Implicit Optimizer for Diffeomorphic Image Registration

Kun Han  Shanlin Sun
University of California, Irvine

Abstract. Diffeomorphic image registration is the underlying technology in medical image processing which enables the invertibility and point-to-point correspondence. Recently, numerous learning-based methods utilizing convolutional neural networks (CNNs) have been proposed for registration problems. Compared with the speed boosting, accuracy improvement brought by the complicated CNN-based methods is minor. To tackle this problem, we propose a rapid and accurate Implicit Optimizer for Diffeomorphic Image Registration (IDIR) which utilize the Deep Implicit Function as the neural velocity field (NVF) whose input is the point coordinate \( p \) and output is velocity vector at that point \( v \). To reduce the huge memory consumption brought by NVF for 3D volumes, a sparse sampling is employed to the framework. We evaluate our method on two 3D large-scale MR brain scan datasets, the results show that our proposed method provide faster and better registration results than conventional image registration approaches and outperforms the learning-based methods by a significant margin, while maintaining the desired diffeomorphic properties.

Keywords: Diffeomorphic image registration · Deep Implicit Function · Neural Velocity Field · Sparse sampling · Self-supervised learning.

1 Introduction

Diffeomorphic image registration has been widely used in medical image analysis and diagnosis \([12]\), which aims to establish invertible and continuous transformations between a pair of images, such as 3D Magnetic Resonance (MR) brain scans. The mathematical rigor of diffeomorphism comes at an important cost which needs to integrate over the velocity field to obtain the deformation. In order to efficiently calculate the diffeomorphisms, stationary velocity fields (SVF) \([8]\) have been proposed to decrease the degree of freedom and are widely used in many researches.

Conventional methods solve the registration task as an optimization problem. SVF has been modeled on the grid points to generate the deformation following the framework, known as Large Deformation Diffeomorphic Metric Mapping (LDDMM) \([5]\). The high accuracy resulting from the iterative non-linear optimization usually comes with the cost of slow process.

Recent years, many learning-based approaches have been proposed for image registration, which are trained to predict spatial transformation map between
two images using convolutional neural network (CNN), such as VoxelMorph and SYMNet. Multi-scale methods are also proposed, where the input images are registered in a coarse-to-fine manner with several cascaded networks. These methods are substantially faster than conventional methods requiring just one pass through the network. Yet, the performance improvement brought by the deep-learning image registration (DLIR) is minor. Moreover, DLIR requires vast 3D volumes as training data in order to predict a accurate mapping. For the dataset with limited available volumes and special modality attributes, DLIR often do not perform as well as for the large-scale dataset. However, optimization-based methods don’t have this kind of limitation for its running only require the two input images.

In our paper, we propose a rapid and accurate optimization-based method, IDIR, which implicitly models the velocity field using a multi-layer perceptron (MLP). This neural velocity field (NVF) can be viewed as the spatially continuous function $v(p) \leftarrow F(p)$, where the weights of the MLP are the parameters of the function. Unlike previous DLIR methods, the input of NVF is not image volumes but the coordinates. Our goal is to optimize the weights of MLP during the optimization so that it can build the accurate diffeomorphic mapping between the input coordinate set $P : \Omega \subset \mathbb{R}^3$ and the velocity field $V : \Omega \subset \mathbb{R}^3$. Furthermore, the velocity field $V$ can be integrated to generate the forward and inverse deformation field using scaling and squaring (SS) method.

In order to use SS to generate deformation field, we need the complete velocity field (i.e. shape $D \times H \times W$). However, the overall memory consumption will explode if NVF takes the entire coordinate set $P$ as input and the proposed model can not be fit into any modern GPU. To tackle this problem, instead of inputting entire coordinate set $P$, we extract a sparse coordinate set $P'$ and recover the resolution afterward. Besides the reduction of memory consumption, using sparse coordinate sampling can also improve the speed compared with the random sampling.

The main contribution of this work can be summarized as follows:

- We’re the first to propose a rapid and accurate diffeomorphic image registration method which utilize the deep implicit function to represent the velocity field.
- The sparse sampling strategy substantially reduces the memory consumption during the registration and improves the speed compared with random sampling.
- Extensive experiments on two large scale MRI brain datasets show that our proposed method requires less running time than conventional optimization methods and brings better registration results. With 6-minute running, IDIR outperforms the state-of-the-art DLIR methods by 6.3% and 0.9% respectively.

