

Delta⁴

by ScandiDos

Be confident implementing
Radixact Synchrony clinically



THE ONLY PLATFORM MOVING IN 7 DIMENSIONS*

We provide you with a realistic pre-treatment verification of the delivered treatment for Radixact Synchrony.

In collaboration with Accuray, ScandiDos has developed a solution that improves the quality assurance (QA) of radiotherapy treatments of moving targets. The solution independently simulates the breathing motion of patients, therefore, adding a seventh dimension to the tumor movement simulation provided by the Delta4 HexaMotion.

*longitudinal, lateral, height, roll, tilt, time and breathing motion.

Learn more ▶

Delta4family.com

LDVoxelMorph: A precise loss function and cascaded architecture for unsupervised diffeomorphic large displacement registration

Jing Yang^{1,2} | Yinghao Wu¹ | Dong Zhang^{1,2} | Wenting Cui² | Xiaoli Yue¹ | Shaoyi Du² | Hongmei Zhang³

¹ School of Automation Science and Engineering, Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China

² Institute of Artificial Intelligence and Robotics, College of Artificial Intelligence, Xi'an Jiaotong University, Xi'an, China

³ Xijing Hospital, Xi'an, China

Correspondence

Shaoyi Du, 28 Xianning West Road, Xi'an 710049, China.

Email: dushaoyi@xjtu.edu.cn

Hongmei Zhang, 127 Changle West Road, Xi'an 710032, China.

Email: zhm@fmmu.edu.cn

Jing Yang, Yinghao Wu, and Dong Zhang contributed equally to this work.

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 62073257, 61971343; Research and Development Program of Shaanxi Province of China; 2020GXLY-Y-008

Abstract

Purpose: The traditional learning-based non-rigid registration methods for medical images are trained by an invariant smoothness regularization parameter, which cannot satisfy the registration accuracy and diffeomorphic property simultaneously. The diffeomorphic property reflects the credibility of the registration results.

Method: To improve the diffeomorphic property in 3D medical image registration, we propose a diffeomorphic cascaded network based on the compressed loss (CL), named LDVoxelMorph. The proposed network has several constituent U-Nets and is trained with deep supervision, which uses a different spatial smoothness regularization parameter in each constituent U-Nets for training. This cascade-variant smoothness regularization parameter can maintain the diffeomorphic property in top cascades with large displacement and achieve precise registration in bottom cascades. Besides, we develop the CL as a penalty for the velocity field, which can accurately limit the velocity field that causes the deformation field overlap after the velocity field integration.

Results: In our registration experiments, the dice scores of our method were 0.892 ± 0.040 on liver CT datasets SLIVER³⁷, 0.848 ± 0.044 on liver CT datasets LiTS³⁸, 0.689 ± 0.014 on brain MRI datasets LPBA³⁸, and the number of overlapping voxels of deformation field were 325, 159, and 0, respectively. Ablation study shows that the CL improves the diffeomorphic property more effectively than increases.

Conclusion: Experiment results show that our method can achieve higher registration accuracy assessed by dice scores and overlapping voxels while maintaining the diffeomorphic property for large deformation.

KEY WORDS

convolutional neural network, diffeomorphism, fluid registration, non-rigid registration, recursive cascade network

1 | INTRODUCTION

Witnessed by the rapid development of computer vision and digital image processing, image registration has become an important tool for image comparison, fusion, three-dimensional visualization, target recognition, and medical image analysis. Image registration matches two or more image information obtained

from different times, different sensors, or different environments, which can make the spatial expression of the matched image information consistent.^{1,2} It is generally divided into rigid registration and non-rigid registration. Due to the large deformation in medical images caused by the activity of soft tissues or organs, non-rigid registration is widely used in medicine.^{3,4}

The traditional non-rigid registration algorithm is generally based on iterative optimization to solve the registration problem, which is easy to fall into the local extreme value and greatly affected by the initial value, such as the coherent point drift algorithm (CPD)⁵ and the thin plate spline robust point matching (TPS-RPM).⁶ These methods have high registration accuracy, but high time cost. The optimization method based on deep learning uses a large amount of data for learning. Although the optimization time is high when learning, the registration time is short when used, which is suitable for clinical real-time registration tasks. Since medical images are too variable for imaging some organs or soft tissues, there may not be a static deformation field. It needs continuous and inverse mapping to ensure the homeomorphic property, and the mapping is differentiable. The velocity field is used to calculate the deformation field and thus achieve and satisfy the diffeomorphic property for the alignment of medical images with large deformations. For example, Yang⁶ proposed a registration algorithm with similarity and affine transformations based on bidirectional a bidirectional kernel mean p-power error loss, named ANTs SyN1, which can guarantee diffeomorphic property through geometric or variational methods. However, traditional registration algorithms require more iterative computations compared to deep learning, which causes the problem of low time efficiency. These algorithms are computationally intensive and not more difficult for practical clinical cases with more large liver deformations than the brain.

In recent years, the registration method based on deep learning has gradually attracted researcher's attention due to its advantages of rapid registration. It estimates non-parametric spatial transformation to warp a moving image to a fixed image. Due to the complexity of generating ground truth for medical images registration, unsupervised method is widely used. Unsupervised methods^{7,8} minimize the similarity loss function between two images and the smoothness regularization loss in the deformation field. In testing, the learning-based methods require fewer or no iterations compared to traditional optimization-based methods. So, these methods are more productive than traditional methods. However, they all use a single smoothness regularization parameter to pursue the smoothness of the deformation fields but ignore the inevitable overlap of the deformation fields in practical cases. It is difficult to pursue a good registration accuracy and diffeomorphic property simultaneously by balancing the effect between the smoothness and folding area.

VoxelMorph⁹ was proposed for deformable medical image registration including a convolutional neural network and a spatial transformation network.¹⁰ Its registration accuracy performs comparably to ANTs and achieved the average dice of 0.75, which is higher than a baseline of only global affine alignment, but it has

poor diffeomorphic property. Later, VoxelMorph was reimplemented with an integral affine network based on the Bayesian probability registration model of the fixed velocity field.⁸ The registration results satisfy the diffeomorphic property in brain MRI datasets but shown deficiency on other datasets such as liver CT scans by later work. Besides, John Ashburner¹¹ mentioned the fixed velocity field used by VoxelMorph to encode the whole trajectory of an evolving diffeomorphism, which may force the diffeomorphism to take very circuitous and high energy trajectories for good correspondence between images. However, in the case of large deformations, only the fixed velocity field cannot be well registered, which will cause the registration accuracy to decrease. VoxelMorph can only guarantee the characteristics of diffeomorphism under the condition of a small range of deformation. The reason is that the generated velocity field only considers smoothness. When deformed in a large range, the velocity field will converge and diverge. At this time, this smoothness is difficult to maintain, which will cause the mesh to move in this speed field and collide. In experiments with clinical image data, we observe that the registration results will be inaccurate if we excessively encourage local smoothness of velocity using a large spatially invariant regularization parameter in VoxelMorph. Meanwhile, there will be many overlapping voxels in the deformation field and yielding unreasonable registration if we excessively encourage accurate registration.¹²

Recently, to improve the registration accuracy, a recursive cascade network was proposed by Zhao,¹³ which can be used in any registered basic network to learn the information through the deep cascade. The result shows a higher registration accuracy but more overlapped voxels, which means diffeomorphism fails. Guha Balakrishnan⁹ presented that it easily guaranteed the diffeomorphic transformation by increasing spatial regularization parameters, but the registration accuracy decreased. Therefore, how to improve registration accuracy while making overlapping voxels in the deformation field as few as possible is the focus of the current research.

To address the above-mentioned problems, we propose a cascaded network composed of multiple outputs and a loss function that can maintain the characteristics of diffeomorphic property in a large displacement deformation registration environment. The main contributions are in three aspects.

First, to take full advantage of the multilevel feature information from coarse to fine, we propose a cascaded architecture that has three constitute U-Net outputs in the feature extraction network, named LDVoxelMorph. The fixed 3D image data will gradually approximate the real moving 3D image by combining weak deformation field feature information from multiple feature extractors, which can improve the registration accuracy.

Second, to balance this contradiction between registration accuracy and diffeomorphic property, we design a new anti-folding loss function based on fluid mechanics, called compressed loss (CL), as a part of smoothness regularization loss. It can predict where the deformation fields are compressed violently through the velocity field and suppress the voxel overlap of the deformation field.

