Bridging Classical and Learning-Based Iterative Registration Through Deep Equilibrium Models

Yi Zhang, Yidong Zhao, and Qian Tao(✉)

Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands
{y.zhang-43,y.zhao-8,q.tao}@tudelft.nl

Abstract. Deformable medical image registration is traditionally formulated as an optimization problem. While classical methods solve this problem iteratively, recent learning-based approaches use weight-tied neural networks to mimic this process by unrolling the prediction of deformation fields in a fixed number of steps. However, classical methods typically converge after sufficient iterations, but learning-based unrolling methods lack a theoretical convergence guarantee and show instability empirically. In addition, unrolling methods have a practical bottleneck at training time: GPU memory usage grows linearly with the unrolling steps due to backpropagation through time (BPTT). To address both theoretical and practical challenges, we propose DEQReg, a novel registration framework based on Deep Equilibrium Models (DEQ), which formulates registration as an equilibrium-seeking problem, establishing a natural connection between classical optimization and learning-based unrolling methods. DEQReg maintains constant memory usage, enabling theoretically unlimited iteration steps. Through extensive evaluation on the public brain MRI and lung CT datasets, we show that DEQReg can achieve competitive registration performance, while substantially reducing memory consumption compared to state-of-the-art unrolling methods. We also reveal an intriguing phenomenon: the performance of existing unrolling methods first increases slightly then degrades irreversibly when the inference steps go beyond the training configuration. In contrast, DEQReg achieves stable convergence with its inbuilt equilibrium-seeking mechanism, bridging the gap between classical optimization-based and modern learning-based registration methods.

Keywords: Deformable Image Registration · Deep Equilibrium Models

1 Introduction

Deformable image registration (DIR) is a fundamental task in medical image analysis that aims to discover the spatial correspondence between pairs of images

J. C. Gee et al. (Eds.): MICCAI 2025, LNCS 15962, pp. 88–98, 2026.
https://doi.org/10.1007/978-3-032-04947-6_9

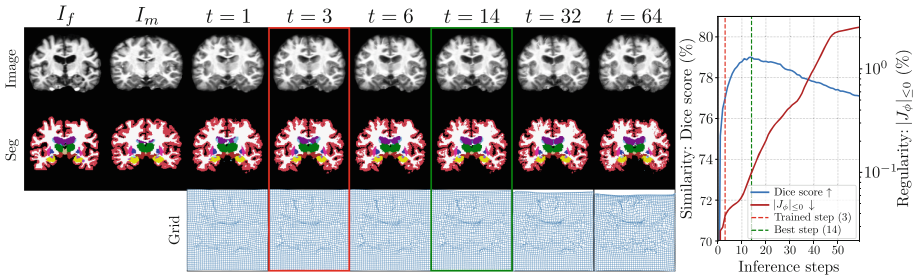

Fig. 1. Performance of a learning-based unrolling method (3 resolution level, 3 steps per level, 9 steps in total). Inference steps t denotes steps on the highest level.

[14,27,28]. DIR can be formulated as an optimization problem to find a parameterized deformation field that aligns the images. The objective typically comprises two terms: image similarity for alignment accuracy, and regularization for deformation smoothness and anatomical plausibility. Historically, this optimization problem has been solved by iterative algorithms such as adaptive gradient descent [2,3,20]. While successful, the iterative nature of these methods leads to long computation time seeking convergence, limiting their real-time clinical applications. With recent advances in deep learning (DL), data-driven approaches have emerged as a promising alternative, by directly inferring the transformation field in a single forward-pass through parametrized neural networks [6,8,23]. More recently, inspired by classical iterative methods, unrolling methods [9,15,23,24,29,30] were proposed to learn the optimization process in multiple steps. Compared to single-step inference, these unrolling methods demonstrated further improved performance [9,24,30].

