Gaussian Primitive Optimized Deformable Retinal Image Registration

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

[1] University of Oxford, Oxford, UK
[2] Cornell University, Ithaca, USA
hz459@cornell.edu
[3] Hunan University, Changsha, China
[4] Shanghai Jiao Tong University, Shanghai, China

Abstract. Deformable retinal image registration is notoriously difficult due to large homogeneous regions and sparse but critical vascular features, which cause limited gradient signals in standard learning-based frameworks. In this paper, we introduce Gaussian Primitive Optimization (GPO), a novel iterative framework that performs structured message passing to overcome these challenges. After an initial coarse alignment, we extract keypoints at salient anatomical structures (e.g., major vessels) to serve as a minimal set of descriptor-based control nodes (DCN). Each node is modelled as a Gaussian primitive with trainable position, displacement, and radius, thus adapting its spatial influence to local deformation scales. A K-Nearest Neighbors (KNN) Gaussian interpolation then blends and propagates displacement signals from these information-rich nodes to construct a globally coherent displacement field; focusing interpolation on the top (K) neighbors reduces computational overhead while preserving local detail. By strategically anchoring nodes in high-gradient regions, GPO ensures robust gradient flow, mitigating vanishing gradient signal in textureless areas. The framework is optimized end-to-end via a multi-term loss that enforces both keypoint consistency and intensity alignment. Experiments on the FIRE dataset show that GPO reduces the target registration error from 6.2 px to 2.4 px and increases the AUC at 25 px from 0.770 to 0.938, substantially outperforming existing methods. The source code can be accessed via https://github.com/xintian-99/GPOreg.

Keywords: Retinal Vessel Alignment · Deformable Image Registration · Gaussian Primitive Parametrization · Sparse Feature Propagation

1 Introduction

Retinal image registration is central to many clinical and research applications, including longitudinal monitoring of diseases such as diabetic retinopathy or age-related macular degeneration, as well as multi-modal fusion for enhanced diagnostic accuracy [18,24]. However, deformable retinal registration remains

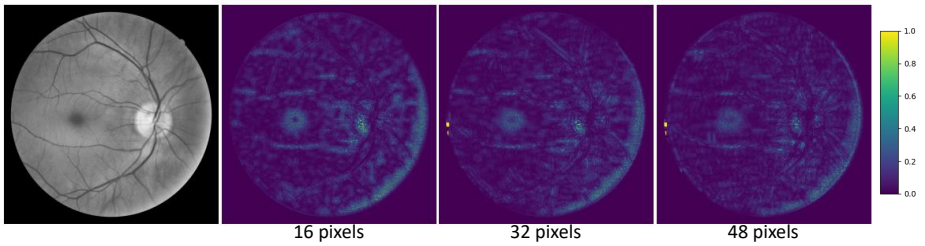

Fig. 1. Visualization of gradient backflow in a retinal image under normalized cross-correlation (NCC). The original image (left) is preprocessed and normalized to [0,1]. Heatmaps (right) show the absolute NCC gradients for x-axis shifts of 16, 32, and 48 pixels. High responses indicate effective gradient propagation; low responses correspond to homogeneous or vessel-sparse regions.

difficult due to the dominance of large homogeneous (textureless) regions and the sparse distribution of salient vasculature (less than 15% of the area) [10], which provides limited gradient signals and hampers accurate correspondence. As shown in Fig. 1, the scarcity of vessel edges leads to weak gradients, while large flat regions offer little to no signal, presenting a challenge for both classical and learning-based registration methods.

Traditional registration methods iteratively optimize a similarity metric (e.g., normalized cross-correlation (NCC), or mutual information (MI)) using gradient-based approaches [1,12]. However, they are prone to local minima, and even advanced discrete optimization techniques [8,23,25] often fail when images contain the extensive homogeneous regions and sparse vascular structures characteristic of the retina.

