

中图法分类号: 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-15

论文引用格式: Sun Jiaying, Yang Shuangyan, Ying Shihui. XXXX. Research on deep unfolding network for brain Image registration. Journal of Image and Graphics, XX(XX):0001-0015(孙嘉颖, 杨双燕, 应时辉. XXXX. 面向脑影像配准的深度展开网络研究. 中国图象图形学报, XX(XX):0001-0015)[DOI:10.11834/jig.250283]

面向脑影像配准的深度展开网络研究

孙嘉颖¹, 杨双燕², 应时辉³

1. 上海大学理学院数学系, 上海 200444; 2. 同济大学附属上海市肺科医院放疗科, 上海 200433; 3. 上海大学上海市应用数学与力学研究所, 上海 200072

摘要: 目的 医学图像配准在手术导航、疾病监测和多模态诊断等临床应用中具有关键作用。然而传统迭代优化方法可解释性强但计算效率低下, 深度学习方法存在黑盒特性导致模型的可解释性不足。对此, 本研究提出了一种基于模型解耦的深度展开配准网络, 旨在兼顾高效计算与模型可解释性。方法 首先通过变量松弛将配准问题解耦为相似性约束和形变正则化两个子问题。在此基础上, 受深度展开网络的启发, 进一步将迭代优化过程映射为多层神经网络。保真项模块用于解决相似性约束子问题, 采用基于TransUnet的架构, 实现融合局部特征与全局上下文。去噪模块用于解决形变正则化子问题, 采用基于残差连接的去噪网络架构近似正则项。两模块协同完成单次交替迭代, 并通过级联构成深度展开网络。所提出的深度展开配准网络与迭代优化步骤严格对应, 每层操作具有明确物理意义和数学依据, 显著提升了模型的可解释性。结果 本文方法在LPBA40和OASIS公开脑部数据集上的戴斯相似系数(Dice Similarity Coefficient, Dice)分别为70.1%和79.9%, 负雅可比矩阵行列式百分比为0.010, 验证了所提方法的有效性。结论 提出的基于模型解耦的深度展开配准网络有效融合了传统迭代优化方法与深度学习方法的优点, 在保持较高计算效率和精度的同时显著提升了可解释性。

关键词: 医学图像配准; 展开网络; 模型解耦; 正则项学习; 脑影像

Research on deep unfolding network for brain Image registration

Sun Jiaying¹, Yang Shuangyan², Ying Shihui³

1. Department of Mathematics, College of Sciences, Shanghai University, Shanghai 200444, China; 2. Department of Radiation Oncology, Shanghai Pulmonary Hospital Affiliated To Tongji University, Shanghai 200433, China; 3. Shanghai Institute of Applied Mathematics and Mechanics, Shanghai University, Shanghai 200072, China

Abstract: Objective Medical image registration plays an important role in clinical applications such as surgical navigation, disease progression monitoring, and multimodal diagnosis. The accuracy of image registration is critical to ensuring the reliability of subsequent image analysis. Traditional methods formulate the registration process as optimization problem by maximizing the similarity between the fixed and warped moving images, and employ iterative optimization strategies to gradually update the deformation field. While such approaches provide clear theoretical interpretability, their high computational costs fail to meet real-time clinical demands. In contrast, deep learning-based methods leverage large-scale datasets to learn the deformation field directly. Although these data-driven approaches have demonstrated remarkable improvements in registration speed and accuracy, they operate as black-box models with limited interpretability. Moreover, most

收稿日期: 2025-06-24; 修回日期: 2025-12-29

基金项目: 国家重点研发计划(2021YFA1003004); 白求恩·中青年医师肿瘤研究基金项目(YDTR-008)

Supported by: National Key Research and Development Project of China (2021YFA1003004); Bethune Young and Middle-aged Physicians Oncology Research Fund(YDTR-008)

current approaches continue to depend on manually designed regularization terms, which may lack adaptability to diverse anatomical structures. To address these challenges, we propose an unfolding-based registration network that combines the theoretical guarantees of traditional optimization with the high efficiency and adaptive learning capabilities of deep neural networks. **Method** Specifically, by introducing variable splitting, we first decouple the registration problem into two subproblems: a similarity-constraint subproblem and a deformation-regularization subproblem. This decomposition reduces optimization complexity and allows each subproblem to be addressed separately. Inspired by deep unfolding network, the alternating update process between the two subproblems is mapped to a multi-stage cascaded architecture. Each stage corresponds to one complete iteration of the optimizer, providing clear physical and mathematical foundations and enhancing interpretability compared to traditional end-to-end models. Within this framework, two specialized modules are designed to address each subproblem. The fidelity module (FM) is employed to solve the similarity-constraint subproblem. Convolutional neural network (CNN) is widely used in deep learning-based registration. However, the limited receptive field of the convolution kernel leads the network to focus on local feature information. While the Transformer uses the self-attention mechanism to effectively capture global dependencies. By integrating the strengths of these two architectures, the FM is designed based on the TransUnet framework. This hybrid design allows the FM to combine local features with global context. The denoising module (DM) is designed to solve the deformation-regularization subproblem. Instead of using hand-crafted regularization terms, we approximate the regularization term using a trainable neural network. Concretely, the DM adopts the DnCNN (denoising convolutional neural network) architecture and incorporates residual connections. This allows the model to adaptively learn priors in a data-driven manner, improving generalization across different anatomical structures. These two modules jointly complete a single alternating iteration and are cascaded to form the deep unfolding network. The overall network is trained with a loss function that balances image similarity, deformation consistency between the outputs of the FM and DM, and smoothness of the deformation field. The proposed network is implemented using the PyTorch framework and trained on an NVIDIA Tesla V100 GPU with 32 GB of memory. The model is optimized using the Adam optimizer with a learning rate of $1e-4$, a batch size of 1, and is trained for 500 epochs. The Transformer component in the FM was configured with 4 attention heads, 8 encoder layers, and an MLP dimension of 96. The unfolding network was constructed with 3 cascaded stages. **Result** Experiments were conducted on two publicly available brain magnetic resonance imaging (MRI) datasets: LPBA40 and OASIS. We compared our method with a range of state-of-the-art models, including traditional registration methods LDDMM (large deformation diffeomorphic metric mapping) and SyN (symmetric normalization), as well as learning-based methods such as VoxelMorph, TransMorph, UTSRMorph and H-SGANet. The quantitative evaluation metrics contained dice similarity coefficient (Dice), the percentage of non-positive Jacobian determinants and inference time. The proposed method achieved an average Dice of 70.1% on LPBA40 and 79.9% on OASIS. The percentage of non-positive Jacobian determinants remained at a low level of 0.01% on both datasets. The inference time was 0.55 s on LPBA40 and 0.37 s on OASIS. Although slightly higher than some deep learning-based models due to the cascaded unfolding architecture and Transformer modules, the time remained clinically feasible. Visualized results also indicated that the deformation fields produced by the proposed method was smoother and more consistent with anatomical priors. Additionally, we carried out a set of ablation studies and reported corresponding quantitative results. These ablation studies explicitly demonstrated the effectiveness of individual modules in our proposed model, including the variable splitting strategy, the TransUnet-based fidelity module and the residual connection-based denoising module, and assessed the influence of key hyperparameters. **Conclusion** In this study, we proposed a deep unfolding network for medical image registration that effectively combines the interpretability of traditional iterative methods with the efficiency and adaptability of deep learning. By unrolling a decoupled optimization model into a trainable network with fidelity and denoising modules, the proposed approach strictly corresponds to a specific step in iterative algorithms. Experiments demonstrate that the proposed method significantly enhances interpretability while maintaining high efficiency and accuracy.

