计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
[doi: 10.15888/j.cnki.csa.009760] [CSTR: 32024.14.csa.009760]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 Transformer 的多尺度可变形三维医学图像配准①

陈璐莹[1,2], 喻国荣[1,2], 鲍海洲[1,2], 边小勇[1,2], 陈聪鹏[1,2]

[1](武汉科技大学 计算机科学与技术学院, 武汉 430065)
[2](智能信息处理与实时工业系统湖北省重点实验室, 武汉 430065)
通信作者: 喻国荣, E-mail: yuguorong190@wust.edu.cn

摘　要: 由于人体器官的不规则形变, 可变形三维医学图像配准仍然是医学图像处理中的难题. 针对该问题, 本文提出了一种基于 Transformer 的多尺度可变形三维医学图像配准方法. 该方法首先采用多尺度策略来实现多层次的连接, 以捕捉不同层次的信息. 通过自注意力机制提取全局特征, 并利用膨胀卷积捕获更广泛的上下文信息和更细节的局部特征, 从而增强配准网络对全局和局部特征的融合能力. 其次, 本文根据图像梯度的稀疏性先验, 引入了归一化总梯度作为损失函数, 有效减少了噪声和伪影对配准过程的干扰, 更好地适应不同模态的医学图像. 在公开的脑 MRI 数据集 (OASIS 和 LPBA) 上评估本文所提方法的性能. 综合结果表明, 该方法不仅能保持基于学习的方法在运行时间上的优势, 还在均方误差和结构相似性等指标上表现出较高的性能. 此外, 消融实验的结果进一步证明了本文所提方法和归一化总梯度损失函数设计的有效性.

关键词: 医学图像; 图像配准; Transformer; 多尺度; 可变形

引用格式: 陈璐莹,喻国荣,鲍海洲,边小勇,陈聪鹏.基于 Transformer 的多尺度可变形三维医学图像配准.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/9760.html

Multi-scale Deformable 3D Medical Image Registration Based on Transformer

CHEN Lu-Ying[1,2], YU Guo-Rong[1,2], BAO Hai-Zhou[1,2], BIAN Xiao-Yong[1,2], CHEN Cong-Peng[1,2]

[1](School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China)
[2](Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan 430065, China)

Abstract: Deformable 3D medical image registration remains challenging due to irregular deformations of human organs. This study proposes a multi-scale deformable 3D medical image registration method based on Transformer. Firstly, the method adopts a multi-scale strategy to realize multi-level connections to capture different levels of information. Self-attention mechanism is employed to extract global features, and dilated convolution is used to capture broader context information and more detailed local features, so as to enhance the registration network's fusion capacity for global and local features. Secondly, according to the sparse prior of the image gradient, the normalized total gradient is introduced as a loss function, effectively reducing the interference of noise and artifacts on the registration process, and better adapting to different modes of medical images. The performance of the proposed method is evaluated on publicly available brain MRI datasets (OASIS and LPBA). The results show that the proposed method can not only maintain the advantages of the learning-based method in run-time but also well performs in mean square error and structural similarity. In addition, ablation experiment results further prove the validity of the method and normalized total gradient loss function design proposed in this study.

Key words: medical image; image registration; Transformer; multi-scale; deformable

① 基金项目: 湖北省教育厅青年人才项目 (Q20221108)
收稿时间: 2024-06-13; 修改时间: 2024-07-10, 2024-08-20; 采用时间: 2024-08-27; csa 在线出版时间: 2024-11-15

1 介绍

医学图像配准是一种将不同时间、来源或模态的医学图像进行对齐和匹配的技术. 其目标是使两幅图像的特征点在空间位置和解剖结构上完全一致, 以确保所有具有诊断意义以及位于手术区域内的解剖点都能够匹配[1]. 然而, 医学图像数据容易受到不同受体、不同姿势、不同位置以及不同成像设备等多种因素的影响. 这些因素导致医学图像数据在形态和强度上存在较大的差异, 增加了配准的挑战性. 有效的配准技术可以消除或减少医学图像间的这些差异, 为后续医学图像处理和分析提供准确、可靠的数据基础. 医学图像配准是疾病诊断、手术导航以及器官运动分析的基础, 具有重要的理论研究价值和实际应用价值.

传统的医学配准方法可以根据几何变换的类型分为两类: (1) 基于刚体变换的配准方法, 例如基于特征的配准方法[2]和基于灰度的配准方法[3]; (2) 基于非刚体变换的配准方法, 例如薄板样条 (thin-plate spline, TPS)[4]和基于 B 样条的自由形变模型 (free-form deformation, FFD)[5]. 传统的配准方法在 2D-2D 单模态图像配准方面效果不错, 但对于 3D 医学图像配准而言, 其效率和性能有待进一步提升. 此外, 这些方法需要大量的迭代优化, 这不仅耗费大量的时间和资源, 还极大地限制了医学图像配准的实时性和有效性.

随着深度学习技术的不断发展, 它在医学图像处理领域也得到了广泛应用. 许多学者在基于深度学习的医学图像配准方面进行了大量研究. 根据训练过程是否依赖标签, 基于深度学习的医学图像配准方法可以分为两类: 基于监督学习和基于无监督学习. 基于监督学习的配准方法需要使用标签来训练模型. 然而, 在医学图像中获取解剖结构或特定器官的标签往往非常困难. 为了解决标签缺失的问题, 研究人员提出了使用传统配准方法生成标签或进行人工标注. Fan 等[6]提出了一种基于双监督全卷积网络的脑图像配准方法. 该方法引入双重监督机制和全卷积网络架构, 利用传统配准方法生成的配准结果作为真实标签. Sokooti 等[7]结合多尺度特征提取和 3D 卷积神经网络, 在胸部 CT 图像上实现了高效且准确的配准. 他们的研究证明了在训练模型时利用人工标注的变形场是有效的. 但在实际应用中, 传统算法生成的标签可能不准确, 人工标注也耗费大量的时间和精力. 因此, 获取大量且有效的标签面临很大的挑战.

为了缓解医学图像数据集中标签信息稀缺和不准确的问题, 无监督学习在医学图像配准中被广泛应用. 这种方法主要利用图像间的空间信息相关性来训练模型, 比如灰度、结构、纹理等特征. 最有效的方法之一是 VoxelMorph[8], 它是一种类似 U-Net 架构的无监督配准方法. 但该方法通过堆叠卷积块构建单尺度可变形配准网络, 难以解决医学图像中的大位移和复杂变形等问题. XMorpher[9]使用双并行的 U 形网络分别提取运动图像和固定图像的特征, 然后通过交叉注意力机制交换特征信息, 并寻找多层次的语义对应关系, 最终实现精细配准. CycleMorph[10]是一种基于循环一致性的配准方法. 该方法引入双向的循环一致性约束, 显著提升了配准精度. SYMNet[11]是一种高效的对称图像配准方法, 该方法最大限度地提高了差分映射空间内图像间的相似性, 并同时估计了正变换和逆变换.

