Applied Mathematical Modelling 151 (2026) 116506

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

A hybrid L-BFGS and Gauss–Newton approach for multimodal diffeomorphic image registration

Yanyan Li [a], Daoping Zhang [b], Jianping Zhang [c,*]

[a] *School of Science, Hunan University of Technology, Zhuzhou, 412007, Hunan, China*
[b] *School of Mathematical Sciences and LPMC, Nankai University, Tianjin, 300071, Tianjin, China*
[c] *School of Mathematics and Computational Science, Xiangtan University, Xiangtan, 411105, Hunan, China*

ARTICLE INFO

Keywords:
Multimodal image registration
Jacobian determinant
Diffeomorphic mappings
L-BFGS method
Gauss–Newton method

ABSTRACT

Multimodal image registration involves aligning images derived from various modalities. Accurately obtaining a diffeomorphic mapping is challenging due to the distinct features these images present. In this work, we propose a bi-variant multimodal diffeomorphic image registration model with flexible deformation. We begin by applying the constraint of the Jacobian determinant, $\det(I_d + \nabla \boldsymbol{u}(\boldsymbol{x})) = v(\boldsymbol{x})$, where $v(\boldsymbol{x}) > 0$ is a flexible function, to multimodal image registration. This approach ensures the accurate alignment of images from different modalities while preserving diffeomorphic mappings. A penalty-splitting method models the constraint as an unconstrained optimization problem, which we solve using a discretize-then-optimize technique and a hybrid optimization strategy. To optimize the variable $\boldsymbol{u}(\boldsymbol{x})$ for the mutual information similarity measure, where Hessian computation is computationally expensive, we employ the L-BFGS algorithm combined with the Armijo line search. For the variable $v(\boldsymbol{x})$, where Hessian is easier to compute, we apply the Gauss–Newton method with the Armijo line search. Numerical results on 2D and 3D medical and synthetic images from different modalities demonstrate the flexibility and effectiveness of our approach.

1. Introduction

Image registration plays a crucial role in the medical imaging community. One of the key challenges arises when registering images acquired from different imaging devices or modalities. The challenge lies in the complex relationship between gray values in multimodal images, which lacks a direct functional connection. Despite this, gray-value patterns are generally not arbitrary. This observation has motivated the development of numerous similarity measures for multimodal image registration [1]. Among these, mutual information (MI) [2,3] is one of the earliest and significant measure due to its ability to capture statistical dependence between image intensities, though it is affected by the extent of image overlap. Normalized mutual information (NMI) [4] was subsequently developed to overcome this limitation and has gained widespread use. In addition to intensity statistics, normalized gradient fields (NGF) [5] utilize the fact that intensity changes across modalities are often aligned, making gradient information a robust descriptor. Correlation-based measures, like cross-correlation (CC) [6], offer simplicity but are sensitive to scaling; normalized cross-correlation (NCC) [7] addresses contrast variation. Although many similarity measures exist, MI and its variants remain the most widely adopted and commonly used in multimodal image registration.

Introduced by Collignon et al. [2] and Viola et al. [3] in 1995, mutual information serves as an effective similarity measure for multimodal image registration. Collignon et al. implemented it in an information-theoretic framework for 3D medical image

---

[*] Corresponding author.
E-mail address: jpzhang@xtu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.apm.2025.116506
Received 22 May 2025; Received in revised form 18 September 2025; Accepted 13 October 2025

registration, whereas Viola et al. applied it to determine object pose in images, remaining unaffected by illumination changes and not requiring prior object surface data. Furthermore, this measure is applicable to parameter transformations in lower-dimensional spaces, such as rigid or affine types [3,8,9]. However, for applications requiring local morphology analysis, affine or rigid transformations alone are no longer sufficient. To address this, several methods have been proposed for nonrigid registration using mutual information, differing in how the deformation is represented and how mutual information varies with the deformation parameters. A common approach is to represent the deformation using smooth basis functions, often with a global or local radial basis function. For example, Meyer et al. [10] used affine and thin-plate spline deformations to demonstrate the accuracy and clinical applicability of mutual information for automatic multimodal image fusion. Later, they [11] employed thin-plate splines to model nonlinear deformations while maximizing mutual information for volume registration into an atlas. Rueckert et al. [12] proposed a nonrigid registration method for breast magnetic resonance (MR) imaging that uses normalized mutual information as the similarity measure. Rohde et al. [13] utilized compactly supported radial basis functions to estimate local deformations. However, a few drawbacks remain: these methods often require manual or semi-automatic landmark point marking when modeling deformation, which can result in unrealistic deformations if few landmark points are used, and they incur high computational complexity as the number of landmark points increases.

Later, Gaens et al. [14] introduced a fully automated non-rigid image registration method based on a mutual information measure. They employed nonparametric deformations and assigned a local deformation to each voxel. Hermosillo et al. [15] proposed a variational method for multimodal image registration, which defines the forces driving the deformation of each voxel in order to maximize mutual information. This approach incorporates linear elastic regularization [16,17] to constrain the deformation. Periaswamy et al. [18] developed a general-purpose image/volume registration technique based on both global and local variations in image intensities. However, the elastic registration is only effective for small deformations and does not necessarily result in a diffeomorphic transformation. In certain applications, large local deformations may need to be recovered. To address this issue, Agostino et al. [19] extended Hermosillo's approach by replacing the elastic regularization term with Christensen's [20] viscous fluid regularization term. They proposed a multimodal registration model based on viscous fluids, which requires regridding the deformation during registration to ensure a positive Jacobian determinant of the transformation throughout, allowing the method to handle large deformations. Kroon et al. [21] suggested using joint histogram peaks to transform a T1 scan into a T2 scan, enabling fast intensity-based local image registration methods such as Demons registration. However, both the viscous fluid and Demons registrations are heuristic and non-physical, lacking a formal optimization foundation. Lu et al. [22] described a diffeomorphic registration method using point-wise integrated mutual information in the Demons framework, which allows for multimodal registration. They subsequently proposed a variational approach for multimodal image registration based on the diffeomorphic Demons algorithm. Lam et al. [23] later built on the mutual transformation [21] to measure the similarity between images with different modalities, proposing a diffeomorphic hybrid multimodal registration method capable of handling large deformations.

Recently, several methods have been developed for multimodal image registration. Cao et al. [24] introduced a bi-directional image synthesis-based approach for computed tomography (CT) and magnetic resonance imaging (MRI) registration, enhancing the accuracy of prostate cancer radiation therapy. Simonovsky et al. [25] proposed a deep convolutional neural network metric for multimodal registration, which, when combined with a continuous optimization framework, significantly outperformed mutual information-based methods in T1- and T2-weighted MR image registration. Oktay et al. [26] developed a probabilistic edge map using a structure decision forest for multimodal ultrasound image registration. While this learning-based method relies on manually engineered features and requires separate training, it limits broader applicability. Ceranka et al. [27] compared various registration methods for whole-body MR image mosaicing and functional-to-anatomical whole-body image registration in a multimodal context. Li et al. [28] proposed a two-step (coarse and fine) registration approach for multimodal retinal image registration, specifically for color and scanning laser ophthalmoscope (SLO) images, addressing alignment challenges between different modalities.

Optimization plays a crucial role in image registration. To address large deformation diffeomorphic image registration, Andreas et al. [29] introduced a preconditioned, globalized, matrix-free, inexact Newton-Krylov method for numerical optimization. They analyzed the spectral properties of the Hessian, grid convergence, numerical accuracy, computational efficiency, and deformation regularity of the proposed scheme. Additionally, they compared the Newton-Krylov method with a globalized Picard method (preconditioned gradient descent). Subsequently, they [30] developed a two-level preconditioned, globalized, matrix-free Newton-Krylov scheme for solving a stationary velocity field. They explored two approaches for inverting the preconditioner on the coarse grid: a nested preconditioned conjugate gradient method (exact solve) and a nested Chebyshev iterative method (inexact solve) with a fixed number of iterations. Meanwhile, Andreas et al. [31] proposed a method to eliminate the partial differential equation (PDE) constraint using a Lagrangian hyperbolic PDE solver, which approximates characteristic curves with a fourth-order Runge-Kutta method. They also introduced an efficient algorithm for computing the derivatives of the system's final state with respect to the velocity field, enabling fast Gauss–Newton methods with spectral preconditioning. Their approach is integrated into the flexible algorithms for image registration (FAIR) framework [1]. Furthermore, Hernandez [32] proposed a novel preconditioned optimization method within the large deformation diffeomorphic metric mapping (LDDMM) framework. This method uses preconditioning matrices inspired by the Hessian approximation employed in the Gauss–Newton method, with derivatives computed using Fréchet differentials. Ruthotto et al. [33] introduced a stable multigrid solver for hyperelastic image alignment, employing a Galerkin-multigrid scheme to discretize and solve the Hessian system. They also presented a stability technique by setting a threshold for the coefficients.

However, many of the optimization methods mentioned above are primarily used for image registration problems where the similarity measure is the sum of squared differences. The challenge in optimizing mutual information lies in the high computational complexity involved. Hermosillo et al. [15] presented a formal calculation of variational gradients for statistical similarity measures and extended an optical flow algorithm to the multimodal image registration case. Henn et al. [34] introduced an iterative method to minimize distance functionals by explicitly discretizing a classical gradient flow approach. Additionally, they employed a multigrid method and a multiresolution minimization strategy to compute an approximate solution and derive a suitable initial guess. Yang et al. [35] proposed a hybrid approach combining the limited memory Broyden-Fletcher-Goldfarb-Shanno with boundaries (L-BFGS-B) and cat swarm optimization for multimodal image registration, using the normalized mutual information measure and the free-form deformations model.

Numerous studies have addressed multimodal image registration, many existing models encounter difficulties with large deformations and commonly fail to ensure diffeomorphic mappings, which are critical for preserving anatomical topology in medical applications. Moreover, mutual information, a widely used similarity metric, involves a complex function requiring the calculation of marginal and joint probabilities across all pixels, resulting in high computational expenses during optimization. These limitations motivate the development of a multimodal diffeomorphic image registration model capable of handling large deformations while guaranteeing topology conservation via diffeomorphic mappings. Thus, our work is therefore to design such a registration method alongside an effective optimization strategy to enhance both the accuracy and computational efficiency of multimodal registration.

Motivated by this, we propose a multimodal diffeomorphic image registration model constrained by the Jacobian determinant. Additionally, we provide an efficient solver for this model. It is particularly effective for CT, T1-, T2-, and PD-weighted MR images, and for handling images with large deformations. Our key contributions are as follows.

1. We introduce a bi-variant multimodal diffeomorphic image registration model with flexible deformation. By incorporating the constraint of the Jacobian determinant, $\det(I_d + \nabla \boldsymbol{u}(\boldsymbol{x})) = v(\boldsymbol{x})$, where $v(\boldsymbol{x}) > 0$ is a flexible and optimizable function function, our model enables accurate alignment of images from different modalities while maintaining diffeomorphic mappings.
2. We employ a penalty-splitting method to address the bi-variant variational model with equality constraints, reformulating it as an unconstrained optimization problem. Then it is solved using a Ifirst-discretize-then-optimize methodology, which allows the use of standard optimization techniques to tackle the highly non-convex and non-linear mutual information measure.
3. We apply the L-BFGS method with Armijo line search to optimize the variable $\boldsymbol{u}(\boldsymbol{x})$ in the problem containing mutual information measures, and the Gauss–Newton method with Armijo line search to optimize the variable $v(\boldsymbol{x})$. Numerical experiments demonstrate that our algorithm is convergent and effectively handles images from different modalities and those with large deformations, while preserving diffeomorphic mapping.

The remainder of this paper is structured as follows. In Section 2, we provide a brief overview of the basic mathematical formulation of the mutual information measure. In Section 3, we introduce a novel multimodal image registration model. To solve the corresponding variational model, we discuss effective discretization and numerical schemes in Section 4. Experimental results involving CT, T1-, T2-, and PD-weighted MR medical images, as well as synthetic images, are presented in Section 5. Finally, we conclude the paper in Section 6.

2. Preliminaries

Let $R(\boldsymbol{x})$ and $T(\boldsymbol{x})$ represent reference and template images from different modalities, with $\boldsymbol{x} \in \Omega \subset \mathbb{R}^d$, where d is either 2 or 3. Image registration involves determining a spatial transformation $\boldsymbol{\varphi}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})$, associated with the displacement field $\boldsymbol{u}(\boldsymbol{x})$, to effectively align $T(\boldsymbol{\varphi}(\boldsymbol{x})) = T(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}))$ with $R(\boldsymbol{x})$. Performing accurate calculations of displacement $\boldsymbol{u}(\boldsymbol{x})$ requires solving an optimization problem with an appropriate similarity measure. For images from different modalities, negative mutual information serves to assess similarity, with higher values indicating greater similarity [36].

Our model considers that the pixel values of the reference and template images are treated as independent random variables. The mutual information between these variables is formulated as follows

$$\mathrm{MI}(R, T; \boldsymbol{u}) = -\int_{\mathcal{R}} p_{[R]}(r) \log p_{[R]}(r) \mathrm{d}r - \int_{\mathcal{T}} p_{[T;\boldsymbol{u}]}(t) \log p_{[T;\boldsymbol{u}]}(t) \mathrm{d}t + \int_{\mathcal{R}} \int_{\mathcal{T}} p_{[R,T;\boldsymbol{u}]}(r,t) \log p_{[R,T;\boldsymbol{u}]}(r,t) \mathrm{d}t \mathrm{d}r, \tag{1}$$

where $\mathcal{R} := [\min_{\boldsymbol{x}\in\Omega} R(\boldsymbol{x}), \max_{\boldsymbol{x}\in\Omega} R(\boldsymbol{x})]$ and $\mathcal{T} := [\min_{\boldsymbol{x}\in\Omega} T(\boldsymbol{x}), \max_{\boldsymbol{x}\in\Omega} T(\boldsymbol{x})]$ represent the sampling domains for images R and T. The joint probability distribution for R and T is defined as

$$I p_{[R,T;\boldsymbol{u}]}(r,t) = \frac{1}{|\Omega|} \int_{\Omega} \kappa_{\sigma_r}\big(r - R(\boldsymbol{x})\big) \kappa_{[\sigma_t;\boldsymbol{u}]}\big(t - T(\boldsymbol{x} + \boldsymbol{u})\big) \mathrm{d}\boldsymbol{x}. \tag{2}$$

The marginal probability distributions for $R(\boldsymbol{x})$ and $T(\boldsymbol{x})$ are accordingly

$$p_{[R]}(r) = \int_{\mathcal{T}} p_{[R,T;\boldsymbol{u}]}(r,t) \mathrm{d}t, \quad \text{and} \quad p_{[T;\boldsymbol{u}]}(t) = \int_{\mathcal{R}} p_{[R,T;\boldsymbol{u}]}(r,t) \mathrm{d}r.$$

The above joint distribution can be estimated using the Parzen-window method [37], with a kernel $\kappa(x)$ satisfying $\int_{\mathbb{R}} \kappa(x) dx = 1$ and $\kappa_\sigma(x) = \frac{1}{\sigma}\kappa(\frac{x}{\sigma})$, where σ denotes the window width. A limited support kernel reduces computational cost, unlike kernels with infinite support [38]. In image registration, using a cubic spline kernel in the Parzen-window method provides smooth and accurate density estimation with compact support, leading to robust mutual information measures and stable optimization. Therefore, our estimation utilizes a cubic spline kernel formulated by

$$\kappa(x) = \begin{cases} \frac{1}{2}|x|^3 - |x|^2 + \frac{2}{3}, & \text{if } 0 \le |x| < 1, \\ -\frac{1}{6}|x|^3 + |x|^2 - 2|x| + \frac{4}{3}, & \text{if } 1 \le |x| \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

Mutual information (1) serves as a metric to evaluate the statistical dependence or redundancy of the information between the intensity values of the corresponding voxels across the images. The geometric alignment of the images is assumed to maximize this statistical dependence [39], necessitating the maximization of $\mathrm{MI}(R, T; \boldsymbol{u})$, or equivalently, the minimization of $\mathcal{D}[R, T; \boldsymbol{u}]$ given by

$$\mathcal{D}[R,T;\boldsymbol{u}] = \int_{\mathcal{R}} p_{[R]}(r)\log p_{[R]}(r)\mathrm{d}r + \int_{\mathcal{T}} p_{[T;\boldsymbol{u}]}(t)\log p_{[T;\boldsymbol{u}]}(t)\mathrm{d}t - \int_{\mathcal{R}}\int_{\mathcal{T}} p_{[R,T;\boldsymbol{u}]}(r,t)\log p_{[R,T;\boldsymbol{u}]}(r,t)\mathrm{d}t\mathrm{d}r. \tag{3}$$

Direct minimization of (3) can lead to ill-posed problems [37,40]. To address this, a regularizer is introduced to penalize undesirable and irregular solutions. Thus, the image registration problem is formulated as

$$\min_{\boldsymbol{u}} \left\{ \mathcal{D}[R,T;\boldsymbol{u}] + \alpha \mathcal{S}[\boldsymbol{u}] \right\},$$

where the regularization parameter $\alpha > 0$ balances the terms $\mathcal{D}$ and $\mathcal{S}$, the diffusion regularizer is given by $\mathcal{S}[\boldsymbol{u}] = \frac{1}{2}\int_\Omega \|\nabla \boldsymbol{u}\|^2 \mathrm{d}\boldsymbol{x}$. For more on classical regularization terms, we refer the reader to [41–47].

3. The proposed multimodal diffeomorphic registration

In this work, a flexible multimodal diffeomorphic image registration model is introduced. It incorporates the constraint $\det(I_d + \nabla \boldsymbol{u}(\boldsymbol{x})) = v(\boldsymbol{x}) > 0$ into a multimodal registration that ensures accurate alignment with preserved diffeomorphic properties, which is formulated as follows

$$\begin{aligned} &\min_{\boldsymbol{u}\in\mathcal{U},0<v\in L^2(\Omega)} \left\{ \mathcal{D}[R,T;\boldsymbol{u}] + \alpha \mathcal{S}[\boldsymbol{u}] \right\}, \\ &\text{s.t.} \quad \det(I_d + \nabla \boldsymbol{u}(\boldsymbol{x})) = v(\boldsymbol{x}), \quad I\forall\, \boldsymbol{x} \in \Omega, \end{aligned} \tag{4}$$

where $\mathcal{U} := \left\{ \boldsymbol{u} \mid \boldsymbol{u} \in [\mathcal{H}^1(\Omega)]^d \text{ and } \boldsymbol{u}_{\partial\Omega} = \boldsymbol{0} \right\}$ represents a subspace of the Sobolev space $[\mathcal{H}^1(\Omega)]^d$.

We observe from the model (4) that ensuring a positive Jacobian determinant is equivalent to maintaining $v(\boldsymbol{x}) > 0$. However, guaranteeing this during optimization is challenging. To address this, we propose incorporating a restriction term $\Psi[v] := \int_\Omega \psi[v(\boldsymbol{x})]\,\mathrm{d}\boldsymbol{x}$ into (4) to facilitate automatic optimization of $v(\boldsymbol{x})$. The reformulated multimodal registration model becomes

$$\begin{aligned} &\min_{\boldsymbol{u}\in\mathcal{U},v\in L^2} \left\{ \mathcal{D}[R,T;\boldsymbol{u}] + \alpha \mathcal{S}[\boldsymbol{u}] + \beta\Psi[v] \right\}, \\ &\text{s.t.} \quad \det(I_d + \nabla \boldsymbol{u}(\boldsymbol{x})) = v(\boldsymbol{x}), \quad I\forall\, \boldsymbol{x} \in \Omega. \end{aligned} \tag{5}$$

where $\beta > 0$ is the penalty parameter. The restriction term $\Psi[v]$ prevents transformation folds by ensuring $v(\boldsymbol{x}) > 0$. Thus, the choice or design of the control function $\psi[\cdot]$ is crucial. In our recent work [48], we proposed three choices that not only guarantee $v(\boldsymbol{x}) > 0$, but also allow the control term $\psi[\cdot]$ to reach its minimum value when $v(\boldsymbol{x}) = 1$, thus achieving volume-preserving deformation on average. Among these, Eq. (6) was shown to outperform the other two in terms of registration accuracy and robustness. Therefore, in this work, we focus only on control function

$$\psi[v(\boldsymbol{x})] = \begin{cases} \frac{(v(\boldsymbol{x})-1)^2}{v(\boldsymbol{x})}, & \text{if } v(\boldsymbol{x}) > 0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{6}$$

To enhance the smoothness of the relaxation function $v(\boldsymbol{x})$ and thereby refine the model solution for (5), a diffusion regularizer for $v(\boldsymbol{x})$ is incorporated. This leads to an advanced registration model that efficiently manages multimodal images with large deformations while preserving diffeomorphic mappings:

$$\begin{aligned} &\min_{\boldsymbol{u}\in\mathcal{U},v\in\mathcal{V}} \left\{ \mathcal{D}[R,T;\boldsymbol{u}] + \alpha \mathcal{S}[\boldsymbol{u}] + \beta\Psi[v] + \gamma\mathcal{S}[v] \right\}, \\ &\text{s.t.} \quad \det(I_d + \nabla \boldsymbol{u}(\boldsymbol{x})) = v(\boldsymbol{x}), \quad I\forall\, \boldsymbol{x} \in \Omega, \end{aligned} \tag{7}$$

where $\mathcal{S}[v] = \frac{1}{2}\int_\Omega \|\nabla v(\boldsymbol{x})\|^2\,\mathrm{d}\boldsymbol{x}$, and $\mathcal{V} := \left\{ v \mid v \in \mathcal{H}^1(\Omega), v_{\partial\Omega} = 1 \right\}$, with $\beta > 0$ and $\gamma > 0$ controlling the smoothness and negativity of $v(\boldsymbol{x})$.

