

基于多尺度感知损失与注意力的脑图像配准^{*}

马添翼 姜大帅 朱 东 张林涛 李国强

(临沂大学信息科学与工程学院 临沂 276000)

摘要: 当前基于深度学习的方法在可变形医学图像对齐任务中得到了广泛的应用,其中利用新颖的损失函数和有效的网络架构提高配准效果是常用的做法。本文设计了多尺度特征感知损失和注意力模块 ECA-D,改进了只用均方误差或归一化互相关的设计思路。本文利用多站点医学图像数据训练获得一个分类神经网络,并构建多尺度特征的学习过程,提高了分类网络的准确性,进而设计多尺度的感知损失函数提高配准的准确性。为了提高对齐网络的学习能力,设计了一种新的注意力模块 ECA-D,更有效地利用空间和通道信息。在 LPBA40 数据集上训练后,与最先进的方法相比,提出的模型在未经训练的 Neurite OASIS 上的 Dice 评分提高了 3%。实验结果表明,本文的方法具有更高的配准精度和更好的鲁棒性。

关键词: 知觉损失;脑 MRI 图像;深度学习;3D 数据

中图分类号: TN391 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Medical image registration based on multiscale feature-perceptual loss and attention mechanism

Ma Tianyi Jiang Dashuai Zhu Dong Zhang Lintao Li Guoqiang

(School of Information Science and Engineering, Linyi University, Linyi 276000, China)

Abstract: Recently, deep learning-based methods have been widely applied in deformable medical image alignment tasks. Among them, utilizing novel loss functions and effective network architectures to improve registration performance is a common approach. This article proposes a multi-scale feature perception loss and attention module ECA-D, which improves the design approach of using only mean square error (MSE) or normalized cross-correlation (NCC). Inspired by the current popular large language model (LLM), this paper trains a classification neural network using multi-site medical image data and constructs a multi-scale feature learning process to improve the accuracy of the classification network. Subsequently, a multi-scale perceptual loss function is designed to enhance the accuracy of registration. To improve the learning ability of the alignment network, a new attention module ECA-D was designed to more effectively utilize spatial and channel information. After training on the LPBA40 dataset, our model showed a 3% improvement in Dice score on the untrained Neurite OASIS compared to the most advanced methods. The experimental results show that compared with other popular registration methods, our method has higher registration accuracy and better robustness.

Keywords: perceptual loss; brain MRI images; deep learning; 3D data

0 引言

可变形图像配准(deformable image registration, DIR)在医学影像分析中起着至关重要的作用,旨在建立固定图像和运动图像之间的空间对应关系。在医学成像中,该技术被广泛用于比较同一患者在不同时间点或不同患者的图像,以进行疾病诊断、治疗计划和结果评估。近年来,基于

深度学习的 VoxelMorph^[1]和 Transformer^[2]配准方法显示出巨大的潜力。卷积神经网络(convolutional neural networks, CNN)被广泛用作医学图像配准领域的主流框架,然而由于卷积操作的局部性,传统的 ConvNet 架构在建模图像中远距离空间关系时存在局限性^[3]。为了解决这一问题,U-Net(或 V-Net)通过引入下采样和上采样操作,在理论上扩大了感受野,能够考虑图像中点之间的远距离

关系。然而,U-Net 仍面临一些问题:首先,前几层的感受野受到卷积核大小限制,只有在较深层才能看到全局信息;其次,随着卷积层加深,远处体素的影响迅速衰减^[4],导致其有效感受野小于理论感受野,限制了其对长距离关系的建模能力,尤其是在医学图像处理中。但是,深度学习方法与传统的图像处理或机器学习方法相比,它们在大脑配准方面仍具有先进的性能。然而,将深度学习算法应用于大脑配准任务仍然存在一些挑战。目前的研究通常在特定的数据集上进行训练并在其中进行测试,而在单个数据集上获得的训练结果尚未在其他数据集上得到改进,配准模型的鲁棒性较差。医学图像配准领域有很多公共数据集,很难实现大规模配准训练的硬件要求。如何高效地利用公共数据集也是一个需要解决的问题。与之前的研究相比,本文的贡献总结如下:

1) 提出一种新的多尺度特征感知损失(multiscale feature-perceptual loss, MFPL),提高了图像配准模型的鲁棒性,并保证了变形场的平滑度。

2) 构建了空间通道注意机制(efficient channel attention-3D deformable convolutions, ECA-D),以帮助基于 CNN 的配准模型更好地捕获远距离空间信息,提高配准精度。

