Deformable Medical Image Registration Under Distribution Shifts with Neural Instance Optimization

Tony C. W. Mok*[1,2], Zi Li[1,2], Yingda Xia[1], Jiawen Yao[1,2], Ling Zhang[1], Jingren Zhou[1,2], and Le Lu[1]

[1] DAMO Academy, Alibaba Group
[2] Hupan Lab, 310023, Hangzhou, China
mokchi-wing.mcw@alibaba-inc.com

Abstract. Deep-learning deformable image registration methods often struggle if test-image characteristic shifts from the training domain, such as the large variations in anatomy and contrast changes with different imaging protocols. Gradient descent-based instance optimization is often introduced to refine the solution of deep-learning methods, but the performance gain is minimal due to the high degree of freedom in the solution and the absence of robust initial deformation. In this paper, we propose a new instance optimization method, Neural Instance Optimization (NIO), to correct the bias in the deformation field caused by the distribution shifts for deep-learning methods. Our method naturally leverages the inductive bias of the convolutional neural network, the prior knowledge learned from the training domain and the multi-resolution optimization strategy to fully adapt a learning-based method to individual image pairs, avoiding registration failure during the inference phase. We evaluate our method with gold standard, human cortical and subcortical segmentation, and manually identified anatomical landmarks to contrast NIO's performance with conventional and deep-learning approaches. Our method compares favourably with both approaches and significantly improves the performance of deep-learning methods under distribution shifts with 1.5% to 3.0% and 2.3% to 6.2% gains in registration accuracy and robustness, respectively.

1 Introduction

Deformable medical image registration aims to establish non-linear correspondence of anatomy between image scans, which is essential in a comprehensive medical image processing and analysis pipeline. Deep learning-based image registration (DLIR) methods can achieve remarkable results on training and testing images from the same distribution, as evidenced by tremendous medical image registration benchmarks [10,6]. However, DLIR method remains notoriously vulnerable to distribution shifts. Distribution shift refers to the existence

* Corresponding author.

of significant divergence between the distributions of the training and the test data [29]. Registration accuracy and robustness of DLIR method could be degraded substantially when the test-image characteristic shifts from the training domain. Different from the learning-based methods, conventional image registration methods [22,1] often register images with an iterative optimization strategy, which circumvents the need for learning. Although conventional image registration methods excel in robustness and diffeomorphic properties under distribution shifts, the resulting solutions are often suboptimal and can be time-consuming with high-resolution 3D image volumes, as indicated by recent studies[3,19].

Combining an initial deformation field predicted by DLIR with instance optimization [27,9,10,3] or further learning fine-tuning steps [28] has drawn growing attention in the community. Instance optimization refers to the process of iteratively optimizing the deformation field along with the optimizer, which is similar to classic optical flow estimation [21]. Instance optimization often serves as a post-processing step to refine the solution from DLIR method during the inference phase. Nevertheless, there is only minimal performance gain using instance optimization under distribution shifts. First, the effectiveness of instance optimization relies on the assumption that the model in DLIR is capable of estimating a robust initial deformation prediction for instance optimization. This assumption cannot be held when test-image characteristic shifts from the training domain, such as large variations in brain anatomy and contrast changes from diverse imaging protocols. Second, taking into account that non-linear image registration is an ill-posed problem [23], the searching space of instance optimization with a high degree of freedom is intractable, resulting in a minimal performance gain. Third, while the multilevel optimization technique has been proven very efficient in conventional variational registration approaches to avoid local minima [2], this technique is not applicable to instance-specific optimization when initialized with the dense and fine-resolution solution from the DLIR method as the fine details of the deformation field from the DLIR method will be distorted during downsampling in the coarse-to-fine optimization pipeline.

While there are vast research studies [20,13,17,7,26,14] on deformable image registration, the robustness and generalizability of DLIR method against distribution shifts are less explored in the context of image registration. Zhu et al.[31] use a cascaded U-Net structure to minimize the distribution shifts between training and target domain by performing test-time training. To against the diverse contrast change in brain MR registration, Hoffmann et al. [11] propose to train networks with synthetic images generated with synthetic anatomical label maps, which requires access to the anatomical delineation of the training set. Different from the above studies, we focus on maximizing the registration accuracy and robustness of instance (individual) image pair under distribution shifts without access to the delineation of the anatomical structures in the training or test set.