4. Numerical implementation

In this section, we provide a numerical computational strategy to address the variational model (7). Initially, we use the penalty method to transform the proposed variational model into an unconstrained optimization framework, subsequently dividing it into subproblems for $\boldsymbol{u}(\boldsymbol{x})$ and $v(\boldsymbol{x})$. The first-discretize-then-optimize methodology is adopted, where subproblems for $\boldsymbol{u}(\boldsymbol{x})$ and $v(\boldsymbol{x})$ are discretized prior to applying the L-BFGS and Gauss–Newton methods for resolution.

4.1. Optimization

The penalty method is a classical method for solving the equality-constrained optimization problem, which adds a penalty term to the objective functional and penalizes violation of the equality constraints. Therefore, the multimodal registration model with equality constraints (7) can be reformulated as the following unconstrained minimization problem

$$\min_{\boldsymbol{u}\in\mathcal{U},v\in\mathcal{V}}\left\{\mathcal{L}_{\lambda}[\boldsymbol{u},v]:=\mathcal{D}[R,T;\boldsymbol{u}]+\alpha\mathcal{S}[\boldsymbol{u}]+\beta\Psi[v]+\gamma\mathcal{S}[v]+\lambda\mathcal{C}[\boldsymbol{u},v]\right\},\tag{8}$$

where $\mathcal{C}[\boldsymbol{u},v]:=\frac{1}{2}\int_{\Omega}\left(\det(I_d+\nabla\boldsymbol{u}(\boldsymbol{x}))-v(\boldsymbol{x})\right)^2\mathrm{d}\boldsymbol{x}$ and $\lambda>0$ is a positive penalty parameter.

The functional (8) involves two optimizing variables, $\boldsymbol{u}(\boldsymbol{x})$ and $v(\boldsymbol{x})$. For simplicity, we split it into two subproblems:

- $\boldsymbol{u}$-subproblem

$$\min_{\boldsymbol{u}\in\mathcal{U}}\left\{\mathcal{L}^1_{\lambda,v}[\boldsymbol{u}]:=\mathcal{D}[R,T;\boldsymbol{u}]+\alpha\mathcal{S}[\boldsymbol{u}]+\lambda\mathcal{C}[\boldsymbol{u},v]\right\},\tag{9}$$

- v-subproblem

$$\min_{v\in\mathcal{V}}\left\{\mathcal{L}^2_{\lambda,\boldsymbol{u}}[v]:=\beta\Psi[v]+\gamma\mathcal{S}[v]+\lambda\mathcal{C}[\boldsymbol{u},v]\right\}.\tag{10}$$

An alternating iterative method is employed to solve these subproblems. Given known values $(\boldsymbol{u}^k,v^k,\lambda^k)$ in the k-th iteration, we update $(\boldsymbol{u}^{k+1},v^{k+1},\lambda^{k+1})$ as:

$$\begin{cases}\boldsymbol{u}^{k+1}=\arg\min\limits_{\boldsymbol{u}\in\mathcal{U}}\mathcal{L}^1_{\lambda^k,v^k}[\boldsymbol{u}]\quad=\arg\min\limits_{\boldsymbol{u}\in\mathcal{U}}\left\{\mathcal{D}[R,T;\boldsymbol{u}]+\alpha\mathcal{S}[\boldsymbol{u}]+\lambda^k\mathcal{C}[\boldsymbol{u},v^k]\right\},\\ v^{k+1}=\arg\min\limits_{\boldsymbol{v}\in\mathcal{V}}\mathcal{L}^2_{\lambda^k,\boldsymbol{u}^{k+1}}[v]=\arg\min\limits_{v\in\mathcal{V}}\left\{\beta\Psi[v]+\gamma\mathcal{S}[v]+\lambda^k\mathcal{C}[\boldsymbol{u}^{k+1},v]\right\},\\ \lambda^{k+1}=\rho\lambda^k,\end{cases}\tag{11}$$

where $\rho>1$ denotes the coefficient of increase. Updating λ enhances constraint enforcement, leading solutions of (8) to meet the optimization condition in (7).

4.2. Discretizations

In this section, we concentrate solely on the discretization of the 2D multimodal registration problem, with the 3D case being analogous and specific differences highlighted in the remarks. To present, we apply finite difference discretization to the problem (11) in the image domain $\Omega:=[0,1]\times[0,1]$. The discretized pixel location denoted by $\boldsymbol{x}$, along with the displacement field $\boldsymbol{u}(\boldsymbol{x})$ and the relaxation function $v(\boldsymbol{x})$, are centered on the grid. The discrete domain is thus expressed as:

$$\Omega_h:=\left\{\boldsymbol{x}_{ij}\in\Omega\mid\boldsymbol{x}_{ij}=(x^1_i,x^2_j)=((i-0.5)h,(j-0.5)h),i,j=1,\dots,n,h=1/n\right\}.$$

In this setting, the discretized displacement $\boldsymbol{u}_{ij}$ and the relaxation function v_{ij} are denoted as $\boldsymbol{u}_{ij}=(u^1_{ij},u^2_{ij})=(u^1(x^1_i,x^2_j),u^2(x^1_i,x^2_j))$ and $v_{ij}=v(x^1_i,x^2_j)$. Using a lexicographical order, we arrange

$$X=\begin{pmatrix}X_1\\X_2\end{pmatrix}=\left(x^1_1,\dots,x^1_1,\dots,x^1_n,\dots,x^1_n,x^2_1,\dots,x^2_n,\dots,x^2_1\dots,x^2_n\right)^\top\in\mathbb{R}^{2n^2\times1},$$

$$U=\begin{pmatrix}U_1\\U_2\end{pmatrix}=\left(u^1_{11},\dots,u^1_{nn},u^2_{11},\dots,u^2_{nn}\right)^\top\in\mathbb{R}^{2n^2\times1},\quad V=\left(v_{11},\dots,v_{nn}\right)^\top\in\mathbb{R}^{n^2\times1}.$$

To proceed, we introduce the discretization of each item (i.e., $\mathcal{D}[R,T;\boldsymbol{u}]$, $\mathcal{S}[\boldsymbol{u}]$, $\mathcal{S}[v]$, $\mathcal{C}[\boldsymbol{u},v]$, and $\Psi[v]$) in the subproblems (9) and (10) separately.

4.2.1. Discretization of the similarity measure

Before discretizing the similarity measure (3), we assume that the pixel values for images $R(\boldsymbol{x})$ and $T(\boldsymbol{x})$ are confined within a known interval i.e., $(r_\imath,t_\nu)\in\mathcal{R}\times\mathcal{T}=[r_a,r_b]\times[t_a,t_b]$. These intervals are partitioned into n_r and n_t bins, with respective bin widths $\sigma_r=\frac{r_b-r_a}{n_r}$ and $\sigma_t=\frac{t_b-t_a}{n_t}$. These widths are utilized as the kernel function's window width. Thus, pixel values at each bin's center can be computed by

$$r_\imath=r_a+(\imath-0.5)\sigma_r,\ \imath=1,2,\dots,n_r\quad\text{and}\quad t_\nu=t_a+(\nu-0.5)\sigma_t,\ \nu=1,2,\dots,n_t.$$

To simplify the representation of the discrete joint probability distribution and the marginal probability distribution, we first define the following vectors

$$K_{[\sigma_t;\boldsymbol{u}]}(t)=\left(\kappa_{[\sigma_t;\boldsymbol{u}]}(t-t_{11}),\dots,\kappa_{[\sigma_t;\boldsymbol{u}]}(t-t_{1n}),\dots,\kappa_{[\sigma_t;\boldsymbol{u}]}(t-t_{nn})\right)^\top\in\mathbb{R}^{n^2\times1},$$

$$K_{\sigma_r}(r)=\left(\kappa_{\sigma_r}(r-r_{11}),\dots,\kappa_{\sigma_r}(r-r_{1n}),\dots,\kappa_{\sigma_r}(r-r_{nn})\right)^\top\in\mathbb{R}^{n^2\times1},$$

$$P_{[R,T;U]} = \left(p_{[R,T;\boldsymbol{u}]}(r_1,t_1), \dots, p_{[R,T;\boldsymbol{u}]}(r_1,t_{n_t}), \dots, p_{[R,T;\boldsymbol{u}]}(r_{n_r},t_{n_t})\right)^\top \in \mathbb{R}^{(n_r n_t)\times 1},$$
$$P_{[T;U]} = \left(p_{[T;\boldsymbol{u}]}(t_1), \dots, p_{[T;\boldsymbol{u}]}(t_{n_t})\right)^\top \in \mathbb{R}^{n_t\times 1}, \quad P_{[R]} = \left(p_{[R]}(r_1), \dots, p_{[R]}(r_{n_r})\right)^\top \in \mathbb{R}^{n_r\times 1},$$

where $r_{ij} := R(\boldsymbol{x}_{ij})$ and $t_{ij} := T(\boldsymbol{x}_{ij} + \boldsymbol{u}_{ij})$.

To preceed, we first introduce the discrete forms of both joint and marginal probability distributions. The expression for the joint probability distribution function Eq. (2) is given by

$$p_{[R,T;\boldsymbol{u}]}(r_\iota, t_\nu) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \kappa_{\sigma_r}(r_\iota - r_{ij})\kappa_{[\sigma_t;\boldsymbol{u}]}(t_\nu - t_{ij}) = \frac{1}{n^2} K_{\sigma_r}^\top(r_\iota) K_{[\sigma_t;\boldsymbol{u}]}(t_\nu).$$

The marginal distributions are approximated using the joint distribution as follows:

$$p_{[T;\boldsymbol{u}]}(t_\nu) = \int_{\mathcal{R}} p_{[R,T;\boldsymbol{u}]}(r, t_\nu)\, dr \approx \sigma_r \sum_{\iota=1}^{n_r} p_{[R,T;\boldsymbol{u}]}(r_\iota, t_\nu),$$
$$p_{[R]}(r_\iota) = \int_{\mathcal{T}} p_{[R,T;\boldsymbol{u}]}(r_\iota, t)\, dt \approx \sigma_t \sum_{\nu=1}^{n_t} p_{[R,T;\boldsymbol{u}]}(r_\iota, t_\nu).$$

Define $S_T = \boldsymbol{e}_r \otimes I_{n_t} \in \mathbb{R}^{n_t\times(n_r\times n_t)}$ and $S_R = I_{n_r} \otimes \boldsymbol{e}_t \in \mathbb{R}^{n_r\times(n_r\times n_t)}$, with $\boldsymbol{e}_r = (1,\dots,1) \in \mathbb{R}^{1\times n_r}$, $\boldsymbol{e}_t = (1,\dots,1) \in \mathbb{R}^{1\times n_t}$, I_{n_t} and I_{n_r} as identity matrices of corresponding sizes and $\otimes$ the Kronecker product. We obtain

$$P_{[T;U]} = \sigma_r S_T P_{[R,T;U]} \quad \text{and} \quad P_{[R]} = \sigma_t S_R P_{[R,T;U]}.$$

Hence, based on the above discussion, the integrals of the similarity measure (3) can be approximated by a midpoint quadrature rule on the $n_r \times n_t$ bins. The discretization of the similarity measure is as follows

$$\begin{aligned}\mathcal{D}[R,T;U] &= \sigma_r \sum_{\iota=1}^{n_r} p_{[R]}(r_\iota)\log p_{[R]}(r_\iota) + \sigma_t \sum_{\nu=1}^{n_t} p_{[T;\boldsymbol{u}]}(t_\nu)\log p_{[T;\boldsymbol{u}]}(t_\nu) - \sigma_r\sigma_t \sum_{\nu=1}^{n_t}\sum_{\iota=1}^{n_r} p_{[R,T;\boldsymbol{u}]}(r_\iota,t_\nu)\ \log\ p_{[R,T;\boldsymbol{u}]}(r_\iota,t_\nu)\\
&= \sigma_r P_{[R]}^\top \mathrm{Log}(P_{[R]}) + \sigma_t P_{[T;U]}^\top \mathrm{Log}(P_{[T;U]}) - \sigma_r\sigma_t P_{[R,T;U]}^\top \mathrm{Log}(P_{[R,T;U]})\\
&= \sigma_r\sigma_t\Big((S_R P_{[R,T;U]})^\top \mathrm{Log}(\sigma_t S_R P_{[R,T;U]}) + \big(S_T P_{[R,T;U]}\big)^\top \mathrm{Log}(\sigma_r S_T P_{[R,T;U]}) - P_{[R,T;U]}^\top \mathrm{Log}(P_{[R,T;U]})\Big),\end{aligned}$$

where the vector logarithmic function $\mathrm{Log}(\cdot)$ is defined as

$$\mathrm{Log}(Q) = (\log q_1, \log q_2, \dots, \log q_n)^\top, \text{with } Q = (q_1, q_2, \dots, q_n)^\top.$$

4.2.2. Discretization of the regularization term

The diffusion regularizer penalizes large gradients in the displacement field $\boldsymbol{u}$, promoting smoothness by minimizing its spatial derivatives. To implement numerically the diffusion regularizer of the displacement field $\boldsymbol{u}$, which is rewritten as

$$\mathcal{S}[\boldsymbol{u}] = \frac{1}{2}\int_\Omega \|\nabla\boldsymbol{u}\|^2 \mathrm{d}\boldsymbol{x} = \frac{1}{2}\int_\Omega \sum_{\ell=1}^{2} \|\nabla u^\ell\|^2 \mathrm{d}\boldsymbol{x},$$

we discretize it using the forward difference and the midpoint quadrature rule on a grid (e.g., pixel/voxel grid for images), and obtain

$$\mathcal{S}[U] = \frac{1}{2} h^2 U^\top G^\top G U,$$

where $G = I_2 \otimes \tilde{G} \in \mathbb{R}^{4n^2\times 2n^2}$, $\tilde{G} = \left(\partial_{x^1}^{+\ \top}, \partial_{x^2}^{+\ \top}\right)^\top \in \mathbb{R}^{2n^2\times n^2}$, $\partial_{x^1}^{+} = \partial_n^h \otimes I_n \in \mathbb{R}^{n^2\times n^2}$, $\partial_{x^2}^{+} = I_n \otimes \partial_n^h \in \mathbb{R}^{n^2\times n^2}$, and

$$\partial_n^h = \frac{1}{h}\begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ & & & -1 \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Especially, we assume the ghost points $u_{n+1,\cdot}^\ell = u_{\cdot,n+1}^\ell = 0$ for the boundary condition $\boldsymbol{u}_{\partial\Omega} = \boldsymbol{0}$.

Remark 1. In the 3D case, these operators can be extended as

$$G = I_3 \otimes \tilde{G} \in \mathbb{R}^{9n^3\times 3n^3}, \qquad \tilde{G} = \left(\partial_{x^1}^{+\ \top}, \partial_{x^2}^{+\ \top}, \partial_{x^3}^{+\ \top}\right)^\top \in \mathbb{R}^{3n^3\times n^3},$$

where $\partial_{x^1}^{+} = I_n \otimes \partial_n^h \otimes I_n$, $\partial_{x^2}^{+} = I_n \otimes I_n \otimes \partial_n^h$, $\partial_{x^3}^{+} = \partial_n^h \otimes I_n \otimes I_n \in \mathbb{R}^{n^3\times n^3}$.

Similarly, for the diffusion regularizer $\mathcal{S}[v] = \frac{1}{2}\int_\Omega \|\nabla v(\boldsymbol{x})\|^2 \mathrm{d}\boldsymbol{x}$ of the relaxation function $v(\boldsymbol{x})$, its discretization can be written as

$$\mathcal{S}[V] = \frac{1}{2} h^2 V^\top \tilde{G}^\top \tilde{G} V. \tag{12}$$

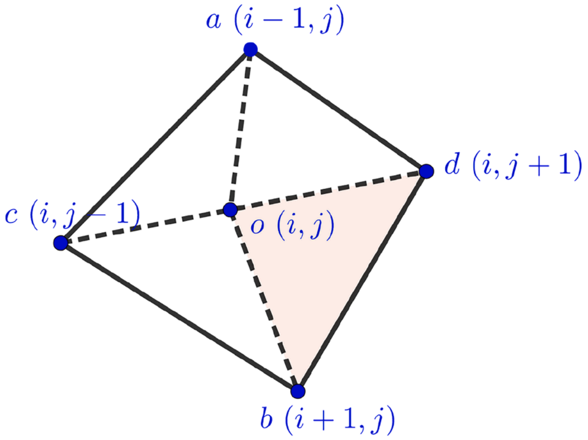


Fig. 1. Finite difference computation involved the Jacobian determinant $\det(I_2+\nabla \boldsymbol{u})\big|_o$ of the deformation $o := (\boldsymbol{x}+\boldsymbol{u})_{ij}$ at the cell center (i,j).