Key words: Medical image registration; unfolding network; model decoupling; regularization learning; brain image

0 引言

医学图像配准作为现代医学图像分析的关键技术,在临床实践中发挥着不可替代的作用。目前,医学图像配准技术广泛应用于临床诊断、疾病治疗、手术导航等场景(Sotiras等,2013;Brock等,2017)。脑部医学图像配准因其解剖结构的特殊性及临床应用的迫切性,已成为该领域的核心研究议题之一。大脑结构高度复杂,不仅具有精细的多尺度结构特征,还常因病理状态引发局部或全局的形态变化。因此,开发能够适应脑部图像特点的高精度配准方法,对于提升脑疾病诊断、手术规划及治疗效果评估具有重要的临床意义。

从技术实现层面而言,图像配准的本质在于建立不同图像之间的空间对应关系,使得不同图像的相应解剖结构能够在同一坐标系下实现精准对齐。在数学上,通常将上述实现过程建模为一个能量泛函优化问题。传统的医学图像配准方法通过显式定义图像相似性度量和形变正则项,采用迭代优化的策略逐步更新形变场,经典代表方法包括:大变形微分同胚度量映射(large deformation diffeomorphic metric mapping, LDDMM)(Beg等,2005)、对称归一化算法(symmetric normalization, SyN)(Avants等,2008)。尽管上述方法拥有明确的可解释性并能生成高质量形变场,但是其计算复杂性导致求解效率低下,无法满足临床场景的实时性要求。

近年来,深度学习的快速发展为医学图像配准带来了突破性进展。基于深度学习的方法能够显著提升配准速度,有效学习复杂解剖变化,并展现出强大的泛化能力,迅速成为该领域的研究热点(施俊等,2020)。代表性深度学习方法VoxelMorph(Balakrishnan等,2019)通过构建端到端的神经网络,实现直接从数据中学习形变映射。在深度学习驱动的配准方法中,卷积神经网络(convolutional neural network, CNN)凭借其强大的局部特征提取能力,成为此类方法的主流架构。然而,由于卷积核有限的感受野导致网络在长程依赖建模方面存在不足。为弥补上述缺陷,研究者提出了许多基于Transformer的模型,包括Transmorph(Chen等,2022)和UTSRMorph(Zhang等,2025),并展现出良好的性能。尽管目前已有诸多配准方法,但是这类方法固有的

黑箱特性,缺乏对过程的显式建模,导致模型的可解释性不足等问题(Chen等,2024)。此外,当前方法中手工设计的正则项也存在显著局限。这些正则项通常基于预设的数学公式,这样单一或固定形式难以捕捉和适应不同解剖部位。尤其是在处理大形变或形变复杂的区域时,影响了配准的精度和泛化能力(马露凡等,2021)。

与此同时,深度展开网络在反问题领域得到广泛应用,相比于一般的神经网络,深度展开网络具有更好的可解释性与性能(Monga等,2021;Xiang等,2021)。算法展开技术将经典优化算法中的迭代步骤进行展开,使其映射为一个网络层级,通过级联这些层级,即可构建深度神经网络架构。网络的前馈过程相当于对原算法进行有限次数的迭代求解。原算法中的超参数均被重构为可学习的网络参数,在端到端训练中实现优化。这种方法不仅保留了神经网络强大的表征能力,又通过显式建模优化迭代过程,提高了算法的可解释性,为医学图像配准提供了新思路。

受深度展开网络优化思想启发,针对当前深度学习方法在脑影像配准中可解释性差、卷积感受野有限的问题,本文提出了一种基于模型解耦的深度展开配准方法。具体而言,通过引入松弛变量将原问题解耦为相似性约束和形变正则化这两个物理意义明确的子问题。并通过展开技术,嵌入正则项学习,分别将两个子问题映射为可训练的神经网络模块,分别称为保真项模块和去噪模块,从而使得网络架构与迭代优化过程严格对应。

本文主要创新点如下:

(1)提出模型解耦策略,将医学图像配准问题解耦成两个复杂度较低的子问题,从而降低模型求解复杂度。同时,为每个子问题设计物理意义明确的可训练神经网络模块,实现更精确高效的配准。

(2)构建深度展开配准网络架构,将传统优化算法的迭代过程显式映射为神经网络的层级结构。每层网络对应一次迭代更新,显式实现解耦后子问题的交替优化,增强了模型的可解释性。

(3)设计结合CNN局部特征提取与Transformer全局上下文建模的保真项模块用于配准任务。该模块突破了卷积核局部感受野的限制,实现跨区域长程依赖建模,增强模型处理复杂形变的能力。与此同时,设计基于残差学习的去噪模块用于逼近正则

项,避免手工设计正则项的局限性,以数据驱动的方式实现自适应学习。

在公开脑部数据集 LPBA40 和 OASIS 上,将所提方法与主流方法进行对比实验,验证了本方法的有效性。

1 相关工作

1.1 传统配准方法

传统医学图像配准方法基于物理模型构建形变场,具备良好的可解释性。早期研究利用弹性力学原理设计数学函数描述形变,如 Bookstein 等人(1989)提出薄板样条插值函数、Rohde 等人(2003)提出径向基函数模型、Xia 等人(2004)提出 B 样条插值函数,通过光滑函数或基函数组合逼近形变场,确保形变连续性和平滑性,但对大尺度非线性形变建模能力有限。为突破上述限制,微分同胚变换理论提供了新的思路。Beg 等人(2005)提出大形变微分同胚度量映射,将配准问题转化为测地线计算问题,通过速度场的积分生成微分同胚变换。Avants 等人(2008)提出对称归一化算法,通过联合优化正向形变和逆向形变,构建对称能量泛函。Vercauteren 等人(2009)在 Demons 算法(Thirion 等人,1998)的基础上引入了微分同胚约束,提出了 Diffeomorphic Demons 算法,通过速度场的积分生成可逆形变场。上述方法具有强大的数学理论支持,并且能生成高质量形变场,但是它们计算成本高、求解速度慢,难以满足实时性需求。

1.2 基于深度学习的配准方法

基于深度学习的配准方法主要可分为有监督方法和无监督方法。

在监督学习框架下,以源图像和目标图像之间的形变场或形变参数作为真值标签。通过构建网络输出与真值标签之间的误差作为损失函数,以此来驱动神经网络的训练。Miao 等人(2016)率先将深度学习应用于刚性配准参数预测,构建了适用于 2D/3D 影像配准的卷积神经网络。由于形变场真值获取困难,研究者提出采用传统配准方法合成形变场作为监督信号,如 Yang 等人(2016)采用 LDDMM 生成形变场,Cao 等人(2017)则使用 SyN 算法获取形变场。Uzunova 等人(2017)进一步提出基于统计外观模型的形变场生成方法。但此类方法存在以下问

题:真实医学图像形变场的获取成本极高;使用传统方法生成的形变场作为监督,会使得训练出的模型精度受限于传统方法的精度;随机生成的形变场与真实形变场存在偏差,缺乏合理性。

空间变换网络(spatial transformer network, STN)(Jaderberg 等,2015)的提出推动研究转向无监督方法。Balakrishnan 等人(2018)提出的 VoxelMorph 框架是无监督配准的代表性工作,通过联合优化图像相似性损失与形变场正则化损失实现端到端训练。其后续工作(Balakrishnan 等,2019)进一步开发了微分同胚版本。de Vos 等人(2019)则构建了多阶段深度学习框架处理单模态配准问题。Zhou 等人(2020)使用级联的思想,提出了嵌套的 U-Net 模型,进一步提高了配准精度。Kim 等人(2021)提出 CycleMorph 模型,通过生成正向和逆向的形变场,并强制形变后的图像经过逆向网络后恢复到原始图像,从而有效保证了形变场的拓扑特征。Wang 等人(2024)提出了递归可变形的纯卷积金字塔网络,利用分步递归策略和高层语义信息,从粗到精地预测形变场。在无监督医学图像配准领域,基于 CNN 的方法长期占据研究主流。然而,此类方法受限于其自身的结构特性。CNN 设计的关键因素之一是感受野的大小,尽管通过堆叠连续卷积层和池化操作能够扩大网络的理论感受野范围,但其有效感受野仍存在显著局限性。这种局部感受野的特性难以捕捉长距离空间对应关系,而这正是医学图像配准中大范围形变建模所必需的。

