Multi-scale Neural ODEs for 3D Medical Image Registration

Junshen Xu[1(✉)], Eric Z. Chen[2], Xiao Chen[2], Terrence Chen[2], and Shanhui Sun[2]

[1] Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, USA
`junshen@mit.edu`
[2] United Imaging Intelligence, Cambridge, MA, USA
{`zhang.chen,xiao.chen01,terrence.chen,shanhui.sun`}`@united-imaging.com`

Abstract. Image registration plays an important role in medical image analysis. Conventional optimization based methods provide an accurate estimation due to the iterative process at the cost of expensive computation. Deep learning methods such as learn-to-map are much faster but either iterative or coarse-to-fine approach is required to improve accuracy for handling large motions. In this work, we proposed to learn a registration optimizer via a multi-scale neural ODE model. The inference consists of iterative gradient updates similar to a conventional gradient descent optimizer but in a much faster way, because the neural ODE learns from the training data to adapt the gradient efficiently at each iteration. Furthermore, we proposed to learn a modal-independent similarity metric to address image appearance variations across different image contrasts. We performed evaluations through extensive experiments in the context of multi-contrast 3D MR images from both public and private data sources and demonstrate the superior performance of our proposed methods.

Keywords: Multi-modal image registration · Neural ordinary differential equations · Disentangled representation · Self-supervised learning

1 Introduction

Image registration is an essential step in many tasks such as motion correction and atlas-based image segmentation. It is indispensable in many clinical applications such as surgical planning [24] and radiogenomics analysis [16], where 3D

---

images are commonly used. Conventional image registration methods solve an optimization problem by minimizing the dissimilarity between the transformed image and the target image. Although the conventional methods often achieve high accuracy, the slow process and expensive computation due to the use of iterative non-linear optimization algorithms hinder their clinical translation.

Recently deep learning based approaches have been proposed for image registration [5,8,25,30], which learn to map from image or feature space to a spatial transformation space using neural networks trained on large datasets. Since the registration during the inference is just one forward pass of the network, the deep learning methods are intrinsically faster than the conventional methods. Supervised methods [8,25,31] require ground truth transformations for training, which are typically difficult to obtain in clinical practice especially for deformable registration. Unsupervised methods, such as DIRNet [30] and VoxelMorph [5], directly regress deformation fields by minimizing dissimilarity between input and target images. However, for large motion, the learn-to-map based methods often do not perform well [27]. Multi-stage methods are proposed [29,32], where several networks are cascaded to gradually refine the estimated transformation. Deep reinforcement learning is also applied to image registration, especially for rigid registration [14,20,28]. In terms of non-rigid registration, Krebs *et al.* [18] proposed an agent-based method for prostate MR registration, which is limited to low dimensional B-spline.

Multi-modal images are frequently used in clinic. The main challenge of multi-modal image registration is to find a proper dissimilarity cost function to distinguish motions from contrast changes. Some classical methods use mutual information (MI) [11,21] and modality-independent neighborhood descriptor (MIND) [13]. Some learning-based methods convert the problem to mono-modal registration by image-to-image translation [1,7], which are prone to synthetic artifacts. UMDIR [23] measured the difference between multi-modal images in a feature space; however, only 2D deformable image registration between two modalities was addressed.

Recently, neural ordinary differential equations (ODEs) are proposed to represent more complex dynamics over classical ODEs. Compared to the common deep learning models such as ResNet and UNet, neural ODE models are more memory and parameter efficient and provide the benefits of adaptive computation [10], which are potentially suitable for medical applications [9]. The optimization dynamics are also inherently continuous. These merits motivate us to learn the optimization in medical image registration using neural ODEs. To the best of our knowledge, our study is the first work to apply neural ODEs to image registrations.

The contributions of our proposed methods are summarized as follows: 1) We proposed a new direction of modeling image registration optimizer as a continuous optimization dynamics via neural ODEs. 2) We introduced multi-scale architecture to neural ODEs to reduce searching space by performing registration iteratively on different scales. 3) Our proposed method is a general learn-to-learn image registration framework and is not limited to specific transformations. 4) Our framework can handle multiple contrasts with a single trained network attribute to proposed contrast-independent similarity metric for $n(\geq 2)$ modalities.