1 相关工作

1.1 图像配准

DIR 是一种通过优化能量函数来建立图像对之间空间对应关系的方法,如式(1)所示。

$$L(I_m, I_f, \phi) = D(I_m \circ \phi, I_f) + \lambda R(\phi) \quad (1)$$

式中: I_m 和 I_f 分别表示运动图像和固定图像。 ϕ 表示将作用于运动图像的变形场,即(i. e., $I_m \circ \phi$)。 $R(\phi)$ 表示变形场正则化项。 λ 是一个超参数,在图像相似性和变形场平滑度之间起着权重作用。DIR 通过最小化能量函数来获得最佳变形场 $\hat{\phi}$ 。

$$\hat{\phi} = \operatorname{argmin} L(I_m, I_f, \phi) \quad (2)$$

在各种能量函数中,相似性度量 $D(I_m \circ \phi, I_f)$ 主要用于评估形变后图像和固定图像之间的对齐程度。除了前面提到的均方误差(mean-square error, MSE)和归一化互相关系数(normalized cross-correlation, NCC)之外,其他常见的相似性度量函数包括结构相似性指数度量(structural similarity, SSIM)^[5]和互信息(mutual information, MI)。

1.2 基于深度神经网络的图像配准

传统的图像配准方法通常使用迭代来最小化每对图像的能量函数(也称为配准误差)。相比之下,基于深度神经网络(deep neural networks, DNN)的方法通过优化能量函数来学习全局表示。这使得基于 DNN 的方法能够对未见过的图像进行配准。VoxelMorph 是基于无监督 DNN 方法的代表,这些方法在图像配准领域取得了极大的成功。

Transformer 的自注意力机制能够捕获远距离的空间信息^[4],因此在医学图像配准领域,一些研究使用了 Transformer 架构来提高配准性能。其中包括 TransMorph^[6]、Swin VoxelMorph^[7] 和 Vit-V-Net^[8],它们通过 Transformer 的自注意力机制促进特征学习。此外,也有一些网络采用交叉注意力来拟合运动图像和固定图像之间的对应关系,如 AttentionReg^[9] 和 Xmorpher^[10]。研究表明,注意力机制^[11]可以提高模型的配准效果。

1.3 转风格损失

在图像转换研究中,研究人员使用监督方法来训练前馈卷积神经网络。他们使用图像中每个像素的损失函数来量化形变后图像和固定图像之间的差异^[12-14]。在 2D 图像转换风格化领域,Mahendran 和 Vedaldi^[15]通过最小化特征重建损失来反转卷积网络的特征,使网络能够理解不同卷积层中的图像特征信息。Dosovitskiy 和 Brox^[16]通过预训练的前馈神经网络来反转卷积特征,使用每像素重建损失进行网络训练。在此领域中,Johnson 等^[17]采用了预训练前馈变换网络来提取高级特征,并构建感知损失进行训练。实验结果表明,感知损失比每像素损失更有效的度量图像之间的相似性。受 Czolbe 等^[18]的启发,基于 Johnson 等^[17]的工作,本文提出了一种多尺度特征感知损失,并用其取代图像配准中常用的损失函数。实验结果表明,提出的方法取得了很好的效果。

2 方 法

2.1 多尺度特征感知损失函数

本文基于多尺度^[19-20]损失函数的思想对感知损失进行调整,并对从图像分类网络不同层获得的特征图应用不同的相似性计算方法。图像分类网络是在多个大脑数据集上训练得到的卷积神经网络,该网络包含 3 个卷积层。在实际操作过程中,本文首先冻结了该网络的参数,以确保在特征提取阶段不会对网络的权重进行更新。接着,本文从分类网络的第一、第二和第三卷积层分别提取输入图像对的浅层、中层和深层特征。浅层特征通常包含图像中的细节信息,如边缘、纹理和局部结构;中层特征捕捉到较为复杂的局部形状和结构特征;而深层特征则能够提取更为高级的语义信息,如物体或区域的整体形态和关联关系。通过这种方式,感知损失可以在不同尺度上充分利用图像中的细节和全局信息,并对不同尺度上的特征信息施加贴合的度量方法,为后续的图像配准提供多层次的特征支持。本文的网络架构如图 1 所示。

$O_i(x)$ 是预训练的医学图像分类网络第 i 层卷积操作中提取得到的特征图,其维度为 $C_i \times H_i \times W_i \times D_i$ 。在浅层卷积操作中,提取得到的特征图通常包含丰富的局部结构信息。为了有效地度量浅层特征之间的相似性,基于这些特征图构建了一种浅层特征感知损失(shallow feature perceptual loss)。该损失函数结合了均方误差(MSE)和