相比之下,Transformer 模型凭借其自注意力机制,在长距离依赖建模方面具有显著优势(Li 等,2023)。因此,研究者们开始将其引入图像配准领域。Chen 等人(2021)最早将 Transformer 用于配准的任务,提出了 ViT-V-Net。在后续研究中,他们进一步提出了 TransMorph(Chen 等,2022),在编码器结构中使用 Swin Transformer(Liu 等,2021)取代了原有的卷积特征提取器与视觉变换器(vision transformer, ViT)模块,在三维脑部磁共振影像(magnetic resonance imaging, MRI)和腹部 CT(computed tomography)图像配准中取得了更好的结果。Chen 等人(2023)提出了 TransMatch,通过独立的特征提取和基于 Transformer 自注意力机制的多级特征匹配,实现了图像间的显式对应。Zhang 等人(2025)提出了一种名为 UTSRMorph 的无监督医学图像配准网络,

它通过融合Transformer和卷积网络的优势构建高效编码器,并利用超分辨率模块生成高分辨率位移场,在多个3D医学数据集上实现了优于现有方法的性能。杨小龙等人(2025)提出双分支网络,通过CNN局部分支提取细节特征、Transformer全局分支捕捉全局位置与上下文信息,再经注意力融合模块自适应整合特征并结合辅助配准头强化监督,实现单锥体的2D/3D配准。

此外,近年来对抗学习、扩散模型和神经微分方程(neural ordinary differential equations, NODEs)等模型也被引入医学图像配准领域,极大推动了配准方法的发展。大多数应用于图像配准的对抗学习都依赖于生成对抗网络(generative adversarial network, GAN)的基本原理,Fan等人(2018)将配准网络作为生成器产生形变场,判别器作为形变图像和固定图像之间的相似性度量。Mahapatra等人(2018)将GAN应用于多模态配准,并额外实现了CycleGAN(Zhu等,2017)。Kim等人(2022)提出DiffuseMorph,由一个扩散网络和一个变形网络组成。扩散网络学习移动图像与固定图像之间的条件分数函数,获得二者的空间对应关系。形变网络通过条件分数的潜在特征生成形变场。Xu等人(2021)首次将NODEs应用于无监督医学图像配准领域,通过NODEs将图像配准优化问题建模为一个连续的动态过程。

2 基于解耦的深度展开网络

本研究在基于无监督的医学图像配准框架上,提出一种基于模型解耦的深度展开配准网络。本章将具体介绍所提出的方法。

2.1 模型解耦

在医学图像配准领域中,经典的求解方法是通过最小化能量泛函,来实现固定图像与移动图像的空间对齐。具体而言,给定固定图像 I_f 与移动图像 I_m ,对应的能量泛函如下:

$$E(\phi) = L_{sim}(I_m \circ \phi, I_f) + \lambda R(\phi). \quad (1)$$
 其中, $L_{sim}(\cdot, \cdot)$ 是保真项,用于衡量变换后的移动图像与固定图像的相似性; $R(\phi)$ 是正则项,通常采用手工设计的光滑性约束,确保形变场的合理性; λ 是权衡图像相似性与形变正则化的超参数。

但是由于目标函数中保真项与正则项的耦合,计算复杂度高,并且在优化过程极易收敛至局部最优解。对此,本文提出采用模型解耦策略,通过将原问题解耦成两个子问题,从而有效降低模型求解复杂度。具体而言,通过引入辅助变量 v ,将原问题转化为带有等式约束的形式:

$$\arg \min_{u,v} L_{sim}(I_m \circ u, I_f) + \lambda R(v) \text{ s.t. } u = v. \quad (2)$$

对于上式,采用惩罚函数引入二次惩罚项,将其转换为无约束优化问题:

$$\arg \min_{u,v} L_{sim}(I_m \circ u, I_f) + \lambda R(v) + \rho \|u - v\|^2, \quad (3)$$

其中 ρ 为惩罚系数。进一步,将此无约束问题转化为可分离的多变量优化问题,通过交替迭代以下子问题实现联合优化:

$$u^{k+1} = \arg \min_u L_{sim}(I_m \circ u, I_f) + \alpha \|u - v^k\|^2 \quad (4)$$

$$v^{k+1} = \arg \min_v R(v) + \beta \|u^{k+1} - v\|^2 \quad (5)$$

其中 α 和 β 是平衡参数。 u 子问题的目标是优化形变场,使得移动图像经过形变后与固定图像相似,同时保持与辅助变量 v 的一致性。相当于在固定辅助变量 v 时,可将其视为一个配准问题。 v 子问题的目标是约束形变场,同时保持与 u 的一致性。相当于在固定 u 时,可将其视为一个去噪问题。上述操作实现了保真项与正则项的显式解耦,显著提升了计算效率,并为后续算法的设计提供了清晰的物理意义与理论可解释性。在后面小节中将进一步介绍子问题的具体解法。

2.2 正则项学习

在上一节中明确了解耦后的 v 子问题承担着约束形变场光滑性和合理性的任务。传统方法通过手工设计正则项显式约束形变场,这样的做法存在显著缺陷:依赖固定数学形式,难以适应复杂多变的医学图像特征;正则项权重需反复手动调试,泛化能力不足。

为了突破上述瓶颈,我们将解耦后的 v 子问题重构为数据驱动的去噪问题,通过神经网络逼近正则项,实现正则项的自适应学习。因此, v 子问题可以重写为:

$$v^{k+1} = N_\theta(u^{k+1}) \quad (6)$$

其中 N_θ 是参数化的神经网络, θ 是可学习的参数。

这样的设计赋予了算法更强大的表达能力和灵
© 中国图象图形学报版权所有

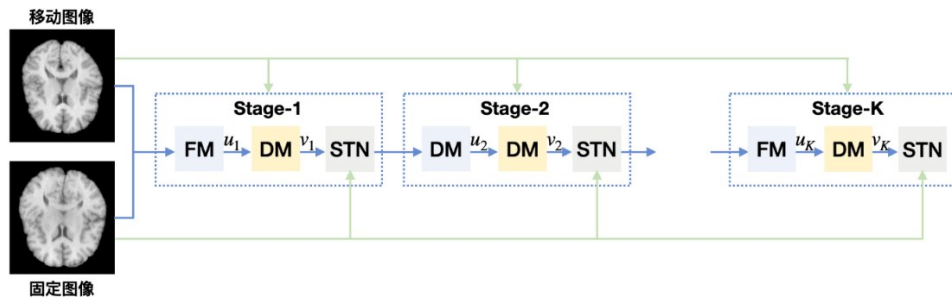


图1 基于模型解耦的深度展开配准网络架构

Fig. 1 Deep unfolding registration network architecture based on model decoupling

活性,尤其在复杂病理形变中能提供更精准的约束,显著提升模型的泛化能力。同时,也为后续深度展

开网络的构建奠定了理论与架构基础。后面的小节中,将具体介绍实现正则项学习的神经网络架构。

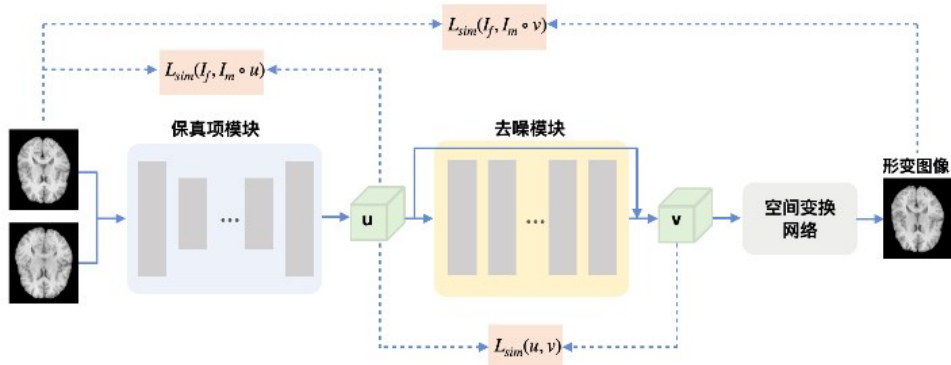


图2 深度展开配准网络单阶段流程

Fig. 2 Single stage of the unfolding registration network

2.3 深度展开配准网络

2.3.1 展开网络架构

在2.1节中,将配准问题解耦为两个子问题。为了减少迭代所需的计算成本,受深度展开网络的启发,考虑将其应用到图像配准问题中。本文提出的深度展开配准网络架构如图1所示,该网络由 K 个阶段级联而成,每个阶段对应一次子问题的交替求解过程。具体而言,将解耦后的 u 子问题和 v 子问题映射到两个可学习的神经网络模块:保真项模块(fidelity module, FM)和去噪模块(denoising module, DM),再联合空间变换网络,共同完成一次迭代。