## 2 Methods

### 2.1 Overview

Let $F \in \mathbb{R}^{D \times H \times W}$ and $M \in \mathbb{R}^{D \times H \times W}$ denote the fixed and moving volumes in the task of image registration, $\Phi_{MF} : \mathbb{R}^3 \mapsto \mathbb{R}^3$ and $\Phi_{FM} : \mathbb{R}^3 \mapsto \mathbb{R}^3$ be the
Fig. 1: Overall framework of the proposed method IDIR. The complete coordinate set $\mathbf{P}$ is downsized to generate a sparse coordinate set $\mathbf{P}'$. Neural Velocity Field (NVF) maps the $\mathbf{P}'$ to the sparse velocity field $\mathbf{V}'$ which is then upsampled to recover the original resolution. Both forward and inverse transformation are generated using the scaling and squaring (SS) method. A bi-directional similarity loss $L_{\text{sim}}$ is applied to insure the accuracy and invertibility. $L_{\text{reg}}$ and $L_{\text{Jac}}$ are used to force the smoothness and local orientation.

forward and inverse deformation field mapping coordinates between $\mathbf{F}$ and $\mathbf{M}$. In the domain of diffeomorphic transformation, $\Phi_{MF}$ and $\Phi_{FM}$ are calculated through the integral of stationary velocity field $\mathbf{V}_{MF}$ and $\mathbf{V}_{FM}$. A neural implicit function $\mathbf{V}_{\theta}$ is utilized to parameterize such forward and inverse velocity field that $\mathbf{V}_{MF}(\cdot) = -\mathbf{V}_{FM}(\cdot) = \mathbf{V}_{\theta}(\cdot)$. To solve the pairwise registration problem, we update the neural velocity field weight $\hat{\theta}$ by minimizing $L(\mathbf{F}, \mathbf{M}, \mathbf{V}_{\theta})$ via gradient decent. We will discuss the design of neural velocity field $\mathbf{V}_{\theta}$ and optimization goal $L$ in Sec. 2.2 and Sec. 2.3 respectively.

### 2.2 Neural Velocity Field

**Deep implicit function** Instead of formulating an explicit optimization term between the source volume and the target, we use a neural velocity field (NVF) parameterization. NVF is basically a Deep implicit function (DIF) that non-linearly maps a 3D coordinate to a 3D velocity vector, consisting of multi-layer artificial neurons. Doing so has been shown to improve the optimization in various works [12,13] as the network serves as an "internal prior" in the optimization. Specifically, we construct the NVF using a simple 5-layer MLP with $\text{sine}$ activation functions. The experiments in [14] shows that compared with popular relu-family activation function, $\text{sin}$ function is more stable and results in faster convergence and higher accuracy.

**Sparse coordinates Sampling** The forward and inverse deformation field are generated through the integration of a complete velocity field (i.e. with shape...
D × H × W × 3). However, directly put entire coordinate set \( P \) into the NVF is unpractical for the exploding memory consumption. In our work, we first divide the coordinate grid into non-overlapping \( L \times L \times L \) patches. Then we extract the center coordinate from each small patch and input the downsized coordinate set \( P' (D \times H \times M) \) to NVF. NVF generate the sparse velocity field \( V' \) followed by a upsampling operation to recover the original resolution. By doing this, the memory consumption is substantially reduced and the integrated deformation fields are smoother compared with upsampling deformation field.

### 2.3 Optimization

To enforce the topology preservation and invertibility, we follow [10] to integrate over the stationary velocity using scaling and squaring method and generate the forward and inverse deformation field. Then we implement the differentiable spatial transformer to warp the 3d volumes and use local normalized cross-correlation (NCC) as our similarity metric to compute the alignment results in both directions:

\[
L_{sim} = NCC(F, M \circ \Phi_{MF}) + NCC(F \circ \Phi_{FM}, M)
\]  

(1)

Besides, in order to secure the local orientation consistency constraints, we follow [4] to impose a selective Jacobian determinant regularization. If the Jacobian determinant at a given point \( p \) is positive, then the deformation field preserves the orientation near \( p \). With a relu function, we can penalize the local region with a negative Jacobian determinant.

\[
L_{Jdet} = \frac{1}{N} \left( \sum relu(-|J_{\Phi_{FM}(p)}|) + \sum relu(-|J_{\Phi_{MF}(p)}|) \right)
\]  

(2)

Finally, we still apply a global regularizer to enforce the smoothness of deformation field \( L_{reg} = \sum (||\nabla \Phi_{FM}||^2 + ||\nabla \Phi_{FM}||^2) \). Therefore, we present the complete loss function as following:

\[
L = L_{sim} + \lambda_1 L_{Jdet} + \lambda_2 L_{reg}
\]  

(3)