Third, deep supervision is performed to train all the constitute U-Nets simultaneously by the regularization loss integrated spatial smoothing and compressed losses calculated in each cascade, making the generated local velocity field more flexible, obtaining higher registration accuracy than the original VoxelMorph when the image has a large deformation.

The paper is organized as follows. In Section 2, the related work is reviewed and the framework of the proposed method is introduced. In Section 3, the performance of our method is verified by two experiments, including the registration of 2D images, and the registration of 3D liver CT scans and brain MRI scans. Finally, discussion and conclusion are given.

2 | MATERIALS AND METHODS

2.1 | Diffeomorphic image registration

For non-rigid medical image registration, there are the following definitions: m , f are two 2D or 3D images, f is a fixed image, and m is a moving image to be deformed. ψ is a variable that determines a transformation function: $\phi_\psi: R^n \rightarrow R^n$, n is the dimension of the input image. Generally, image registration is mainly based on the following typical optimization form:

$$\psi = \arg \min [L_{\text{Sim}}(m \circ \phi_\psi \cdot f) + \lambda L_{\text{Reg}}(\phi_\psi)] \quad (1)$$

where \circ means the resampling operation, which combines the predicted deformation field and moving image to generate $m \circ \phi_\psi$. L_{Sim} measures the alignment between the fixed image f and the transformed moving image $m \circ \phi_\psi$ and sum-of-squared differences, cross-correlation or mutual information is the explicit form of it. L_{Reg} measures the smoothness of transformation, which is a penalty encouraging the generation of spatially smooth deformation field. For low-dimensional parameterizations of ϕ_ψ , for example, for affine or B-spline models, L_{Reg} may not be necessary. However, for non-parametric registration models, L_{Reg} is necessary for a well-posed optimization problem. λ balances the trade-off between the similarity term L_{Sim} and the regularization term L_{Reg} . The first term represents the similarity between the registered image and the fixed image. The higher the similarity, the higher the registration accuracy. The second term represents the regularization term of the deformation field. The lower the value of the regu-

larization term, the smoother the mesh, that is, the lower the number of voxel overlaps.

For non-parametric registration models,¹⁴ displacement and velocity usually represent deformation. They require an excellent regularization parameter to replace the derivative of the differential by the difference and approximate the differential by the resulting difference, which guarantee diffeomorphic transformation. For example, Kuang¹⁵ introduced a negative Jacobian loss on the deformation field to penalize the folding area, called FAIM. Zhang¹⁶ introduced anti-folding constraints to the deformation field, which is similar to FAIM. Both of them designed a loss function directly applied to the deformation field and can precisely decrease the spatial folding area in deformation field. Different from giving restrain on displacement, VoxelMorph⁹ introduces diffeomorphic velocity layers with a velocity smoothness loss for diffeomorphic registration. However, a fixed smoothness regularization parameter penalized in the velocity field is not precise enough in large deformation, which may make the diffeomorphic property fail. Besides, simply improving the smoothness in the velocity field may cause the registration accuracy to decrease.^{7,8} Tony¹² adds a negative Jacobian loss function to the deformation field integrated by the velocity field. Our method is to directly and accurately limit the generation of the velocity field where cause deformation field overlap. To our best knowledge, no one has designed a precisely anti-folding constraint on the velocity field. Here, we design a compressed loss to velocity field, which combines the advantages of scaling and squaring layer in VoxelMorph and precise anti-folding constraints based on the velocity field.

2.2 | Cascaded architecture and skip connections

Cascade architecture has been used in a variety of registration methods. Traditional algorithms iteratively optimize some common energy functions.^{17–23} These methods perform an alignment with a similar function on the current deformed image during the iteration. In rigid registration, the iterative closest point algorithm^{24,25} matches the closest point pair in each iteration and gradually approaches the correct registration parameters by minimizing iteration errors. In the learning-based registration method, VTN²⁶ uses an end-to-end cascading scheme that resolves large displacement. Zhao¹³ uses a recursive cascade network recursively warping a moving image to a fixed image and achieves registration accuracy of 0.944 to 0.950 from 5 to 15 cascades. However, both methods use a whole base network as cascades and are computationally intensive with GPU time of 0.41 to 1.09 s and CPU time of 69 to 201 s, which weakens the advantages of fast registration in learning-based registration. Besides, information

transmission within cascades is only connected by the warped image in each cascade, and the feature layer information lacks connection in each cascade. Recently, U-Net++²⁷ was proposed with several constituent U-Nets and trained simultaneously with deep supervision, which acquires segmentation accuracy of 0.8933, 0.9121, and 0.9244 in electron microscopy, brain tumor, and nuclei datasets, respectively. The proposed method introduces cascaded architecture based on the multi-output of U-Net++, which takes each output as input and the warp moving image sequentially to approach the fixed image. Our network combines the information of the previous cascade by skip connections and has fewer training parameters in bottom cascades, which facilitate fast registration.

Skip connections were first proposed in fully convolutional networks (FCN)²⁸ for semantic segmentation. U-Net architectures²⁹ with some skip connections were proposed for semantic segmentation in medical images. The skip connections help recover the full spatial resolution and achieve finer segmentation which combines feature layers of different depths. Later, U-Net++ introduces a built-in ensemble of U-Net of varying depths and uses forward-skip connections between every two convolution blocks in some depth, which achieves state-of-the-art accuracy in medical image segmentation. This type of skip connection is computationally intensive for the 3D medical image but not suitable for feature transfer in our cascaded architecture. Therefore, our work redesigns a new type of skip connection on U-Net++, which is suitable for our cascaded architecture.

2.3 | Time-varying regularization parameter

Regularization aims to prevent overfitting and essentially to optimize the target parameters by the constraint condition. The spatial-varying and time-varying regularizations are common methods used in registration tasks. The former focuses on the spatial structure relationship between feature points, while the latter focuses on the position relationship of themselves. There are several works in spatial-varying regularization parameter, which mainly focuses on traditional optimization-based non-rigid registration.^{30–33} Recently, a learning-based method based on metric learning³⁴ is proposed, which use neural network to learn a spatial-varying regularization parameter for a stationary velocity field registration. This method only considers the spatial difference of regularization parameters without the time-varying parameters. However, time-varying regularization parameters are important, because larger regularization parameters should be used in the early registration stages to learn the global displacement and lower regularization parameters used in the later registration stages to refine the details. In Equation (1), a larger regularization parameter

λ will produce a smoother non-overlapping deformation field ϕ_ψ and a lower similarity between $m \circ \phi_\psi$ and f , while a smaller λ may make deformation field ϕ_ψ overlap and a finer similarity for them. In our cascaded architecture, each cascade output is affected by the independent optimization Equation (1) with a hyperparameter λ . We set a larger λ to prevent the deformation field ϕ_ψ from overlapping in top cascades and a smaller λ to ensure registration accuracy in bottom cascades. In the proposed method, cascade-varying regularization is used to simulate time-varying regularization in registration based on deep learning.

2.4 | Cascaded architecture in LDVoxelMorph

For non-rigid medical image registration based on deep learning, there are the following definitions: f and m are two 2D or 3D images, and m means a moving image to be deformed, and f denotes a fixed image. A transformation function is definite by $\phi_v : R_n \rightarrow R_n$, v means a hidden variable, and n indicates the dimension of the input image. Here, we use velocity field and a generative model⁷ to represent this transformation function, and the warped image is definite $m \circ \phi_v$, which means m through ϕ_v and make f and $m \circ \phi_v$ as similar as possible. Figure 1 is the overall architecture named LDVoxelMorph, which implements VoxelMorph⁷ with a cascaded architecture and a compressed loss.