To this end, we present neural instance optimization (NIO), a new instance optimization method that is capable of significantly improving the registration performance of deep-learning methods during the inference phase, leading to robust generalizability under domain shift.

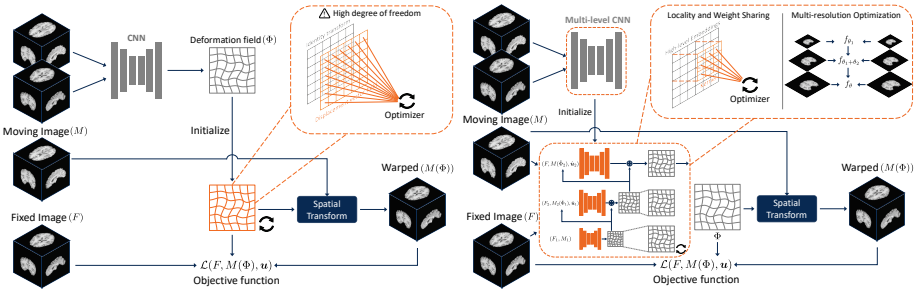

Fig. 1. Overview of the (a) deep-learning method followed by instance optimization approach and (b) the proposed neural instance optimization method. Instead of directly manipulating the deformation field with instance-specific optimization, our proposed method implicitly updates the deformation field by adapting the CNN to each target image pair with the multiresolution optimization strategy.

2 Methods

Deformable image registration aims to establish a dense non-linear correspondence between a fixed image F and a moving image M, subject to a smoothness regularization to penalize implausible solutions. Existing DLIR methods often formulate deformable image registration as a learning problem $\Phi = f_\theta(F, M)$, where f_θ is parameterized with a convolutional neural network (CNN) and θ denotes the set of learning parameters in CNN. Inherit from the learning nature, DLIR methods are prone to distribution shifts and the registration performance could be degraded. Our goal is to refine the deformation field in the inference phase to avoid failure in registration due to distribution shifts and maintain the sub-pixel accuracy of the registration simultaneously.

2.1 Neural Instance Optimization

Formulation of Instance Optimization Recall that the instance optimization method can be adopted as a post-processing step to refine the predicted deformation field predicted from the DLIR method in the inference phase. Formally, the formulation of instance optimization is defined as follows:

$$\Phi^* = \arg\min_{\Phi_\theta} \mathcal{L}_{sim}(F, M(\Phi_\theta)) + \lambda \mathcal{L}_{reg}(\boldsymbol{u}_\theta), \tag{1}$$

where $\mathcal{L}_{sim}$ denotes the dissimilarity measure, $\mathcal{L}_{reg}$ is the smoothness regularization and λ is a hyperparameter. The deformation field is defined as $\Phi_\theta = \boldsymbol{u}_\theta + \boldsymbol{Id}$, where $\boldsymbol{Id}$ denotes the identity transform in mutual space Ω. Specifically, Φ_θ in instance optimization is first initialized with the displacement from a feed-forward prediction of DLIR network, followed by an update on $\boldsymbol{u}_\theta$ with gradient descent to maximize the similarity between F and $M(\Phi_\theta)$ in an iterative manner, as shown in Fig. 1(a). While instance optimization could be used to refine DLIR's

solution, the performance gain to registration is minimal under moderate or severe distribution shifts due to the high degree of freedom in Φ_θ and a robust initial deformation prediction cannot be guaranteed by DLIR under distribution shifts, violating the underlying assumption of instance optimization approach.

