

# 结构先验引导的多模态腰椎 MRI 图像分割算法

李伟豪<sup>1,2,3</sup>, 王苹苹<sup>1,2,3</sup>, 许万博<sup>1,2,3,4</sup>, 魏本征<sup>1,2,3\*</sup>

(1. 山东中医药大学医学人工智能研究中心, 山东 青岛 266112; 2. 山东中医药大学青岛中医药科学院, 山东 青岛 266112; 3. 青岛市中医人工智能技术重点实验室, 山东 青岛 266112; 4. 山东大学齐鲁医院德州医院, 山东 德州 253046)

**摘要:**为挖掘腰椎磁共振成像(magnetic resonance image, MRI)图像中多种模态信息的相关性、腰椎结构间的相互依赖关系、腰椎结构先验知识对腰椎精准分割和疾病辅助诊断的重要价值,提出一种结构先验引导的多模态信息融合分割算法。设计的多模态图像编码模块(multi-modality encoding module, MMEM)可同时对 T1 和 T2 加权图像做语义特征提取;跨模态体素融合模块(cross-modality voxel fusion module, CMVF)可在融合过程中为各模态图像特征自适应分配融合权重。根据腰椎内部各组织结构间的先验知识构建图模型,利用图卷积神经网络分割模块(graph convolutional networks segmentation module, GCNSM)实现图模型上的语义信息传播。采用多模态图像解码模块(multi-modality decoding module, MMDM)对特征图进行解码,实现对椎体及椎间盘的精准图像分割。对山东大学齐鲁医院德州医院采集的 190 组患者 MRI 数据集进行试验验证,所设计算法的平均骰子系数  $D_{ice}$ 、交并比  $I_{ou}$ 、95% Hausdorff 距离  $H_{D95}$  和平均对称表面距离  $A_{SSD}$  分别为 90.3%、82.31%、4.40 mm 和 1.21 mm,结果表明了算法的有效性及其先进性。

**关键词:**腰椎;磁共振成像;多模态;多类别分割;图卷积神经网络

**中图分类号:**TP391 **文献标志码:**A

**引用格式:**李伟豪,王苹苹,许万博,等. 结构先验引导的多模态腰椎 MRI 图像分割算法[J]. 山东大学学报(工学版), 2025, 55(1): 66-76.

LI Weihao, WANG Pingping, XU Wanbo, et al. Multimodal lumbar MRI image segmentation algorithm guided by structure priori[J]. Journal of Shandong University (Engineering Science), 2025, 55(1): 66-76.

## Multimodal lumbar MRI image segmentation algorithm guided by structure priori

LI Weihao<sup>1,2,3</sup>, WANG Pingping<sup>1,2,3</sup>, XU Wanbo<sup>1,2,3,4</sup>, WEI Benzhen<sup>1,2,3\*</sup>

(1. Center for Medical Artificial Intelligence, Shandong University of Traditional Chinese Medicine, Qingdao 266112, Shandong, China; 2. Qingdao Academy of Chinese Medical Science, Shandong University of Traditional Chinese Medicine, Qingdao 266112, Shandong, China; 3. Qingdao Key Laboratory of Artificial Intelligence Technology of Traditional Chinese Medicine, Qingdao 266112, Shandong, China; 4. Qilu Hospital of Shandong University Dezhou Hospital, Dezhou 253046, Shandong, China)

**Abstract:** To explore the relevance of multi-modality information in lumbar magnetic resonance image (MRI) images, the interdependence among lumbar structures, and the significant value of prior knowledge of lumbar structures for precise segmentation of lumbar spine and disease-assisted diagnosis, a segmentation algorithm guided by structural priori for multi-modality information fusion was proposed. A multi-modality encoding module (MEM) was designed, which could simultaneously extract semantic features from T1-weighted and T2-weighted images. A cross-modality voxel fusion module (CMVF) could adaptively allocate fusion weights to the features of each modality image during the fusion process. The graph model was constructed according to the prior knowledge of the internal organizational structures of the lumbar, and the graph convolutional networks segmentation module (GCNSM) was utilized to propagate semantic information on the graph model. Multi-modality decoding module (MMDM) was employed to decode the feature maps, which achieved precise image segmentation of the vertebrae and intervertebral discs. The MRI data set of 190 groups of patients collected from the Qilu Hospital of Shandong University Dezhou Hospital was verified by experiments. The average dice coefficient  $D_{ice}$ , intersection of union  $I_{ou}$ , 95% Hausdorff distance  $H_{D95}$ , and average symmetric surface distance  $A_{SSD}$  of the designed algorithm were 90.3%, 82.31%, 4.40 mm, and 1.21 mm, respectively, demonstrated the effectiveness and advancement of the algorithm.

**Keywords:** lumbar; magnetic resonance image; multi-modality; multi-class segmentation; graph convolutional networks

收稿日期:2023-12-28

基金项目:国家自然科学基金资助项目(62372280, 61872225);山东省自然科学基金资助项目(ZR2020KF013, ZR2019ZD04);青岛市科技惠民示范专项资助项目(23-2-8-smjk-2-nsh);齐鲁卫生与健康领军人才资助项目

第一作者简介:李伟豪(1999—),男,山东菏泽人,硕士研究生,主要研究方向为医学图像处理与分析。E-mail: 986053032@qq.com

\* 通信作者简介:魏本征(1976—),男,山东临沂人,教授,博士生导师,博士,主要研究方向为医学人工智能、医学影像智能分析和医学信息工程。E-mail: wbz99@sina.com

## 0 引言

腰椎在维持身体稳定、承担负重等方面具有极其重要的作用。腰椎疾病长期以来一直是全球范围内常见的健康问题,对个人、医疗保健系统和社会都带来负担<sup>[1]</sup>。诊断是治疗的基础,腰椎疾病的诊疗问题逐渐成为人们关注的重点。腰椎分割是腰椎疾病辅助诊断的前提,有助于检测和定位腰椎解剖或病理结构变化,加快诊断效率,把握最佳治疗时期。

影像学检查为疾病辅助诊断提供重要的数据基础<sup>[2-3]</sup>。磁共振成像(magnetic resonance image, MRI)通过结合不同加权序列对病变情况进行反映,在显示腰椎软组织成分及病变侵袭范围方面较其他成像方法更具优势,已得到广泛应用。MRI 利用 T1 加权图像显示腰椎解剖结构,利用 T2 加权图像反映腰椎周围组织的含水量,从而精确发现水肿、炎症、肿瘤等组织变化<sup>[4]</sup>,两者相互补充,为评估腰椎的形态和病变提供更全面的信息。

近年来,基于深度学习的 MRI 图像腰椎分割方法准确率显著提升<sup>[5]</sup>。在利用单模态图像分割算法研究方面,为实现在分割过程中结合腰椎结构的远距离依赖关系,文献[6]提出一种由空洞卷积自编码器模块、基于长短时记忆(long short-term memory, LSTM)模块的空间动态建模模块和对抗模块构成的递归生成对抗网络(recurrent generative adversarial network, RGAN),用于医学报告生成过程中对 T1 或 T2 加权 MRI 图像椎间盘、神经孔和椎体的语义分割及放射学分类;文献[7]将生成对抗网络(generative adversarial networks, GAN)和 LSTM 同卷积神经网络相结合,设计一种新型循环生成对抗网络,用于对 T1 或 T2 加权 MRI 图像中椎间盘、椎体和神经孔的同时分割和分类;文献[8]通过并行双通道架构融合卷积神经网络(convolutional neural networks, CNN)和 Transformer,在 Transformer 的自注意力模块中引入相对位置嵌入,增强模型对腰椎结构远距离作用关系的建模能力,实现对 T2 加权 MRI 图像中椎体和椎间盘的语义分割。上述方法中使用的 LSTM 和 Transformer 忽略了结构先验信息,未能捕获特定解剖区域的空间相关性,分割性能受到限制。为解决此问题,部分模型通过多任务学习方式或图神经网络模型将结构先验知识添加到图像分割过程中。文献[9]设计一种迭代的全卷积神经网络模型,以滑动窗口方式定位和识别每个椎