We observe that the registration results will be inaccurate if we excessively encourage local smoothness of velocity using a large spatially invariant regularization parameter in VoxelMorph. Meanwhile, there will be many overlapping voxels in the deformation field and yielding unreasonable registration if we excessively encourage accurate registration.¹¹ To compromise between registration accuracy and diffeomorphic property, we design a cascaded architecture named LDVoxelMorph which contains three parts, as shown in Figure 1. The first part is a U-Net++ style encoder-decoder network, which contains four down-sampling layers, six up-sampling layers and some skip connections. We increase the sample to half of the image resolution to save memory and restore the resolution by bilinear interpolation. The first part has three sampled outputs of velocity fields v_1, v_2, v_3 and compute deformation fields $\phi_{v_1}, \phi_{v_2}, \phi_{v_3}$ in the second part using scaling and squaring layers proposed by VoxelMorph.⁷ The third part is cascaded architecture, we successively warp moving image m_0 as follows:

$$\begin{cases} m_1 = m_0 \circ \phi_{v_1} \\ m_2 = m_1 \circ \phi_{v_2} \\ m_3 = m_2 \circ \phi_{v_3} \end{cases} \quad (2)$$

In each block output of U-Net++. The total training loss is the sum of a similarity loss and regularization

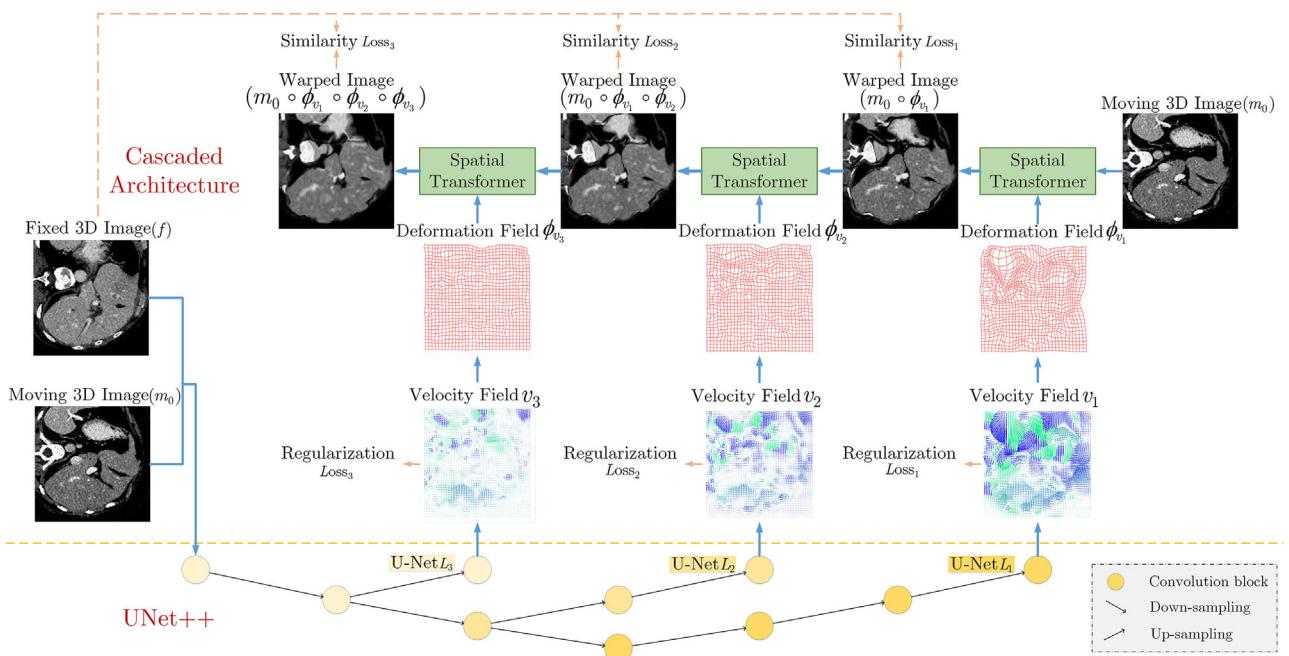


FIGURE 1 Illustration of the overall structure of LDVoxelMorph and how gradients back-propagate. The network contains three parts. The first part is a U-Net++, which takes the input images and simultaneously output three velocity fields. The second part transforms velocity field to deformation field using squaring and scaling layers proposed in VoxelMorph. Finally, the spatial transformer networks sequentially warp the moving image to approach a fixed image. The network trained by deep supervision and each constituent of U-Net layer has independent loss function, including similarity loss and regularization loss

loss. The similarity loss is measured by each warped image and fixed image. The regularization loss contains spatial smoothing loss and compressed loss, which we explain in the next section. We use mean square error as similarity loss in this paper. The total loss in each U-Net layer is as follow:

$$\text{Loss}_i = \text{MSE}(m_{i-1} \circ \phi_{v_i}, f) + \lambda_i (\|Dv_i\|_2 + CL_i) \quad (3)$$

where i indicates the layer number of constitute U-Net++. Dv_i means the spatial smoothing loss, CL_i denotes the compressed loss. λ_i is fixed hyper-parameter in each cascaded constitute U-Net.

2.5 | The configuration and rationale about the regulation parameters

Generally, we can use coarse-to-fine method combined with different regulation parameters to preserve diffeomorphic property and registration accuracy when handle a large deformation registration problem. In our study, we introduce this principle on deep-learning-based registration and design the LDVoxelMorph with three cascaded blocks for registration.

In top cascade block, the difference between moving image m and fixed image f is large, so the diffeomorphic property is more likely to fail and cause the deformation field grid to overlap. Therefore, we use a larger

regularization parameter to penalize the grid overlap. On the contrary, the difference between moving image m and fixed image f is small in the bottom cascade block. It is enough to use a relatively small regularization parameter to prevent grid overlap and smaller regularization parameter can improve the registration accuracy. In short, as the output of the cascade increases, the regularization parameter used to punish the overlapping of the deformation field should gradually decrease. The changed regularization parameter makes the network find a balance between registration accuracy and diffeomorphic property. We train the LDVoxelMorph end-to-end by deep supervision.

2.6 | Skip connections in LDVoxelMorph

The feature extraction network in LDVoxelMorph and U-Net++²⁷ is similar except the skip connections. U-Net++ introduces a built-in ensemble of U-Net of varying depths and use forward-skip connections between all of convolution blocks in some depth for medical segmentation. It computes intensively for 3D medical images and is not suitable for our cascaded architecture. We redesign the U-Net++ with some forward- and backward-skip connections, as shown in Figure 2. The backward-skip connection can make the current cascade get the information from the previous cascade, which is important for the network outputting current

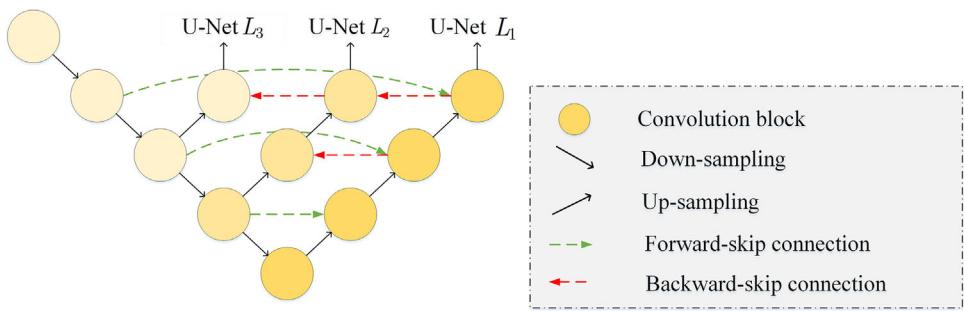


FIGURE 2 Feature extraction modules. The figure illustrates feature transfer in U-Net++. Several backward-skip connections adopted to integrate previous cascaded features into current cascade

velocity combined with precious cascade. First, in the spatial dimension, the different layers produce various features, the low-level layers contain more local detailed features, and the high-level layers include more global semantic features. The fusion of detailed and semantic features enriches the feature representations, which in turn promotes the model performance. The backward-skip connections in Figure 2 obtain the better feature representation while adapting to the data feature distribution and delivers the global semantic information to the velocity field contained more detailed features. Second, in the time dimension, the cascade aims to use similarity features and velocity fields to combine the gradually refined registration process while ensuring the diffeomorphic property. The cascade-L1 tends to tackle the orientation problem of the pixel points to be aligned in an extensive range, that is, these points are constrained by high-level semantic information with strong global directionality. The cascade-L3 prefers to deal with the local adjustment problem in the refined registration, that is, the alignment points need more local detailed feature information to ensure model accuracy. Therefore, the forward connections in L1 to L3 combine with semantic and detailed information to optimize the registration gradually, which provides global directionality and local accuracy and avoids disturbances from similar and noisy velocity features.