Neural Instance Optimization (NIO) To circumvent the pitfalls in the instance optimization approach, we implicitly refine the resulting deformation field by updating the parameters in the CNN instead of directly manipulating the deformation field as in the instance optimization approach. To exemplify the idea, we first parametrize an example of the function f_θ with the deep Laplacian pyramid image registration network (LapIRN) [18], i.e., $\Phi = f_\theta(F, M)$. The overview of our proposed method is illustrated in Fig. 1(b). Mathematically, our method reformulates the problem from Eq. 1 as follows:

$$\theta^* = \arg\min_\theta \mathcal{L}_{sim}(F_t, M_t \circ f_\theta(F_t, M_t)) + \lambda \mathcal{L}_{reg}(\boldsymbol{u}), \tag{2}$$

where F_t and M_t belong to the fixed and moving image pair in the target dataset $\mathcal{D}_{test}$, and $\circ$ denotes the composition operator, i.e., $M_t \circ f_\theta(F_t, M_t) = M_t(\Phi)$. During the inference phase of each image pair, the parameters in the network θ are first initialized with the pre-trained θ_0 on the training dataset. Then, we use gradient descent for η iterations searching for the optimal set of θ^* that adapt the model f_θ for individual (F_t, M_t) pair in $\mathcal{D}_{test}$. Compared with the instance optimization approach, our proposed method is two-fold: first, the intrinsic inductive bias, i.e., weight sharing and locality, embedded in the CNN implicitly reduces the degree of freedom of the solution, reducing the searching space in the optimization problem; second, the prior knowledge of registering the images in the training set can be transferred to the individual test image pair during the optimization, avoiding the sub-optimal solution in the instance optimization.

2.2 Multi-resolution Optimization Strategy with NIO

Optimizing the CNN f_θ during the inference phase could be computationally intensive, especially for DLIR methods with multilevel network architecture. This restricts the maximum number of iterations of NIO in the inference phase. To alleviate this issue, we propose a multi-resolution optimization strategy for NIO. Specifically, given a L-level LapIRN framework f_θ with $\theta \in \{\theta_1, \ldots, \theta_L\}$ such that θ_i represents the parameters of the network in level i. We first create the image pyramid for input images by downsampling the input images with trilinear interpolation to obtain $F_i \in \{F_1, \ldots, F_L\}$ (and $M_i \in \{M_1, \ldots, M_L\}$), where F_i denotes the downsampled F with a scale factor $0.5^{(L-i)}$ and $F_L = F$. The network f_θ is then optimized in a coarse-to-fine manner, starting with the coarsest input pair (F_1 and M_1) and network f_{θ_1}. Then, we progressively add the network in the next level $f_{\theta_{i+1}}$ into optimization until the optimization of the final level f_θ is completed, as shown in Fig. 1 (b). The output displacement field $\boldsymbol{u}$ of f_θ is formed by aggregating the upsampled displacement fields $\hat{\boldsymbol{u}}_i$ from each level via element-wise addition. We set $L = 3$ throughout this paper. By

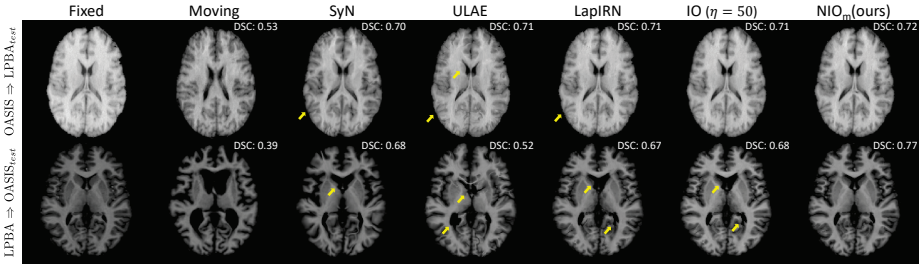

Fig. 2. Example axial MR slices from the moving, atlas and resulting warped images from SyN, ULAE, LapIRN, LapIRN followed by IO and NIO$_m$ for the cross-dataset brain MR registration. Major artifacts are highlighted with yellow arrows.

adjusting the number of iterations η_i in each level of the optimization, most of the computation is deployed with images at a lower resolution and the optimization naturally inherits the advantage of the conventional multi-resolution strategy, enabling a better trade-off between registration accuracy and runtime.