4.2.3. Discretizations of the constraint term and the restriction term

To discretize the constraint term $C[\boldsymbol{u},v] := \frac{1}{2}\int_\Omega \left(\det(I_d+\nabla\boldsymbol{u}(\boldsymbol{x}))-v(\boldsymbol{x})\right)^2 \mathrm{d}\boldsymbol{x}$, the key component is how to discretize the Jacobian determinant $\det(I_2+\nabla\boldsymbol{u}(\boldsymbol{x}))$. Here, we exploit the idea of our previous work [49]. For the Jacobian determinant $\det(I_2+\nabla\boldsymbol{u}(\boldsymbol{x}))\big|_o$ of the deformation $o := (\boldsymbol{x}+\boldsymbol{u})_{ij}$ at the cell center (i,j) (see Fig. 1), its discretization is given by

$$\begin{aligned}\det(I_2+\nabla\boldsymbol{u}(\boldsymbol{x}))\big|_o = \frac{1}{4h^2}\Bigg(&\begin{vmatrix} h+u^1_{i+1,j}-u^1_{i,j} & u^1_{i,j+1}-u^1_{i,j} \\ u^2_{i+1,j}-u^2_{i,j} & h+u^2_{i,j+1}-u^2_{i,j}\end{vmatrix} + \begin{vmatrix} u^1_{i,j+1}-u^1_{i,j} & -h+u^1_{i-1,j}-u^1_{i,j} \\ h+u^2_{i,j+1}-u^2_{i,j} & u^2_{i-1,j}-u^2_{i,j}\end{vmatrix} \\ &+\begin{vmatrix} -h+u^1_{i-1,j}-u^1_{i,j} & u^1_{i,j-1}-u^1_{i,j} \\ u^2_{i-1,j}-u^2_{i,j} & -h+u^2_{i,j-1}-u^2_{i,j}\end{vmatrix} + \begin{vmatrix} u^1_{i,j-1}-u^1_{i,j} & h+u^1_{i+1,j}-u^1_{i,j} \\ -h+u^2_{i,j-1}-u^2_{i,j} & u^2_{i+1,j}-u^2_{i,j}\end{vmatrix}\Bigg) \\ =:\ &\frac{1}{2h^2}\left(A^{ij}_{obd}+A^{ij}_{oda}+A^{ij}_{oac}+A^{ij}_{ocb}\right),\end{aligned}$$

where A^{ij}_{obd} denotes the signed area of the triangle obd at the cell center (i,j), i.e.,

$$A^{ij}_{obd} := \frac{1}{2}\begin{vmatrix} h+u^1_{i+1,j}-u^1_{i,j} & u^1_{i,j+1}-u^1_{i,j} \\ u^2_{i+1,j}-u^2_{i,j} & h+u^2_{i,j+1}-u^2_{i,j}\end{vmatrix}.$$

To simplify the Jacobian determinant constraint discretization at central grid points, we introduce specific discrete differential operators, facilitating the computation of partial derivatives in a compact form. These operators are defined as:

$$\begin{aligned}&\partial^+_{x^1} = \partial^h_n \otimes I_n, \quad \partial^+_{x^2} = I_n \otimes \partial^h_n, \quad \partial^-_{x^1} = (\partial^h_n)^\top \otimes I_n, \quad \partial^-_{x^2} = I_n \otimes (\partial^h_n)^\top \in \mathbb{R}^{n^2\times n^2}, \\ &D^+_{\ell,1} = (\partial^+_{x^\ell}, \mathbf{0}), \quad D^+_{\ell,2} = (\mathbf{0}, \partial^+_{x^\ell}), \quad D^-_{\ell,1} = (\partial^-_{x^\ell}, \mathbf{0}), \quad D^-_{\ell,2} = (\mathbf{0}, \partial^-_{x^\ell}) \in \mathbb{R}^{n^2\times 2n^2},\end{aligned}$$

where $\ell = 1,2$. Then the discretized Jacobian determinant given by $\mathrm{Det}(U) := \det\left(\nabla(X+U)\right)$ can be expressed compactly as

$$\begin{aligned}\mathrm{Det}(U) = \frac{1}{4h^2}\Big[&(he+D^+_{1,1}U)\odot(he+D^+_{2,2}U) - D^+_{2,1}U\odot D^+_{1,2}U \\ &+ D^+_{2,1}U\odot D^-_{1,2}U - (-he+D^-_{1,1}U)\odot(he+D^+_{2,2}U) \\ &+ (-he+D^-_{1,1}U)\odot(-he+D^-_{2,2}U) - D^-_{2,1}U\odot D^-_{1,2}U \\ &+ D^-_{2,1}U\odot D^+_{1,2}U - (he+D^+_{1,1}U)\odot(-he+D^-_{2,2}U)\Big] \\ =:\ &\frac{1}{2h^2}\left(\mathcal{A}_{obd}+\mathcal{A}_{oda}+\mathcal{A}_{oac}+\mathcal{A}_{ocb}\right) =: \frac{1}{2}\left(\Theta_{obd}+\Theta_{oda}+\Theta_{oac}+\Theta_{ocb}\right),\end{aligned} \tag{13}$$

where $\odot$ denotes the Hadamard product and $\boldsymbol{e} = (1,\dots,1)^\top \in \mathbb{R}^{n^2\times 1}$. The term $\mathcal{A}_{obd}$ represents the signed area of the triangle obd in the center of each cell (specifically, $\mathcal{A}_{obd}(l) = A^{ij}_{obd}$, for $l = (i-1)n+j$) and the term $\Theta_{obd} := \frac{1}{h^2}\mathcal{A}_{obd}$ indicates the ratio of the area of $\mathcal{A}_{obd}$ with respect to the area element h^2.

Remark 2. In the 3D case, similar to the 2D scenario, the Jacobian determinant of the deformation $o := (\boldsymbol{x}+\boldsymbol{u})_{ijk}$ at the cell center (i,j,k) is given by

$$\begin{aligned}&\det(I_3+\nabla \boldsymbol{u}(\boldsymbol{x}))\big|_o\\ &= \frac{1}{8h^3}\sum_{i',j',k'\in\{-1,+1\}}\begin{vmatrix} i'h+u^1_{i+i',j,k}-u^1_{i,j,k} & u^1_{i,j+j',k}-u^1_{i,j,k} & u^1_{i,j,k+k'}-u^1_{i,j,k}\\ u^2_{i+i',j,k}-u^2_{i,j,k} & j'h+u^2_{i,j+j',k}-u^2_{i,j,k} & u^2_{i,j,k+k'}-u^2_{i,j,k}\\ u^3_{i+i',j,k}-u^3_{i,j,k} & u^3_{i,j+j',k}-u^3_{i,j,k} & k'h+u^3_{i,j,k+k'}-u^3_{i,j,k}\end{vmatrix} = \frac{1}{8h^3}\sum_{i',j',k'\in\{-1,+1\}}\mathfrak{det}_{i',j',k'},\end{aligned}$$

where $\mathfrak{det}_{i',j',k'} := \begin{vmatrix} i'h+u^1_{i+i',j,k}-u^1_{i,j,k} & u^1_{i,j+j',k}-u^1_{i,j,k} & u^1_{i,j,k+k'}-u^1_{i,j,k}\\ u^2_{i+i',j,k}-u^2_{i,j,k} & j'h+u^2_{i,j+j',k}-u^2_{i,j,k} & u^2_{i,j,k+k'}-u^2_{i,j,k}\\ u^3_{i+i',j,k}-u^3_{i,j,k} & u^3_{i,j+j',k}-u^3_{i,j,k} & k'h+u^3_{i,j,k+k'}-u^3_{i,j,k}\end{vmatrix}$.

In a similar manner, we also introduce specific discrete differential operators for the 3D case:

$$\begin{aligned}&\partial^+_{x^1} = I_n\otimes\partial^1_n\otimes I_n, \qquad \partial^+_{x^2} = I_n\otimes I_n\otimes\partial^1_n, \qquad \partial^+_{x^3} = \partial^1_n\otimes I_n\otimes I_n\in\mathbb{R}^{n^3\times n^3},\\ &\partial^-_{x^1} = I_n\otimes(\partial^1_n)^\top\otimes I_n, \quad \partial^-_{x^2} = I_n\otimes I_n\otimes(\partial^1_n)^\top, \quad \partial^-_{x^3} = (\partial^1_n)^\top\otimes I_n\otimes I_n\in\mathbb{R}^{n^3\times n^3},\\ &\boldsymbol{D}^+_{\ell,1} = (\partial^+_{x^\ell},\boldsymbol{0},\boldsymbol{0}), \quad \boldsymbol{D}^+_{\ell,2} = (\boldsymbol{0},\partial^+_{x^\ell},\boldsymbol{0}), \quad \boldsymbol{D}^+_{\ell,3} = (\boldsymbol{0},\boldsymbol{0},\partial^+_{x^\ell})\in\mathbb{R}^{n^3\times 3n^3},\\ &\boldsymbol{D}^-_{\ell,1} = (\partial^-_{x^\ell},\boldsymbol{0},\boldsymbol{0}), \quad \boldsymbol{D}^-_{\ell,2} = (\boldsymbol{0},\partial^-_{x^\ell},\boldsymbol{0}) \quad \boldsymbol{D}^-_{\ell,3} = (\boldsymbol{0},\boldsymbol{0},\partial^-_{x^\ell})\in\mathbb{R}^{n^3\times 3n^3}.\end{aligned}$$

Accordingly, the discretized Jacobian determinant is given by

$$\mathrm{Det}(U) = \frac{1}{8h^3}\sum_{i',j',k'\in\{-1,+1\}} V_{i'j'k'},$$

where

$$\begin{aligned}V_{i'j'k'} = {}&(i'h\boldsymbol{e}+\boldsymbol{D}^{\omega_{i'}}_{1,1}U)\odot\left[(j'h\boldsymbol{e}+\boldsymbol{D}^{\omega_{j'}}_{2,2}U)\odot(k'h\boldsymbol{e}+\boldsymbol{D}^{\omega_{k'}}_{3,3}U)-(\boldsymbol{D}^{\omega_{k'}}_{3,2}U)\odot(\boldsymbol{D}^{\omega_{j'}}_{2,3}U)\right]\\ &-(\boldsymbol{D}^{\omega_{i'}}_{1,2}U)\odot\left[(\boldsymbol{D}^{\omega_{j'}}_{2,1}U)\odot(k'h\boldsymbol{e}+\boldsymbol{D}^{\omega_{k'}}_{3,3}U)-(\boldsymbol{D}^{\omega_{k'}}_{3,1}U)\odot(\boldsymbol{D}^{\omega_{j'}}_{2,3}U)\right]\\ &+(\boldsymbol{D}^{\omega_{i'}}_{1,3}U)\odot\left[(\boldsymbol{D}^{\omega_{j'}}_{2,1}U)\odot(\boldsymbol{D}^{\omega_{k'}}_{3,2}U)-(\boldsymbol{D}^{\omega_{k'}}_{3,1}U)\odot(j'h\boldsymbol{e}+\boldsymbol{D}^{\omega_{j'}}_{2,2}U)\right]\end{aligned}$$

with the sign convention $\omega_1 =$‘ + ’ and $\omega_{-1} =$‘ - ’.

Utilizing the midpoint quadrature rule and derived analysis from the previous discussion, the discretization of the Jacobian determinant constraint $C[\boldsymbol{u}, v]$ is reformulated as

$$C[U,V] = \frac{h^2}{2}(\mathrm{Det}(U)-V)^\top\big(\mathrm{Det}(U)-V\big).$$

Following the approach in [49], a grid folding indicator $\Theta\in\mathbb{R}^{n^2\times 1}$ is assigned to each cell center (i,j) to track the risk of grid folding. It is formulated as

$$\Theta(l) := \min\left\{\Theta_{obd}(l),\Theta_{oda}(l),\Theta_{oac}(l),\Theta_{ocb}(l)\right\}, \quad l=(i-1)n+j. \tag{14}$$

Here, it can be reminded that a negative $\Theta(l)$ denotes the folding in the l-th cell at (i,j). In such cases, we use [49, Algorithm 4.1] to correct the deformation, thereby deriving a new U.

Utilizing the rectangular integral formula, the discrete form of the restriction term $\Psi[v]$ is succinctly derived as follows:

$$\Psi[V] = h^2[(V-\boldsymbol{e})\odot(V-\boldsymbol{e})]^\top\frac{1}{V},$$

where $\frac{1}{V} = \left(\frac{1}{v_{11}},\dots,\frac{1}{v_{1n}},\dots,\frac{1}{v_{n1}},\dots,\frac{1}{v_{nn}}\right)^\top\in\mathbb{R}^{n^2\times 1}$.

4.2.4. *Discretization of two subproblems*

From the preceding analysis, we derive the discrete version of subproblems in (11). This discrete formulation incorporates the similarity measure, regularization components, Jacobian determinant constraint, and restriction term within a finite-dimensional framework, as follows,

$$\begin{cases} U^{k+1} = \underset{U}{\arg\min} \mathcal{L}^1_{\lambda^k,V^k}[U] := \left\{\mathcal{D}[R,T;U] + \alpha \mathcal{S}[U] + \lambda^k \mathcal{C}[U,V^k]\right\} \\ \quad = \underset{U}{\arg\min} \Big\{ \sigma_r\sigma_t((S_R P_{[R,T;U]})^\top \mathrm{Log}(\sigma_t S_R P_{[R,T;U]}) + (S_T P_{[R,T;U]})^\top \mathrm{Log}(\sigma_r S_T P_{[R,T;U]}) \\ \quad - P^\top_{[R,T;U]} \mathrm{Log}(P_{[R,T;U]})) + \frac{\alpha h^2}{2} U^\top G^\top G U + \frac{\lambda^k h^2}{2}\left(\mathrm{Det}(U) - V^k\right)^\top \left(\mathrm{Det}(U) - V^k\right) \Big\}, \\ V^{k+1} = \underset{V}{\arg\min} \mathcal{L}^2_{\lambda^k,U^{k+1}}[V] := \left\{\beta\Psi[V] + \gamma\mathcal{S}[V] + \lambda^k \mathcal{C}[U^{k+1},V]\right\} \\ \quad = \underset{V}{\arg\min} \Big\{ \beta h^2 [(V - \boldsymbol{e}) \odot (V - \boldsymbol{e})]^\top \frac{1}{V} + \frac{\gamma h^2}{2} V^\top \tilde{G}^\top \tilde{G} V \\ \qquad + \frac{\lambda^k h^2}{2}\left(\mathrm{Det}(U^{k+1}) - V\right)^\top \left(\mathrm{Det}(U^{k+1}) - V\right) \Big\}, \end{cases} \tag{15}$$

enabling the formulation of effective numerical algorithms to solve these subproblems.

4.3. Numerical solvers for two discrete subproblems

This subsection focuses on the numerical solvers for the discrete subproblems mentioned above. Given their specific properties, we utilize the L-BFGS method for the U subproblem and the Gauss–Newton method for the V subproblem. To guarantee that the function value sequences $\mathcal{L}^1_{\lambda,V}[U]$ and $\mathcal{L}^2_{\lambda,U}[V]$ decrease monotonically, the Armijo line search is used [50]. Moreover, to improve algorithm speed and avoid local minima, a multilevel strategy is incorporated [1].

Algorithm 1 L-BFGS two loop recursion (compute $\delta_{U^i} = -H^{-1}_{U^i} g$).

```
Input: $i, m, Z, Y, H_{U^{i-m}}, g$;
for $s = i-1, i-2, \dots, i-m$ do
    Let $z_s = Z(:, s-i+m+1),\ y_s = Y(:, s-i+m+1)$;
    Compute $\eta_s = 1/(y_s^\top z_s)$;
    Compute $\alpha_s = \eta_s z_s^\top g$;
    Update $g = g - \alpha_s y_s$;
end
Compute $r$ by solving $H_{U^{i-m}} r = g$;
for $s = i-m, i-m+1, \dots, i-1$ do
    Compute $\beta = \eta_s y_s^\top r$;
    Update $r = r + (\alpha_s - \beta) z_s$;
end
Output: $\delta_{U^i} = -r$.
```

4.3.1. The L-BFGS method to solve the discretized U subproblem

Employing a similarity measure based on mutual information leads to significant computational demands in approximating the Hessian matrix. To address both precision and computational efficiency, the L-BFGS method is a viable choice, as it avoids the need for explicit Hessian calculations. This approach only necessitates the storage and management of the last m updates of the displacement field increment matrix Z and the gradient increment matrix U. Let U^i denote the displacement field for the current i-th step. The search direction δ_{U^i} is derived as

$$\delta_{U^i} = -H^{-1}_{U^i} g_{U^i}. \tag{16}$$

In the upcoming iteration, the displacement field U^{i+1} is updated using the formula $U^{i+1} = U^i + \mu_{U^i}\delta_{U^i}$, where μ_{U^i} is the step size determined via Armijo line search [50, Algorithm 3.1]. The L-BFGS update for the inverse Hessian $H^{-1}_{U^{i+1}}$ is expressed as:

$$\begin{aligned} H^{-1}_{U^i} = & \left((\Lambda^{i-1})^T \dots (\Lambda^{i-m})^T\right) H^{-1}_{U^{i-m}} \left(\Lambda^{i-m} \dots \Lambda^{i-1}\right) \\ & + \eta^{i-m}\left((\Lambda^{i-1})^T \dots (\Lambda^{i-m+1})^T\right) \boldsymbol{z}^{i-m} (\boldsymbol{z}^{i-m})^T \left(\Lambda^{i-m+1} \dots \Lambda^{i-1}\right) \\ & + \eta^{i-m+1}\left((\Lambda^{i-1})^T \dots (\Lambda^{i-m+2})^T\right) \boldsymbol{z}^{i-m+1} (\boldsymbol{z}^{i-m+1})^T (\Lambda^{i-m+2} \dots \Lambda^{i-1}) + \dots + \eta^{i-1} \boldsymbol{z}^{i-1} (\boldsymbol{z}^{i-1})^T, \end{aligned} \tag{17}$$

where

$$\Lambda^i = I - \eta^i \boldsymbol{y}^i (\boldsymbol{z}^i)^\top,\ \eta^i = 1/\left((\boldsymbol{y}^i)^\top \boldsymbol{z}^i\right),\ \boldsymbol{z}^i = U^{i+1} - U^i,\ \boldsymbol{y}^i = \nabla\mathcal{L}^1_{\lambda,V}(U^{i+1}) - \nabla\mathcal{L}^1_{\lambda,V}(U^i),$$

and $m = \min\{i, m_{\max}\}$ with $m_{\max} = 5$. The inverse of $H_{U^{i-m}}$ is not calculated explicitly; an approximation is employed. The L-BFGS algorithm improves efficiency and saves memory by avoiding direct computation of (17) through matrix and vector operations to evaluate (16). Details of the two-loop recursion applied in L-BFGS to derive (16) are presented in Algorithm 1.

Remark 3. For the L-BFGS method to produce a positive definite matrix $H_{U^{i+1}}^{-1}$, it is necessary that $(\boldsymbol{z}^i)^\top \boldsymbol{y}^i > 0$ and $H_{U^i}^{-1}$ remain positive definite. The Armijo line search does not consistently guarantee the curvature condition. Hence, to resolve this, we retain $H_{U^{i+1}}^{-1} = H_{U^i}^{-1}$ if $(\boldsymbol{z}^i)^\top \boldsymbol{y}^i \leq 0$, Ithen avoiding the L-BFGS update.

To optimize the U subproblem in (15) using the L-BFGS method, the gradient $g_{_U}$ and an initial Hessian matrix H_{U^1} approximating $H_{U^{i-m}}$ in (17) are required:

$$g_{_U} = g_{_U}^{\mathcal{D}[R,T;U]} + \alpha g_{_U}^{\mathcal{S}[U]} + \lambda g_{_U}^{\mathcal{C}[U,V]}, \quad H_{U^1} = G^\top G + \varepsilon I_{2n^2}. \tag{18}$$

To ensure that $H_{U^i}^{-1}$ remains symmetric and positive definite in subsequent steps without explicit computation (see Algorithm 1), the initial matrix $H_{U^1}^{-1}$ should also have these properties. Direct computation of second-order derivatives is costly; therefore, the initial matrix H_{U^1} uses the Hessian $G^\top G$ of the regularization term. To enhance the positive definiteness of H_{U^1}, a scaled identity matrix is added, with ε being a small positive constant (set to $\varepsilon = 10^{-4}$ in our implementation).

To proceed, we compute $g_{_U}^{\mathcal{D}[R,T;U]}$, $g_{_U}^{\mathcal{S}[U]}$, and $g_{_U}^{\mathcal{C}[U,V]}$ within (18). For the discretized similarity metric $\mathcal{D}[R,T;U]$, the chain rule is employed to yield

$$\begin{aligned} g_{_U}^{\mathcal{D}[R,T;U]} = \sigma_r\sigma_t \frac{\partial T}{\partial U}\Bigg(& \left(S_R \frac{\partial P_{[R,T;U]}}{\partial T}\right)^\top \left(\mathrm{Log}\left(\sigma_t S_R P_{[R,T;U]}\right) + \boldsymbol{e}_r^\top\right) \\ & + \left(S_T \frac{\partial P_{[R,T;U]}}{\partial T}\right)^\top \left(\mathrm{Log}\left(\sigma_r S_T P_{[R,T;U]}\right) + \boldsymbol{e}_t^\top\right) - \left(\frac{\partial P_{[R,T;U]}}{\partial T}\right)^\top \left(\mathrm{Log}\left(P_{[R,T;U]}\right) + (\boldsymbol{e}_t \otimes \boldsymbol{e}_r)^\top\right)\Bigg). \end{aligned}$$

Next, for the discretized diffusion regularizer $\mathcal{S}[U]$ and the constraint of the Jacobian determinant $\mathcal{C}[U,V]$, their gradients are given by

$$g_{_U}^{\mathcal{S}[U]} = h^2 G^\top G U \quad \text{and} \quad g_{_U}^{\mathcal{C}[U,V]} = h^2 M^\top(U)(\mathrm{Det}(U) - V),$$

where $M(U)$ denotes the Jacobian matrix of $\mathrm{Det}(U)$ with its explicit form given as follows

$$M(U) := \frac{1}{4h^2} \sum_{s,t=1}^{2} (-1)^{s+t} \mathrm{diag}\left(\mathrm{D}_{\mathrm{s,t}}^{+}\mathrm{U} - \mathrm{D}_{\mathrm{s,t}}^{-}\mathrm{U}\right)\left(\mathrm{D}_{\mathrm{3-s,3-t}}^{+} - \mathrm{D}_{\mathrm{3-s,3-t}}^{-}\right) + \frac{1}{2\mathrm{h}} \sum_{\mathrm{s}=1}^{2} \left(\mathrm{D}_{\mathrm{s,s}}^{+} - \mathrm{D}_{\mathrm{s,s}}^{-}\right),$$

and the detailed derivation presented in Appendix A. Consequently, following the gradient computation for the U subproblem, we summarize the L-BFGS solution process in Algorithm 2.