体,利用二元分割神经网络分割 T2 加权 MRI 图像中的椎体,但容易发生级联故障;文献[10]提出的检测引导混合监督分割网络(detection-guided mixed-supervised segmentation network, DGMSNet)设计混合监督学习策略和检测引导的标签融合方法,实现在 T2 加权 MRI 图像中对椎体、椎间盘的分割,但将三维矩阵转化为多个二维图像完成分割任务,计算复杂度较高;文献[11]提出的多任务多结构相关学习网络(multi-task multi-structure correlation learning network, MMCL-Net)采用在网络中聚合 3 个任务的学习方法挖掘潜在结构信息,依靠任务间的信息融合增强学习效果,实现对 T1 或 T2 加权 MRI 图像中椎间盘、椎体和椎间孔的同时分割、检测和分类。上述模型通过多任务学习的方式结合结构腰椎先验信息。在结合图神经网络的方法中,文献[12]提出的多椎体分割(multi-vertebrae segmentation, MVSeg)模型利用图卷积神经网络(graph convolutional networks, GCN)<sup>[13]</sup>捕获椎体间的结构先验信息和全局空间信息,将其与卷积层构成的标签注意网络结合,实现 T1 或 T2 加权 MRI 图像中的多椎体分割,但对椎体边界的鲁棒性有待提高;文献[14]利用三维 GCN 结合腰椎各结构间的依赖关系,提取 T2 加权 MRI 图像中的低分辨率特征,利用二维残差 U-Net 通过高分辨率图像切片细化分割结果,但对不确定性较高区域(如棘突)的鲁棒性较差;文献[15]提出的神经符号学习(neural-symbolic learning, NSL)框架在生成对抗网络中加入包含结构先验知识的符号图推理模块,实现对 T1 或 T2 加权 MRI 图像中具有高复杂性和可变性的组织结构的语义分割。但此类结合多任务学习或图神经网络的方法对腰椎图像各结构间长距离作用关系的捕捉能力有待进一步加强。

利用单模态图像分割算法仅将 T1 或 T2 加权 MRI 图像中的一种作为图像输入,未能将二者的互补信息结合。在利用多模态图像分割算法研究方面:文献[16]将 T1 加权和 T2 加权 MRI 图像同时作为输入,结合生成对抗网络和多阶段学习策略的优势,实现对三维 MRI 图像中椎体、椎间盘和椎间孔的同时分割;文献[17]将 T1 加权矢状位和 T2 加权横断位腰椎 MRI 图像结合,实现对中央椎管周围结构的分割和椎管狭窄的分级;文献[18]提出一种基于全卷积网络的椎间盘自动定位和分割方法,以同相位和反相位等 4 种序列的 MRI 图像同时作为输入,采用随机模态元素丢失策略提高模型的识别能力,最终分割结果的骰子系数为 91.2%,平均定位

误差为 0.62 mm;文献[19]设计的基于脊柱电子计算机断层扫描(computer tomography, CT)和 MRI 图像的协同分割方法,在椎体分割任务中精确率和准确率分别达到 74%和 95%。此类利用多模态图像的分割方法将多种模态的图像同时作为输入,但对于腰椎结构远距离依赖关系和解剖结构与位置先验信息的结合能力有待进一步提升。

上述所提方法基于 MRI 图像完成对椎体、椎间盘、椎间孔等腰椎结构的分割,在深度学习分割算法研究方面取得出色表现,但主要存在以下问题: T1、T2 加权图像对比度存在差异,可提供不同结构信息,但二者的互补关系未被充分结合;部分模型在建模腰椎图像远距离依赖关系方面具有局限性;腰椎解剖结构与位置先验信息未得到充分利用。

针对上述问题,本研究提出一种结构先验引导的多模态腰椎 MRI 图像分割算法,实现对腰椎 MRI 图像中椎体和椎间盘的同时分割。在采集于山东大学齐鲁医院德州医院的 190 组患者 MRI 数据集上进行试验,验证本研究算法作为临床工具的有效性和潜在能力。

### 1 本研究分割算法原理

本研究分割算法由多模态图像编码-解码模块、跨模态体素融合模块(cross-modality voxel fusion module, CMVF)和图卷积神经网络分割模块(graph convolutional networks segmentation module, GCNSM)3 部分组成,网络框架如图 1 所示,其中  $D$ 、 $H$  和  $W$  分别为输入图像的深度、高度和宽度,  $C_1$ 、 $C_2$ 、 $C_3$  和  $C_4$  为图像的通道数,此处分别设置为 32、64、128 和 256。图 1 中多模态图像编码模块(multi-modality encoding module, MMEM)用于对 T1、T2 加权图像同时编码,形成特征图;CMVF 模块用于实现特征融合,在融合过程中自适应分配权重,获得数据的综合特征表示;GCNSM 模块用于在图结构上传播融合后的特征,旨在充分结合腰椎结构先验;多模态图像解码模块(multi-modality decoding module, MMDM)将处理后的特征图映射为分割结果。本研究分割算法在编码和解码路径之间引入跳跃连接,使 2 个部分能够共享信息。

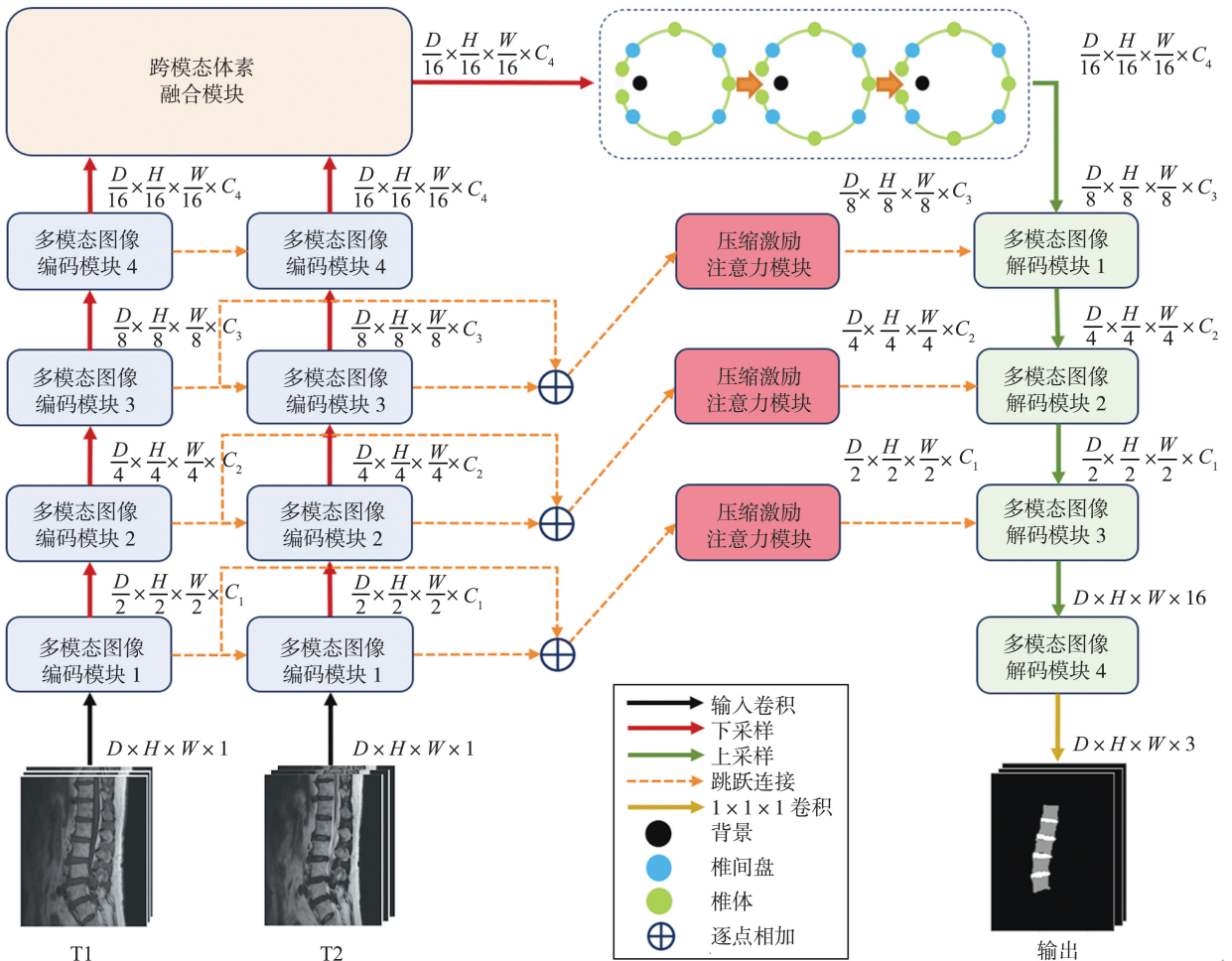


图 1 网络框架示意图  
Fig.1 Network frame diagram

### 1.1 多模态图像编码-解码模块

为充分提取各模态数据的空间信息,多模态图像编码-解码模块利用 2 条路径编码不同模态的输入数据,在不同编码路径之间设置跳跃连接进行信息交互,设置 1 条解码路径生成预测结果,其结构如图 2 所示。