2.3 Loss Function

To further accelerate the runtime in the inference phase, we discard the similarity pyramid in the vanilla LapIRN [19] and adopt the local normalized cross-correlation (LNCC) with window size 7^3 as similarity function $\mathcal{L}_{sim}$. We adopt a diffusion regularizer as $\mathcal{L}_{reg}$ to encourage smooth solutions and penalise implausible solutions. We follow [13] to further normalize the cost function with λ, i.e., $\lambda \in [0, 1]$. The loss function for level i in the proposed multi-resolution optimization scheme is defined as:

$$\mathcal{L}_i = (1-\lambda)\text{LNCC}(F, M(\hat{\Phi}_i)) + \lambda||\nabla \hat{\boldsymbol{u}}_i|| \quad i \in \{1, 2, \ldots, L\}, \tag{3}$$

where $\hat{(\cdot)}$ is the upsampling operator with trilinear interpolation, which upsample $\boldsymbol{u}_i$ and Φ_i to match the size of F and M, maintaining the consistency of the objective function among all the levels.

3 Experiments

Data and Pre-processing We evaluate our method on cross-dataset brain atlas registration and intra-patient inspiration-expiration Chest CT registration. For brain atlas registration, we use 414 T1-weighted brain MR scans from the OASIS dataset [15,16] and 40 brain MR scans from the LPBA dataset [24,25]. Three experiments are conducted to assess the registration accuracy, robustness and plausibility of our method with dataset shifts and insufficient training data on brain atlas registration: 1) training on OASIS dataset and testing on LPBA dataset; 2) training on LPBA and testing on OASIS; and 3) train on LPBA and testing on LPBA datasets. We follow the pre-processing pipeline in

Table 1. Quantitative results of brain MR atlas registration. $X^{n}_{train} \Rightarrow Y_{test}$ represents the experiment of training on n scans from the X dataset and registering images in the test set of Y dataset to pre-defined atlases. The subscript of DSC indicates the number of anatomical structures involved. Initial: Affine spatial normalization. ↑: higher is better, and ↓: lower is better. †: $p < 0.001$, in comparison to NIO_m. *: $p < 0.05$, in comparison to NIO_m.

Method	$\text{OASIS}^{n=250}_{train} \Rightarrow \text{LPBA}_{test}$			$\text{LPBA}^{n=20}_{train} \Rightarrow \text{OASIS}_{test}$			$\text{LPBA}^{n=20}_{train} \Rightarrow \text{LPBA}_{test}$			T_{test} (sec) ↓
	DSC_{54} ↑	DSC30_{54} ↑	$\text{std}(\lvert J_\phi \rvert)$ ↓	DSC_{35} ↑	DSC30_{35} ↑	$\text{std}(\lvert J_\phi \rvert)$ ↓	DSC_{54} ↑	DSC30_{54} ↑	$\text{std}(\lvert J_\phi \rvert)$ ↓	
Initial	0.535 ± 0.05†	0.476 ± 0.02†	-	0.598 ± 0.08†	0.498 ± 0.05†	-	0.535 ± 0.05†	0.476 ± 0.02†	-	-
(Iterative) NiftyReg [22]	0.690 ± 0.02†	0.664 ± 0.02*	0.345 ± 0.04	0.763 ± 0.04†	0.714 ± 0.03†	0.655 ± 0.11	0.690 ± 0.02†	0.664 ± 0.02†	0.345 ± 0.04	126.7 ± 23.0
(Iterative) SyN [1]	0.692 ± 0.02†	0.666 ± 0.01*	0.255 ± 0.02	0.772 ± 0.03†	0.729 ± 0.03†	0.342 ± 0.04	0.692 ± 0.02†	0.666 ± 0.01†	0.255 ± 0.02	892.4 ± 56.0
MPR [14]	0.639 ± 0.03†	0.598 ± 0.02†	**0.224** ± 0.02	0.662 ± 0.08†	0.561 ± 0.05†	**0.325** ± 0.01	0.656 ± 0.03†	0.620 ± 0.02†	**0.206** ± 0.07	**0.08** ± 0.01
ULAE [26]	0.665 ± 0.03†	0.632 ± 0.02†	0.471 ± 0.04	0.607 ± 0.05†	0.543 ± 0.03†	0.675 ± 0.04	0.700 ± 0.02†	0.676 ± 0.01†	0.382 ± 0.04	0.20 ± 0.00
LapIRN [18]	0.694 ± 0.02†	0.665 ± 0.02*	0.435 ± 0.05	0.793 ± 0.04†	0.743 ± 0.04†	0.654 ± 0.06	0.710 ± 0.02†	0.688 ± 0.01*	0.410 ± 0.04	0.14 ± 0.01
LapIRN + IO ($\eta = 50$)	0.696 ± 0.02†	0.667 ± 0.02*	0.440 ± 0.05	0.801 ± 0.04†	0.752 ± 0.04†	0.684 ± 0.05	0.712 ± 0.02*	0.690 ± 0.01*	0.426 ± 0.04	33.37 ± 0.01
LapIRN + IO ($\eta = 200$)	0.697 ± 0.02†	0.668 ± 0.02*	0.450 ± 0.04	0.805 ± 0.04†	0.759 ± 0.04†	0.714 ± 0.06	0.713 ± 0.02*	0.692 ± 0.01*	0.441 ± 0.04	60.80 ± 0.11
NIO ($\eta = 50$)	0.708 ± 0.03	0.678 ± 0.02	0.443 ± 0.04	0.815 ± 0.02	0.785 ± 0.02	0.605 ± 0.04	0.719 ± 0.01	0.701 ± 0.01	0.432 ± 0.03	47.22 ± 0.91
NIO_m ($\eta = [100, 60, 30]$)	**0.711** ± 0.02	**0.685** ± 0.02	0.434 ± 0.03	**0.817** ± 0.02	**0.789** ± 0.02	0.600 ± 0.04	**0.721** ± 0.01	**0.704** ± 0.01	0.425 ± 0.03	52.71 ± 0.57