2   Method

Let x_{mov} and x_{fix} denote the moving and fixed images, respectively, and let ϕ_θ be the transformation field between two images parameterized by θ. The image registration optimization problem can be written as:

$$\hat{\theta} = \arg\min_\theta \mathcal{L}(x_{\mathrm{mov}} \circ \phi_\theta, x_{\mathrm{fix}}) + \mathcal{R}(\theta), \tag{1}$$
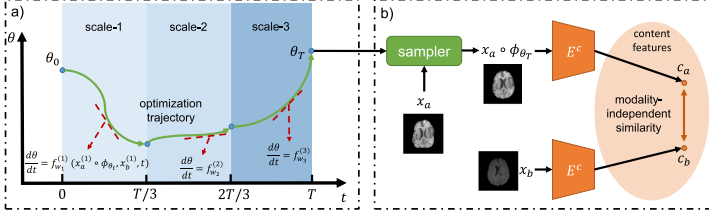

Fig. 1. Overview of our proposed method: a) Image registration optimization is modeled as a continuous optimization dynamics via neural ODEs, b) A modality-independent similarity metric is realized via a pretrained content encoder.

where $x_{\mathrm{mov}} \circ \phi_\theta$ is the transformed image, $\mathcal{L}$ is the dissimilarity cost function, and $\mathcal{R}$ is the regularization term. The form of θ is determined by the types of transformation.

We propose a novel multi-modal image registration method by learning an image registration optimizer via neural ODEs, and deriving loss function from a pretrained image content encoder. Figure 1 illustrates an overview of our method.

2.1   Learn Registration Optimizer via Neural ODEs

The optimization problem for image registration in Eq. 1 can be solved via gradient descent optimizer. Thus, the update of θ at each step is $\theta_{t+1} = \theta_t - \eta_t \partial(\mathcal{L} + \mathcal{R})/\partial\theta_t \triangleq \theta_t + f(\theta_t, t)$, where t represents t^{th} step and η_t is the step size. If the time difference is small enough, the difference equation can be re-written as an ODE function, $\frac{d\theta_t}{dt} = f(\theta_t, t), t \in [0, T]$. Given the initial parameter θ_0, the final parameter θ_T is the solution to this ODE initial value problem. Some conventional methods such as LDDMM [6] and SyN [3] describe the evolution of transformation as a differential equation. These methods solve differential equations where system dynamics are described by predefined functions, which is less flexible. Neural ODEs use trainable network to replace the predefined function, which can be considered as learning an optimizer to compute the gradient update. θ_T is the inference of the neural ODE model. We further assume that for two 3D images x_a and x_b, the evolution of θ_t follows a neural ODE:

$$\frac{d\theta_t}{dt} = f_w(x_a \circ \phi_{\theta_t}, x_b, t), t \in [0, T], \tag{2}$$

where f_w is a neural network with parameter w which takes the current warped image $x_a \circ \phi_{\theta_t}$, the target image x_b and the time variable t as inputs. Therefore, the final output at time T can be computed by integrating function f over time interval $[0, T]$. In practice, the integral is evaluated by an ODE solver, such as Runge–Kutta methods and adaptive step size solver. To train the neural ODE model, we adopt the adaptive checkpoint adjoint method [33].

Multi-scale ODE network: Empirically, we found that solving neural ODE for image registration problem requires an extensive number of function evaluations (NFE) by ODE solver, which leads to prolonged training and inference time. To address this problem, we propose a multi-scale ODE network (MS-ODENet) which performs registration by solving ODEs at different resolutions. Specifically, let $\{x_i^l\}_{l=1}^L (i = a, b)$ be an image pyramid with L different resolutions, where $x_i^{(L)} = x_i$, and $x_i^{(l-1)}$ is generated by down sampling $x_i^{(l)}$ with a factor of 2 on all 3 axes. We divide the whole time interval $[0, T]$ into L congruent segments. In each segment, we solve the ODE in Eq. 3 at the corresponding resolution as shown in Fig. 2a and b.

$$\frac{d\theta_t}{dt} = f_{w_l}^{(l)}(x_a^{(l)} \circ \phi_{\theta_t}, x_b^{(l)}, t), t \in [\frac{l-1}{L}T, \frac{l}{L}T], \tag{3}$$

where $f_{w_l}^{(l)}$ is the network at the l-th scale. The output parameter θ_T is the integral over all scales. The benefits of this design are two-fold: 1) The time cost for function evaluations is much smaller at low resolutions, allowing a larger number of steps to reach the desired accuracy. 2) The searching space for image registration is largely reduced and therefore the convergence of the ODE network is faster, which also makes the model less sensitive to local optimal.