图2展示了深度展开配准网络单阶段的具体流程。首先,将上一阶段得到的形变后图像与固定图像拼接后输入至保真项模块。保真项模块依据数据相似性驱动,对形变场进行优化,确保形变后的移动图像与固定图像对齐,从而输出更新后的形变场。接着,将形变场输入至去噪模块,对形变场进行正则化处理,约束形变场的光滑性和合理性,输出优化后

的形变场。最终空间变换网络将优化后的形变场应用于移动图像,生成形变后的图像。

与传统的端到端深度学习网络不同,这样的设计使得网络中的每一层都具有明确的物理意义和数学依据,确保网络在拥有神经网络强大的学习能力的同时,具备了传统方法的可解释性。同时可以根据具体需求,独立设计各个模块,充分发挥各模块在特点任务上的优势,增强了模型兼容性,提升了计算效率。在下面的章节中将详细阐述各模块的架构。

2.3.2 保真项模块

针对 u 子问题,其本质为配准问题,因此我们采用U-Net类似的结构,并在此基础上加入Transformer模块。本文提出了一种基于TransUnet的配准网络,核心设计思想在于结合CNN的局部特征提取能力和Transformer的全局上下文建模优势,网络架构如图3所示。

在编码阶段,通过多层下采样模块逐层提取多尺度特征。每个下采样模块包含以下核心组件:使

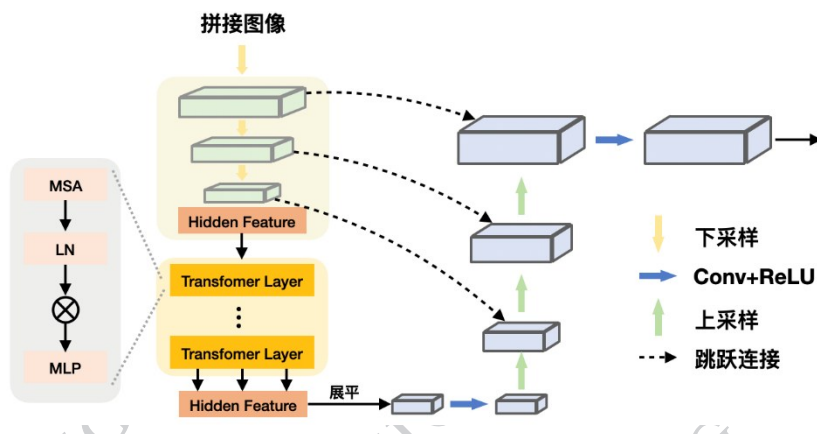


图3 保真项模块网络架构

Fig. 3 Network architecture of the fidelity module

用 $3 \times 3 \times 3$ 卷积核的卷积层、批归一化 (batch normalization, BN) 和激活函数 (rectified linear unit, ReLU)。在编码器末端, 将生成的深层特征 (hidden feature) $F \in R^{C \times D \times H \times W}$ 展平为序列 $S \in R^{(D \times H \times W) \times C}$, 输入至 Transformer 模块, 建立模跨区域长程依赖。; Transformer 层可划分为多头自注意力 (multi-head self-attention, MSA)、层归一化 (layer normalization, LN) 和多层感知机 (multilayer perceptron, MLP) 部分。在解码阶段, 由一系列上采样模块逐步恢复空间分辨率。并且通过跳跃连接, 将每个上采样层的特征与编码器中对应尺度的特征进行融合。最终经过卷积操作和 ReLU 激活函数输出预测形变场, 该形变场维度与图像空间维度相匹配。

2.3.3 去噪模块

如 2.2 节所述, 传统方法中正则项依赖手工设计, 而本文通过神经网络逼近正则项。在实现细节上, 去噪模块采用图像去噪领域经典的 DnCNN (denoising convolutional neural network) 网络架构 (Zhang 等, 2016), 并融入了深度学习中常用的 ResNet (residual network) 残差学习机制 (He 等, 2016)。该模块的网络架构如图 4 所示。具体而言, 网络级联了多个去噪模块, 每个模块都遵循 DnCNN 的设计范式, 包含使用 $3 \times 3 \times 3$ 卷积核的卷积层、BN 和 ReLU。最后, 采用残差连接机制, 将输入直接与多层变换后的输出相加。这样的设计使网络专注于学习形变场中非物理形变或非光滑形变, 避免了直接对完整形变场进行重构, 从而降低了网络训练的难度, 减少了可能出现的梯度消失或过拟合问题。

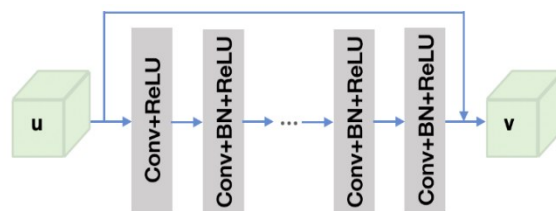


图4 去噪模块网络架构

Fig. 4 Network architecture of the denoising module

2.3.4 损失函数

在整个训练过程中, 本文使用均方误差来评估固定图像与变换后的移动图之间的相似度, 形变场梯度的 L_2 范数用于确保形变场的合理性以及一致性损失约束解耦变量的一致性。因此, 总损失为

$$L_{total} = \lambda_1 L_{sim}(I_m \circ u, I_f) + (1 - \lambda_1) L_{sim}(I_m \circ v, I_f) + \lambda_2 \|\nabla u\|^2 + \lambda_3 \|u - v\|^2 \quad (7)$$

3 实验

3.1 数据集

为证明所提出的基于模型解耦和正则项学习的深度展开配准方法的有效性, 我们在公开的 LPBA40 (Shattuck 等, 2008) 与 OASIS (Daniel 等, 2007) 脑部数据集展开实验。

LPBA40 包含 40 例脑部 MRI 扫描数据, 手动分割了 54 个解剖区域。对于 LPBA40 数据集, 所有图像先使用核磁共振数据处理软件 FreeSurfer 工具包进行颅骨去除操作, 随后重采样至体素尺寸为 $1\text{mm} \times 1\text{mm} \times 1\text{mm}$, 再将图像进行仿射变换, 最后所有图像裁剪为 $160 \times 192 \times 160$ 大小。在实验中,

从数据集中随机选取1例作为固定图像,其余作为移动图像,最终构建30对训练样本和9对测试样本。

OASIS数据集包含414例脑部MRI扫描数据,手动分割了35个解剖区域。对于OASIS数据集进行预处理,主要流程涵盖基于FreeSurfer的强度校正、颅骨去除以及空间标准化等步骤,随后被重新采样为 $1mm \times 1mm \times 1mm$,并将大小裁剪为 $160 \times 192 \times 224$ 。在实验中将,将数据集随机划分为254例训练样本、20例验证样本及140例测试样本。

3.2 实验设置

3.2.1 实验环境

本文实验基于PyTorch框架,GPU为NVIDIA Tesla V100,显存大小为32GB。在实验中,选择Adam优化器进行网络训练,批处理大小为1,学习率为 1×10^{-4} ,迭代次数为500。Transformer模块参数设置如下:注意力头数为4、编码器层数为8、MLP维度为96。展开网络层数为3。

3.2.2 比较算法

为了证明所提出算法的有效性,实验选取以下代表性算法作为对比,包括传统方法和基于深度学习的方法。首先,考虑LDDMM(Beg等,2005)和对称归一化算法(Avants等,2008),该方法通过迭代更新来寻找最优形变。同时使用了基于CNN的配准方法:VoxelMorph(Balakrishnan等,2019),该方法是深度学习配准的代表性工作。此外,比较了近年来将Transformer融入配准网络的模型,分别是TransMorph(Chen等,2022)、UTSRMorph(Zhang等,2025)、H-SGANet(Zhou等,2025)。LDDMM和SyN方法使用Python的Ants工具包实现。除了H-SGANet,基于深度学习的方法都采用作者公开的官方代码及默认参数设置进行训练。由于H-SGANet的源代码尚未公开,本研究直接引用了原文献中报告的结果,与本文方法在训练集、验证集和测试集的比例设置上保持一致,但具体样本的分配不同。