Despite improved accuracy and generalizability over single-step inference, a fundamental gap exists between learning-based unrolling methods and classical optimization methods. While classical optimization methods pursue convergence for optimal performance and often require >100 iterations, current unrolling methods rely on a limited number of steps (<10) at both training and inference time. At training time, this limitation stems from the reliance on back-propagation through time (BPTT) [26], which causes a linearly growing memory footprint that quickly surpasses the practical GPU capacity. At inference time, extended iterations can be made beyond the training steps, but existing implementations maintain the same iteration steps in both phases. This raises a curious question as to what would happen when inference steps go beyond training ones, which we show in Fig. 1. It can be observed that extended inference steps initially improve the Dice score and then degrade it, while the deformation regularity expressed by percentage of negative Jacobian determinants of deformation field (folded voxels) deteriorates consistently. The empirical results show that learning-based unrolling methods lack a convergence guarantee, which is well-established in classical methods and highly desirable for DIR.

To bridge this conceptual and empirical gap, we propose DEQReg, which seeks a converged estimation of registration fields based on recent Deep Equilibrium Models (DEQ) [5]. DEQReg is the first attempt to formulate the entire learning-based registration process as an end-to-end unsupervised fixed-point problem. It offers both theoretical and practical advantages: first, by explicitly seeking equilibrium points during iterative registration, it establishes a natural connection between classical optimization and learning-based unrolling methods. Second, DEQ's implicit differentiation [5,21] enables potentially unlimited iteration steps with constant memory usage, bypassing the memory limitations of BPTT. Unlike recent works that apply DEQ as a post-processing step for pre-trained models [18] and semi-supervised registration [19], or as a solver for supervised regression and inverse problems [4,13], our work presents two new contributions for DIR:

- We propose DEQReg, the first end-to-end unsupervised registration framework that formulates DIR as a fixed-point equilibrium problem. Extensive experiments on the public brain MRI and lung CT datasets demonstrate that our approach achieves competitive accuracy with constant memory complexity through implicit differentiation.
- We empirically demonstrate that DEQReg achieves stable convergence at inference time while revealing an intriguing phenomenon in current unrolling methods: Their performances are bound to deteriorate after initial slight improvement beyond the predefined training step, highlighting the significance of our principled equilibrium-seeking design.

2 Methods

2.1 Iterative Learning-Based Image Registration via Unrolling

Given a fixed image I_f and a moving image I_m, DIR aims to find a dense displacement field $\mathbf{u}$ that aligns I_m with I_f. The deformation field ϕ is defined as $\phi(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$ for any spatial coordinate $\mathbf{x}$. For notational simplicity, we use ϕ hereafter. DIR can be formulated as an optimization problem:

$$\phi^* = \underset{\phi}{\operatorname{argmin}}\, \mathcal{L}_{\text{sim}}(I_f, I_m \circ \phi) + \lambda \mathcal{L}_{\text{reg}}(\phi), \tag{1}$$

where $\mathcal{L}_{\text{sim}}$ measures image dissimilarity, $\mathcal{L}_{\text{reg}}$ enforces smoothness of the deformation, and λ balances these competing terms. Classical methods typically solve this optimization through gradient descent:

$$\phi_{t+1} = \phi_t - \eta_t \nabla_\phi (\mathcal{L}_{\text{sim}}(I_f, I_m \circ \phi_t) + \lambda \mathcal{L}_{\text{reg}}(\phi_t)), \quad t = 0, 1, \ldots, L, \tag{2}$$

where η_t denotes an adaptive step size. The optimization typically continues until convergence or reaches a maximum iteration count L, which often exceeds 100 steps for sufficient optimization. In iterative learning-based registration, this optimization is often reformulated by unrolling into a fixed number of steps T.