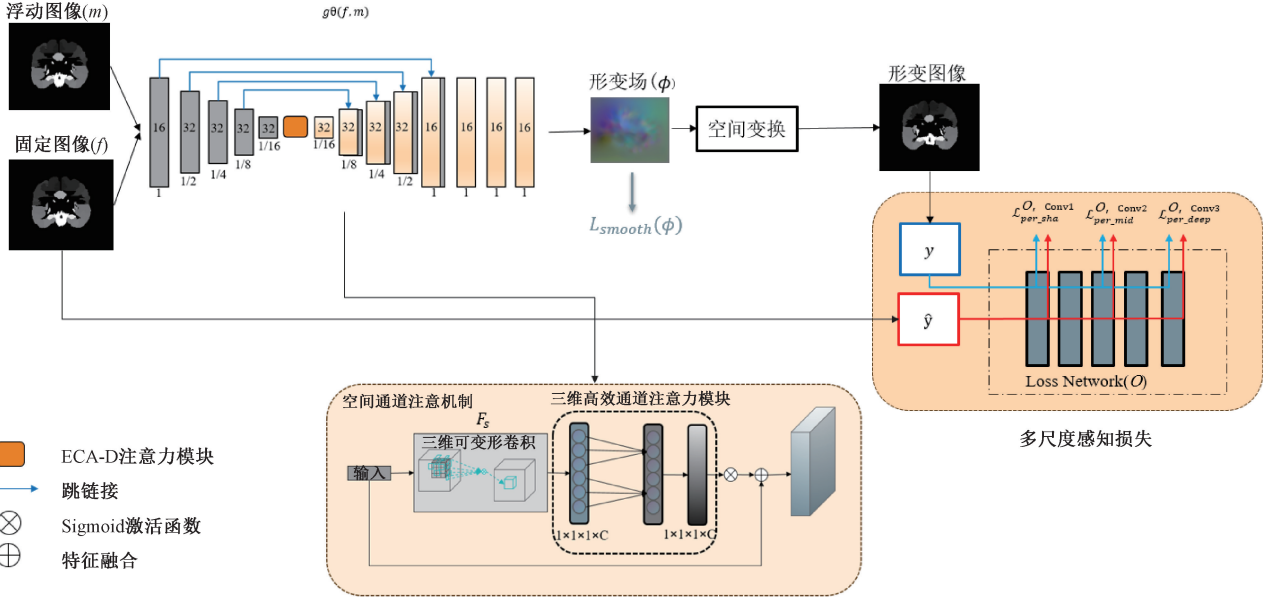


图 1 网络架构

Fig. 1 Network architecture

L_1 损失,以捕捉特征图中局部差异和全局一致性的综合特性,从而提升对浅层特征的感知能力。

$$\begin{aligned} \mathcal{L}_{per_sha}^{O,i}(\hat{y}, y) &= \frac{1}{c_i H_i W_i D_i} \|O_i(\hat{y}) - O_i(y)\|_2^2 + \\ &\alpha \frac{1}{c_i H_i W_i D_i} |O_i(\hat{y}) - O_i(y)| \end{aligned} \quad (3)$$

中间层的特征图蕴含了丰富的语义信息,能够有效捕捉输入数据中的高级语义模式和上下文关系。为了增强网络在处理异常值时的鲁棒性,在中间层特征的损失计算中引入了更稳健的 Charbonnier 损失函数。相比传统的损失函数,Charbonnier 损失通过对误差平方和添加一个较小的平滑项,避免了过度惩罚异常值,从而在保持优化稳定性的同时提高了网络对异常情况的适应能力。

$$\begin{aligned} \mathcal{L}_{per_mid}^{O,i}(\hat{y}, y) &= \frac{1}{C_i H_i W_i D_i} \|O_i(\hat{y}) - O_i(y)\|_2^2 + \\ \beta \ell_{cha}(\hat{y}, y) \end{aligned} \quad (4)$$

Charbonnier 损失的具体公式为^[21]:

$$\ell_{cha}(\hat{y}, y) = \frac{1}{N} \sqrt{\|O_i(\hat{y}) - O_i(y)\|_2^2 + \epsilon^2} \quad (5)$$

式中: ϵ 是一个用于避免梯度消失的值,在本文中设定为 1×10^{-6} 。 N 表示层间特征的总和,用于归一化损失值。深层特征则包含更抽象且更丰富的多尺度信息,能够有效地捕捉输入数据中的全局模式和复杂语义关系。通过计算特征间的欧氏距离来构建深度特征的感知损失,这种方法不仅能够衡量深层特征间的相似性,还可以更全面地评估特征分布在高维空间中的差异,从而提升网络对深层特征的感知能力。

$$\mathcal{L}_{per_deep}^{O,i}(\hat{y}, y) = \frac{1}{C_i H_i W_i D_i} \|O_i(\hat{y}) - O_i(y)\|_2^2 \quad (6)$$

本文通过对来自不同卷积层的特征信息进行个性化定制,成功构建了最终的多尺度特征感知损失。该方法能够有效地指导配准后图像与固定图像之间的感知匹配,帮助优化网络输出的视觉质量。通过结合浅层、中间层和深层特征的优势,该损失在语义和结构层面实现了对真实图像的良好对齐,从而显著提升了扭曲图像的感知质量,使其在视觉效果和语义一致性上更加接近真实图像。

$$\begin{aligned} \mathcal{L}_{per}^{O,i}(\hat{y}, y) &= \mathcal{L}_{per_sha}^{O,i}(\hat{y}, y) + \mathcal{L}_{per_mid}^{O,i}(\hat{y}, y) + \\ &\mathcal{L}_{per_deep}^{O,i}(\hat{y}, y) \end{aligned} \quad (7)$$