现有的医学图像配准研究主要基于卷积神经网络 (convolutional neural network, CNN), 但这类模型往往在远程依赖建模能力上受到限制[12]. 由于 Transformer[13]在自然语言处理中的出色表现, 它逐渐被应用于医学影像分析等任务, 并取得了优于 CNN 的结果. Liu 等[14]提出了一种具有层次化窗口的模型 Swin Transformer, 该模型利用类似 CNN 的方法将注意力局限在窗口范围内, 降低了模型的复杂度和参数量. Chen 等[15]首次将 Swin Transformer 应用到医学图像配准领域, 提出了一种无监督配准架构 TransMorph. 该框架超越了当前 CNN 在配准上的表现, 但在特征融合方面的处理还不够完善. Transformer 模型更关注学习全局特征, 而医学图像中许多微小的解剖结构和细微的变化需要更细致的局部特征来捕捉. 因此, 在医学图像配准中, 需要综合考虑局部和全局特征, 从而提高配准的准确性和稳定性.

在三维医学图像配准领域, 大多数基于深度学习的方法使用相似性度量来评估图像之间的相似性, 如归一化互信息 (normalized mutual information, NMI)、归一化相关系数 (normalized correlation coefficient, NCC)[16]和互信息 (mutual information, MI)[17]. 然而, 这些方法通常基于像素强度来比较相似性, 在处理医学图像时并非总是有效. 医学图像通常通过 X 射线、CT 扫描和 MRI 等设备采集, 而这些设备可能在成像过程中引入噪声. 噪声会扰动像素值, 影响相似性度量的准确性. 因此, 基于强度的度量对于医学图像中的局部噪声变化非常敏感, 可能无法准确地反映出医学图像的

解剖结构和生理特征.

为了解决基于深度学习的医学图像配准任务中的内在挑战, 本文设计了一种基于 Transformer 的多尺度可变形三维医学图像配准方法. 本文的技术贡献有两个方面: (1) 提出了一种基于膨胀卷积的 Transformer 模型 (dilation based vision Transformer, Dil-ViT), 该模型将 Transformer 和膨胀卷积结合, 能够捕捉更广泛的全局特征和更细致的局部特征. (2) 引入了归一化总梯度 (normalized total gradient, *NTG*) 作为新的相似性度量, 利用差分图像的梯度稀疏性来度量图像间的相似性, 能够降低噪声和异常值对配准结果的影响.

2 相关工作

2.1 膨胀卷积

膨胀卷积[18]是一种通过膨胀率来扩大感受野范围的卷积方式, 在许多计算机视觉任务中表现出色. 与普通卷积不同, 膨胀卷积在卷积核中插入了一定数量的零值 (即膨胀), 进而增大了卷积核的有效窗口, 扩大了感受野的范围. 膨胀卷积的卷积核有效窗口大小和感受野范围的计算如下:

$$K' = (D-1)\times(K-1)+K \tag{1}$$

$$RF = (K'-1)\times L+1 \tag{2}$$

其中, K 表示普通卷积核的尺寸, D 表示膨胀率, K'表示膨胀后卷积核的有效尺寸, L 表示卷积层数, RF 表示感受野. 当膨胀率 D 为 1 时, 式 (1) 和式 (2) 表示普通卷积的卷积核有效窗口大小和感受野范围.

膨胀卷积在增大感受野的同时, 还具有多重优势. 首先, 膨胀卷积没有增加卷积核的实际窗口大小和网络的参数数量, 这使得网络能够在较少的层数下捕获更大范围的上下文信息. 因此, 在需要处理全局信息的图像问题或依赖较长序列信息的语音和文本问题中, 膨胀卷积能够发挥重要作用. 其次, 膨胀卷积能够保持特征图的分辨率, 从而在提取局部特征方面更具优势. 此外, 通过使用不同膨胀率的膨胀卷积, 网络可以在不同尺度上提取特征, 保留更多的细节信息. 这种多尺度特征提取能力使得膨胀卷积能够更好地处理具有层次结构的特征, 并在不同尺度上进行信息融合.

2.2 Vision Transformer

Vision Transformer (ViT)[19]是首个将 Transformer 应用于图像并展示出最佳配准性能的模型. ViT 的工作原理是将图像分割成多个图像块, 然后利用 Transformer 模型对这些块进行建模, 从而捕捉到全局的图像语义信息. 具体而言, ViT 将图像分割为固定大小的图像块, 每个块代表图像中的一个局部区域. 然后, 这些图像块将被展平为一维向量, 并输入到 ViT 模型中进行处理. ViT 模型包含多个 Transformer 编码器, 每个 Transformer 编码器主要由两个模块组成: 多头自注意力机制 (multi-head self-attention, MSA) 和前馈神经网络 (feed forward network, FFN). 多头自注意力机制能够对不同图像块之间的关系进行建模, 而前馈神经网络则对输入图像块进行非线性变换和特征提取[20]. 通过 Transformer 编码器, ViT 逐渐提取图像中的高级语义特征. 最后, 这些特征向量将被传递给一个全连接层, 用于分类或回归任务.

2.3 多头自注意力机制

Transformer 中最重要的模块是多头自注意力机制. 给定一个由多个图像块构成的输入向量 X, 通过线性映射得到查询 (Q)、键 (K)、值 (V) 3 个子向量, 即 $Q = XW^Q$, $K = XW^K$, $V = XW^V$, 其中, W^Q、W^K、W^V 是对应的权重矩阵. 然后将子向量拆分成 h 组, 每组向量维度为d_h, 计算每组向量的注意力权重和加权求和:

$$Attention(Q_i, K_i, V_i) = Softmax\left(\frac{Q_i K_i^{\mathrm{T}}}{\sqrt{d_h}}\right)V_i \tag{3}$$

其中, i 表示第 i 组向量, $Softmax(\cdot)$函数用于计算注意力权重.

最后将 h 个注意力的输出向量拼接在一起, 并乘以一个权重矩阵, 得到多头自注意力机制的最终输出.

2.4 前馈神经网络

前馈神经网络通常包含两个线性层和一个非线性激活函数. 多头自注意力机制输出的特征首先通过第一个线性层映射到高维空间, 再经过激活函数进行非线性转换, 最后通过第 2 个线性层恢复到原始维度. 整个过程实现了特征的升维和降维, 有助于特征的提取与融合. 这个过程可以表示为:

$$MLP(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{4}$$

其中, x是输入向量, W_1和b_1是第 1 个线性变换的权重和偏置, W_2和b_2是第 2 个线性变换的权重和偏置.

3 研究方法

针对人体器官的不规则形变, 本文提出了一种基于 Transformer 的多尺度可变形三维医学图像配准方

法. 设 F 和 M 为三维空间里的固定图像和运动图像, 该方法的目标是学习 F 和 M 的最佳仿射矩阵 A, 通过 A 将 M 变形为与 F 相似的图像. 首先, 给定一对固定图像和运动图像作为输入, 将它们送入配准网络得到初始仿射变换矩阵. 然后, 使用空间变换网络 (spatial Transformer network, STN)[21]对运动图像进行变换, 得到变形后的运动图像. 最后, 通过不断迭代调整配准网络的参数, 使得损失函数最小, 从而得到最优仿射变换矩阵, 实现准确的图像配准.

