

PET-CT 多模态融合在图像语义分割的应用进展

周信宏¹, 黄钢²

1. 上海理工大学健康科学与工程学院, 上海 200093; 2. 上海健康医学院附属嘉定中心医院上海市分子影像学重点实验室, 上海 201318

【摘要】近年来,深度学习在各种医学图像语义分割任务中表现出了优异的性能。然而,在癌症的淋巴结转移以及一些细小病灶的分割任务中,基于单模态图像的语义分割仍然存在局限性。PET-CT这一多模态成像模式提供了有关结构和生理病理的解剖和功能信息,被认为是癌症诊断、分期和治疗反应评估的首选成像设备。PET-CT携带的多模态互补信息被引入深度卷积神经网络的语义分割模型中辅助病灶的分割。首先,聚焦于PET-CT多模态图像特征融合在卷积神经网络图像语义分割中的应用,介绍多模态成像的特点以及常用的医学图像分割网络。然后,归纳和总结了现阶段基于深度学习的多模态融合的3种思路,并划分为早期融合、后期融合和混合融合,并从分割的性能、参数量的大小、方法的亮点与不足等维度对多模态融合分割方法进行优缺点分析。最后,对现阶段多模态医学图像分割存在的问题进行讨论。

【关键词】PET-CT; 卷积神经网络; 语义分割; 多模态; 综述

【中图分类号】R318; TP391

【文献标志码】A

【文章编号】1005-202X(2023)06-0683-12

Multimodal fusion of PET-CT for semantic image segmentation: a review

ZHOU Xinhong¹, HUANG Gang²

1. School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2. Shanghai Key Laboratory of Molecular Imaging, Jiading District Central Hospital, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China

Abstract: In recent years, deep learning has exhibited excellent performance in various semantic image segmentation tasks. However, the semantic segmentation based on monomodal image still has limitations in the segmentations of cancer with lymph node metastasis and some small lesions. PET-CT which is a multimodal imaging mode, providing anatomical and functional information about structure, physiology and pathology, is considered to be the preferred imaging choice for cancer diagnosis, staging and treatment response evaluation. The multimodal complementary information carried by PET-CT is introduced into the semantic segmentation model of deep convolutional neural network to assist the segmentation of lesions. The study focuses on the application of PET-CT multimodal image feature fusion in semantic image segmentation using convolutional neural network, and introduces the characteristics of multimodal imaging and commonly-used medical image segmentation networks. Three ideas for multimodal fusion based on deep learning at this stage are summarized and divided into early fusion, late fusion and hybrid fusion. The pros and cons of multimodal fusion based segmentation methods are analyzed in terms of segmentation performance, parameter size, method highlights and shortcomings. Finally, the current issues existing in multimodal medical image segmentation are discussed.

Keywords: PET-CT; convolutional neural network; semantic segmentation; multimodal; review

【收稿日期】2022-12-20

【基金项目】国家自然科学基金(81830052);上海市分子影像学重点实验室建设项目(18DZ2260400)

【作者简介】周信宏, 硕士, 研究方向: 医学图像处理, E-mail: zhou_xh163@163.com

【通信作者】黄钢, 博士, 教授, 研究方向: 核医学分子影像, E-mail: huanggang@sumhs.edu.cn

前言

医学影像在临床诊断、术前计划、癌症放射治疗等临床应用中起着至关重要的作用^[1]。随着医学影像技术的发展,单模态的医学影像可以很好地表征特定信息,如使用氟代脱氧葡萄糖(¹⁸F-FDG)的PET图像可以反映患者的代谢信息,而CT图像可以反映患者的解剖结构信息。尽管PET图像提供的代谢信息可以观测肿瘤的吸收代谢情况,但其分辨率较低,存在一定局限性;

CT图像具有很好的成像质量,但无法提供肿瘤的代谢信息。而PET/CT这一多模态医学影像弥补了单一模态下的不足,很好地结合CT图像的解剖结构信息和PET图像的病灶代谢信息^[2]。随着深度学习技术的发展,涌现出许多基于深度学习的医学图像分割与检测算法模型^[3]。基于单模态图像的深度学习模型在病灶分割、目标检测与疾病分类任务中均取得了不错的成绩,但在细小病灶/组织上存在分割性能不足的问题^[4-6]。一些细小病灶在CT图像上表现为其占位性病变与相邻生理组织无显著差异,这使得深度学习分割模型在这一单模态图像中的任务性能存在较大的提升空间^[7]。而PET-CT这一多模态医学图像提供的病灶代谢信息可以有效帮助解决这类问题。PET成像具有高对比度,但由于其有限的空间分辨率,常常导致肿瘤边缘模糊,而CT成像具有高分辨率,但肿瘤与其周围正常组织之间的差异性较低。利用PET和CT之间的互补信息进行联合分割可以有效提高模型的分割精度。

基于多模态医学图像的深度学习模型需要结合两个或多个成像模态下的互补信息,与传统深度学习图像分割模型所用单一模态数据略有不同。U-net等模型均基于单模态的图像进行分割,通常利用编码器-解码器和跳跃连接的结构对网络进行训练-预测-分割^[8]。而使用多模态数据的模型往往需要借助多模态的特征融合策略融合多个模态的互补特征辅助模型进行分割,从而提高分割性能。这就使得多模态特征融合的质量对最后模型的性能具有重要作用。

基于深度学习的多模态医学图像分割思路多样,所用模型和方法各有异同,为帮助医学图像处理领域的科研人员和医生进行高效探索和开发应用,本文聚焦于PET-CT这一多模态医学图像在深度学习图像分割模型的应用,对近年来基于深度学习的多模态医学图像分割算法进行阐述。同时,本文侧重于各研究中的多模态特征融合策略进行重点分析与讨论。最后对各方法在医学图像分割任务中的性能和优缺点进行对比分析,讨论多模态医学图像分割目前面临的机遇和挑战。

1 医学图像分割算法及PET/CT图像概述

图像分割是计算机视觉领域的经典任务之一,旨在对图像中的感兴趣区域进行划分。图像分割在医学图像中的应用则是对患者医学图像中的肿瘤病灶、组织、器官等区域进行自动分割,分割结果可用于辅助医生的临床诊断、辅助肿瘤患者放疗的靶区勾画、辅助患者的预后分析。

传统的图像分割方法包括阈值分割^[9]、模糊联通^[10]和聚类^[11]、随机游走^[12]等方法。其中,基于多模态医学

图像的传统分割方法近年来也有所更新迭代,取得一定进展。本文主要针对深度学习方法进行讨论。基于深度学习的医学图像分割方法已经成为医学图像分割领域的重要研究方向,其中,卷积神经网络(CNN)及其改进模型(如FCN^[13]、U-net等)对医学图像的分割性能得到了证实。在医学图像分割领域,U-net、3D-Unet^[14]和V-net^[15]是最受欢迎的网络,近年来,其简单高效的CNN架构和跳跃连接等设计理念得到诸多学者的扩展和改进^[16-18]。因此,本节对医学图像分割算法中的经典模型U-net网络进行简要介绍。并对PET/CT图像的意义、通信格式、成像特点进行概括。

1.1 U-net医学图像分割网络

U-net于2015年的MICCAI会议提出,最初应用于2015年的ISBI细胞追踪挑战赛并获得冠军^[8]。U-net结构主要由编码器-解码器结构(Encoder-Decoder)和跳跃连接组成。通过连续两次 3×3 大小的卷积核配合ReLU激活函数对图像进行卷积,再通过 2×2 最大池化操作进行降采样组成左侧编码器。右侧解码器的卷积核大小与编码器相同,经过上采样以及 2×2 卷积后,与左侧对应维度的特征图进行拼接(即跳跃连接),最后再连续进行两次 3×3 卷积和ReLU激活函数完成解码操作。最后使用 1×1 卷积将特征向量映射到目标分类数,在医学图像分割中通常为二分类。