2.7 | Compressed loss as anti-folding constraint

The compressed loss is inspired by fluid compression and expansion. In our algorithm, the registration field ϕ_v is obtained by integrating the stationary velocity field v over $t = [0, 1]$. Generally, folding area in ϕ_v is more likely to occur where the grid density is high and it can be judged by a negative Jacobian. We treat the deformation field grid intersection as fluid molecules. When fluid molecules gather, the density increases, and vice versa. We make an analogy between fluid density and deformation field density. Before integrating, fluid density is uniform just like a uniform grid, which means

fluid density ρ_0 is constant. After integrating, fluid density changes with coordinates just like grid points in deformation field. So, we design a compressed penalty on velocity field where the grid density may change greatly in deformation field after integrating. The changes in grid density are just like the compressed air by different pressure in each area. In physics, the velocity of fluid conforms to the law of conservation of mass which can be formulated as

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \vec{u}) = 0 \quad (4)$$

where \vec{u} denotes velocity vector defined in the Euler coordinate system and ρ is Euler field that denotes fluid density on each coordinate. $\rho(x, y, z)$ reflects the number of fluid molecules gathered in coordinate (x, y, z) and we treat deformation field grid intersections as fluid molecules in this paper. ∇ denotes divergence operator in space. Integrate the formula over $t = [0, 1]$ to obtain the final density ρ_1 :

$$\rho_1 = - \left(\int_0^1 (\vec{u} \cdot \nabla \rho) dt + \int_0^1 (\rho \cdot \nabla \vec{u}) dt + \rho_0 \right) \quad (5)$$

where ρ_0 indicates constant density at $t = 0$. We approximate the density changes in a certain area are the same, that is, $\nabla \rho = 0$. The density changes rate formulated as

$$\frac{\rho_1}{\rho_0} = - \left(\left(\int_0^1 \frac{\rho}{\rho_0} dt \right) \cdot \nabla \vec{u} + 1 \right) \quad (6)$$

where \vec{u} means a stationary velocity field and only changes in space.

Equation (6) shows that the change of $\frac{\rho_1}{\rho_0}$ is proportional to the negative value of $\nabla \vec{u}$, and the greater the change in density, the greater the inverse coefficient $\int_0^1 \frac{\rho}{\rho_0} dt$. $\rho_1 > \rho_0$ means deformation field grid density increase in specific coordinate (i, j, k) . Only when $\rho_1(i, j, k) > \rho_0(i, j, k)$, the deformation field may overlap because of the increased gathered deformation field

grid intersections in coordinate (i, j, k) . So, we only punish the velocity field where $\rho_1 > \rho_0$. The formula is as follows:

$$-\left(\left(\int_0^1 \frac{\rho}{\rho_0} dt\right) \cdot \nabla \vec{u} + 1\right) > 0 \quad (7)$$

where $\int_0^1 \frac{\rho}{\rho_0} dt$ is the integral of the density ratio over $t = [0, 1]$ and set $A = \int_0^1 \frac{\rho}{\rho_0} dt$. After simplification:

$$\nabla \vec{u} < -\frac{1}{A} \quad (8)$$

Due to our fixed Euler velocity field, a specific location will maintain gather or diverge fluid molecules over $t = [0, 1]$, which means ρ will increase or decrease monotonously. Equation (7) is derived under the assumption $\rho_1 > \rho_0$ of Equation (6). So, we can infer that ρ is always greater than ρ_0 over $t = [0, 1]$. In other words, $A \in (1, +\infty]$. So, we can infer that the overlap of the voxel may occur where $\nabla \vec{u} < 0$. We design a loss function named compressed loss (CL) that precisely penalizes where density increase dramatically in velocity field. The formula is as follows:

$$CL = \sum_{p \in \Omega} \left(\frac{|\nabla \vec{u}(p)| - |\nabla \vec{u}(p)|}{2} \right)^\gamma \quad (9)$$

where p denotes pixel location in image, γ indicates hyperparameter depended on deformation degree. When the image has a large deformation, CL has a larger value, which has a greater effect on the local velocity field of the deformed part. For the small deformation part of the image, CL has a small value, and the local velocity field is smooth and suppresses the voxel overlap of the deformation field.

2.8 | Dataset

We evaluate our method and VoxelMorph on a c-shape dataset and two types of 3D medical images, including large deformation liver CT scans and small deformation brain MRI scans. For liver CT scans, we use subject-to-subject registration, which means any image in the dataset can be used as a fixed image. For brain MRI scans, we use atlas-based registration, letting each moving image be registered to a fixed atlas.

2.8.1 | C-shape image dataset

On a wholly controlled c-shape dataset, the effects of the proposed method on velocity fields, differential fea-

tures can be visualized more clearly and expediently, and further validated for differential homogeneous properties. Compared to complex medical image registration, such an experimental transition is necessary and advisable. We use atlas-based registration in the c-shape dataset, which uses a full c-shape image as a fixed image. We also use varying degrees of incomplete c-shape images as moving images. In the test, we apply the deformation field to warp marked different color c-shape pictures to track pixel movement.

2.8.2 | Medical image dataset

We preprocessed liver CT scans and brain MRI scans via the method proposed by Zhao.¹³ As required in VoxelMorph,⁹ we use ANTs³⁵ to pre-affine the datasets before deformable registration. For liver CT scans, the training datasets contain parts of MSD³⁶ which contain CT scans of liver tumors (70 scans excluding LiTS), hepatic vessels (443 scans), and pancreas tumors (420 scans). To sum, there are 933 (933 = 70 + 443 + 420) image pairs for unsupervised training. We use SLIVER³⁷ and LiTS³⁸ for evaluation, where SLIVER contains 20 scans and LiTS contains 131 scans, both have liver segmentation ground truth. For brain MRI scans, the training datasets contain: ADNI¹⁸ with 66 scans, ABIDE³⁹ with 1287 scans, and ADHD⁴⁰ with 949 scans, all of them are preprocessed by affine spatial normalization using ANTs. The LPBA⁴¹ dataset was used for evaluation with 40 scans, each of which has segmentation ground truth of 56 anatomical structures. Besides, all the above scans are resampled to $128 \times 128 \times 128$.

2.9 | Evaluation metrics

For registration accuracy, we use the dice similarity coefficient to evaluate the overall performance, which can be formulated as

$$Dice(A, B) = 2 \cdot \frac{|A \cap B|}{|A| + |B|} \quad (10)$$

where A, B mean corresponding anatomical structure between warped moving image and fixed image.

For diffeomorphic property, we use the Jacobian matrix $J_{\phi(p)} = \nabla \phi(p) \in \mathbb{R}^{3 \times 3}$ as our evaluation metric, which captures the local degree of overlap of ϕ around voxel p . The local deformation is diffeomorphic, both invertible and orientation preserving, only at locations where $|J_{\phi(p)}| > 0^4$. We count the average number of voxels N in each deformation field where $|J_{\phi(p)}| < 0$.

$$N = \sum (|J_{\phi(p)}| \leq 0) \quad (11)$$

3 | RESULTS

3.1 | Parametric settings of the compressed loss

We describe the parameters to be tuned during experiments, including λ in smoothness regularization and γ in the compressed loss. Generally, for the smoothness regularization parameter λ , the larger the value, the greater the penalty for speed smoothing. Due to a large difference at the initial cascade for the large deformation, a larger regularization parameter λ is required to force the smoothing of the velocity field. With increasing the cascade, the similarity of the cascaded output images is higher and closer to the fixed image, so the λ is also designed to be reduced to increase the flexibility of the velocity field to increase the registration accuracy. The influence of the parameter γ in the compression loss is similar to that of λ . In the initial cascaded, the γ is larger to increase the penalty for the unsmooth speed field, and then gradually decrease. In c-shape image experiment: we set λ as 20 and 100, respectively, in VoxelMorph and set $\lambda_1 = 100$, $\lambda_2 = 50$, $\lambda_3 = 25$, $\gamma_1 = 4$, $\gamma_2 = 3$, $\gamma_3 = 2$ in LDVoxelMorph. In liver CT scans experiment: we set λ as 200 in VoxelMorph and set $\lambda_1 = 600$, $\lambda_2 = 300$, $\lambda_3 = 100$, $\gamma_1 = 5$, $\gamma_2 = 4$, $\gamma_3 = 3$ in LDVoxelMorph. In brain MRI scans experiment: we set $\lambda = 100$ in VoxelMorph and set $\lambda_1 = 100$, $\lambda_2 = 50$, $\lambda_3 = 25$, $\gamma_1 = 3$, $\gamma_2 = 2$, $\gamma_3 = 1$ in LDVoxelMorph. In the ablation study, we set $\gamma = 3$ in compressed loss and the same λ in the comparative experiment.

3.2 | Experiment on C-shape image dataset

In the test, we use incomplete c-shape images as a moving image registered to a fixed full c-shape image. Figure 3 visualizes the results of the baseline network (VoxelMorph) with hyperparameters $\lambda = 20$ and $\lambda = 100$, respectively. The first column of Figure 3 visualizes the area using yellow dot where overlapping voxels occur ($|J\phi(p)| \leq 0$). The second column is registration results with rainbow markers. The third column is deformation field from the integral of the velocity field and the last is flow field.