### 3 Experiments and Results

#### Data and Pre-processing

We evaluate our proposed method on two 3D brain MRI datasets: the Mindboggle101 dataset and OASIS dataset. Mindboggle101 consists of 101 T1-weighted MR scans from 5 datasets, e.g. HLN-12, MMRR-21 and NKI-RS. We follow [11] to fuse corresponding left and right cortical regions and evaluate the performance on 31 remaining segmentations. OASIS dataset contains 425 T1-weighted MR images aging from 18 to 96. 35 anatomical structures were used to compare the registration performance with traditional and DLIR methods. For both two dataset, we carry out the standard preprocessing methods. With skull stripped, all scans are resampled to same resolution (\( 1mm \times 1mm \times 1mm \)), cropped to (162 × 198 × 144) and normalized by the maximum intensity of each volume.
Experimental Settings  For mindboggle 101 dataset, we sampled 20 scans evenly from 5 individual sub-datasets as our test images. And 3 scans are randomly selected as the atlas. Similarly for OASIS dataset, 20 and 3 volumes are randomly sampled as test images and atlas. Our model is implemented with Pytorch and trained on a GTX 2080Ti GPU. We set the learning rate = $10^{-4}$, $\lambda_{reg} = 0.1$ and $\lambda_{J_{det}} = 100$, with the coordinate patch size set to 3.

| Method          | Mindboggle | OASIS |
|-----------------|------------|-------|
| Affine          | 0.356 (0.019) | 0.543 (0.069) |
| ANTs-SYN(best)  | 0.548 (0.019) | 0.777 (0.027) |
| IDIR (2 min)    | 0.579 (0.014) (+2.1%) | 0.774 (0.013) (-0.3%) |
| IDIR (6 min)    | 0.603 (0.016) (+6.3%) | 0.784 (0.015) (+0.9%) |
| IDIR (best)     | 0.619 (0.017) (+9.1%) | 0.791 (0.021) (+1.8%) |
| IDIR_{grid} (best) | 0.569 (0.019) | 0.757 (0.043) |
| VoxelMorph      | 0.555 (0.018) | 0.763 (0.033) |
| SYMNet          | 0.567 (0.018) | 0.777 (0.028) |

Table 1: Quantitative results for image registration on Mindboggle and OASIS datasets. We compare the performance of proposed IDIR at different running-time stages. For ANTs-SYN and IDIR_{grid}, we report their best score in the table and present the running-time comparison in Fig. 2.

Measurement  Three metrics are used to evaluate the performance of our proposed method. We use the dice score (DSC) to evaluate the accuracy which is calculated based on anatomical segmentation. Secondly, as the topology preservation and invertibility are of great value in medical analysis, we measure the diffeomorphism using the percentage of voxels with non-positive Jacobian determinant ($|J_{\Phi}| \leq 0$). However, with proper regularization for the local orientation, all methods achieve a < 0.0% results, so we skip reporting the $|J_{\Phi}| \leq 0$ in this paper. Last, speed is critical in real medical application. As our method is a optimization-based method, we calculate the DSC and $|J_{\Phi}| \leq 0$ at different time stages and compare them with other methods.

Results Comparison  We compare our model with several popular image registration methods, including conventional methods: Affine Transformation and ANTs-Syn, and two DLIR methods: VoxelMorph and SYMNet. In order to measure the improvement brought by deep implicit function, we set up a new baseline IDIR_{grid} where the NVF is replaced by a trainable grid-structure parameter.

As shown in Table. 1, with a 2-minute running, the performance of our proposed method achieve a mean 0.579 Dice Score on Mindboggle dataset which outperforms the Syn by 5.6% and DLIR by 2.1%. 6-min running leads to a 10.0%
and 6.3% margin to Syn and learning-based methods respectively. As for the OA-SIS dataset, though the 2-min performance is not ideal, from the Fig. 2 we can see that the performance of proposed method are better than ANTs-Syn at the same running-time stage. Compared with DLIR methods, the proposed method outperforms by 0.9% at 6-min training stage. Based on these observations, we get the conclusion that our new proposed method achieve a new state-of-art performance and assures a faster and better results than other optimization-based methods.

![Fig. 2: Dice Score comparison at different running stages on two datasets. Green and Red stars are the results of VoxelMorph and SYMNet. Yellow, brown and blue curves are the performance of proposed IDIR, IDIR\textsubscript{grid} and ANTs-SYN at different running time.](image)

**Ablation Study** Three ablation studies are conducted to compare the variants of setting in our proposed method.

1. **Using deep implicit function as neural velocity field** Instead using NVF, we use a trainable grid-structure parameter as the velocity field. From the Table 1 and Fig. 2 we can see that using NVF provides a faster and better registration results.
| $\lambda_{J_{det}}$ | Dice Score | $|J_{\Phi}| \leq 0$ |
|-----------------|------------|----------------|
| 1               | 0.618 (0.017) | 9936.5 |
| 10              | 0.618 (0.017) | 1807.0 |
| 100             | 0.619 (0.016) | 165.3  |
| 1000            | 0.618 (0.015) | 6.9    |

Table 2: Dice Score and number of negative Jacobian determinant $|J_{\Phi}| \leq 0$ comparison across different $\lambda_{J_{det}}$ on Mindboggle Dataset.

