

基于多空间信息提取的级联分块配准模型^①

王南南¹, 程远志^{1,2}

¹(青岛科技大学 信息科学技术学院, 青岛 266061)

²(哈尔滨工业大学(威海) 信息科学与工程学院, 威海 264209)

通信作者: 程远志, E-mail: yzchengqust2007@163.com



摘要: 在基于深度学习的医学图像配准中, 当医学图像中包含多种组织类型时, 不同组织之间结构的不同可能会导致网络配准的精度下降, 特别是在复杂形变区域, 如组织的交界处和病变区域, 精准的配准变得更加困难, 现有的配准算法对复杂形变区域的关注度不高, 导致配准精度较低. 同时现有的配准网络无法同时对图像的局部和全局空间信息进行捕获, 导致网络的鲁棒性不够, 在迁移到其他器官的配准工作中时配准准确率低. 为了解决上述的问题, 本文提出一种基于多空间信息提取的级联分块配准模型, 本模型可以有效利用输入图像的局部和空间信息, 并通过分块融合的技术, 将医学图像进行分块并依次对每个图像进行精细配准生成相应的形变场块, 在模型的最后阶段将生成的形变场块进行融合还原, 以增强网络对局部复杂形变区域的配准强度. 实验结果表明, 所提方法不仅在脑部配准上有所提升, 并且在其他人体部位的配准中也有较好的表现, 提高了医学图像配准的准确性和可靠性, 为临床医生提供更好的诊断和治疗支持.

关键词: 医学影像; 级联网络; 自注意力机制; 分块融合; 信息提取; 注意力机制; 图像融合

引用格式: 王南南, 程远志. 基于多空间信息提取的级联分块配准模型. 计算机系统应用, 2024, 33(2): 125-133. <http://www.c-s-a.org.cn/1003-3254/9401.html>

Cascaded Block Registration Model Based on Multi-spatial Information Extraction

WANG Nan-Nan¹, CHENG Yuan-Zhi^{1,2}

¹(School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

²(School of Information Science and Engineering, Harbin Institute of Technology, Weihai, Weihai 264209, China)

Abstract: In medical image registration based on deep learning, when the medical image contains multiple tissue types, the structural difference between different tissue may lead to a decrease in the accuracy of network registration, especially in complex deformation regions, such as the junction of tissue and the lesion region, and accurate registration becomes more difficult. The existing registration algorithms have low registration accuracy for complex deformation regions. At the same time, the existing registration network cannot capture the local and global spatial information of the image at the same time, resulting in insufficient robustness of the network and low accuracy when it is transferred to other organs for registration. In order to solve the above problems, this study creates a cascaded block registration model based on multi-spatial information extraction. This model can effectively use the local and spatial information of input images, divide medical images into blocks through block fusion technology, and perform fine registration for each image in turn to generate corresponding deformation field blocks. In the last stage of the model, the generated deformation field blocks are fused and restored to enhance the registration strength of the network for the local complex deformation region. The experimental results show that the proposed method not only improves the registration of the brain but also performs well in the registration of other human body parts, which improves the accuracy and reliability of medical image registration.

① 收稿时间: 2023-06-24; 修改时间: 2023-07-27; 采用时间: 2023-10-08; csa 在线出版时间: 2023-12-26

CNKI 网络首发时间: 2023-12-28

and provides better diagnosis and treatment support for clinicians.

Key words: medical image; cascading network; self-attention mechanism; block fusion; information extraction; attention mechanism; image fusion

1 引言

医学图像配准的目标是通过寻找空间变化的算法,将一幅图像中的区域识别并对齐另一幅图像中相似或者相同的解剖结构^[1],使两幅图像在空间中相同的位置可以一一对应,从而可以对不同的图像进行比较和融合.医学图像配准可分为刚体配准和非刚体配准,其中非刚体配准又称可变形图像配准^[2].刚体配准是指目标物体内部任意两点之间的距离在变换过程中保持不变,刚体变换主要由平移和旋转操作组成.由于刚体配准的参数有限,无法适应于形态变量复杂的图像.故目前应用的主流配准方法是可形变图像配准^[3].

可形变图像配准按照训练方式的不同,分为基于传统算法的配准方法和基于深度学习的配准方法.目前,基于扩散模型的 Demons 算法^[4]及其变体是传统算法中的主流算法. Demons 算法将图像配准视为一个扩散过程,利用参考图像灰度梯度中产生的局部信息来驱动待配准图像向参考图像的扩散,并对待配准图像进行迭代扭曲,直到单次迭代中的形变位移足够小时完成迭代配准.由于 Demons 算法在参考图像灰度梯度较小时迭代次数明显增加且效率较低,因此 Wang 等人^[5]将每次迭代变形后的待配准图像与参考图像结合作为扩散过程中的驱动力. Zhang 等人^[6]在每次迭代时观察两幅图像的相似性变化,当值达到最大时停止迭代优化.然而,这些迭代优化的配准算法非常耗时,难以用于实时的医学图像配准应用. Jaderberg 等人^[7]在反向传播算法的基础上提出空间转换网络 (spatial Transformer network, STN), STN 生成的形变场可以通过网络的反向传播进行优化并具有可微属性,通过使用配准后图像之间的相似性损失函数,可以避免复杂的损失函数设计. Balakrishnan 等人^[8]通过融合 STN 结构提出了一种新的端到端的无监督图像配准模型 (VoxelMorph),此模型的自定义损失函数可以有效提升图像配准的速度. Zhao 等人^[9]将级联网络运用到图像配准中,并证明级联做法可以提升配准的准确率.因此本文方法模型采用级联的设计对图像配准,通过对图像块进行精细

配准提升小区域的配准精度,从而提高整体配准结果的稳定性.