Loss Functions: The proposed registration network is trained in an unsupervised manner by minimizing a loss function similar to Eq. 1. The utilized similarity metric $\mathcal{L}_{\mathrm{sim}}$ (Fig. 1b) between two images (x_a and x_b) is defined in Eq. 4:

$$\mathcal{L}_{\mathrm{sim}} = \mathbb{E}_{x_a, x_b, a, b}||E_{3D}^c(x_a \circ \phi_{\theta_T}) - E_{3D}^c(x_b)||_2^2, \tag{4}$$

where $E_{3D}^c(\cdot)$ is a 3D content feature extractor. The 3D content feature is composed of 2D content features generated from N randomly selected 2D images of different axes from the image utilizing a 2D feature extractor E^c (Sect. 2.2). In addition, we perform a random perturbation to a given image x_a such that $\tilde{x}_a = x_a \circ \phi_{\tilde{\theta}}$, where $\tilde{\theta}$ is randomly sampled from a distribution of parameter $\mathbb{P}_\theta$. The pair x_a and $\tilde{x}_a$ are fed to the registration network resulting transformation parameters $\tilde{\theta}_T = $ MS-ODENet$(\theta_0, x_a, \tilde{x}_a, L, T)$. Our network can align two images from the same modality so that we expect $\tilde{\theta}_T$ approximating to the purturbation $\tilde{\theta}$ by minimizing L2 loss: $\mathcal{L}_{\mathrm{self}} = \mathbb{E}_{x_a, a, \tilde{\theta}}||\tilde{\theta} - \tilde{\theta}_T||_2^2$. The total loss function is summarized as $\mathcal{L} = \mathcal{L}_{\mathrm{sim}} + \lambda_{\mathrm{self}}\mathcal{L}_{\mathrm{self}} + \lambda_{\mathrm{reg}}\mathcal{L}_{\mathrm{reg}}$, where $\mathcal{L}_{\mathrm{reg}} = \mathbb{E}_{x_a, x_b, a, b}||\nabla\phi_{\theta_T}||_2^2$ is the regularization term enforcing a smooth motion field in deformable registration. λ_{self} and λ_{reg} are weighting coefficients.

2.2 Pretrained Feature Extraction Network

Different contrast images from the same patient share the same anatomical structures (content features) but have different style features. This inspires us decomposing the image into content features and style features in the latent space. The realization of this feature domain disentanglement is extended from the diverse image-to-image translation framework [19]. We trained content encoder E^c, style encoder E^s and generator G to perform image translation from one contrast to another. Note that only E^c is used in our registration task. Figure 2c illustrates an overview of our feature extraction framework utilizing image-to-image translation. Suppose two different groups $\mathcal{X}_a$ and $\mathcal{X}_b$ sampled from M modality groups ($\{\mathcal{X}_i\}_{i=1}^M$). Given x_a and x_b two image samples in these two groups, we perform cross-modality translation as follows: (a) Extract the content and style features, i.e., $s_a = E^s(x_a, a)$, $c_a = E^c(x_a)$, $s_b = E^s(x_b, b)$, $c_b = E^c(x_b)$. (b) Swap style features and generate translated images, i.e., $\hat{x}_{ab} = G(c_a, s_b, b)$, $\hat{x}_{ba} = G(c_b, s_a, a)$. (c) Encode the translated images to reconstruct content and style features of the original images, $\hat{s}_a = E^s(\hat{x}_{ba}, a)$, $\hat{c}_a = E^c(\hat{x}_{ab})$, $\hat{s}_b = E^s(\hat{x}_{ab}, b)$, $\hat{c}_b = E^c(\hat{x}_{ba})$. (d) Swap the style features again to generate the original images, $\hat{x}_{aba} = G(\hat{c}_a, \hat{s}_a, a)$, and $\hat{x}_{bab} = G(\hat{c}_b, \hat{s}_b, b)$. We trained the network using adversial loss in [19]. Moreover, we expect that $\hat{x}_{aba}$ and $\hat{x}_{bab}$ are consistent to the original images x_a and x_b respectively. We use L1 loss $\mathcal{L}_{\mathrm{cyc}}$ to enforce this cycle consistency. We also introduce the feature reconstruction loss for content/style features, $\mathcal{L}_{\mathrm{rec}}^c/\mathcal{L}_{\mathrm{rec}}^s$, which are the L1 loss between $c_a, c_b/s_a, s_b$ and $\hat{c}_a, \hat{c}_b/\hat{s}_a, \hat{s}_b$. Similarly, we define mono-modal reconstruction loss $\mathcal{L}_{\mathrm{rec}}^x$ as the L1 loss between the reconstructed images $\hat{x}_a = G(c_a, s_a, a)$, $\hat{x}_b = G(c_b, s_b, b)$ and the original images x_a, x_b. Since training is not trivial in 3D due to high dimensionality, we use a 2D multi-channel (adjacent slices in 3D volume) network.
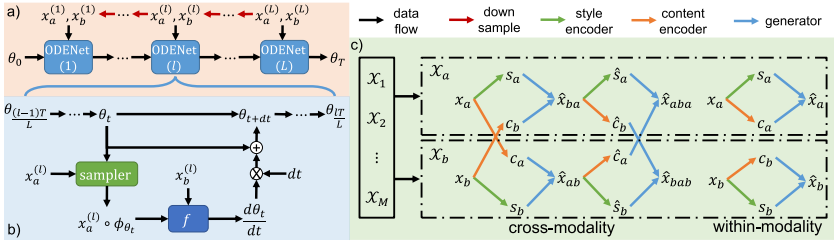