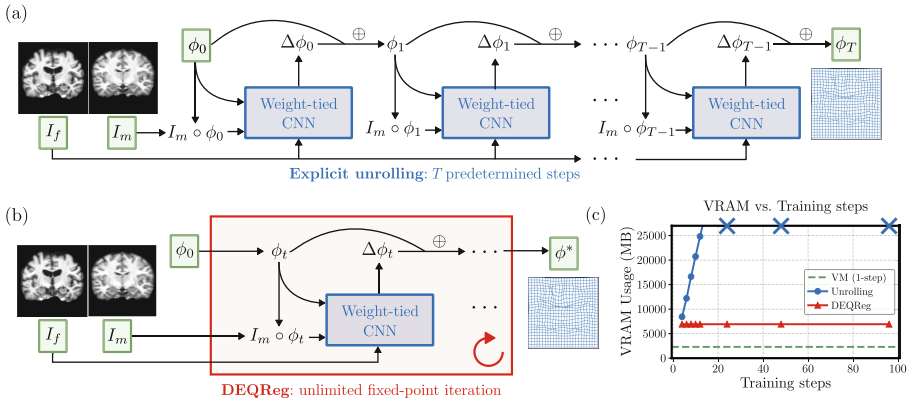

Fig. 2. (a) Illustration of a typical learning-based registration method that uses explicit unrolling of T steps in training. (b) DEQReg uses implicit fixed-point iteration, enabling theoretically unlimited steps in training. (c) Memory usage when training on a 3D brain MRI pair, where blue crosses indicate memory overflow for a 24 GB GPU. (Color figure online)

The update rule is modeled using a weight-tied neural network as shown in Fig. 2, where parameters θ are shared across steps to learn the optimal update function f_θ:

$$\phi_{t+1} = \phi_t + f_\theta(I_f, I_m, \phi_t), \quad t = 0, 1, \ldots, T-1. \tag{3}$$

The update function f_θ is learned by minimizing the expected loss over a training dataset $\mathcal{D} = \{(I_f^i, I_m^i)\}_{i=1}^M$:

$$\min_\theta \mathbb{E}_{(I_f, I_m)\sim\mathcal{D}}[\mathcal{L}(I_f, I_m, \phi_T) + \sum_{t=0}^{T-1} w_t \mathcal{L}(I_f, I_m, \phi_t)], \tag{4}$$

where $w_t > 0$ is the weight for the correction of intermediate states.

Training such unrolling networks through BPTT [26] requires storing *all* intermediate states, leading to $\mathcal{O}(T)$ memory usage and typically restricting steps to $T < 10$ due to hardware constraints.

2.2 Deep Equilibrium Models for Deformable Image Registration

To overcome the limitations of unrolling approaches that can only allow for a fixed number of iterations, we reformulate the iterative process using the Deep Equilibrium Model (DEQ) [5] to find the equilibrium of a fixed-point equation. In its general form, a fixed point $\mathbf{x}^*$ of a function g satisfies $\mathbf{x}^* = g(\mathbf{x}^*)$. Here, we design our network f_θ to learn residual updates, where DEQ seeks an equilibrium solution ϕ^* through residual updates that satisfy this fixed-point condition:

$$\phi^* = g_\theta(I_f, I_m, \phi^*) = \phi^* + f_\theta(I_f, I_m, \phi^*), \tag{5}$$

which is equivalent to finding ϕ^* where $f_\theta(I_f, I_m, \phi^*) = 0$. This formulation provides a principled way to enforce convergence through vanishing updates, analogous to classical optimization but with the learned update by f_θ.