计算机视觉领域中卷积神经网络 (convolutional neural network, CNN) 通常被用来提取局部特征,但提取过程中是卷积层的感受野受到卷积核尺寸的限制,这意味着只能捕捉到局部特征而难以学习到全局特征,导致在处理复杂的医疗图像时出现配准误差,并且随着卷积网络层数的加深,图像中距离较远的像素点之间的依赖关系会急剧减小.在处理复杂的医疗图像时,这种依赖关系非常重要,因为图像中不同区域间可能具有复杂的依赖关系.为了解决这个问题, Liu 等人^[10]提出了拥有移位窗口的层次化 Transformer,可以动态的调整网络对不同区域的关注度,从而更好地学习全局特征. Chen 等人^[11]首次将 Transformer 构造成编码器-解码器的结构应用到无监督配准中,在实验结果中可以说明对比 CNN 网络的高效性和准确性.但由于自注意力机制本身的计算特性,导致纯 Transformer 构成的配准网络计算代价高昂,并存在训练显存过大的问题. gConv 网络是 Rao 等人^[12]总结了 Transformer 机制的成功之处,并将其应用于普通 CNN 中,以 CNN 的参数体量实现了接近 Transformer 网络的全局信息提取能力, gConv 提出的方法借助门控卷积和递归运算,可以高效地实现任意阶的空间交互,然而 gConv 在医学图像配准中的应用仍处于起步阶段.

因此本文在以上观点的基础上提出一种基于多空间信息提取的级联多阶层配准模型,该模型的主要特点是使用了多空间信息提取技术,可以在不同尺度上提取图像的特征信息,从而提高配准的准确度和鲁棒性.同时,该模型还采用了级联多阶层网络和分块模块,可以降低计算复杂度和内存占用量,适用于多种医学图像配准场景.该模型包含:① 构建由 gConv 网络组成的编码器,构建 Transformer、CNN 组成的解码器,并将二者组成 U 型基础配准网络.② 构造分块融合模块,对图像进行分块处理操作,并按照一定规则进行拼接融合,形成最终的形变场.③ 将多个基础配准网络通

过级联的方式连接起来,并在每层网络对不同大小的图像进行优化,从而解决图像中复杂局部区域的形变问题。

2 方法

2.1 总体架构

在计算机视觉领域中,图像分块是一种常用的处理策略.将图像分块后对每个图像块分别进行相应的处理,并在处理完毕后再按照一定的规则进行拼接融合.这种分块合并的方法可以更精确地处理图像的局部细小区域^[13],提高配准的准确性和准确性.同时,由于医学图像分辨率高、数据尺寸大的特点,通过将图像分块,可以逐个读取和处理从而可以加快处理速度。

此模型由3层子网络和1个融合模块组成.模型的整体架构如图1所示,模型共有3层子网络和1个融合模块,并在每层子网络中都包含分块模块和STN模块,3个子网络中的配准网络的结构是相同的,可以实现参数共享,从而提升配准的健壮性.模型的整体架构按照由粗到精^[14]的配准原则,自底向上的阶层顺序依次对图像进行配准,每层子网络的输入为一对待配准的图像,输出为对应的形变场块,将此形变场块与输入的形变场块一起输入到分块模块和STN模块后,输入到下一阶层子网络。