3.1 总体结构

图 1 为配准方法示意图, (a) 为整体结构, (b) 为 Transformer 编码器结构, (c) 为回归器结构. 配准方法的整体结构分为 3 个阶层，如图 1(a) 所示. 该方法采用多尺度策略, 每个阶层都将待配准的图像对作为输入, 并输出相应的仿射变换矩阵. 这种多尺度策略有助于逐步细化配准过程, 从而提高结果的精确度. 此外, 所有阶层都使用基于膨胀卷积的 Transformer 网络模型 (dilation based vision Transformer, Dil-ViT). 该模型由卷积图像块嵌入层、膨胀 Transformer 编码层以及回归器组成. 其中, 膨胀 Transformer 编码层包含 4 个编码器, 每个编码器主要由多头自注意力机制和前馈神经网络组成, 编码器结构如图 1(b) 所示.

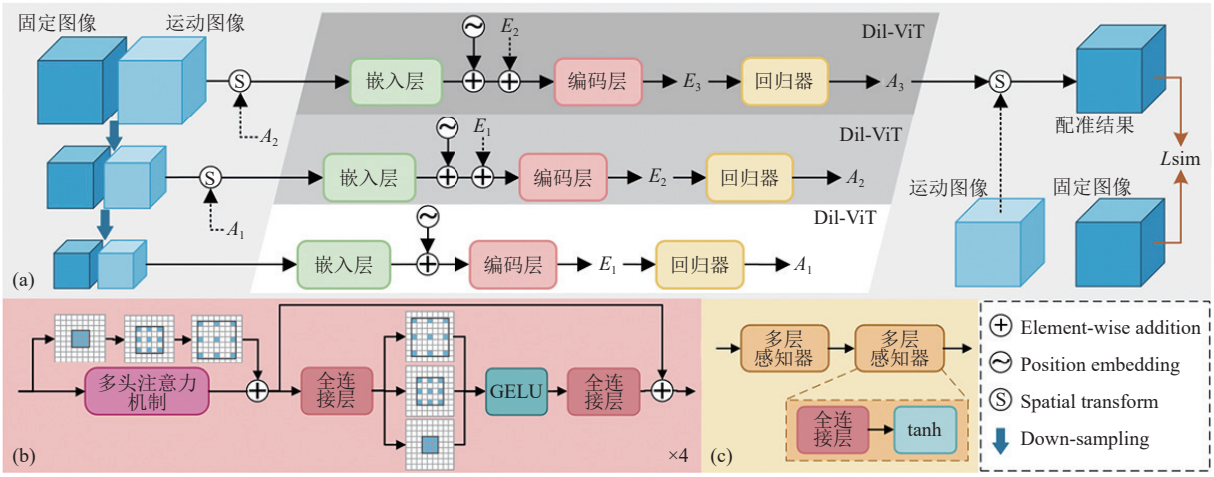


图 1 配准方法示意图

具体而言, 首先给定一个固定图像F和一个运动图像M, 利用三线性插值对图像进行下采样, 创建图像金字塔. 本文设置 3 层金字塔, 采样后的图像表示为F_i和M_i. 其中, i表示当前阶层数, 即$i\in\{1,2,3\}$, 下采样比例因子为0.5^{3-i}. 将F_i和M_i按照通道维度进行连接, 作为第i阶层的输入图像. 第 3 阶层的图像尺寸最大, 为原始图像尺寸, 即$F_3=F$、$M_3=M$. 第 1 阶层的图像尺寸最小, 为原始图像尺寸乘以对应的比例因子0.5^2. 然后将金字塔的输入图像送入 Dil-ViT 的卷积图像块嵌入层, 得到块嵌入矩阵. 紧接着, 将该块嵌入矩阵输入膨胀 Transformer 编码层, 得到特征嵌入E_i. 接下来, 利用 Dil-ViT 的回归器将E_i映射为仿射变换矩阵A_i. 该回归器由 2 个采用 tanh 激活函数的多层感知器 (multi layer perceptron, *MLP*) 构成, 回归器结构如图 1(c) 所示. 此外, 第i $(i<3)$阶层的 Dil-ViT 输出特征嵌入E_i和仿射变换矩阵A_i后将额外进行 2 个操作: (1) E_i通过残差连接添加到下一阶层的图像块中; (2) A_i将对下一阶层的M_{i+1}进行空间变换, 并与F_{i+1}一起作为下一阶层的输入. 最后, 第 3 阶层输出的仿射变换矩阵A_3即为最终的仿射变换矩阵.

3.2 基于膨胀卷积的 Transformer 模型

ViT 模型由于自注意力机制在建模一系列非重叠图像块中的长程依赖关系方面表现出色, 但它缺乏局部机制来建模输入块与相邻块之间的关系[22]. 考虑到 CNN 的工作原理适合对局部关系建模, 本文使用膨胀卷积对 ViT 进行调整优化, 并提出了 Dil-ViT 网络模型. 在每个阶层中, Dil-ViT 相比传统 ViT 模型主要进行了 3 项优化.

第一, 本文使用卷积图像块嵌入代替线性图像块嵌入, 连续的卷积层能够更好地编码像素级空间信息. 各阶层的图像输入尺寸及其对应的卷积参数如表 1 所示. 给定一个图像对的输入$I\in\mathbb{R}^{H\times W\times D\times C}$, 其中 H、

W 和 D 表示图像 I 的空间维度, 均设为 128, C 表示通道数, 设为 3. 卷积图像块嵌入层利用卷积计算 I 的图像块嵌入映射 $X \in \mathbb{R}^{H\times W\times D\times C}$, 将 X 展平后得到尺寸为 $N\times d$ 的特征图, 其中 $N=H\times W\times D$ 表示图像块数, d 表示嵌入维度. 本文将 N 和 d 分别设为 4096 和 256.

表 1 各阶层的图像尺寸及和卷积操作参数

阶层	输入图像尺寸	卷积核	步长	填充
第1层	32×32×32	3×3×3	2	1
第2层	64×64×64	5×5×5	4	3
第3层	128×128×128	7×7×7	8	7

第二, 本文提出了一种结合膨胀卷积的自注意力模块, 用于替代 ViT 模型中的自注意力模块. 虽然 ViT 模型在建模非重叠图像块间的长程依赖关系方面表现出色, 但缺乏建模输入块与其邻近块之间关系的局部机制. 因此, 本文在多头自注意力模块中添加了一个与其平行的膨胀卷积模块, 即一组串联的膨胀卷积 (dilated convolution in series, *DCS*), 这有助于 ViT 在图像块中同时建模局部性和远程依赖关系.

第三, 本文提出了一种结合膨胀卷积的前馈模块, 用于将多尺度上下文和局部信息嵌入到图像块中. 在传统的 ViT 模型中, 前馈神经网络由一个具有两个隐藏层的多层感知器组成. 为了解决前馈神经网络在块嵌入映射中缺乏局部关系建模的问题, 本文在两个隐藏层之间增加了一组并联的膨胀卷积 (dilated convolution in parallel, *DCP*), 进一步将局部性引入到 ViT 中.