Unlike BPTT that requires storing and backpropagating throughout the forward trajectory, a key advantage of DEQ is its ability to compute derivatives directly from the derived equilibrium state. When the system converges to a fixed point ϕ^* where $f_\theta(I_f, I_m, \phi^*) = 0$, we can leverage the implicit function theorem (IFT) [5,21] to compute gradients without tracking intermediate states:

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial \mathcal{L}}{\partial \phi^*} \left(I - \frac{\partial g_\theta}{\partial \phi^*} \right)^{-1} \frac{\partial g_\theta(I_f, I_m, \phi^*)}{\partial \theta}, \tag{6}$$

where I is an identity matrix of the same size as $\frac{\partial g_\theta}{\partial \phi^*}$. Leveraging IFT, equilibrium points can be derived using any black-box root-finding algorithm, *e.g.* Broyden's method [7]. Figure 2 provides a schematic comparison between DEQ and existing unrolling methods. By focusing on the equilibrium state rather than the iteration trajectory, DEQ achieves constant memory complexity while enabling theoretically unlimited iterations, with convergence design (i.e. fixed point condition).

2.3 Training the DEQ

Despite the memory efficiency of DEQ, directly computing the inverse Jacobian term $(I - \frac{\partial g_\theta}{\partial \phi^*})^{-1}$ in Eq. (6) remains computationally intensive. To address this challenge, we use the phantom gradient approach [11,12], which offers a structured approximation of IFT that preserves the same descent direction while being computationally efficient. The key idea is to approximate the inverse Jacobian through a sequence of damped iterations. Starting from the converged fixed point $\phi^0 = \phi^*$, we generate this sequence with a damping factor $\tau \in (0, 1)$:

$$\phi^{p+1} = \tau g_\theta(I_f, I_m, \phi^p) + (1 - \tau)\phi^p, \quad p = 0, \ldots, K - 1. \tag{7}$$

Using the generated sequence $\{\phi^p\}_{p=0}^{K-1}$, we can approximate the gradient as:

$$\frac{\partial \mathcal{L}}{\partial \theta} \approx \frac{\partial \mathcal{L}}{\partial \phi^*} \mathbf{A}, \quad \text{where} \quad \mathbf{A} = \tau \sum_{k=0}^{K-1} \prod_{s=k+1}^{K-1} \left(\tau \frac{\partial g_\theta}{\partial \phi} \bigg|_{\phi^s} + (1 - \tau) I \right) \frac{\partial g_\theta}{\partial \theta} \bigg|_{\phi^k}. \tag{8}$$

This truncated gradient converges to the exact implicit gradient in Eq. 6 as $K \to \infty$ under mild conditions (see [12] for detailed theoretical analysis). Generally, larger K adds higher-order Neumann terms as in Eq. 8, improving estimation accuracy at a cost of $\mathcal{O}(K)$ total memory consumption. In the case of K=1, the estimation recovers Jacobian-free backpropagation [10], with $\mathcal{O}(1)$ memory complexity but less stability. In DEQReg, we set $K = 2$ to balance accuracy and efficiency.

Additionally, we can leverage intermediate state correction during the forward process to regularize the iteration behavior [4,9,29] in training. Given a

training trajectory $[\phi_0, \phi_1, \ldots, \phi_N, \phi^*]$, we sample S intermediate states with equal intervals, where $\mathcal{S} \subset \{0, 1, \ldots, N\}$ with $|\mathcal{S}| = S$, to compose the total loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}(I_f, I_m, \phi^*) + \gamma \sum_{t \in \mathcal{S}} \mathcal{L}(I_f, I_m, \phi_t), \quad \gamma > 0, \tag{9}$$

The intermediate state gradients are computed via phantom gradients, enabling effective training of the equilibrium-seeking network f_θ.

3 Experiments

Datasets. We evaluated our method on two public datasets: (1) **OASIS** [22], a brain MRI dataset containing 414 T1-weighted scans for inter-subject registration. Following [17], scans were preprocessed and resampled to a size of $128 \times 128 \times 128$, with 30 anatomical segmentation labels (300/30/84 for train/val/test). (2) **NLST** [1], a lung CT dataset from Learn2Reg challenge [16] with 210 intra-subject inhale-exhale pairs. The affinely prealigned scans are of size $224 \times 192 \times 224$, with keypoint annotations and lung masks (170/10/30 for train/val/test).