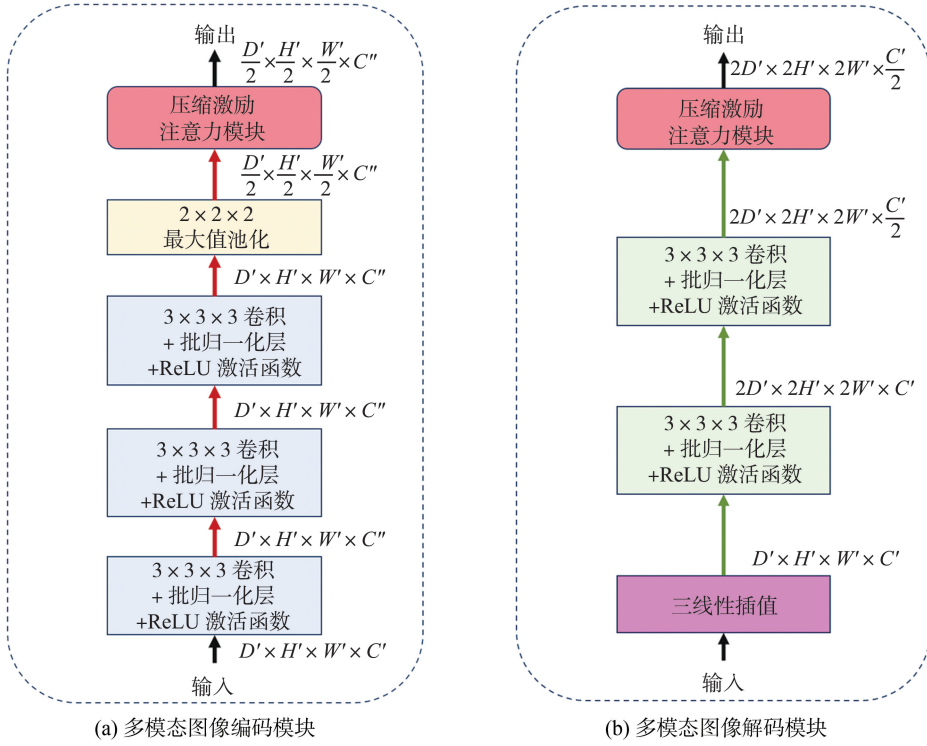


图 2 多模态图像编码-解码模块示意图

Fig.2 Schematic diagram of multi-modality encoding-decoding module

每条编码路径包含 4 个子编码模块,子编码模块如图 2(a) 所示。假设输入特征为  $\mathbf{X}_{\text{Input}}$ ,  $\mathbf{X}_{\text{Input}} \in \mathbf{R}^{D' \times H' \times W' \times C'}$ ,其中  $D'$ 、 $H'$ 、 $W'$  和  $C'$  分别为输入特征  $\mathbf{X}_{\text{Input}}$  的深度、高度、宽度和通道数,每个子模块进行 3 次连续卷积,生成一组特征映射,将特征映射输入批归一化 (batch normalization, BN) 层和线性整流

单元 (rectified linear unit, ReLU) 进一步处理,利用最大值池化操作进行下采样。为使网络自适应学习各特征通道的重要性,实现对特征图中不同通道贡献的加权调整,部署压缩激励 (squeeze-and-excitation, SE)<sup>[20]</sup> 注意力模块对下采样后的特征权重进行调整。SE 注意力模块的结构如图 3 所示。

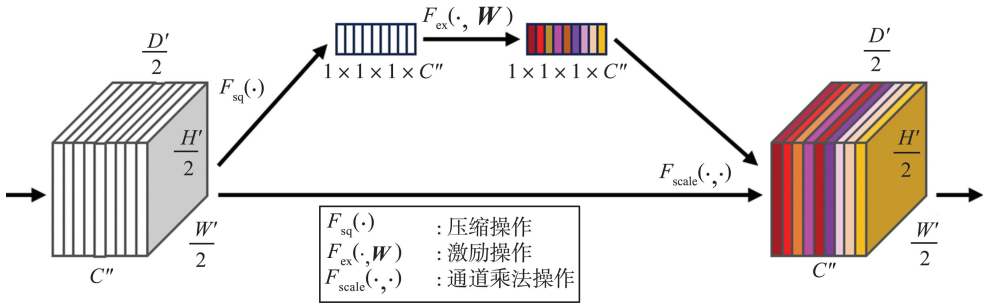


图 3 SE 注意力模块示意图

Fig.3 Schematic diagram of SE attention module

对于输入特征图  $\mathbf{X}_{\text{En}} \in \mathbf{R}^{\frac{D'}{2} \times \frac{H'}{2} \times \frac{W'}{2} \times C''}$ ,压缩阶段将  $\mathbf{X}_{\text{En}}$  压缩为特征向量  $\mathbf{Z}$ ,  $\mathbf{Z} \in \mathbf{R}^{1 \times 1 \times 1 \times C''}$ ,用于捕捉全局特征信息,其中  $C''$  为输入特征  $\mathbf{X}_{\text{En}}$  的通道数。激励阶段通过全连接层和非线性激活函数实现对特征向量的自动加权,挖掘不同通道特征在后续网络层中的重要性。将输入特征与加权后的特征向量相

乘,以调整输入特征中各通道的贡献值。在经 SE 注意力模块处理后,该子编码模块以特征图  $\mathbf{X}_{\text{Ense}}$  作为输出,  $\mathbf{X}_{\text{Ense}} \in \mathbf{R}^{\frac{D'}{2} \times \frac{H'}{2} \times \frac{W'}{2} \times C''}$ 。

解码路由 4 个子解码模块构成,用于对处理后的特征图进行上采样,子解码模块结构如图 2(b) 所示。假设该部分输入特征为  $\mathbf{X}_{\text{De}}$ ,  $\mathbf{X}_{\text{De}} \in$

$\mathbf{R}^{D' \times H' \times W' \times C'}$ , 在每个子模块中对输入特征进行三线性插值, 利用2次连续卷积操作对通道数进行调整, 并跟随批归一化层和线性整流单元。添加一个同图3结构的SE注意力模块, 目的是增强解码阶段各特征通道间的信息交互, 在生成分割结果的同时恢复图像细节。在经SE注意力模块处理后, 该子解码模块输出特征图  $\mathbf{X}_{\text{Dese}} \in \mathbf{R}^{2D' \times 2H' \times 2W' \times \frac{C'}{2}}$ 。模型通过  $1 \times 1 \times 1$  卷积将各部分特征向量映射到结果所需的类别数。