We observe that VoxelMorph with a low regularization parameter ($\lambda = 20$) can achieve a better registration accuracy (Moved result in line 2 in Figure 3) but have many overlapping voxels (where $|J\phi(p)| \leq 0$). The result in VoxelMorph with a high regularization parameter $\lambda = 100$ has no overlapping voxel, but the accuracy is poor. Therefore, VoxelMorph cannot balance the registration accuracy and diffeomorphic property well.

VoxelMorph directly predicts one deformation field while LDVoxelMorph predicts three cascaded deformation fields to sequentially warp a moving image to

approach a fixed image. The visualized results of our method are better on the same registration task in Figure 3, as shown in Figure 4. We observe that each part of the deformation field has no folding area, so the diffeomorphic property preserves well. Moreover, thanks to the bottom cascades (warped field 3 in Figure 4) as a refinement to warped image, the final registration result (warped image 3 in Figure 4) is accurate enough compared to the fixed image. LDVoxelMorph does well in c-shape experiment because we set a larger regularization parameter in top cascade for preserving diffeomorphic property and smaller in bottom cascade for improving registration accuracy.

3.3 | Experiment on medical image datasets

We compare the proposed method to the traditional method (ANTs SyN³⁵) and baseline networks (VoxelMorph,⁸ VTN²⁶). ANTs SyN is the algorithm with the highest registration accuracy in the traditional registration method and the best maintenance of the differential homeomorphism characteristics. VTN network has the highest registration accuracy in the cascade network, but it can also be found that the deformation field meshes have a large number of overlaps, and the differential homeomorphism characteristics cannot be well guaranteed. VoxelMorph introduces a scaling squaring layer based on the velocity field, which greatly improves the characteristics of the network's differential homeomorphism. However, the characteristics of diffeomorphism cannot be maintained well on the liver dataset with a large range of deformation. Table 1 presents the average dice scores and the average number of overlapping voxels on liver datasets (SLIVER, LiTS) and brain datasets (LPBA). The meaning of the numbers in the bracket are the standard deviation among the test set. In testing experiments, 20 scans of SLIVER dataset form 380 test groups because of scan-to-scan registration. LiTS and LPBA datasets have 130 and 39 test groups, respectively, because of atlas-based registration.

The proposed compressed loss can reduce the number of overlapping voxels (see N in Table 1) by approximately 40% while ensuring the accuracy of Dice scores compared to the baseline network (VoxelMorph). The lower number of overlapping voxels responds to the better property of diffeomorphic property, which reflects the credibility of the registration results. In fifth line, LDVoxelMorph removed the compressed loss and the alternative loss consists of the similarity loss and spatial smoothing regularization item, which denotes "VoxelMorph+Cascades." The results show that cascaded architecture outperforms the VoxelMorph in both registration accuracy and overlapping voxels. A reasonable explanation is that we use a cascade-variant regularization parameter for deep supervision in

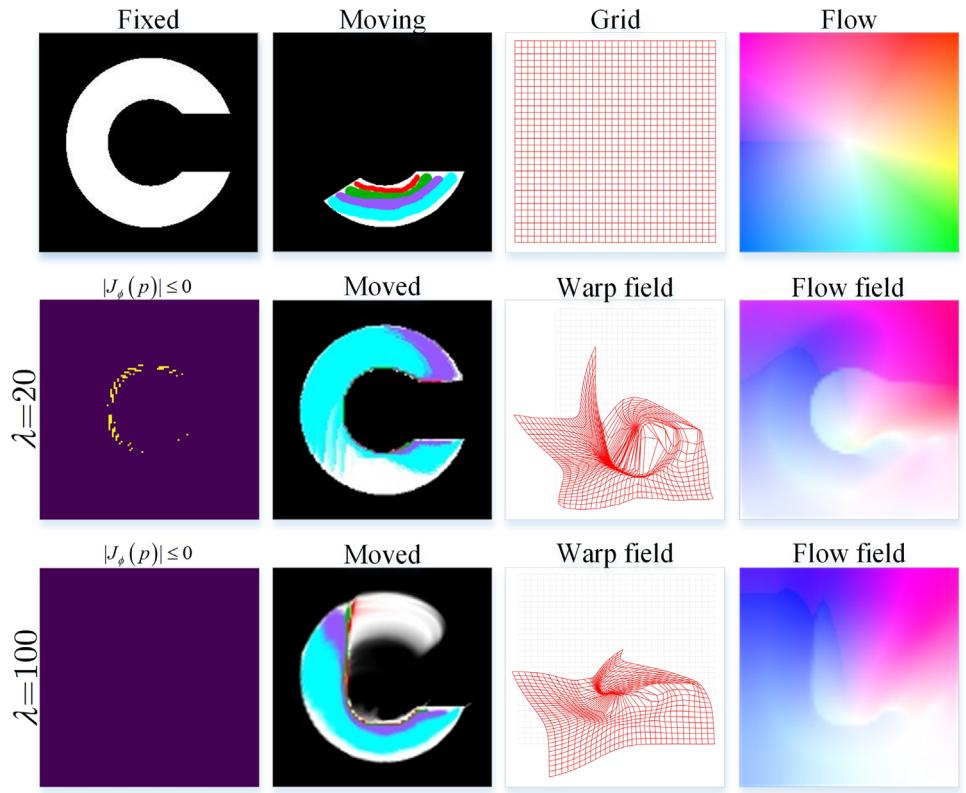


FIGURE 3 The visualization results reflect the effect between registration accuracy and diffeomorphic property in high ($\lambda = 100$) and low ($\lambda = 20$) smoothness regularization parameter in VoxelMorph. Flow fields are drawn by mapping the two components (x, y) in the deformation field into pre-defined flow whose direction is encoded with color and magnitude with color intensity, and the white position represents non-displacement

training. A higher regularization parameter in top cascades can avoid overlapping voxels when the deformation is large, so it improves the diffeomorphic property. A lower regularization parameter in bottom cascades gives the network greater flexibility, so it improves the registration accuracy. The last line shows the comprehensive results of our proposed methods. LDVoxelMorph has higher registration accuracy and better diffeomorphic property for large deformation registration. Compared with the VTN algorithm, although VTN has a higher dice accuracy, it has a large number of overlaps in the deformation field of the liver dataset, which does not meet the ability of the diffeomorphic property. Our network can achieve relatively high accuracy under the condition of satisfying the diffeomorphic property.

3.4 | Ablation study

3.4.1 | Experimental results on compressed loss

To observe the registration effect of the compressed loss, we performed experiments on VoxelMorph and LDVoxelMorph with different regulation parameters λ .

Due to the different value of regulation parameters required for different deformation datasets, we have designed different regularization parameters for experiments. The datasets contain larger scale deformation dataset SLIVER and smaller scale deformation dataset LPBA. We only evaluate registration results after three cascaded blocks output in LDVoxelMorph because of its high accuracy. The experiences were performed on the SLIVER and LBPA datasets, as shown in Table 2. The numbers in the bracket are the standard deviation among the test set.

With the introduction of compressed loss, the change in the number of overlapping voxels N was significant. The larger the Dice value and the smaller the N value, the better. As the scale of hyperparameters increased, N became smaller. However, there was a tendency to decrease in the Dice values. When CL was introduced in the VoxelMorph, N values were reduced to 3862 from 9678 and the Dice was slightly improved. With the introduction of the cascade and the number increased, the multilevel optimization with CL made the overlapping voxels values rapidly less. While keeping the regularization parameter of the second and third cascades constant, the N values gradually decreased as the increased regularization parameter of the first cascade. However, the experimental results show that the best results did