[13] to preprocess the OASIS dataset. We divide the OASIS dataset into 250, 5, and 159 scans, and divide the LPBA dataset into 20, 3, and 17 scans for training, validation, and test sets, respectively. 35 and 54 anatomical structures are included in the evaluation for OASIS and LPBA datasets, respectively. We resample all MR scans to isotropic voxel sizes of 1 mm and center-cropped all scans to 144×160×192. We select Case 285 and (S24, S25) from the test set of OASIS and LPBA as atlases, respectively.

For Chest CT registration, we use publicly available 4D chest CT from the DIR-Lab dataset [4] for the inter-patient chest CT registration task. The dataset consists of ten 4D chest CT scans, which encompass a full breathing cycle in ten timepoints. The axial isotropic resolutions of each scan range from 256×256 to 512×512 (0.97mm to 1.16mm per voxel, with a slice thickness and increment of 2.5mm. We follow [30] to perform leave-one-out cross-validation during evaluation. We use U-net [12] to delineate the lung lobe of each scan and use the segmentation to define the region of interest in this task. The learning-based method was trained with intra-patient registration by taking random timepoints per patient as fixed and moving images. We leverage the 300 manually identified anatomical landmarks annotated at maximum inspiration and maximum expiration for each case to quantify the registration accuracy.

Implementation Our proposed method and the other baseline methods are implemented with PyTorch 1.9 and deployed on the same machine equipped with an Nvidia RTX4090 GPU and an Intel Core (i7-13700) CPU. We build our method on top of conditional LapIRN [18] framework. For the multi-resolution optimization strategy, we set the iteration of each level η to $[100, 60, 30]$ and $[100, 60, 60]$ for brain atlas and chest 4DCT registration, respectively. λ is set to 0.4 for both the training and inference phases. We adopt Adam optimizer with a fixed learning rate of $1e^{-4}$ in both training and inference phases. We train all the deep-learning methods from scratch for 150,000 iterations and select the model with the highest DSC (lowest TRE) in the validation set.