Modern deep learning-based approaches have also struggled to overcome this issue, broadly falling into three paradigms. (i) *Regression-based* methods directly predict transformation parameters in a single forward pass (e.g., affine or flow predictors [2,3,16,29,31]) for a coarse global alignment, but fail to model fine, local deformations and are susceptible to vanishing gradients in textureless regions. (ii) *Descriptor-based* methods [4,5,14,15,19,21,26] detect and match salient keypoints (usually on vessel junctions or other distinctive structures) to guide the transformation, but typically compute a single global transformation (e.g., homography) lacking an explicit data-fidelity term to refine local misalignments. More advanced (iii) *learning and iterative optimization based* frameworks [6,7,17,20,22,32] integrate neural networks into multi-scale or iterative optimization pipelines. However, their reliance on dense image similarity or a simple smoothness regularizer lets loss function become dominated by easily aligned homogeneous regions, causing the crucial gradient signal from fine structures "diluted" or averaged out. Thus, thin vessels and other subtle anatomical features remain insufficiently registered due to restricted gradient backpropagation.

In summary, classical optimization methods can become trapped in local minima, while descriptor-based solutions often compute a single global transforma-

tion, lacking the flexibility for local refinement. Furthermore, modern learning-based approaches suffer from gradient signal dilution, as the ambiguous displacement estimation in vast, textureless regions causes their loss to overwhelm the critical alignment signals from the sparse vasculature. A robust solution must therefore facilitate message passing to propagate displacement information from high-confidence vascular structures to these ambiguous regions in the network design [27,28,30]. To address these challenges, we propose Gaussian Primitive Optimization (GPO), a deformable registration framework designed specifically for sparse-feature medical images. Our key contributions are:

- We address gradient signal dilution by anchoring a minimal set of descriptor-based control nodes at salient keypoints, such as major vessels, which preserves the crucial displacement information provided by vascular structures.
- Our KNN-based Gaussian interpolation serves as a structured message passing mechanism, propagating displacement signals from the control nodes into feature-sparse regions and blending local transformations into a globally coherent and locally precise alignment.
- On the FIRE dataset, GPO lowers the target registration error from 6.20 px to 2.35 px and increases the AUC@25 px from 0.770 to 0.938, demonstrating both higher accuracy and fewer outlier errors compared to previous methods.

2 Methodology

We propose a Gaussian Primitive Optimized (GPO) framework for retinal image registration, organized into four main stages. First, we perform coarse alignment using a descriptor-based network, which also yields matched keypoints to serve as control nodes. Next, each node is initialized as a Gaussian primitive with learnable position, displacement, and an adaptive radius. Then, we compute the deformation field via a structured message passing process, where a KNN Gaussian interpolation blends local transformations to accommodate complex retinal deformations. Finally, all parameters are iteratively optimized under a multi-term loss that enforces both intensity alignment and global consistency.

Formally, given a fixed image $I_f \colon \Omega \to \mathbb{R}$ and a moving image $I_m \colon \Omega \to \mathbb{R}$, we seek a transformation $\mathcal{T} \colon \mathbb{R}^2 \to \mathbb{R}^2$ satisfying

$$I_f(\mathbf{x}) \approx I_m\big(\mathbf{x} + \mathbf{u}(\mathbf{x})\big), \tag{1}$$

where $\mathbf{u}(\mathbf{x})$ is the displacement field our GPO optimizes.

2.1 Coarse Alignment & Control Node Initialization

We first obtain a coarse alignment of I_m to I_f by estimating a global transform:

$$I_m^{(\text{coarse})}(x) = I_m(Ax + b), \tag{2}$$

where A and b represent an affine, homography, or other global parameters learned by a descriptor-based network (e.g., GeoFormer [14]). Concurrently, the