在融合模块中,按照自顶向下的阶层顺序合并每层子网络产生的形变场块,并且在合并过程中以上一层子网络产生的形变场块为参考,以减少边缘效应。

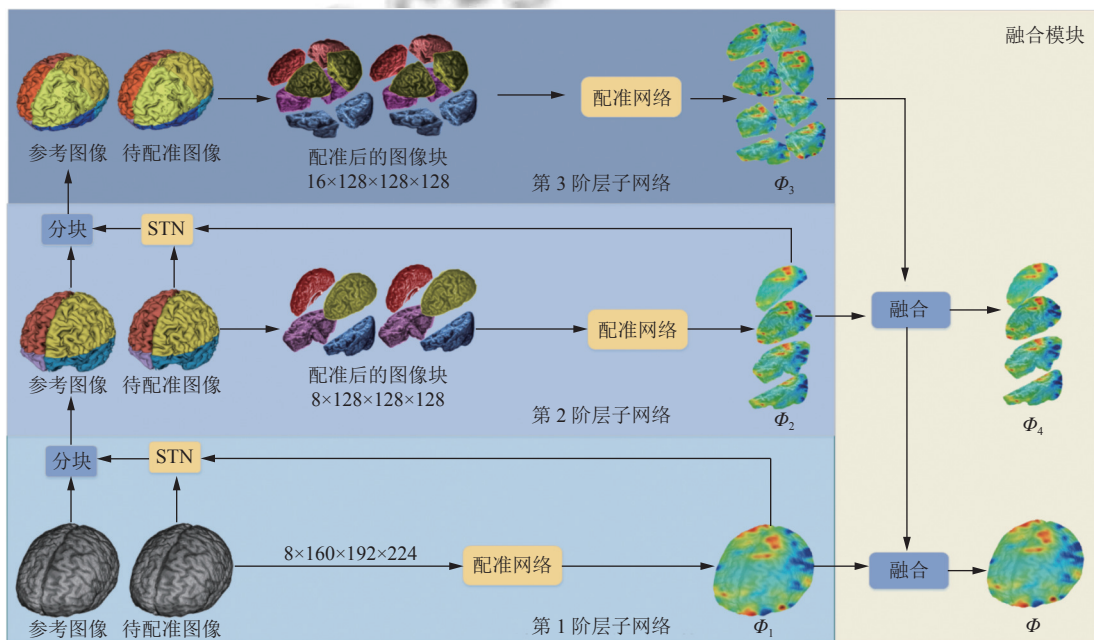


图1 级联多阶层配准网络模型总体图

2.2 基础配准网络

在计算机视觉领域,CNN网络的优点是局部感知性强,可以有效地捕捉到局部空间的特征信息,提高模型对局部特征的感知性.Transformer网络的优势是可以使用自注意力机制对全局空间特征信息进行提取,可以有效捕捉到图像中的长期依赖关系,而gConv网络结合了两网络各自的优点。

MainNet分为3个模块,gConv编码器、Transformer解码器和普通卷积解码器,如图2所示.基础配准网络使用gConv网络构成的编码器进行下采样的同时保留

局部和全局空间信息,使用Transformer和CNN构成的解码器应用于还原全局和局部空间信息,并在形变场生成前,将参考图像和Transformer和CNN的解码输出结合,增强网络中特征图空间上的理解能力。

gConv编码器:传统Transformer的成功因素包括动态权重、长距离建模和高阶的空间交互,gConv提取高阶的空间交互和局部特征,可以获得更空间交互,虽然通过使用卷积线性投影实现和element-wise乘法进行实现,但具有和自注意力随的权重随输入变化而变化的功能.通过使用3D卷积层的目标是在普通卷积

的基础上实现长距离建模和局部特征,可以获得更加精细的细节特征和提取图像的空间信息联系,同时为了解决反向传播中的梯度消失问题,每个gConv后都有一个(layer normalization, LN)模块。

Transformer 解码器:第2阶段由Transformer解码器建立一系列的序列信息还原图像的全局空间特征.此解码器由4个Transformer blocks组成,每个Transformer block由一个Transformer模块和上采样模块构成,每个头都可以对输入的特征图进行注意力加权以提取到不同的特征,再将输出进行拼接和线性变换,从而学习到输入图像各部分的相关性。

CNN 解码器:第3阶段由CNN解码器还原图像的局部特征关系.CNN解码器由一系列的普通3D卷

积构成,作用是还原图像的局部特征信息,通过将同级的Transformer block的输出进行跳跃连接弥补CNN对捕获全局空间信息的不足,可以更好地学习到全局空间信息的相关性和共享性,从而提高网络的性能和稳定性。

Transformer解码器和CNN解码器中上采样使用转置卷积层将低分辨率特征图的尺寸扩大为原尺寸的两倍,相比于向上插值采样,转置卷积学习能力更强,可以学习到更复杂的特征信息,并可以适应不同的数据集.在最后阶段,将网络最后一层的Transformer block和CNN block进行一次上采样操作还原到和原始输入图像相同的尺寸,再将这三者经过一次卷积操作生成最终的形变场。