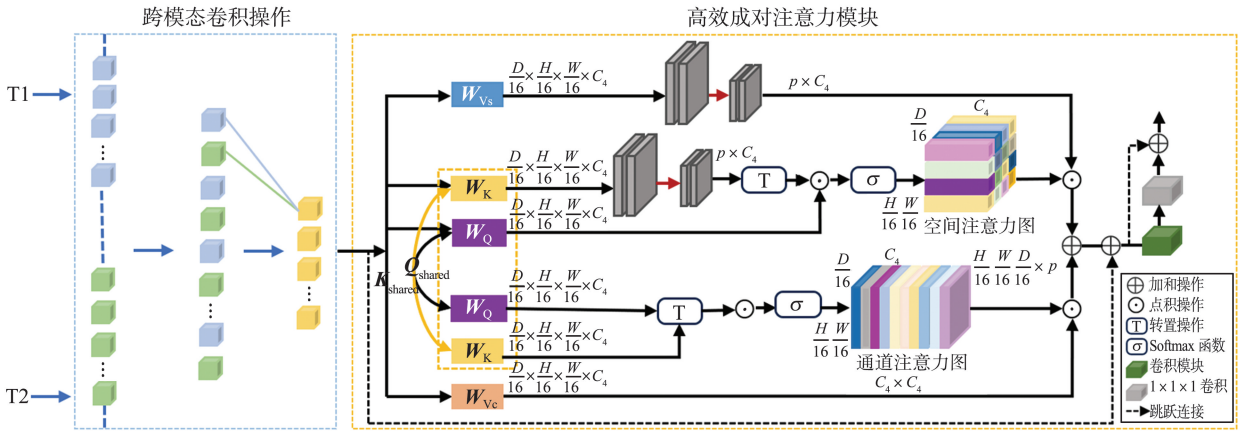


图4 跨模态体素融合模块示意图

Fig.4 Schematic diagram of cross-modality voxel fusion module

跨模态卷积操作可实现对各模态MRI图像特征的高效融合。经编码路径处理后, 模型生成特征映射  $\mathbf{X}_{T1}$  和  $\mathbf{X}_{T2}$ ,  $\mathbf{X}_{T1} \in \mathbf{R}^{\frac{D}{16} \times \frac{H}{16} \times \frac{W}{16} \times C_4}$ ,  $\mathbf{X}_{T2} \in \mathbf{R}^{\frac{D}{16} \times \frac{H}{16} \times \frac{W}{16} \times C_4}$ 。对于包含  $n$  种模态的特征映射, 跨模态卷积操作将同一模态的相同通道特征堆叠到一个堆栈中, 对其整形, 得到整形后的二维特征图  $\mathbf{X}_{\text{Reshape}}$ ,  $\mathbf{X}_{\text{Reshape}} \in \mathbf{R}^{C_4 \times n \times \frac{H}{16} \times \frac{W}{16}}$ ; 进行核大小为  $n \times 1 \times 1$  的3D跨模态卷积操作, 利用  $n \times 1 \times 1$  的平均池化操作对卷积结果进行处理; 将池化后的特征图拼接, 形成拼接后的特征图  $\mathbf{X}_{\text{Concat}} \in \mathbf{R}^{\frac{D}{16} \times \frac{H}{16} \times \frac{W}{16} \times C_4}$ 。上述操作的公式表达为

$$\mathbf{X}_{\text{Stack}} = \text{Stack}(\mathbf{X}_{T1}, \mathbf{X}_{T2}), \quad (1)$$

$$\mathbf{X}_{\text{Reshape}} = \text{Reshape}(\mathbf{X}_{\text{Stack}}), \quad (2)$$

$$\mathbf{X}_{\text{CMC}} = \text{CMC}(\mathbf{X}_{\text{Reshape}}), \quad (3)$$

$$\mathbf{X}_{\text{AvgPool}} = \text{AvgPool}(\mathbf{X}_{\text{CMC}}), \quad (4)$$

$$\mathbf{X}_{\text{Concat}} = \text{Concat}(\mathbf{X}_{\text{AvgPool}}), \quad (5)$$

式中,  $\mathbf{X}_{\text{Stack}}$  为堆叠后的特征图,  $\mathbf{X}_{\text{CMC}}$  为经过3D跨模态卷积操作的特征图,  $\mathbf{X}_{\text{AvgPool}}$  为经过平均池化操作的特征图, Stack为堆叠操作, Reshape为整形操作, CMC为3D跨模态卷积操作, AvgPool为平均池化操作, Concat为拼接操作。

EPA模块包含空间注意力分支和通道注意力分支, 可在特征融合过程中高效捕捉腰椎图像内相

## 1.2 跨模态体素融合模块

为使模型在特征融合过程中充分结合多模态图像的互补信息, 发挥各模态图像的成像优势, 在利用编码路径完成各模态特征提取后, 设置跨模态体素融合模块, 实现多模态图像特征融合。跨模态体素融合模块如图4所示, 由跨模态卷积(cross-modality convolution, CMC)操作<sup>[21]</sup>和高效成对注意力(efficient paired-attention, EPA)模块<sup>[22]</sup>组成。

互依存的空间通道特征表示, 实现对腰椎图像中远距离依赖关系的建模。其中, 空间注意力分支将注意力运算的复杂度从二次型降低到线性型; 通道注意力分支有效学习通道特征映射间的相互依赖关系。EPA模块共享查询矩阵和键矩阵, 并将二者作用于不同的值矩阵, 实现对空间注意力和通道注意力的结合, 提升模型对空间和通道互补特性的学习能力, 产生包含图像全局信息的有效特征, 建模腰椎各结构间的远距离依赖关系。

在空间注意力分支中, 根据输入  $\mathbf{X}_{\text{Concat}}$ , 实现对共享查询矩阵  $\mathbf{Q}_{\text{shared}}$ 、共享键矩阵  $\mathbf{K}_{\text{shared}}$  和空间值矩阵  $\mathbf{V}_{\text{spatial}}$  的计算, 具体计算式为

$$\mathbf{Q}_{\text{shared}} = \mathbf{W}_Q \mathbf{X}_{\text{Concat}}, \quad (6)$$

$$\mathbf{K}_{\text{shared}} = \mathbf{W}_K \mathbf{X}_{\text{Concat}}, \quad (7)$$

$$\mathbf{V}_{\text{spatial}} = \mathbf{W}_{Vs} \mathbf{X}_{\text{Concat}}, \quad (8)$$

式中,  $\mathbf{Q}_{\text{shared}} \in \mathbf{R}^{\frac{D}{16} \times \frac{H}{16} \times \frac{W}{16} \times C_4}$ ,  $\mathbf{K}_{\text{shared}} \in \mathbf{R}^{\frac{D}{16} \times \frac{H}{16} \times \frac{W}{16} \times C_4}$ ,  $\mathbf{V}_{\text{spatial}} \in \mathbf{R}^{\frac{D}{16} \times \frac{H}{16} \times \frac{W}{16} \times C_4}$ ,  $\mathbf{W}_Q$ 、 $\mathbf{W}_K$ 、 $\mathbf{W}_{Vs}$  分别为  $\mathbf{Q}_{\text{shared}}$ 、 $\mathbf{K}_{\text{shared}}$ 、 $\mathbf{V}_{\text{spatial}}$  的投影权重。

完成上述计算后, 通过投影操作将  $\mathbf{K}_{\text{shared}}$  和  $\mathbf{V}_{\text{spatial}}$  降维, 得到矩阵  $\mathbf{K}_{\text{proj}}$  和  $\tilde{\mathbf{V}}_{\text{spatial}}$ ,  $\mathbf{K}_{\text{proj}} \in \mathbf{R}^{p \times C_4}$ ,  $\tilde{\mathbf{V}}_{\text{spatial}} \in \mathbf{R}^{p \times C_4}$ , 其中  $p$  为投影向量的维度。将投影后的  $\mathbf{K}_{\text{proj}}$  转置, 与  $\mathbf{Q}_{\text{shared}}$  相乘, 生成空间注意力图,

Softmax 函数将特征转化为一组归一化的概率分布。将相似度衡量结果与投影后的  $\tilde{V}_{\text{spatial}}$  相乘,生成空间注意力图  $\hat{X}_s$ 。空间注意力机制的计算式为

$$\hat{X}_s = \text{Softmax} \left( \frac{Q_{\text{shared}} K_{\text{proj}}^T}{\sqrt{d}} \right) \cdot \tilde{V}_{\text{spatial}}, \quad (9)$$

式中  $d$  为投影向量的大小。

在通道注意力分支中,通道值矩阵

$$V_{\text{channel}} = W_{Vc} X_{\text{Concat}}, \quad (10)$$

式中,  $V_{\text{channel}} \in \mathbf{R}^{\frac{D}{16} \times \frac{H}{16} \times \frac{W}{16} \times C_4}$ ,  $W_{Vc}$  为  $V_{\text{channel}}$  的投影权重。