DCS 和 *DCP* 分别由 3 个膨胀卷积以串联和并联的形式组成, 膨胀卷积的各项参数如表 2 所示. 膨胀卷积在获取图像特征方面比普通卷积更具优势, 特别是在局部特征提取方面. 以卷积核大小 3×3 的二维卷积为例, 首先, 膨胀卷积比普通卷积具有更大的感受野, 使每个卷积操作能够覆盖更大的图像区域. 如图 2 所示, 普通卷积和膨胀卷积都使用 3×3 的卷积核, 前者的感受野范围是 3×3, 而后者在不改变卷积核大小的情况下, 感受野范围可扩大至 5×5. 利用这一特性, 膨胀卷积可以捕捉更大范围内的特征, 同时不增加额外计算量. 其次, 膨胀卷积能够提取不同尺度的特征. 膨胀卷积的膨胀率决定了输入特征图上的采样点之间的距离. 较小的膨胀率意味着采样点之间的距离较小, 可以更好地捕捉局部细节特征; 而较大的膨胀率则意味着采样点之间的距离较大, 可以更好地捕捉到更广泛的上下文信息. 因此, 膨胀卷积通过调整膨胀率来改变感受野的大小, 从而提取不同尺度的特征. 这对于处理较高分辨率和大尺寸的医学图像尤为重要. 再次, 膨胀卷积能够同时捕捉细粒度的局部特征和粗粒度的全局特征. 相比于普通卷积利用下采样 (如池化操作) 来扩大感受野, 膨胀卷积通过插入零值来实现, 避免了降低特征图分辨率的问题, 从而保留更多细节信息. 图 3 展示了普通卷积和膨胀卷积在提取特征上的不同. 与普通卷积提取的特征 (b) 相比, 膨胀卷积提取的特征 (a) 范围更广泛; 与降低图像分辨率后使用普通卷积提取的特征 (c) 相比, 两者具有相似的感受野范围, 但膨胀卷积保留了更多详细的局部特征.

表 2 膨胀卷积模块的参数

卷积核	膨胀率	步长	填充
3×3×3	1	2	1
3×3×3	3	4	3
3×3×3	5	8	5

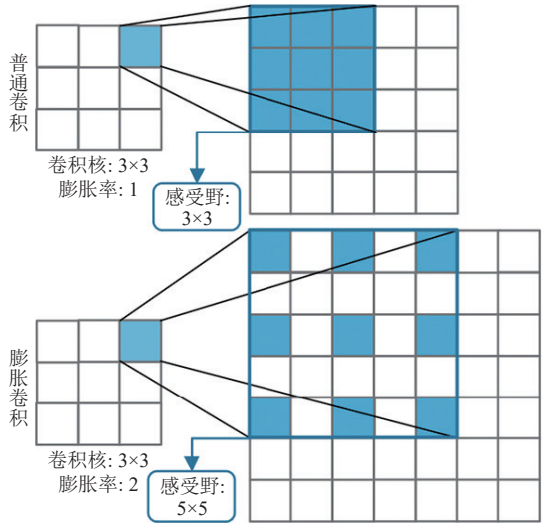


图 2 普通卷积和膨胀卷积的感受野范围对比

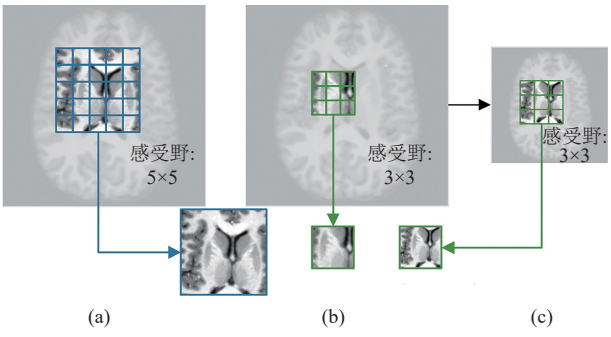


图 3 普通卷积和膨胀卷积提取特征的效果对比

Dil-ViT 的优势在于其基于 *DCS* 的多头自注意力机制和基于 *DCP* 的前馈神经网络. 这种方式可以将图

像的全局和局部特征融合, 将膨胀卷积固有的尺度不变性和局部性引入到 ViT 模型中, 从提高模型对图像的理解能力和多尺度处理能力. 对于第 l 层的输入特征 X^l, Dil-ViT 的建模过程如式 (5) 所示:

$$\begin{cases} \hat{X}^l = MSA(LN(X^l)) + DCS(X^l) \\ X^{l+1} = DCP\text{-}MLP(LN(\hat{X}^l)) + \hat{X}^l \end{cases} \tag{5}$$

其中, MSA 表示多头自注意力机制, $DCP\text{-}MLP$ 表示结合了膨胀卷积模块的多层感知器, LN 表示归一化层.

3.3 损失函数

图像配准是将运动图像与固定图像逐渐对齐的过程, 旨在不断提高两者之间的相似性. 在配准过程中, 损失函数用于衡量图像间的相似性. 优化算法通过迭代更新参数, 使损失函数逐渐减小, 从而找到最佳变换参数. 现有的图像配准方法中常用的相似性度量方法及其相应的损失函数有 NMI、NCC 和 MI 等. 尽管这些方法在一定程度上表现出良好的性能, 但在医学图像中存在一些问题. 由于不同扫描仪或不同扫描参数的影响, 图像之间的灰度分布可能存在差异, 图像更容易受到噪声和伪影等干扰因素的影响. 上述相似性度量方法是基于像素级的比较, 当图像存在噪声时, 噪声会扰动像素值, 从而影响相似性度量的结果.

为了解决这个问题, 本文引入归一化总梯度 (normalized total gradient, NTG)[23]作为配准框架的损失函数. NTG 的核心思想是基于一个重要的先验知识: 对齐后的图像之间的差分图像的梯度比未对齐图像的差分图像的梯度更加稀疏. 具体而言, NTG 利用梯度的稀疏性来衡量图像的对齐程度. 当图像对齐时, 梯度分布更稀疏, 差分图像中梯度的绝对值之和更小. 作为一种基于梯度的相似性度量, NTG 对噪声不敏感. 因为噪声通常会引入局部的随机变化, 这些变化在梯度计算中可能会被平均化或抵消. 因此, NTG 能够在存在噪声和伪影的医学图像中提供更稳定的结果. 配准后的运动图像 $M\circ A$ 与固定图像 F 之间的 NTG 如式 (6) 所示:

$$NTG(M\circ A, F) = \frac{\sum_l (\|\nabla_l (M\circ A - F)\|_1)}{\sum_l (\|\nabla_l M\circ A\|_1 + \|\nabla_l F\|_1)} \tag{6}$$

其中, A 是预测的仿射变换矩阵, $M\circ A - F$ 表示配准后运动图像和固定图像的差分图像, ∇_l 表示沿 $l\in\{x,y\}$ 方向的导数. 式 (6) 很容易验证 $0\leqslant NTG\leqslant 1$. NTG 衡量的是两幅图像的梯度差异, 总梯度差异越小, 表示图像间的相似性越高. 因此, 最小化 NTG 相当于最大化 $M\circ A$ 和 F 的相似性. 图像配准目的是通过最小化损失函数找到最优的仿射变换矩阵 A. 最终的损失函数如式 (7) 所示:

$$L_{\text{sim}}(M\circ A, F) = \frac{\sum_l \sum_{x\in\Omega} |\nabla_l (M(x)\circ A - F(x))|}{\sum_l \sum_{x\in\Omega} (|\nabla_l (M(x)\circ A)| - |\nabla_l F(x)|)} \tag{7}$$

其中, Ω 表示 F 和 $M\circ A$ 的重叠区域, 其定义为经过仿射变换后具有有效值的点集合.