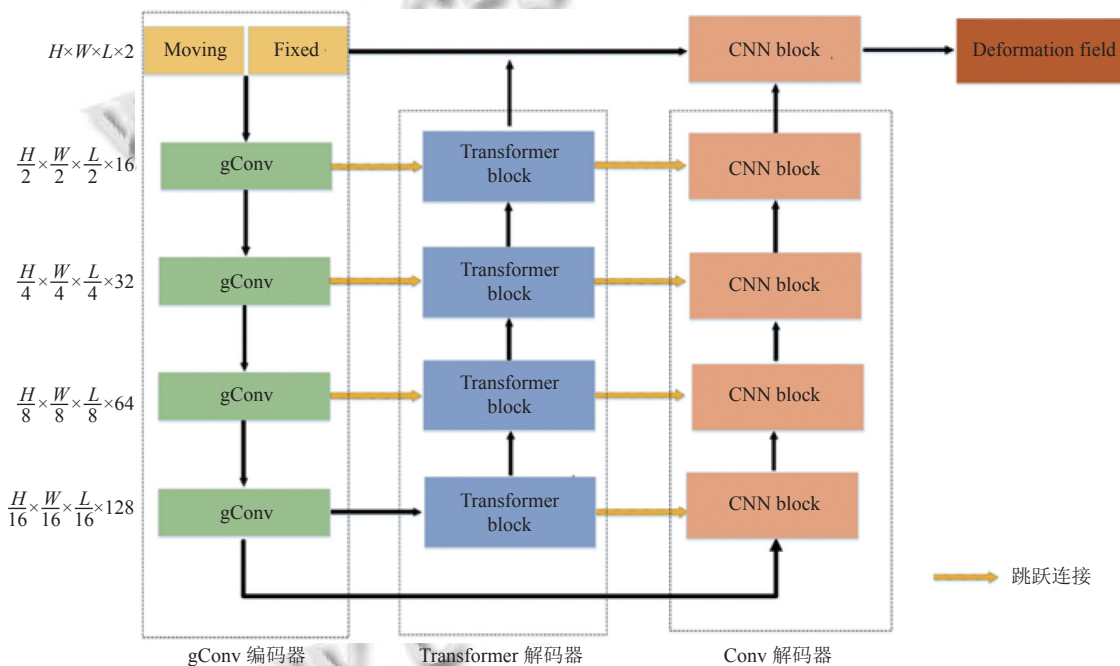


图2 基础配准网络

2.3 图像的分块与合并

医学图像配准的复杂程度体现在配准的结果上,具体表现为不同区域之间的差异性,在图像复杂区域的配准结果形变较大并且精度不高.因此提高图像整体配准精度的一个有效方案是将图像分块并找到配准形变较大的区域进行精细配准。

本模型在每一阶层的子网络中,将产生的形变场和本层的输入图像进行形变扭曲生成配准后的图像,将生成的图像和本层的输入图像在每个维度上同时使用滑动窗口计算两者相同位置差的平方值.因为图像

中越复杂的区域越难形变对齐,所以选取差的平方值最小的位置为分割位置,选取好分割位置后,在选定的位置利用浮动窗口计算并得到最终的位置.如图3所示,在滑动窗口阶段,通过滑动计算选定了各个维度的裁切位置,在浮动窗口阶段在选定的位置进行浮动,通过指定的浮动步长进行收缩或者扩张,从而确定最后的裁切位置.同时在图像进行裁切分块时,需要在图像块的边缘预留一些重叠区域,可以在后续的融合过程中可以平滑地处理边缘效应,减少合并处理结果的误差。

具体在本模型第1阶层子网络中,分块模块将完成初步配准的浮动图像和待配准图像在 x 和 y 轴上进行分块,将图像分成4个图像块.在第2阶层子网络中,分块模块将完成二次配准的浮动图像和待配准图像在 z 轴上进行分块,最终将图像分成8个子块.通过滑动窗口分块策略可以提高配准网络对复杂形变区域的关注度.同时可以帮助避免一些块或子区域复杂区域过度密集或过于稀疏的问题^[15],从而保持配准网络的均衡性和稳定性.通过对比滑动窗口分块策略的配准结果和平均分块策略的配准结果,可以发现滑动窗口分块策略在配准结果上最大提高了0.3%的准确率.

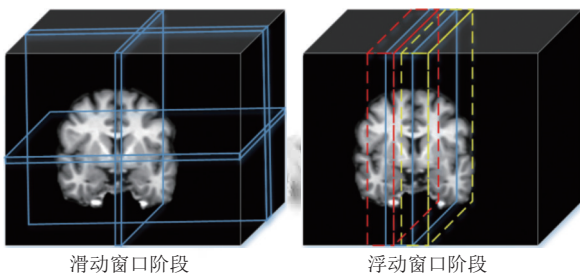


图3 图像分块过程

在每一层子网络中,将图像分块成有重叠区域的子块并输入到相应的配准网络完成计算后,参照上一层子网络输出的形变场,将本层配准网络输出的形变场块的重叠区域进行合并融合操作,来消除重叠区域并生成整体的形变场.如图4所示,若直接将生成的子块进行拼接,会导致在拼接的边缘部分出现边缘效应并损失大量的边缘信息,因此需要对重叠的边缘部分进行加权拼接处理.

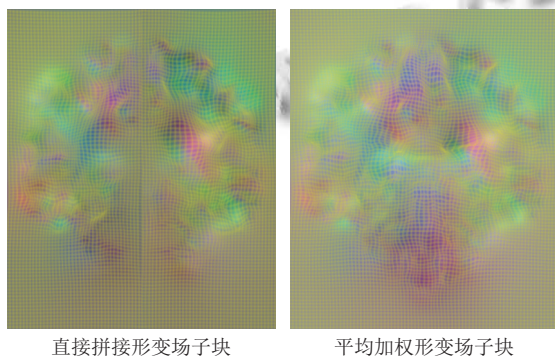


图4 拼接模式对比图

形变场块的重叠区域如图5所示,Region A 和 Region B 分别代表两块形变场块,各有一块重叠区域,将两块形变场的重叠区域的像素值按一定的权值相加

合成以形成平滑的过度.重叠边界区域经过融合后,最终的权值分配结果如图6所示.