2. $\lambda_{J_{det}}$ for local orientation loss Table 2 shows the effect of different $\lambda_{J_{det}}$, where the dice scores are stable and the number of points with negative jacobian determinant decreases as expected. However, from Fig. 3 the convergence rate is much slower if we progressively select a large coefficient, which is 1000. Therefore, in our model, we choose $\lambda_{J_{det}}=100$, which preserves the diffeomorphic property and also provide a fast speed.

3. Patch Size The last ablation study presents the effect of using different batch size. As shown from Table 3 small patch size results in better performance according to the dice score. But the memory consumption decreases with larger patch size. As we want our framework can be fit in most modern GPUs, which typically have memory size 11 GB, such as GTX 1080Ti or GTX 2080Ti. Considering both the accuracy and memory consumption, we use patch size 3 in our framework. One more thing to mention is from the Table 3, the performance of our proposed methods still provides a good result (outperforms other methods by 4.2%) even we use a much larger patch size as 6.

| Patch Size | Dice Score | MC(MiB) |
|------------|------------|---------|
| 2          | 0.622 (0.014) | 16633   |
| 3          | **0.619 (0.016)** | **8880** |
| 4          | 0.613 (0.016) | 6770    |
| 5          | 0.600 (0.017) | 6386    |
| 6          | 0.591 (0.015) | 5972    |
| 9          | 0.547 (0.016) | 5826    |
| 14         | 0.494 (0.013) | 5130    |

Table 3: Dice Score and Memory Consumption (MC) analysis across different patch size On Mindboggle Dataset.

4 Conclusion

This paper presents the a rapid and accurate optimization-based diffeomorphic image registration model named IDIR. By leveraging the power of implicit neural representation and sparse sampling. Extensive quantitative and qualitative
evaluations are conducted on two large-scale MR brain datasets. The results show that the proposed method is able to generate a faster and better registration result than conventional methods, and outperforms the state-of-the-art learning-based methods by a significant margin. Sufficient ablation study clearly demonstrate the choice of hyper parameters. In the future, we will continue on this topic and improve the speed by utilizing CNN encoder for initialization.

References

1. Incoronato, M., Aiello, M., Infante, T., Cavaliere, C., Grimaldi, A.M., Mirabelli, P., Monti, S., Salvatore, M.: Radiogenomic analysis of oncological data: a technical survey. International journal of molecular sciences 18(4), 805 (2017)
2. Risholm, P., Golby, A.J., Wells, W.: Multimodal image registration for preoperative planning and image-guided neurosurgical procedures. Neurosurgery Clinics 22(2), 197–206 (2011)
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. Mok, T. C., Chung, A. (2020). 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.
5. J. Ashburner, “A fast diffeomorphic image registration algorithm,” NeuroImage, vol. 38, no. 1, pp. 95–113, 2007.
6. V. Arsigny, O. Commowick, X. Pennec, and N. Ayache, “A log-Euclidean framework for statistics on diffeomorphisms,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2006, pp. 924–931.
7. N. J. Higham, “The scaling and squaring method for the matrix exponential revisited,” SIAM J. Matrix Anal. Appl., vol. 26, no. 4, pp. 1179–1193, 2005.
8. 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)
9. M. Corvi and G. Nicchiotti, "Multiresolution image registration," Proceedings., International Conference on Image Processing, 1995, pp. 224-227 vol.3, doi: 10.1109/ICIP.1995.537621.
10. A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, (2018). "Unsupervised learning for fast probabilistic diffeomorphic registration." Medical Image Computing and Computer Assisted Intervention (MICCAI), Springer, Cham (2018), pp. 729-738
11. Xu, Zhenlin, and Marc Niethammer. "DeepAtlas: Joint semi-supervised learning of image registration and segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019.
12. Williams, F., Schneider, T., Silva, C., Zorin, D., Bruna, J. & Panozzo, D. Deep geometric prior for surface reconstruction. Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition. pp. 10130-10139 (2019)
13. Hanoeka, R., Metzer, G., Giryes, R. & Cohen-Or, D. Point2mesh: A self-prior for deformable meshes. ArXiv Preprint arXiv:2005.11082 (2020)
14. Sitzmann, Vincent, et al. "Implicit neural representations with periodic activation functions." Advances in Neural Information Processing Systems 33 (2020): 7462-7473.