Algorithm 2 Compute U^{k+1} by using L-BFGS (U-Subproblem-Optimizer).

```
Input: $U^k, V^k, T, R, \alpha, \beta, \gamma, \lambda^k, \epsilon_u$, MaxItU;
Initialization: $U^{k,1} = U^k, m_{\max} = 5, Z(:, m_{\max}) = 0, Y(:, m_{\max}) = 0, m = 0$;
Compute initial approximation Hessian $H_{U^{k,1}}$ by (18);
for $i = 1, \ldots,$ MaxItU do
    Compute gradient $g_{_{U^{k,i}}}$ by (18);
    Compute search direction $\delta_{U^{k,i}} = -H_{U^{k,i}}^{-1} g_{_{U^{k,i}}}$ by using Algorithm 1, i.e.,
      $\delta_{U^{k,i}} = \text{L-BFGS}(i, m, Z, Y, H_{U^{k,1}}, g_{_{U^{k,i}}})$;
    Update $U^{k,i+1} = U^{k,i} + \mu_{U^{k,i}} \delta_{U^{k,i}}$ by Armijo line search (see [50, Algorithm 3.1]);
    if $|\mathcal{L}_{\lambda^k,V^k}^1(U^{k,i+1}) - \mathcal{L}_{\lambda^k,V^k}^1(U^{k,i})| \leq \epsilon_u |\mathcal{L}_{\lambda^k,V^k}^1(U^{k,1})|$ or $\|U^{k,i+1} - U^{k,i}\| \leq \epsilon_u \|U^{k,1}\|$ then
        break;
    end
    Compute $z^{k,i} = U^{k,i+1} - U^{k,i}$, $y^{k,i} = \nabla\mathcal{L}_{\lambda^k,V^k}^1(U^{k,i+1}) - \nabla\mathcal{L}_{\lambda^k,V^k}^1(U^{k,i})$;
    if $(z^{k,i})^\top y^{k,i} > 0$ then
        Update $m = m + 1$;
        if $m > m_{\max}$ then
            Overwrite variables $Z(:, 1), Y(:, 1)$ and perform data migration:
              $Z(:, 1 : m_{\max} - 1) = Z(:, 2 : m_{\max}),\ Y(:, 1 : m_{\max} - 1) = Y(:, 2 : m_{\max})$;
            Set $m = m_{\max}$;
        end
        Store variables $Z(:, m) = z^{k,i},\ Y(:, m) = y^{k,i}$;
    end
end
$U^{k+1} = U^{k,i+1}$;
Output: $U^{k+1}$.
```

4.3.2. The Gauss–Newton method solving the discretized V subproblem

Due to the simpler computation of the Hessian and gradient for the V subproblem in (15), compared to the U subproblem, the Gauss–Newton method is preferred over L-BFGS due to its efficiency. The most demanding computational effort involves finding the search direction. At iteration i, the search direction δ_{V^i} is derived from solving the Gauss–Newton equation:

$$H_V^i \delta_{V^i} = -g_{V^i}. \tag{19}$$

The V update is given by $V^{i+1} = V^i + \mu_{V^i}\delta_{V^i}$, with the step size μ_{V^i} determined by the Armijo line search [50, Algorithm 3.1].

The implementation of the Gauss–Newton method requires the calculation of the gradient g_V and the approximate Hessian H_V for the V subproblem. Below, we present the gradients and Hessians of each term in the V subproblem, followed by the combined gradient and Hessian necessary for optimizing this subproblem. For the diffusion term $S[V]$, the gradient and Hessian are expressed as

$$g_V^{S[V]} = h^2\tilde{G}^\top\tilde{G}V \quad \text{and} \quad H_V^{S[V]} = h^2\tilde{G}^\top\tilde{G}.$$

For the restricted term $\Psi(V)$, the gradient is given by

$$g_V^{\Psi[V]} = h^2\Big(\boldsymbol{e} - \frac{1}{V}\odot\frac{1}{V}\Big).$$

To facilitate computations, we simplify the Hessian of $\Psi(V)$ by ignoring second-order terms, resulting in

$$H_V^{\Psi[V]} = 2h^2\mathrm{diag}\Big(\frac{1}{\mathrm{V}}\odot\frac{1}{\mathrm{V}}\odot\frac{1}{\mathrm{V}}\Big).$$

For the Jacobian determinant constraint in V, the gradient and Hessian become

$$g_V^{C[U,V]} = -h^2(\mathrm{Det}(U) - V), \quad H_V^{C[U,V]} = h^2 I_{n^2}. \tag{20}$$

Thus, the combined gradient and approximate Hessian of V in (15) are:

$$g_V = \gamma g_V^{S[V]} + \beta g_V^{\Psi[V]} + \lambda g_V^{C[U,V]}, \quad H_V = \gamma H_V^{S[V]} + \beta H_V^{\Psi[V]} + \lambda H_V^{C[U,V]}. \tag{21}$$

The explicit formation or storage of H_V is computationally costly. To address this, we adopt an iterative strategy to resolve the Gauss–Newton system (19) without directly constructing H_V. Krylov subspace techniques are well-suited for this purpose [51–54], as they only require matrix-vector multiplications. Given that H_V is both symmetric and positive definite, the conjugate gradient (CG) method is used to solve these systems [55,56]. In our implementation, we compute the matrix-vector product $H_V\delta$ as follows:

$$\begin{aligned} H_V\delta &= \gamma h^2(\tilde{G}^\top\tilde{G})\delta + 2\beta h^2\mathrm{diag}\Big(\frac{1}{\mathrm{V}}\odot\frac{1}{\mathrm{V}}\odot\frac{1}{\mathrm{V}}\Big)\updelta + \uplambda \mathrm{h}^2\updelta \\ &= \gamma h^2(\tilde{G}^\top\tilde{G})\delta + 2\beta h^2\frac{1}{V}\odot\frac{1}{V}\odot\frac{1}{V}\odot\delta + \lambda h^2\delta. \end{aligned}$$

The term $\tilde{G}^\top\tilde{G}$ is precomputed during the regularization step (12). Thus, the Gauss–Newton method with Armijo line search for solving the V subproblem is summarized in Algorithm 3.

Algorithm 3 Compute V^{k+1} by using Gauss–Newton method (V-Subproblem-Optimizer).

```
Input: $V^k, U^{k+1}, T, R, \beta, \gamma, \lambda^k, \epsilon_v$, MaxItV;
Initialization: $V^{k,1} = V^k$;
for $i = 1, \ldots,$ MaxItV do
    Compute gradient $g_{V^{k,i}}$ and Hessian $H_{V^{k,i}}$ by (21);
    Compute search direction $\delta_{V^{k,i}}$ by solving $H_{V^{k,i}}\delta_{V^{k,i}} = -g_{V^{k,i}}$;
    Update $V^{k,i+1} = V^{k,i} + \mu_{V^{k,i}}\delta_{V^{k,i}}$ by Armijo line search [50, Algorithm 3.1];
    if $|\mathcal{L}^2_{\lambda^k,U^{k+1}}(V^{k,i+1}) - \mathcal{L}^2_{\lambda^k,U^{k+1}}(V^{k,i})| \le \epsilon_v|\mathcal{L}^2_{\lambda^k,U^{k+1}}(V^{k,1})|$ or $\|V^{k,i+1} - V^{k,i}\| \le \epsilon_v\|V^{k,1}\|$ then break;
end
$V^{k+1} = V^{k,i+1}$;
Output: $V^{k+1}$.
```

4.3.3. *Multimodal diffeomorphic image registration algorithm*

In Sections 4.3.1 and 4.3.2, the utilization of L-BFGS and Gauss–Newton techniques to address subproblems in (15) is explored within the iterative penalty method. Consequently, Algorithm 4 summarizes the solution procedure for the proposed multimodal diffeomorphic image registration model (7).

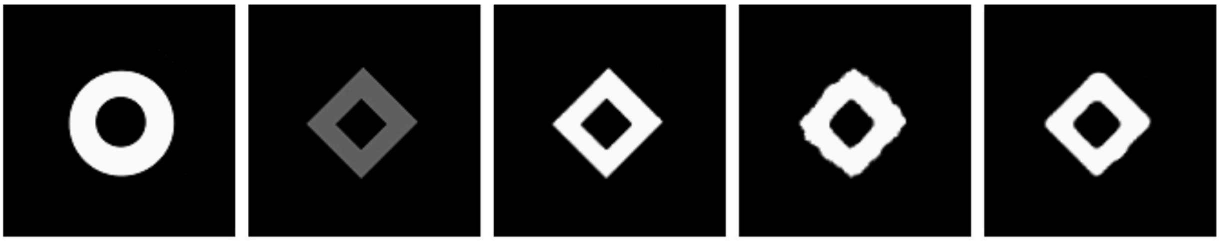
Due to the highly non-convex nature of mutual information with numerous local minima, the multilevel strategy effectively avoids getting trapped in false local minima, reducing total computation time. This method successively simplifies the original problem across levels. Initially, the subproblems in (15) are solved at the coarsest level using Algorithm 4. The results are then refined with the interpolation operator I_H^h to provide initial guesses for the subproblems of the next level [57]. This refinement continues until an adequately fine level is achieved, customized to application-specific needs. Detailed steps of the multilevel method are in Algorithm 5.

Algorithm 4 Multimodal diffeomorphic image registration based on penalty method (MDIRPM).

```
Input: $R, T, U^1, V^1, \alpha, \beta, \gamma, \lambda^1, \rho$, MaxIt, MaxItU, MaxItV, $\epsilon, \epsilon_u, \epsilon_v$;
for $k = 1, \dots,$ MaxIt do
    Compute displacement field $U^{k+1}$ by Algorithm 2, i.e.,
      $U^{k+1}$ = U-Subproblem-Optimizer($U^k, V^k, T, R, \alpha, \beta, \gamma, \lambda^k, \epsilon_u$, MaxItU);
    Compute grid unfolding indicator $\Theta$ by (14) and $\Theta_{\min} := \min\{\Theta\}$;
    if $\Theta_{\min} < 1e-2$ then Correct $U^{k+1}$ by the deformation correction [49, Algorithm 4.1];
    Compute $V^{k+1}$ by Algorithm 3, i.e.,
      $V^{k+1}$ = V-Subproblem-Optimizer($V^k, U^{k+1}, T, R, \beta, \gamma, \lambda^k, \epsilon_v$, MaxItV);
    Update the penalty factor $\lambda^{k+1} = \rho\lambda^k$;
    Update template image $T^{k+1} = T(X + U^{k+1})$;
    if $|\mathcal{L}_{\lambda^{k+1}}(U^{k+1}, V^{k+1}) - \mathcal{L}_{\lambda^{k+1}}(U^k, V^k)| \le \epsilon|\mathcal{L}_{\lambda^{k+1}}(U^1, V^1)|$ or $\|(U^{k+1} - U^k, V^{k+1} - V^k)\| \le \epsilon\|(U^1, V^1)\|$ then break;
end
$T^{new} = T^{k+1},\ U^{new} = U^{k+1},\ V^{new} = V^{k+1}$;
Output: $T^{new}, U^{new}, V^{new}$.
```

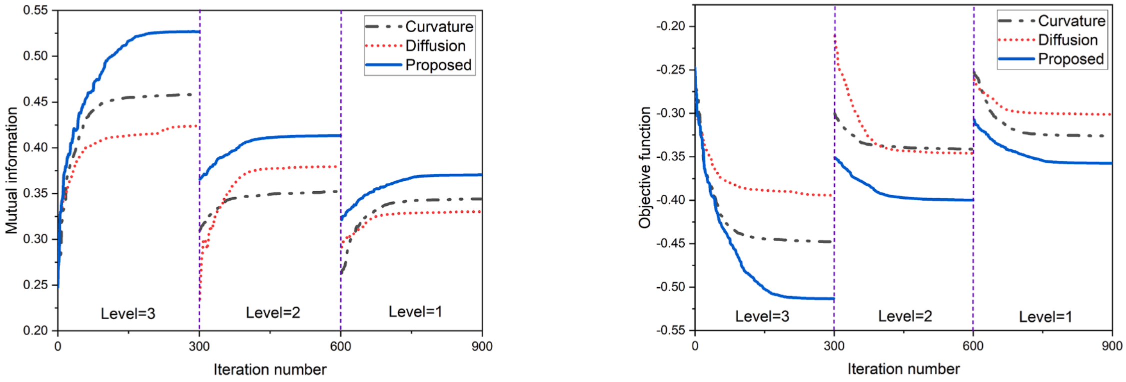
Algorithm 5 Multilevel Registration.

```
Input: $R, T, L, \alpha, \beta, \gamma, \lambda, \rho$, MaxIt, MaxItU, MaxItV, $\epsilon, \epsilon_u, \epsilon_v$;
Initialization: $R_1 = R, T_1 = T, U_L = 0, V_L = 1$;
for $\ell = 2, \dots, L$ do
    Compute the reference and template images for the $\ell$-level coarse grid:
      $R_\ell = I_h^H R_{\ell-1}, T_\ell = I_h^H T_{\ell-1}$;
end
for $\ell = L, \dots, 1$ do
    Call the diffeomorphic image registration Algorithm 4 based on penalty method:
      $[T_\ell^{new}, U_\ell^{new}, V_\ell^{new}]$ = MDIRPM($R_\ell, T_\ell, U_\ell, V_\ell, \alpha, \beta, \gamma, \lambda, \rho$, MaxIt, MaxItU, MaxItV, $\epsilon, \epsilon_u, \epsilon_v$);
    if $\ell > 1$ then $U_{\ell-1} = I_H^h U_\ell^{new}, V_{\ell-1} = I_H^h V_\ell^{new}$;
end
Output: $T_1^{new}, U_1^{new}$.
```



(a) Template (b) Reference (c) Proposed (d) Diffusion (e) Curvature

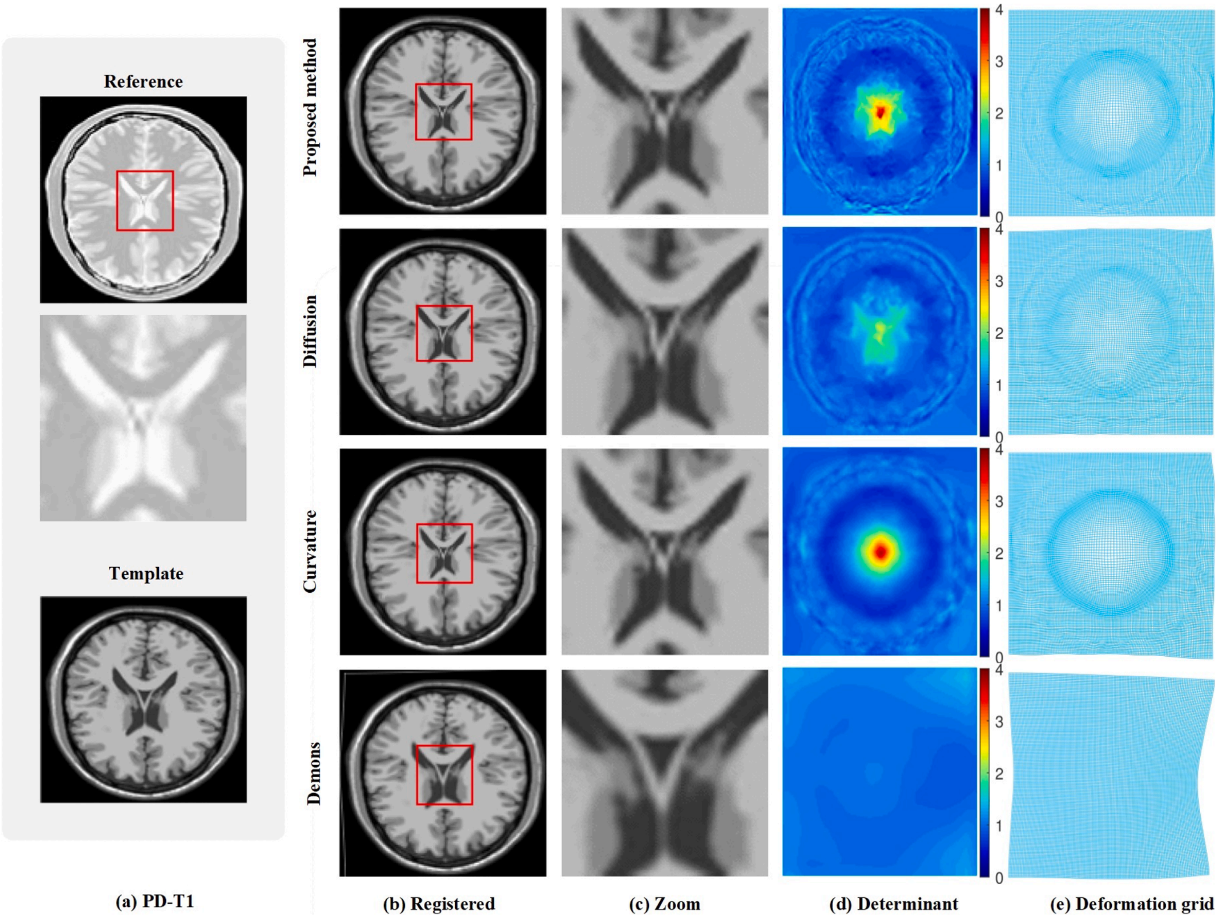
Fig. 2. Registration results for synthetic images across three methods. (a) template image; (b) reference image; (c) registered image of the proposed method with the optimal parameters: $\alpha = 0.4, \beta = 0.001, \gamma = 0.01, \lambda = 0.1, \rho = 1.01, n_r = n_t = 40$; (d) registered image of the diffusion method with the optimal parameters: $\alpha = 0.0001$, $n_r = n_t = 40$; (e) registered image of the curvature method with the optimal parameters: $\alpha = 0.0001$, $n_r = n_t = 40$.



(a) The trend of mutual information measure changes

(b) The trend of objective function changes

Fig. 3. Circle-Square example. The black double-dotted line, red dashed line, and blue solid line indicate the results of the curvature, diffusion, and proposed models, respectively.



(a) PD-T1   (b) Registered   (c) Zoom   (d) Determinant   (e) Deformation grid

Fig. 4. Registration results of PD-T1 images across four methods. (a) reference and template images; (b) registered images using optimal parameters (Proposed method: $\alpha = 4, \beta = 0.001, \gamma = 0.01, \lambda = 0.04, \rho = 1.05, n_r = n_t = 80$; Diffusion method: $\alpha = 0.0001$, $n_r = n_t = 60$; Curvature method: $\alpha = 0.001$, $n_r = n_t = 30$; Demons method: $\alpha = 2$); (c) zoomed regions of interest; (d) Jacobian determinant hotmaps of deformation fields; (e) deformation grids.

