

基于多模态融合的脊柱图像分割方法

代广鑫¹, 王辉², 王连雷², 刘新宇², 张梦华¹, 黄伟杰¹

(1. 济南大学自动化与电气工程学院, 山东 济南 250024; 2. 山东大学齐鲁医院脊柱外科, 山东 济南 250012)

摘要: **目的** 结合脊柱 CT 和 MR 多模态医疗图像的互补信息, 综合利用骨骼和软组织的详细特征, 改善识别的准确性, 提高脊柱医疗图像的分割精度, 进而提供更全面的脊柱病变评估。 **方法** 构建一个多模态医疗图像融合网络模型和一个半监督分割网络模型, 分别用于脊柱 CT 和 MR 图像的融合以及基于融合图像的分割任务。多模态融合网络通过共享编码器保留不同模态的共同特征, 基础编码器提取全局特征, 细节编码器专注于局部细节。半监督分割网络模型采用双子网络架构, 并引入对比差异评审模块和动态竞争伪标签生成模块来纠正和约束网络训练。 **结果** 多模态融合网络在图像信息保留和特征保持方面表现优异, 融合图像的高频信息噪声更少。半监督分割网络在 Dice 系数和 Jaccard 系数上均表现优异, 改善了脊柱软组织与骨组织之间的清晰度。 **结论** 多模态医疗图像融合网络和半监督分割网络有效地提升了脊柱图像的融合和分割精度。通过对比差异评审和动态竞争伪标签生成模块的引入, 进一步提高分割结果的准确性, 为脊柱疾病的评估提供更加清晰和可靠的图像信息。

关键词: 多模态; 图像融合; 图像分割; 半监督; 脊柱

中图分类号: R445.6

文献标志码: A

Spinal images segmentation method based on multimodal fusion

DAI Guangxin¹, WANG Hui², WANG Lianlei², LIU Xinyu², ZHANG Menghua¹, HUANG Weijie¹

(1. School of Electrical Engineering, University of Jinan, Jinan 250024, Shandong, China;

2. Department of Spinal Surgery, Qilu Hospital of Shandong University, Jinan 250012, Shandong, China)

Abstract: **Objective** By combining the complementary information from spinal CT and MR multimodal medical images, and utilizing detailed features of both bone and soft tissues, to improve the accuracy of identification and enhance the segmentation precision of spinal medical images by soft tissues, thereby providing a more comprehensive assessment of spinal lesions. **Methods** This paper proposed a multimodal medical image fusion network model for the fusion of spinal CT and MR images and a semi-supervised segmentation network model for the segmentation tasks based on the fused images. The multimodal fusion network retained the shared features of different modalities through a shared encoder, with a basic part extracting global features and a detail part focusing on local details. A dual-network architecture was employed to the segmentation network, which was corrected and constrained by a contrastive difference review module and a dynamic competitive pseudo-label generation module when the network was training. **Results** The proposed fusion network performed well in preserving image information and features, with less high-frequency noise in the fused images. The semi-supervised segmentation network excelled in both the Dice coefficient and Jaccard index, improving the clarity between spinal soft tissues and bone tissues. **Conclusion** The proposed multimodal medical image fusion network and semi-supervised segmentation network effectively enhance the fusion and segmentation accuracy of spinal images. The introduction of the contrastive difference review and dynamic competitive pseudo-label generation modules further improved the accuracy of the segmentation results, providing clearer and more reliable image information for the assessment of spinal diseases.

Key words: Multimodal; Image fusion; Image segmentation; Semi-supervised; Spine

收稿日期: 2024-07-29

基金项目: 山东省重点研发计划(重大科技创新工程)(2022CXGC010503)

通信作者: 黄伟杰。E-mail: cse_huangwj@ujn.edu.cn

刘新宇。E-mail: newyuliu@163.com

脊柱疾病在各年龄段人群中的发病率较高,近年来其发病年龄呈现出逐渐降低的趋势。以往,腰椎间盘突出、颈椎病和脊柱侧弯等脊柱疾病主要发生在老年人群中,但近年来,青少年和中青年人群的发病率显著上升^[1]。现代社会生活方式的变化,如长时间的不良坐姿、缺乏运动以及过度依赖电子设备,导致脊柱疾病在年轻人群中愈加普遍^[2]。随着脊柱疾病的发病率不断上升,医学图像处理技术在脊柱疾病的诊断和治疗中变得尤为重要。医学影像帮助医生及时发现病变区域并评估疾病严重程度。医学图像处理不仅能够辅助早期诊断,还为个性化治疗方案的制定提供科学依据,特别是在脊柱疾病的手术规划和随访中具有重要价值^[3]。

脊柱图像分割在医学图像处理领域中占据重要地位,特别是在脊柱疾病的诊断和治疗计划中发挥着关键作用^[4-5]。脊柱由多个椎骨、椎间盘、韧带和周围软组织组成,其形态和位置在不同个体中存在较大差异^[6]。通过脊柱图像分割,医生能够直观地观察脊柱的解剖学和病理学变化,包括椎骨的形态异常、椎间盘的退变状态以及周围软组织的病理改变等,有助于医生评估病变的范围和性质,从而为诊断、治疗和随访提供依据^[7-8]。然而,脊柱的复杂解剖结构以及病变的多样性,常常使得单一模态的影像无法满足临床诊断和治疗的需求^[9]。计算机断层扫描(computer tomography, CT)图像能够详细显示脊柱骨骼结构,但难以区分脊柱周围的软组织结构和病变,因此需要磁共振(magnetic resonance, MR)图像来清晰显示脊柱周围的软组织^[10]。通过结合CT和MR影像,提供更全面和准确的影像信息,有助于提高脊柱疾病的诊断和治疗效果^[11]。本研究旨在提出一种基于多模态融合的脊柱图像分割方法,利用神经网络模型有效提取和融合脊柱的骨骼和软组织特征,并采用半监督学习方法,引入对比差异评审和动态伪标签生成模块,以增强分割精度。

1 资料与方法

1.1 数据集

数据来源于2021年12月-2023年6月在山东大学齐鲁医院脊柱外科就诊的47例患者。本研究获得山东大学齐鲁医院科研伦理委员会批准[伦理号:KYLL-2021(KS)-055],所有患者均签署知情同意书。多模态融合网络采用CT和MR的配对数据集,使用3D Slicer^[12]中的Landmark模块对同一患者的脊柱CT和MR进行手动配准。将患者的CT

和MR在世界坐标系中居中,使用transforms模块将二者固定,随后使用landmark模块在椎体上标定7~8组固定点,使得椎体位置对齐,配对数据集一共完成924组,部分数据集见图1。



图1 部分脊柱MR和CT的配对数据集

Figure 1 Partial paired dataset of spinal MR and CT

脊柱图像分割数据集是在脊柱CT和MR融合后使用3D Slicer进行标注。调整图像的窗宽窗位,以便更清晰分辨脊柱和周围组织的区别,使用标注工具进行勾画制作脊柱标签,流程见图2。