在两个形变场块的重叠区域使用一个滑动窗口计算合并加权值,根据滑动窗口的大小将重叠区域分成 N 个子区域,对于重叠区域 S 内任意一个子区域 i 来说, S 最终确认为式(1)–式(5)所示:

$$S_i = S_{iA} \times P_{iA} + S_{iB} \times P_{iB} \tag{1}$$

$$P_{iA} = \frac{D_{iA}}{D_{iA} + D_{iB}} \tag{2}$$

$$P_{iB} = \frac{D_{iB}}{D_{iA} + D_{iB}} \tag{3}$$

$$D_{iA} = \text{MSE}(m_{if} - S_{iA} \circ O_{iA}) \tag{4}$$

$$D_{iB} = \text{MSE}(m_{if} - S_{iB} \circ O_{iB}) \tag{5}$$

其中, S_{iA} 是形变场块 A 中所包含的重叠区域中的子区域, S_{iB} 是形变场块 B 所包含的重叠区域, P_{iA} 表示为在最终的重叠区域 S_{iA} 中形变场块 A 所占的比例,其经过计算后,其中 m_{if} 为参考形变场对应的重叠区域, O_A 为 S_A 所对应的形变场, O_B 为 S_B 所对应的形变场.

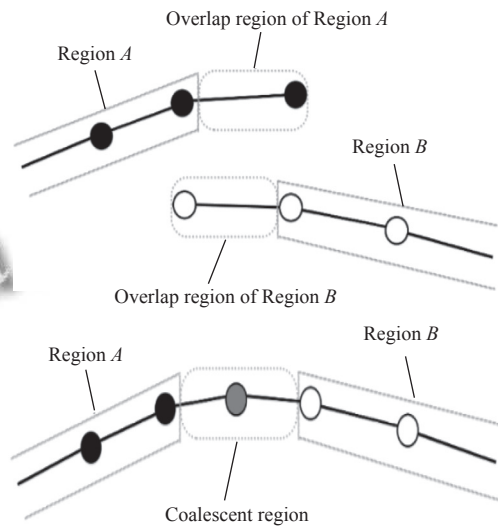


图5 形变场块合并示意图

	0.42	0.56	0.58	0.44	
Region A	0.48	0.73	0.52	0.27	Region B
	0.78	0.88	0.22	0.12	
	0.27	0.66	0.73	0.34	

图6 融合的两个形变场块在重叠区域的占比权重示意图

2.4 损失函数

本模型为无监督学习的方法,因此不需要在损失函数中计算预测值与真实标签的损失项,只需在训练过程中最小化图像相似损失项即可.模型总体损失函数如式(6)所示:

$$L(F, M, \varphi) = \lambda_1 L_{\text{sim}}(F, M \circ \varphi) + \lambda_2 R_{\text{diff}}(\varphi) \quad (6)$$

其中,在总体损失函数中分为两部分:第1部分为衡量图像间相似度的损失函数 $L_{\text{sim}}(F, M \circ \varphi)$,其定义如式(7)所示,其中 F 和 M 各自代表参考图像和待配准图像, φ 为网络生成的形变场;第2部分是保持形变场平滑度的扩散正则化项 $R_{\text{diff}}(\varphi)$,目的是为了防止在最小化 L_{sim} 时导致生成不平滑的形变场,其定义如(8)所示,其中 $\Delta u(p)$ 是形变场中空间位移的前后差分,其说明了各体素位置间的区别, $\Delta u(p)$ 的定义如式(9)所示. λ_1 和 λ_2 为权重系数,用来决定不同特征对模型输出结果的贡献大小.该权重系数的最佳值为1和0.02,因此将 λ_1 和 λ_2 大小各设置为1和0.02.

$$L_{\text{sim}}(F, M \circ \varphi) = \text{MSE}(F, M \circ \varphi) \quad (7)$$

$$R_{\text{diff}}(\varphi) = \sum_{p \in \Omega} \|\Delta u(p)\|^2 \quad (8)$$

$$\Delta u(p) = u(p+1) - u(p) \quad (9)$$

3 实验设置和评估标准

3.1 实验数据

为了测试此配准模型在不同器官的配准效果,本文共使用3个数据集,其中包括两个脑部数据集和一个膝盖软骨数据集:LPBA40数据集包括40个正常人的MRI扫描图像,每个个体的图像数据包括T1加权和T2加权扫描图像.此外,该数据集还包括手动标注的脑部结构分割和3D形变场;OASIS-1数据集包含了416位参与者的MRI图像和临床信息.每个参与者的MRI图像包括T1加权、T2加权和PD加权扫描图像,以及手动标注的脑部结构分割和病变标注.OAI数据集包含了多种类型的膝盖软骨数据,包括MRI图像、X光图像、病史信息、临床评估等,MRI图像包括膝盖、髌关节和手指等部位的图像.