5. Numerical experiments

To assess the registration performance of the proposed method in comparison with other approaches, we conduct experiments using both synthetic and real medical data. The synthetic data includes English characters and simple geometric shapes across different modalities, while the medical data comprises CT and MR (T1-, T2-, and PD-weighted) images. We quantitatively evaluate image matching using mutual information and the Dice coefficient. Additionally, we report the minimum value of the grid unfolding indicator $\Theta_{\min}$, as well as the minimum and maximum values of the Jacobian determinant, to verify the diffeomorphism of the deformation.

i.) The Dice coefficient is defined as

$$\mathrm{Dice}(R, T) = 2\frac{|R(X) \cap T(X+U)|}{|R(X)| + |T(X+U)|},$$

where $R(X)$ and $T(X+U)$ represent the segmented regions of interest in the reference and deformed template images, respectively. These segments are obtained using MATLAB's built-in *imsegkmeans* function, which is based on the K-means clustering image segmentation method [58]. A Dice coefficient of $\mathrm{Dice} = 1$ indicates perfect alignment of the segmentation boundaries, while $\mathrm{Dice} = 0$ signifies no overlap of the segmented regions after registration.

ii.) The minimum of the grid unfolding indicator $\Theta_{\min}$ is defined as

$$\Theta_{\min} = \min\{\Theta\},$$

where the calculation of the grid unfolding indicator Θ is provided in (14).

iii.) The Jacobian determinant measures are defined as

$$\overline{\det} = \text{mean } \mathrm{Det}(U), \quad \det_{\min} = \min \mathrm{Det}(U), \quad \det_{\max} = \max \mathrm{Det}(U),$$

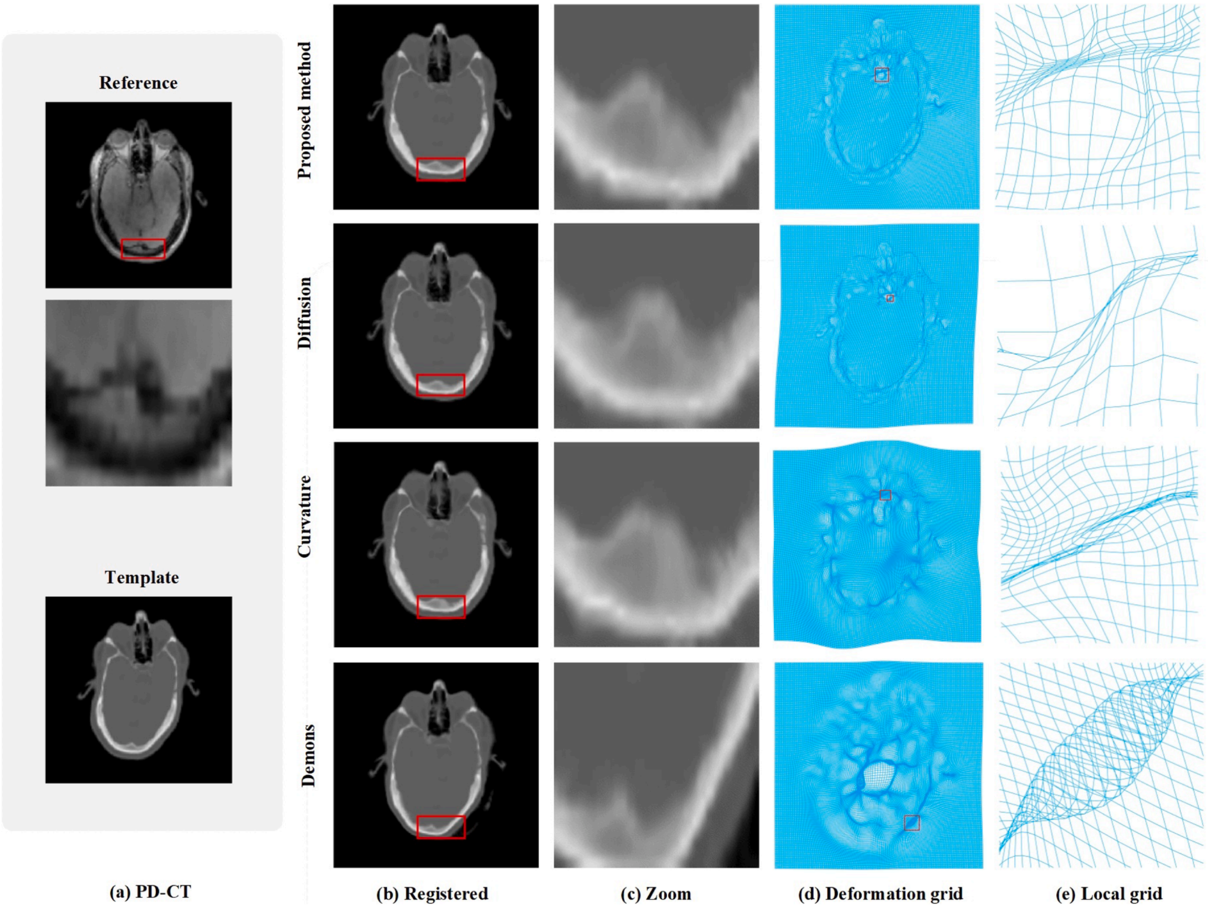


Fig. 5. Registration results of PD-CT images across four methods. (a) reference and template images; (b) registered images using optimal parameters (Proposed method: $\alpha = 1.5, \beta = 0.001, \gamma = 0.01, \lambda = 0.01, \rho = 1.05$, $n_r = n_t = 80$; Diffusion method: $\alpha = 0.00005$, $n_r = n_t = 40$; Curvature method: $\alpha = 0.0001$, $n_r = n_t = 40$; Demons method: $\alpha = 1$); (c) zoomed regions of interest; (d) deformation grids; (e) local enlargement of deformation grids.

where mean $\mathrm{Det}(U)$ denotes the average value of the Jacobian determinant $\mathrm{Det}(U)$, and the calculation of $\mathrm{Det}(U)$ is described in (13).

In the following numerical experiments, we aim to demonstrate the effectiveness of our method by comparing it with several state-of-the-art multimodal approaches. The implementations of three such methods are as follows: (a) the Diffusion method[1] [59], (b) the Curvature method[1] [41], and (c) the Demons method[2] [21]. Additionally, the parameters for all methods are selected based on the principle of optimizing mutual information (MI). In our method, the maximum number of inner iterations for the two subproblems is set to MaxItU = MaxItV = 30, and for the global registration problem, it is set to MaxIt = 300. The tolerance for the two subproblems is set to $\epsilon_u = \epsilon_v = 10^{-4}$, while for the global registration problem, it is set to $\epsilon = 10^{-5}$. Finally, our program is implemented in MATLAB R2022a, and the experiments are performed on a machine equipped with an Intel(R) Core(TM) Ultra 7 155H CPU at 3.80 GHz and 32 GB of RAM.

5.1. Comparisons for the numerical convergence of the proposed algorithm

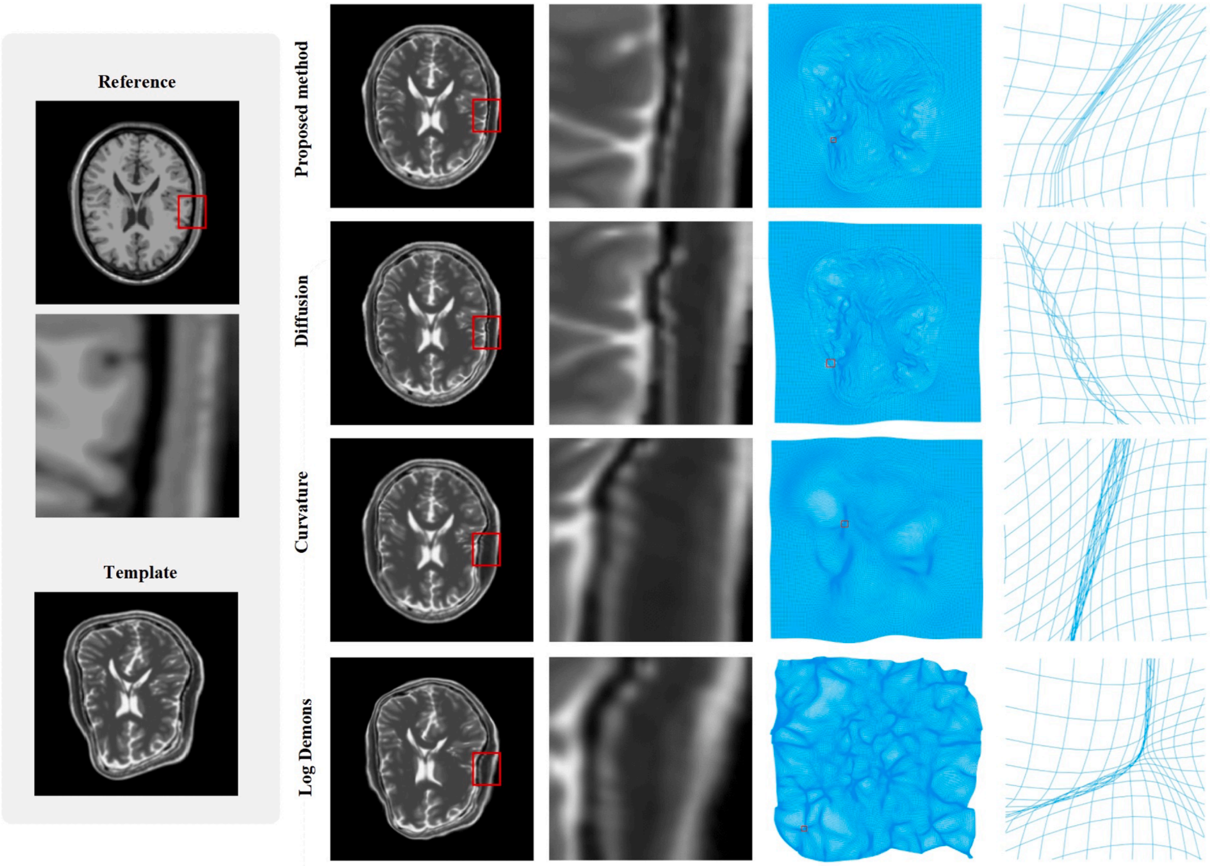
In this section, we mainly verify the numerical convergence of the proposed algorithm through numerical experiments on a pair of synthetic multimodal images with a resolution of 128×128. We also compare our method with the diffusion- and curvature-based registration algorithms. For the sake of comparison, all three methods are forced to run 300 iterations at each level.

Fig. 2 shows the deformation results for the Circle-Square example using our proposed model, as well as the diffusion- and curvature-based models [41,59]. Here, we observe that our method produces results that closely match the ground truth reference image, while the diffusion and curvature models exhibit unsmooth corners and edges that do not align well with the reference

---

[1] https://github.com/C4IR/FAIR.m

[2] https://www.mathworks.com/matlabcentral/fileexchange/21451-multimodality-non-rigid-demon-algorithm-image-registration



(a) T1-T2 (b) Registered (c) Zoom (d) Deformation grid (e) Local grid

Fig. 6. Registration results of T1-T2 images across four methods. (a) reference and template images; (b) registered images using optimal parameters (Proposed method: $\alpha = 1.5, \beta = 0.001, \gamma = 0.01, \lambda = 0.02, \rho = 1.02$, $n_r = n_t = 80$; Diffusion method: $\alpha = 0.00003$, $n_r = n_t = 30$; Curvature method: $\alpha = 0.0001$, $n_r = n_t = 40$; Demons method: $\alpha = 3$); (c) zoomed regions of interest; (d) deformation grids; (e) local enlargement of deformation grids.

Table 1
A quantitative comparison of the proposed method, along with the diffusion and curvature methods, for the Circle-Square example. The best metric values are highlighted in **bold**.

Example	Methods	$\overline{\det}$	$\Theta_{\min}$	$\det_{\min}$	$\det_{\max}$	Dice	MI
Circle-Square	Proposed	1.000	+	0.43	2.91	**0.9878**	**0.37**
	Diffusion	1.031	+	0.23	2.14	0.9511	0.31
	Curvature	1.042	+	0.75	2.24	0.9632	0.34

image. Table 1 presents quantitative results, demonstrating that the deformations produced by all three methods are diffeomorphic, as indicated by $\Theta_{\min} > 0$. Furthermore, our proposed model achieves the highest mutual information and Dice coefficient, indicating superior alignment between the registered image and the reference image. Therefore, in this experiment, our method delivers better results both qualitatively and quantitatively.

Fig. 3 illustrates the convergence performance of the three methods on the Circle-Square example. The horizontal axis represents the number of iteration steps, while the vertical axis shows both the mutual information measure and the objective function value. At the coarse level, the image retains more structural information, making registration easier and allowing for higher mutual information and objective function values. At the fine level, the image contains additional detailed information, which complicates the registration process, resulting in slightly lower mutual information and objective function values. As the number of iterations increases at each level, we observe that the mutual information measure from our method steadily increases and eventually stabilizes, indicating improved registration accuracy that gradually approaches the ideal state. Additionally, the mutual information value for our model consistently outperforms those of the curvature and diffusion models. At the same time, the objective function value decreases monotonically and stabilizes, indicating that all three registration algorithms are numerically

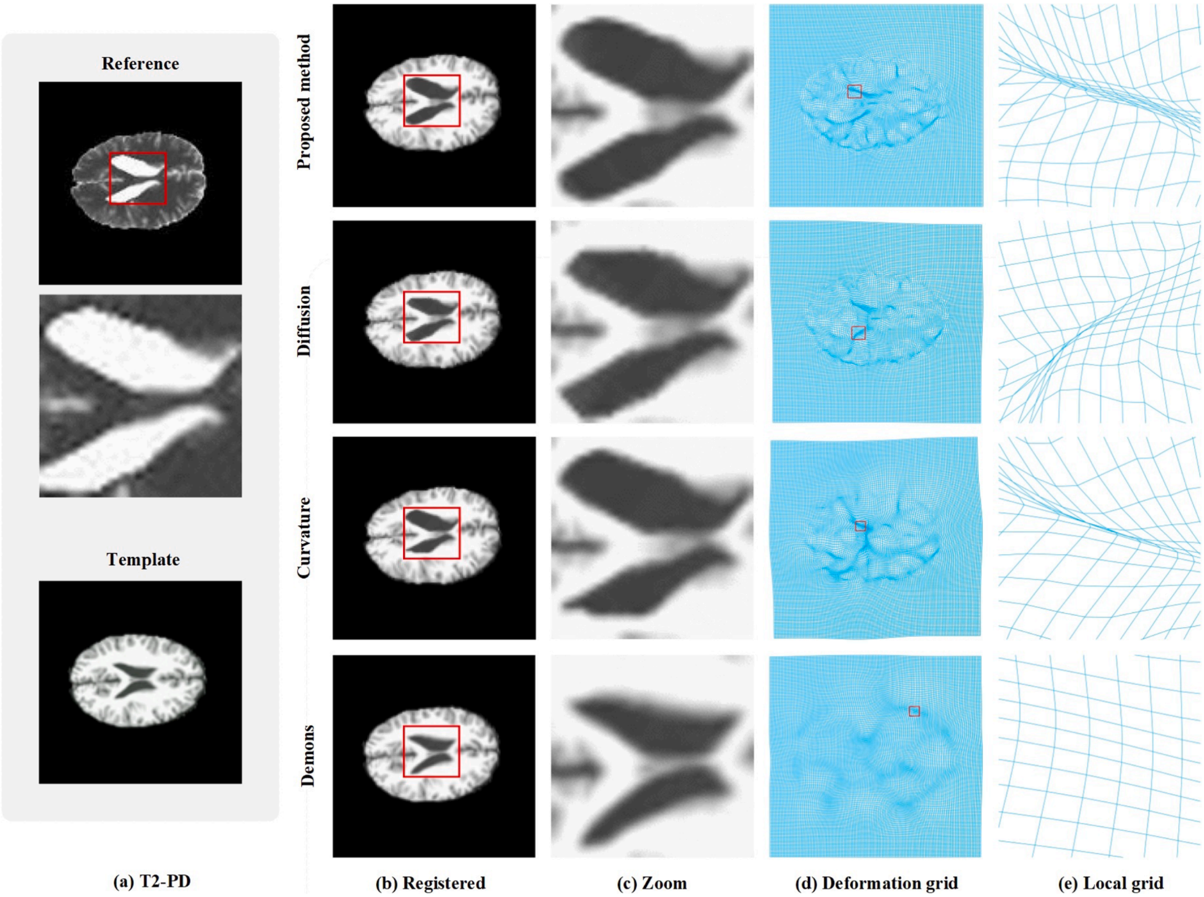


Fig. 7. Registration results of T2-PD images across four methods. (a) reference and template images; (b) registered images using optimal parameters (Proposed method: $\alpha = 2, \beta = 0.001, \gamma = 0.001, \lambda = 0.01, \rho = 1.03$, $n_r = n_t = 32$; Diffusion method: $\alpha = 0.0001$, $n_r = n_t = 14$; Curvature method: $\alpha = 0.001$, $n_r = n_t = 20$; Demons method: $\alpha = 4$); (c) zoomed regions of interest; (d) deformation grids; (e) local enlargement of deformation grids.

convergent. This demonstrates that our algorithm not only exhibits numerical convergence but also achieves superior registration accuracy.

5.2. Comparison with state-of-the-art registration methods

To assess the effectiveness of the proposed method in registering multimodal images, we conducted several experiments and compared our approach with the diffusion, curvature, and Demons methods [21]. The Demons method referenced here, developed by Kroon et al., involves an MRI modality transformation that converts a T1 scan into a T2 scan using the peaks in a joint histogram. In these experiments, we utilized T1- and T2-weighted MR images with a resolution of 256×256, CT and PD-weighted MR images with a resolution of 200×200, PD- and T1-weighted MR images with a resolution of 128×128, and PD- and T2-weighted MR images with a resolution of 128×128 for medical image registration. Additionally, synthesized CG and RA images, each with a resolution of 128×128, were used for large deformation registration. For all methods, a multilevel strategy with $L = 3$ was applied, and $\text{MaxIt} = 300$ iterations were performed.

Figs. 4 to 9 show visualizations of the registration results for all methods, including the reference image, the registered template image, magnified views of regions of interest, and deformation grids. Specifically, we applied local averaging to the Jacobian determinant using a mean filter, identified the window with the minimum average Jacobian determinant, and zoomed in on it, as shown in Figs. 5(e) to 9(e). The formula for locating the window with the minimum mean Jacobian determinant is given by:

$$(i^*, j^*) = \arg\min_{(i,j)} \frac{1}{(2w+1)^2} \sum_{k=-w}^{w} \sum_{l=-w}^{w} \text{Det}(map(i+k, j+l)), \tag{22}$$

where $map(i,j) = (i-1)n + j$, (i^*, j^*) denotes the center of the zoomed region with a window size of $(2w+1) \times (2w+1)$, and $w = 4$.