1.2 方法

1.2.1 网络模型概述

本研究提出两个网络模型利用CT和MR的优势特征,得到更全面和准确的脊柱多模态融合图像,并根据融合图像对脊柱进行半监督的分割,通过对比差异评审模块和动态竞争伪标签生成模块提高分割的精度,网络模型的整体结构,见图3。

在多模态医疗图像融合网络中引入一种双分支编码器结构,以提取和融合脊柱CT和MR图像的全局及局部特征。该结构分为高频信息提取和低频信息提取两个部分。在高频信息提取阶段,利用共享编码器提取CT和MR的共同特征,防止信息丢失。在低频信息提取阶段,分别采用基础编码器和细节编码器对CT和MR进行特征提取,以减少特征之间的干扰,确保信息完整性。

在半监督的互校正分割网络中,本研究提出一种对比差异评审模块,通过比较两个子网络的分割标签差异并施加约束,使差异趋近于零,从而引导子

网络关注并纠正潜在偏见,提高伪标签的质量。此外,本研究设计动态竞争伪标签生成模块,以减少学

生网络对教师网络的负面约束,优化网络对脊柱标签的学习过程。

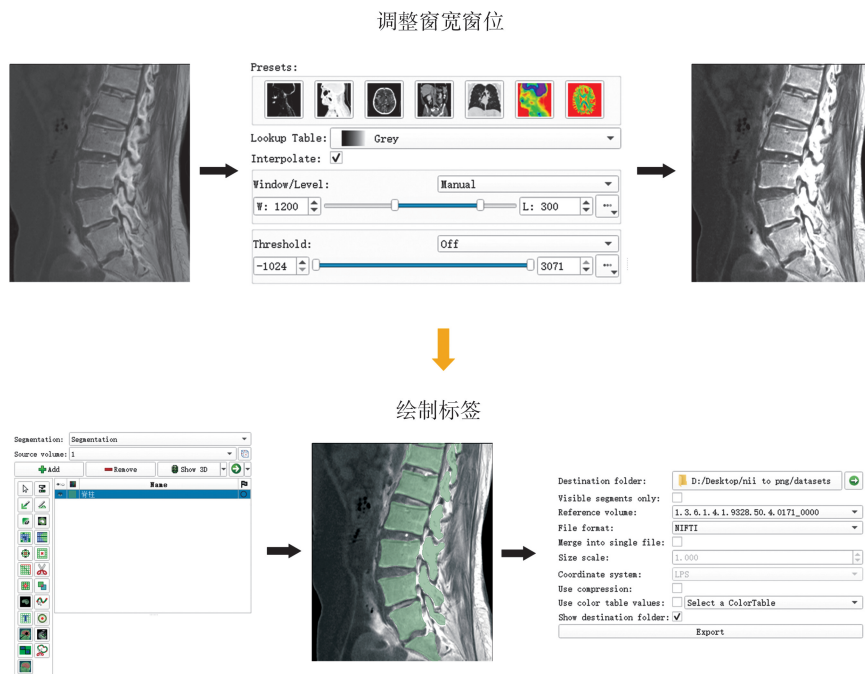
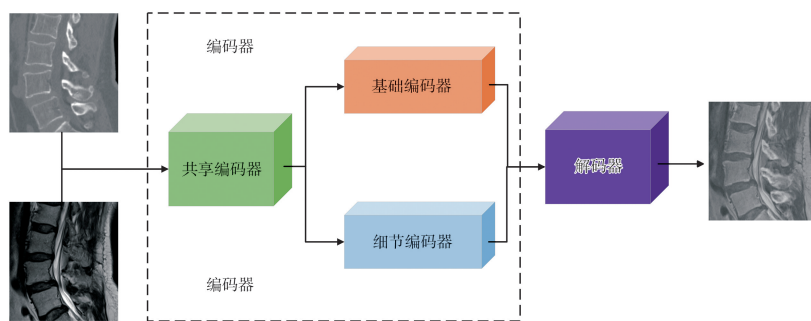
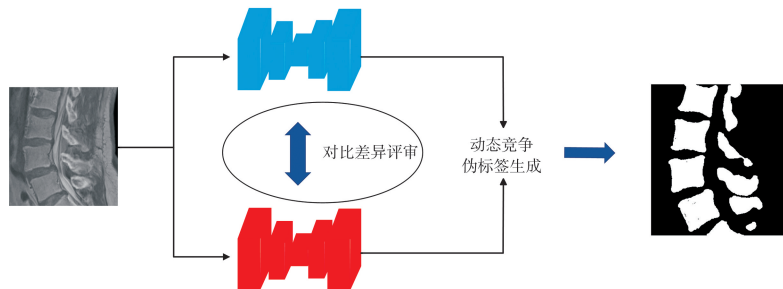


图2 脊柱图像分割数据集的标注
Figure 2 Labeling of spinal segmentation dataset



多模态医疗图像的融合算法



互校正的半监督分割算法

图3 网络模型的整体结构
Figure 3 The comprehensive architecture of the network model

1.2.2 多模态医疗图像融合网络

多模态医疗图像融合网络的核心结构包括3部

分:①提取 CT 和 MR 共同特征的共享编码器;②提取 CT 和 MR 各自特征的双分支编码器(包括基础

编码器和细节编码器);③提取的特征进行融合的解码器,网络结构见图4。

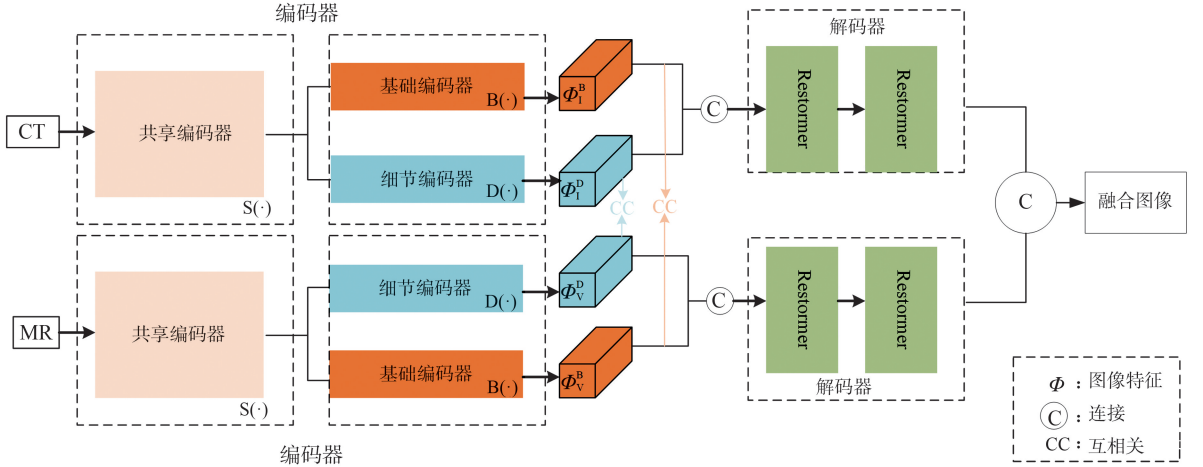


图4 多模态医疗图像的融合网络结构图

Figure 4 The fusion network structure diagram of multimodal medical images

本研究的编码器部分由共享编码器和双分支编码器组成。共享编码器的设计目的是保留不同模态图像中的共同特征信息,以便后续能够提取CT和MR各自特有的特征信息。在共享编码器中,选择Restormer块^[13],因为Restormer能够在特征维度上从高分辨率输入图像中提取全局特征。通过引入多尺度注意力机制,Restormer可以在特征提取过程中有效捕捉不同尺度的信息,增强模型对图像细节和整体结构的理解,提高特征的表达能力。此外,Restormer能够有效捕获特征之间的全局关联性和上下文信息,减少特征之间的冗余信息,因此它可以在不显著增加计算量的情况下提取跨模态的浅层特征。

