.P., Heinrich, M.P., Modat, M., Rueckert, D., Ourselin, S., Abramson, R.G., Landman, B.A.: Evaluation of six registration methods for the human abdomen on clinically acquired ct. IEEE Transactions on Biomedical Engineering **63**(8), 1563–1572 (2016)
29. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime tv-l 1 optical flow. In: Pattern Recognition: 29th DAGM Symposium, Heidelberg, Germany, September 12-14, 2007. Proceedings 29. pp. 214–223. Springer (2007)
30. Zhang, H., Chen, X., Hu, R., Liu, D., Li, G., Wang, R.: Memwarp: Discontinuity-preserving cardiac registration with memorized anatomical filters. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 671–681. Springer (2024)
31. Zhang, H., Chen, X., Hu, R., Wang, R., Zhang, J., Liu, M., Wang, Y., Li, G., Cheng, X., Duan, J.: Unsupervised deformable image registration with structural nonparametric smoothing. arXiv preprint arXiv:2506.10813 (2025)
32. Zhang, H., Chen, X., Wang, R., Hu, R., Liu, D., Li, G.: Slicer networks. arXiv preprint arXiv:2401.09833 (2024)
33. Zhang, H., Wang, R., Zhang, J., Liu, D., Li, C., Li, J.: Spatially covariant lesion segmentation. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. pp. 1713–1721 (2023)
34. Zhang, J., Spincemaille, P., Zhang, H., Nguyen, T.D., Li, C., Li, J., Kovanlikaya, I., Sabuncu, M.R., Wang, Y.: Laro: Learned acquisition and reconstruction optimization to accelerate quantitative susceptibility mapping. NeuroImage **268**, 119886 (2023)