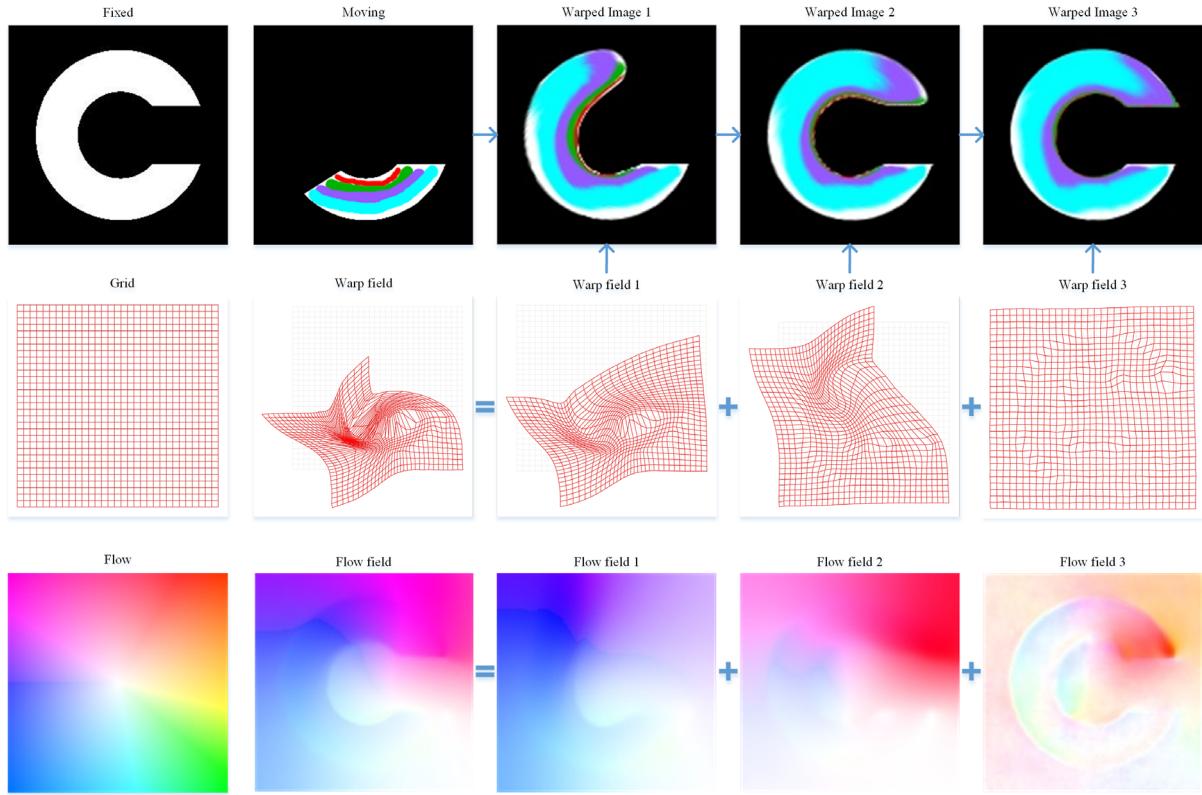


FIGURE 4 Visualization of an example in c-shape datasets. Our method achieves state-of-the-art result in both registration accuracy and diffeomorphic property

TABLE 1 Experiment results on medical image datasets

Method	SLIVER (liver)		LiTS (liver)		LPBA (brain)		Time (s)	
	Avg. Dice	N	Avg. Dice	N	Avg. Dice	N	GPU	CPU
ANTs SyN ³⁴	0.892 (0.041)	10780	0.836 (0.058)	7562	0.683 (0.157)	5426	—	748
VTN (ADDD) ²⁶	0.901 (0.042)	19430	0.856 (0.049)	13728	0.695 (0.132)	9620	0.65	56
VoxelMorph ⁹	0.880 (0.041)	9678	0.838 (0.048)	830	0.683 (0.015)	844	0.20	20
VoxelMorph ¹³ + CL	0.881 (0.040)	2549	0.837 (0.052)	278	0.684 (0.013)	165	0.20	20
VoxelMorph ¹³ + Cascades	0.890 (0.039)	594	0.847 (0.048)	294	0.689 (0.015)	82	0.38	26
LDVoxelMorph	0.892 (0.040)	325	0.848 (0.044)	159	0.689 (0.014)	0	0.38	26

Notes: The comparative experiments include a traditional method, such as ANTs SyN, our baseline network inspired by VoxelMorph and VTN, with and without the proposed compressed loss and cascaded architecture. For liver datasets (SLIVER and LiTS), the dice score measures the overlap of liver segmentation. For brain dataset (LBPA), Avg. Dice measures an average Dice score among all 56 segmented anatomical structures, and the number in the bracket means the standard deviation. N represents the average number of statistically overlapping voxels in the deformation field, which responses to diffeomorphic property (lower is better).

not benefit from larger parameters. The best N values were obtained from LDVoxelMorph on LPBA at $\lambda = 100, 50, 25$ and SLIVER at $\lambda = 600, 300, 100$ respectively.

3.4.2 | Experimental results on cascade components

To observe the registration effect of the cascade structure, the proposed LDVoxelMorph removed the

cascade structure and degraded it to a network consisting of the compressed loss and VoxelMorph. The experiences were performed on the SLIVER and LPBA datasets. Cascade-v1(U1 + U2, $\lambda = 50, 50$) denotes two stages in a U-shaped structure with the respective regularization parameters set to $\lambda = 50$ and $\lambda = 50$, respectively. The experimental results at different numbers of cascade structures are shown in Table 3. The meaning of the numbers in the bracket are the standard deviation among the test set. It can

TABLE 2 Experimental results on compressed loss and different scaling hyperparameters. The larger the Dice value and the smaller the N value, the better

Method	SLIVER (liver)		LPBA (brain)		Time (s)	
	Avg. Dice	N	Avg. Dice	N	GPU	CPU
VoxelMorph ⁹ +MSE and D_V	0.880 (0.041)	9678	0.683 (0.015)	844	0.24	20
VoxelMorph ⁹ + CL, $\lambda = 50$	0.881 (0.040)	3862	0.687 (0.015)	319	0.24	20
VoxelMorph ⁹ + CL, $\lambda = 100$	0.883 (0.040)	2938	0.688 (0.013)	165	0.24	20
VoxelMorph ⁹ + CL, $\lambda = 200$	0.881 (0.040)	2549	0.684 (0.015)	170	0.24	20
LDVoxelMorph (CL, $\lambda = 25, 50, 25$)	0.891 (0.042)	881	0.677 (0.016)	153	0.38	26
LDVoxelMorph (CL, $\lambda = 50, 50, 25$)	0.889 (0.041)	704	0.684 (0.013)	33	0.38	26
LDVoxelMorph (CL, $\lambda = 100, 50, 25$)	0.890 (0.040)	437	0.689 (0.014)	0	0.38	26
LDVoxelMorph (CL, $\lambda = 600, 300, 100$)	0.892 (0.040)	325	0.681 (0.015)	17	0.38	26

TABLE 3 Experimental results of the proposed algorithm with different cascade architectures. The larger the Dice value and the smaller the N value, the better

Method	SLIVER (liver)		LPBA (brain)		Time (s)	
	Dice	N	Dice	N	GPU	CPU
VoxelMorph ¹³ + CL, $\lambda = 50$	0.880 (0.040)	3862	0.687 (0.015)	319	0.24	20
Cascade-v1 (U1 + U2, $\lambda = 50, 50$)	0.875 (0.039)	1376	0.683 (0.015)	67	0.33	23
Cascade-v2 (U1 + U2, $\lambda = 100, 100$)	0.878 (0.039)	825	0.684 (0.014)	40	0.33	23
Cascade-v3 (U1 + U2 + U3, $\lambda = 50, 50, 50$)	0.888 (0.041)	709	0.685 (0.015)	36	0.38	26
Cascade-v4 (U1 + U2 + U3, $\lambda = 100, 100, 100$)	0.890 (0.040)	389	0.685 (0.014)	5	0.38	26

be found from Table 3 that as the number of cascades increases, the registration accuracy increases. We add a higher regularization parameter to the cascade-v2 and cascade-v4, which makes the number of overlapping voxels also decrease. As the cascade increases, the registration accuracy gradually improves.

3.4.3 | Hypothesis tests

To demonstrate that our proposed algorithm outperforms another statistically significantly, we do hypothesis tests on the difference of mean Dice score between VoxelMorph and our representative experiment results from Tables 1 and 2. We use t hypothesis test and hypothesize the mean Dice score of VoxelMorph is greater than our methods, which is unilateral hypothesis. The results of Dice score and its p -value based on the hypothesis are shown in Table 4. The meaning of the numbers in the bracket are the standard deviation among the test set. The third and fourth rows of Table 4 are evaluated from Table 2 which do not do the experiment on LiTS dataset. Since the regulation parameter group setting $\lambda = 600, 300, 100$ is also the experiment on last line of Table 1, we add hypothesis tests on LiTS experiments LDVoxelMorph (CL, $\lambda = 600, 300, 100$) from the last line

of Table 1. Therefore, LDVoxelMorph (CL, $\lambda = 600, 300, 100$) has LiTS evaluation while LDVoxelMorph (CL, $\lambda = 100, 50, 25$) does not.