在双分支编码器的选择上,采用基础编码器和细节编码器相配合。共享编码器提取共同特征后,基础编码器和细节编码器分别用于CT和MR图像的特征提取。基础编码器主要提取图像的全局特征,而细节编码器则专注于提取图像的局部细节信息。这种分支结构的设计有助于保留每种模态图像的特有信息,从而在后续的融合过程中能够更好地综合利用这些信息。

为了提取多模态图像的长距离依赖特征,本研究使用Lite-Transformer(LT)块^[14]作为基础编码器的基本单元。通过平坦前馈网络的结构,LT块缩小嵌入,减少参数数量,同时保持相同的性能。这样设计的双分支编码器能够有效提取和融合脊柱CT和MR图像的全局和局部特征,提高特征提取和融合的精度与效率。

LT块包含两个主要分支:自注意力机制分支和卷积分支。自注意力机制分支用于捕捉序列中的长

距离依赖关系。通过计算序列中各个位置之间的注意力权重,模型能够有效地理解不同位置之间的关联性。在此过程中,模型根据学习到的查询、键、值的映射关系,计算每个位置与其他位置之间的注意力权重,并将值与注意力权重进行加权求和,从而生成每个位置的表示。这种机制使得模型能够更深入地理解序列的整体结构和语义信息。

卷积分支则通过一系列卷积核对输入序列进行卷积操作,提取序列中每个位置的局部特征表示。随后,采用池化等操作对特征进行压缩和整合,生成更高级别的特征表示。卷积操作能够有效地捕捉序列中的局部结构和细节信息,并通过参数共享和局部感受野的特性,在捕获局部特征时具备显著优势。

基础编码器从共享特征中提取低频特征,

$$\Phi_{CT}^B = B(\Phi_{CT}^S), \quad \Phi_{MR}^B = B(\Phi_{MR}^S), \quad (1)$$

其中, $B(\cdot)$ 是基础编码器, Φ^B 是图像低频特征, Φ^S 是图像共享特征。

细节编码器采用卷积神经网络(convolutional neural networks, CNN)结构,通过卷积操作有效地提取输入序列中的局部特征和细节信息。具体而言,卷积核在输入序列上执行滑动操作,对每个位置的局部感受野进行特征提取。这种方式使得细节编码器能够捕捉序列中的局部结构,并对基础编码器提取的全局特征进行补充,从而增强模型对局部细节的表达能力。

细节编码器主要用于从共享特征中提取高频细节信息。鉴于边缘和纹理等细节特征在图像融合任务中的重要性,细节编码器采用CNN结构,以最大限度地保留细节信息。该结构能够有效捕

提图像中的细微特征,确保在融合过程中保留尽可能多的高频细节信息,从而提升图像融合的精度和质量,

$$\Phi_{CT}^D = D(\Phi_{CT}^S), \Phi_{MR}^D = D(\Phi_{MR}^S), \quad (2)$$

其中, $D(\cdot)$ 是细节编码器, Φ^D 是图像高频特征。

在共享编码器和双分支编码器提取特征后,解码器负责将两个分支提取的特征进行融合,并生成最终的融合特征表示。这些融合特征表示将用于后续图像分割任务。解码器的设计旨在有效地融合

来自不同编码器的特征信息,充分利用 CT 和 MR 成像技术的优势,提供更丰富、更全面的信息,使融合图像既拥有软组织信息,又保留完整的骨组织信息,从而为后续的分割任务提供支持。这些特征的融合将显著提高图像分割的精度和准确性。

1.2.3 互校正的半监督分割网络

为了提高半监督分割算法的精度,本研究提出了一种拥有双子网络结构的互校正半监督分割网络,其网络结构见图 5。

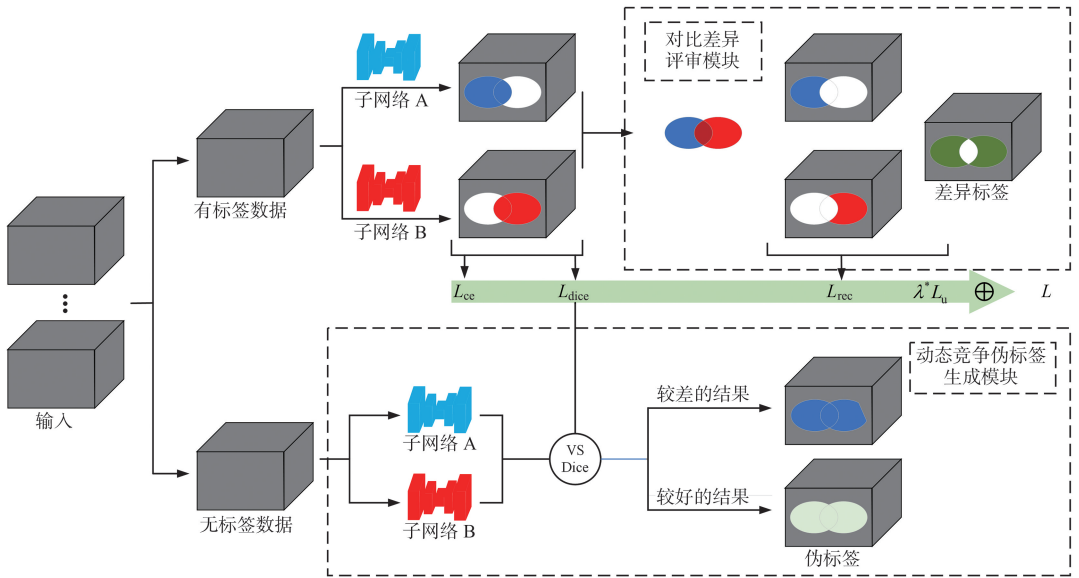


图 5 互校正的半监督分割网络结构

Figure 5 The structure of a semi-supervised segmentation network with mutual correction

两个子网络分别为 UNet^[15] 和 VNet^[16]。针对这两个网络,提出一种对比差异评审模块,用于识别和评估预测结果中的不一致区域。该模块通过对有标签数据和无标签数据同时输入两个子网络,分别生成有标签的预测结果和无标签的预测结果。损失函数由有监督损失函数和无监督损失函数组成,其中有监督损失函数(包含 L_{ce} 、 L_{dice} 、 L_{rec})通过对比差异评审模块来纠正潜在的预测误差,而无监督损失函数(L_u)则用于指导无标签数据的伪标签生成。不同结构的子网络和独立的参数更新策略引入更多的灵活性,充分发挥双子网络结构的潜力。