For the PD- and T1-weighted MR images in Fig. 4, we observe that our proposed method, along with the curvature method, yields visually more satisfactory registration results compared to the diffusion and Demons models. Furthermore, as shown in Table 2, all

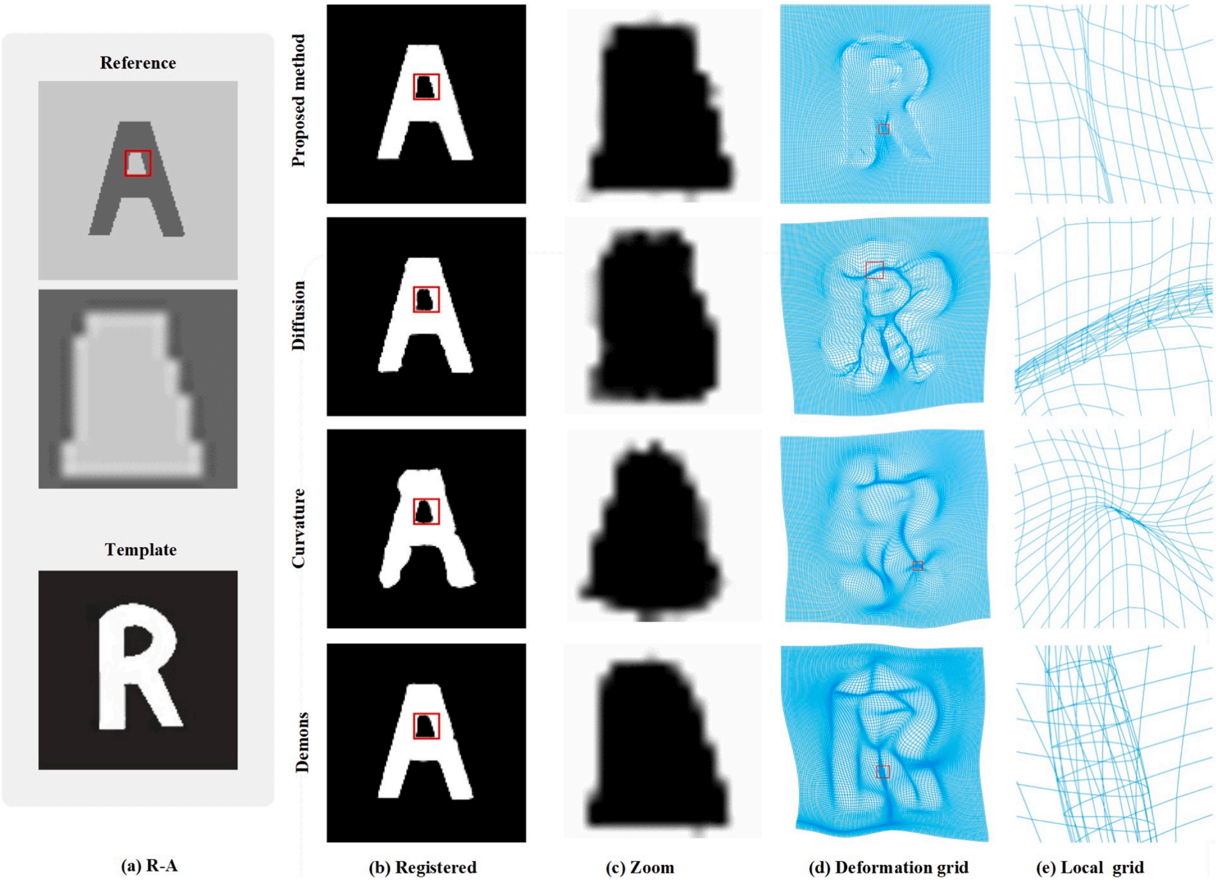


Fig. 8. Registration results of R-A images across four methods. (a) reference and template images; (b) registered images using optimal parameters (Proposed method: $\alpha = 0.45, \beta = 0.001, \gamma = 0.01, \lambda = 0.1, \rho = 1.02,\ n_r = n_t = 40$; Diffusion method: $\alpha = 0.0001,\ n_r = n_t = 8$; Curvature method: $\alpha = 0.0001,\ n_r = n_t = 20$; Demons method: $\alpha = 4$); (c) zoomed regions of interest; (d) deformation grids; (e) local enlargement of deformation grids.

models produce diffeomorphic mappings. Notably, our method achieves a mutual information value of 1.65 and a Dice coefficient of 0.8 for the region of interest. For the Demons model, despite having $\Theta_{\min} > 0$, the template image undergoes minimal deformation, resulting in the poorest registration performance. For the PD-weighted MR and CT images in Fig. 5, the proposed method, as well as the diffusion and curvature models, successfully complete the registration. Although the Dice coefficient of the diffusion model is slightly higher than that of our method, our approach achieves the highest mutual information. Additionally, the deformation grids generated by the diffusion and curvature methods exhibit folds. By inspecting the zoomed-in regions of the registered T1- and T2-weighted MR images and T2- and PD-weighted MR images in Figs. 6(c) and 7(c), we can see that the deformation template images produced by our method are closer to the reference image compared to the other methods, and our deformation grid is free from folding, as shown in Figs. 6(e) and 7(e). Although the Demons method aims to map T1 scans to T2 scans for registration, its performance is inferior in the T1-T2 experiment shown in Fig. 6. Additionally, Table 2 confirms that our method performs well quantitatively. It is noteworthy that the T2- and PD-weighted MR image experiments conducted with the diffusion model exhibit a phenomenon where $\det_{\min} = 0.001$ but $\Theta_{\min} < 0$, indicating that the mapping is non-diffeomorphic, as seen in Fig. 7(e). This suggests that the Jacobian determinant value obtained by the diffusion method may not always accurately reflect the diffeomorphism of the mapping. In contrast, our method for calculating the Jacobian determinant is more reliable, as it better detects whether deformation has occurred. Overall, for the four sets of experiments mentioned above, our proposed method achieves excellent results, demonstrating that our algorithm performs well in handling multimodal medical images with complex structures.

Figs. 8–9 present the registration results of synthetic images with large deformations produced by the four methods. For the CG image pairs, it is clear that both our proposed model and the curvature model produce more visually satisfactory registration results. As shown in Table 2, our method achieves the highest mutual information and Dice coefficient among the models. For the RA image pairs, a close inspection of the zoomed-in regions reveals that our proposed model provides more accurate registration results compared to the other methods. Additionally, as indicated in Table 2, our proposed method achieves the highest mutual information and Dice coefficient within the region of interest. While the diffusion and Demons models produce mutual information and Dice coefficient values close to ours, the deformed template images in the zoomed-in regions do not align well with the reference image, and their handling of image details is less accurate than our method. This discrepancy may stem from insufficient sampling and the

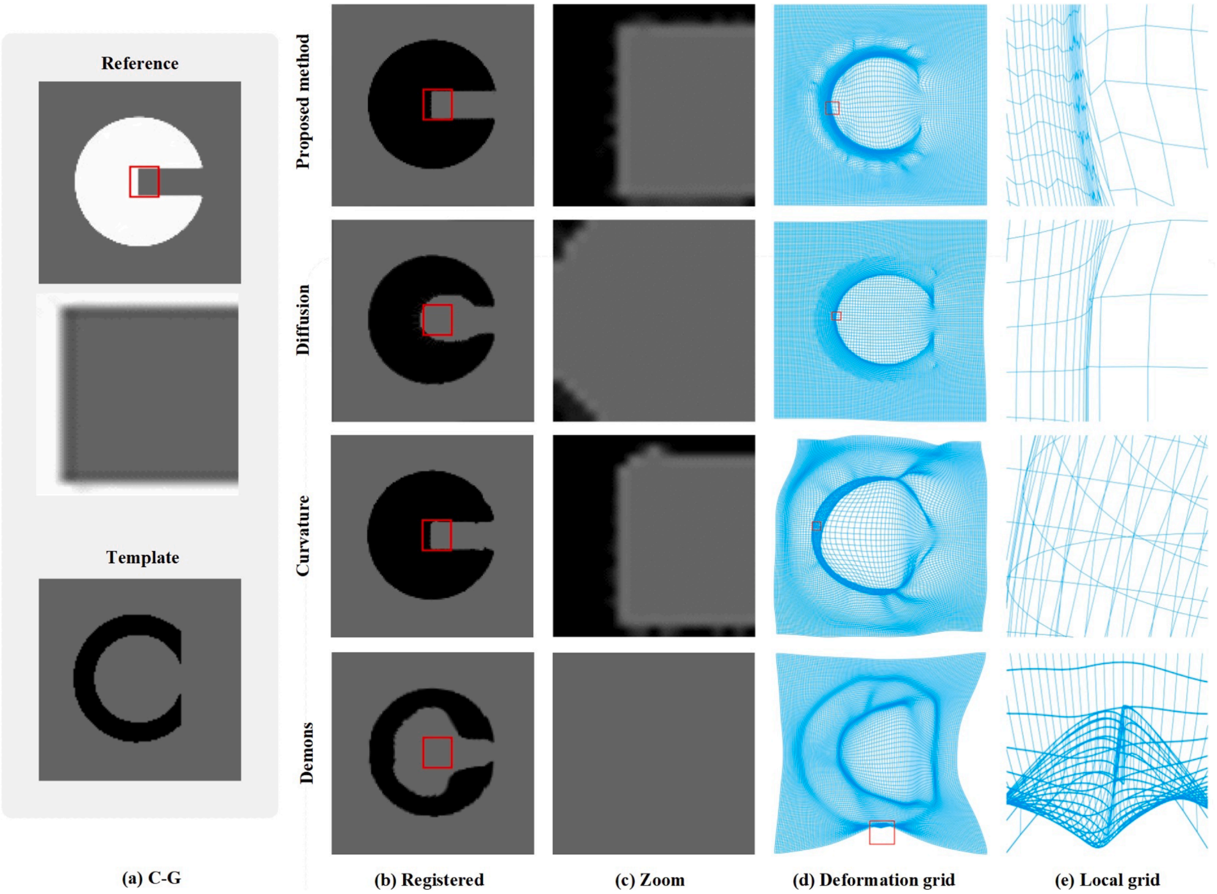


Fig. 9. Registration results of C-G images across four methods. (a) reference and template images; (b) registered images using optimal parameters (Proposed method: $\alpha = 0.1, \beta = 0.001, \gamma = 0.1, \lambda = 0.1, \rho = 1.02, n_r = n_t = 40$; Diffusion method: $\alpha = 0.0001, n_r = n_t = 10$; Curvature method: $\alpha = 0.0001$, $n_r = n_t = 40$; Demons method: $\alpha = 3$); (c) zoomed regions of interest; (d) deformation grids; (e) local enlargement of deformation grids.

presence of local maxima in the mutual information, which prevent these methods from achieving the desired deformation field. In contrast, our model offers precise control over transformations, allowing local deformations to shrink or expand within a flexible range. This level of control contributes to achieving the ideal deformation. Furthermore, for both pairs of synthetic images with large deformations, the deformation grids generated by the diffusion, curvature, and Demons models exhibit folding with $\Theta_{\min} < 0$, indicating that their deformations are non-diffeomorphic.

The numerical experiments conducted on four pairs of medical images and two pairs of synthetic images with large deformations demonstrate that our proposed model delivers the best registration results without grid foldings. This indicates that our multimodal image registration method is highly applicable. In terms of computational time, the Demons method exhibits the lowest computational cost (see Table 2). However, its registration accuracy and diffeomorphism cannot be guaranteed. Among the remaining three methods, our approach consistently requires the least computational time while achieving higher mutual information and preserving diffeomorphism. This emphasizes the computational efficiency advantage of our method.

5.3. Parameter selection and sensitivity analysis

This section aims to evaluate the sensitivity of the parameters $\alpha, \beta, \gamma, \lambda, \rho$, and n_r (where $n_t = n_r$ consistently) within our model, focusing on their impact on registration accuracy and convergence speed. To ensure reliability, parameters must be tested and adjusted appropriately, acknowledging that optimal values can differ across images. We employed a synthetic pair of C-G images with a 128×128 resolution (refer to Fig. 9) and utilized a grid search method to outline parameter selection. For comparison, a three-level multilevel scheme ($L = 3$) was used, allowing up to MaxIt $= 100$ iterations per level. The parameter tuning and evaluation methodology is outlined below.

The influence of various parameter configurations on registration results is shown in Fig. 10, which illustrates the variation of Dice, MI, and cumulative iterations (Iter represents total iterations across levels). For optimal parameters, we focus on minimizing iterations while maintaining registration precision. For parameter α, values between 10^{-2} and 10^{2} were examined, keeping $\beta = 0.001,\ \gamma = 0.1,\ \lambda = 0.1,\ \rho = 1.02,\ n_r = 40$ constant. As shown in Fig. 10, effective model solutions occur when α is within 10^{-2} to 1,

Table 2
A quantitative comparison of the proposed method, along with the diffusion, curvature methods, and Demons methods, for six examples. The best metric values are highlighted in **bold**. The underline represents cases where the grid is folded, but the Jacobian determinant remains positive.

Examples	Methods	$\overline{\det}$	$\Theta_{\min}$	$\det_{\min}$	$\det_{\max}$	Dice	MI	Time (s)
PD-T1	Proposed	0.999	+	0.08	3.86	**0.8000**	**1.65**	17.98
	Diffusion	0.981	+	0.57	2.29	0.7871	1.01	20.30
	Curvature	0.981	+	0.55	3.74	0.7993	1.21	52.96
	Demons	1.038	+	0.88	1.40	0.6643	0.78	12.39
PD-CT	Proposed	1.000	+	0.02	3.98	0.9390	**0.90**	43.46
	Diffusion	0.963	−	−0.16	2.81	**0.9393**	0.83	86.06
	Curvature	1.013	−	−0.24	3.01	0.9368	0.84	95.84
	Demons	1.00	−	−1.92	6.11	0.8916	0.61	30.60
T1-T2	Proposed	0.999	+	0.03	2.88	**0.7952**	**1.18**	70.20
	Diffusion	1.014	−	−0.45	2.66	0.7923	0.95	74.97
	Curvature	0.998	−	−0.18	2.78	0.7770	0.92	74.30
	Demons	1.133	−	−1.18	4.04	0.6529	0.67	38.44
T2-PD	Proposed	1.000	+	0.03	2.88	**0.8972**	**0.84**	47.54
	Diffusion	0.968	−	$\underline{0.001}$	2.12	0.8766	0.73	80.14
	Curvature	0.989	−	−0.002	2.08	0.8925	0.77	51.60
	Demons	1.002	+	0.48	1.56	0.8447	0.67	10.43
R-A	Proposed	0.999	+	0.29	4.36	**0.9953**	**0.40**	19.79
	Diffusion	1.199	−	−1.70	10.00	0.9951	0.39	94.45
	Curvature	1.001	−	−0.19	3.86	0.9476	0.35	14.40
	Demons	0.948	−	−0.88	5.78	0.9942	0.39	14.46
C-G	Proposed	1.002	+	0.04	5.67	**0.9957**	**0.62**	15.23
	Diffusion	0.904	−	−0.06	3.45	0.9543	0.49	88.91
	Curvature	0.859	−	−0.96	12.69	0.9940	0.60	50.95
	Demons	0.735	−	−0.15	4.16	0.7691	0.29	14.25

with peak Dice and MI at $\alpha = 10^{-1}$ and fewer iterations. Thus, we suggest using this as the regularization parameter in this case. The same approach is applied to identify optimal parameters for β, γ, λ, ρ, n_r.

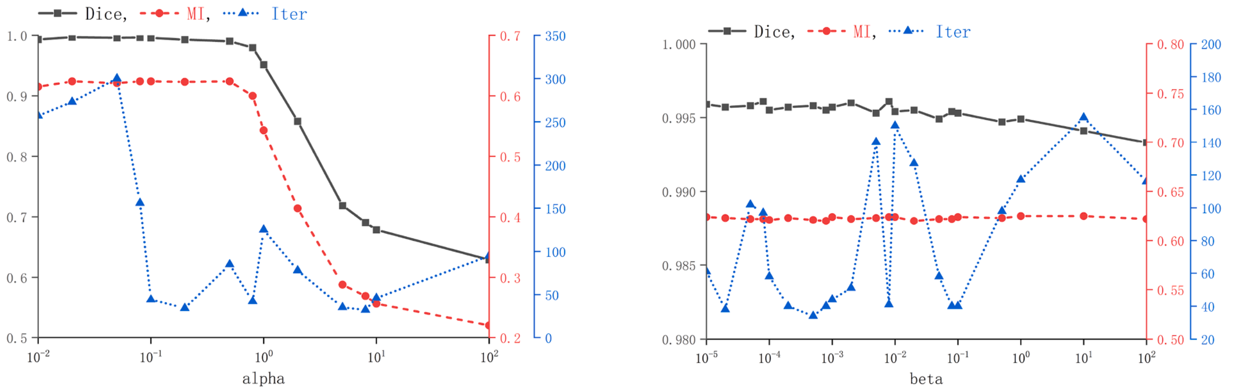
From analyzing the six parameters in Fig. 10, it is clear that the regularization parameter α and penalty parameter λ significantly affect both the sensitivity and convergence of the algorithm. Specifically, high values result in a noticeable decline in Dice and MI metrics and a sharp increase in iteration count, leading to reduced accuracy and slower convergence. However, within moderate thresholds, these parameters maintain high registration accuracy and facilitate relatively fast convergence. In contrast, the penalty parameter β, the regularization parameter γ, and the growth factor ρ remain stable over a wide range; they have little effect on Dice and MI, and only slight fluctuations in the iteration count. This indicates the algorithm's insensitivity to these parameters. The number of histogram bins n_r has an intermediate impact: increasing the bin count slightly reduces Dice and MI while gradually increasing iterations, implying that excessive bins may introduce noise and computational burden. Thus, n_r should be kept moderate to balance accuracy and efficiency. Consequently, parameter tuning should prioritize optimizing α and λ, with secondary control over n_r, while β, γ, and ρ can remain within empirical ranges without substantial adjustment.

5.4. Generalization ability of the proposed model

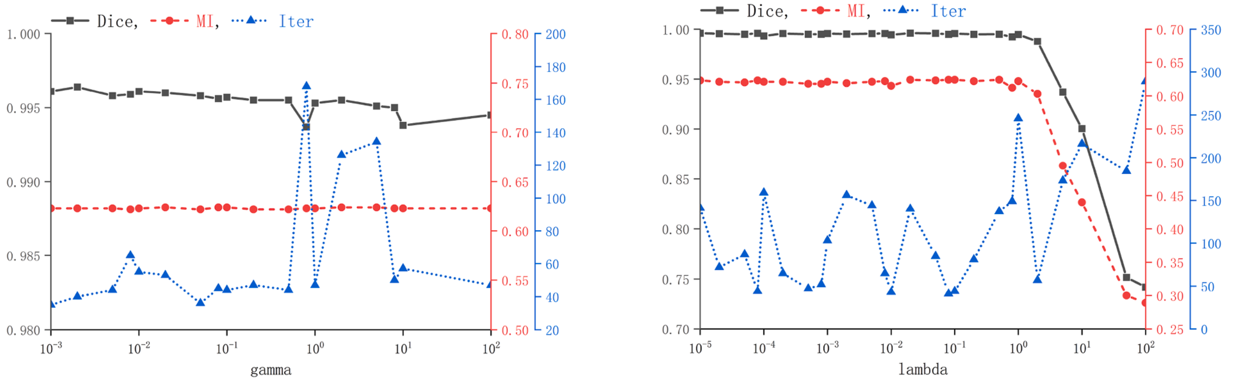
Traditional methods often face limitations in determining both model and algorithmic parameters, which may weaken their ability to generalize. In this context, generalization refers to how well a model can accurately predict outcomes on unseen data, even when such data are derived from a specific dataset. It is widely acknowledged in the deep learning community that both model and algorithmic parameters can be learned and optimized from large-scale training data. In view of the data-related limitations of traditional methods, we evaluated the generalization capability of the proposed model by computing a set of universal parameters, obtained by averaging the parameters from the seven experimental cases discussed above. Formally, each universal parameter is defined as

$$p^{\star} = \frac{1}{n}\sum_{i=1}^{n} p_i, \quad p_i \in \{\alpha_i, \beta_i, \gamma_i, \lambda_i, \rho_i\},\ n = 7.$$

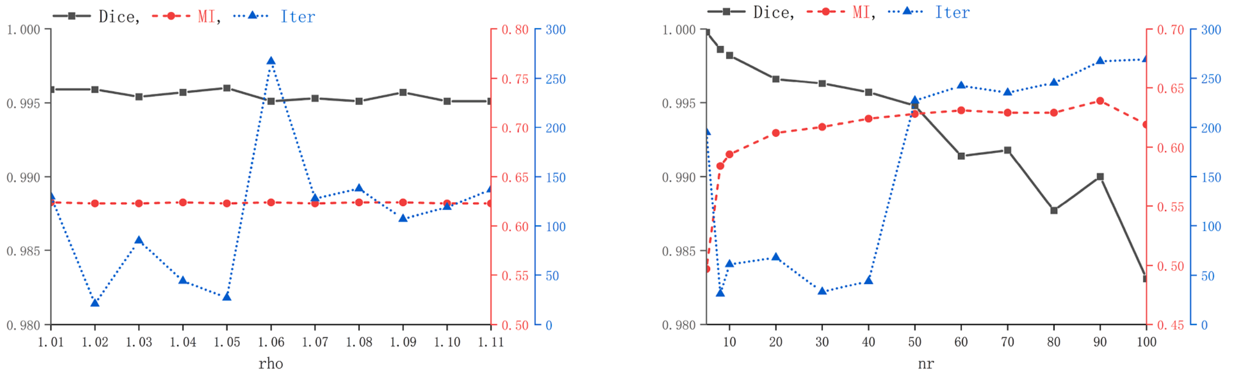
The resulting parameter configuration was then applied to predict outcomes on two types of new images from different sources: a synthetic image (Double-Beans, resolution 512×512) and a medical image (Brain-T1-PD, resolution 512×512). Notably, the resolutions of these new images differ from those used in the previous experiments, thereby providing a more rigorous evaluation of generalization capability. For fair comparison, the same procedure was also applied to the diffusion and curvature methods. In all experiments, a five-level multilevel scheme ($L = 5$) was employed, with a maximum of MaxIt $= 100$ iterations at each level, and the number of bins was consistently fixed at $n_r = n_t = 40$. The relevant numerical results are presented in Fig. 11 and Table 3.