Measurement In brain atlas registration, we register each scan in the test set to an atlas, propagate the anatomical segmentation map of the moving scan using the resulting deformation field, and measure the overlap of the segmentation maps using the Dice similarity coefficient (DSC). We also quantify the robustness of each method by measuring the 30% lowest DSC (DSC30) of all cases. In chest 4DCT registration, we register scans of the maximum expiration phase to that of the maximum inspiration phase and measure the target registration error (TRE) using the manually identified anatomical landmarks. The standard deviation of Jacobian determinant on the deformation field (std($|J_\Phi|$)) is measured, which quantifies the smoothness and local orientation consistency of the deformation field. Furthermore, we measure the average registration time per case (T_{test}).

Baseline Methods We compare our method with two conventional methods (denoted as NiftyReg [22] and SyN [1]), and three state-of-the-art learning-based methods (denoted as MPR [14], ULAE [26] and LapIRN [18,19]) for brain atlas registration. NiftyReg and SyN use the multilevel iterative optimization strategy and do not suffer from the distribution shifts issue. For MPR and ULAE, we adopt the official implementation maintained by the authors and use the best hyper-parameters reported in the paper [14,26]. Besides, we compare the baseline LapIRN with the Adam-based instance optimization approach [27](denoted as IO). Empirically, the performance gains of IO saturated within 80-150 iterations. We set the number of iterations for IO to [50, 200] and 200 for brain atlas registration and chest 4DCT registration tasks, respectively. NIO with the proposed multiresolution strategy is denoted as NIO_m. We follow [27] to induce additional smoothness of the displacement field by adding a B-spline deformation model and affine augmentation to LapIRN, IO and NIO_m during the inference phase for chest 4DCT registration.

Results and Discussions Table 1 presents comprehensive results of the brain atlas registration. Fig. 2 illustrates the qualitative results of the cross-dataset brain atlas registration. Two out of three learning-based methods (MPR and ULAE) fail spectacularly in the cross-dataset brain atlas registration tasks and achieve consistently inferior registration accuracy and robustness to the conventional methods (NiftyReg and SyN) among three tasks in brain atlas registra-

Table 2. Mean ± standard deviation of the target registration error in millimetre (mm) determined on DIR-Lab 4D-CT dataset. *: method trained on external data.

Scan	Initial	DLIR [30]	B-splines [5]	LungRegNet*[8]	LapIRN [19]	IO ($\eta = 200$)	NIO_m (ours)
Case 1	3.89 ± 2.78	1.27 ± 1.16	1.2 ± 0.6	**0.98** ± 0.54	1.05 ± 0.47	1.00 ± 0.46	0.99 ± 0.46
Case 2	4.34 ± 3.90	1.20 ± 1.12	1.1 ± 0.6	0.98 ± 0.52	1.04 ± 0.52	0.97 ± 0.47	**0.97** ± 0.47
Case 3	6.94 ± 4.05	1.48 ± 1.26	1.6 ± 0.9	1.14 ± 0.64	1.18 ± 0.62	1.11 ± 0.61	**1.10** ± 0.61
Case 4	9.83 ± 4.85	2.09 ± 1.93	1.6 ± 1.1	1.39 ± 0.99	1.40 ± 0.99	1.35 ± 0.98	**1.33** ± 0.95
Case 5	7.48 ± 5.50	1.95 ± 2.10	2.0 ± 1.6	1.43 ± 1.31	1.41 ± 1.21	**1.34** ± 1.20	**1.34** ± 1.21
Case 6	10.89 ± 6.96	5.16 ± 7.09	1.7 ± 1.0	2.26 ± 2.93	1.36 ± 0.77	1.20 ± 0.67	**1.16** ± 0.66
Case 7	11.03 ± 7.42	3.05 ± 3.01	1.9 ± 1.2	1.42 ± 1.16	1.31 ± 0.68	1.20 ± 0.64	**1.16** ± 0.62
Case 8	14.99 ± 9.00	6.48 ± 5.37	2.2 ± 2.3	3.13 ± 3.77	1.73 ± 2.31	1.59 ± 2.20	**1.19** ± 0.96
Case 9	7.92 ± 3.97	2.10 ± 1.66	1.6 ± 0.9	1.27 ± 0.94	1.34 ± 0.73	1.20 ± 0.69	**1.16** ± 0.65
Case 10	7.30 ± 6.34	2.09 ± 2.24	1.7 ± 1.2	1.93 ± 3.06	1.25 ± 0.77	1.14 ± 0.74	**1.08** ± 0.58
Mean	8.46 ± 6.58	2.64 ± 4.32	1.66 ± 1.14	1.59 ± 1.58	1.31 ± 0.19	1.21 ± 0.87	**1.15** ± 0.71