数据集图像由于采集的时间、机器和其他原因导致数据间存在一定的差异,不利于网络的训练和泛化能力^[16],故对数据集进行对齐、归一化和裁剪等标准预处理操作,将图像裁剪到 $160 \times 192 \times 160$,并将3个数

据集通过随机配对的方式生成训练集、验证集和测试集,分配比例为6:2:2,并在训练过程中,对训练图像进行5折训练.

3.2 实验配置

在验证实验中使用传统算法中准确率较高的几个算法:对称归一化(symmetric normalization, SYN)^[17]、CycleMorph^[18]、VoxelMorph和TransMorph对本文方法进行对比.SYN是传统配准算法中性能最好的配准算法,其主要思想是利用基于正态分布的模型来估计图像间的相似性度量^[19],设置其最大迭代次数为160.CycleMorph是一种多尺度配准模型,通过优化互信息的值来实现两幅图像之间的精确配准.VoxelMorph是目前在工业生产中常用的无监督配准方法,本文使用其推荐的参数设置进行对比实验.TransMorph是将Transformer网络和无监督学习方法集合的配准网络,使用其默认参数进行对比实验.

深度学习框架使用:PyTorch-GPU 1.7.2,实验所使用的计算机配置如下:Nvidia Geforce RTX 3090, Intel Core i9-10900K CPU @ 3.70 GHz.学习率初始化为 1×10^{-4} ,训练次数设为500,训练时先随机初始化网络参数,定义网络参数的初始值后进行训练.

3.3 实验评价指标

本方法使用的评估指标为戴斯相似系数(Dice similarity coefficient, DSC),DSC常用于测试医疗图像的相似度,其取值在0和1之间,表示两个样本集合的相似程度,DSC值越高,说明图像相似度越高,配准性能和准确性越好.DSC的具体定义如式(10)所示:

$$DSC(S_F, S_M) = \frac{|S_F \cap S_M|}{(|S_F| + |S_M|)/2} \quad (10)$$

其中, S_F 和 S_M 分别表示为参考图像的分割图和配准后图像的分割图, $S_F \cap S_M$ 为两个分割图的交集结果.

负雅可比行列式(negative Jacobian determinant, NJD)是用于描述一个变换对图像局部伸缩性影响的数学指标^[20],NJD可以用来评估图像变形程度,特别是在非刚性变形(例如弯曲和扭曲)的情况下.在医学图像配准中,负雅可比行列式通常用于评估配准算法的效果.雅可比行列式的值越小,表示变形程度越大,即图像发生了较大的缩小或扭曲.在医学图像配准中,通常会选择最小化负雅可比行列式的优化算法,以最大程度地减小图像变形程度,从而实现精确配准.NJD系

数的定义如式 (11) 所示:

$$NJD(\varphi(p)) = - \begin{vmatrix} \frac{\partial i}{\partial x} & \frac{\partial j}{\partial x} & \frac{\partial k}{\partial x} \\ \frac{\partial i}{\partial y} & \frac{\partial j}{\partial y} & \frac{\partial k}{\partial y} \\ \frac{\partial i}{\partial z} & \frac{\partial j}{\partial z} & \frac{\partial k}{\partial z} \end{vmatrix} \quad (11)$$

图像相似性评估系数是用来比较两幅图像之间程度的指标,常用的图像相似性评估系数有:误差平方根 (root mean square error, *RMSE*), *RMSE* 是指两幅图像像素点之间差值的平方和的开方平均值,通常用于评估图像的重建质量,其定义如式 (12) 所示:

$$RMSE(F, M) = \sqrt{MSE(F, M)} \quad (12)$$

结构相似性系数 (structural similarity, *SSIM*), *SSMI* 是指两幅图像之间亮度、对比度和结构的相似程度,通常用于评估图像的质量和相似程度^[21],其具体计算过程如式 (13)–式 (16) 所示,式中 $L(F, M)$ 、 $C(F, M)$ 、 $S(F, M)$ 分别代表着参考图像 F 和待配准图像 M 两者的亮度、对比度和结构之间的比值差异。 μ_F 、 σ_F 、 μ_M 和 σ_M 分别为参考图像和待配准图像的平均值和标准差,同时为了避免分母过小造成误差在分母上加上非零常数 c_1 、 c_2 、 c_3 ,并都设置为 1.

$$SSIM(F, M) = [L(F, M)^a] [C(F, M)^b] [S(F, M)^c] \quad (13)$$

$$L(F, M) = \frac{2\mu_F\mu_M + c_1}{\mu_F^2 + \mu_M^2 + c_1} \quad (14)$$

$$C(F, M) = \frac{2\sigma_{FM} + c_2}{\sigma_F^2 + \sigma_M^2 + c_2} \quad (15)$$

$$S(F, M) = \frac{\sigma_{FM} + c_3}{\sigma_F\sigma_M + c_3} \quad (16)$$

互信息 (mutual information, *MI*), *MI* 是一种用于度量两幅图像之间相似性的指标,它是指两幅图像之间像素值的相关程度。*MI* 的定义如式 (17) 所示, $g(\cdot, \cdot)$ 为图像间的联合概率密度函数, $g(F(p))$ 和 $g(M(p))$ 分别表示图像的边缘概率密度函数.