U-net网络的结构形似字母U,并提出跳跃连接这一融合操作,很好地融合医学图像的位置特征(深层特征)和图像的细节特征(在医学图像中对应为病灶的细节特征)。这使得U-net非常契合医学图像的图像特点,成像中的人体结构和位置较为固定,病灶部位处于图像中的确定范围。同时,U-net被作为许多分割网络的基础对照网络(baseline),它的编码器-解码器结构和跳跃连接在多个医学图像分割网络中被沿用。

1.2 PET/CT图像简介

PET/CT结合了PET成像和CT成像的影像学检查方式。PET通过检测人体事先服用的同位素示踪剂在人体各部位代谢区域的信号浓度来获得图像。由于同位素在各部位的积聚程度不同,从而获得不同的信号强度,在图像上直接反映的是具有更“亮”的像素值,同理,信号强度弱的地方则显得更“暗”。CT通过检测X线经过不同物质衰减后的X射线光子数来获得断层成像,现有的CT成像仪器具有很高的分辨率和成像精度,通常用来观察病人的解剖结构,还可通过后期合成获得三维人体成像。如图1所示,图为某肺癌病人的某一帧PET/CT图像。第一行从左至右分别为PET图像的横断面、矢状面和冠状面,第二行为CT图像的横断面、矢状面和冠状面。PET图像中的较亮区域代表摄入同位素浓度较高,临床以SUV摄入值表示(Standardized

Uptake Value, SUV),该图中较亮区域为肺部肿瘤。PET图像的特点为分辨率较低,在图像上体现为像素较少,将图像放大时可观察到明显的模糊现象。CT图像分辨率较高,可以很好地观察各组织的细节和纹理特征,但肿瘤的边界和周围正常组织的区分度不明显,仅依靠人眼很难对病灶边缘进行准确识别,如图中红框处标出部分。因此,阅片过程中通常将CT图像与PET图像进行对比观察,并结合专业知识对病人做进一步的判断。

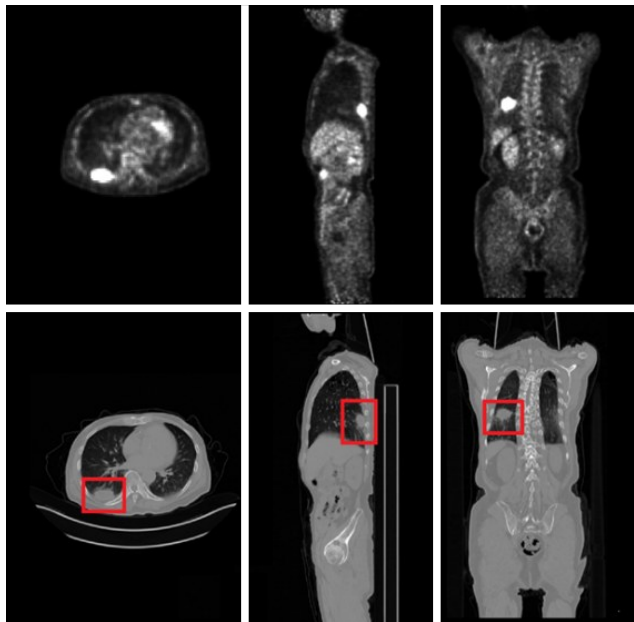


图1 PET/CT多模态医学图像

Figure 1 PET/CT multimodal medical images

PET-CT图像与计算机的通信保存过程中遵循DICOM标准^[19],即医学数字成像和通信标准。保存的文件以.dcm为后缀,一个病例由几百到几千个dcm文件组成,每个dcm文件保存了病人某一帧的像素信息以及详细的拍摄信息,如患者信息、中央位置信息、窗宽以及窗位信息、所用的仪器型号和仪器设置参数等。在深度学习模型的构建过程中,数据集的构建需要读取对应的图像数据。为了提取DICOM数据中的图像信息,通常需要借助对应的工具包对其进行处理。以Python计算机语言为例,可通过Pydicom工具包对dcm文件进行解码并读取,然后剔除文件中包含病人的隐私信息,符合可供研究的标准后,才读取对应的图像数据,并进行预处理、归一化等一系列数据集的构建操作。

2 基于深度学习的多模态医学图像分割研究

基于深度学习的医学图像分割在许多病灶分割、器官分割等任务中已经取得不错的成绩,但目前的卷积神经网络均基于单模态的图像进行研究,与使用多

模态图像的网络相比,分割性能仍有差距。利用多模态医学图像的分割任务相对较少,且侧重点不同。利用单模态图像的任务侧重于分割网络本身的架构与性能,而利用多模态数据的任务重点除了分割网络本身的架构与性能外,更侧重对多模态图像的互补特征进行融合,即多模态特征融合策略。卷积神经网络结构较为复杂,任务场景存在差异,所使用的数据类型各有异同。本节以多模态医学图像的特征融合策略作为主要划分依据,对该领域近4年的文献进行归纳,并重点讨论文献中的多模态特征融合策略问题,同时从实现方式、分割性能等维度对各方法的优缺点进行分析。此外,以分割的任务类型、所用网络模型(目标改进的模型)、所用数据类型作为辅助划分的依据,对多模态医学图像分割的文献进行归纳和阐述。

2.1 多模态特征融合策略的3种思路

如图2所示,在利用多模态数据的任务中,对现有文献的方法进行总结,多模态特征融合思路可划分为以下3个层面:(a)早期融合/浅层融合(Early Fusion, EF);(b)后期融合/决策级融合(Late Fusion, LF);(3)混合融合/超融合(Hybrid Fusion, HF)。

本文将多模态特征融合的思路结合分割模型的训练和求解过程进行分析和讨论。在求解最优分割模型的过程中,将数据的输入、模型的训练以式(1)进行描述,可得到如下目标函数的优化问题。给定某个病例数为 N 的一组PET-CT图像作为网络的输入,以目标函数 f_{seg} 代表分割模型,将数据输入分割网络 f_{seg} 进行训练,使其满足以下优化问题:

$$EF = \operatorname{argmin}_{f_{seg}} \left(\frac{1}{N} \sum_{i=1}^N L \{ f_{seg} [b_{fused} (x_{CT}, x_{PET})], y \} \right) \quad (1)$$

其中, b_{fused} 代表网络中的融合模块; x_{CT} 、 x_{PET} 分别代表输入某一病例的CT图像和PET图像; f_{seg} 为分割网络; L 代表网络中的Loss损失函数; y 为医生标注的金标准。如图2a所示,EF将 x_{CT} 、 x_{PET} 输入融合模块 b_{fused} 得到融合后的中间结果,再输入分割网络 f_{seg} 得到分割结果,将分割结果与金标准 y 输入损失函数Loss计算得到分割结果与金标准的损失值,最后通过计算所有病例损失的均值获得当前分割网络 f_{seg} 的性能。当分割结果与标注金标准的损失值越小,则代表分割结果与真实值差距越小,分割模型的性能越优。网络的训练迭代过程结束,选取分割性能最优的网络并保存网络中各层的卷积核参数,即为最后的分割网络。

$$LF = \operatorname{argmin}_{f_{seg}} \left(\frac{1}{N} \sum_{i=1}^N L \{ b_{fused} [f_{seg} (x_{CT}), f_{seg} (x_{PET})], y \} \right) \quad (2)$$