式中: O 代表分类网络, i 是卷积层的索引, \hat{y} 是配准后的图像, y 是输入的固定图像, α 和 β 都是设置为 0.01 的超参数。为了生成平滑的形变场并实现有效的网络训练,在形变场 ϕ 的空间梯度上引入了扩散正则化。该正则化项旨在鼓励形变场的平滑性,从而避免过度形变或不连续现象的出现。

$$\mathcal{L}_{smooth}(\phi) = \sum_{p \in \Omega} \|\nabla \phi(p)\|^2 \quad (8)$$

式中: p 表示 3D 图像(Ω)中的体素。

完整的损失函数为:

$$\mathcal{L}(f, m, \phi) = \mathcal{L}_{per}^{O,i}(\hat{y}, y) + \lambda \mathcal{L}_{smooth}(\phi) \quad (9)$$

式中: λ 是设置为 0.01 的归一化参数。

2.2 ECA-D 模块

空间信息包括图像中结构和解剖特征的精确定位。通过保留和最大化这些空间信息,配准算法可以有效地实现结构对齐。医学图像通常包括多个通道,代表不同的信息水平。通过充分整合通道信息,配准算法可以更全面地考虑图像的多方面特征。ECA-D 模块结合了三维高效通道

注意力模块 (ECA) 和三维可变形卷积 (3D deformable convolutions, D), 有效提升了网络对长距离空间信息的捕捉能力。ECA 模块通过全局池化操作生成通道级注意力, 能够自适应地加强对关键特征的关注, 提升模型对长距离空间依赖的感知能力。与此同时, 三维可变形卷积通过动态调整卷积核的感受野, 在不同位置自适应地捕捉图像中的远距离空间关系, 增强了局部与全局信息的融合^[22]。这一结合使得 ECA-D 模块能够高效地在三维医学图像中建模复杂的空间依赖关系, 提升了配准精度。具体结构如图 2 所示。

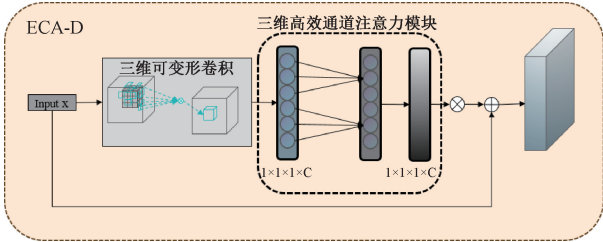


图 2 ECA-D 模块架构
Fig. 2 ECA-D module architecture

首先, 通过三维可变形卷积 (3D deformable convolutions, D) 提取空间信息, 得到空间特征图 F_s 。

$$F_s = \sum_{p_n \in R} w_n x(p_0 + p_n + \Delta p_n) \tag{10}$$

式中: R 是常规卷积的采样网格。对于输入特征图 p_n 上的每个位置 x 和采样网格中的位置 p_0 , 通常使用具有相同大小内核的额外卷积层预测偏移量 $\Delta p_n, n = 1, \dots, |R|$ ^[23]。然后, 可变形卷积的采样网格会根据预测的偏移量进行扩展。随后, 获取的空间特征图 F_s 将在通道注意力模块中进行处理, 从而得到具有空间信息的通道特征。

$$\omega = \sigma(C1D_k(F_s)) \tag{11}$$

式中: “C1D” 表示一维卷积, 仅包含 K 个参数信息, σ 表示 Sigmoid^[24] 函数。最后, 空间通道特征图 ω 与原始输入 x 组合以获得最终输出。

$$x = x + \omega \tag{12}$$

ECA-D 模块能够同时捕捉空间和通道特征, 同时保留输入图像中的原始信息, 这对变形场的生成具有重要意义。通过三维可变形卷积提取空间信息, ECA-D 模块可以适应图像中复杂的结构变化, 捕捉局部区域的精细特征。同时, 利用高效通道注意力机制, 该模块能够动态调整不同通道的权重, 强化对关键信息的表达。通过结合空间和通道特征的优势, ECA-D 模块在生成变形场时能够更全面地考虑输入图像的多尺度特性和多模态信息, 从而生成更加精准且平滑的变形场, 显著提升配准效果。