4 实验和结果

4.1 数据集及预处理

本文在两个公开数据集上 OASIS (open access series of imaging studies)[24] 和 LPBA[25]评估了所提方法的性能. 其中, OASIS 数据集来源于阿尔茨海默病神经影像学倡议 (Alzheimer's disease neuroimaging initiative, ADNI), 该数据集包含 414 个正常老人和患有不同程度阿尔茨海默病患者的脑图像和分割图 (https://www.oasis-brains.org). LPBA 由神经影像研究所 (Laboratory of Neuro Imaging, LONI) 开发, 该数据集收集了 40 个健康成年人的脑图像和手动分割图 (https://www.loni.usc.edu/research/atlas_downloads). 图 4 展示了从 OASIS 和 LPBA 数据集中获得的横截面 (Axial)、冠状面 (Coronal) 和矢状面 (Sagittal) 的切片实例, 以及相应的分割图切片.

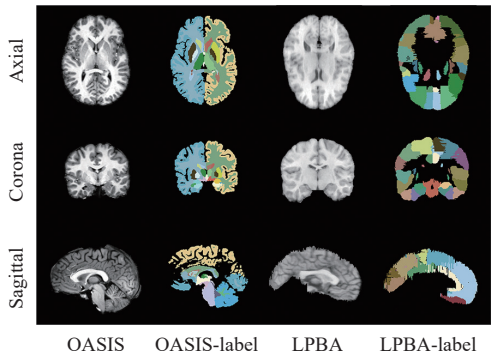


图 4 OASIS 和 LPBA 数据集的脑 MRI 切片及相应分割图的示例

对于 OASIS 数据集, 本文将图像裁剪为 128×128×128, 分辨率保持为 1 mm×1 mm×1 mm, 然后对图像进行标准的预处理, 包括运动矫正、脑提取和归一化等. 对于 LPBA 数据集, 本文使用 1.2 mm×1.2 mm×1.2 mm 的分辨率对所有图像进行重采样, 裁剪为 128×128×128, 并进行归一化. 为了构建配准数据集, 本文首

先将 OASIS 数据集按照 8:2 的比例分为训练集 331 例和测试集 83 例. 然后, 在 OASIS 的训练集、测试集以及 LPBA 数据集中进行随机不重复的两两配对. OASIS 训练集中生成 4 000 例待配准数据对用于训练, OASIS 测试集和 LPBA 数据集中分别生成 1 000 例待配准数据对用于测试.

4.2 实验设置

本文方法与 5 种配准方法进行对比, 包括传统配准方法质心 (center of mass, CoM) 初始化以及基于深度学习的配准方法 SYMNet[11]、VoxelMorph (VM)[8]、NICE-Trans[26]和 C2FViT[22]. CoM 是一种简单有效的配准方法, 在三维医学图像配准中常用于确定初始平移变换参数. SYMNet 是一种快速对称可变形配准方法, 该方法通过对称性约束来实现高效且精确的图像配准 (https://github.com/cwmok/Fast-Symmetric-Diffeomorphic-Image-Registration-with-Convolutional-Neural-Networks). VoxelMorph 是近年来提出的最前沿的无监督可变形配准方法, 也是文献中常用的基准配准方法 (https://github.com/balakg/voxelmorph). NICE-Trans 是一种非迭代粗到精的配准方法, 该方法将仿射和可变形配准结合, 使网络能够同时学习和优化两种变换 (https://github.com/MungoMeng/Registration-NICE-Trans). C2FViT 是一种由粗到精的配准方法, 该方法利用卷积 ViT 的全局连通性和局部性以及多分辨率策略来学习图像之间的全局仿射变换关系 (https://github.com/cwmok/C2FViT).

对于基于深度学习的方法, 本文使用它们的官方在线实现, 从头开始训练. 为了获得最佳性能, 所有超参数设置均遵循各自文献中的最佳参数. 在 SYMNet 中, 损失函数的方向一致性损失、正则化损失和幅度损失的权重分别设置为 100、3 和 0.1. 对于 VoxelMorph 和 NICE-Trans 的损失函数, 其正则化损失权重分别设置为 1.5 和 1. C2FViT 只使用 NCC 作为损失函数, 权重设为 1. 本文所提方法采用 Adam 优化器, 初始学习率设置为 1×10^{-4}, 批处理大小设置为 1, 训练迭代次数为 160 000 次, 与基于深度学习的对比方法保持一致. 训练模型的编程语言使用 Python, 深度学习框架 PyTorch 版本为 1.13.0, 实验使用独立显卡 Nvidia Geforce RTX 3090 训练和测试模型.

4.3 评估指标

为了量化本文方法的配准性能, 本文使用戴斯相似系数 (Dice similarity coefficient, *DSC*)[27]计算图像之间在相同解剖结构区域的体积重叠程度. *DSC* 取值范围为 0–1, 数值越大表示重叠程度越大, 配准性能越好. 本文还采用分割图边界像素点的豪斯距离第 95 百分位数 (Hausdorff distance 95th percentile, *HD*95)[28]来评估配准结果的精度. *HD*95 用于衡量两个图像间的形状差异, 较小的 *HD*95 值表示两个图像的形状或边界更相似, 即图像之间的一致性更高. 本文用 X 表示固定图像 F, 用 Y 表示配准后的图像 $M\circ A$, *DSC* 和 *HD*95 的定义如式 (8)、式 (9) 所示:

$$DSC(X,Y)=2\times\frac{V_X\cap V_Y}{|V_X|+|V_Y|} \tag{8}$$

$$\begin{aligned}HD95(X,Y)&=\max_{95\%}\{d_{XY},d_{YX}\}\\&=\max_{95\%}\{\max_{x\in B_X}\min_{y\in B_Y}\|x-y\|,\max_{y\in B_Y}\min_{x\in B_X}\|y-x\|\}\end{aligned} \tag{9}$$

其中, V_X和V_Y分别表示图像 X 和 Y 的像素点集合, B_X和B_Y分别表示图像 X 和 Y 的边界像素点集合, d_{XY}表示对于图像 X 的每个边界像素点$x\in B_X$, 计算其到图像 Y 中所有边界像素点的最小距离的最大值, d_{YX}表示对于图像 Y 的每个边界像素点$y\in B_Y$, 计算其到图像 X 中所有边界像素点的最小距离的最大值. $\max_{95\%}$表示对距离集合进行升序排序, 取集合中的第 95 百分位数作为 *HD*95 的值.