对比差异评审模块的设计灵感源于一个基本事实:在二元分类任务中,如果两个分类器对同一体素给出不同的预测结果,则至少有一个预测必然存在误差。因此,将这些预测不一致的区域视为潜在的预测误差区域,并使网络对这些区域进行审查,以提高预测的准确性。预测失误区域掩码可以通过对两个子网络的 softmax 输出进行二值化的异或操作来

生成,其具体形式为

$$M_{diff} = O_1(\hat{Y}_{UNet}^L) \oplus O_1(\hat{Y}_{VNet}^L), \quad (3)$$

其中, \hat{Y} 是有标签的数据, \oplus 表示异或操作, $O_1(\cdot)$ 表示二值化操作。

接着,可以利用掩码来识别潜在的错误预测区域,其过程为

$$\hat{Y}_{diff}^L = O_2(M_{diff}, \hat{Y}^L), \quad (4)$$

其中, $O_2(\cdot)$ 表示执行与掩码区域相对应的预测提取操作。

本研究设计一种校正损失函数,用于指导网络对潜在的预测失误区域进行审查。该校正损失函数的具体形式为

$$L_{rec} = O_3(\hat{Y}_{diff}^L, \hat{Y}_{diff}^L), \quad (5)$$

其中, $O_3(\cdot)$ 表示均方误差损失函数, \hat{Y}_{diff}^L 是与这些区域相对应的真实标签。

动态竞争伪标签生成模块旨在实时评估两个子网络的分割性能,并根据评估结果选择表现更优的子网络作为另一个网络的伪标签生成器。在该设计中,首先使用 Dice 损失来衡量两个子网络在图像分

割任务中的性能。Dice 损失作为一种评估分割准确度的指标,能够直接反映预测结果的相似性。接下来,通过熵最小化方法处理这些预测结果,将其转换为软伪标签,从而增强标签的信息含量。随后,应用锐化函数对软伪标签进行进一步处理,以提升标签的质量。最后,再次使用 Dice 损失来评估处理后的伪标签,选择表现最优的预测结果作为最终的伪标签。

表 1 实验超参数设置
Table 1 Setting of experimental hyperparameters

实验名称	batch_size	初始学习率	Epoch	权重衰减
融合实验	16	0.000 1	100	0.025
分割实验	4	0.01	200	0.000 1

本研究对于图像融合实验采用熵 (entropy, EN)、互信息 (mutual information, MI)、空间频率 (spatial frequency, SF) 和视觉保真度 (visual information fidelity, VIF) 4 个评价标准进行验证^[17-20]。采用 Dice 系数和 Jaccard 系数来作为评价指标对本研究的实验结果进行评估^[21-22]。

有效性验证评价指标计算如下:

(1) EN 是一种常用的评估图像质量和信息量的指标。EN 在信息论中用于衡量信源的不确定性或信息量。在图像处理中, EN 被广泛应用于图像质量评价和图像融合等任务中, 以量化图像的信息丰富程度和信息分布均匀性, 其定义为

$$EN = - \sum_{i=1}^N p(i) \log_2 p(i), \quad (6)$$

其中, N 代表图像的灰度级数, $p(i)$ 是图像中灰度级 i 出现的概率。

当图像的灰度级分布较为均匀时, EN 的值会较高; 当图像的灰度级分布不均匀, 即某些灰度级出现的概率较高, 而另一些灰度级出现的概率较低时, EN 的值会较低。通过计算图像的 EN, 可以评估图像的信息量和信息分布的均匀性。在图像融合中, 通常会 will 多幅图像融合成一幅新的图像, EN 用来评估融合后图像的质量和是否得到了有效保留和利用。

(2) MI 是一种重要的评估图像融合效果的指标。MI 用于度量两幅图像之间的信息相关性和相互依赖程度, 进而评估图像融合的质量和效果, 其数学定义为

$$MI(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right), \quad (7)$$

1.3 实验设置和评价指标

本研究的融合实验使用两张 GPU NVIDIA GeForceRTx 3090 进行训练, 半监督的分割实验则是使用四张 GPU NVIDIA GeforceRTx 3090 进行训练。此外, 在训练之前先将脊柱的 CT 和 MR 图像填充至 512×512 的大小, 实验结束之后, 输出的大小会重新回到输入图像的大小, 以避免得到的结果与事实不符。融合实验和分割实验的部分超参数见表 1。

其中, X 和 Y 分别表示 CT 和 MR, $p(x, y)$ 是图像对中灰度值为 x 和 y 的像素同时出现的概率, $p(x)$ 和 $p(y)$ 分别是图像 X 和 Y 中灰度值为 x 和 y 的像素出现的概率。

MI 的值越大, 表示两幅图像之间的相关性越高, 信息量的重叠程度也越大。在图像融合任务中, 通过计算融合后图像与原始图像之间的 MI, 可以评估融合后图像的信息丢失程度和信息重叠情况。

(3) SF 是用于描述图像中变化率的指标, 常被用作图像融合的评估标准之一。SF 反映了图像中灰度值随着空间位置的变化而变化的速率。在二维图像中, SF 可以通过傅立叶变换来计算。具体而言, 二维离散傅立叶变换 (discrete Fourier transform, DFT) 可以将图像转换到频域中, 从而得到图像在频率域上的表示。在频率域中, 图像的 SF 可以通过计算图像的幅度谱来获取, 其数学表达为

$$SF(u, v) = \sqrt{u^2 + v^2}, \quad (8)$$

其中, u 和 v 是频域中的水平和垂直方向的频率。

在图像融合中, SF 常被用来评估融合后图像的细节信息和结构保留情况。通常情况下, 较高的 SF 对应于图像中的细节和纹理, 而较低的 SF 则对应于图像中的整体结构和背景信息。因此, 通过分析融合后图像在不同 SF 上的分布情况, 可以评估融合算法对图像细节和结构的处理效果。

(4) VIF 是一种用于评估图像质量的指标, 特别适用于图像融合的评估。VIF 是一种全局性的评价指标, 它综合考虑了图像在空间和频率域的结构信息, 能够有效地评估图像融合后的视觉保真度, VIF 的计算基于两幅图像之间的视觉差异, 其数学表达为

$$VIF = \frac{\sum_{k=1}^K \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}}{K}, \quad (9)$$

其中, K 是图像块的数量, μ_x 和 μ_y 分别是 CT 和 MR 的均值, σ_x^2 和 σ_y^2 分别是 CT 和 MR 的方差, σ_{xy} 是它们的协方差, C_1 和 C_2 是为了稳定性而引入的常数。