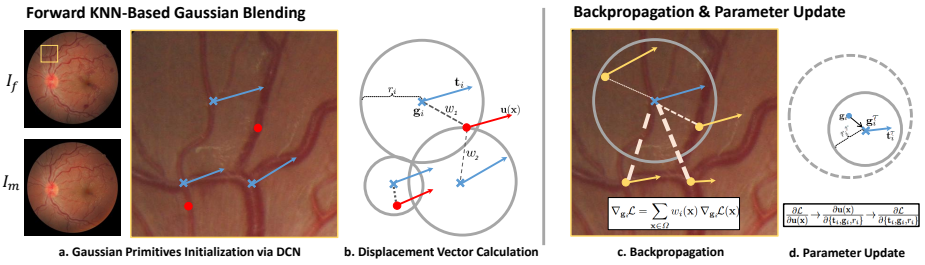

Fig. 2. Overview of GPO: control node initialization, KNN-based Gaussian blending, and iterative parameter updates.

network provides N matched keypoints $\{(g_i^f, g_i^m)\}_{i=1}^N$ in I_f and $I_m^{(\text{coarse})}$, forming *descriptor-based control nodes (DCN)* (Fig. 2), which act as the primary sources for propagating displacement signals. If descriptors are unavailable or sparse, we sample *grid-based control nodes (GCN)* on an $n \times n$ lattice; each lattice point $\{g_i\}_{i=1}^{n^2}$ in I_f is mapped to the same coordinate in $I_m^{(\text{coarse})}$.

2.2 KNN-Based Gaussian Blending for Deformation Estimation

Gaussian Primitive Initialisation After coarse alignment, each matched pair (g_i^f, g_i^m) is used to initialise a Gaussian primitive centered at $\mathbf{g}_i \equiv g_i^f$. Every node i possesses three sets of learnable parameters:

i. **Position** $\mathbf{g}_i \in \mathbb{R}^2$: Refined during training to allow local anchors to shift toward anatomically salient regions.

ii. **Displacement Vector** $\mathbf{t}_i \in \mathbb{R}^2$: Encodes the local translation of node $\mathbf{g}_i$. Initialized to $\mathbf{t}_i^{(0)} = g_i^m - g_i^f$ for DCN or $\mathbf{0}$ for GCN.

iii. **Radius** $r_i \in \mathbb{R}^+$: Adjusts each node's spatial influence and parametrize r_i by a learnable scalar β_i, via mapping $r_i = r_{\min} + (r_{\max} - r_{\min})\,\sigma(\beta_i) + 0.1$, where $\sigma(\cdot)$ is the sigmoid function. The 0.1 offset ensures r_i never becomes zero, preventing vanishing gradients.

KNN-Based Gaussian Blending To propagate displacement signals and construct a smoothly varying displacement field $\mathbf{u}(\mathbf{x})$ from the sparse set of control nodes (Fig 2 b), we employ a KNN-based Gaussian weighting scheme that functions as a message passing mechanism. Specifically, the displacement field is computed as a weighted sum over the K-nearest control nodes. We define the displacement at $\mathbf{x}$ as:

$$\mathbf{u}(\mathbf{x}) = \sum_{i=1}^{K} w_i(\mathbf{x})\mathbf{t}_i, \quad \text{where} \quad w_i(\mathbf{x}) = \frac{\exp\left(-\frac{\|\mathbf{x}-\mathbf{g}_i\|^2}{2r_i^2}\right)}{\sum_{j=1}^{K}\exp\left(-\frac{\|\mathbf{x}-\mathbf{g}_j\|^2}{2r_j^2}\right)}, \tag{3}$$

and $\mathbf{t}_i$ is the displacement vector associated with the i-th control node. The $w_i(\mathbf{x})$ is a Gaussian kernel weight that decays exponentially, thus giving greater

influence to control nodes closer to $\mathbf{x}$. The denominator ensures $\sum_{i=1}^{K} w_i(\mathbf{x}) = 1$. Consequently, each pixel is influenced by its K nearest nodes, and each node, in turn, receives gradient signals from those pixels during backpropagation (see Sec. 2.3). In descriptor-based control nodes (DCN), these nodes are placed on anatomically distinctive features (e.g., major vessels), preserving high-gradient signals to pixel and during optimization vice versa and mitigating vanishing gradients in homogeneous regions. Notably, by restricting on the top K nearest nodes instead of all nodes, we reduce computational overhead while retaining accurate local detail.