Fig. 2. Implementation of our proposed framework: a) Overview of multi-scale ODENet: several cascaded ODENets in different resolutions, b) a computational graph of a single scale ODENet with Euler's solver, c) overview of the utilized image translation towards metric learning.

3   Experiments and Results

3.1   Experiment Setup

Dataset: Currently, there is a lack of large scale medical image dataset dedicated to multi-modal image registration. We use a public dataset and a separate acquired volunteer dataset. The public dataset is the brain tumor segmentation (BraTS) 2020 dataset [22] which consists of multi-modal 3D brain MRI with four distinctive contrasts (T1, T2, T2-FLAIR, and T1Gd) from 494 subjects with glioblastomas, resulting in various multi-modal image registration tasks (12 pairs per subject). Besides, tumor masks provide a clinically meaningful evaluation metric for registration. For each subject, all modalities were normalized into 3D volumes with a size of $240 \times 240 \times 160$ and a resolution of $1 \times 1 \times 1$ mm^3. The dataset is split into 444 subjects (5,328 pairs) for training and validation and 50 subjects (600 pairs) for testing. Since different modalities have been registered, we simulate motions by applying random rigid transformation, random control point deformation, or both. Note that the simulated transformation fields are only used for evaluation, not for training. We also acquired additional multi-modal 3D MR brain data on 25 volunteers using a special MR technique that can acquire multiple contrasts simultaneously. This private dataset is used only for testing the generalizability of the proposed methods. The motion is simulated as described above. Images were preprocessed to remove skull using DeepBrain[1] and resized as the public data.

Table 1. Quantitative results for image registration on BraTS, where R, D, and B represent rigid, dense and B-spline parameterization in MS-ODENet. The mean (standard deviation) are reported.

Transformations	Methods	Dice/%	RMSE(x) $\times 10^{-2}$	RMSE (ϕ)/mm	Time/s
Rigid	ANTs	63.0 (40.5)	8.34 (7.64)	7.28 (6.66)	17.17 (4.05)
	ANTs+I2I	60.6 (40.8)	8.83 (7.81)	7.54 (6.78)	30.69 (4.70)
	MS-ODENet (R)	**90.6 (17.2)**	**3.89 (2.95)**	**3.57 (5.18)**	**0.55 (0.08)**
Deformable	ANTs	81.9 (8.8)	6.31 (1.98)	1.21 (0.37)	55.35 (7.07)
	ANTs+I2I	81.1 (8.3)	6.34 (1.82)	1.06 (2.35)	68.47 (6.18)
	VM	79.4 (8.7)	8.81 (3.03)	1.61 (0.69)	**0.24 (0.05)**
	VM+I2I	80.1 (7.8)	8.52 (2.30)	1.26 (0.31)	0.34 (0.06)
	MS-ODENet (D)	81.6 (8.1)	6.63 (2.22)	1.11 (0.18)	1.13 (0.10)
	MS-ODENet (B)	**83.0 (8.7)**	**6.17 (2.43)**	**0.99 (0.32)**	0.31 (0.06)
Rigid + deformable	ANTs	73.6 (22.2)	6.79 (2.51)	2.99 (2.95)	70.87 (7.97)
	ANTs+I2I	71.1 (22.1)	6.97 (2.26)	2.83 (2.80)	87.25 (12.2)
	RCN	64.9 (10.1)	8.74 (3.21)	5.48 (1.97)	2.49 (0.31)
	VM*	78.1 (9.8)	8.62 (3.02)	2.92 (2.31)	0.60 (0.23)
	VM*+I2I	78.4 (9.2)	7.07 (1.96)	2.04 (1.38)	1.01 (0.38)
	MS-ODENet (R+D)	79.6 (9.1)	6.70 (2.28)	1.82 (1.22)	1.39 (0.12)
	MS-ODENet (R+B)	**81.1 (9.9)**	**6.28 (2.38)**	**1.52 (1.02)**	**0.56 (0.09)**