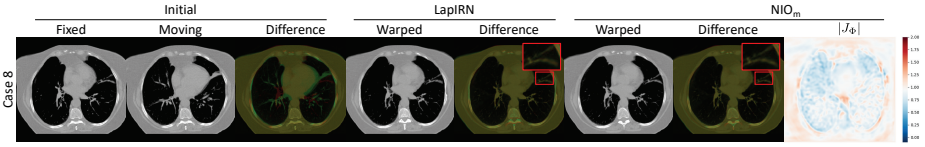

Fig. 3. Example axial slices from the inspiration-expiration registration results. From left to right: the fixed image, moving image, and the results from LapIRN and NIO$_m$. The difference map overlays the fixed (in red) and moving/warped images (in green).

tion, suggesting the distribution shift can significantly degrade the solution's quality of learning-based methods. Interestingly, the registration performance of LapIRN diverged from the other learning-based methods and achieved comparable results to the conventional methods under distribution shifts. Comparing NIO$_m$ to LapIRN, our method significantly improves the registration accuracy (+1.5% to +3.0% gain) and robustness (+2.3% to +6.2% gain) of LapIRN in brain atlas registration, reaching state-of-the-art results among three tasks with severe distribution shifts at the cost of registration time. As shown in Fig. 2, our method can drastically improve the registration result for extreme cases with large inter-variation in anatomical structures presented in fixed and moving images. Comparing NIO to the IO, NIO achieves consistently superior registration results with less number of iterations, suggesting the effectiveness of NIO and capable of avoiding sub-optimal solutions in IO. The results of each case on the DIR-Lab 4DCT dataset are shown in Table 2. The results demonstrate that our method not only significantly improves the registration accuracy (an average of 12% drop in TRE compared to the LapIRN), but also corrects the outlier with large initial registration errors. For instance, the average TRE in case 8 of the baseline decreased from 1.73mm to 1.19mm using our method. Our method can also correct subtle misalignment inside the lung lobe, as highlighted in Fig. 3. The fraction of landmark pairs with < 1.2mm error of LapIRN improves from 55% to 64% with NIO. The results suggested that our method enables learning-based methods to achieve state-of-the-art registration results even with insufficient training data (20 and 90 scans in LPBA and DIR-Lab 4DCT dataset, respectively) and distribution shifts.

4 Conclusion

We have presented a novel neural instance optimization to improve the registration accuracy and robustness of learning-based deformable image registration methods under distribution shifts. Our method naturally leverages the inductive bias of the convolutional neural network, the prior knowledge learned from the training domain and the multi-resolution optimization strategy to optimize the solution during the inference phase. Extensive experiments on brain atlas and chest 4DCT registration have been carried out, demonstrating that our proposed method achieves state-of-the-art results even with severe distribution shifts or limited training data.