2.3 Neural Iterative Optimization

Forward Pass & Loss Function To find an optimal set of node parameters $\{\mathbf{g}_i, \mathbf{t}_i, r_i\}_{i=1}^{N}$, we adopt an iterative, gradient-based framework indexed by $\tau = 1, \ldots, \tau_{\max}$. At iteration τ, we compute the displacement field $\mathbf{u}_\tau(\mathbf{x})$ from Eq. (3) and warp the coarse-aligned moving image via bilinear interpolation for pixel resampling.

$$I_{w,\tau}(\mathbf{x}) \;=\; I_m^{(\text{coarse})}\big(\mathbf{x} + \mathbf{u}_\tau(\mathbf{x})\big), \tag{4}$$

We learn the parameters $\{\mathbf{g}_i, \mathbf{t}_i, r_i\}_{i=1}^{N}$ and optimise the displacement field by minimising a two-term loss function:

$$\mathcal{L}_\tau \;=\; \alpha_1\, \mathcal{L}_{\text{gcc}} \;+\; \alpha_2\, \mathcal{L}_{\text{ncc}}, \tag{5}$$

where $\mathcal{L}_{\text{gcc}}$ is global cross-correlation loss with matched control nodes, $\mathcal{L}_{\text{ncc}}$ aligns overall intensity patterns in I_f and $I_w^{(\tau)}$.

Backpropagation & Parameter Update As $\mathbf{u}(\mathbf{x})$ is a weighted sum over K nearest nodes (Eq. 3), each node $\mathbf{g}_i$ accumulates gradient signals from multiple pixels during backpropagation via $\frac{\partial \mathcal{L}}{\partial \mathbf{u}(\mathbf{x})} \rightarrow \frac{\partial \mathbf{u}(\mathbf{x})}{\partial \{\mathbf{t}_i, \mathbf{g}_i, r_i\}} \rightarrow \frac{\partial \mathcal{L}}{\partial \{\mathbf{t}_i, \mathbf{g}_i, r_i\}}$ to update $\{\mathbf{g}_i, \mathbf{t}_i, r_i\}$ (Fig 2 c & d). If $\nabla_{\mathbf{g}_i}\mathcal{L}(\mathbf{x})$ denotes the pixel-level gradient at $\mathbf{x}$ for node i, then its total gradient:

$$\nabla_{\mathbf{g}_i}\mathcal{L} = \sum_{\mathbf{x}\in\Omega} w_i(\mathbf{x})\, \nabla_{\mathbf{g}_i}\mathcal{L}(\mathbf{x}), \tag{6}$$

where $w_i(\mathbf{x})$ is the Gaussian blending weight. Thus, even if vessels occupy only a small portion of the image, a subset of pixels near each node typically lies on or around high-gradient vessel edges, ensuring that every node still receives nonzero gradient signal for optimization and enabling robust convergence despite sparse vascular structures.

The $\{\mathbf{g}_i, \mathbf{t}_i, r_i\}$ is then updated by subtracting their respective gradient terms scaled by distinct learning rates η_g, η_t, η_r at each iteration τ. After $\tau_{\max}$ iterations, we obtain the final displacement field $\mathbf{u}_{\text{final}}$ for the coarse-aligned image to produce the fully registered result $I_{w,\text{final}}(\mathbf{x}) = I_m^{(\text{coarse})}\big(\mathbf{x} + \mathbf{u}_{\text{final}}(\mathbf{x})\big)$.