3.2.3 评价指标

1 戴斯相似系数(dice similarity coefficient, Dice)

Dice系数用于计算两幅图像相同解剖结构的重合程度,是医学图像配准中常用的评价指标。Dice系数的取值范围为0-1,值越接近1表明配准精度

越高。其计算公式如下:

$$Dice(X,Y) = 2 \cdot \frac{|X \cap Y|}{|X| + |Y|} \quad (8)$$

其中, X 和 Y 分别表示配准后的移动图像和固定图像的标签图像。

2 雅可比行列式

为了定量地评估形变场 ϕ 的质量,可以计算形变场的雅可比行列式。理想的形变场应保持雅可比行列式值均为正数,若出现负值则表明该区域存在拓扑结构破坏。通过计算负雅可比矩阵行列式百分比,可以有效检测形变场是否存在的折叠现象。雅可比行列式的计算公式如下:

$$J_{\phi} = \begin{pmatrix} \frac{\partial \phi_x}{\partial x} & \frac{\partial \phi_x}{\partial y} & \frac{\partial \phi_x}{\partial z} \\ \frac{\partial \phi_y}{\partial x} & \frac{\partial \phi_y}{\partial y} & \frac{\partial \phi_y}{\partial z} \\ \frac{\partial \phi_z}{\partial x} & \frac{\partial \phi_z}{\partial y} & \frac{\partial \phi_z}{\partial z} \end{pmatrix} \quad (9)$$

3.3 结果与讨论

表1展示了在LPBA40和OASIS数据集上,所提出的方法与其他几种对比方法上的结果。其中Dice分别是LPBA40数据集54个区域Dice值的平均结果和OASIS数据集35个区域Dice值的平均结果。在实验数据集上,本文方法在配准精度上处于较高水平,负雅可比矩阵行列式百分比也处于较低水平,说明所生成的形变场更符合解剖学规律。在时间效率方面,本文方法配准耗时略高于某些深度学习对比方法,这主要由于深度展开架构中Transformer模块和级联设计所带来的计算开销,但仍在可接受范围内,远快于传统迭代优化算法。实现模型可解释性、精度与效率之间的良好平衡。

图5展示了所用方法的可视化结果。第一列为原始图像,包括固定图像和移动图像。前三行为LPBA40数据集上的结果,后三行为OASIS数据集上的结果。对于每个数据集,第一行展示了所使用方法得到的形变后图像,第二行展示了残差图,第三行展示了形变场网格。从形变后图像及其残差图来看,相较于VoxelMorph,本文方法在内部结构的对齐上更贴合固定图像,残差分布更为均匀且数值更小,说明引入Transformer模块能更精准地捕捉图像间的

表1 不同方法在LPBA40和OASIS上的实验结果
Table 1 Experimental results of different methods on LPBA40 and OASIS

方法	LPBA40			OASIS		
	Dice	$\% J_\phi \leq 0$	时间	Dice	$\% J_\phi \leq 0$	时间
LDDMM	67.3	0.000	89.24	75.2	0.000	132.99
SyN	69.5	0.000	18.74	76.0	0.000	30.98
VoxelMorph	68.3	0.030	0.45	77.6	0.002	0.17
TransMorph	69.9	0.004	0.35	79.8	0.005	0.36
UTSRMorph	69.5	0.010	0.35	78.1	0.005	0.27
H-SGANet	69.1	0.080	0.28	81.4	0.020	0.26
Our	70.1	0.010	0.55	79.9	0.010	0.37

注:加粗字体表示各列最优结果。

融合上的优势。从形变场的可视化来看,本文方法形变场颜色分布更均匀,网格变化相对平滑,未出现异常强烈的颜色突变,很少出现局部的过度扭曲或折叠,表明形变更符合解剖学先验。尽管形变场网格变化未呈现过度复杂的模式,但结合配准图像的高精度,说明我们的方法在保持整体结构对齐的同时,能精准捕捉必要的局部形变。与其他方法,如UTSRMorph、TransMorph相比,其形变场更简洁有效,反映了深度展开网络通过迭代优化逐步细化形变的特性,避免了冗余的形变预测。综上,本文方法在配准精度与形变场合理性上取得了更好的平衡。配准精度体现了保真项模块对特征的有效利用,形变场合理性体现了去噪模块对解剖先验的融入,从而验证了本方法的架构设在医学图像配准中的有效性。

空间变换关系,体现了保真项模块全局和局部特征

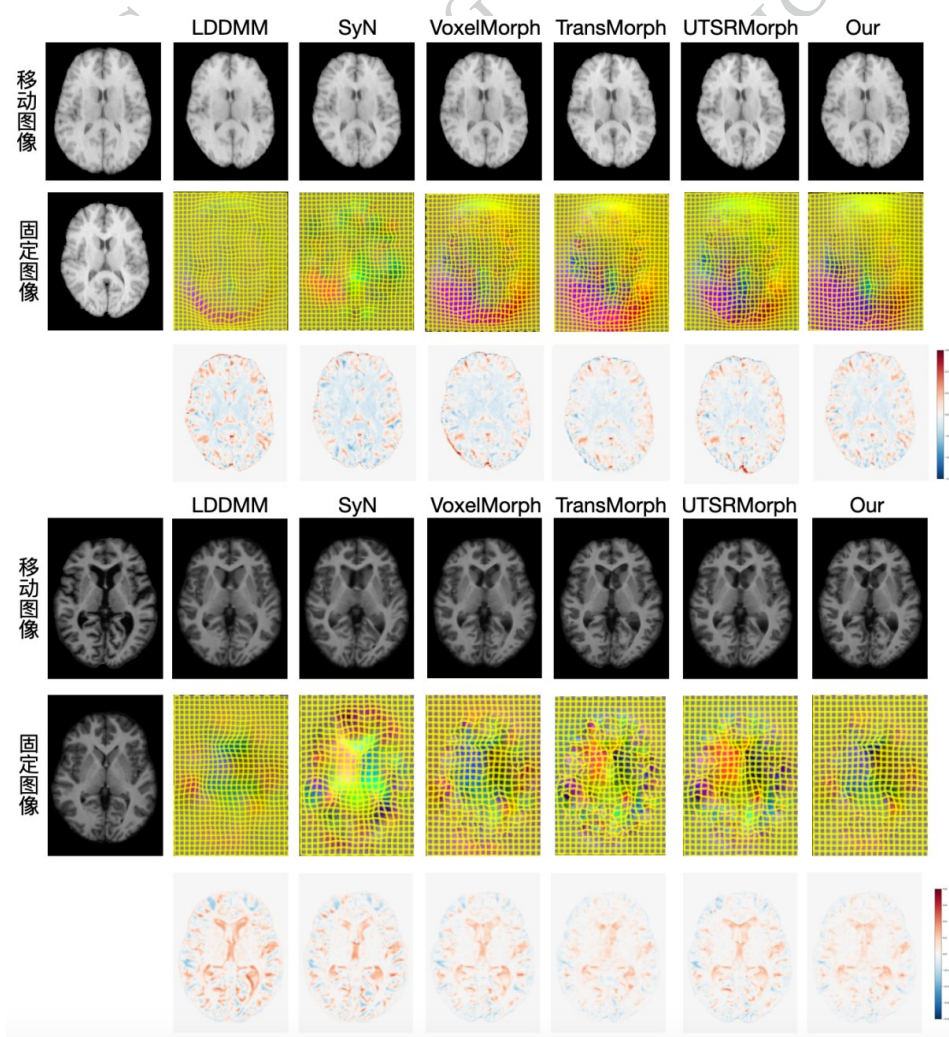


图5 不同方法在LPBA40(上)和OASIS(下)数据集的可视化结果

Fig. 5 Visualization results of different methods on LPBA40(top) and OASIS(bottom) datasets