3 实 验

本文使用公开的实验数据集 LPBA40 和 Neurite-

OASIS 进行实验。在 LPBA40 数据集中, 每个 MRI 容积包含 54 个标记的感兴趣区; 而在 Neurite-OASIS 数据集中, 每个 MRI 容积包含 35 个标记的感兴趣区。实验中未使用数据增强, 仅对 Neurite-OASIS 数据集进行了重新采样, 以节省训练资源。LPBA40 数据集被分为 30 次扫描和 9 次扫描用于训练和测试, 同时根据 MICCAI Learn2Reg 配准挑战赛从 Neurite-OASIS 选取了 20 副图用于测试。将图像大小裁剪为 $160 \times 192 \times 160$, 批次 (batch-size) 设为 1, 并仅在 LPBA40 上使用随机配对进行训练, 未直接在 Neurite-OASIS 数据集上进行训练。最后, 用两个数据集验证了训练结果。

实验采用平均 Dice 分数作为主要的评价指标, 将在 LPBA40 上得到最好 Dice 的训练结果直接用 Neurite-OASIS 进行测试, 以验证模型的鲁棒性。此外, 计算了形变场的 Jacobian 标准偏差 ($std(|J_\phi|)$), 以量化变形场的差异同质性和平滑度, 其值越小代表生成的形变场越平滑。最后, 计算了 SSIM 与 HD95 (95% Hausdorff 距离)。SSIM 的范围为 $[0, 1]$, 其值越大, 表示配准的效果越好。当两幅图像一模一样时, 此时 $SSIM = 1$ 。HD95 是一种用于评估图像分割精度的评价指标, 主要用于衡量预测结果与真实结果之间的边界相似度, 其数值越小, 表示配准效果越好。

模型是在配备 24 G 显存的 NVIDIA RTX A5000 上训练的, 使用 PyTorch 实现, 并使用 Adam 优化器进行了优化。目前流行的几种配准方法进行了比较, 包括 SyN、VIT-V-Net、VoxelMorph、CycleMorph^[25] 和 Deformer^[26]。应用于 LPBA40 和 Neurite OASIS 数据集的各种方法的数值结果如表 1 和 2 所示, 其中 SyN 的结果来自 Deformer。在 LPBA40 中, 本文的方法取得了最高的 Dice, 比 Deformer 提高了 1%。

表 1 LPBA40 数据集的结果

Table 1 Numerical results on the LPBA40 dataset				
算法	Dice	$Std(J_\phi)$	SSIM	HD95
Deformer	66.7*	0.073	0.794	4.336
SyN	66.5*	0.126	—	—
VIT-V-Net	66.3**	0.262	0.782	4.341
CycleMorph	64.8**	0.067	0.793	4.422
VoxelMorph	62.6**	0.039	0.785	4.259
本研究	67.4	0.281	0.787	4.105

注: “*” 表示算法之间存在显著差异, $P < 0.05$, “**” 表示算法与 $P < 0.001$ 存在显著差异。

加入 ECA-D 模块后, 在 SSIM 指标上与以 Transformer 为主要框架的模型表现相同, 甚至更高, 说明该模块增强了网络捕捉远距离空间信息的能力。此外, 结合 ECA-D 模块与多尺度特征感知损失后, 提出的模型获得了最低的 HD95 值, 表明这种组合能够提升图像配准质量,

并且增强图像边界的配准精度。在 Neurite-OASIS 数据集中,本文的方法还取得了最高的 Dice 系数,比 Vit-V-Net 提高了 3%,同时保持了较高的 SSIM 和几乎最低的 HD95,进一步验证了该方法的有效性和更好的鲁棒性。使用 T-test 来评估组间差异。P 值小于 0.05 被认为具有显著性,P 值小于 0.001 表示高度显著差异,差异越高代表方法越具有有效性。

为了更好地定量比较不同方法,为两个数据集的每个标注区域生成了 Dice 系数的箱形图。由于 LPBA40 数据集的标注区域较多,将图分成了两部分,分别展示在图 3~5 中。从图中可以看出提出的方法在绝大多数标签上 Dice 系数波动更小,在其他算法出较大 Dice 波动标签,本文的方法都有出色的配准表现。综合来看,所提的算法相较于其他方法几乎在所有解剖结构(包括复杂细微区域)的配准上都表现优异。

表 2 未经直接训练的 Neurite OASIS 数据集上不同方法的结果

Table 2 Numerical results of different methods on the Neurite OASIS dataset without direct training				
算法	Dice	$Std(J_g)$	SSIM	HD95
Deformer	65.0**	0.136	0.787	2.611
VIT-V-Net	67.4**	0.301	0.784	2.566
CycleMorph	53.9**	0.111	0.743	3.126
VoxelMorph	62.6**	0.019	0.744	3.007
本研究	69.6	0.132	0.775	2.647