可通过这 3 个指标评判图像的相似性,具体来说若两张图像间的 *RMSE* 高,而 *MI* 和 *SSIM* 变低,则表示两张图像相似性变差^[21].

$$MI(F, M) = \sum_{p \in \Omega} g(F(p), M(p)) \log_2 \frac{g(F(p), M(p))}{g(F(p))g(M(p))} \quad (17)$$

4 实验结果与定量分析

通过实验对本文方法和所有对比方法在脑部和膝关节的数据集上进行量化,得到相关的量化结果对比如表 1 所示,可以得知本文方法在两个脑部的数据集上的精准度均大于所有的对比方法,并且 *NJD* 系数优于除 *SYN* 的算法.

表 1 脑部数据集中不同方法的配准结果比较

方法	LPBA40		OASIS-1	
	<i>DSC</i>	<i>NJD</i>	<i>DSC</i>	<i>NJD</i>
参考图像	0.548±0.05	—	0.576±0.065	—
<i>SYN</i>	0.548±0.050	<0.000 01	0.767±0.033	0.000 15
<i>CycleMorph</i>	0.688±0.024	0.060 32	0.778±0.033	0.424 98
<i>VoxelMorph</i>	0.667±0.046	0.042 57	0.790±0.027	0.294 27
<i>TransMorph</i>	0.659±0.044	0.040 85	0.804±0.024	0.277 08
本文方法	0.672±0.046	0.036 25	0.813±0.025	0.299 88

图 7 和图 8 是将方法作用在膝关节和脑部数据集上的对比图,展示切面为矢状面,图 7 中红色箭头指示的黄色区域所示可以观察到 *SYN* 和 *VoxelMorph* 方法对于胫骨外侧软骨的还原不完整,并在软骨的中间位置还原的厚度比原图像单薄;同时 *CycleMorph* 方法在软骨的边缘部位产生过度的扭曲,导致生成的图像形态特征与参考图像相比有较大差距;*TransMorph* 在软骨的两端较好地还原了形状,但在软骨的中断位置产生过量的扭转偏移致使软骨的整体体积偏大,而本文方法较好地还原了软骨的两端,并在整体形态和大小方面也更符合固定图像.图 7 中 *SYN* 和 *TransMorph* 没有将细节图中白质中圆形结构区域很好地还原为尖锐结构区域, *VoxelMorph* 在一定程度上还原了尖锐部分结构区域,但是对于白质密集区域的还原能力较差, *CycleMorph* 方法对于细节的处理最好,但没有很好地对齐白质的长枝方向,可以观察到本文方法生成的配准图像更加接近参考图像,配准质量相比于其他配准方法精准度更高.

5 结论与展望

本文利用多空间信息提取的技术把多个子网络进行级联,生成医学图像配准模型.每层子网络都含有一个基于 *gConv-Transformer* 的基础配准网络,此网络能够有效理解输入图像的空间信息,并对局部和全局空间特征信息进行建模.同时在模型中加入了分块融合模块,对图像分块并依次进行精细配准,在完成配准后