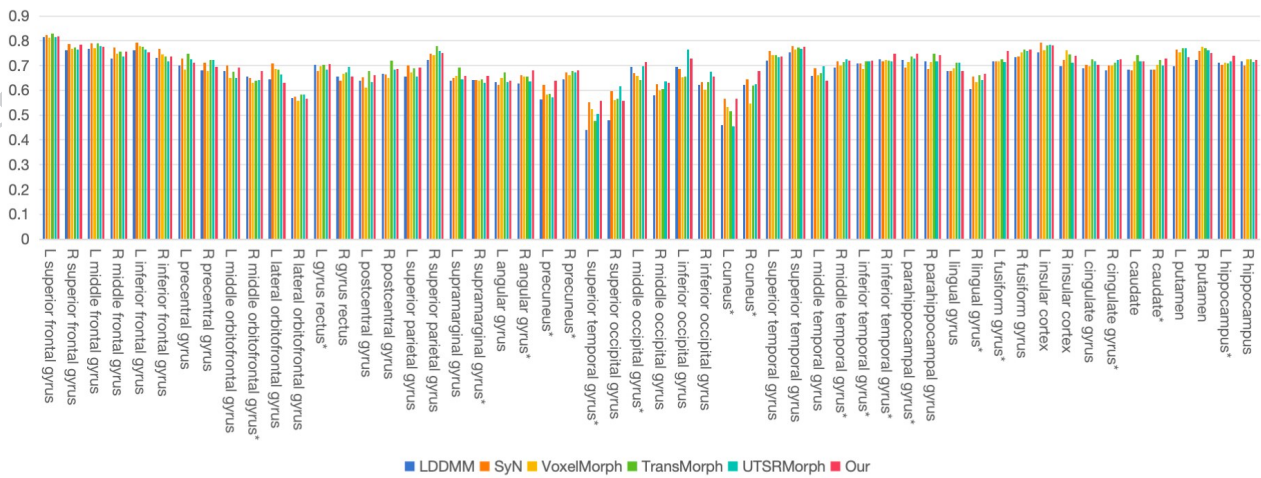


图6 不同方法在LPBA40每个解剖结构的Dice值

Fig. 6 Dice scores of different methods for each anatomical structure in LPBA40

图6和图7分别展示了两个数据集上,本文方法与其他深度学习的方法在每个解剖结构上的Dice值的直方图。其中标星号的脑区表明本文方法在该脑区的Dice值高于其他方法。经统计在LPBA数据集上共有19个脑区标星号,在OASIS数据集上共有13个脑区标星号。本文方法在皮层脑区与深部结构等关键脑区中取得了更优的配准表现。尤其是在结构形态差异大、体积较小且通常难以准确配准的

区域,如海马、丘脑、扣带回、楔前叶等表现出优势,体现了其在复杂脑区配准任务中的准确性。

图8展示了 $K = 3$ 时形变场逐步优化的过程以及对应的Dice和负雅可比行列式。其中,Dice值随迭代次数单调上升。此外,负雅可比行列式先下降后上升,但是始终控制在0.01以内。结果表明,本文架构能够在配准精度与形变场合理性之间实现平衡。

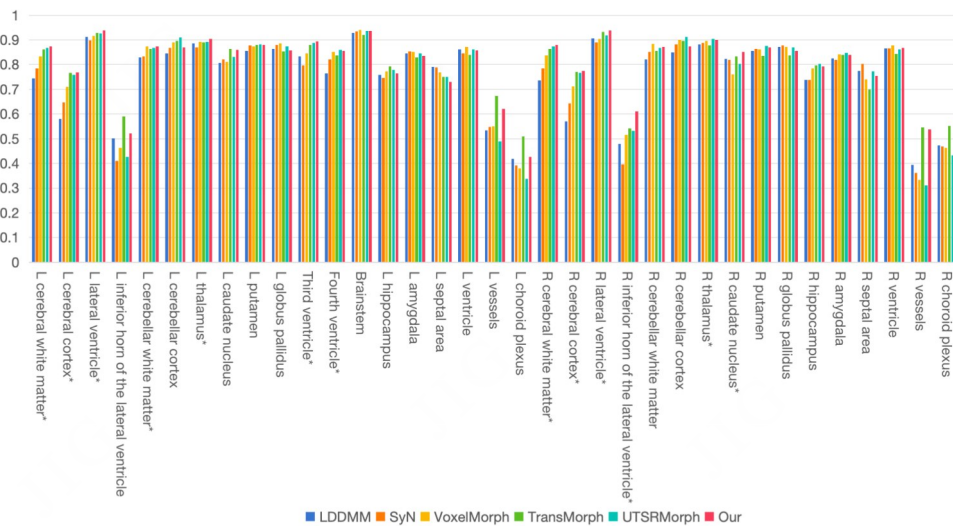


图7 不同方法在OASIS每个解剖结构的Dice值

Fig. 7 Dice scores of different methods for each anatomical structure in OASIS

3.4 消融实验

3.4.1 模块组件的影响

为验证各模块组件的必要性及其交互作用,避免功能冗余,本文进行了一系列消融实验。首先验证保真项模块中Transformer结构与跳跃连接的贡

献,进而探究该保真项模块与去噪模块的协同效应。

实验初始在固定去噪模块为DnCNN的条件下,设计了以下三组对照实验:TransUnet(基准模型,含CNN特征提取、Transformer结构以及跳跃连接);仅CNN模型(移除Transformer结构,保留原CNN结构

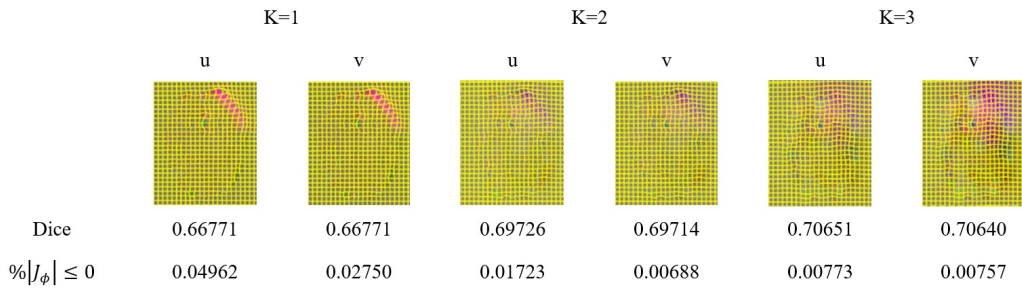


图8 K=3时各Stage形变场可视化

Fig. 8 Visualization of deformation fields for each stage at K=3

及跳跃连接); TransUnet 去掉跳跃连接(保留完整的 TransUnet, 但移除各层特征之间的跳跃连接)。表2展示了各模型在配准性能与训练时间的对比结果, 可以看出: 虽然引入 Transformer 增加了额外的计算开销, 导致时间略有增加, 但其能通过全局注意力机制有效捕捉图像长程依赖关系, 弥补 CNN 局部感受野对大尺度形变建模的不足, 从而提升配准精度。同时, 跳跃连接以微量增加时间开销的代价, 实现 CNN 多尺度特征融合, 进一步优化配准精度。上述实验充分验证了二者在保真项模块中的必要性。

进一步, 为深入分析模块间的交互作用, 固定保真项模块为 TransUnet 的条件下, 将去噪模块中的 DnCNN 替换为简单的多层卷积。从表2结果可以看出: 使用 DnCNN 作为去噪模块, 其性能优于卷积去噪模块 Conv, 这表明使用 DnCNN 作为去噪模块能够对保真项模块的不足进行有效补偿与协同增强。

表2 模块组件的消融实验结果

Table 2 Ablation experimental results on Module components

保真项模块			去噪模块		LPBA40		
CNN	Transformer	跳跃连接	DnCNN	Conv	Dice	$\% J_\phi \leq 0$	时间
√	√	√	√		70.1	0.010	0.55
√		√		√	68.1	0.019	0.36
√	√			√	67.7	0.005	0.54
√	√	√		√	68.9	0.016	0.50