如图2b所示,LF先将 x_{CT} 、 x_{PET} 分别输入分割网络 f_{seg} 获得分割结果,再将分割后的结果通过融合模块 b_{fused}

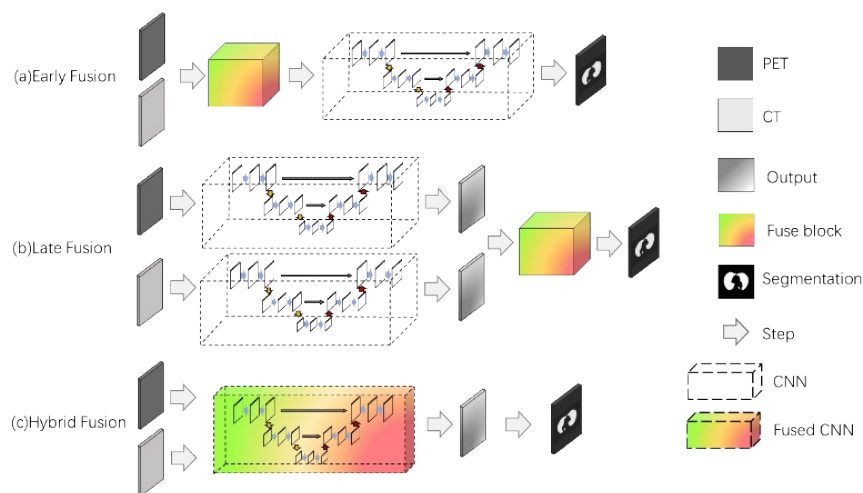


图2 多模态医学图像特征融合的3种思路

Figure 2 Three ideas for feature fusion of multimodal medical images

进行融合,然后将融合结果与医生标注的金标准 y 输入损失函数 L 计算损失,取均值后得到当前分割模型的平均分割性能。

$$HF = \operatorname{argmin}_{f_{\text{seg_fused}}} \left(\frac{1}{N} \sum_{i=1}^N L \left\{ \left[f_{\text{seg_fused}}(x_{\text{CT}}, x_{\text{PET}}) \right], y \right\} \right) \quad (3)$$

如图2c所示, HF直接将 x_{CT} 、 x_{PET} 输入具有融合功能的分割网络 $f_{\text{seg_fused}}$ 获得分割结果,再将分割结果与医生标注的金标准 y 输入损失函数 L 计算损失,取均值后得到当前模型的性能。

2.2 3种融合思路的比较和分析

EF的融合操作体现在数据输入分割网络之前进行,此时的图像具有丰富的细节特征,在这一层面的融合目标更侧重于为深度学习模型引入多模态信息。值得注意的是,如果通过传统图像融合的方式直接将多模态图像进行像素层面的融合得到一幅图像,再送入网络,会过多丢失原始图像所带来的丰富的细节信息,进而导致深度学习模型在浅层的卷积层对图像进行卷积时直接丢失高级特征。因此,为保留多模态图像的原始特征,采用多通道叠加的方式将多模态数据送入网络是最常用的手段,既保留了原始特征,也间接将多模态数据引入神经网络。具体的实现方式通常将多模态图像通过 concat 函数叠加为三维张量,再通过卷积操作将病人的CT图像和PET图像融合到二维空间,最后输入分割网络。

LF与EF的不同之处在于融合策略 b_{fused} 在整体网络架构中所处的位置不同。从式(2)可知, LF的输入是各模态分别输入分割网络后的输出结果,以分割后的结果作为函数的输入再进行融合操作。此时的分割结果对应为每个像素被划分为病灶的概率(对应为 True positive 的概率),以矩阵形式保存每个像素的概率。将

概率标准化为0~255的灰度值区间可以进行图像的可视化。若通过某个阈值进行处理后得到的图像则体现为黑白二值图(黑白掩膜),白色代表病灶区,黑色则为正常组织。此时的图像更为直观,携带的信息更多代表了病灶的位置和细节特征。在此基础上的LF通过一些决策算法或多层卷积模块将分割结果进行进一步的“精炼”,从而提高模型的分割性能。

HF侧重在分割网络中加入多尺度、多步骤的具有融合功能的操作对多模态图像的互补信息进行融合,在第3.3节中通过具体的方法进行详细讨论。

由于3种融合思路在网络的构建中并不冲突,一种分割方法可以同时包含3种融合的思路,不好直接度量3种思路的优劣性。因此,为了直观比较3种融合思路之间的差异,以及同一种融合思路中各种方法的优劣程度,本文第3节首先将体现各思路方法的文献进行归纳,并根据文献中不同融合思路的具体实现方式和对分割性能的提升进行阐述。然后,将该方法与其他仅使用单模态数据的网络以及使用相同主干网络下的其他方法进行对比,从方法的亮点、分割的性能、参数量的大小、方法的不足等维度对各方法进行归纳总结,展示了不同方法的优缺点。

3 3种融合思路的研究进展与讨论

3.1 EF融合思路

如图2a所示, EF泛指在模型早期阶段(即多模态图像输入网络之前)对多模态图像进行的一系列操作。图3为3种多通道叠加的方式举例。方式1为多通道的方式叠加。以图3a两个通道为例,通过将两个模态叠加到两个通道得到三维张量^[20-24], $512 \times 512 \times 2$ 的三维张量中,第一维与第二维对应512,代表某一病例一帧图

像的长度与宽度,即保存了第 n 帧的像素信息;第三维则是代表不同的成像模态,代表CT的像素信息与PET的像素信息。

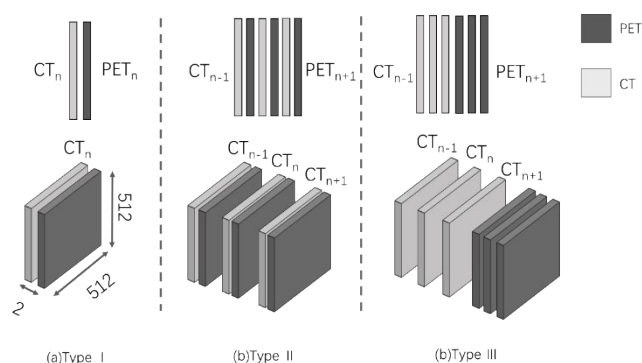


图3 多模态图像的3种叠加形式

Figure 3 Three superposition forms of multimodal images

Ren等^[25]将CT、PET、MRI 3种模态以不同的组合通过图3a多通道方式进行叠加并送入网络训练,并观察不同组合对分割模型的性能影响。通过对比CT+PET、CT+MRI、PET+MRI、CT+PET+MRI 4种组合,以Dice系数、豪斯多夫距离(Hausdorff, HD)和平均表面距离(Mean Surface Distance, MSD)作为分割的评价指标。CT-PET-MRI三模态组合效果最优,Dice评分为 0.74 ± 0.13 ;其次是CT+PET,Dice评分为 0.73 ± 0.13 ;CT+MRI效果最差,Dice评分为 0.58 ± 0.18 。由结果可知,随着多模态信息的引入,不同模态的贡献不同。MRI图像有反应 T_1W 、 T_2W 等序列,反映不同高级信息,直接通过叠加融合可能适得其反。PET图像携带了组织对放射造影剂的摄入信息,在这一叠加方式中的效果较好。

如图3b所示,Zhao等^[26]扩展了叠加的方式,以3个相邻的样本组作为一次输入,每个样本组由一个对应的PET切片和一个CT切片叠加而来,输入维度为 $512 \times 512 \times 6$ 。如图3c所示,Li等^[27]将三帧相邻的CT图像与三帧PET图像叠加作为输入,每次样本长度为6,与图3b相比,由于叠加的顺序有所改变,同一模态下的切片保留了原始的空间顺序,相较而言引入更多的互补信息和空间信息。

叠加的操作在代码实现的过程中对应为拼接或者堆叠操作。这种以多维张量的形式送入网络可以使得网络同时接受两种模态的信息,间接将特征进行融合,是现有的方法中最常用也是数量最多的融合方式。

EF的3种方式间接为模型引入了多模态医学图像的互补信息,同时可以最大限度地保留图像的原始信息,可复现性高、运行效率高。值得注意的是,由于叠加操作通常为无损叠加,使得“干扰”信息和“无用”的