---

[1] https://github.com/iitzco/deepbrain.

Registration Networks: We evaluated our methods on rigid, deformable and a hybrid of rigid and deformable motions. For 3D rigid motion, f includes convolution layers followed by fully connected layers. The output size of f is the same as the degree of freedom in the transformation. For deformable motion, fully convolutional network is used. One variation of the deformable registration is kernel based method (B-spline). Parameter θ is the grid of control points which can also be regarded as an image. For dense motion, the UNet as in [26] is used for voxel-wise estimation. For hybrid motion, we cascade the rigid and deformable MS-ODENets sequentially.

Feature Extraction Network: The backbone network is similar to [15]. The content encoder is a fully convolutional network while the style encoder has convolution layers followed by a global average pooling and a fully connected layer, which forces the network to extracts global style features. The generator is a CNN with deconvolution layers to generate images of the original size. Besides, to condition on modality $\mathcal{X}_i$ for style encoder and image generator, the modality code i is converted into a one-hot vector concatenated with the input tensor along the channel dimension.

Implementation Details: We follow the protocol in [2] for training GAN with number of slices set to 5. For registration, we set $\lambda_{\text{self}} = 1$, and let λ_{reg} be 10 and 2 for dense and B-spline respectively. We use $L = T = 3$ for MS-ODENet. Let $F(dt)$ be Euler's method with fixed step size dt and $A(\epsilon)$ be adaptive Heun's method with tolerance of error ϵ. We use adaptive solvers when the search space is small (Rigid: $A(10^{-3})$-$A(10^{-3})$-$F(0.1)$) to avoid extensive NFE and use fixed step size solver for large search space (B-spline: $F(0.2)$-$F(0.2)$-$F(0.25)$, Dense: $F(0.2)$-$F(0.2)$-$F(0.5)$). All networks are trained using Adam [17] optimizer with a learning rate of 10^{-4} on an NVIDIA Tesla V100 GPU.
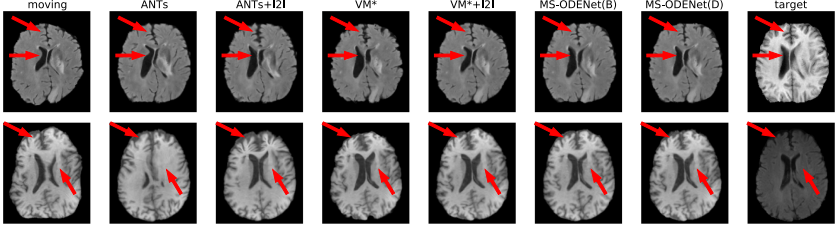

Fig. 3. Examples of compared methods on deformable registration. Top row: T2-FLAIR to T1 registration. Bottom row: T1 to T2-FLAIR. Red-arrows highlight example different areas.

Evaluation Metrics: For evaluation, Dice scores [12] are computed over tumor masks. With available synthetic transformation fields and ground truth images, root mean square errors are calculated and denoted as RMSE (ϕ) and RMSE (x), respectively. The metrics are averaged over all pairs of test data.