On first line of the results, we can see that the p -value is greater than 0.05, which means that CL has little improvement on the registration accuracy. The second line shows the confidence of the improvement on cascade architecture. Because the p -value is less than 0.05 in SLIVER, LiTS, and LPBA datasets, we can infer that our cascade architecture outperforms the baseline model VoxelMorph on registration accuracy. The third line is the experiments results on LDVoxelMorph with regularization parameters $\lambda = 100, 50, 25$. Both p -values are less than 0.05, we can infer that the parameter setting $\lambda = 100, 50, 25$ can make LDVoxelMorph outperform VoxelMorph on accuracy. Besides, the last line shows the experiment on LDVoxelMorph with $\lambda = 600, 300, 100$. From p -value, we can infer that the parameter setting $\lambda = 600, 300, 100$ can make LDVoxelMorph outperform VoxelMorph on liver dataset while not on brain dataset. The reason for these results is the large regularization parameters are just suitable for the large deformed liver dataset while they are too large for the small deformed brain dataset. Therefore, these results show that LDVoxelMorph can improve registration accuracy under the premise of carefully setting the regularization parameter λ .

TABLE 4 The results of Dice score and its *p*-value between VoxelMorph and our method. The smaller the *p*-value, the better confidence for the improvement registration accuracy of our method

Method (vs. VoxelMorph + MSE and D_v)	SLIVER (liver)		LiTS (liver)		LPBA (brain)	
	Dice	<i>p</i> -value (<i>n</i> = 378)	Dice	<i>p</i> -value (<i>n</i> = 128)	Dice	<i>p</i> -value (<i>n</i> = 37)
VoxelMorph ¹³ + CL	0.881 (0.040)	0.367	0.837 (0.052)	0.563	0.684 (0.013)	0.380
VoxelMorph ¹³ + Cascades	0.890 (0.039)	3.12*e-4	0.847 (0.048)	0.0484	0.689 (0.015)	0.0448
LDVoxelMorph (CL, λ = 100, 50, 25) (0.040)	0.890	3.62*e-4	/	0.689 (0.014)	0.0398	
LDVoxelMorph (CL, λ = 600, 300, 100)	0.892 (0.040)	2.57*e-5	0.848 (0.044)	0.0418	0.681 (0.015)	0.716

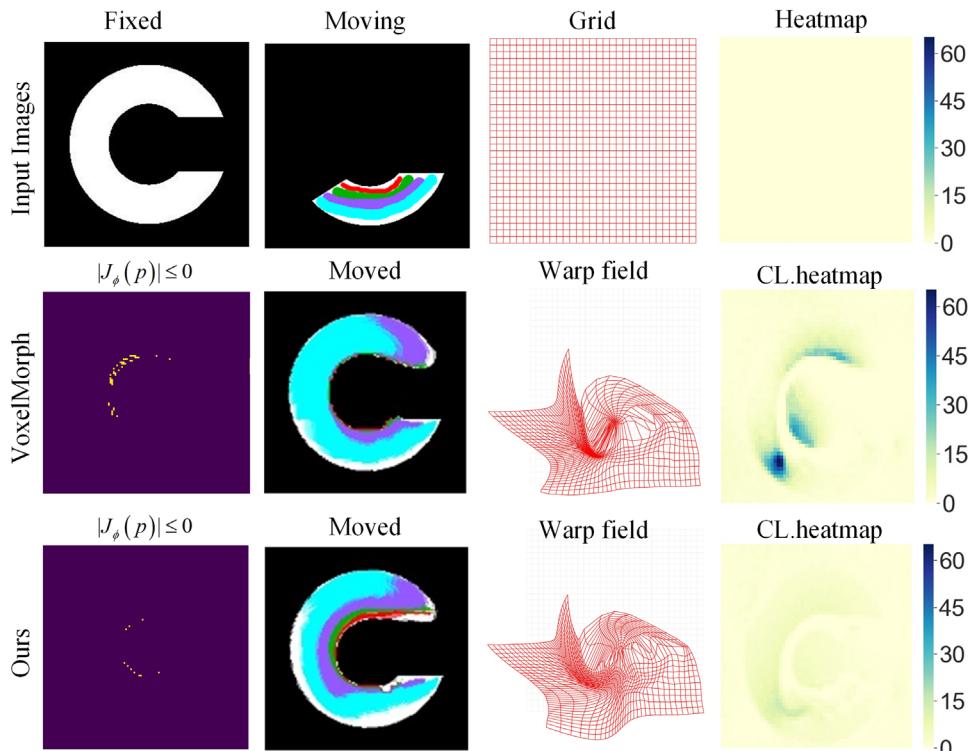


FIGURE 5 Visualization of an example in the c-shape image dataset in ablation study. CL heatmap is drawn by calculating compressed loss in every pixel based on velocity field

3.4.4 | Visualization results

To observe the registration effect of the compressed loss, the proposed LDVoxelMorph removed the cascade structure and degraded to a network that consists of the compressed loss and VoxelMorph. The experiments were performed on the c-shape and SLEVER datasets.

On the c-shape dataset, the representative testing results are shown in Figure 5. In first column, we observe that our method has fewer overlapping voxels. In second column, compared with the VoxelMorph, some rainbow colors are missing, and our method saves the rainbow colors more completely, which means diffeomorphic property preserves well. The last column shows compressed loss heatmap obtained by velocity field. Before

using CL, the VoxelMorph has a high value of CL and has many overlapping voxels in the area where CL is high. After using CL in VoxelMorph, the reduction of the CL value directly leads to the reduction of the number of overlapping voxels in the corresponding area.

On the SLIVER dataset, we visualized the effect of CL on the on the deformation field. The testing results are visualized and shown in Figure 6, the second column means the VoxelMorph without CL. The third column shows our method with CL and without cascade module. We can clearly see that the warp field generated by our method becomes smoother and has less grid extrusion in column 3. A reasonable explanation is that CL can predict the local degree of overlap in warp field through the velocity field generated by VoxelMorph

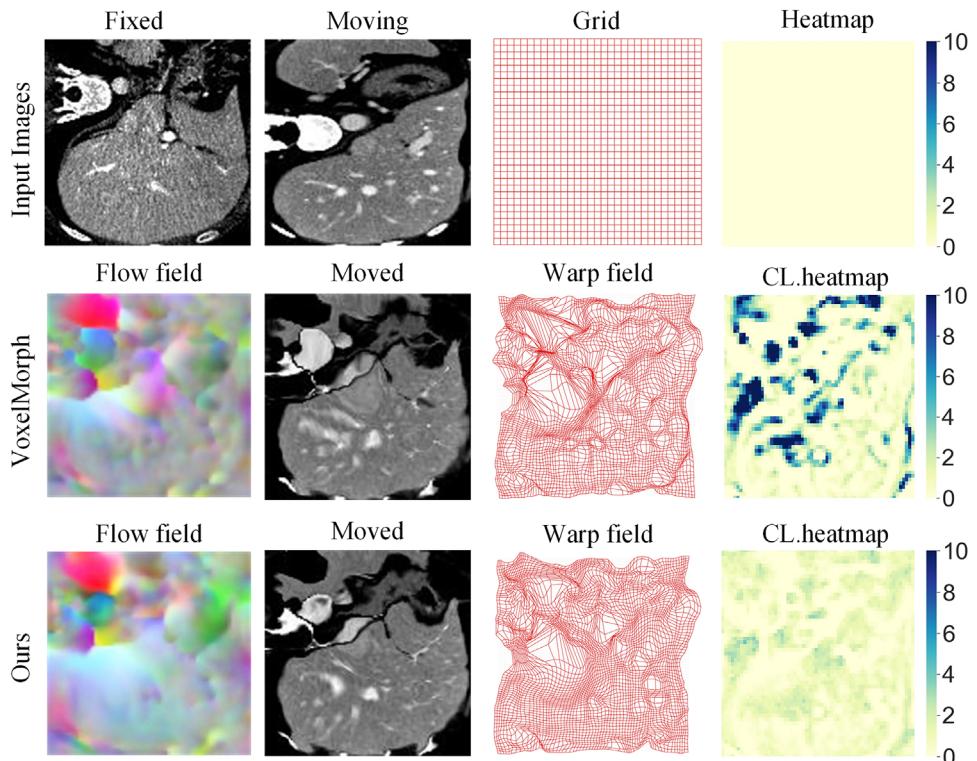


FIGURE 6 Visualization of an example in the SLIVER dataset in ablation experiment. Flow fields are drawn by mapping the three components (x, y, z) in deformation field into color channels (R, G, B), respectively. Yellow position in heatmap corresponds to no displacement

and add spatial variation loss based on degree of overlap, shown in column 4. In summary, compressed loss improves diffeomorphic property by limiting the large CL values.