Evaluation Metrics. We evaluate both anatomical alignment and deformation smoothness. For alignment, we compute the average Dice score and Hausdorff distance (HD) on OASIS using the structures in [17], and keypoint target registration error (TRE) and lung Dice score on NLST. For deformation smoothness, we compute two Jacobian-based metrics ($J_\phi = \nabla\phi$): the percentage of non-diffeomorphic (folded) voxels $|J_\phi|_{\leq 0}$ and the standard deviation of log-Jacobian determinant $\text{std}(\log |J_\phi|)$ quantifying the spatial consistency of volume changes.

Baseline Methods. We compare our method against the following approaches: (1) **elastix** [20], a widely-used classical registration method, configured with 3-level B-spline transformation and grid spacing of 4 voxels; (2) **VoxelMorph** [6], a well-established one-step DL method; and two state-of-the-art unrolling methods: (3) **GraDIRN** [24], which integrates explicit similarity gradients alongside neural network predictions in 3 resolution levels with $T = 3$ per level ($w_t = 0$), and (4) **RIIR** [29] with single resolution and $T = 6$ iterations using exponentially increasing weights ($w_t = 10^{\frac{t-1}{T-1}}$). Our proposed DEQReg employs a fixed-point iteration strategy with maximum $T = 48$ steps with $S = 3$ and $\gamma = 0.5$ in Eq. 9.

Convergence Analysis Setup. In this study, we investigate how DL-based unrolling methods behave when varying inference steps while keeping training steps fixed. Besides GraDIRN and RIIR, we evaluate two additional variants: (1) Unrolling (the vanilla version of GraDIRN without incorporating similarity gradients) and (2) DEQReg-12, DEQReg with $T = 12$ and $S = 0$. For GraDIRN and Unrolling, we vary the steps at the highest resolution.

Implementation Details. For fair comparison, we use the same U-Net [25] configuration as VoxelMorph and RIIR for DEQReg. For each step t, our network takes concatenated $[I_f, I_m \circ \phi_t, \phi_t]$ as input, consistent with other iterative

Table 1. Quantitative comparison of registration methods. Mem. denotes GPU memory usage and Time shows training/inference time per pair. Best performance is in **bold** and second best is underlined, with * denotes statistical significance ($p < 0.05$, Wilcoxon).

OASIS	Dice (%)↑	HD (mm)↓	$\lvert J_\phi\rvert_{\leq 0}$ (%)↓	std($\log \lvert J_\phi\rvert$) ↓	Mem. (MB)	Time (s)
Affine	61.2(7.3)	8.56(1.34)	–	–	–	–
elastix	75.2(3.8)	2.89(0.58)	0.068(0.041)	0.551(0.134)	–	35.45
VoxelMorph	77.8(3.7)	2.74(0.49)	0.063(0.031)	0.601(0.121)	2645	0.20/0.14
GraDIRN	80.3(3.2)	2.47(0.41)	0.071(0.029)	0.568(0.114)	16619	0.59/0.24
RIIR	80.6(3.1)	**2.43**(0.39)	0.058(0.031)	0.556(0.118)	12542	0.60/0.30
DEQReg	**80.7***(3.1)	2.43(0.40)	**0.053***(0.033)	**0.542***(0.123)	6925	1.09/0.98
NLST	TRE (mm)↓	Dice (%)↑	$\lvert J_\phi\rvert_{\leq 0}$ (%)↓	std($\log \lvert J_\phi\rvert$) ↓	Mem. (MB)	Time (s)
Affine	8.43(3.97)	87.3(4.1)	–	–	–	–
elastix	3.24(1.74)	94.8(0.7)	0.110(0.136)	0.636(0.370)	–	32.51
VoxelMorph	4.08(2.31)	95.6(1.2)	0.143(0.290)	0.671(0.506)	1522	0.24/0.12
GraDIRN	2.26(1.87)	96.5(1.2)	0.165(0.278)	0.738(0.512)	9313	0.41/0.18
RIIR	**2.21***(1.78)	96.5(1.1)	0.120(0.250)	0.598(0.481)	7029	0.40/0.16
DEQReg	2.23(2.11)	**96.6**(1.5)	**0.079***(0.078)	**0.586***(0.290)	3983	0.74/0.53