3 Experiments, Results, and Discussion

3.1 Experimental Setup and Baselines

Dataset: We evaluated our approach on the FIRE dataset [9], which contains 134 retinal image pairs with 10 expert-annotated landmarks each for evaluation at a resolution of 2912×2912 pixels. The dataset was previously collected and published in accordance with institutional ethical standards and the Declaration of Helsinki; no new human data collection was conducted for this study. The dataset has three subgroups: 71 pairs with minimal distortion (Category S), 4 pairs with anatomical changes (A), and 49 pairs with perspective distortion (P). Each image pair includes 10 expert-annotated landmark points for evaluation. For all experiments, we resized images to 1024×1024 and applied Gaussian blur for anti-aliasing.

Baseline Methods: We benchmark GPO against two categories of methods, with full results in Table 1: (1) descriptor-based methods for global transformation, including SuperPoint [4], R2D2 [21], RoMa [5], GeoFormer [14], and others; and (2) learning-based deformable registration frameworks, such as GraDIRN [20], PDD-Net [6], VoxelMorph++ [7], and RetinaRegNet [22]. For trainable methods, we used a 7:1:2 stratified train-val-test split. In qualitative comparisons, we highlight GeoFormer against our GPO-GCN and GPO-DCN variants.

Evaluation Metrics: We used two common metrics to evaluate model performance: Target Registration Error (TRE) and Area Under the Curve (AUC) for TRE thresholds. TRE measures the L_2 distance between corresponding points in the fixed and warped moving images annotated by clinical experts, with lower values indicating better alignment. AUC quantifies the percentage of TRE values below error thresholds (15, 25, and 50 pixels), normalizing accumulated rates to provide a comprehensive measure of registration success.

Implementation Details: The experiments were conducted in PyTorch and optimized on an NVIDIA A100 GPU using the Adam optimizer. The node position updates used an initial learning rate of $\eta_g = 1.0$, while both radii and displacement vectors used $\eta_r = \eta_t = 0.01$, enabling rapid coarse adjustments while preserving local fidelity. We fixed the KNN interpolation parameter to $K = 10$ for all experiments. For GPO-DCN, we sampled $N = 1000$ keypoints and optimized for 100 iterations; for GPO-GCN, we placed a 20×20 grid of nodes and converged within 200 iterations. The loss weights in Eq. (5) were set to $\alpha_{\text{gcc}} = 0.4$ and $\alpha_{\text{ncc}} = 1.0$, providing balanced guidance from both keypoint consistency and intensity similarity.

3.2 Results and Analysis

Quantitative Analysis Table 1 shows that both GPO variants substantially outperform selected descriptor-based and learning-based methods on the FIRE dataset. Specifically, GPO-DCN achieves the lowest TRE (**2.352** px) and consistently ranks highest across all AUC thresholds(0.938@25 px and 0.964@50 px). Compared to the best-performing descriptor-based method (GeoFormer [14],

Table 1. Quantitative comparison on the FIRE dataset. Methods are categorized into Descriptor-based methods (left) and Learning & Iterative Optimization methods (right). Best results are in **bold**, second-best are underscored.

Descriptor-based Methods					*Learning & Iterative Optimization Methods*				
Method	**TRE ↓**	**AUC ↑**			**Method**	**TRE ↓**	**AUC ↑**		
		@15	@25	@50			@15	@25	@50
XFeat [19]	10.858	0.560	0.637	0.794	GraDIRN [20]	6.344	0.657	0.774	0.885
R2D2 [21]	7.926	0.553	0.701	0.850	PDD-Net [6]	5.765	0.688	0.792	0.893
LightGlue [13]	7.802	0.575	0.710	0.855	VoxelMorph++ [7]	5.400	0.710	0.808	0.902
SuperPoint [4]	6.641	0.612	0.757	0.879	VR-Net [11]	4.974	0.705	0.823	0.911
Glampoints [26]	6.608	0.595	0.757	0.879	RetinaRegNet [22]	2.766	0.852	0.910	0.955
SuperRetina [15]	6.382	0.622	0.767	0.884					
RoMa [5]	6.388	0.605	0.763	0.881	GPO-GCN (Ours)	2.649	0.888	0.914	0.926
GeoFormer [14]	6.201	0.625	0.770	0.887	GPO-DCN (Ours)	**2.352**	**0.906**	**0.938**	**0.964**