注: 加粗字体表示各列最优结果。

3.4.2 DnCNN 去噪的影响

为验证将 DnCNN 作为形变场去噪模块的合理性, 本文设计了以 U-Net 去噪为对照的消融实验。由表3结果可知, 在使用 DnCNN 时, 不仅在两个数

据集上的配准精度优于 U-Net, 还能更高效约束形变场的合理性, 同时其计算耗时更短。上述结果验证了本研究选择 DnCNN 作为去噪模块的合理性与优势。

表3 不同去噪模块的实验结果

Table 3 Experimental results of different denoising modules

去噪模块	LPBA40			OASIS		
	Dice	$\% J_\phi \leq 0$	时间	Dice	$\% J_\phi \leq 0$	时间
Unet	68.5	0.017	0.60	78.3	0.015	0.44
DnCNN	70.1	0.010	0.55	79.9	0.010	0.37

注: 加粗字体表示各列最优结果。

表4 不同超参数下的实验结果

Table 4 Experimental results of different hyperparameters

λ_1	λ_3	LPBA40	
		Dice	$\% J_\phi \leq 0$
$\lambda_1 = 0$	$\lambda_3 = 0.1$	67.3	0.003
$\lambda_1 = 0.2$	$\lambda_3 = 0.1$	67.8	0.006
$\lambda_1 = 0.5$	$\lambda_3 = 0.1$	68.7	0.008
$\lambda_1 = 0.8$	$\lambda_3 = 0.1$	69.5	0.010
$\lambda_1 = 1$	$\lambda_3 = 0.1$	69.1	0.011
$\lambda_1 = 0.8$	$\lambda_3 = 1$	69.3	0.007
$\lambda_1 = 0.8$	$\lambda_3 = 10$	68.7	0.007

注: 加粗字体表示各列最优结果。

3.4.3 超参数分析

根据经验设置 $\lambda_2 = 1$ 和 $\lambda_3 = 0.1$, 通过调整 λ_1 以找到最佳平衡权重。然后, 再固定 $\lambda_1 = 0.8$, 通过

调整 λ_3 以找到最佳一致性权重。表4展示了不同超参数组合下的消融实验结果,分别比较了Dice和负雅可比行列式百分比。

3.4.4 Transformer 参数影响分析

表5 Transformer模块的不同关键参数的实验结果

Table 5 Experimental results of different key parameters of the transformer module

Transformer关键参数	LPBA40			
	Dice	$\% J_\phi \leq 0$	时间	
注意力头数 (num_layers=8, MLP_dim=96)	2	68.5	0.030	0.55
	4	70.1	0.010	0.55
	8	70.2	0.020	0.57
编码器层数 (num_heads=4, MLP_dim=96)	4	69.7	0.002	0.47
	8	70.1	0.010	0.55
MLP维度 (num_heads=4, num_layers=8)	64	69.1	0.040	0.58
	96	70.1	0.010	0.55
	256	70.3	0.003	0.50

注:加粗字体表示各列最优结果。

Transformer模块的关键参数包括多头自注意力头数(num_heads)、编码器层数(num_layers)、MLP维度(MLP_dim)等。这些参数共同决定模型的特征建模能力、计算复杂度与训练效率。本文将通过控制变量法,仅调整Transformer的相关参数,依次考察上述参数在固定深度展开阶段数($K=3$)时的作用效果。根据表5结果明确了各个关键参数对医学图像配准性能的影响,最终确定最优参数配置为:注意力头数4、编码器层数8、MLP维度96。该配置在本文所用数据集上取得了配准精度、形变场合理性与推理效率之间的良好平衡。在实际应用时,可依据目标数据集的特性与可用硬件资源,对参数进行适当调整,以进一步提升模型性能,从而更好地适应不同临床场景下的医学图像配准需求。

3.4.5 展开网络层数分析

表6展示了不同展开网络层数下的实验结果。当 $K=2$ 时,配准精度较低。当 $K=3$ 时,配准性能显著提升。而 $K=5$ 时,配准效果却未如预期进一步提升,可能是由于过多的展开层数导致模型在训练过程中出现过拟合。同时随着展开层数增加,负雅可比矩阵行列式百分比也随之下降,这表明增加

展开层数有助于抑制不合理形变。综上,网络展开层数应结合实际任务需求进行选择,并非越多越好,合适的层数才能实现配准精度、形变合理性、计算效率的最大化平衡。

表6 不同展开网络层数下的实验结果

Table 5 Experimental results of different numbers of unfolding network layers

展开网络层数	LPBA40		
	Dice	$\% J_\phi \leq 0$	时间
2	69.2	0.040	0.54
3	70.1	0.010	0.55
5	68.3	0.003	0.56

注:加粗字体表示各列最优结果。

4 结论

针对当前基于深度学习的医学图像配准方法可解释不足,本文提出了一种基于模型解耦和正则项学习的深度展开配准方法,该方法融合传统优化理论的可解释性与深度学习的数据驱动优势。通过将原始配准问题解耦为图像对齐子问题与正则项约束子问题,降低了模型求解的复杂度。并通过深度展开技术,将两个子问题分别映射为保真项模块和去噪模块这两个可学习的神经网络模块。两模块协同完成单次交替迭代,并通过级联构成深度展开网络。这样的展开策略,使得每一层网络严格对应迭代算法的优化步骤,增强了网络的可解释性。在方法的具体实现上,保真项模块结合CNN局部特征提取与Transformer长程依赖建模,提升处理复杂结构能力;去噪模块借鉴残差学习的思想构建神经网络逼近正则项,实现在数据驱动下自适应学习正则项,突破了手工设计正则项的局限性。最后,在公开脑部数据集上的实验验证了所提方法的有效性。

尽管本文方法展现出较好的性能,但是仍存在关键局限性需进一步突破:其一,当前方法仅针对单模态图像;其二,级联构成的深度展开网络增加了参数数量与计算量。基于上述局限,未来研究可围绕以下几个方面展开:1. 自适应确定展开深度层数,以实现计算资源根据配准难度的动态分配;2. 研究适配医学图像配准的轻量化Transformer,优化网络

结构。

参考文献(References)