methods. The loss combines local normalized cross correlation (LNCC) with window size 5 as $\mathcal{L}_{\text{sim}}$ and diffusion regularization as $\mathcal{L}_{\text{reg}}$, with $\lambda = 0.1$ (OASIS) and 0.125 (NLST). Models were trained using AdamW ($\text{lr} = 1 \times 10^{-4}$) for 100 epochs. Code and pretrained weights are made publicly available[1].

4 Results and Discussions

Main Results. Table 1 presents a comprehensive comparison of our DEQReg against classical and learning-based approaches, with visual results shown in Fig. 4. DEQReg achieves a statistically significant improvement in Dice scores while maintaining competitive HD on OASIS, and demonstrates comparable performance in TRE and Dice score on NLST. Notably, by decomposing deformation into more steps, DEQReg achieves the lowest percentage of $\lvert J_\phi\rvert_{\leq 0}$ and std($\log \lvert J_\phi\rvert$), indicating superior deformation regularity. Though DEQReg requires relatively higher inference time, this can be mitigated by using solver acceleration, which can largely reduce iterations. Remarkably, DEQReg substantially reduces memory consumption compared to other DL methods.

Convergence Analysis. Figure 3 and Table 2 present the convergence behavior of different methods. The vanilla unrolling method shows a severe performance drop beyond its trained steps. Although GraDIRN and RIIR show modest improvement when inference steps exceed training configuration due to their image similarity-driven updates, they eventually suffer from deteriorating deformation regularity (shown in Fig. 4, row 5) and Dice. In contrast, both DEQReg

[1] https://gitlab.tudelft.nl/ai4medicalimaging/deqreg.

Table 2. Comparison of registration performance at trained steps versus best achievable steps on validation set. Best performance is in **bold** and second best is underlined.

Method	Trained Steps			Best Steps		
	Steps	Dice (%)	$\lvert J_\phi\rvert_{\leq 0}$ (%)	Steps	Dice (%)	$\lvert J_\phi\rvert_{\leq 0}$ (%)
Unrolling	3	77.95(4.67)	0.064(0.031)	3	77.95(4.67)	0.064(0.031)
GraDIRN	3	78.69(4.20)	0.057(0.029)	14	79.44(3.57)	0.078(0.033)
RIIR	6	79.30(3.72)	0.066(0.031)	12	79.39(3.58)	0.062(0.028)
DEQReg-12	12	78.19(4.68)	0.077(0.041)	96	**79.53**(3.62)	0.087(0.039)
DEQReg	48	**79.33**(3.97)	**0.053**(0.030)	96	79.45(3.84)	**0.056**(0.031)

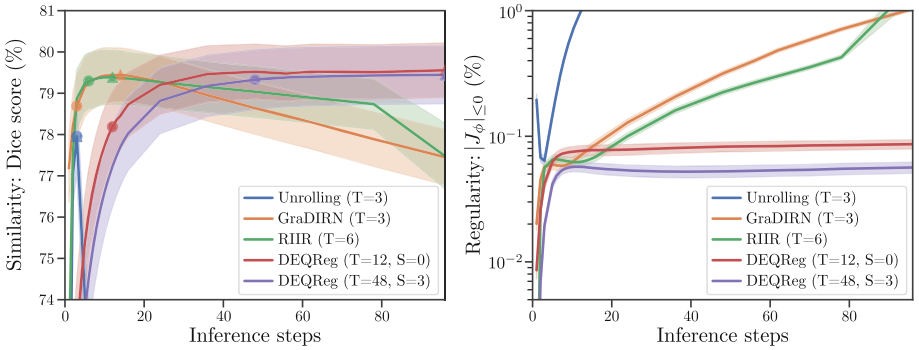

Fig. 3. Convergence analysis for different registration methods. Left: Dice score (%) where • indicates the results at the trained step count and ▲ marks the steps achieving best Dice performance; Right: negative Jacobian determinant ratio (%). Shaded areas show standard deviation. DEQReg exhibits convergence beyond training steps.