3.2   Results

Table 1 shows the quantitative results for rigid, deformable and hybrid registration (rigid+deformable). **Rigid:** Random rotation and translation were synthesized along all three axes and were sampled from $U(-75°, 75°)$ and $U(-20, 20)$mm, respectively. We used the rigid registration method with MI metric in Advanced Normalization Tools (ANTs) [4] as the baseline. We also used the trained GAN to perform image translation followed by mono-modal image registration with ANTs (ANTs+I2I). **Deformable:** We made synthetic image pairs through elastic transformations by perturbing B-spline control points with noise from $\mathcal{N}(0, (8mm)^2)$ on three axes. For comparison, we use SyN [3] in ANTs, VoxelMorph (VM) [5] with MI metric, and also their variants with image translation (ANTs+I2I and VM+I2I). For MS-ODENet, we use two different parameterizations, namely dense deformation (D) and B-spline (B). Figure 3 shows example results. **Rigid + deformable:** Random rotations, translations and control point noise are from $U(-40°, 40°)$, $U(-10, 10)$mm, and $\mathcal{N}(0, (8mm)^2)$, respectively. We compared our method to SyN, RCN [32] and VM. RCN is an iterative deep learning method consisting of an affine network and n deformable networks. We use MI metric for training RCN and set $n = 2$ due to memory limit. Since VM was proposed to solve deformable registration, we performed a rigid registration using our rigid MS-ODENet prior to applying VM. This approach is denoted as VM*.

Ablation: Figure 4(a) summarizes the results that evaluate models with no learned content loss (replaced by MI), no multi-scale ODE, no self-supervision or no multi-slice GAN respectively and compare them with the full model on rigid + B-spline deformable registration. To investigate the necessity of iterative registration for large motions, we conducted a rigid registration experiment with $L = T = 1$ and various numbers of steps in ODE solver. Figure 4(b) shows the corresponding results.
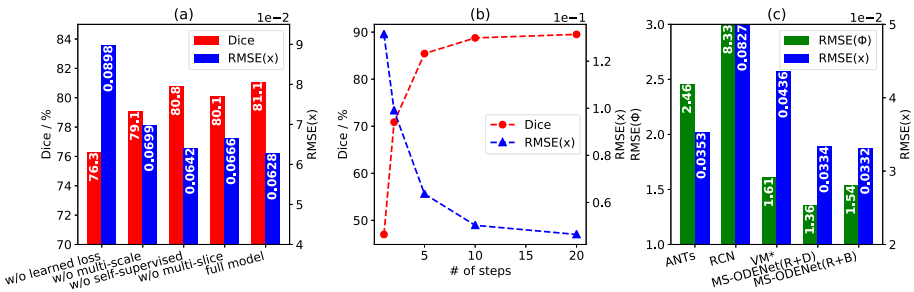

Fig. 4. a) Quantitative results for the ablation study. b) Performance on large motion for neural ODE models with different number of solver steps. c) Results for the generalizability study.

Generalization: We performed the rigid+deformable test on our private dataset. Figure 4(c) shows bar chart among compared methods.

4   Discussions and Conclusions

Table 1 shows that our proposed MS-ODENet outperforms classical methods under various transformations. In all experiments, MS-ODENet is much faster than ANTs due to the fast inference of neural networks and the smaller number of iterations needed in neural ODEs. In rigid registration (Fig. 1), MS-ODENet greatly outperforms ANTs with or without the domain translation. Classical methods only consider local gradient information and tend to get stuck at local minima. The MS-ODENet learns the optimization process via neural ODEs, which utilizes not only the current local information but also the experiences learned from the dataset, and is thus more likely to reach the global minimum. In deformable registration, our proposed methods achieve similar or better accuracy compared with classical methods. For rigid+deformable registration, ANTs suffers greatly from the additional rigid transformation, indicating that traditional optimization-based methods rely heavily on the initialization of parameters.

Our methods also outperform other deep learning methods (Table 1) on deformable registration. The iterative updates in MS-ODENet improve registration accuracy progressively (Fig. 4(c)). When the step number is one, the MS-ODENet is equivalent to a conventional deep learning model. Note that all the ablated methods (Fig. 4(a)) outperform the other deep learning methods. Furthermore, with the adjoint method, the training of our neural ODE model does not backpropagate through the operations of the solver [10,33] and thus is more memory efficient than traditional deep learning models. Unlike other coarse-to-fine methods that need to be trained in separate stages [29], our multi-scale neural ODE model can be trained end-to-end. In the generalization study, the proposed MS-ODENets show consistent improvement over the other methods. The RCN does not employ iterative networks for affine transformation and therefore has poor generalizability for large transformation.