VIF 的计算过程主要包括两个步骤: 首先, 计算图像的均值、方差和协方差; 然后, 利用这些统计量计算视觉保真度。VIF 越高, 表示两幅图像之间的

相似度越高, 即融合后的图像质量越好。

2 结果

2.1 实验结果

在融合实验中, 本研究使用了 740 组 CT 和 MR 数据进行训练。为了直观地展示所提出网络的优越性, 本研究选用了 4 例患者的 CT 和 MR 数据进行融合实验, 实验结果见图 6。

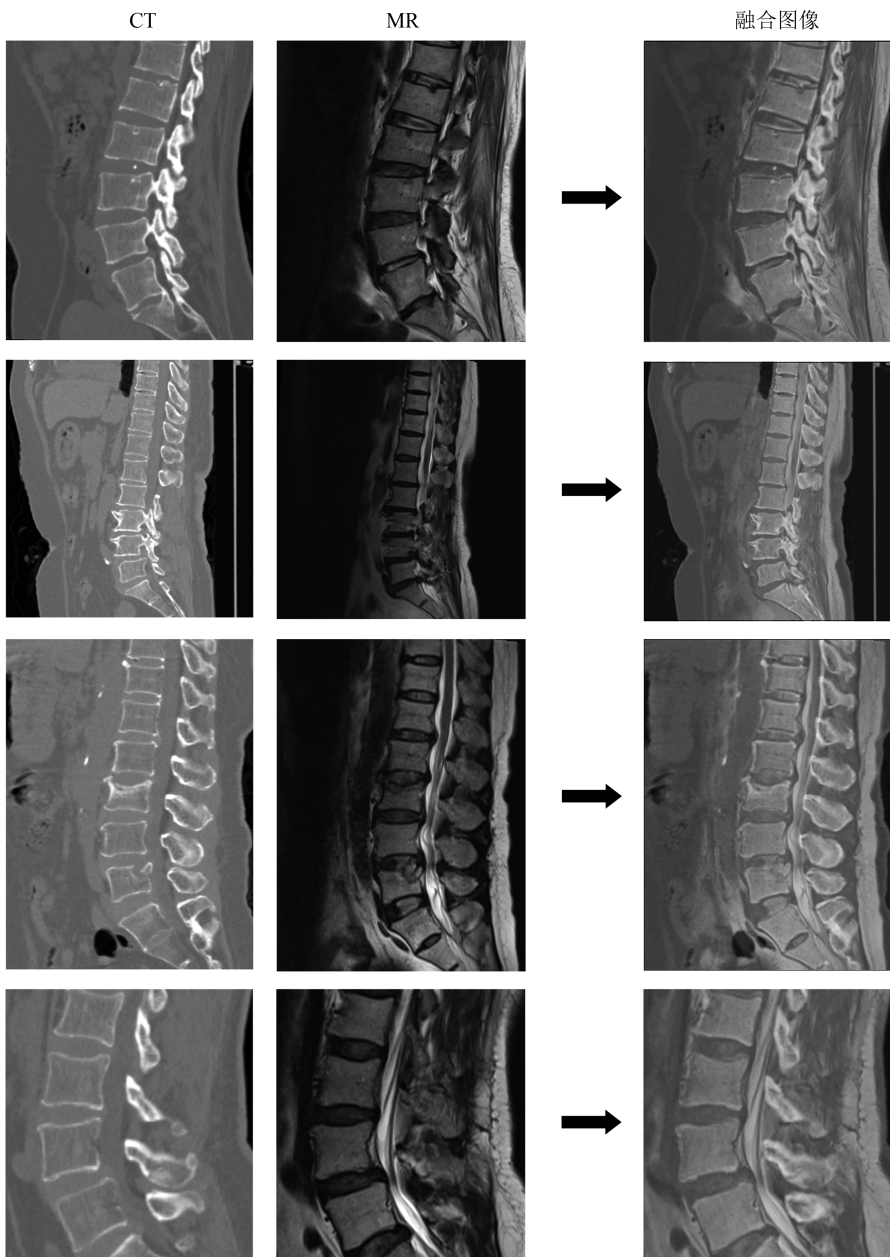


图 6 多模态医疗图像的融合结果

Figure 6 The fusion result of multimodal medical images

实验结果表明, 融合图像能够清晰地显示椎体和棘突等骨组织结构, 同时呈现椎间盘等软组织结构形态和信号强度。

在分割实验中, 为了展示所提出的半监督分割算法在少量标签数据集情况下的优越性, 本研究使用相同的少量标签数据集对 VNet 和 UNet 进行了

有监督训练,并将其结果与本研究提出的半监督分割网络的训练结果进行了对比,结果见图7。

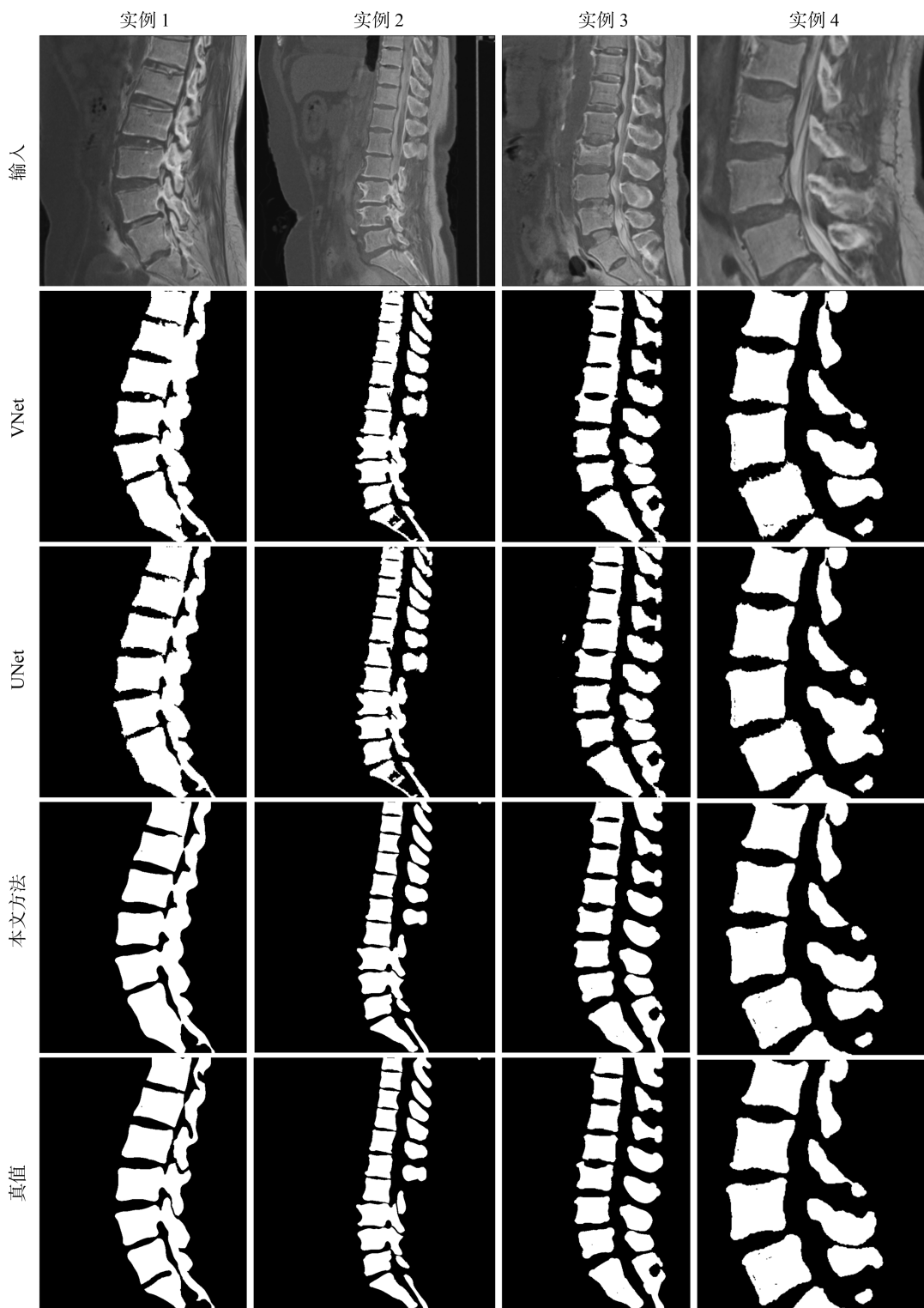


图7 有监督的脊柱图像分割对比实验

Figure 7 Supervised spinal segmentation comparative experiment

由于融合后的脊柱医疗图像结构更加复杂,且高频信息(骨组织部分)被弱化,脊柱低频信息的干扰增大。然而,本研究提出的网络在骨组织的分割结果上与真值接近,其边缘较VNet和UNet的实验结果更加平滑,且椎体间的边界更

加清晰。

此外,本研究使用相同的数据集对目前主流的三种半监督分割网络,UA-MT^[23]、SASSNet^[24]和Min-Max Similarity^[25]进行训练,并将其结果与本研究提出的网络进行对比,结果见图8。