TRE 6.201 px) and the strongest learning-based method (RetinaRegNet [22], TRE 2.766 px), GPO-DCN reduces alignment error by 3.8 px and 0.4 px, respectively. These gains reflect the benefit of anatomically guided nodes and KNN-based blending in modelling local deformations beyond homography or uniform grids. Moreover, GPO-DCN attains higher AUC values at all thresholds (0.906@15 px vs. 0.625 for GeoFormer), indicating improved spatial precision and robustness to outlier errors through anatomically guided node placement.

By contrast, GPO-GCN achieves slightly higher TRE (2.649 px) and lower AUCs (0.914@25 px), but still exceeds other baselines, underscoring the benefit of the iterative Gaussian-primitive optimization. However, its uniform-grid sampling may miss vessel junctions or small-scale variations, resulting in lower AUC compared to GPO-DCN—especially at moderate thresholds (e.g., AUC@25 px = 0.914). Consequently, GPO-DCN's descriptor-based node acquisition preserves more fine-grained vascular structure and ensures more robust alignment.

Qualitative Analysis Fig 3 compares I_m, I_f, and registration outputs from GeoFormer, GPO-GCN, and GPO-DCN on two challenging retinal cases. In *top row*, we highlight a region at the optic disc where vessel geometry is highly tortuous and subject to complex localized deformations. GeoFormer and GPO-GCN struggle to preserve these fine details, leading to partial misalignment. In contrast, GPO-DCN leverages descriptor-based control nodes for the message passing framework to capture complex localized distortions more accurately, preserving vessel continuity. The *bottom row* highlights shadowing artifacts that obscure sections of the vasculature. Here, both GeoFormer and GPO-GCN exhibit residual misalignment in shadowed regions, whereas GPO-DCN demonstrates tighter correspondence of vascular edges. By iteratively refining Gaussian primitives near salient features, GPO-DCN effectively compensates for local intensity variations, resulting in sharper vessel alignment even under low contrast.

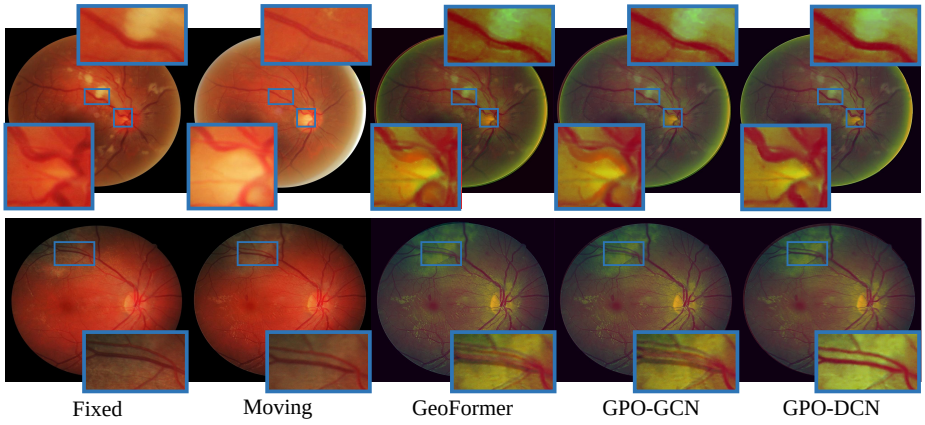

Fig. 3. Qualitative comparison on the FIRE dataset.