通道注意力机制的计算式为

$$\hat{X}_c = V_{\text{channel}} \cdot \text{Softmax} \left( \frac{Q_{\text{shared}}^T K_{\text{shared}}}{\sqrt{d}} \right), \quad (11)$$

式中  $\hat{X}_c$  为通道注意力图。

EPA 模块利用卷积操作对两分支的输出特征进行融合,将其作为图卷积神经网络分割模块的输入。EPA 模块的输出特征图

$$\hat{X} = \text{Conv}_1 (\text{Conv}_3 (\hat{X}_s + \hat{X}_c)), \quad (12)$$

式中,  $\text{Conv}_1$  和  $\text{Conv}_3$  分别为  $1 \times 1 \times 1$  卷积操作和  $3 \times 3 \times 3$  卷积操作。

### 1.3 图卷积神经网络分割模块

为使模型在分割过程中充分结合腰椎解剖结构与位置先验信息,设计 GCNSM 模块,依靠邻接矩阵对腰椎结构间的固有依赖关系进行高效建模,使用堆叠 GCN 在融合后的特征图中传播全局信息,用于实现不同腰椎结构间的信息传播。

腰椎由 5 个椎体和 4 个椎间盘构成,其中  $L_1$ 、 $L_2$ 、 $L_3$ 、 $L_4$ 、 $L_5$  代表 5 个椎体,  $L_1-L_2$ 、 $L_2-L_3$ 、 $L_3-L_4$ 、 $L_4-L_5$  代表 4 个椎间盘,各结构间的连接方式如图 5 所示。在构建腰椎邻接矩阵  $A$  时,将 5 个椎体和 4 个椎间盘抽象为节点,将椎体和椎间盘的邻接关系抽象为边,当且仅当 2 个节点由图 5 中的白线连接时,椎体与椎间盘之间存在邻接关系,可以进行信息交互,实现对腰椎结构间相关性的定义。

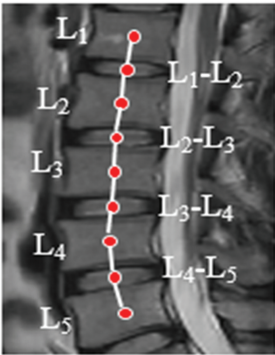


图5 腰椎各结构连接关系示意图

Fig.5 Schematic diagram of the connection relationship of various structures of the lumbar spine

在完成邻接矩阵的构建后,GCN 通过聚合邻接矩阵中相邻节点的空间信息对目标节点特征进行更新。对于输入特征矩阵  $\hat{X}$ ,利用单层 GCN 对其进行更新,更新后的节点特征矩阵

$$F_1 = f(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \hat{X} W_0), \quad (13)$$

式中:  $F_1 \in \mathbf{R}^{N \times k}$ , 其中  $N$  为节点数,  $k$  为节点特征维度;  $\tilde{A} = A + I$ , 其中  $I$  为单位矩阵;  $\tilde{D}$  为对角矩阵,  $\tilde{D}$  中的元素  $\tilde{d}_{i,i}$  与  $\tilde{A}$  中的元素  $\tilde{a}_{i,j}$  的关系为  $\tilde{d}_{i,i} = \sum_j \tilde{a}_{i,j}$ ;  $W_0$  为可学习的权重矩阵;  $f(\cdot)$  为激活函数。单层 GCN 只实现对目标节点近邻信息的捕捉,为提取更多的邻域信息,多层 GCN 的节点特征矩阵

$$F_{i+1} = f(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} F_i W_i), \quad (14)$$

式中  $W_i$  为第  $i$  层权重矩阵。经试验验证,叠加 3 层 GCN 可实现对全局信息建模,即  $F_3$  中每个节点特征都已经过所有节点特征的加权聚合。

将堆叠 GCN 的输出重构为与  $\hat{X}$  相同的大小,并与  $\hat{X}$  相加,以融合原始特征。该方法利用图卷积操作实现对全局和局部空间信息的同时捕捉,所获特征具有更强的表示能力。

## 2 试验过程与结果分析

### 2.1 数据与预处理

本试验所用数据集为采集自山东大学齐鲁医院德州医院的 190 组临床患者腰椎矢状位 MRI 图像数据集,每组图像同时包含 T1 加权图像和 T2 加权图像,层厚为 4.0 mm,层间距为 0.5 mm,层数为 11,图像大小为 (408~512) 像素  $\times$  (489~384) 像素。由 2 名资深放射科医生对图像进行标注,可清晰展示椎体和椎间盘的相对位置。在预处理阶段对数据尺度进行统一,设置每组图像层数为 48,图像大小为 256 像素  $\times$  256 像素。通过随机裁剪、随机旋转等方法对训练数据进行增强,以提高网络泛化能力。

### 2.2 试验设置

将试验所用的图像数据以 8:1:1 的比例划分为训练集、验证集和测试集。本研究所提方法使用 Pytorch 1.13.0 库设计实现,在配备 10 GB 内存的 NVIDIA GeForce RTX 3080 GPU 工作站上迭代训练,最大迭代次数为 500 次。试验过程中设定初始学习率为 0.001,采用 Tversky 损失函数、Adam 优化器对网络进行训练,结合权重衰减策略,在每 30 次后将学习率递减为当前学习率的 80%。在图

卷积神经网络分割模块利用3层图卷积神经网络进行全局信息传播,设置图节点特征维度为128。

### 2.3 评价指标

本试验采用骰子系数  $D_{ice}$  和交并比  $I_{ou}$  衡量分割结果准确性,通过95% Hausdorff 距离  $H_{D95}$  和平均对称表面距离  $A_{SSD}$  进一步评估不同模型对椎体、椎间盘边缘的分割能力。

作为在体素级混淆矩阵上计算的标准度量,  $D_{ice}$  和  $I_{ou}$  的计算式分别为

$$D_{ice} = \frac{2T_p}{2T_p + F_p + F_n}, \quad (15)$$

$$I_{ou} = \frac{T_p}{T_p + F_p + F_n}, \quad (16)$$

式中,  $T_p$  和  $F_p$  分别为真阳性和假阳性,  $F_n$  为假阴性。

$H_{D95}$  通过计算前95%的 Hausdorff 距离评估模型对图像边界的分割能力,其定义为

$$H_{D95}(Y, P) = \max \{ d_{yp}, d_{py} \}, \quad (17)$$

式中,  $Y$  为真实掩码的体素集合,  $P$  为分割结果边界的体素集合,  $d_{yp}$  为分割结果边界体素与真实掩码体素间前95%的 Hausdorff 距离,  $d_{py}$  为真实掩码体素与分割结果边界体素间前95%的 Hausdorff 距离。

$A_{SSD}$  通过计算平均对称表面距离衡量模型对图像边界的分割能力。设  $S(A)$  为分割结果  $A$  的表面体素集合,任意体素  $v$  到  $S(A)$  的最短距离

$$d(v, S(A)) = \min_{s_A \in S(A)} \|v - s_A\|, \quad (18)$$

式中,  $s_A$  为  $S(A)$  中的表面体素,  $\|\cdot\|$  为欧几里得距离。则  $A_{SSD}$  的定义式为

$$A_{SSD}(A, B) = \frac{1}{|S(A) + S(B)|} \cdot \left( \sum_{s_A \in S(A)} d(s_A, S(B)) + \sum_{s_B \in S(B)} d(s_B, S(A)) \right), \quad (19)$$