信息也得以保留,会使输入的数据引入更多噪声和干扰特征。

3.2 LF融合思路

如图2b所示,LF泛指经过特征提取后的融合或将多模态分割的结果通过某种决策方式进行的融合。本小节对采用LF思路的文献进行阐述,首先对文献中网络的架构进行简要介绍,然后对文献中的融合思路以及实现过程进行阐述,若文献分割的病灶为多部位,则选取其一作为性能改进的参考。Zhao等^[28]提出级联卷积融合模块对高级特征进行融合。首先使用两个并行的V-net网络,将PET图像和CT图像分别输入特征提取模块,再将得到的特征图通过级联的卷积融合模块二次提取特征,通过softmax函数得到最后的分割结果。最终的Dice评分为85%,与使用单模态数据时的分割性能(仅使用CT:76%,仅使用PET:83%)以及多模态数据集的类似方法进行对比(OTSU:78%,FCM:81%)均取得更高的性能提升。Zhong等^[29]分别将PET-CT送入3D U-net获得各自的分割概率图,并提出一种优化后的基于图切割的方法将两种模态得到概率图进行融合获得最后的分割图^[30]。文献[26]在Zhong等^[29]基础上进行改进,提出一种模糊变分模型,将概率图和PET强度图像相结合。先将CT图像输入基于FCN的3D全卷积网络生成分割概率图,并将得到的概率图作为新设计的变分模型的隶属度先验,再将概率图和PET图像的肿瘤摄入强度信息作为模型的输入,实现准确的多模态肿瘤分割,同时提高了该方法的灵活性和稳定性^[31]。单模态数据和类似改进方法的比较见表1。

Zhao等^[26]提出3层2.5D的U-net来自动检测和分割局部前列腺癌骨转移和淋巴结转移。该网络以3个相邻的切片作为输入(排列方式见图3b),每个输入切片由一个对应的PET切片和一个CT切片组成两个通道作为输入来预测中间切片的结果,属于前期融合。值得注意的是,该网络将PET和CT图像中的横切面、冠状面和矢状面分别输入网络,通过多数投票策略获得最终的分割结果。该方法综合了EF和LF两种思路,同时分别提取冠状面、矢状面、横切面的图像数据输入分割网络对其进行处理和分割,从3个解剖面的角度更好地利用不同维度的空间信息,进一步辅助高级图像特征在深度学习中的迁移^[32],得到3种分割结果后,通过多数投票的策略对结果进行融合。即当某一像素,3种切面下被分割为阳性的数量 ≥ 2 时,这一像素的分割结果即为阳性。这使得最后的分割结果中被-划分为阳性的像素所占比例更多,导致真阳性率(True Positive)和假阳性率(False Positive)的提高,进而导致分割结果在模型评价的指标中对应数值的占比提高。在性能提升方面,以骨

表 1 不同 LF 方法比较

Table 1 Comparison among different methods for late fusion

作者	研究网络	研究数据	研究方法	单模态对比 Dice	多模态对比 Dice	Dice/%
Zhao 等 ^[28] , 2018 年	并行 V-net	84 例 PET-CT 非小细胞肺癌	级联的卷积融合模块+softmax	(CNN)CT: 76% (CNN)PET: 83%	W-net: 83% V-net: 81%	85.00
Zhong 等 ^[29] , 2018 年	3D U-net	32 例 PET-CT 非小细胞肺癌	优化的基于图切割的方法	-	Song: 62.4% Zhong: 78.1%	86.90
Li 等 ^[27] , 2020 年	基于 3D-FCN	84 例 PET-CT 非小细胞肺癌	用于特征融合的模糊变分模型	(FVM)CT: 76% (FVM)PET: 82%	W-net: 83% 3D-Unet: 82%	86.00
Zhao 等 ^[26] , 2020 年	3 层 2.5D U-net	193 例 PET-CT 前列腺癌骨转移	EF+LF (多数投票)	CT: 低于 10% PET: 51.5%	-	64.50
Fu 等 ^[33] , 2021 年	基于 U-net	101 例 PET-CT 非小细胞肺癌	多模态空间注意力模块 (MSAM)	(U-net)CT: 11.92% (U-net)PET: 69.23%	U-net: 68.22% ResNet50: 67.08%	71.44

转移分割为例。单独以 CT 图像进行输入时,得到的分割结果 Dice 评分均小于 10%,分割性能非常糟糕。单独以 PET 图像作为输入时,取得 51.5% 的 Dice 评分。以两种模态作为输入时,取得 64.5% 的 Dice 评分,显著优于单模态下的分割结果。

Fu 等^[33]提出一种多模态空间注意力模块(MSAM),MSAM 学习肿瘤空间区域相关的位置特征,并抑制 PET 图像中其他高 SUV 值的正常区域。主干网络由一个 U-net 网络和一个用于提取 PET 图像特征的子网络组成。首先将 PET 图像输入子网络得到携带了肿瘤位置信息的空间注意力图,然后将得到的空间注意力图的大小通过重采样(采用双线性插值)为主干网络后半部分的解码器跳跃连接的对应维度,再通过逐元素相乘来融合 PET 和 CT 图像的互补信息。

由于 PET 图像中,患者的非病灶区域(如心脏、膀胱、肠道)也会对显像剂具有一定的摄入,使得该部位所携带的特征信息具有一定干扰性。作者将 PET 图像输入 MSAM 模块获得空间注意力图,由于 MSAM 模块是由经过预训练的 U-net 网络构成,相当于对 PET 图像进行一次“预分割”,所以获得的空间注意力图可以有效剔除 PET 图像中的干扰区域(即心脏、膀胱的 SUV 摄入),同时保留了更多对于目标病灶有贡献的区域。然后,将空间注意力图通过重采样、逐像素相乘的跳跃连接来融入主干网络中,可以很好地提升 PET 图像携带的高级特征对主干网络的贡献,进而提升多模态图像的融合效果和模型的分割性能。将其与仅使用单模态数据集相比也取得了更优的性能。以非小细胞肺癌数据集为例,单独使用 CT 在 U-net 上的 Dice 分割精度为 11.92%,单独使用 PET 时为 69.23%;使用多模态 PET-CT 数据在 U-net 和 ResNet50 模型中的精度分别为 68.22% 和 67.08%。加入 MSAM 模块后使用多模态 PET-CT 图像的精度为 71.44%,见表 1。

3.3 HF 融合思路

如图 2c 所示,超融合/混合融合的思路泛指直接将多模态数据输入具有多模态特征融合功能的图像分割网络。该网络通常具有多尺度、多步骤的融合策略,以融合多模态图像的互补信息。表 2 展示了各方法的思路,基于单模态及多模态方法的对比。本文从各方法的主要亮点、研究的不足以及其他贡献 3 个方面对各研究方法进行归纳和总结,如表 3 所示。Kumar 等^[34]提出一种协同学习特征融合图在网络中对多模态图像进行融合,该方法在神经网络的每一个深度中(具体为卷积+池化后)进行融合,为每个深度的多模态图像提取融合图,再将各深度的融合图依次参与解码器的上采样和卷积层。

由于网络在各深度中均进行了融合操作,图 4 以网络中某一深度融合图的生成为例,融合过程如图 4 所示:(1)首先将 PET 和 CT 通过编码器(两次卷积+一次池化+降采样)得到各自的特征图;(2)然后将 PET 和 CT 的特征图进行通道上的连接(Concatenate)和 3D 卷积操作(卷积核大小:3×3×2),得到堆叠特征图和 3D 卷积后的多模态特征图(fusion map);(3)最后将二者进行逐元素相乘,得到当前深度下的特征融合图(co-learned fusion map)。