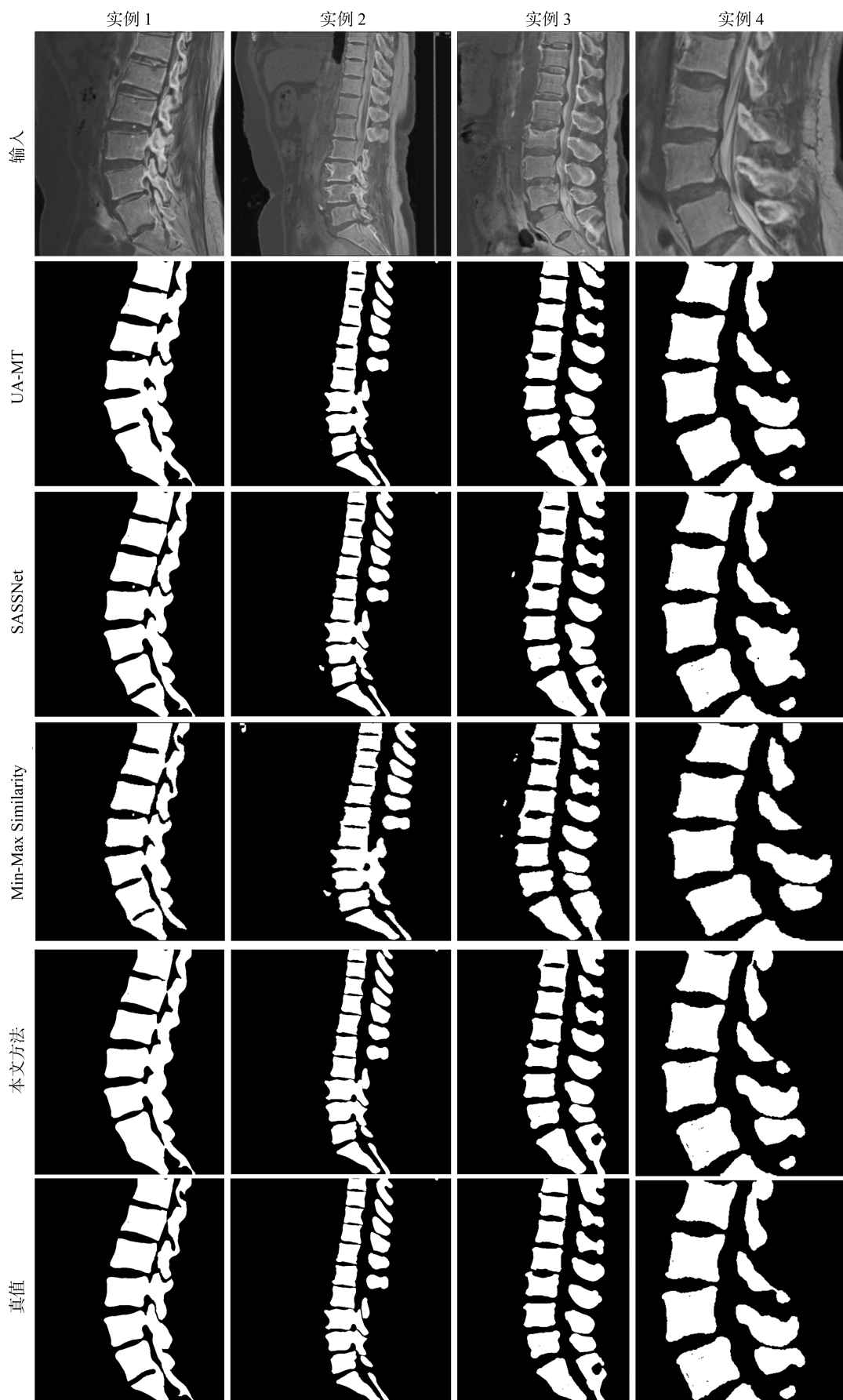


图 8 半监督的脊柱图像分割对比实验

Figure 8 Semi-supervised spinal segmentation comparative experiment

从图 8 中可以清晰地看出,与现有的两种半监督分割网络相比,本研究提出的半监督分割网络在实验结果中,除了椎体和棘突以外,其他骨组织或类似骨组织的高频信息噪声更少。

2.2 有效性验证

本研究分别对融合实验结果和分割实验结果使用评估指标进行了验证,并与对比实验的结果一同进行评估,以验证本研究所提出两个网络的有效性和优越性。在融合实验中,采用在图像融合领域表

表 2 图像融合的评价结果

Table 2 The evaluation results of image fusion

方法	EN	MI	SF	VIF
U2F	4.83	2.14	28.14	0.38
本文方法	5.02	2.60	33.83	0.51

分割实验中,使用 Dice 系数和 Jaccard 系数对本研究所提出的半监督分割网络,以及 5 个网络对 15 组数据推理的结果进行评估,结果见表 3。

表 3 脊柱图像分割的评价结果

Table 3 Evaluation results of spinal segmentation

方法	Dice/%	Jaccard/%
VNet	64.18±0.07	49.26±0.08
UNet	66.53±0.02	51.25±0.07
UA-MA	74.43±0.03	60.21±0.03
SASSNet	73.57±0.02	59.19±0.03
Min-Max Similarity	74.92±0.01	61.19±0.04
本文方法	75.21±0.02	61.53±0.03
<i>F</i>	566708.395	27424.589
<i>P</i>	<0.001	<0.001