式中,  $S(B)$  为分割结果  $B$  的表面体素集合,  $s_B$  为  $S(B)$  中的表面体素。

### 2.4 结果与分析

#### 2.4.1 各模块的有效性分析

为实现对分割过程中各模块有效性的研究,以 T1 加权图像和 T2 加权图像共同作为输入,对图1所示网络进行训练,用于消融试验,试验结果如表1、2所示。相比仅利用 MMEM+MMDM 模块产生的分割结果,CMVF 和 GCNSM 模块对分割精度均有所提升。对于平均  $D_{ice}$ , MMEM+CMVF+MMDM 模块提升约 0.22 百分点, MMEM+GCNSM+MMDM 提升约 0.09 百分点,本研究算法提升约 0.28 百分点。对于平均  $I_{ou}$ , MMEM+CMVF+MMDM 模块提升约 0.37 百分点, MMEM+GCNSM+MMDM 提升约 0.16 百分点,本研究算法提升约 0.46 百分点。为进一步实现对上述方法边缘分割能力的准确评估,以  $H_{D95}$  和  $A_{SSD}$  作为评价指标, MMEM+CMVF+MMDM 模块分别降低约 0.13 mm 和 0.05 mm, MMEM+GCNSM+MMDM 分别降低约 0.16 mm 和 0.05 mm, 本研究算法分别降低约 0.38 mm 和 0.12 mm。

各类模块的分割结果如图6所示,以二维标注掩码切片作为背景,利用上色后的二维分割结果切片对背景进行覆盖。

表1 消融试验  $D_{ice}$  与  $I_{ou}$  结果统计表

Table 1 Statistical table of  $D_{ice}$  and  $I_{ou}$  results of ablation tests

单位: %

方法	$D_{ice}$			$I_{ou}$		
	椎体	椎间盘	平均	椎体	椎间盘	平均
MMEM+MMDM	90.91±0.29	89.13±0.28	90.02±0.29	83.35±0.50	80.40±0.45	81.85±0.48
MMEM+CMVF+MMDM	91.19±0.10	89.28±0.31	90.24±0.11	83.81±0.16	80.63±0.51	82.22±0.19
MMEM+GCNSM+MMDM	91.11±0.38	89.12±0.58	90.11±0.46	83.67±0.64	80.38±0.93	82.01±0.76
本研究算法(T1)	91.04±0.39	89.11±0.24	90.08±0.28	83.56±0.65	80.37±0.39	81.95±0.47
本研究算法(T2)	91.19±0.02	89.16±0.01	90.18±0.02	83.80±0.04	80.44±0.02	82.10±0.02
本研究算法(T1+T2)	<b>91.20±0.30</b>	<b>89.40±0.25</b>	<b>90.30±0.16</b>	<b>83.82±0.50</b>	<b>80.83±0.40</b>	<b>82.31±0.26</b>

表2 消融试验  $H_{D95}$  与  $A_{SSD}$  结果统计表

Table 2 Statistical table of results of  $H_{D95}$  and  $A_{SSD}$  ablation tests

单位: mm

方法	$H_{D95}$	$A_{SSD}$
MMEM+MMDM	4.78±1.65	1.33±0.35
MMEM+CMVF+MMDM	4.65±1.37	1.28±0.30
MMEM+GCNSM+MMDM	4.62±1.55	1.28±0.32
本研究算法(T1)	4.84±1.51	1.41±0.49
本研究算法(T2)	4.53±1.21	1.27±0.28
本研究算法(T1+T2)	<b>4.40±1.16</b>	<b>1.21±0.26</b>

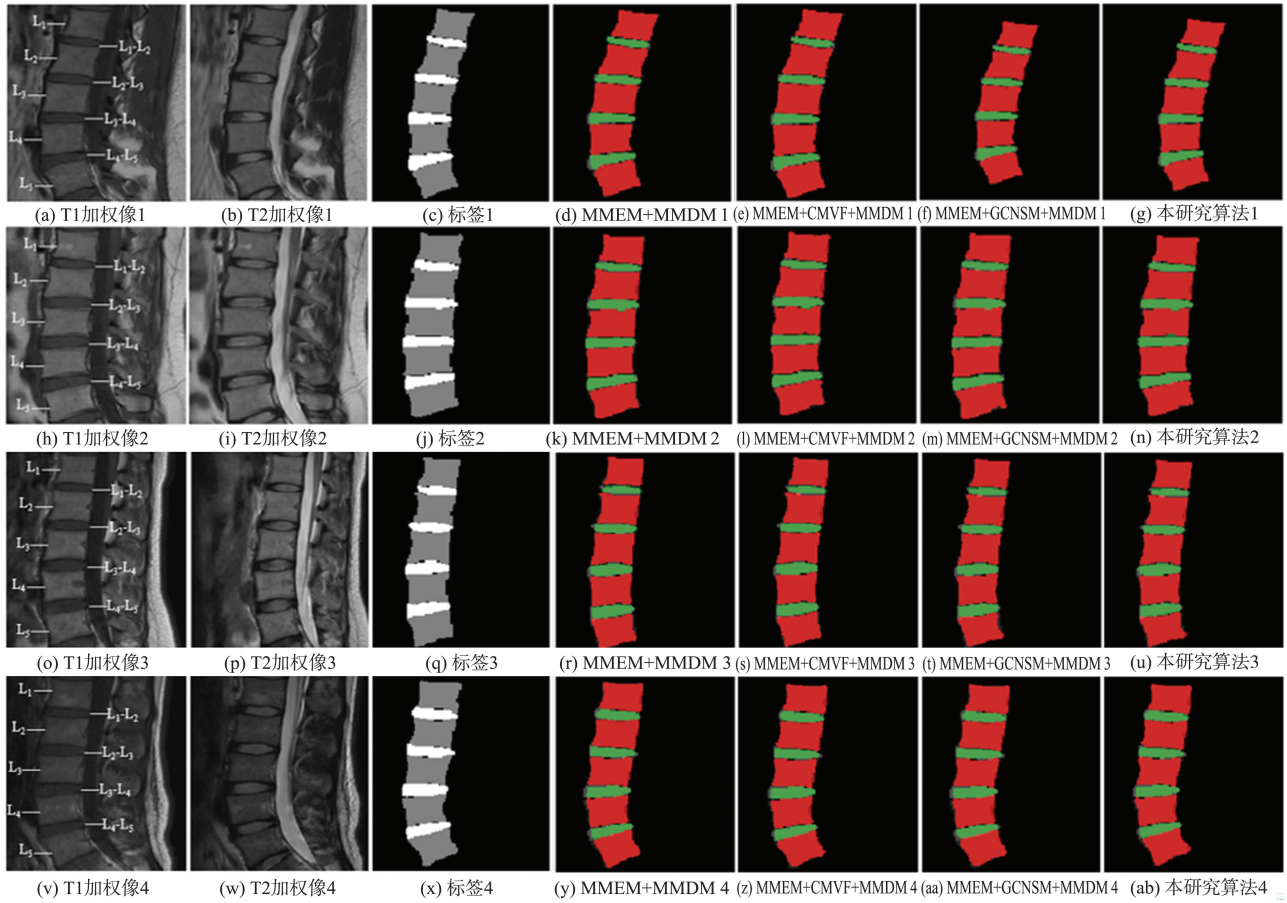


图 6 各类模块分割结果示例图

Fig.6 Examples of segmentation results of various modules

相较于使用 MMEM+MMDM 模块,MMEM+MMDM 与 CMVF 模块的结合获得更优的分割表现,  $H_{D95}$  和  $A_{SSD}$  分别降低 0.13 mm 和 0.05 mm, 可能的原因:跨模态卷积操作自适应地对不同模态特征分别赋予权重,EPA 模块实现在特征融合过程中对远距离依赖关系的捕捉,从而获得更精准的特性融合结果。GCNSM 模块将腰椎结构先验同分割过程相结合,对于椎间盘  $L_1-L_2$  和  $L_2-L_3$  边界部分的分割结果比利用 MMEM+MMDM 模块的结果更精确,  $H_{D95}$  和  $A_{SSD}$  分别降低 0.16 mm 和 0.05 mm。综上所述,将 MMEM+MMDM 模块同 CMVF 和 GCNSM 模块相结合,不仅能保证每种模态的特征不被其他模态特征干扰,而且能保证特征融合过程中对不同模态特征融合权重的自适应分配,获得更好的分割结果。

#### 2.4.2 多模态图像输入的有效性分析

为实现对多模态图像输入有效性的定量分析,试验使用 T1 加权图像、T2 加权图像和二者结合的方法分别训练图 1 所示网络。如表 1、2 所示,在多模态输入网络产生的试验结果中,平均  $D_{ice}$  比单模