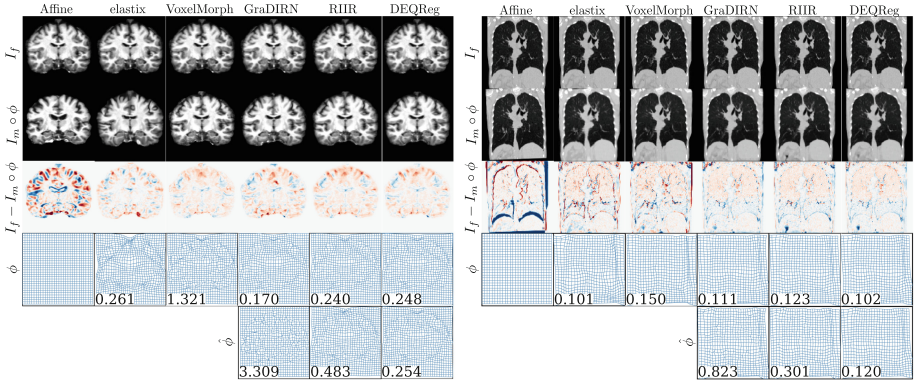

Fig. 4. Qualitative comparison of registration methods on brain MRI and lung CT images. In the last row, all unrolling methods go to the excessive step count of $T = 96$ to evaluate the convergence of ϕ to $\hat{\phi}$. The percentage of $|J_\phi|_{\leq 0}$ is reported at the bottom-left corner to indicate deformation regularity.

variants demonstrate stable convergence. Notably, DEQReg shows potential for improvement with increased iterations, as its optimal Dice occur at maximum steps. Comparing the gap between trained-step and best-achievable performance, methods with intermediate state correction as in Eq. 9 (RIIR and DEQReg) achieve near-optimal results at trained steps, confirming their effectiveness. In particular, the full DEQReg consistently maintains superior deformation regularity while achieving competitive accuracy, as shown in Fig. 3.

5 Conclusion

For deformable medical image registration, classical optimization methods typically converge, but modern learning-based unrolling methods lack the theoretical guarantee of convergence and suffer from memory bottlenecks. This creates a gap between the principled convergence of classical methods and the lack of convergence in learning-based methods. In this work, we propose DEQReg to bridge this gap by reformulating learning-based registration as an equilibrium-seeking problem. Our experiments show that this reformulation enables stable convergence, in contrast to existing unrolling methods that tend to degrade when further unrolled beyond the training steps. DEQReg achieves competitive accuracy and improved regularity with constant memory usage, effectively unifying classical optimization principles and learning-based advantages in speed and flexibility.

Acknowledgments. We gratefully acknowledge the Dutch Research Council (NWO) and Amazon Science for financial and computing support.

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

References

1. Aberle, D.R., et al.: Reduced lung-cancer mortality with low-dose computed tomographic screening. N. Engl. J. Med. **365**(5), 395–409 (2011)
2. Ashburner, J.: A fast diffeomorphic image registration algorithm. Neuroimage **38**(1), 95–113 (2007)
3. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage **54**(3), 2033–2044 (2011)
4. Bai, S., Geng, Z., Savani, Y., Kolter, J.Z.: Deep equilibrium optical flow estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 620–630 (2022)
5. Bai, S., Kolter, J.Z., Koltun, V.: Deep equilibrium models. Adv. Neural Inf. Process. Syst. **32** (2019)
6. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38**(8), 1788–1800 (2019)
7. Broyden, C.G.: A class of methods for solving nonlinear simultaneous equations. Math. Comput. **19**(92), 577–593 (1965)