依次将图像融合得到最后的形变场。本文最后的对比实验结果表明,与目前流行的无监督配准方法相比,本文方法能够提高配准精度,能够获得更准确的配准结

果可以为医生提供更准确的图像数据支持,有助于医生进行更准确的临床诊断,从而帮助医生做出更准确的诊断和治疗决策。

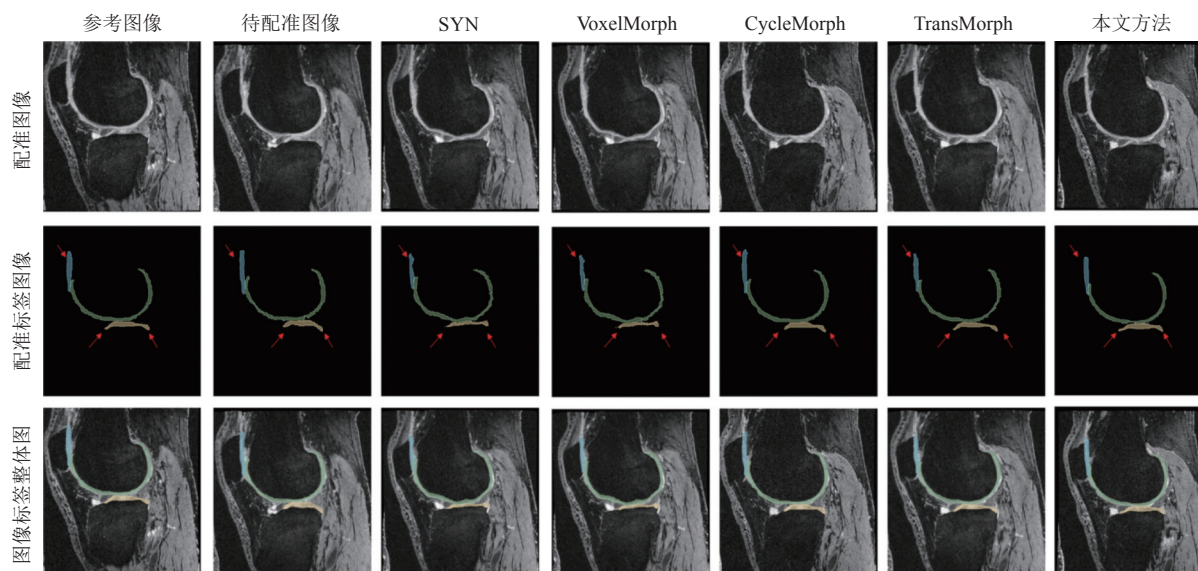


图7 膝软骨配准结果对比

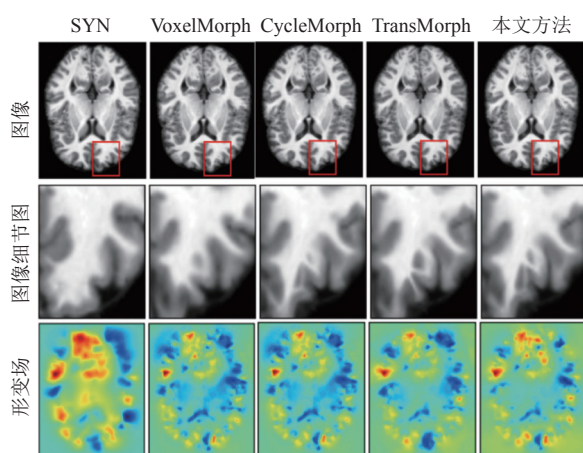


图8 脑部配准结果细节示意图

参考文献

- 1 邹茂扬, 杨昊, 潘光晖. 深度学习在医学图像配准上的研究进展与挑战. 生物医学工程学杂志, 2019, 36(4): 677-683. [doi: 10.7507/1001-5515.201810004]
- 2 潘英杰, 程远志, 刘豪, 等. 基于视觉变换器的级联多阶层医学影像配准方法. 生物医学工程学杂志, 2022, 39(5): 876-886.
- 3 Viola P, Wells WM. Alignment by maximization of mutual information. Proceedings of the 1995 IEEE International

Conference on Computer Vision. Cambridge: IEEE, 1995. 16-23. [doi: 10.1109/ICCV.1995.466930]

4 Vercauteren T, Pennec X, Perchant A, *et al.* Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage, 2009, 45(S1): S61-S72.

5 Wang SQ, Rehman A, Wang Z, *et al.* SSIM-motivated rate-distortion optimization for video coding. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(4): 516-529. [doi: 10.1109/TCSVT.2011.2168269]

6 Zhang J. Inverse-consistent deep networks for unsupervised deformable image registration. arXiv:1809.03443, 2018.

7 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.

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 Zhao SY, Dong Y, Chang E, *et al.* Recursive cascaded networks for unsupervised medical image registration. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 10599-10609.

10 Liu Z, Lin YT, Cao Y, *et al.* Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the

- 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 10012–10022.
- 11 Chen JY, He YF, Frey EC, *et al.* ViT-V-Net: Vision transformer for unsupervised volumetric medical image registration. arXiv:2104.06468, 2021.
 - 12 Rao YM, Zhao WL, Tang YS, *et al.* HorNet: Efficient high-order spatial interactions with recursive gated convolutions. arXiv:2207.14284, 2022.
 - 13 Yang X, Kwitt R, Niethammer M. Fast predictive image registration. Proceedings of the 1st International Workshop on Deep Learning in Medical Image Analysis. Athens: Springer, 2016. 48–57.
 - 14 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.
 - 15 Stauffer C, Grimson WEL. Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747–757. [doi: [10.1109/34.868677](https://doi.org/10.1109/34.868677)]
 - 16 Liu YH, Zuo LR, Han S, *et al.* Coordinate translator for learning deformable medical image registration. Proceedings of the 3rd International Workshop on Multiscale Multimodal Medical Imaging. Singapore: Springer, 2022. 98–109.
 - 17 Wang PSP, Zhang YY. A fast serial and parallel thinning algorithm. IEEE Transactions on Computers, 1989, 38(5): 741–745.
 - 18 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](https://doi.org/10.1016/j.media.2021.102036)]
 - 19 Mok TCW, Chung ACS. Conditional deformable image registration with convolutional neural network. Proceedings of the 24th International Conference on Medical Image Computing and Computer-assisted Intervention. Strasbourg: Springer, 2021. 35–45.
 - 20 Tang K, Li Z, Tian LL, *et al.* ADMIR—Affine and deformable medical image registration for drug-addicted brain images. IEEE Access, 2020, 8: 70960–70968. [doi: [10.1109/ACCESS.2020.2986829](https://doi.org/10.1109/ACCESS.2020.2986829)]
 - 21 Rueckert D, Sonoda LI, Hayes C, *et al.* Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging, 1999, 18(8): 712–721. [doi: [10.1109/42.796284](https://doi.org/10.1109/42.796284)]

(校对责编: 孙君艳)