态输入网络高出约 0.17 百分点,平均  $I_{OU}$  高出约 0.29 百分点,  $H_{D95}$  降低约 0.29 mm,  $A_{SSD}$  降低约 0.13 mm, 说明与单模态数据相比,多模态数据可以提供更丰富的补充信息。使用 T1 加权图像作为输入的分割结果比使用 T2 加权图像获得更高的分割精度,其中一个主要原因是相较于 T2 加权图像,T1 加权图像在椎体、椎间盘区域的像素与其邻近区域像素间的对比度较大,使网络更容易分辨椎体、椎间盘的相对位置。

不同试验设置下的分割结果如图 7 所示,包括使用 T1 加权图像、T2 加权图像和 T1+T2 加权图像进行训练。

由图 7 可知,所有试验设置均能以合理的精度实现椎体和椎间盘的分割,多模态图像输入可获得更为精确的分割边界。相较于图 7(d)、(j)、(p)、(v) 中对椎间盘  $L_1-L_2$  和  $L_2-L_3$  的分割结果和图 7(e)、(k)、(q)、(w) 中对椎间盘  $L_2-L_3$  和  $L_3-L_4$  的分割结果,图 7(f)、(l)、(r)、(x) 对上述椎间盘的分割边界更精确。上述试验结果表明,相比使用单模态图像作为输入,使用多模态 MRI 图像进行网络训练可以获得更精确的分割结果。

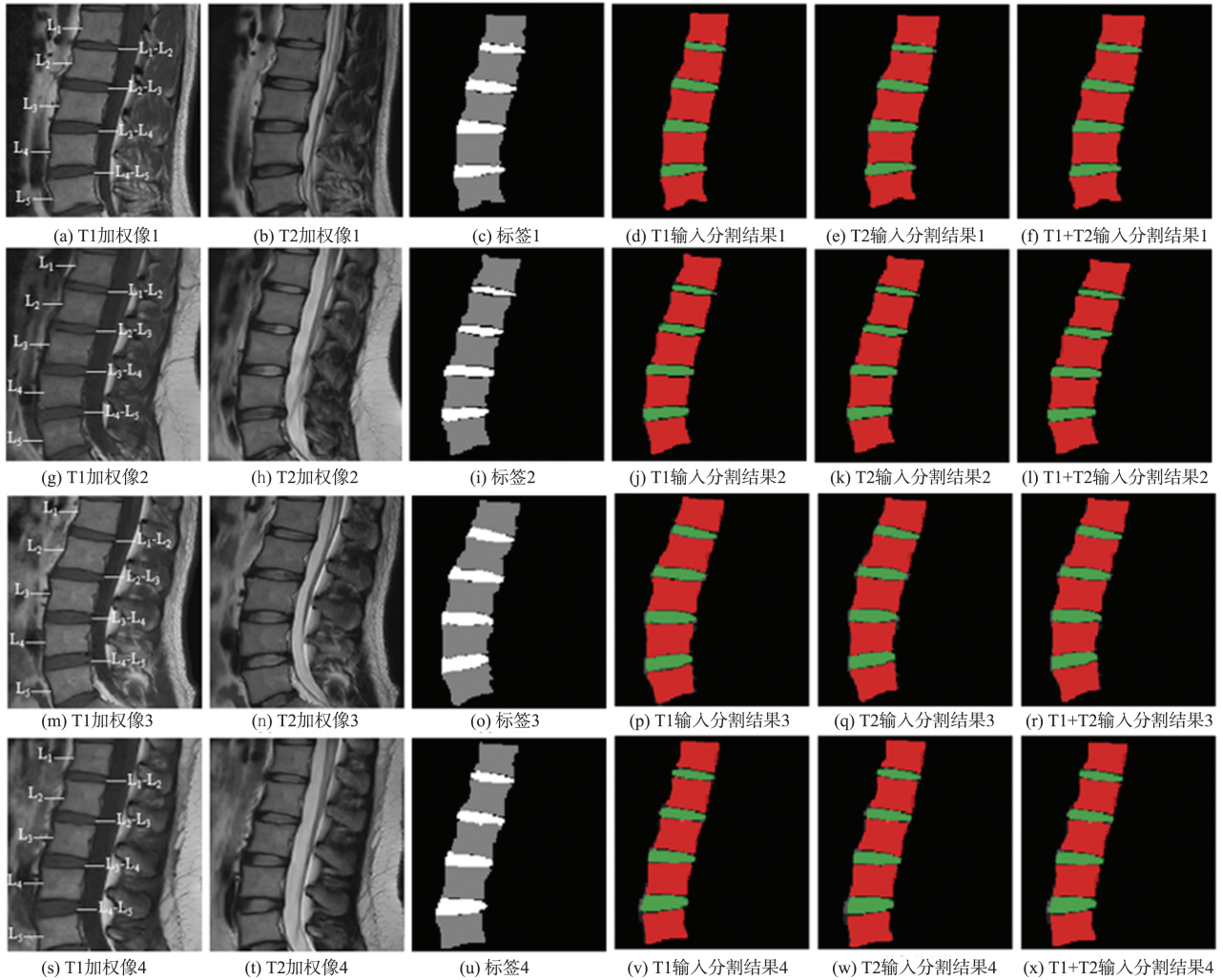


图7 T1、T2 加权图像输入分割结果示例图

Fig.7 T1, T2 weighted image input segmentation results sample graph

### 2.4.3 对比试验

将本研究算法与 UNet3+(3D)<sup>[23]</sup>、脊柱解析网络 (SpineParseNet)<sup>[14]</sup>、体积 Transformer-UNet (volumetric Transformer UNet, VT-UNet)<sup>[24]</sup>、分层上下文注意力网络 (hierarchical context attention network, HCA-Net)<sup>[25]</sup> 和共生 Transformer-CNN (symbiotic Transformer-CNN, SymTC)<sup>[8]</sup> 进行比较, 结果如表 3、4 所示。将本研究算法与经典网络 UNet3+(3D) 进行对比, 发现本研究算法的平均  $D_{ice}$  提升约 0.49 百分点, 平均  $I_{ou}$  提升约 0.80 百分点,  $H_{D95}$  降低 0.42 mm,  $A_{SSD}$  降低 0.16 mm。SpineParseNet 利用图神经网络完成对椎体、椎间盘的分割, 将本研究算法与其进行对比, 发现本研究算法的平均  $D_{ice}$  提升约 5.72 百分点, 平均  $I_{ou}$  提升约 8.95 百分点,  $H_{D95}$  降低 0.36 mm,  $A_{SSD}$  降低 0.29 mm。VT-UNet 是一种基于 Transformer 构建的医学图像分割模型, 将本研究算法与其进行对比, 发现本研究算法的平均  $D_{ice}$  提升约 6.56 百分点, 平均  $I_{ou}$  提升

约 10.28 百分点,  $H_{D95}$  降低 0.85 mm,  $A_{SSD}$  降低 0.43 mm, 具有更高的分割精度。HCA-Net 将腰椎结构分割问题建模为位姿估计问题, 设计骨骼损失函数以增强模型对脊柱的几何依赖性, 实现将模型的预测结果与人类脊椎骨骼的一般结构相匹配, 以提高分割结果的准确性。相较于 HCA-Net 算法, 本研究算法的平均  $D_{ice}$  提升约 0.20 百分点, 平均  $I_{ou}$  提升约 0.30 百分点,  $H_{D95}$  降低 0.13 mm,  $A_{SSD}$  降低 0.09 mm。SymTC 通过并行双通道架构融合 CNN 和 Transformer, 在 Transformer 的自注意力机制模块中集成一种位置信息嵌入, 可以增强该模型对腰椎各结构间位置信息的利用能力, 实现 MRI 图像中椎体、椎间盘的分割。将本研究算法与 SymTC 进行对比, 发现本研究算法的平均  $D_{ice}$  提升约 0.03 百分点, 平均  $I_{ou}$  提升约 0.05 百分点,  $H_{D95}$  降低 0.05 mm,  $A_{SSD}$  降低 0.03 mm。综上, 本研究算法在腰椎结构分割性能上具有较强优势。

表 3 对比试验  $D_{ice}$  与  $I_{ou}$  结果统计表  
Table 3 Statistical table of  $D_{ice}$  and  $I_{ou}$  results of comparative tests