注:“*”表示算法之间存在显著差异, $P<0.05$,”**”表示算法与 $P<0.001$ 存在显著差异。

为了验证所提模块的有效性,在两个数据集上进行了消融实验,结果如表 3 和 4 所示。从结果中可以看出,包含所有模块的模型在 Dice 系数上表现最佳。在多尺度特征感知损失函数中加入一定比例的 L_1 Loss 有助于提高 Dice 系数,尽管会导致折叠率的增加,但 ECA-D 模块能够有效抑制这一负面影响,从而提高 Dice 系数。最后,将提出的损失分别与 MSE 和 NCC 进行组合训练。结果如表 5 和 6 所示,从表中可以明显看出,单独使用多尺度特征感知损失能够获得更好的配准结果,进一步验证了所提出方法的有效性。

表 3 和 4 显示了所提出的模块的消融实验。表 5、6 显示了所提出的方法与 NCC、MSE 相结合的结果。

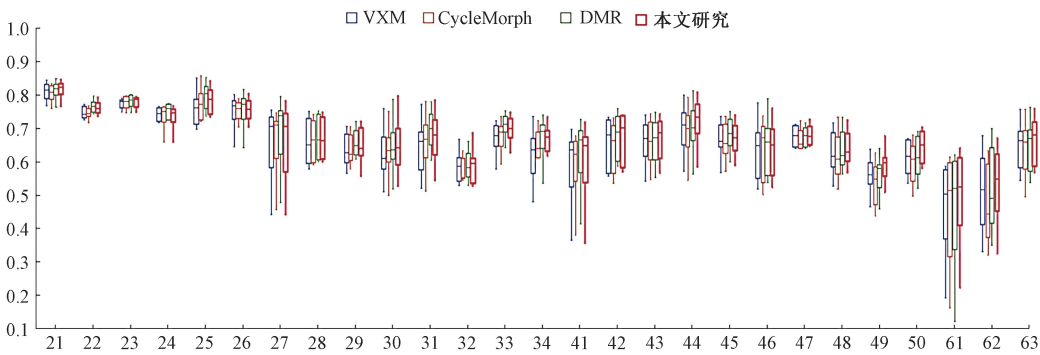


图 3 LPBA40 数据集中前 27 个标记区域的结果
Fig. 3 Results for the first 27 labeled regions in the LPBA40 dataset

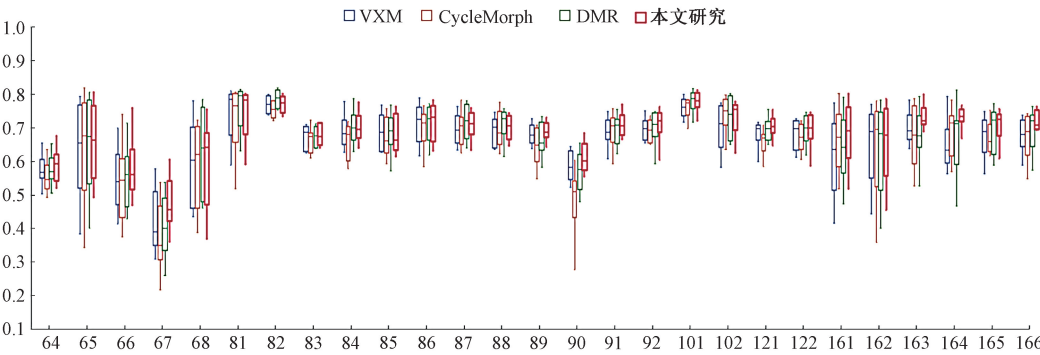


图 4 LPBA40 数据集后 27 个标记区域的结果
Fig. 4 Results for the last 27 labeled regions in the LPBA40 dataset

图 6 和 7 则展示了不同方法在两个数据集上得到的配准后图像。本文的方法生成的配准图像更为精确,能够更

好地保留图像的内部结构和边界信息,更好实现复杂小区域的配准。同时,本文还制作了所提方法在两个数据集上

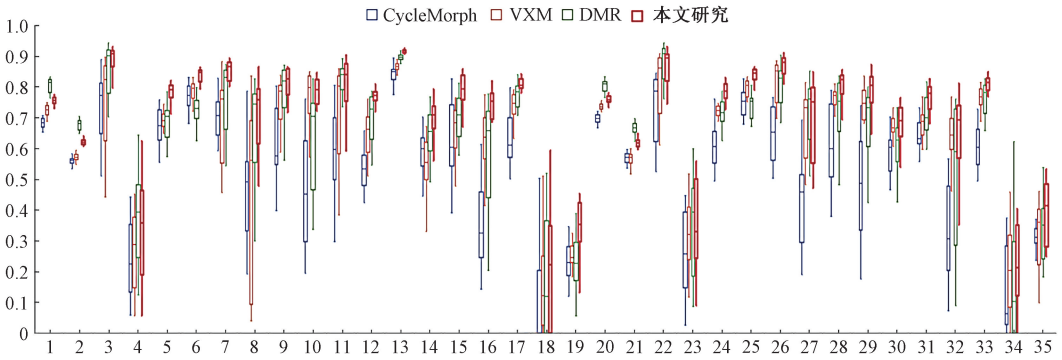


图 5 Neurite OASIS 数据集 35 个标记区域的结果

Fig. 5 Results for the 35 labeled regions in the Neurite OASIS

的误差热图,如图 8 所示。图中的颜色深浅反映了配准误差的大小,颜色越深,表明误差越小。从热图中可以清晰地看出,所提方法在绝大多数区域实现了较低的配准误差,尤其在复杂小区域的配准任务中,证明了该方法能较好地处理细节和结构复杂区域。这一结果进一步验证了方法在提升配准精度和稳定性方面的有效性。