该方法引入 3D 卷积对特征进行融合,并通过逐元素相乘操作将特征图中的每个像素逐一进行加权。融合后得到的特征图在感兴趣区域的像素权重更高,在 PET 中非感兴趣区域但同样有高摄入部位(干扰部位)的权重降低。这使得特征图中的每一个像素表达了对病灶分割的贡献,同时削弱了 PET 图像中正常 SUV 摄入部位(例如膀胱尿液、肠道及部分人群的心脏摄入)对分割模型的干扰性,进一步强化 PET 图像中互补信息的表达,也削弱了 PET 在分割模型中的干扰信息。该方法在肺实质、纵膈肌和肿瘤等分割任务中进行验证。与其他单模态方法和多模态方法的比较如表 2 所

表2 不同HF方法比较

Table 2 Comparison among different methods for hybrid fusion

作者	研究网络	研究部位	研究方法	单模态方法 Dice	多模态方法 Dice	Dice/%
Kumar等 ^[34] , 2019年	基于CNN	50例PET-CT 非小细胞肺癌	提出一种协同学习特征融合图在 网络中多次融合互补信息	(Zhong)CT: 45.62% (Zhong)PET: 62.37%	MB: 52.16%, MC: 49.31%, FS: 53.07%	63.0
Zhong等 ^[35] , 2019年	3d U-net	60例PET-CT 非小细胞肺癌	双通道并行3D-Unet, 共享网络 同深度高级特征融合互补信息	(3D-Unet)CT: 81.1% (3D-Unet)PET: 79.4%	Jin: 75.9% Zhong: 78.3%	86.1
Jin等 ^[36] , 2021年	PSNN	148例PET-CT 食管癌	结合EF融合和LF融合的PHNN 改进的PSNN分割网络	(DenseUNet)CT: 70.3% (PHNN)CT: 74.3%	Zhong: 73.2% Kumar: 74.2%	79.0
Bi等 ^[22] , 2021年	改进FCN	120例PET-CT 非小细胞肺癌	提出递归融合网络, 通过递归调 用多模态上/下采样模块对多模 态图像进行分割	-	W-net: 53.4% Zhong: 63.9% Kumar: 74.2%	65.8
Cabrera等 ^[37] , 2021年	U-net	240例PET-CT 乳腺癌淋巴结	基于C-Tree的特征提取+双通道 U-net进行联合分割	(U-net)PET: 79.9%	(U-net)PETCT: 87.2%	89.4

表3 不同方法优缺点比较

Table 3 Comparison of advantages and disadvantages of different methods

作者	方法亮点	研究不足	其他亮点/贡献
Kumar等 ^[34] , 2019年	提出 co-learned 特征图来更好地强化多模态 互补信息的表达, 同时削弱 PET 图像中干扰 信息对分割模型的影响	模型较为复杂, 由于 co-learn 特征图需要多次 进行 3D 卷积和逐像素相乘, 使得模型参数量 大, 复现难度较高	提供了详细数据集制作步骤, 开源了代码
Zhong等 ^[35] , 2019年	双通道的设计可以使 PET 和 CT 共享互补信 息, 具有基于 PET 和基于 CT 的输出, 在结果 上保留了更多的多模态的互补信息	由于使用了过多的跳跃连接来保证互补信息 的贡献, 使得网络训练和分割的时间更长	该网络将两个输出在单模态和多 模态下与各模态下的分割结果进 行详细的对比, 更直观地观察各 方法带来的改进
Jin等 ^[36] , 2021年	提出少参数的 PSNN 分割网络, 网络结构简 单, 仅配合 EF 融合也能取得较高的精度, 相 比于 U-net 缩短三分之一的训练时间	由于不同模态图像均使用固定的 PSNN 进行 融合分割, 没有根据特定模态的特点对网络 进行优化	针对放疗靶区这一分割任务进 行深入研究, 并取得 82.6% 的 精度, 提高临床应用价值
Bi等 ^[22] , 2021年	通过将 CNN 中的上采样和下采样进行模块化 来显著提升各自的性能, 同时将具有融合功能 的操作模块化为 ILM。模块化的网络提高整体 的集成度、灵活性、降低网络内的耦合性	由于网络通过多次递归的方式调用多个模 块, 随着网络深度的提升, 可能导致运行效率 降低。代码未开源, 复现难度较高	通过实验来比较 EF、LF、HF 的 性能和特点
Cabrera等 ^[37] , 2021年	结合图像分析中的 C-Tree 方法将 PET 图像在 5 个维度上高效地提取特征, 并通过双输入网 络与 CT 图像进行融合分割, 有效提高淋巴结 转移的分割性能	由于引入了专门提取 PET 图像特征的 C-Tree 模型, 故转换到其他病灶的分割任务时需要 单独建模, 提高复现的成本	通过 3D map 的形式实现了 C-Tree 在多模态图像分割中的迁移应用, 深入分析 C-Tree 对 PET 图像提取 的 5 个维度特征对分割的贡献

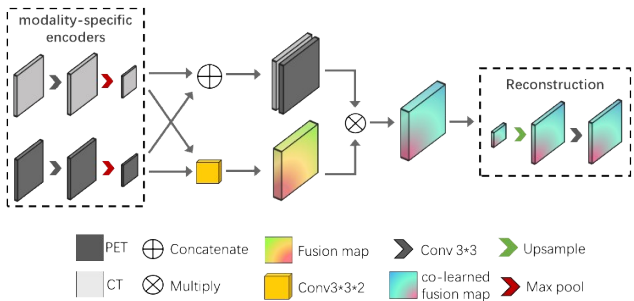


图4 协同学习特征融合图

Figure 4 Co-learning feature fusion map

示,对比了多分支方法(Multi-Branch, MB)、多通道方法(Multi-Channel, MC)和融合输入方法(Fused, FS)的结果。

Zhong等^[35]提出一种双通道并行3D-Unet,网络结构如图5所示,将多模态数据进行重采样+标准化后输入网络,对应特征通过跳跃连接与解码器中的特征进行拼接,再进行上采样+卷积的操作进行融合,实现特征的共同学习。由于该网络是由两个独立的编码-解码器分支组成,同步对图像进行运算,故

网络会得到两种输出:基于PET的分割结果和基于CT的分割结果。相对于其他单输出模型而言,该方法保留了更多的多模态信息供医生及研究人员进行比对。同时,该方法在两个并行的3D-Unet的跳跃连接中共享了另一模态同一深度的信息,使得对应的低、中、高层次的语义特征均得到共享,而不是将各

模态的信息在解码器中进行“融合”,提高信息的利用率,保留了多模态的互补信息。在分割性能方面,PET-CT联合共分割的Dice系数为86.1%,显著优于使用单模态数据的3D-Unet(Dice评分:CT为81.1%,PET为79.4%),以及同类共分割方法(Jin等^[36]Dice为75.9%,Zhong等^[35]Dice为78.3%)。

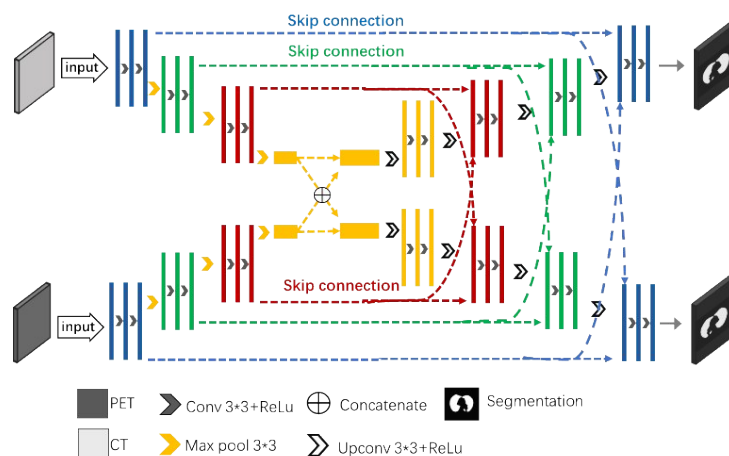


图5 双通道平行3D-Unet
Figure 5 Dual-channel parallel 3D-Unet

Jin等^[36]提出一种双流链式深度网络融合方法应用于食管癌的分割,将分割任务划分为肿瘤总体积(Gross Tumor Volume, GTV)和临床目标体积(Clinical Target Volume, CTV),并针对不同任务构建模型,为了对比多模态融合思路在分割任务中的比较,本文围绕GTV任务的分割性能进行讨论。作者提出一种渐进语义嵌套网络(Progressive Semantically-Nested Network, PSNN)用于辅助GTV的分割。PSNN是作者根据U-net和PHNN(Harrison, 2017)的特点改进而来。PHNN具有高效、少参数的特点,但它将网络中的低级特征反馈到高级特征,在医学图像分割中是不利的。U-net对医学图像分割具有很好的适配性,它将高层特征经过跳跃连接传递到底层空间,但解码器占用了大量参数,提高了模型复杂度。通过翻转PHNN的特征传播路径,使得高级特征可以传递到低层空间,同时保留PHNN的少参数模块,可以有效地结合两者的优点,得到PSNN。