(a) Regularization parameter α (b) Penalty parameter β



(c) Regularization parameter γ (d) Penalty parameter λ

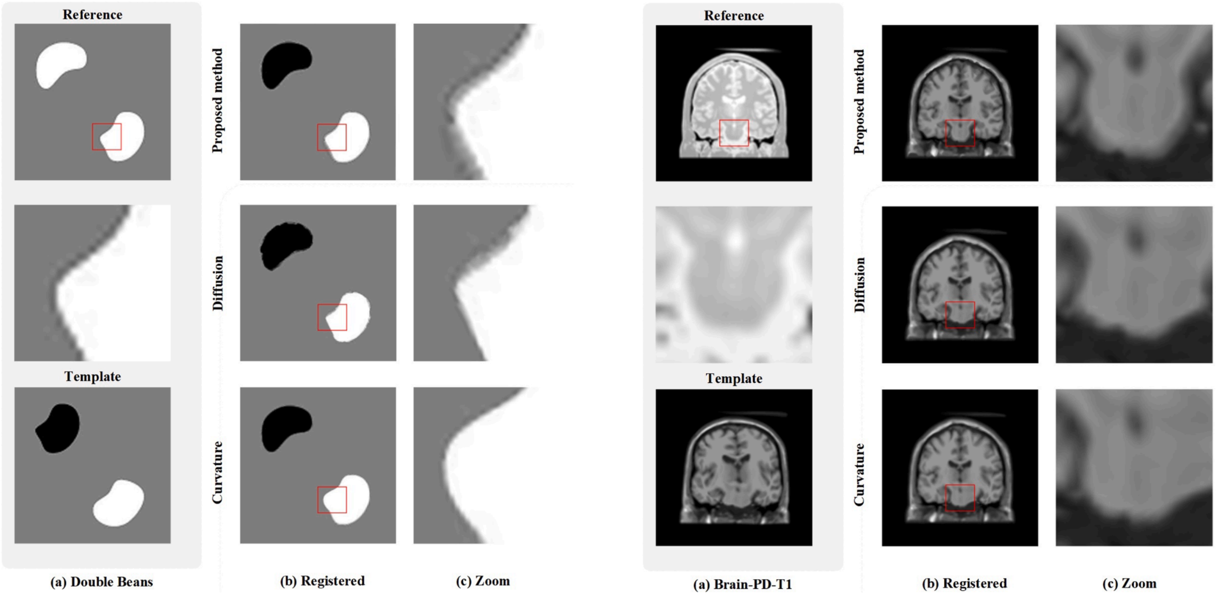


(e) Growth factor ρ (f) Number of bins n_r

Fig. 10. The influence of different parameters on the registration results for the C-G example. The black solid line, red dashed line, and blue dotted line represent Dice, MI, and Iter, respectively.

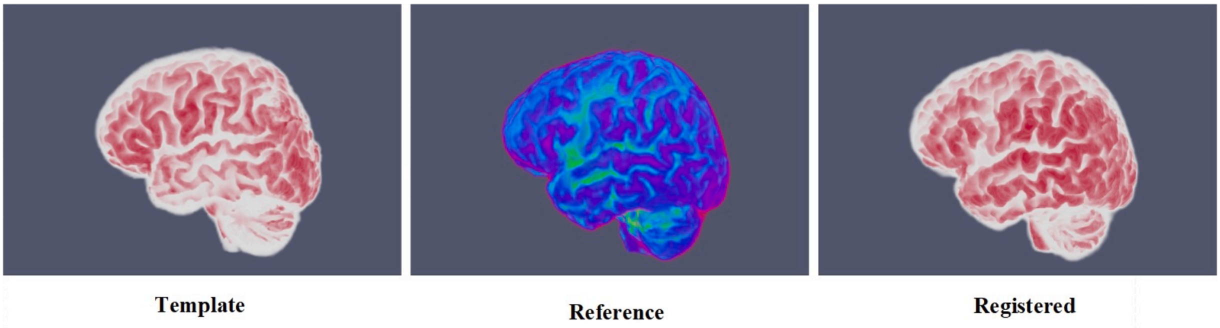
Table 3
The quantitative evaluation comparisons of the proposed, diffusion, and curvature methods. The best metric values are highlighted in **bold**.

Examples	Methods	$\overline{\det}$	$\Theta_{\min}$	$\det_{\min}$	$\det_{\max}$	Dice	MI
Double-Beans	Proposed	1.001	+	0.03	2.26	**0.9638**	**0.45**
	Diffusion	0.949	–	−0.39	3.37	0.9604	0.41
	Curvature	0.871	–	−0.30	3.04	0.9619	0.42
Brain-PD-T1	Proposed	0.999	+	0.02	5.12	**0.8802**	**1.02**
	Diffusion	1.030	+	0.15	3.31	0.8710	0.92
	Curvature	1.014	+	0.27	2.54	0.8741	0.95

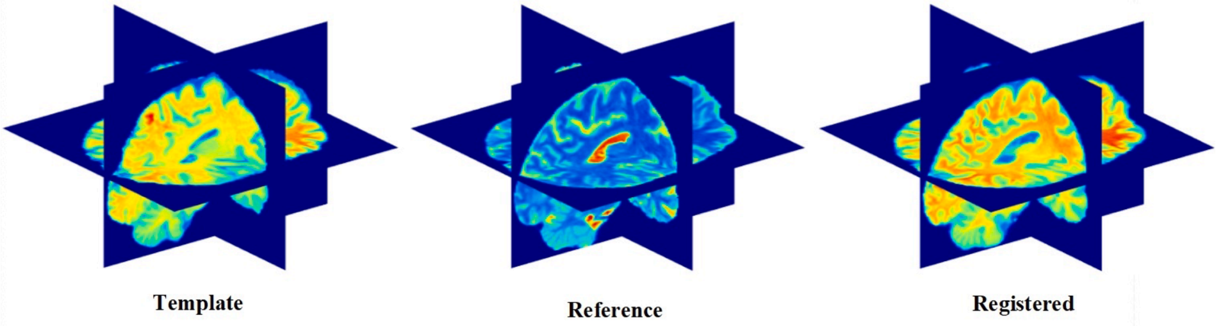


(a) Double-Beans (b) Brain-PD-T1

Fig. 11. Comparison of different models for Double-Beans and Brain-T1-PD using average parameters. (a) reference and template images; (b) registered images using average parameters (Proposed method: $\alpha^\star = 1.42, \beta^\star = 0.001, \gamma^\star = 0.02, \lambda^\star = 0.05, \rho^\star = 1.03$; Diffusion method: $\alpha^\star = 0.00008$; Curvature method: $\alpha^\star = 0.00036$); (c) zoomed regions of interest.



(a) 3D Brain



(b) 3D Brain slice planes

Fig. 12. Multimodal registration results for the proposed method operating on 3D Brain images. (a) shows the reference image, template image, and deformed template image of the 3D cerebral cortex surface. (b) visualize the reference image, template image, and deformed template images from slice planes. The optimal parameters for the 3D Brain images are $\alpha = 0.35, \beta = 0.001, \gamma = 0.001, \lambda = 0.1, \rho = 1.02, n_r = n_t = 60$.

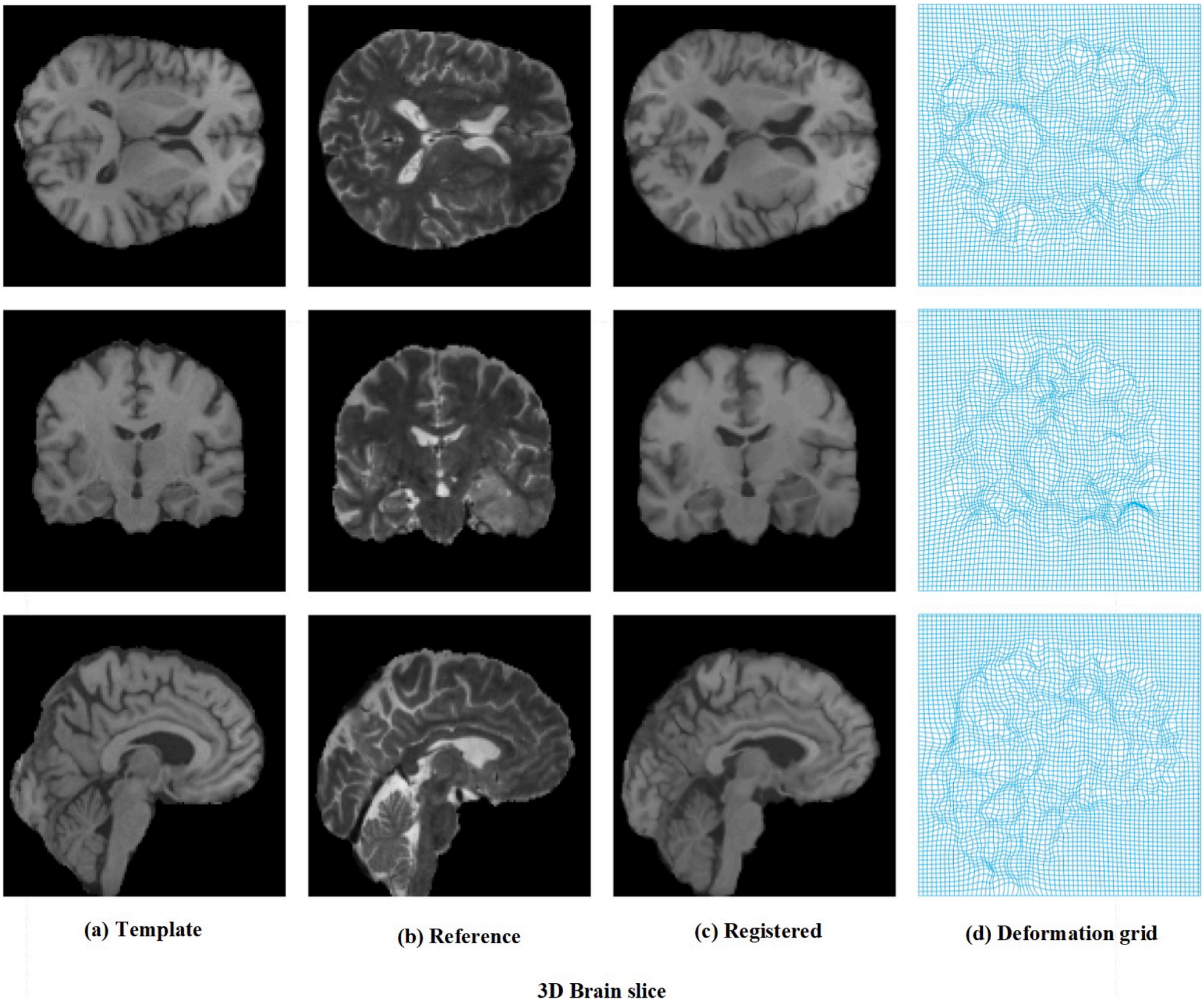


Fig. 13. Multimodal registration results of 3D Brain images. top row: transverse plane, middle row: sagittal plane, bottom row: coronal plane. (a) the template images; (b) the reference images; (c) the deformed template images; (d) the deformation grids.

As shown in Fig. 11, the proposed method yields more visually satisfactory registration results for both the Double-Beans and Brain-T1-PD images compared with the diffusion and curvature methods. In particular, for the structurally complex Brain-T1-PD images, our method achieves superior alignment quality. Furthermore, as reported in Table 3, the proposed method attains the highest mutual information and Dice coefficient within the region of interest. It is worth noting that for the Double-Beans images with large deformations, the deformation grids generated by the diffusion and curvature methods exhibit folding with $\Theta_{\min} < 0$, indicating non-diffeomorphic transformations. By contrast, our approach preserves diffeomorphic deformations under the same conditions. These findings confirm that the proposed method can deliver robust and satisfactory results for new images using averaged parameters, despite potential influences from specific parameter choices. Overall, the proposed model demonstrates stronger generalization ability than the compared state-of-the-art models.

5.5. 3D experiment

In this subsection, we use a pair of brain images with a resolution of $128 \times 128 \times 128$ to evaluate the effectiveness of the proposed approach for 3D image registration. By examining the texture structure of the 3D brain surface in Fig. 12(a) and the internal and edge regions in Fig. 12(b), we observe that the deformed template image closely matches the reference image. To further illustrate the registration results, transverse, sagittal, and coronal slices of the 3D brain images are shown in Fig. 13. It is evident that all deformed template images align almost perfectly with the corresponding reference images, with no evidence of folding in the transformations. Quantitative results for the 3D experiments are presented in Table 4, where $\Theta_{\min} > 0$ and the mutual information reaches 0.86. These results demonstrate that our proposed model effectively handles 3D images while maintaining diffeomorphic deformation.

Table 4
The quantitative evaluation of the proposed method for 3D multimodal experiment.

Example	Method	$\overline{\det}$	$\Theta_{\min}$	$\det_{\min}$	$\det_{\max}$	MI
Brain	Proposed	1.023	+	0.06	8.62	0.86

6. Conclusion

In this paper, we propose a bi-variate multimodal image registration model that utilizes negative mutual information as a similarity metric to quantify the statistical correlation or information redundancy between the image intensities of corresponding pixels/voxels in two images. To ensure a diffeomorphic transformation, we first introduce a relaxation constraint on the Jacobian determinant of the transformation, $\det(I_d + \nabla \boldsymbol{u}) = v(\boldsymbol{x}) > 0$, within the multimodal image registration framework. The positivity of the relaxation function, $v(\boldsymbol{x})$, is automatically maintained through a penalty term. Additionally, the relaxation function $v(\boldsymbol{x})$ serves to constrain the Jacobian determinant, allowing local deformations to shrink or expand within a flexible range. We also employ a penalty method to address the equality constraint problem and combine the L-BFGS and Gauss–Newton methods with Armijo line search in a multilevel strategy to solve the corresponding subproblems. Finally, experimental results demonstrate that our algorithm converges successfully, effectively preventing grid folding, and achieving highly satisfactory registration results for both medical images and images with large deformations across different modalities. Our future work will focus on developing a fast iterative solver.

CRediT authorship contribution statement

Yanyan Li: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Conceptualization; **Daoping Zhang:** Writing – review & editing, Validation, Methodology, Funding acquisition, Conceptualization; **Jianping Zhang:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 12201320, 12471484), as well as by grants from the Science and Technology Innovation Program of Hunan Province (Grant No. 2024RC9008).

Appendix A. Appendix

The main purpose of this appendix is to derive the Jacobian matrix of the discretized Jacobian determinant. To facilitate the presentation of the derivation, we first state the following proposition, which establishes the fundamental relation required for subsequent analysis.

Proposition 1. *Let the vector-valued function* $F : \mathbb{R}^n \to \mathbb{R}^m$ *be defined as*

$$F(U) = (AU) \odot (BU), \tag{A.1}$$

where $U = (u_1, u_2, \dots, u_n)^\top \in \mathbb{R}^n$, $A, B \in \mathbb{R}^{m\times n}$ *are given matrices, and* $\odot$ *denotes the Hadamard product. Then the Jacobian matrix of* $F(U)$ *is*

$$\frac{\partial F(U)}{\partial U} = \operatorname{diag}(AU)\, B + \operatorname{diag}(BU)\, A \in \mathbb{R}^{m\times n},$$

where $\operatorname{diag}(X)$ *denotes the diagonal matrix whose diagonal entries are given by the components of the vector* X.

Proof. Suppose $F(U) := (f_1, f_2, \dots, f_m)^\top$, $W := (w_1, w_2, \dots, w_m)^\top = AU$, $G := (g_1, g_2, \dots, g_m)^\top = BU$, with $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{m\times n}$. According to (A.1), each component f_i can be written as

$$f_i = \left(\sum_{k=1}^{n} a_{ik}u_k\right)\left(\sum_{k=1}^{n} b_{ik}u_k\right), \quad i = 1, 2, \dots, m.$$

Taking the partial derivative of f_i with respect to u_j yields

$$\frac{\partial f_i}{\partial u_j} = a_{ij}\left(\sum_{k=1}^{n} b_{ik}u_k\right) + b_{ij}\left(\sum_{k=1}^{n} a_{ik}u_k\right) = a_{ij}g_i + b_{ij}w_i, \quad j = 1, 2, \ldots, n.$$

Hence, the Jacobian matrix of $F(U)$ is

$$\begin{aligned}\frac{\partial F(U)}{\partial U} &= \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} & \cdots & \frac{\partial f_1}{\partial u_n} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} & \cdots & \frac{\partial f_2}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial u_1} & \frac{\partial f_m}{\partial u_2} & \cdots & \frac{\partial f_m}{\partial u_n} \end{bmatrix} = \begin{bmatrix} a_{11}g_1 + b_{11}w_1 & a_{12}g_1 + b_{12}w_1 & \cdots & a_{1n}g_1 + b_{1n}w_1 \\ a_{21}g_2 + b_{21}w_2 & a_{22}g_2 + b_{22}w_2 & \cdots & a_{2n}g_2 + b_{2n}w_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}g_m + b_{m1}w_m & a_{m2}g_m + b_{m2}w_m & \cdots & a_{mn}g_m + b_{mn}w_m \end{bmatrix} \\ &= \begin{bmatrix} a_{11}g_1 & a_{12}g_1 & \cdots & a_{1n}g_1 \\ a_{21}g_2 & a_{22}g_2 & \cdots & a_{2n}g_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}g_m & a_{m2}g_m & \cdots & a_{mn}g_m \end{bmatrix} + \begin{bmatrix} b_{11}w_1 & b_{12}w_1 & \cdots & b_{1n}w_1 \\ b_{21}w_2 & b_{22}w_2 & \cdots & b_{2n}w_2 \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1}w_m & b_{m2}w_m & \cdots & b_{mn}w_m \end{bmatrix} \\ &= \mathrm{diag(G)A + diag(W)B = diag(BU)A + diag(AU)B.}\end{aligned}$$

This completes the proof. ☐

Similar to Proposition 1, let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be defined by

$$F(U) = (AU) \odot (BU) \odot (CU),$$

then the Jacobian matrix of $F(U)$ is given by

$$\frac{\partial F(U)}{\partial U} = \mathrm{diag}(CU)\,\mathrm{diag}(BU)A + \mathrm{diag}(CU)\,\mathrm{diag}(AU)B + \mathrm{diag}(AU)\,\mathrm{diag}(BU)C \in \mathbb{R}^{m\times n}. \tag{A.2}$$

Next, we derive the Jacobian matrix $M(U)$ of the discretized Jacobian determinant $\mathrm{Det}(U)$ (see (13)). Before proceeding, we simplify (13) as follows:

$$\begin{aligned}\mathrm{Det}(U) = \frac{1}{4h^2}\Big[&4h^2e + hD^{+}_{2,2}U + hD^{+}_{1,1}U + (D^{+}_{1,1}U)\odot(D^{+}_{2,2}U) - (D^{+}_{2,1}U)\odot(D^{+}_{1,2}U) \\ &+ (D^{+}_{2,1}U)\odot(D^{-}_{1,2}U) + hD^{+}_{2,2}U - hD^{-}_{1,1}U - (D^{-}_{1,1}U)\odot(D^{+}_{2,2}U) \\ &- hD^{-}_{2,2}U - hD^{-}_{1,1}U + (D^{-}_{1,1}U)\odot(D^{-}_{2,2}U) - (D^{-}_{2,1}U)\odot(D^{-}_{1,2}U) \\ &+ (D^{-}_{2,1}U)\odot(D^{+}_{1,2}U) - hD^{-}_{2,2}U + hD^{+}_{1,1}U - (D^{+}_{1,1}U)\odot(D^{-}_{2,2}U)\Big] \\ = \frac{1}{4h^2}\Big[&4h^2e + 2hD^{+}_{1,1}U - 2hD^{-}_{1,1}U + 2hD^{+}_{2,2}U - 2hD^{-}_{2,2}U \\ &+ (D^{+}_{1,1}U)\odot(D^{+}_{2,2}U) - (D^{+}_{2,1}U)\odot(D^{+}_{1,2}U) + (D^{+}_{2,1}U)\odot(D^{-}_{1,2}U) - (D^{-}_{1,1}U)\odot(D^{+}_{2,2}U) \\ &+ (D^{-}_{1,1}U)\odot(D^{-}_{2,2}U) - (D^{-}_{2,1}U)\odot(D^{-}_{1,2}U) + (D^{-}_{2,1}U)\odot(D^{+}_{1,2}U) - (D^{+}_{1,1}U)\odot(D^{-}_{2,2}U)\Big]\end{aligned}$$