4 | DISCUSSION

As shown in Table 1, the VoxelMorph⁹ can acquire the diffeomorphic property well for small deformation datasets such as brain MRI while it is confusing for large deformation datasets such as liver CT scans. On the LPBA dataset, we obtained same registration accuracy with FDRN⁴² which is a recent registration method and Dices are 0.684 (our) and 0.684 (FDRN), respectively. Under the premise of ensuring registration accuracy, our method produces fewer overlapping voxels. Therefore, our proposed networks and compressed loss can improve the diffeomorphic property for large deformation. At the same time, we also find some problems and improvement ideas worthy of further reflection in the analysis and experiment of the proposed method. In the liver CT scans registration experiment, since the training data contains the gallbladder tissue around the liver, unsupervised training methods will introduce errors, reducing the registration accuracy. The image features of the gallbladder are weird, so pixel-based registration will make the gallbladder misaligned. Our registration purpose is to register the liver, and the gallbladder exists

as an error. The larger the gallbladder difference in the image to be registered, the more the liver is registered, the greater the interference. Future research directions can consider adding auxiliary supervision signals to train the network to improve registration accuracy.

Affected by the U-Net++ network structure, the output of cascades is limited. Since the simple stacked cascaded network can also increase the cascaded output, the proposed network can be simply stacked to improve the registration accuracy if computing resources allow. In the function blocks of network, we use a fixed size convolution, which may be limited in feature extraction in cascaded non-rigid registration tasks. Since deformable convolution can adjust the convolution position according to the output feature map, it dynamically adjusts the position of convolution according to the deformation process of cascaded non-rigid registration, which can improve the disadvantages of fixed convolution to a certain extent. This is an excellent idea worthy of further study.

5 | CONCLUSION

We propose LDVoxelmorph, which incorporates a cascaded architecture and several backward-skip connections using a coarse-to-fine idea for large deformable registration. Besides, we propose a compressed loss that can precisely limit the local chaotic velocity field

generated by LDVoxelMorph, which is more efficient than spatial smoothness regularization. The experimental results show that our method achieves good performance both on registration accuracy and diffeomorphic property for large deformable registration.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant Nos. 62073257 and 61971343, the Key Research and Development Program of Shaanxi Province of China under Grant No. 2020GXLY-Y-008.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Myronenko A, Song X. Point set registration: coherent point drift. *IEEE Trans Pattern Anal Mach Intell.* 2010;32(12):2262-2275.
- Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. *Comput Vis Image Underst.* 2003;89(2-3):114-141.
- Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.
- Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: a survey. *IEEE Trans Med Imaging.* 2013;32(7):1153-1190.
- Du S, Xu G, Zhang S, et al. Robust rigid registration algorithm based on pointwise correspondence and correntropy. *Pattern Recognit Lett.* 2020;132:91-98.
- Yang Y, Fan D, Du S, et al. Point set registration with similarity and affine transformations based on bidirectional KMPE Loss. *IEEE Trans Cybern.* 2019;12(99):100-126.
- Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. An unsupervised learning model for deformable medical image registration. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018:9252-9260.
- Dalca AV, Balakrishnan G, Guttag J, Sabuncu MR. Unsupervised learning for fast probabilistic diffeomorphic registration. *International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham;* 2018:729-738.
- Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca A. Voxel-morph: a learning framework for deformable medical image registration. *IEEE Trans on Med Imaging.* 2019;38(8):1788-1800.
- Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[J]. In: *Advances in Neural Information Processing Systems.* 2015:28.
- Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage.* 2007;38(1):95-113.
- Mok TCW, Chung ACS. Fast symmetric diffeomorphic image registration with convolutional neural networks. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition; 2020:4644-4653.
- Zhao SY, Dong Y, Eric I-C C, Xu Y. Recursive cascaded networks for unsupervised medical image registration. Proceedings of the IEEE International Conference on Computer Vision; 2019:10600-10610.
- Modersitzki J. *Numerical Methods for Image Registration.* Oxford University Press; 2004.
- Kuang D, Schmah T. FAIM – a convNet method for unsupervised 3D medical image registration. *arXiv:* 181109243, 2018.
- Zhang J. Inverse-consistent deep networks for unsupervised deformable image registration. *arXiv:180903443,* 2018.
- Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labelling of elderly and neurodegenerative brain. *Med Image Anal.* 2008;12(1):26-41.
- Faisal B M, Miller MI, Trouve A, Younes L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int J Comput Vis.* 2005;61(2):139-157.
- Ashburner J, Friston KJ. Voxel-based morphometry the methods. *Neuroimage.* 2000;11(6):805-821.
- Ruzena B, Stane KC. Multiresolution elastic matching. *Comput Vis Graphics Image Process.* 1989;46(1):1-21.
- Shen D, Davatzikos C. Hammer: hierarchical attribute matching mechanism for elastic registration. Proceedings of IEEE Workshop on Mathematical Methods in Biomedical Image Analysis; 2001:29-36.
- Davatzikos C. Spatial transformation and registration of brain images using elastically deformable models. *Comput Vis Image Underst.* 1997;66(2):207-222.
- Glocker B, Komodakis N, Tziritas G, Nassir N, Nikos P. Dense image registration through mrfs and efficient linear programming. *Med Image Anal.* 2008;12(6):731-741.
- Besl PJ, McKay ND. Method for registration of 3-d shapes. In sensor fusion IV: control paradigms and data structures. *Int Soc for Opt Photonics.* 1992;1611:586-607.
- Zhang Z. Iterative point matching for registration of free-form curves and surfaces. *Int J Comput Vis.* 1994;13(2):119-152.
- Lau T, Luo J, Zhao S, et al. Unsupervised 3d end-to-end medical image registration with volume tweening network. *arXiv:1902.05020,* 2019.
- Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. U-Net++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Medical Imaging.* 2019;39(6):1856-1867.
- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015:3431-3440.
- Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer; 2015:234-241.
- Stefanescu R, Pennec X, Ayache N. Grid powered nonlinear image registration with locally adaptive regularization. *Med Image Anal.* 2004;8(3):325-342.
- Risser L, Vialard FX, Baluwala HY, Schnabel JA. Piecewise-diffeomorphic image registration: application to the motion estimation between 3D CT lung images with sliding conditions. *Med Image Anal.* 2013;17(2):182-193.
- Pace DF, Aylward SR, Niethammer M. A locally adaptive regularization based on anisotropic diffusion for deformable image registration of sliding organs. *IEEE Trans Med Imaging.* 2013;32(11):2114-2126.
- Vialard FX, Risser L. Spatially-varying metric learning for diffeomorphic image registration: a variational framework. International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer; 2014:227-234.
- Niethammer M, Kwitt R, Vialard FX. Metric learning for image registration. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2019:8463-8472.
- Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). *Or insight,* 2009;2(365):1-35.
- MSD Medical segmentation decathlon, 2018. <https://decathlon-10.grand-challenge.org/>. Accessed June 2021
- Heimann T, Van Ginneken B, Styner MA, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imaging.* 2009;28(8):1251-1265.
- LiTS. Liver tumor segmentation challenge, 2018. <https://competitions.codalab.org/competitions/15595>. Accessed June 2021

39. Bellec P, Chu C, Chouinard-Decorte F, Benhajali Y, Margulies DS, Cameron CR. The neuro bureau adhd-200 preprocessed repository. *Neuroimage*. 2017;144:275-286.
40. Besl PJ, McKay ND. Method for registration of 3-d shapes. *IEEE Trans Pattern Anal Mach Intell*. 1992;14(2):586-607.
41. Shattuck DW, Mirza M, Adisetiyo V, et al. Construction of a 3d probabilistic atlas of human cortical structures. *Neuroimage*. 2008;39(3):1064-1080.
42. Sun K, Simon S. FDRN: a fast deformable registration network for medical images. *Medical Physics*. 2021;48:6453-6463. <http://doi.org/10.1002/mp.15011>

How to cite this article: Yang J, Wu Y, Zhang D, et al. LDVoxelMorph: A precise loss function and cascaded architecture for unsupervised diffeomorphic large displacement registration. *Med Phys*. 2022;49:2427–2441.
<https://doi.org/10.1002/mp.15515>