PSNN的架构如图6所示,主要由两种卷积块、上采样函数、Sigmoid函数组成。卷积块1由两次连续3×3的3D卷积+一次BN+一次ReLU组成。卷积块2由3次3×3的3D卷积+一次BN+一次ReLU组成。⊕为逐元素相加。上采样为三线性插值,倍率从左至右依次为2、4、8,以保证每一层进行⊕时维度大小相同。每经过一个卷积块后对图像进行1×1×1卷积进行降维,再通过

过上采样与上一层进行相加。随着网络层数的加深,特征图由图像的细节特征递进为病灶的轮廓以及位置特征,保留了更多的深层信息并向下传递。

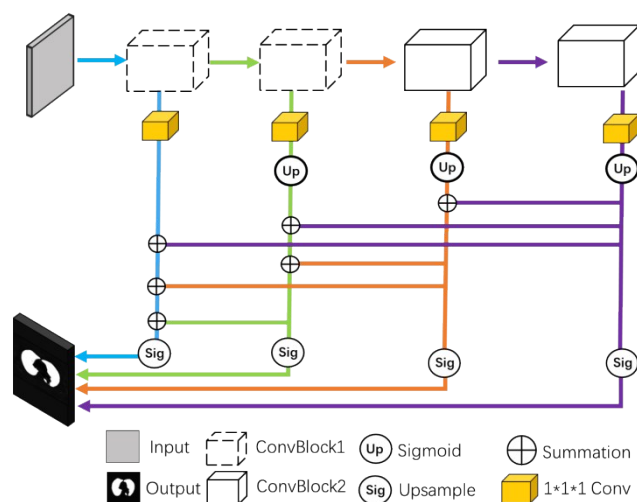


图6 PSNN渐进语义嵌套网络
Figure 6 Progressive semantically-nested network

分割网络架构如图7所示。第一次融合:将PET和CT图像进行通道上的拼接得到多维度张量,为模型引入多模态信息,随后输入PSNN模块进行分割得到分割图,思路为EF;第二次融合之前,将源CT图像单独输入

PSNN模块得到分割图,与第一次融合后的分割图进行拼接,然后输入PSNN进行第二次融合,获得最后的分割结果,思路为LF。该方法总共进行了两次融合操作,3次分割操作。最后得到的分割结果有效避免了基于PET的高对比度导致的分割不足效应,也规避了基于单独CT的低对比度导致的过度分割效应。

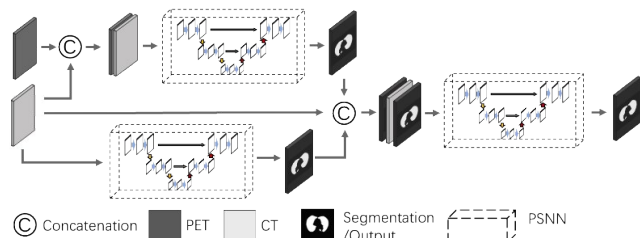


图7 双流链式深度网络融合方法

Figure 7 Two-stream chained deep network fusion method

该方法体现了EF以及LF的思路,同时,作者将PSNN在EF和EF+LF两种策略下的分割性能进行比较。与其他多模态分割方法相比(表2),仅使用PSNN配合第一次EF融合取得的Dice为77.8%,通过EF+LF的Dice为79.0%。该网络仅配合早EF(以图3a的融合方式)也取得了优于其他多模态网络的性能。然而,网络多次使用固定的PSNN架构来对多模态图像以及中间的分割结果进行处理,没有进一步对各融合阶段中不同模态下的图像特征进行优化,没有充分利用分割后的中间模态结果的互补信息。

Cabrera等^[37]根据成份树^[38](Component-tree, C-Tree)在血管造影^[39-40]和多模态图像分析^[41-42]的应用中受到启发,设计了基于C-Tree的双通道U-net架构对乳腺癌的颈部和腋下淋巴结进行分割。C-Tree在数学上是一种用于灰度图像的层次模型,该模型可以有效地用于评估图像中基于区域的高级特征。作者通过C-Tree提取PET图像特征后再形成特征图输入分割网络进行联合分割。

该网络主要由两部分组成:(1)基于C-Tree的特征提取模型。如图8所示,基于C-Tree的特征提取模型主要用于提取PET图像中基于区域的高级特征。处理流程又包括PET图像建模得到C-Tree、基于区域的特征描述、将C-Tree特征进行特征映射得到特征图3个步骤。C-Tree在数学上是一种用于灰度图像的层次模型。树中的每个节点对应为二维图像中的一个连通域,一个连通域对应多个像素。首先,根据两个约束将PET图像划分为若干个连通域:体素体积为5~40 mm³,连通域近似为球形。两个维度均代表了淋巴结的形状和大小特征。进行粗略划分后,得到一个候选节点的集合,再

通过5个维度对每个节点进行评估:a.节点的轮廓在PET上的平均梯度(G),PET图像中淋巴结的摄入相对于背景区域具有较高的梯度;b.节点在CT中的HU平均值(H),淋巴结区域的HU值为0~80;c.该节点HU值平均值的标准差(即H的标准差,作为S),淋巴结在CT上颜色更均匀,故该节点标准差越小,所反映的位置与淋巴结越“相似”;d.节点的相对积分体积($R = \int_N (I(x) - \min_N I) dx$), $I(x)$ 为点 x 在PET上的像素值, $\min_N I$ 为该区域内最小值。该公式反映了划分的节点区域与PET中淋巴结的匹配程度。当 $\min_N I$ 越小,差值越小,反映被划分的区域与PET中淋巴结的区域越“重合”;e.节点在肺内边界的位置(L),为bool型变量,当其位置处于肺部的边界时则为false,从而削弱了心脏等假阳性摄入对模型的贡献。

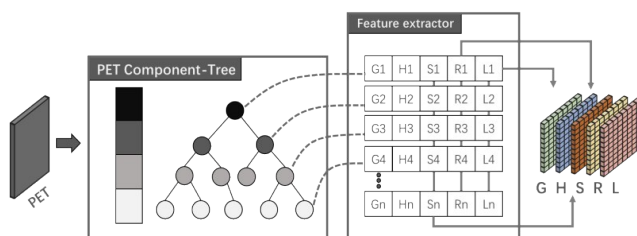


图8 基于C-Tree的特征提取模型

Figure 8 Feature extraction model based on C-Tree

C-Tree建模完成后,每个节点包含了以上5个维度的评价。由于分割模型的输入为逐像素组成的三维矩阵,所以需要进行区域到像素的特征映射,在此不展开叙述。从以上5个维度,PET图像被提取出5种特征图,再将特征图与PET和CT图像送入分割网络参与后续的训练,为网络提供额外的互补信息。