单位: %

方法	$D_{ice}$			$I_{ou}$		
	椎体	椎间盘	平均	椎体	椎间盘	平均
UNet3+(3D)	90.48±0.58	89.13±0.72	89.81±0.61	82.62±1.02	80.41±1.17	81.51±0.91
SpineParseNet	88.85±1.56	80.30±6.44	84.58±2.63	79.98±2.53	67.56±8.69	73.36±3.90
VT-UNet	86.52±0.38	80.96±0.90	83.74±0.64	76.24±0.59	68.02±1.28	72.03±0.95
HCA-Net	93.15±0.69	86.94±0.58	90.10±0.63	87.20±1.20	76.90±0.90	82.01±1.05
SymTC	<b>93.18±0.14</b>	87.19±0.14	90.27±0.11	<b>87.22±0.25</b>	77.29±0.23	82.26±0.19
本研究算法	91.20±0.30	<b>89.40±0.25</b>	<b>90.30±0.16</b>	83.82±0.50	<b>80.83±0.40</b>	<b>82.31±0.26</b>

表 4 对比试验  $H_{D95}$  与  $A_{SSD}$  结果统计表  
Table 4 Statistical table of  $H_{D95}$  and  $A_{SSD}$  results of comparative tests

单位: mm

方法	$H_{D95}$	$A_{SSD}$
UNet3+(3D)	4.82±1.50	1.37±0.36
SpineParseNet	4.76±1.09	1.50±0.30
VT-UNet	5.25±1.36	1.64±0.57
HCA-Net	4.53±0.08	1.30±0.01
SymTC	4.45±0.08	1.24±0.01
本研究算法	<b>4.40±1.16</b>	<b>1.21±0.26</b>

### 3 结论

本研究提出一种结构先验引导的腰椎分割算法,在分割过程中充分考虑 T1、T2 加权腰椎 MRI 图像的相关性,使用 EPA 注意力机制精准捕获腰椎各结构间的长距离依赖关系,依靠图卷积神经网络结合腰椎结构先验,实现对椎体和椎间盘的准确分割。在 190 组患者 MRI 图像上对本研究算法的有效性进行验证,结果表明本研究算法具有更高的分割精度和可靠性,可实现对多种腰椎结构的分割,体现其进行腰椎疾病计算机辅助诊断的潜在效能。

#### 参考文献:

[1] JAMES S L, ABATE D, ABATE K H, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990 – 2017: a systematic analysis for the Global Burden of Disease Study 2017[J]. The Lancet, 2018, 392(10159): 1789-1858.

[2] 刘方旭, 王建, 魏本征. 基于多空间注意力的小儿肺炎辅助诊断算法[J]. 山东大学学报(工学版), 2023, 53(2): 135-142.

LIU Fangxu, WANG Jian, WEI Benzhen. Auxiliary diagnosis algorithm for pediatric pneumonia based on multi-spatial attention[J] Journal of Shandong University (Engineering Science), 2023, 53(2): 135-142.

[3] 张月芳, 邓红霞, 呼春香, 等. 融合残差块注意力机制和生成对抗网络的海马体分割[J]. 山东大学学报(工学版), 2020, 50(6): 76-81.

ZHANG Yuefang, DENG Hongxia, HU Chunxiang, et al. Hippocampal segmentation combining residual attention mechanism and generative adversarial networks [J]. Journal of Shandong University (Engineering Science), 2020, 50(6): 76-81.

[4] SERAI S D. Basics of magnetic resonance imaging and quantitative parameters T1, T2, T2\*, T1rho and diffusion-weighted imaging[J]. Pediatric Radiology, 2022, 52(2): 217-227.

[5] 王国力, 孙宇, 魏本征. 医学图像深度学习分割算法综述[J]. 计算机工程与应用, 2022, 58(12): 37-50.

WANG Guoli, SUN Yu, WEI Benzhen. Systematic review on graph deep learning in medical image segmentation[J] Computer Engineering and Applications, 2022, 58(12): 37-50.

[6] HAN Z, WEI B, LEUNG S, et al. Towards automatic report generation in spine radiology using weakly supervised framework[C]//Proceedings of Medical Image Computing and Computer Assisted Intervention-MICCAI 2018. Granada, Spain: Springer, 2018: 185-193.

[7] HAN Z, WEI B, MERCADO A, et al. Spine-GAN: semantic segmentation of multiple spinal structures [J]. Medical Image Analysis, 2018, 50: 23-35.

[8] CHEN J, QIAN L, MA L, et al. SymTC: a symbiotic Transformer-CNN net for instance segmentation of lumbar spine MRI[EB/OL]. (2024-01-17) [2024-03-17]. <https://arxiv.org/abs/2401.09627>.

[9] LESSMANN N, VAN GINNEKEN B, DE JONG P A, et al. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification [J]. Medical Image Analysis, 2019, 53: 142-155.

[10] PANG S, PANG C, SU Z, et al. DGMSNet: spine segmentation for MR image by a detection-guided mixed-supervised segmentation network [J]. Medical Image Analysis, 2022, 75: 102261.

[11] HONG Y, WEI B, HAN Z, et al. MMCL-Net: spinal disease diagnosis in global mode using progressive multi-

- task joint learning [J]. *Neurocomputing*, 2020, 399: 307-316.
- [12] CHANG H, ZHAO S, ZHENG H, et al. Multi-vertebrae segmentation from arbitrary spine MR images under global view[C]//*Proceedings of Medical Image Computing and Computer Assisted Intervention-MICCAI 2020*. Lima, Peru; Springer, 2020: 702-711.
- [13] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2016-09-09)[2023-11-15]. <https://arxiv.org/abs/1609.02907>.
- [14] PANG S, PANG C, ZHAO L, et al. SpineParseNet: spine parsing for volumetric MR image by a two-stage segmentation framework with semantic image representation[J]. *IEEE Transactions on Medical Imaging*, 2020, 40(1): 262-273.
- [15] HAN Z, WEI B, XI X, et al. Unifying neural learning and symbolic reasoning for spinal medical report generation[J]. *Medical Image Analysis*, 2021, 67: 101872.
- [16] LI T, WEI B, CONG J, et al. S3egANet: 3D spinal structures segmentation via adversarial nets[J]. *IEEE Access*, 2019, 8: 1892-1901.
- [17] LIM D S W, MAKMUR A, ZHU L, et al. Improved productivity using deep learning-assisted reporting for lumbar spine MRI [J]. *Radiology*, 2022, 305(1): 160-166.
- [18] LI X, DOU Q, CHEN H, et al. 3D multi-scale FCN with random modality voxel dropout learning for intervertebral disc localization and segmentation from multi-modality MR images[J]. *Medical Image Analysis*, 2018, 45: 41-54.
- [19] ZOU Y, SHI Y. Co-segmentation of multi-modality spinal image using channel and spatial attention [C]//*Proceedings of Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021*. Strasbourg, France; Springer, 2021: 287-295.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York, USA; IEEE, 2018: 7132-7141.
- [21] TSENG K L, LIN Y L, HSU W, et al. Joint sequence learning and cross-modality convolution for 3D biomedical segmentation [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA; IEEE, 2017: 6393-6400.
- [22] SHAKER A, MAAZ M, RASHEED H, et al. UNETR++: delving into efficient and accurate 3D medical image segmentation[EB/OL]. (2022-12-08)[2023-11-15]. <https://arxiv.org/abs/2212.04497>.
- [23] HUANG H, LIN L, TONG R, et al. UNet 3+: a full-scale connected unet for medical image segmentation [C]//*Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain; IEEE, 2020: 1055-1059.
- [24] PEIRIS H, HAYAT M, CHEN Z, et al. A robust volumetric transformer for accurate 3D tumor segmentation[C]//*Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham, Germany; Springer, 2022: 162-172.
- [25] BOZORGPOUR A, AZAD B, AZAD R, et al. HCA-Net: hierarchical context attention network for intervertebral disc semantic labeling [EB/OL]. (2023-11-21)[2024-03-17]. <https://arxiv.org/abs/2311.12486>.

(编辑:孙亚彤)