表 3 LPBA40 数据集的结果

Table 3 Results of the LPBA40 dataset					
序号	ECA-D	MFPL	L_1 Loss	Dice	Std ($ J_\# $)
1	✓	✓	✓	67.4	0.281
2	✓	✓		67.1	0.090
3		✓		67.0	0.292

表 4 Neurite OASIS 数据集的结果

Table 4 Results of the Neurite OASIS dataset					
序号	ECA-D	MFPL	L_1 Loss	Dice	Std ($ J_\# $)
1	✓	✓	✓	69.6	0.132
2	✓	✓		68.8	0.080
3		✓		66.5	0.323

表 5 LPBA40 数据集的结果

Table 5 Results of the LPBA40 dataset					
序号	本研究	NCC	MSE	Dice	Std ($ J_\# $)
1	✓			67.4	0.281
2	✓	✓		64.6	0.507
3	✓		✓	65.2	0.189

表 6 Neurite OASIS 数据集的结果

Table 6 Results on the Neurite OASIS dataset					
序号	本研究	NCC	MSE	Dice	Std ($ J_\# $)
1	✓			69.6	0.132
2	✓	✓		68.0	0.597
3	✓		✓	66.3	0.284

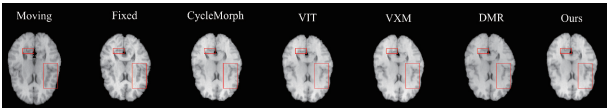


图 6 不同方法在 LPBA40 数据集上的可视化结果

Fig. 6 Visualization results of different methods on LPBA40 dataset

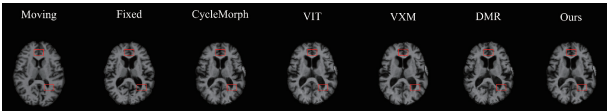


图 7 Neurite OASIS 数据集上不同方法的可视化结果

Fig. 7 Visualization results of different methods on Neurite OASIS dataset

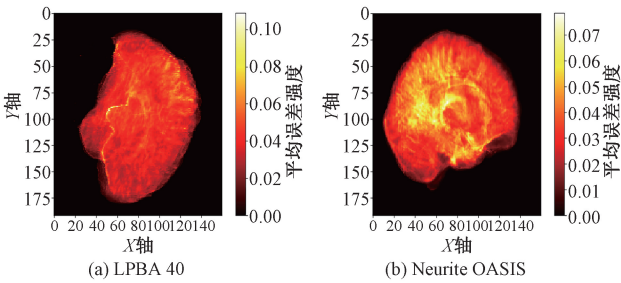


图 8 误差热图

Fig. 8 Error heatmap

4 结 论

本文通过引入多尺度损失理念,构建了多尺度特征感知损失,取代了依赖于每像素相似度测量的传统损失函数(如均方误差 MSE 和归一化互相关 NCC)。此外,通过设计新的注意力模块,进一步增强了基于卷积神经网络的配准网络捕获长距离特征的能力。在脑部 MR 数据集 LPBA40 和 Neurite-OASIS 上的实验表明,与其他基于深度学习的配准方法相比,该模型在提升配准精度方面具有显著优势,尤其在未经训练的 Neurite-OASIS 数据集上表现尤为出色。这些改进显著增强了算法的配准性能和鲁棒性。