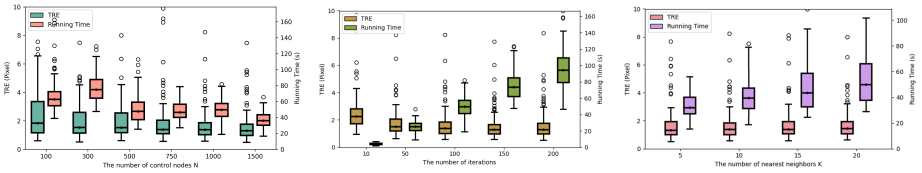

Fig. 4. Ablation study on key parameters on TRE and running time. **Left:** Influence of the number of control nodes N. **Middle:** Effect of the number of iterations. **Right:** Impact of the number of nearest neighbors K.

Ablation Studies We conducted a three-way ablation to examine how the number of control nodes N, the number of nearest neighbors K, and the number of optimization iterations τ affect both accuracy and runtime (Fig. 4).

Number of Control Nodes N: Increasing N from 300 to 1000 reduces the median TRE from ~ 2.60 px to ~ 2.35–2.40 px but raises runtime from $\sim$ 18s to $\sim$ 34s. Beyond 1000 nodes, further improvements (~ 2.22–2.35 px) come at the expense of a longer runtime ($\sim$ 45s), yielding diminishing returns.

Number of Nearest Neighbors K: For $K = 5$, the average TRE remains at ~ 2.60–2.65 px. Increasing to $K = 10$ lowers it to ~ 2.40–2.45 px, with a moderate $\sim$ 30s runtime. Although going to $K = 15$ or $K = 20$ can yield minor accuracy gains (~ 2.35–2.38 px), runtime increases by 20–30%. Thus, $K = 10$ strikes the best balance between precision and efficiency.

Number of Iterations τ: With $\tau = 50$, the median TRE hovers around 2.55–2.60 px in $\sim$ 15s. Doubling to $\tau = 100$ reduces TRE to ~ 2.40–2.45 px (a 0.1–0.2 px gain) at $\sim$ 30s. Beyond 100 iterations, further gains (≤ 0.05–0.07 px) come at a steep runtime cost ($\sim$ 45s or more).

Overall, $N = 1000$, $K = 10$, and $\tau = 100$ achieve a median TRE of ~ 2.3–2.4 px within $\sim$ 30s, offering a favorable trade-off between alignment accuracy and computational overhead.

4 Conclusion

We introduced GPO, an iterative deformable registration framework that addresses the critical challenge of gradient signal dilution in retinal images. By leveraging descriptor-based control nodes (or uniform-grid nodes) to blend and propagate their local transformations via KNN-weighted Gaussian primitives, GPO mitigates vanishing gradients in homogeneous regions while accurately modelling localized vessel deformations. On the FIRE dataset, GPO outperforms conventional homography-based and learning-based methods in both alignment accuracy and robustness. Future work will explore extending GPO to multi-modal retinal registration and incorporating multi-scale optimization strategies for broader medical imaging applications.

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

References

1. Avants, B.B., Tustison, N., Song, G., et al.: Advanced normalization tools (ANTS). Insight j **2**(365), 1–35 (2009)
2. Chen, X., Liu, M., Wang, R., Hu, R., Liu, D., Li, G., Wang, Y., Zhang, H.: Spatially covariant image registration with text prompts. IEEE Transactions on Neural Networks and Learning Systems (2024)
3. De Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis **52**, 128–143 (2019)
4. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 224–236 (2018)
5. Edstedt, J., Sun, Q., Bökman, G., Wadenbäck, M., Felsberg, M.: Roma: Robust dense feature matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19790–19800 (2024)
6. 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)
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., 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)
9. Hernandez-Matas, C., Zabulis, X., Triantafyllou, A., Anyfanti, P., Douma, S., Argyros, A.A.: Fire: fundus image registration dataset. Modeling and Artificial Intelligence in Ophthalmology **1**(4), 16–28 (2017)
10. Hu, J., Wang, H., Cao, Z., Wu, G., Jonas, J.B., Wang, Y.X., Zhang, J.: Automatic artery/vein classification using a vessel-constraint network for multicenter fundus images. Frontiers in cell and developmental biology **9**, 659941 (2021)