在图像的相似性评估上, 本文使用结构相似性 (structural similarity, *SSIM*)[29]和均方误差 (mean square error, *MSE*)[30]对图像的相似性进行评价. *SSIM* 不仅考虑了亮度的差异, 还考虑了结构和纹理的差异, 全面地反映了图像间的视觉相似性. *SSIM* 的取值范围为 0–1, 其值越接近 1, 表明两个图像在视觉上越接近, 从而更准确地模拟人眼对图像质量的感知. *MSE* 计算了两个图像之间每个像素的差异, 并将这些差异的平方值求和后取平均值. *MSE* 的数值越小, 表示两个图像越相似. 本文用 N 表示图像中的体素总数, x_i和y_i表示图像 X 和 Y 在第 i 个体素的数值, *MSE* 和 *SSIM* 的定义如式 (10)、式 (11) 所示:

$$MSE(X,Y)=\frac{1}{N}\sum_{i=1}^{N}(x_i-y_i)^2 \tag{10}$$

$$SSIM(X,Y)=\frac{(2\mu_X\mu_Y+C_1)(2\sigma_{XY}+C_2)}{(\mu_X^2+\mu_Y^2+C_1)(\sigma_X^2+\sigma_Y^2+C_2)} \tag{11}$$

其中，C_1和C_2是为了数值稳定性而添加的常数项, 通常取值为$C_1=(K_1L)^2$和$C_2=(K_2L)^2$, 其中, L 是图像的

像素值动态范围, K_1和K_2是两个小常数, 用于平滑亮度相似性的计算. μ_X和μ_Y分别是图像 X和 Y的均值, σ_X^2和σ_Y^2分别是图像 X和 Y的方差, σ_{XY}是图像 X和 Y之间的协方差, 计算公式如下:

$$\begin{cases}\mu_X=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}x_i\\ \mu_Y=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}y_i\end{cases} \tag{12}$$

$$\begin{cases}\sigma_X^2=\dfrac{1}{N-1}\displaystyle\sum_{i=1}^{N}(x_i-\mu_X)^2\\ \sigma_Y^2=\dfrac{1}{N-1}\displaystyle\sum_{i=1}^{N}(y_i-\mu_Y)^2\end{cases} \tag{13}$$

$$\sigma_{XY}=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y) \tag{14}$$

同时, 为了评估配准方法的计算效率, 本文使用 Ttest 指标量化每个配准方法在测试集上的平均预测时间, 即预测一对待配准图像的仿射变换矩阵或变形场所需要的时间.

4.4 实验结果及分析

表 3 展示了本文所提方法和所有对比方法在 OASIS 数据集和 LPBA 数据集上的配准性能. CoM 在两个数据集中都显示出较差的性能, 特别在 *DSC* 和 *HD*95 方面, 配准精度较低且边界误差较大. 虽然它的 *MSE* 和 *SSIM* 指标相对较好, 但与其他方法相比仍存在显著差距. CoM 通过计算图像的质心来进行图像配准, 它仅利用图像的几何中心信息, 而忽略了图像内部的细节和结构信息. 因此, 在处理复杂形变和非刚性形变时, 该方法无法提供足够的精度. CoM 不需要在 GPU 上运行, 所以不与其他方法在时间上进行比较. SYMNet 在 LPBA 数据集中整体表现较好, 尤其在 *DSC* 和 *HD*95 方面. SYMNet 是一种基于对称性网络结构的配准方法, 这种结构有助于减少不对称变形, 但在处理复杂形变时可能不够灵活, 导致它在 OASIS 数据集上的配准精度不高. VoxelMorph 利用 CNN 实现快速图像配准, 在两个数据集中都表现出较小的 *MSE*. 然而, CNN 在处理大范围形变时可能存在局限性, 整体配准性能不如本文方法. NICE-Trans 在 OASIS 数据集和 LPBA 数据集上的性能同样出色, 仅测试时间稍长. 相比 SYMNet 和 VM, NICE-Trans 在 *DSC* 和 *SSIM* 方面表现更好, 但整体性能略低于本文方法. 该方法利用联合仿射和可变形进行配准, 能够处理复杂的形变, 但计算效率相对较低. C2FViT 在两个数据集上表现出较高的配准精度和计算效率, 在 OASIS 数据集上运行时间仅为 0.0793 s, 在 LPBA 数据集上仅为 0.0370 s. 然而, 由于 C2FViT 采用 NCC 作为相似性度量, 无法很好地捕捉非线性特征和局部细节, 影响了最终的配准效果, 所以整体表现不及本文方法.

表 3 各方法在 OASIS 和 LPBA 数据集上的配准结果

配准方法	OASIS					LPBA				
	DSC↑	*HD*95↓	*MSE*↓	*SSIM*↑	Ttest↓	*DSC*↑	*HD*95↓	*MSE*↓	*SSIM*↑	Ttest↓
CoM	0.7592	23.5419	0.01074	0.6432	—	0.7153	12.7895	0.03512	0.6948	—
SYMNet	0.8706	18.2607	0.00567	0.8561	0.3696	0.9306	7.7041	0.01619	0.7814	0.3175
VM	0.8569	16.9353	0.00260	0.8608	0.3841	0.9268	8.0385	0.01475	0.8070	0.2302
NICE-Trans	0.9385	7.0131	0.00219	0.9024	0.1691	0.9309	7.8078	0.01318	0.8101	0.2916
C2FViT	0.9359	7.3169	0.00194	0.9032	**0.0793**	0.9152	9.2411	0.01957	0.7515	**0.0370**
Ours	**0.9473**	**6.5246**	**0.00185**	**0.9068**	0.0876	**0.9370**	**7.0706**	**0.01254**	**0.8242**	0.0751

本文方法在 OASIS 数据集和 LPBA 数据集上均表现出最优配准性能, 同时还保持了较高的计算效率. 这些优势来源于本文的方法在特征提取、全局依赖关系捕捉和损失函数设计上的创新和改进. 相比 VoxelMorph 和 SYMNet 的 CNN 结构, 本文的方法采用基于膨胀卷积的 Transformer 结构, 能够更好地捕捉全局和局部特征, 减少信息丢失, 处理更复杂的形变. 在两个数据集中, 本文方法与 VoxelMorph 和 SYMNet 相比在 *DSC* 和 *HD*95 方面展示了优越的性能, 这表明误差显著减少, 配准精度大幅提升. NICE-Trans 的基本结构是基于 CNN 和 Transformer 的组合. 然而, CNN 在处理图像时具有局部感知性, 即卷积的感受野范围有限. 这意味着网络在处理远程相关的图像区域时可能无法获取足够的信息, 导致配准的准确性受到影响. 本文方法将膨胀卷积和 Transformer 结合, 可以扩大卷积的感受野范围, 同时帮助 Transformer 处理远程依赖关系. 因此, 本文方法能够在图像配准任务中取得更好的性能. 相比基于 NCC 的 C2FViT, 本文方法采用 *NTG* 能够更全面地考虑图像的梯度信息, 降低噪声等因素对配准的影响, 捕捉更多的细节和非线性特征, 因此整体

配准效果更佳. 为了进一步展示所提方法的优势, 本文绘制了评价指标柱状图, 展示了不同配准方法在每个评价指标中的性能, 如图 5 所示.