(2)基于U-net的双通道图像分割网络。如图9所示,该网络基于U-net改进,主要通过重建步骤中的叠加多模态图像结合上采样的方式对多模态特征进行融合,并通过跳跃连接多次将两个模态的浅层特征与深层特征进行融合。该网络在结构上与文献[35]相似,但区别在于文献[35]的网络在两个模态的解码器部分具有独立的分支,共享了另一模态下同一深度的高级特征,也获得两种输出,而文献[37]的网络在解码器部分将两个模态通过通道上的拼接进行融合,再进行上采样和卷积,从而使模型得到一种分割输出。

该方法尝试从图像分析的角度将C-Tree方法迁移至多模态融合分割模型,为多模态图像提取出更高级的特征,是一种新的思路。同时,提出3D map特征映射方法,为解决二维特征映射到三维特征图问题提供了

具体的实现思路。基于C-Tree的方法针对淋巴结在图像上的特点,从5个维度上进行建模,并验证每一维度的特征对分割性能的提升。然而,由于C-Tree的建模需要根据分割的目标病灶进行定制化的约束,与其他多模态融合分割方法相比,需要借助额外的临床经验对病灶特点制定约束条件,使其复现的成本增加,降低方法的可推广性。

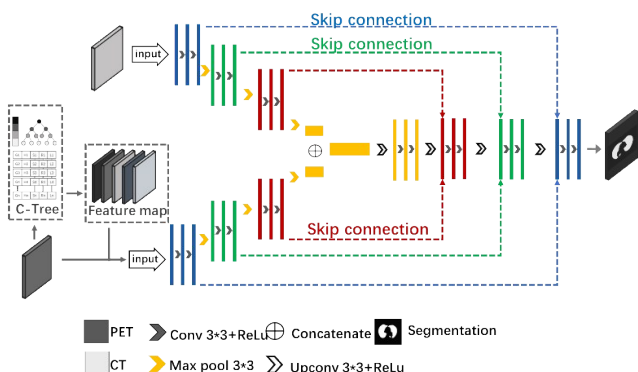


图9 基于双输入U-net的CNN结构

Figure 9 CNN architecture based on U-net with two inputs

Bi等^[22]提出一种递归融合网络用于肺部肿瘤的分割,该方法将卷积神经网络中的上采样层、下采样层进行改进,形成模块化的递归多模态上采样模块(Recurrent Multi-Modality Upsampling, RMU)以及递归多模态下采样模块(Recurrent Multi-Modality Downsampling, RMD),并提出互链路模块(Interconnect Link Modules, ILM)将多模态特征进行融合。多模态下采样模块RMD的主干网络是根据He等^[43]残差神经网络改进而来,如图10所示,RMD主要由4个残差块组成,每个残差块跟随一个跳跃连接,以保证RMD可以将不同维度的特征图与对应的ILM和RMU进行连接。RMD的主要作用是处理输入的PET/CT多模态图像并进行特征提取。每个残差块的输出具有不同的维度,每个残差块的输出为当前模态的中间结果。互链路模块ILM由卷积层、ReLU激活函数以及跳跃连接组成,它的输入为上一层的多模上采样模块RMU的中间结果,输出连接到下一层的多模降采样RMD进行特征提取。多模态上采样模块RMU由4个解码块Decoder组成。以一个融合周期 t 为例,它的3个输入分别为RMD处理后的基于PET的中间结果RMD(PET)、基于CT的结果RMD(CT),以及上一层RMU的输出。首先将PET和CT的多模态特征映射通过SUM逐像素求和进行融合,再通过多个卷积层对融合后的特征图进行处理,提取多模态图像特征。之后,上层RMU的输出结果 $Output_{t-1}$ 通过线性插值进行上采样,使其与多模态特征图具有相同的维

度,再将两者进行叠加,输入链式残差池化层^[44]。该模块通过3次残差的最大池化-卷积-叠加的操作来优化分割的结果,同时缩小了特征图的维度。

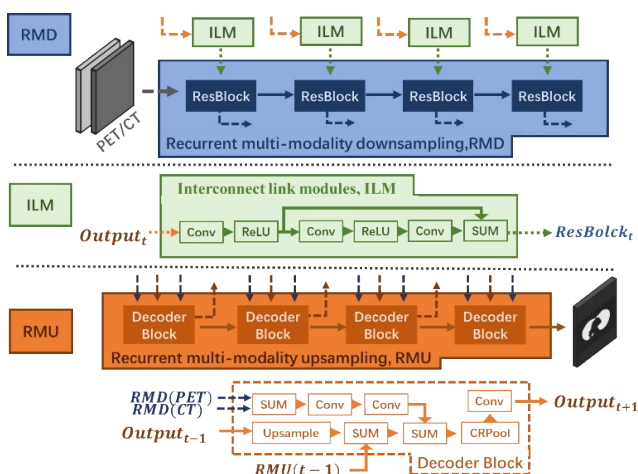


图10 RMD、RMU和ILM模块的处理流程

Figure 10 Workflows of RMD, RMU and ILM modules

如图11所示,分割网络由多个递归多模态降采样RMD、上采样RMU和互联链路模块ILM组成。每经过一轮RMD降采样和RMU上采样后都会通过互联模块ILM将当前尺度的特征图传递到下一层的RMD和RMU。每一层RMD得到基于PET和基于CT的特征图,将多模态特征图输入RMU进行融合并得到中间结果再由ILM进行进一步细化,连接到下一层的RMD进行特征提取。RMU将上一层的中间结果通过线性插值进行上采样,得到相同尺度的特征图,参与后续的步骤。通过递归调用以上3个模块进行多次融合-分割得到分割结果。该方法将卷积神经网络中的上采样层、下采样层从CNN网络中剥离出来,同时结合递归的思想和多模态图像的特点,将其模块化为递归多模态上采样层RMU和递归多模态下采样RMU来强化各层的编码-解码性能。同时将具有传递多模态信息功能的跳跃连接以及融合操作集成为单独的互联链路模块ILM,配合递归思想在网络中多次调用,在每个阶段中对多模态信息进行融合。模块化的网络能提高编码器和解码器的集成度,降低网络的耦合性,使网络可以通过灵活调度各个模块来达到预期设计的功能。但递归思想受限于内存堆栈和大量参数的传递,随着网络层数的加深,多次递归的操作可能会导致网络的运行效率大大降低。

3.4 3种融合思路的讨论

在3种融合思路中,EF是出现次数最多且最通用的思路。在第3.2及3.3节中列举的LF和HF的方法中,大多数都结合了EF的思路。早期阶段融合的实现方式简单而高效,只需要将图像在通道上进行

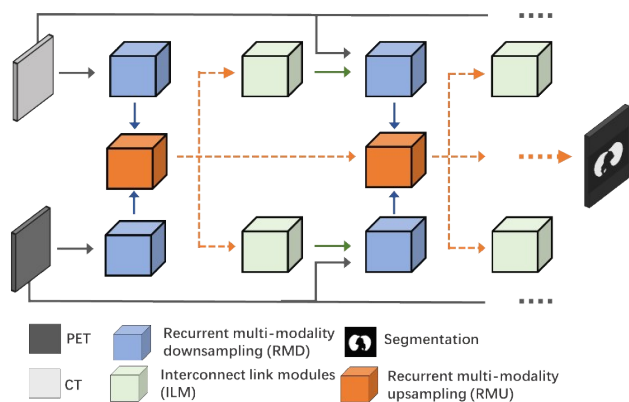


图 11 递归融合网络结构图

Figure 11 Recurrent fusion network

叠加即可。同时,考虑到卷积神经网络具有训练时间长、复现难度高等特点,EF思路凭借高效的融合方式在3种思路中体现出一定的优越性。此外,EF的作用更多是为网络引入多模态特征,在PET-CT中是为图像引入病灶的位置信息,有利于帮助模型对病灶进行定位。但是,这也导致更多的高级信息可能得不到有效的利用,也引入了更多的假阳性干扰信息,例如在PET图像中的心脏、消化系统及膀胱的正常摄入。因此,模型往往需要引入其他处理步骤来弱化或抵消这些干扰信息。例如Kumar等^[34]提出的co-learned特征融合图通过“权重”的方式有效弱化了PET图像带来的干扰信息。