参考文献

- [1] BALAKRISHNAN G, ZHAO A, SABUNCU M R, et al. An unsupervised learning model for deformable medical image registration[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 9252-9260.
- [2] KAISER L, GOMEZ A N, SHAZEER N, et al. One model to learn them all[J]. ArXiv preprint arXiv: 1706.05137, 2017.
- [3] LUO W J, LI Y J, URTASUN R, et al. Understanding the effective receptive field in deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2016, 29: 4898-4906.
- [4] LI SH H, SUI X C, LUO X D, et al. Medical image segmentation using squeeze-and-expansion transformers [J]. ArXiv preprint arXiv: 2105.09511, 2021.
- [5] CHEN J Y, LI Y, DU Y, et al. Generating anthropomorphic phantoms using fully unsupervised deformable image registration with convolutional neural networks[J]. Medical Physics, 2020, 47(12): 6366-6380.
- [6] CHEN J Y, FREY E C, HE Y F, et al. Transmorph: Transformer for unsupervised medical image registration[J]. Medical Image Analysis, 2022, 82: 102615.
- [7] ZHU Y P, LU SH. Swin-voxiormorph: A symmetric unsupervised learning model for deformable medical image registration using swin transformer [C]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022: 78-87.
- [8] CHEN J Y, HE Y F, FREY E C, et al. Vit-v-net: Vision transformer for unsupervised volumetric medical image registration[J]. ArXiv preprint arXiv: 2104.06468, 2021.
- [9] SONG X R, GUO H T, XU X ANG, et al. Cross-modal attention for MRI and ultrasound volume registration [C]. Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, 2021: 66-75.
- [10] SHI J CH, HE Y T, KONG Y Y, et al. Xmorpher: Full transformer for deformable medical image registration via cross attention [C]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022: 217-226.
- [11] 单慧琳, 吕宗奎, 付相为, 等. 基于DFECANet的遥感图像飞机目标检测方法[J]. 电子测量与仪器学报, 2024, 38(2): 19-29.
SHAN H L, LYU Z K, FU X W, et al. Aircraft target detection method in remote sensing images based on DFECANet [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38 (2): 19-29.
- [12] DONG CH, LOY CH CH, HE K M, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2): 295-307.
- [13] CHENG Z ZH, YANG Q X, SHENG B. Deep colorization [C]. IEEE International Conference on Computer Vision, 2015: 415-423.
- [14] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [15] MAHENDRAN A, VEDALDI A. Understanding deep image representations by inverting them [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5188-5196.
- [16] DOSOVITSKIY A, BROX T. Inverting visual representations with convolutional networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4829-4837.
- [17] JOHNSON J, ALAHI A, LI F F. Perceptual losses for real-time style transfer and super-resolution[C]. Computer Vision-ECCV 2016: 14th European Conference, 2016: 694-711.
- [18] CZOLBE S, KRAUSE O, FERAGEN A. DeepSim: Semantic similarity metrics for learned image registration[J]. ArXiv preprint arXiv:2011.05735, 2020.
- [19] 张娜, 吴信元, 赵强, 等. 基于多尺度特征的路面不平度识别方法[J]. 电子测量技术, 2024, 47(12): 28-36.
ZHANG N, WU X Y, ZHAO Q, et al. A road surface roughness recognition method based on multi-scale features [J]. Electronic Measurement Technology, 2024, 47(12): 28-36.
- [20] 刘罡, 侯恩翔, 黄孙港, 等. 多尺度特征融合增强的行人翻越护栏检测[J]. 电子测量技术, 2024, 47(14): 127-138.
LIU G, HOU EN X, HUANG S G, et al. Multi-scale feature fusion enhanced pedestrian barrier-crossing detection[J]. Electronic Measurement Technology, 2024, 47(14): 127-138.
- [21] LAI W S, HUANG J B, AHUJA N, et al. Fast and

accurate image super-resolution with deep laplacian pyramid networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(11): 2599-2613.

[22] 王琦,张涛,徐超炜,等. 多尺度注意力融合与视觉 Transformer 方法优化的电阻抗层析成像深度学习方
法[J]. 仪器仪表学报,2024,45(7):52-63.
WANG Q, ZHANG T, XU CH W, et al. Multi-scale
attention fusion and vision Transformer method-
optimized deep learning approach for electrical
impedance tomography [J]. Chinese Journal of
Scientific Instrument, 2024, 45(7): 52-63.

[23] POMINOVA M, KONDRATEVA E, SHARAEV
M, et al. 3D deformable convolutions for MRI
classification [C]. 2019 18th IEEE International
Conference on Machine Learning and Applications
(ICMLA). IEEE, 2019: 1710-1716.

[24] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient
channel attention for deep convolutional neural
networks[C]. IEEE/CVF Conference on Computer
Vision and Pattern Recognition, 2020: 11534-11542.

[25] KIM B, KIM D H, PARK S H, et al. CycleMorph:
Cycle consistent unsupervised deformable image
registration [J]. Medical Image Analysis, 2021,
71: 102036.

[26] CHEN J SH, LU D H, ZHANG Y, et al. Deformer:
Towards displacement field learning for unsupervised
medical image registration [C]. International
Conference on Medical Image Computing and
Computer-Assisted Intervention. Cham: Springer
Nature Switzerland, 2022: 141-151.

作者简介

马添翼, 硕士研究生, 主要研究方向为医学图像配准。
E-mail: 1149602575@qq.com

姜大帅, 硕士研究生, 主要研究方向为医学图像配准。

朱东, 硕士研究生, 主要研究方向为医学图像分割。

张林涛, 博士, 副教授, 主要研究方向为人工智能、医学
图像处理。
E-mail: zhanglintao@lyu.edu.cn

李国强(通信作者), 博士, 副教授, 主要研究方向为人工
智能、医学图像处理。
E-mail: liguoqiang@lyu.edu.cn