- Avants B B, Epstein C L, Grossman M and Gee J C. 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12 (1) : 26-41 [DOI: 10.1016/j.media.2007.06.004]
- Balakrishnan G, Zhao A, Sabuncu M R, Dalca A V, Guttag J and Dalca A V. 2018. An unsupervised learning model for deformable medical image registration//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 9252-9260 [DOI:10.1109/cvpr.2018.00964]
- Balakrishnan G, Zhao A, Sabuncu M R, Guttag J and Dalca A V. 2019. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8) : 1788-1801 [DOI:10.1109/TMI.2019.2897538]
- Beg M F, Miller M I, Troune A and Younes L. 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2) : 139-157 [DOI: 10.1023/B:VISI.0000043755.93987.aa]
- Bookstein F L. 2002. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 11(6) : 567-585 [DOI:10.1109/34.24792]
- Brock K K, Mutic S, Menutt T R, Li H and Kessler M L. 2017. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM radiation therapy committee task group no. 132. *Medical Physics*, 44(7) [DOI:10.1002/mp.12256]
- Cao X H, Yang J H, Zhang J, Nie D, Kim M, Wang Q and Shen D G. 2017. Deformable image registration based on similarity-steered CNN regression//Proceedings of the 20th Medical Image Computing and Computer-Assisted Intervention. Quebec City, Canada: Springer: 300-308 [DOI:10.1007/978-3-319-66182-7_35]
- Chen J Y, Frey E C, He Y F, Segars W P, Li Y and Du Y. 2022. Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis*, 82: 102615 [DOI:10.1016/j.media.2022.102615]
- Chen J Y, He Y, Frey E C, Li Y and Du Y. 2021. ViT-V-Net: vision transformer for unsupervised volumetric medical image registration [EB/OL]. [2025-12-10].
<https://arxiv.org/pdf/2104.06468.pdf>
- Chen J Y, Liu Y H, He Y F and Du Y. 2023. Spatially-varying regularization with conditional transformer for unsupervised image registration [EB/OL]. [2025-4-18].
<https://arxiv.org/pdf/2303.06168.pdf>
- Chen J Y, Liu Y H, Wei S W, Bian Z X, Subramanian S, Carass A, Prince J L and Du Y. 2024. A survey on deep learning in medical image registration: new technologies, uncertainty, evaluation metrics, and beyond. *Medical Image Analysis*, 100: 103385 [DOI:10.1016/j.media.2024.103385]
- de Vos B D, Berendsen F F, Viergever M A, Sokooti H, Staring M and Išgum I. 2019. A deep learning framework for unsupervised affine and deformable image registration. *Medical Image Analysis*, 52: 128-143 [DOI:10.1016/j.media.2018.11.010]
- Fan J F, Cao X H, Xue Z, Yap P T and Shen D G. 2018. Adversarial similarity network for evaluating image alignment in deep learning based registration//Proceedings of the 21st International Conference on Medical Image Computing and Computer-Assisted Intervention. Granada, Spain: Springer: 739-747 [DOI: 10.1007/978-3-030-00928-1_83]
- He K M, Zhang X Y, REN S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI:10.1109/CVPR.2016.90]
- Kim B, Han I and Ye J C. 2022. DiffuseMorph: unsupervised deformable image registration using diffusion model// Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 347-364 [DOI: 10.1007/978-3-031-19821-2_20]
- Kim B, Kim D H, Park S H, Kim J, Lee J G and Ye J C. 2021. CycleMorph: cycle consistent unsupervised deformable image registration. *Medical image analysis*, 71: 102036 [DOI: 10.1016/j.media.2021.102036]
- Li J, Chen J Y, Tang Y C, Wang C, Landman B A and Zhou S K. 2023. Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives. *Med Image Anal*, 85: 102762 [DOI:10.1016/j.media.2023.102762]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin Z and Guo B N. 2021. Swin Transformer: hierarchical vision transformer using shifted windows//Proceedings of 2021IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE: 9992-10002. [DOI: 10.1109/ICCV48922.2021.00986]
- Ma L F, Luo F, Yan J P, Xu Z, Luo J and Li X. 2021. Deep-learning based medical image registration pathway: towards unsupervised learning, 026(009): 2037-2057 (马露凡, 罗凤, 严江鹏, 徐哲, 罗捷, 李秀. 2021. 深度医学图像配准研究进展: 迈向无监督学习. *中国图象图形学报*, 026(009): 2037-2057) [DOI:10.11834/jig.200361]
- Mahapatra D, Antony B, Sedai S and Garnavi R. 2018b. Deformable medical image registration using generative adversarial networks// The 15th IEEE International Symposium on Biomedical Imaging. Washington, USA: IEEE: 1449-1453 [DOI: 10.1109/isbi.2018.8363845]
- Marcus D S, Wang T H, Parker J, Csernansky J G, Morris J C and Buckner R L. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented,

- and demented older adults. *Journal of Cognitive Neuroscience*, 2007, 19(9): 1498-1507 [DOI:10.1162/jocn.2007.19.9.1498]
- Miao S, Wang Z J and Liao R. 2016a. A CNN regression approach for real-time 2D/3D registration. *IEEE Transactions on Medical Imaging*, 35(5): 1352-1363 [DOI:10.1109/TMI.2016.2521800]
- Monga V, Li Y L and Eldar Y C. 2021. Algorithm Unrolling: interpretable, efficient deep learning for signal and Image processing. *IEEE Signal Processing Magazine*, 38(2): 18-44 [DOI:10.1109/MSP.2020.3016905]
- Rohde G K, Aldroubi A and Dawant B M. 2003. The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Transactions on Medical Imaging*, 22(11): 1470-1479 [DOI:10.1109/TMI.2003.819299]
- Shattuck D W, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr K L, Poldrack R A, Bilder R M and Toga A W. 2008. Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage*, 2008, 39(3): 1064-1080 [DOI:10.1016/j.neuroimage.2007.09.031]
- Shi J, Wang L L, Wang S S, Chen Y X, Wang Q, Wei D M, Liang S J, Peng J L, Yi J J, Liu S F, Ni D, Wang M L, Zhang D Q and Shen D G. 2020. Applications of deep learning in medical imaging: a survey. *Journal of Image and Graphics*, 25(10): 1953-1981 (施俊, 汪琳琳, 王珊珊, 陈艳霞, 王乾, 魏冬铭, 梁淑君, 彭佳林, 易佳锦, 刘盛锋, 倪东, 王明亮, 张道强, 沈定刚. 2020. 深度学习在医学影像中的应用综述. *中国图象图形学报*, 25(10): 1953-1981) [DOI:10.11834/jig.200255]
- Sotiras A, Davatzikos C and Paragios N. 2013. Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging*, 32(7): 1153-1190 [DOI:10.1109/TMI.2013.2265603]
- Thirion J P. 1998. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3): 243-260 [DOI:10.1016/S1361-8415(98)80022-4]
- Uzunova H, Wilms M, Handels H and Ehrhardt J. 2017. Training CNNs for image registration from few samples with model-based data augmentation//*Proceedings of the 20th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Quebec City, Canada: Springer: 223-231 [DOI:10.1007/978-3-319-66182-7_26]
- Vercauteren T, Pennec X, Perchant A and Ayache N. 2009. Diffeomorphic demons: efficient non-parametric image registration. *Academic Press*, 45(1): 61-72 [DOI:10.1016/J.NEUROIMAGE.2008.10.040]
- Wang H Q, Ni D and Wang Y. 2024. Recursive deformable pyramid network for unsupervised medical image registration. *IEEE Transactions on Medical Imaging*, 43(6): 2229-2240 [DOI:10.1109/TMI.2024.3362968]
- Xia M H and Liu B D. 2004. Image registration by "super-curves". *IEEE Transactions on Image Processing a Publication of the IEEE Signal Processing Society*, 13(5): 720-732 [DOI:10.1109/TIP.2003.822611]
- Xiang J X, Dong Y G and Yang Y J. 2021. FISTA-Net: learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Transactions on Medical Imaging*, 40(5): 1329 - 1339 [DOI:10.1109/TMI.2021.3054167]
- Xu J S, Chen E Z, Chen X, Chen T and Sun S H. 2021. Multi-scale neural ODEs for 3D medical image registration//*Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Strasbourg, France: Springer: 213-223 [DOI:10.1007/978-3-030-87202-1_21]
- Yang X, Kwitt R and Niethammer M. 2016. Fast predictive image registration//*Proceedings of the 1st International Workshop on Deep Learning in Medical Image Analysis*. Athens, Greece: Springer: 48-57 [DOI:10.1007/978-3-319-46976-8_6]
- Yang X L, Zhang Z C, Xu S K, Zhang B C, Luo X Q and Hu F Y. 2025. Single vertebra 2D/3D registration with fusion of local and global features. *Journal of Image and Graphics*, 30(8): 2851-2865 (杨小龙, 张战成, 徐少康, 张宝成, 罗晓清, 胡伏原. 2025. 融合局部与全局特征的单椎体2D/3D配准网络. *中国图象图形学报*, 30(8): 2851-2865) [DOI:10.11834/jig.240502]
- Zhang K, Zuo W M, Chen Y J, Meng D Y and Zhang L. 2016. Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 2016, 26(7): 3142-3155 [DOI:10.1109/TIP.2017.2662206]
- Zhang R S, Mo H, Wang J C, Jie B M, He Y, Jin N G and Zhu L. 2025. Utsrmp: a unified transformer and superresolution network for unsupervised medical image registration. *Medical Imaging, IEEE Transactions on*, 44(2): 891-902 [DOI:10.1109/TMI.2024.3467919]
- Zhou Y F and Cao W M. H-SGANet: hybrid sparse graph attention network for deformable medical image registration [J]. *Neurocomputing*, 2025, 633(000) [DOI:10.1016/j.neucom.2025.129810]
- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2020. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6): 1856-1867 [DOI:10.1109/TMI.2019.2959609]
- Zhu J Y, Park T, Isola P and Efros A A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 2242-2251 [DOI:10.1109/ICCV.2017.244]

作者简介

孙嘉颖, 2000年生, 女, 硕士研究生, 主要研究方向为医学图像配准。E-mail: sjying@shu.edu.cn

杨双燕, 共同通讯作者, 女, 主要研究方向为临床放疗计划设计与呼吸运动管理技术实施。E-mail: yang_shuang_yan@163.com

应时辉, 共同通讯作者, 男, 教授, 主要研究方向为医学影像 处理与智能分析。E-mail: shying@shu.edu.cn