11. Jia, X., Thorley, A., Chen, W., Qiu, H., Shen, L., Styles, I.B., Chang, H.J., Leonardis, A., De Marvao, A., O'Regan, D.P., et al.: Learning a model-driven variational network for deformable image registration. IEEE Transactions on Medical Imaging **41**(1), 199–212 (2021)
12. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. IEEE transactions on medical imaging **29**(1), 196–205 (2009)
13. Lindenberger, P., Sarlin, P.E., Pollefeys, M.: LightGlue: Local feature matching at light speed. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 17627–17638 (2023)
14. Liu, J., Li, X.: Geometrized transformer for self-supervised homography estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9556–9565 (2023)
15. Liu, J., Li, X., Wei, Q., Xu, J., Ding, D.: Semi-supervised keypoint detector and descriptor for retinal image matching. In: European Conference on Computer Vision. pp. 593–609. Springer (2022)
16. Mok, T.C., Chung, A.: Affine medical image registration with coarse-to-fine vision transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20835–20844 (2022)
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. Noyel, G., Thomas, R., Bhakta, G., Crowder, A., Owens, D., Boyle, P.: Superimposition of eye fundus images for longitudinal analysis from large public health databases. Biomedical Physics & Engineering Express **3**(4), 045015 (2017)
19. Potje, G., Cadar, F., Araujo, A., Martins, R., Nascimento, E.R.: XFeat: Accelerated features for lightweight image matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2682–2691 (2024)
20. Qiu, H., Hammernik, K., Qin, C., Chen, C., Rueckert, D.: Embedding gradient-based optimization in image registration networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 56–65. Springer (2022)
21. Revaud, J., De Souza, C., Humenberger, M., Weinzaepfel, P.: R2D2: Reliable and repeatable detector and descriptor. Advances in neural information processing systems **32** (2019)
22. Sivaraman, V.B., Imran, M., Wei, Q., Muralidharan, P., Tamplin, M.R., Grumbach, I.M., Kardon, R.H., Wang, J.K., Zhou, Y., Shao, W.: Retinaregnet: A zero-shot approach for retinal image registration. Computers in Biology and Medicine **186**, 109645 (2025)
23. Tian, X., Anantrasirichai, N., Nicholson, L., Achim, A.: Optimal transport-based graph matching for 3d retinal oct image registration. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 2791–2795. IEEE (2022)
24. Tian, X., Anantrasirichai, N., Nicholson, L., Achim, A.: Tagat: Topology-aware graph attention network for multi-modal retinal image fusion. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 775–784. Springer (2024)
25. Tian, X., Zheng, R., Chu, C.J., Bell, O.H., Nicholson, L.B., Achim, A.: Multi-modal retinal image registration and fusion based on sparse regularization via a

generalized minimax-concave penalty. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1010–1014. IEEE (2019)
26. Truong, P., Apostolopoulos, S., Mosinska, A., Stucky, S., Ciller, C., Zanet, S.D.: Glampoints: Greedily learned accurate match points. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10732–10741 (2019)
27. 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)
28. 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)
29. Zhang, H., Chen, X., Wang, R., Hu, R., Liu, D., Li, G.: Slicer networks. arXiv preprint arXiv:2401.09833 (2024)
30. Zhang, H., Wang, R., Zhang, J., Liu, D., Li, C., Li, J.: Spatially covariant lesion segmentation. arXiv preprint arXiv:2301.07895 (2023)
31. Zhao, S., Lau, T., Luo, J., Eric, I., Chang, C., Xu, Y.: Unsupervised 3D end-to-end medical image registration with volume tweening network. IEEE journal of biomedical and health informatics **24**(5), 1394–1404 (2019)
32. Zheng, J.Q., Wang, Z., Huang, B., Vincent, T., Lim, N.H., Papież, B.W.: Recursive deformable image registration network with mutual attention. In: Annual Conference on Medical Image Understanding and Analysis. pp. 75–86. Springer (2022)