本文展示了基于学习的配准方法在 OASIS 数据集上的配准结果, 如图 6 所示. 由图 6 可知, 本文方法在任何切面上的配准结果都与固定图像更为相似. 为了更清晰地比较本文方法与对比方法在配准性能上的差异, 本文将运动图像或配准结果叠加到固定图像上, 以便直观地观察图像之间的差异. 以 OASIS 数据集的脑 MRI 横截面为例, 叠加效果如图 7 所示. 其中, 红色代表固定图像的非黑色像素, 绿色代表运动图像和配准结果的非黑色像素. 图 6(a) 展示了固定图像和运动图像之间的初始差异, 显示出明显的位置和形变差异. 图 6(b) 和图 6(c) 分别显示了 SYMNet 和 VoxelMorph 的差异图, 脑部边缘和内部结构明显存在大量未重合区域, 表明具有较差的结构相似性和较大的边界误差. 图 6(d) 和图 6(e) 分别显示了 NICE-Trans 和 C2FViT 的差异图, 尽管大脑内部结构的区域重合度较好, 但脑部边缘存在明显的红色和绿色未重合区域. 图 6(f) 显示了本文方法的差异图, 与其他方法相比, 图 6(f) 不仅更准确地对齐图像, 而且在外观上更接近固定图像的体积大小. 从图 7 可以看出, 对比方法的配准结果均存在不同程度的缩放和旋转角度差异, 而本文方法在配准精度和重合度方面表现最佳, 特别是在处理复杂的大脑内部结构时, 能够实现更高的配准精度和一致性.

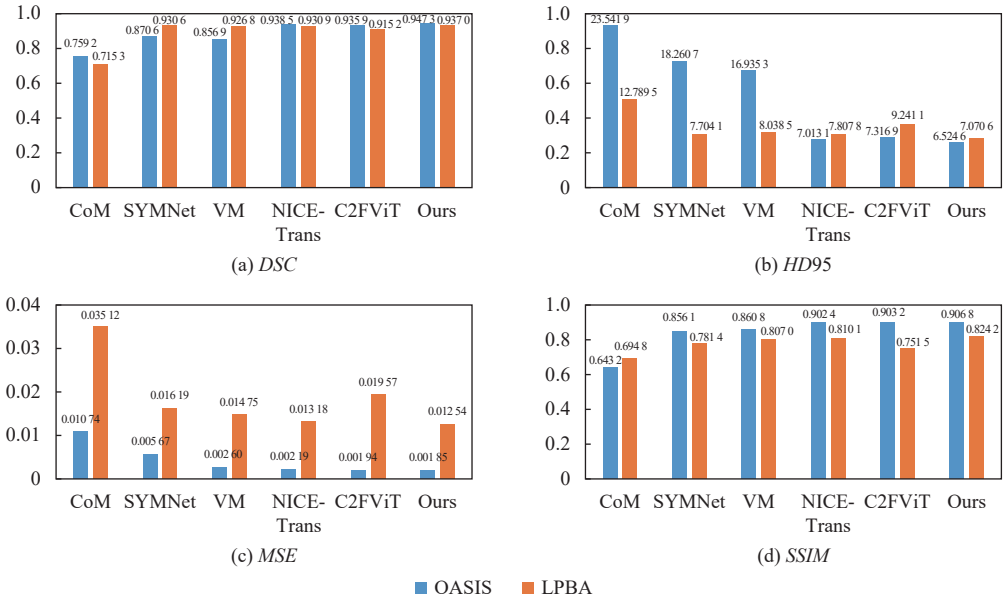


图 5 不同配准方法的评价指标柱状图

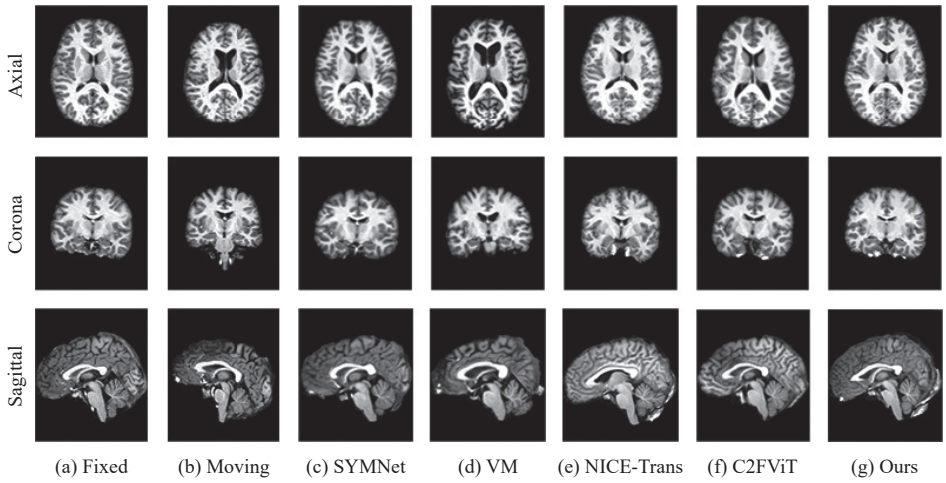
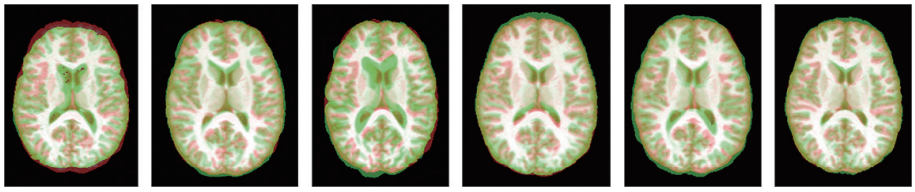


图 6 基于学习的配准方法在 OASIS 数据集上的配准结果



(a) Initial (b) SYMNet (c) VM (d) NICE-Trans (e) C2FViT (f) Ours

图 7 不同配准方法的结果对比

表 4 展示了本文方法基于 OASIS 数据集的消融实验结果. 本文在 Dil-ViT 和 *NTG* 这两个创新点进行了消融实验. 结果显示, 虽然使用 ViT 模型与 NCC 作为相似性度量的基线方法表现出较高的配准性能, 但仍有改进空间. 结合 Dil-ViT 和 NCC 的配准方法以及结合 ViT 和 *NTG* 的配准方法在性能上不分上下, 但相比基线方法, 各项指标均有所提升, 表明这些改进能够进一步提高本文方法的配准性能.

表 4 消融实验

序号	Dil-ViT	*NTG*	*DSC*↑	*HD*95↓	*MSE*↓	*SSIM*↑
0	—	—	93.5975	7.3169	0.001 943	0.903 2
1	√	—	94.5056	6.8998	0.001 867	0.906 6
2	—	√	94.4141	6.8873	0.001 871	0.906 8
3	√	√	**94.7332**	**6.5246**	**0.001 858**	**0.906 8**

一方面, 膨胀卷积不仅能够帮助网络感知更广泛的图像内容, 还可以更好地保留图像中微小的结构变化或特征. Dil-ViT 将膨胀卷积与具有强大全局特征提取能力的 ViT 结合, 从而增强了配准方法对全局信息和局部特征的感知能力. 另一方面, 基于 *NTG* 的损失函数利用图像的梯度信息引导配准过程, 通过度量图像灰度值的变化强度, 有助于模型更精确地对齐图像. 结合 Dil-ViT 和 *NTG* 的配准方法能够同时利用两者的优势, 显著提升图像配准的性能. Dil-ViT 和 *NTG* 在 *DSC*、*HD*95、*MSE* 和 *SSIM* 指标上的出色表现证明了在实际应用中的优越性和广泛适用性.