References

1. 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)
2. Bajcsy, R., Kovačič, S.: Multiresolution elastic matching. Computer vision, graphics, and image processing **46**(1), 1–21 (1989)
3. 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)
4. Castillo, R., et al.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Physics in Medicine & Biology **54**(7), 1849 (2009)
5. Delmon, V., et al.: Registration of sliding objects using direction dependent b-splines decomposition. Physics in Medicine & Biology **58**(5), 1303 (2013)
6. Eisenmann, M., et al.: Biomedical image analysis competitions: The state of current participation practice. arXiv preprint arXiv:2212.08568 (2022)
7. Falta, F., Hansen, L., Heinrich, M.P.: Learning iterative optimisation for deformable image registration of lung ct with recurrent convolutional networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI. pp. 301–309. Springer (2022)
8. Fu, Y., Lei, Y., Wang, T., Higgins, K., Bradley, J.D., Curran, W.J., Liu, T., Yang, X.: Lungregnet: an unsupervised deformable image registration method for 4d-ct lung. Medical physics **47**(4), 1763–1774 (2020)
9. Heinrich, M.P., Hansen, L.: Voxelmorph++ going beyond the cranial vault with keypoint supervision and multi-channel instance optimisation. In: Biomedical Image Registration: 10th International Workshop, WBIR 2022, Munich, Germany, July 10–12, 2022, Proceedings. pp. 85–95. Springer (2022)
10. Hering, A., Hansen, L., Mok, T.C., et al.: Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. IEEE Transactions on Medical Imaging (2022)
11. Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V.: Synthmorph: learning contrast-invariant registration without acquired images. IEEE transactions on medical imaging **41**(3), 543–558 (2021)
12. Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., Langs, G.: Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. European Radiology Experimental **4**(1), 1–13 (2020)
13. Hoopes, A., Hoffmann, M., Fischl, B., Guttag, J., Dalca, A.V.: Hypermorph: Amortized hyperparameter learning for image registration. In: International Conference on Information Processing in Medical Imaging (2021)
14. Liu, R., Li, Z., , et al.: Learning deformable image registration from optimization: Perspective, modules, bilevel training and beyond. IEEE Transactions on Pattern Analysis Machine Intelligence **44**(11), 7688–7704 (2022)
15. 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. Journal of cognitive neuroscience **19**(9), 1498–1507 (2007)
16. Marcus, D.S., Wang, T.H., et al.: Oasis brains - open access series of imaging studies. `https://www.oasis-brains.org/`, accessed: 01.03.2021

17. Mok, T.C., Chung, A.: Fast symmetric diffeomorphic image registration with convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4644–4653 (2020)
18. Mok, T.C., Chung, A.: Conditional deformable image registration with convolutional neural network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 35–45. Springer (2021)
19. Mok, T.C., Chung, A.C.: Large deformation diffeomorphic image registration with laplacian pyramid networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 211–221. Springer (2020)
20. Mok, T.C., Chung, A.C.: Unsupervised deformable image registration with absent correspondences in pre-operative and post-recurrence brain tumor mri scans. In: MICCAI 2022. pp. 25–35. Springer (2022)
21. Papenberg, N., Bruhn, A., Brox, T., Didas, S., Weickert, J.: Highly accurate optic flow computation with theoretically justified warping. International Journal of Computer Vision **67**, 141–158 (2006)
22. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: application to breast mr images. IEEE transactions on medical imaging **18**(8), 712–721 (1999)
23. Ruthotto, L., Modersitzki, J.: Non-linear image registration. Handbook of Mathematical Methods in Imaging: Volume 1, Second Edition pp. 2005–2051 (2015)
24. Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W.: Construction of a 3d probabilistic atlas of human cortical structures. Neuroimage **39**(3), 1064–1080 (2008)
25. Shattuck, D.W., Mirza, M., et al.: Lpba40 atlases download. `https://resource.loni.usc.edu/resources/atlases-downloads/`, accessed: 01.03.2021
26. Shu, Y., Wang, H., Xiao, B., Bi, X., Li, W.: Medical image registration based on uncoupled learning and accumulative enhancement. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV. pp. 3–13. Springer (2021)
27. Siebert, H., Hansen, L., Heinrich, M.P.: Fast 3d registration with accurate optimisation and little learning for learn2reg 2021. In: Biomedical Image Registration, Domain Generalisation and Out-of-Distribution Analysis: MICCAI 2021 Challenges: MIDOG 2021, MOOD 2021, and Learn2Reg 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27–October 1, 2021, Proceedings, pp. 174–179. Springer (2022)
28. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. pp. 402–419. Springer (2020)
29. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: CVPR 2011. pp. 1521–1528. IEEE (2011)
30. 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)
31. Zhu, W., Huang, Y., Xu, D., Qian, Z., Fan, W., Xie, X.: Test-time training for deformable multi-scale image registration. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 13618–13625. IEEE (2021)