从两个评价指标上定量比较显示,在少量标签数据集的情况下,半监督策略带来的增益明显。其中,本研究提出的半监督分割网络在 Dice 系数和 Jaccard 系数上均表现最佳($P<0.001$)。表明对比差异评审模块和动态竞争伪标签生成模块在网络训练中的纠正和约束作用是有效的。

3 讨论

脊柱图像分割任务中,脊柱骨组织和软组织的准确分割一直是医学影像分析中的难点。脊柱 CT 与 MR 图像的不同模态特性决定了它们各自在结构和软组织分割中的优势,传统的单模态方法往往无法兼顾两者的特点,导致脊柱软组织和骨组织边界模糊,尤其是细小结构(如棘突)的分割精度较低。单一模态脊柱图像的分割方法在软组织和细小结构分割上的准确性普遍较低,尤其是在高噪声、低分辨率图像的情况下,分割精度更是受到极大影响。因

现优异的 U2F^[26] 网络进行对比,评估结果见表 2。

从 4 个评价指标来看,本研究提出的多模态医疗图像融合网络的实验结果表现出色。EN 和 MI 的优异表现表明,该网络在图像信息保留方面具有重要作用;SF 反映网络在特征保持方面的有效性;而 VIF 则展示网络在图像质量融合结果上的优越性。这些指标均体现网络在 CT 和 MR 特征提取及特征保留方面的可靠性,验证多种编码器设计的有效性。

此,结合多模态医学影像的优势,能够有效弥补单一模态方法的局限性。多模态医学图像融合通过利用不同模态的优点,可以得到更全面和详细的脊柱影像^[27-28]。另外,CT 和 MR 各自在影像中可能出现的伪影和噪声,通过融合处理可以相互补偿,减少伪影干扰,提高图像质量^[29-31]。融合后的图像包含了更丰富的解剖信息,使得分割算法能够更准确地区分不同组织,提高分割的精确度^[32]。

在多模态医学图像分割领域中,Das 等^[33]提出一种名为 RIMNet 的新型深度学习模型,通过融合多模态特征,显著提升分割精度,并在 MICCAI 2018 挑战数据集上展示其处理复杂医学图像任务的潜力。然而,该模型在应对不同模态图像时可能存在对特定模态的依赖性,从而限制泛化能力,尤其在多样化的临床场景中表现出鲁棒性不足的问题。为进一步提升分割性能,Pang 等^[34]提出 SpineParseNet,这是一种结合 3D 图卷积网络(GCSN)与 2D 残差 U-Net(ResUNet)优点的两阶段框架。该模型通过区域池化和区域反池化模块增强图像与图表示之间的可解释性,同时有效捕捉脊柱结构间的空间依赖关系,从而显著提高分割精度。然而,SpineParseNet 在处理细小脊柱结构(如棘突)及图像顶部区域时存在一定的局限性。

为了结合脊柱 CT 与 MR 图像的优势特征,本研究提出一种基于多模态融合的脊柱图像分割方法,实现精准的图像融合和分割任务。在设计上,本研究通过共享编码器的方式,保留两种不同模态图像的共同特征,同时引入细节编码器来专注于提取局部细节信息,从而增强分割模型对脊柱细节和骨组织边界的敏感性。通过多模态图像融合,能够清晰地显示椎体和棘突等骨组织结构,同时呈现椎间

盘等软组织结构的形态和信号强度。椎间盘作为连接相邻椎体的重要软组织结构,对脊柱运动和冲击吸收起着关键作用。通过将MR图像的软组织信息与CT图像的骨组织信息进行融合,医生能够在同一图像中全面了解椎间盘与周围骨结构之间的关系。这为评估椎间盘退变、脊柱畸形等疾病提供了更准确的诊断依据。利用CT图像提供的脊柱骨结构的清晰信息和MR图像提供的软组织细节,解决脊柱骨组织和软组织之间界限模糊的问题,显著提高分割精度。此外,某些患者存在血管硬化等疾病,这些因素会干扰网络对骨组织信息的学习,尤其是在半监督框架下,教师网络会受到学生网络的干扰,容易导致分割结果偏差,对棘突组织的学习带来影响,导致其分割不够精细。本研究提出的半监督分割网络有效地避免此类错误,为脊柱疾病的诊断与治疗提供更加精确的图像信息支持。

本研究存在一些局限性:①脊柱疾病的诊断和治疗往往依赖于精确的影像数据,而脊柱图像的复杂性和不同模态影像的差异性可能导致分割精度不稳定;②在实际临床环境中,CT与MR图像可能来自不同设备和医院,导致影像质量和分辨率的不一致。

为了在更广泛的临床应用中推广本方法,未来的研究需要进一步扩展数据集,尤其是在多中心数据融合和跨域适应性方面。不同医院的影像数据可能存在不同的设备偏差、采集协议和图像质量,这可能影响模型的泛化能力。未来可以探索自适应网络架构,以增强模型在异构医疗影像中的表现,进一步提升分割精度和鲁棒性。此外,随着医学影像技术的不断发展,新的影像类型(如PET、超声图像等)将逐步成为医学诊断的重要手段。因此,未来的研究应致力于如何将不同类型的医学影像结合起来,提升跨模态数据融合的效果,推动多模态影像技术在医学影像学中的广泛应用。

综上所述,本研究提出的基于多模态融合的脊柱图像分割方法,通过共享及细节编码器,增强对脊柱细节和边界的敏感性,清晰呈现骨与软组织结构,为疾病诊断提供准确依据,显著提升分割精度。同时,有效避免因血管硬化等因素干扰造成的分割偏差,为脊柱疾病诊断与治疗提供精确图像信息支持。

参考文献:

- [1] Li Y, Zheng S, Wu YX, et al. Trends of surgical treatment for spinal degenerative disease in China: a cohort of 37, 897 inpatients from 2003 to 2016 [J]. *Clin Interv Aging*, 2019, 14: 361-366. doi:10.2147/CIA.S191449
- [2] Salsali M, Sheikhhoseini R, Sayyadi P, et al. Association between physical activity and body posture: a systematic review and meta-analysis [J]. *BMC Public Health*, 2023, 23(1): 1670. doi:10.1186/s12889-023-16617-4
- [3] Samartzis D, Borthakur A, Belfer I, et al. Novel diagnostic and prognostic methods for disc degeneration and low back pain [J]. *Spine J*, 2015, 15(9): 1919-1932.
- [4] Lenchik L, Heacock L, Weaver AA, et al. Automated segmentation of tissues using CT and MRI: a systematic review [J]. *Acad Radiol*, 2019, 26(12): 1695-1706.
- [5] Martín-Noguerol T, Oñate Miranda M, Amrhein TJ, et al. The role of artificial intelligence in the assessment of the spine and spinal cord [J]. *Eur J Radiol*, 2023, 161: 110726. doi:10.1016/j.ejrad.2023.110726
- [6] Qu B, Cao JP, Qian C, et al. Current development and prospects of deep learning in spine image analysis: a literature review [J]. *Quant Imaging Med Surg*, 2022, 12(6): 3454-3479.
- [7] 马信龙, 马剑雄, 杜育任, 等. 医工结合引领骨科“精”“智”发展之路 [J]. *交通医学*, 2019, 33(6): 545-547. MA Xinlong, MA Jianxiong, DU Yuren, et al. The combination of medicine and work leads the development of orthopedics “essence” and “intelligence” [J]. *Medical Journal of Communications*, 2019, 33(6): 545-547.
- [8] Lee S, Jung JY, Mahatthanatrakul A, et al. Artificial intelligence in spinal imaging and patient care: a review of recent advances [J]. *Neurospine*, 2024, 21(2): 474-486.
- [9] Azam MA, Khan KB, Salahuddin S, et al. A review on multimodal medical image fusion: compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics [J]. *Comput Biol Med*, 2022, 144: 105253. doi:10.1016/j.compbiomed.2022.105253
- [10] 任博文, 韩振川, 吴剑慧, 等. 腰椎椎间融合影像学评价方法的研究进展 [J]. *解放军医学院学报*, 2022, 43(8): 901-906. REN Bowen, HAN Zhenchuan, WU Jianhui, et al. Research advances in radiographic evaluation methods for lumbar intervertebral fusion [J]. *Academic Journal of Chinese PLA Medical School*, 2022, 43(8): 901-906.
- [11] Zhang YD, Dong ZC, Wang SH, et al. Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation [J]. *Inf Fusion*, 2020, 64: 149-187. doi:10.1016/j.inffus.2020.07.006
- [12] Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D slicer as an image computing platform for the Quantitative Imaging Network [J]. *Magn Reson Imaging*, 2012, 30(9): 1323-1341.
- [13] Zamir SW, Arora A, Khan S, et al. Restormer: efficient

- transformer for high-resolution image restoration [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 5718-5729. doi: 10.1109/CVPR52688.2022.00564
- [14] Fournier Q, Caron GM, Aloise D. A practical survey on faster and lighter transformers [J]. *ACM Comput Surv*, 2023, 55(14s): 1-40.
- [15] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]//Medical image computing and computer-assisted intervention-MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015: 234-241.
- [16] Abdollahi A, Pradhan B, Alamri A. VNet: an end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data [J]. *IEEE Access*, 2020, 8: 179424-179436. doi: 10.1109/ACCESS.2020.3026658
- [17] Guo P, Xie GQ, Li RF, et al. Multimodal medical image fusion with convolution sparse representation and mutual information correlation in NSST domain [J]. *Complex Intell Syst*, 2023, 9(1): 317-328.
- [18] Tang L, Tian CG, Li LD, et al. Perceptual quality assessment for multimodal medical image fusion [J]. *Signal Process Image Commun*, 2020, 85: 115852. doi: 10.1016/j.image.2020.115852
- [19] Huang B, Yang F, Yin MX, et al. A review of multimodal medical image fusion techniques [J]. *Comput Math Methods Med*, 2020, 8279342. doi: 10.1155/2020/8279342
- [20] Lévêque L, Outtas M, Liu HT, et al. Comparative study of the methodologies used for subjective medical image quality assessment [J]. *Phys Med Biol*, 2021, 66(15). doi: 10.1088/1361-6560/ac1157
- [21] Masood RF, Ahmad Taj I, Khan MB, et al. Deep learning based vertebral body segmentation with extraction of spinal measurements and disorder disease classification [J]. *Biomed Signal Process Contr*, 2022, 71: 103230. doi: 10.1016/j.bspc.2021.103230
- [22] Liang YW, Fang YT, Lin TC, et al. The quantitative evaluation of automatic segmentation in lumbar magnetic resonance images [J]. *Neurospine*, 2024, 21(2): 665-675.
- [23] Yu LQ, Wang SJ, Li XM, et al. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation [C]//Medical image computing and computer assisted intervention-MICCAI 2019: 22nd international conference, Shenzhen, China, October 13-17, 2019, proceedings, part II 22. Springer International Publishing, 2019: 605-613.
- [24] Li SL, Zhang CY, He XM. Shape-aware semi-supervised 3D semantic segmentation for medical images [C]//Medical image computing and computer assisted intervention-MICCAI 2020: 23rd international conference, Lima, Peru, October 4-8, 2020, proceedings, part I 23. Springer International Publishing, 2020: 552-561.
- [25] Lou AG, Tawfik K, Yao X, et al. Min-max similarity: a contrastive semi-supervised deep learning network for surgical tools segmentation [J]. *IEEE Trans Med Imaging*, 2023, 42(10): 2832-2841.
- [26] Xu H, Ma JY, Jiang JJ, et al. U2Fusion: a unified unsupervised image fusion network [J]. *IEEE Trans Pattern Anal Mach Intell*, 2022, 44(1): 502-518.
- [27] Tan W, Tiwari P, Pandey HM, et al. Multimodal medical image fusion algorithm in the era of big data [J]. *Neural Comput Appl*, 2025, 37. doi: 10.1007/s00521-020-05173-2
- [28] Tran VL, Lin HY, Liu HW. Multitask deep learning for segmentation and lumbosacral spine inspection [J]. *IEEE Trans Instrum Meas*, 2022, 71: 4005910. doi: 10.1109/TIM.2022.3184341
- [29] Fu J, Li WS, Du J, et al. Multimodal medical image fusion via Laplacian pyramid and convolutional neural network reconstruction with local gradient energy strategy [J]. *Comput Biol Med*, 2020, 126: 104048. doi: 10.1016/j.compbimed.2020.104048
- [30] Muhammad G, Alshehri F, Karray F, et al. A comprehensive survey on multimodal medical signals fusion for smart healthcare systems [J]. *Inf Fusion*, 2021, 76: 355-375. doi: 10.1016/j.inffus.2021.06.007
- [31] Dimitri GM, Spasov S, Duggento A, et al. Multimodal and multicontrast image fusion via deep generative models [J]. *Inf Fusion*, 2022, 88: 146-160. doi: 10.1016/j.inffus.2022.07.017
- [32] Zhou SK, Greenspan H, Davatzikos C, et al. A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises [J]. *Proc IEEE Inst Electr Electron Eng*, 2021, 109(5): 820-838.
- [33] Das P, Pal C, Acharyya A, et al. Deep neural network for automated simultaneous intervertebral disc (IVDs) identification and segmentation of multi-modal MR images [J]. *Comput Methods Programs Biomed*, 2021, 205: 106074. doi: 10.1016/j.cmpb.2021.106074
- [34] Pang SM, Pang CL, Zhao L, et al. SpineParseNet: spine parsing for volumetric MR image by a two-stage segmentation framework with semantic image representation [J]. *IEEE Trans Med Imaging*, 2021, 40(1): 262-273.