In this work, we present a new framework for 3D multi-modal image registration. We formulate the optimization in conventional registration methods as a continuous process and learn the optimizer via a neural ODE. Furthermore, for efficient learning and inference, we propose a multi-scale architecture to narrow the searching space from coarse to fine image resolutions. In addition, we employ an image-to-image translation GAN to learn a modality-independent metric between images from different modalities. Experiment results show that our proposed framework is superior to other compared methods. For future work, we will extend our framework to other types of medical registration such as 3D-2D image registration.

References

1. Arar, M., Ginger, Y., Danon, D., Bermano, A.H., Cohen-Or, D.: Unsupervised multi-modal image registration via geometry preserving image-to-image translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13410–13419 (2020)

2. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. arXiv preprint arXiv:1701.07875 (2017)

3. 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. Med. Image Anal. **12**(1), 26–41 (2008)

4. Avants, B.B., Tustison, N., Song, G.: Advanced normalization tools (ANTS). Insight J. **2**(365), 1–35 (2009)

5. 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)

6. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. **61**(2), 139–157 (2005)

7. Cao, X., Yang, J., Gao, Y., Guo, Y., Wu, G., Shen, D.: Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis. Med. Image Anal. **41**, 18–31 (2017)

8. Cao, X., et al.: Deformable image registration based on similarity-steered CNN regression. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 300–308. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_35

9. Chen, E.Z., Chen, T., Sun, S.: MRI image reconstruction via learning optimization using neural ODEs. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 83–93. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_9

10. Chen, R.T., Rubanova, Y., Bettencourt, J., Duvenaud, D.K.: Neural ordinary differential equations. In: Advances in Neural Information Processing Systems, pp. 6571–6583 (2018)

11. De Nigris, D., Mercier, L., Del Maestro, R., Louis Collins, D., Arbel, T.: Hierarchical multimodal image registration based on adaptive local mutual information. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 643–651. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15745-5_79

12. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)

13. Heinrich, M.P., et al.: MIND: modality independent neighbourhood descriptor for multi-modal deformable registration. Med. Image Anal. **16**(7), 1423–1435 (2012)

14. Hu, J., et al.: End-to-end multimodal image registration via reinforcement learning. Med. Image Anal. **68**, 101878 (2020)

15. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 172–189 (2018)

16. Incoronato, M., et al.: Radiogenomic analysis of oncological data: a technical survey. Int. J. Mol. Sci. **18**(4), 805 (2017)

17. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

18. Krebs, J., et al.: Robust non-rigid registration through agent-based action learning. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 344–352. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_40

19. Lee, H.Y., et al.: DRIT++: diverse image-to-image translation via disentangled representations. Int. J. Comput. Vis. **128**(10), 2402–2417 (2020). https://doi.org/10.1007/s11263-019-01284-z

20. Ma, K., et al.: Multimodal image registration with deep context reinforcement learning. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 240–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_28

21. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imaging **16**(2), 187–198 (1997)

22. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2014)

23. Qin, C., Shi, B., Liao, R., Mansi, T., Rueckert, D., Kamen, A.: Unsupervised deformable registration for multi-modal images via disentangled representations. In: Chung, A.C.S., Gee, J.C., Yushkevich, P.A., Bao, S. (eds.) IPMI 2019. LNCS, vol. 11492, pp. 249–261. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20351-1_19

24. Risholm, P., Golby, A.J., Wells, W.: Multimodal image registration for preoperative planning and image-guided neurosurgical procedures. Neurosurg. Clin. **22**(2), 197–206 (2011)

25. Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: SVF-Net: learning deformable image registration using shape matching. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 266–274. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_31

26. 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. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

27. Shen, Z., Han, X., Xu, Z., Niethammer, M.: Networks for joint affine and non-parametric image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4224–4233 (2019)

28. Sun, S., et al.: Robust multimodal image registration using deep recurrent reinforcement learning. In: Jawahar, C.V., Li, H., Mori, G., Schindler, K. (eds.) ACCV 2018. LNCS, vol. 11362, pp. 511–526. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20890-5_33

29. 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)

30. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Staring, M., Išgum, I.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 204–212. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_24

31. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: fast predictive image registration-a deep learning approach. NeuroImage **158**, 378–396 (2017)

32. Zhao, S., Dong, Y., Chang, E.I., Xu, Y., et al.: Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10600–10610 (2019)

33. Zhuang, J., Dvornek, N., Li, X., Tatikonda, S., Papademetris, X., Duncan, J.: Adaptive checkpoint adjoint method for gradient estimation in neural ODE. arXiv preprint arXiv:2006.02493 (2020)