5 总结

本文提出了一种基于 Transformer 的多尺度可变形三维医学图像配准方法, 该方法可以有效处理大位移和复杂形变. 本文在脑 MRI 图像数据上进行了广泛的定性和定量比较, 结果表明所提方法优于传统的配准方法 CoM, 以及基于深度学习的配准方法 SYMNet、NICE-Trans、VoxelMorph 和 C2FViT. 本研究的主要贡献在于提出了一种结合膨胀卷积和 Transformer 的配准模型, 并引入了 *NTG* 作为相似性度量方法.

为了将局部性引入 Transformer 模型, 本文提出了一种基于膨胀卷积的 Transformer 模型. 通过扩大感受野, 膨胀卷积不仅能够捕捉更大范围的全局信息, 还能保留局部特征的细节信息, 从而增强网络模型对全局信息和局部特征的感知能力. 在增大感受野的同时, 膨胀卷积保持计算复杂度基本不变, 使得模型在处理大范围图像信息和高分辨率图像时都能够保持高效. 由于基于强度的相似性度量容易受到噪声干扰, 本文引入了 *NTG* 作为新的相似性度量. *NTG* 基于两幅图像完全对齐时其差分图像的梯度最稀疏这一假设, 能够有效降低医学图像中图像噪声和伪影等因素对配准的影响, 从而实现更精确的医学图像配准. 此外, 本文将 *NTG* 和图像金字塔相结合, 使其在处理大位移和复杂变形时仍能保持高效性和准确性.

未来, 本文将从以下几个方面尝试提高所提方法的效率、准确性和鲁棒性: (1) 进一步优化网络模型, 使其能够应用于更广泛的医学图像配准任务, 不仅限于脑 MRI 图像; (2) 在网络模型内引入尺度空间的概念, 使本文所提的配准方法具有尺度不变性.

参考文献

1 郭艳芬, 崔喆, 杨智鹏, 等. 基于深度学习的医学图像配准技术研究进展. 计算机工程与应用, 2021, 57(15): 1–8. [doi: 10.3778/j.issn.1002-8331.2101-0281]

2 罗雅雯, 王远军. 基于特征提取的脑部医学图像配准研究. 生物医学工程学进展, 2023, 44(3): 226–234. [doi: 10.3969/j.issn.1674-1242.2023.03.002]

3 邱玲娜, 吴经芝, 吴泳兴, 等. 不同 CT-CBCT 配准方法对颈胸膜固定方式下食管癌患者摆位误差的影响. 福建医科大学学报, 2023, 57(3): 217–222. [doi: 10.3969/j.issn.1672-4194.2023.03.014]

4 Ji HZ, Li YS, Dong EQ, *et al*. A non-rigid image registration method based on multi-level B-spline and L2-regularization. Signal, Image and Video Processing, 2018, 12(6): 1217–1225. [doi: 10.1007/s11760-018-1274-0]

5 Chun SY, Fessler JA. A simple regularizer for B-spline nonrigid image registration that encourages local invertibility. IEEE Journal of Selected Topics in Signal Processing, 2009, 3(1): 159–169. [doi: 10.1109/JSTSP.2008.

2011116]
6 Fan JF, Cao XH, Yap PT, *et al*. BIRNet: Brain image registration using dual-supervised fully convolutional networks. Medical Image Analysis, 2019, 54: 193–206. [doi: 10.1016/j.media.2019.03.006]
7 Sokooti H, de Vos B, Berendsen F, *et al*. Nonrigid image registration using multi-scale 3D convolutional neural networks. Proceedings of the 20th International Conference on Medical Image Computing and Computer-assisted Intervention. Quebec City: Springer, 2017. 232–239.
8 Balakrishnan G, Zhao A, Sabuncu MR, *et al*. VoxelMorph: A learning framework for deformable medical image registration. IEEE Transactions on Medical Imaging, 2019, 38(8): 1788–1800. [doi: 10.1109/TMI.2019.2897538]
9 Shi JC, He YT, Kong YY, *et al*. XMorpher: Full Transformer for deformable medical image registration via cross attention. Proceedings of the 25th International Conference on Medical Image Computing and Computer-assisted Intervention. Singapore: Springer, 2022. 217–226.
10 Kim B, Kim DH, Park SH, *et al*. CycleMorph: Cycle consistent unsupervised deformable image registration. Medical Image Analysis, 2021, 71: 102036. [doi: 10.1016/j.media.2021.102036]
11 Mok TCW, Chung ACS. Fast symmetric diffeomorphic image registration with convolutional neural networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 4643–4652.
12 应时辉, 杨菀, 杜少毅, 等. 基于深度学习的医学影像配准综述. 模式识别与人工智能, 2021, 34(4): 287–299.
13 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st Conference on Neural Information Processing Systems. Long Beach: NIPS, 2017. 5998–6008.
14 Liu Z, Lin YT, Cao Y, *et al*. Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9992–10002.
15 Chen JY, Frey EC, He YF, *et al*. TransMorph: Transformer for unsupervised medical image registration. Medical Image Analysis, 2022, 82: 102615. [doi: 10.1016/j.media.2022.102615]
16 Avants BB, Epstein CL, Grossman M, *et al*. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis, 2008, 12(1): 26–41. [doi: 10.1016/j.media.2007.06.004]
17 Viola P, Wells III WM. Alignment by maximization of mutual information. International Journal of Computer Vision,1997,24(2):137–154.[doi:10.1023/A:1007958904918]
18 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. Proceedings of the 4th International Conference on Learning Representations. San Juan: ICLR, 2016.
19 Dosovitskiy A, Beyer L, Kolesnikov A, *et al*. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929, 2021.
20 马明睿. 基于深度学习的医学影像配准算法研究 [博士学位论文]. 长春: 吉林大学, 2023.
21 Jaderberg M, Simonyan K, Zisserman A, *et al*. Spatial transformer networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 2017–2025.
22 Mok TCW, Chung ACS. Affine medical image registration with coarse-to-fine vision Transformer. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 20803–20812.
23 Chen SJ, Shen HL, Li CG, *et al*. Normalized total gradient: A new measure for multispectral image registration. IEEE Transactions on Image Processing, 2018, 27(3): 1297–1310. [doi: 10.1109/TIP.2017.2776753]
24 Marcus DS, Wang TH, Parker J, *et al*. 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]
25 Shattuck DW, Mirza M, Adisetiyo V, *et al*. Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage, 2008, 39(3): 1064–1080. [doi: 10.1016/j.neuroimage.2007.09.031]
26 Meng MY, Bi L, Fulham M, *et al*. Non-iterative coarse-to-fine transformer networks for joint affine and deformable image registration. Proceedings of the 26th International Conference on Medical Image Computing and Computer-assisted Intervention. Vancouver: Springer, 2023. 750–760.
27 Dice LR. Measures of the amount of ecologic association between species. Ecology, 1945, 26(3): 297–302. [doi: 10.2307/1932409]
28 Taha AA, Hanbury A. An efficient algorithm for calculating the exact Hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(11): 2153–2163. [doi: 10.1109/TPAMI.2015.2408351]
29 Wang Z, Bovik AC, Sheikh HR, *et al*. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600–612. [doi: 10.1109/TIP.2003.819861]
30 Beg MF, Miller MI, Trouvé A, *et al*. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision, 2005, 61(2): 139–157. [doi: 10.1023/B:VISI.0000043755.93987.aa]

(校对责编: 王欣欣)