According to Proposition 1, the Jacobian matrix of $\mathrm{Det}(U)$ is given by

$$\begin{aligned}M(U) &= \frac{\partial \mathrm{Det}(U)}{\partial U} \\ &= \frac{1}{4h^2}\Big[2hD^{+}_{1,1} - 2hD^{-}_{1,1} + 2hD^{+}_{2,2} - 2hD^{-}_{2,2} \\ &\quad + \mathrm{diag}(\mathrm{D}^{+}_{1,1}\mathrm{U})\mathrm{D}^{+}_{2,2} + \mathrm{diag}(\mathrm{D}^{+}_{2,2}\mathrm{U})\mathrm{D}^{+}_{1,1} - \mathrm{diag}(\mathrm{D}^{+}_{2,1}\mathrm{U})\mathrm{D}^{+}_{1,2} - \mathrm{diag}(\mathrm{D}^{+}_{1,2}\mathrm{U})\mathrm{D}^{+}_{2,1} \\ &\quad + \mathrm{diag}(\mathrm{D}^{+}_{2,1}\mathrm{U})\mathrm{D}^{-}_{1,2} + \mathrm{diag}(\mathrm{D}^{-}_{1,2}\mathrm{U})\mathrm{D}^{+}_{2,1} - \mathrm{diag}(\mathrm{D}^{-}_{1,1}\mathrm{U})\mathrm{D}^{+}_{2,2} - \mathrm{diag}(\mathrm{D}^{+}_{2,2}\mathrm{U})\mathrm{D}^{-}_{1,1} \\ &\quad + \mathrm{diag}(\mathrm{D}^{-}_{1,1}\mathrm{U})\mathrm{D}^{-}_{2,2} + \mathrm{diag}(\mathrm{D}^{-}_{2,2}\mathrm{U})\mathrm{D}^{-}_{1,1} - \mathrm{diag}(\mathrm{D}^{-}_{2,1}\mathrm{U})\mathrm{D}^{-}_{1,2} - \mathrm{diag}(\mathrm{D}^{-}_{1,2}\mathrm{U})\mathrm{D}^{-}_{2,1} \\ &\quad + \mathrm{diag}(\mathrm{D}^{-}_{2,1}\mathrm{U})\mathrm{D}^{+}_{1,2} + \mathrm{diag}(\mathrm{D}^{+}_{1,2}\mathrm{U})\mathrm{D}^{-}_{2,1} - \mathrm{diag}(\mathrm{D}^{+}_{1,1}\mathrm{U})\mathrm{D}^{-}_{2,2} - \mathrm{diag}(\mathrm{D}^{-}_{2,2}\mathrm{U})\mathrm{D}^{+}_{1,1}\Big] \\ &= \frac{1}{2h}(D^{+}_{1,1} - D^{-}_{1,1} + D^{+}_{2,2} - D^{-}_{2,2}) \\ &\quad + \frac{1}{4h^2}\Big[\mathrm{diag}(\mathrm{D}^{+}_{1,1}\mathrm{U})(\mathrm{D}^{+}_{2,2} - \mathrm{D}^{-}_{2,2}) - \mathrm{diag}(\mathrm{D}^{-}_{1,1}\mathrm{U})(\mathrm{D}^{+}_{2,2} - \mathrm{D}^{-}_{2,2}) \\ &\quad - \mathrm{diag}(\mathrm{D}^{+}_{1,2}\mathrm{U})(\mathrm{D}^{+}_{2,1} - \mathrm{D}^{-}_{2,1}) + \mathrm{diag}(\mathrm{D}^{-}_{1,2}\mathrm{U})(\mathrm{D}^{+}_{2,1} - \mathrm{D}^{-}_{2,1}) \\ &\quad - \mathrm{diag}(\mathrm{D}^{+}_{2,1}\mathrm{U})(\mathrm{D}^{+}_{1,2} - \mathrm{D}^{-}_{1,2}) + \mathrm{diag}(\mathrm{D}^{-}_{2,1}\mathrm{U})(\mathrm{D}^{+}_{1,2} - \mathrm{D}^{-}_{1,2}) \\ &\quad + \mathrm{diag}(\mathrm{D}^{+}_{2,2}\mathrm{U})(\mathrm{D}^{+}_{1,1} - \mathrm{D}^{-}_{1,1}) - \mathrm{diag}(\mathrm{D}^{-}_{2,2}\mathrm{U})(\mathrm{D}^{+}_{1,1} - \mathrm{D}^{-}_{1,1})\Big] \\ &= \frac{1}{4h^2}\sum_{s,t=1}^{2}\Big[(-1)^{s+t}\mathrm{diag}(\mathrm{D}^{+}_{\mathrm{s,t}}\mathrm{U})(\mathrm{D}^{+}_{\mathrm{3-s,3-t}} - \mathrm{D}^{-}_{\mathrm{3-s,3-t}})\end{aligned}$$

$$-(-1)^{s+t}\mathrm{diag}(\mathrm{D}^-_{\mathrm{s,t}}\mathrm{U})(\mathrm{D}^+_{3-\mathrm{s},3-\mathrm{t}}-\mathrm{D}^-_{3-\mathrm{s},3-\mathrm{t}})\Big]+\frac{1}{2h}\sum_{s=1}^{2}(D^+_{s,s}-D^-_{s,s}).$$

$$=\frac{1}{4h^2}\sum_{s,t=1}^{2}(-1)^{s+t}\mathrm{diag}(\mathrm{D}^+_{\mathrm{s,t}}\mathrm{U}-\mathrm{D}^-_{\mathrm{s,t}}\mathrm{U})(\mathrm{D}^+_{3-\mathrm{s},3-\mathrm{t}}-\mathrm{D}^-_{3-\mathrm{s},3-\mathrm{t}})+\frac{1}{2\mathrm{h}}\sum_{\mathrm{s}=1}^{2}(\mathrm{D}^+_{\mathrm{s,s}}-\mathrm{D}^-_{\mathrm{s,s}}).$$

In the 3D case, according to Remark 2, the Jacobian matrix of $\det(U)$ can be written as

$$M(U)=\frac{\partial \mathrm{Det}(U)}{\partial U}=\frac{1}{8h^3}\sum_{i',j',k'\in\{-1,+1\}}\frac{\partial \mathfrak{DET}_{i',j',k'}}{\partial U}.$$

The computation of $\frac{\partial \mathfrak{DET}i',j',k'}{\partial U}$ follows the same procedure as in the 2D case and can be derived by combining formula (A.2) with Proposition 1.

References

[1] J. Modersitzki, FAIR: Flexible Algorithms for Image Registration, SIAM Journal on Imaging Sciences, 2009.

[2] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, G. Marchal, Automated multi-modality image registration based on information theory, Inf. Process. Med. Imaging 3 (6) (1995) 263–274.

[3] P. Viola, W.M. Wells, III, Alignment by maximization of mutual information, Int. J. Comput. Vis. 24 (2) (1997) 137–154. https://doi.org/10.1023/A:1007958904918

[4] A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration: a survey, IEEE Trans. Med. Imaging 32 (7) (2013) 1153–1190. https://doi.org/10.1109/TMI.2013.2265603

[5] E. Haber, J. Modersitzki, Intensity gradient based registration and fusion of multi-modal images (2006) 726–733. https://doi.org/10.1007/11866763_89

[6] A. Andronache, M. Vonsiebenthal, G. Szekely, P. Cattin, Non-rigid registration of multi-modal images using both mutual information and cross-correlation, Med. Image Anal. 12 (1) (2008) 3–15. https://doi.org/10.1016/j.media.2007.06.005

[7] C.S. Moore, G.P. Liney, A.W. Beavis, Quality assurance of registration of CT and MRI data sets for treatment planning of radiotherapy for head and neck cancers, J. Appl. Clinical Med. Phys. 5 (1) (2004) 25–35. https://doi.org/10.1120/jacmp.v5i1.1951

[8] K.V. Leemput, F. Maes, D. Vandermeulen, P. Suetens, Automated model-based tissue classification of MR images of the brain, IEEE Trans. Med. Imaging 18 (10) (1999) 897–908. https://doi.org/10.1109/42.811270

[9] J. West, J.M. Fitzpatrick, M.Y. Wang, et al., Comparison and evaluation of retrospective intermodality brain image registration techniques, J. Comput. Assist. Tomogr. 21 (4) (1997) 554–568. https://doi.org/10.1097/00004728-199707000-00007

[10] C.R. Meyer, J.L. Boes, B. Kim, P.H. Bland, K.R. Zasadny, P.V. Kison, K. Koral, K.A. Frey, R.L. Wahl, Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations, Med. Image Anal. 1 (3) (1997) 195–206. https://doi.org/10.1016/S1361-8415(97)85010-4

[11] C.R. Meyer, J.L. Boes, B. Kim, P.H. Bland, Probabilistic brain atlas construction: thin-Plate spline warping via maximization of mutual information, in: Medical Image Computing and Computer-Assisted Intervention MICCAI'99, Springer Berlin Heidelberg, 1999, pp. 631–637. https://doi.org/10.1007/10704282_68

[12] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, D.J. Hawkes, Nonrigid registration using free-form deformations: application to breast MR images, IEEE Trans. Med. Imaging 18 (8) (1999) 712–721. https://doi.org/10.1109/42.796284

[13] G.K. Rohde, A. Aldroubi, B.M. Dawant, Adaptive free-form deformation for interpatient medical image registration, in: M. Sonka, K.M. Hanson (Eds.), SPIE Proceedings, 4322, SPIE, 2001, pp. 1578–1587. https://doi.org/10.1117/12.431043

[14] T. Gaens, F. Maes, D. Vandermeulen, P. Suetens, Non-rigid multimodal image registration using mutual information, in: Medical Image Computing and Computer-Assisted Intervention MICCAI'98, Springer Berlin Heidelberg, 1998, pp. 1099–1106. https://doi.org/10.1007/BFb0056299

[15] G. Hermosillo, C. Chefd Hotel, O. Faugeras, Variational methods for multimodal image matching, Int. J. Comput. Vis. 50 (3) (2002) 329–343. https://doi.org/10.1023/A:1020830525823

[16] G.E. Christensen, M.I. Miller, M.W. Vannier, U. Grenander, Individualizing neuro-anatomical atlases using a massively parallel computer, Computer 29 (1) (1996) 32–38. https://doi.org/10.1109/2.481434

[17] B. Fischer, J. Modersitzki, Fast inversion of matrices arising in image processing, Numer. Algorithms 22 (1) (1999) 1–11. https://doi.org/10.1023/A:1019194421221

[18] S. Periaswamy, H. Farid, Medical image registration with partial data, Med. Image Anal. 10 (3) (2006) 452–464. https://doi.org/10.1016/j.media.2005.03.006

[19] E. D'agostino, F. Maes, D. Vandermeulen, P. Suetens, A viscous fluid model for multimodal non-rigid image registration using mutual information, Med. Image Anal. 7 (4) (2003) 565–575. https://doi.org/10.1016/S1361-8415(03)00039-2

[20] G.E. Christensen, R.D. Rabbitt, M.I. Miller, Deformable templates using large deformation kinematics, IEEE Trans. Image Process. 5 (10) (1996) 1435–1447. https://doi.org/10.1109/83.536892

[21] D.-J. Kroon, C.H. Slump, MRI Modality transformation in demon registration, IEEE Int. Symposium Biomed. Imaging: From Nano to Macro, ISBI '09 (2009) 963–966. https://doi.org/10.1109/ISBI.2009.5193214

[22] H. Lu, M. Reyes, A. Serifovic, S. Weber, Y. Sakurai, H. Yamagata, P.C. Cattin, Multi-modal diffeomorphic demons registration based on point-wise mutual information, 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (2010). https://doi.org/10.1109/ISBI.2010.5490333

[23] K.C. Lam, L.M. Lui, Quasi-Conformal hybrid multi-modality image registration and its application to medical image fusion, in: Advances in Visual Computing, Springer International Publishing, Cham, 2015, pp. 809–818. https://doi.org/10.1007/978-3-319-27857-5_72

[24] X. Cao, J. Yang, Y. Gao, Y. Guo, G. Wu, D. Shen, Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis, Med. Image Anal. 41 (2017) 18–31. https://doi.org/10.1016/j.media.2017.05.004

[25] M. Simonovsky, B. Gutiérrez-Becker, D. Mateus, N. Navab, N. Komodakis, A deep metric for multimodal registration, in: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016, Springer-Verlag, 2016, pp. 10–18. https://doi.org/10.1007/978-3-319-46726-9_2

[26] O. Oktay, A. Schuh, M. Rajchl, K. Keraudren, A. Gomez, M.P. Heinrich, G. Penney, D. Rueckert, Structured decision forests for multi-modal ultrasound image registration, Lect. Notes Comput. Sci. (2015) 363–371. https://doi.org/10.1007/978-3-319-24571-3_44

[27] J. Ceranka, M. Polfliet, F. Lecouvet, N. Michoux, J. de Mey, J. Vandemeulebroucke, Registration strategies for multi-modal whole-body MRI mosaicing, Magn. Reson. Med. 79 (3) (2017) 1684–1695. https://doi.org/10.1002/mrm.26787

[28] Z. Li, F. Huang, J. Zhang, B. Dashtbozorg, S. Abbasi-Sureshjani, Y. Sun, X. Long, Q. Yu, B.t.H. Romeny, T. Tan, Multi-modal and multi-vendor retina image registration, Biomed. Opt. Express 9 (2) (2018) 410–422. https://doi.org/10.1364/BOE.9.000410

[29] A. Mang, G. Biros, An inexact Newton–Krylov algorithm for constrained diffeomorphic image registration, SIAM J. Imaging Sci. 8 (2) (2015) 1030–1069. https://doi.org/10.1137/140984002

[30] A. Mang, G. Biros, A semi-Lagrangian two-level preconditioned Newton–Krylov solver for constrained diffeomorphic image registration, SIAM J. Sci. Comput. 39 (6) (2017) B1064–B1101. https://doi.org/10.1137/16M1070475

[31] A. Mang, L. Ruthotto, A Lagrangian Gauss–Newton–Krylov solver for mass- and intensity-preserving diffeomorphic image registration, SIAM J. Sci. Comput. 39 (5) (2017) B860–B885. https://doi.org/10.1137/17M1114132

[32] M. Hernandez, Gauss-Newton inspired preconditioned optimization in large deformation diffeomorphic metric mapping, Phys. Med. Biol. 59 (20) (2014) 6085–6115. https://doi.org/10.1088/0031-9155/59/20/6085

[33] L. Ruthotto, C. Greif, J. Modersitzki, A stabilized multigrid solver for hyperelastic image registration, Numer. Linear Algebra Appl. 24 (5) (2017). https://doi.org/10.1002/nla.2095

[34] S. Henn, K. Witsch, Multimodal image registration using a variational approach, SIAM J. Sci. Comput. 25 (4) (2004) 1429–1447. https://doi.org/10.1137/S1064827502201424

[35] F. Yang, M. Ding, X. Zhang, W. Hou, C. Zhong, Non-rigid multi-modal medical image registration by combining L-BFGS-B with cat swarm optimization, Inf. Sci. 316 (2015) 440–456. https://doi.org/10.1016/j.ins.2014.10.051

[36] T.M. Cover, J.A. Thomas, Elements of Information Theory, John Wiley & Sons, Inc., 1991. https://doi.org/10.1002/0471200611

[37] J. Modersitzki, Numerical Methods for Image Registration, Oxford University Press, 2003. https://doi.org/10.1093/acprof:oso/9780198528418.001.0001

[38] R. Xu, Y.-W. Chen, S.-Y. Tang, S. Morikawa, Y. Kurumi, Parzen-window based normalized mutual information for medical image registration, IEICE Trans. Inf. Syst. E91-D (1) (2008) 132–144. https://doi.org/10.1093/ietisy/e91-d.1.132

[39] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens, Multimodality image registration by maximization of mutual information, IEEE Trans. Med. Imaging 16 (2) (1997) 187–198. https://doi.org/10.1109/42.563664

[40] S. Henn, K. Witsch, Iterative multigrid regularization techniques for image matching, SIAM J. Sci. Comput. 23 (4) (2001) 1077–1093. https://doi.org/10.1137/S106482750037161X

[41] B. Fischer, J. Modersitzki, Curvature based image registration, J. Math. Imaging Vis. 18 (1) (2003) 81–85. https://doi.org/10.1023/A:1021897212261

[42] B. Fischer, J. Modersitzki, A unified approach to fast image registration and a new curvature based registration technique, Linear Algebra Appl. 380 (2004) 107–124. https://doi.org/10.1016/j.laa.2003.10.021

[43] M.A. Fischler, R.A. Elschlager, The representation and matching of pictorial structures, IEEE Trans. Comput. C-22 (1) (1973) 67–92. https://doi.org/10.1109/T-C.1973.223602

[44] C. Frohn-Schauf, S. Henn, K. Witsch, Multigrid based total variation image registration, Comput. Vis Sci. 11 (2) (2007) 101–113. https://doi.org/10.1007/s00791-007-0060-2

[45] J. Zhang, K. Chen, Variational image registration by a total fractional-order variation model, J. Comput. Phys. 293 (2015) 442–461. https://doi.org/10.1016/j.jcp.2015.02.021

[46] J. Zhang, K. Chen, B. Yu, An improved discontinuity-preserving image registration model and its fast algorithm, Appl. Math. Model. 40 (23–24) (2016) 10740–10759. https://doi.org/10.1016/j.apm.2016.08.009

[47] J. Zhang, Z. Sun, X. Kong, J. Zhang, A vectorial minimized surface regularizer based image registration model and its numerical algorithm, Appl. Math. Model. 106 (2022) 150–176. https://doi.org/10.1016/j.apm.2022.01.015

[48] Y. Li, K. Chen, C. Chen, J. Zhang, A bi-variant variational model for diffeomorphic image registration with relaxed Jacobian determinant constraints, Appl. Math. Model. 130 (2024) 66–93. https://doi.org/10.1016/j.apm.2024.02.033

[49] J. Zhang, Y. Li, Diffeomorphic image registration with an optimal control relaxation and its implementation, SIAM J. Imaging Sci. 14 (4) (2021) 1890–1931. https://doi.org/10.1137/21M1391274

[50] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, 1999.

[51] M. Benzi, G.H. Golub, J. Liesen, Numerical solution of saddle point problems, Acta Numer. 14 (2005) 1–137. https://doi.org/10.1017/S0962492904000212

[52] G. Biros, O. Ghattas, Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. part i: the Krylov–Schur solver, SIAM J. Sci. Comput. 27 (2) (2005) 687–713. https://doi.org/10.1137/S106482750241565X

[53] R.H. Byrd, F.E. Curtis, J. Nocedal, An inexact SQP method for equality constrained optimization, SIAM J. Optim. 19 (1) (2008) 351–369. https://doi.org/10.1137/060674004

[54] E. Haber, U.M. Ascher, Preconditioned all-at-once methods for large, sparse parameter estimation problems, Inverse Probl. 17 (6) (2001) 1847–1864. https://doi.org/10.1088/0266-5611/17/6/319

[55] R. Barrett, M. Berry, T.F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, H. van der Vorst, Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, SIAM, 1994. https://doi.org/10.1137/1.9781611971538

[56] C.C. Paige, M.A. Saunders, Solution of sparse indefinite systems of linear equations, SIAM J. Numer. Anal. 12 (4) (1975) 617–629. https://doi.org/10.1137/0712047

[57] U. Trottenberg, C.W. Oosterlee, A. Schuller, Multigrid, Academic Press, 2001.

[58] D. Arthur, S. Vassilvitskii, k-means + +: the advantages of careful seeding, in: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07, SIAM, USA, 2007, p. 1027-1035.

[59] B. Fischer, J. Modersitzki, Fast diffusion registration, Inverse Problems, Image Anal. Med. Imaging (2002) 117–127. https://doi.org/10.1090/conm/313/05372