8. 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. Med. Image Anal. **52**, 128–143 (2019)
9. Falta, F., Hansen, L., Heinrich, M.P.: Learning iterative optimisation for deformable image registration of lung CT with recurrent convolutional networks. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13436, pp. 301–309. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_29
10. Fung, S.W., Heaton, H., Li, Q., McKenzie, D., Osher, S., Yin, W.: JFB: Jacobian-free backpropagation for implicit networks. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36 (2022)
11. Geng, Z., Kolter, J.Z.: TorchDEQ: a library for deep equilibrium models. arXiv preprint arXiv:2310.18605 (2023)
12. Geng, Z., Zhang, X.Y., Bai, S., Wang, Y., Lin, Z.: On training implicit models. Adv. Neural. Inf. Process. Syst. **34**, 24247–24260 (2021)
13. Gilton, D., Ongie, G., Willett, R.: Deep equilibrium architectures for inverse problems in imaging. IEEE Trans. Comput. Imaging **7**, 1123–1133 (2021)
14. Haskins, G., Kruger, U., Yan, P.: Deep learning in medical image registration: a survey. Mach. Vis. Appl. **31**, 1–18 (2020)
15. Hering, A., Ginneken, B., Heldmann, S.: mlVIRNET: multilevel variational image registration network. In: Shen, D., et al. (eds.) MICCAI 2019, Part VI. LNCS, vol. 11769, pp. 257–265. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_29
16. Hering, A., et al.: Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. IEEE Trans. Med. Imaging **42**(3), 697–712 (2022)
17. Hoopes, A., Hoffmann, M., Fischl, B., Guttag, J., Dalca, A.V.: HyperMorph: amortized hyperparameter learning for image registration. In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds.) IPMI 2021. LNCS, vol. 12729, pp. 3–17. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78191-0_1
18. Hu, J., Gan, W., Sun, Z., An, H., Kamilov, U.S.: A plug-and-play image registration network. arXiv preprint arXiv:2310.04297 (2023)
19. Jena, R., Chaudhari, P., Gee, J.C.: Deep implicit optimization enables robust learnable features for deformable image registration. Med. Image Anal., 103577 (2025)
20. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging **29**(1), 196–205 (2009)
21. Krantz, S.G., Parks, H.R.: The Implicit Function Theorem: History, Theory, and Applications. Springer, New York (2002). https://doi.org/10.1007/978-1-4614-5981-1
22. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. **19**(9), 1498–1507 (2007)
23. Mok, T.C.W., Chung, A.C.S.: Large deformation diffeomorphic image registration with Laplacian pyramid networks. In: Martel, A.L., et al. (eds.) MICCAI 2020, Part III. LNCS, vol. 12263, pp. 211–221. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_21

24. Qiu, H., Hammernik, K., Qin, C., Chen, C., Rueckert, D.: Embedding gradient-based optimization in image registration networks. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. Lecture Notes in Computer Science, vol. 13436, pp. 56–65. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_6
25. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015, Part III. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
26. Sherstinsky, A.: Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D **404**, 132306 (2020)
27. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: a survey. IEEE Trans. Med. Imaging **32**(7), 1153–1190 (2013)
28. Staring, M., vander Heide, U.A., Klein, S., Viergever, M.A., Pluim, J.P.: Registration of cervical MRI using multifeature mutual information. IEEE Trans. Med. Imaging **28**(9), 1412–1421 (2009)
29. Zhang, Y., Zhao, Y., Xue, H., Kellman, P., Klein, S., Tao, Q.: Recurrent inference machine for medical image registration. arXiv preprint arXiv:2406.13413 (2024)
30. Zhao, S., Dong, Y., Chang, E.I., Xu, Y.: Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10600–10610 (2019)