在LF方法中,每个模态的特征在经过编码-解码器之后、网络输出之前的阶段对特征图进行细化或加入决策算法将多种模态的结果进行“合并”。这种思路有利于在粗分割的结果上更精确地确定分割病灶的边界。同时,得益于经过编码器-解码器这一“粗分割”的过程,分割后得到的特征图也在一定程度上改善了PET图像携带的干扰信息的问题。但由于LF属于多对一的有损融合,势必会导致最终的分割结果丢失一部分多模态互补信息。

与LF思路相比,HF方法融合了多尺度的多模态图像特征,从而保留了有用的图像特征,弱化了干扰特征对模型的贡献,可以很好地结合以上两种思路的优点,具有更优的分割性能。但为了更好地处理多模态图像的互补信息,排除干扰特征,模型不可避免地引入了更多的特征提取和处理的步骤,使得模型的结构更为复杂、模型参数量更庞大,导致方法的可复现难度高、推广性较差。部分模型的方法可能针对某一分割任务有更高的效率,如文献^[37]的方法,该方法针对癌症的淋巴结转移具有较高的分割性能(归功于C-Tree建模中根据淋巴结的特点进行了多约束),但缺陷是将其应用于其他的病灶分割任

务时需要单独确定C-Tree的约束以及对C-Tree建模,提高了该方法的迁移应用成本,降低可复现性。

4 结语

本文首先对常用的医学图像分割算法和多模态医学图像进行举例介绍,然后针对多模态医学图像的互补信息在深度学习模型中的融合策略问题进行讨论,最后根据3种融合思路以及对应思路中各方法的优缺点进行分析。最先进的方法有效地利用了多模态数据,在多模态图像的分割场景应用中取得更先进的性能。然而,目前的多模态融合策略存在以下值得探讨和深入的问题:(1)数据问题:多模态医学图像数据具有数据量大、获取难度高、标注难度大、数据异质性大等特点。而深度学习模型是数据驱动的方法,良好的开源生态可以有效提高研究的效率,促进研究的更新迭代,帮助多模态分割取得进一步进展。在单模态数据开源生态中,单模态开源数据集种类多,数据量大,每年都有会议和竞赛提供开源数据集来供各位学者进行研究和探索。而在现有的多模态医学图像研究当中,大多数研究均通过合作的医院进行多模态数据集的采集,开源数据少。因此,多模态数据的开源问题不可忽视。此外,在单模态数据中的图像病灶标注问题在多模态数据集中仍然存在。相对单模态数据而言,多模态图像的互补信息能有效帮助分割模型取得更优性能,但对图像的精度要求、图像采集的要求也更高,也使得标注难度更大,花费的时间成本更高,可以借助半自动标注方法、少样本训练以及数据扩充技术来缓解这一问题。(2)多模态特征融合策略问题:在PET-CT多模态特征融合策略中,采用EF思路的研究数量最多,而LF融合以及基于模型本身的混合融合的挖掘较少。EF融合效率最高,可以直接为模型引入多模态信息,但方式较为单一。目前较为先进的方法均将多模态图像特征在网络中多次融合,对多模态图像进行多层次、多尺度的融合操作,但可供参考的思路和策略较少。因此,多模态特征融合策略存在探索的空间。(3)优化问题:由于多模态数据的引入,势必会引入多分支、多模块来融合多模态的互补信息,使得网络内卷积层、池化层、上采样层的冗余,进而导致算法模型逐渐复杂化,模型的计算开销和内存需求提高,得到的模型权重参数更多,体积更大,加大模型落地部署的难度。因此,如何在保证模型性能的前提下提高训练的收敛速度、压缩模型的参数大小并获得更高效的多模态分割网络模型是值得进一步深入的研究方向。

【参考文献】

- [1] Bibault JE, Giraud P, Burgun A. Big data and machine learning in

- radiation oncology: state of the art and future prospects[J]. *Cancer Lett*, 2016, 382(1): 110-117.
- [2] Fonti R, Conson M, Del Vecchio S. PET/CT in radiation oncology[J]. *Semin Oncol*, 2019, 46(3): 202-209.
 - [3] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis[J]. *Med Image Anal*, 2017, 42(9): 60-88.
 - [4] Chen L, Bentley P, Mori K, et al. DRINet for medical image segmentation[J]. *IEEE Trans Med Imaging*, 2018, 37(11): 2453-2462.
 - [5] Dogan RO, Doan H, Bayrak C, et al. A two-phase approach using mask R-CNN and 3D U-Net for high-accuracy automatic segmentation of pancreas in CT imaging[J]. *Comput Meth Prog Bio*, 2021, 207(1): 106141.
 - [6] Kumar A, Kim J, Lyndon D, et al. An ensemble of fine-tuned convolutional neural networks for medical image classification[J]. *IEEE J Biomed Health*, 2017, 21(1): 31-40.
 - [7] Hossain S, Najeeb S, Shahriyar A, et al. A pipeline for lung tumor detection and segmentation from CT scans using dilated convolutional neural networks [C]//ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 1348-1352.
 - [8] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention(MICCAI). 2015: 234-241.
 - [9] Leibfarth S, Eckert F, Welz S, et al. Automatic delineation of tumor volumes by co-segmentation of combined PET/MR data[J]. *Phys Med Biol*, 2015, 60(14): 5399-5412.
 - [10] Sbei A, Elbedoui K, Barhoumi W, et al. Gradient-based generation of intermediate images for heterogeneous tumor segmentation within hybrid PET/MRI scans[J]. *Comput Biol Med*, 2020, 119: 103669.
 - [11] Hansen S, Kuttner S, Kampffmeyer M, et al. Unsupervised supervoxel-based lung tumor segmentation across patient scans in hybrid PET/MRI [J]. *Expert Syst Appl*, 2021, 167: 114244.
 - [12] Bagci U, Udupa JK, Mendhiratta N, et al. Joint segmentation of anatomical and functional images: applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images[J]. *Med Image Anal*, 2013, 17(8): 929-945.
 - [13] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. *IEEE Trans Pattern Anal*, 2016, 39(4): 640-651.
 - [14] Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2016: 424-432.
 - [15] Milletari F, Navab N, Ahmadi SA. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3d Vision (3DV). IEEE, 2016: 565-571.
 - [16] Zhou Z, Siddiquee M, Tajbakhsh N, et al. UNet++: redesigning skip connections to exploit multiscale features in image segmentation[J]. *IEEE Trans Med Imaging*, 2020, 39(6): 1856-1867.
 - [17] Ibtehaz N, Rahman MS. MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation [J]. *Neural Netw*, 2020, 121: 74-87.
 - [18] Zhou Y, Huang W, Dong P, et al. D-UNet: a dimension-fusion U shape network for chronic stroke lesion segmentation[J]. *IEEE Acm Trans Comput Bi*, 2019, 18(3): 940-950.
 - [19] Pianykh OS. Digital imaging and communications in medicine (DICOM)[M]. Berlin: Springer, 2012.
 - [20] Lim H, Dewaraja YK, Fessler JA. Joint low-count PET/CT segmentation and reconstruction with paired variational neural networks[C]//Medical Imaging 2020: Physics of Medical Imaging. International Society for Optics and Photonics. 2020.
 - [21] Zhao L, Lu Z, Jiang J, et al. Automatic nasopharyngeal carcinoma segmentation using fully convolutional networks with auxiliary paths on dual-modality PET-CT images[J]. *J Digit Imaging*, 2019, 32(3): 462-470.
 - [22] Bi L, Fulham M, Li N, et al. Recurrent feature fusion learning for multi-modality PET-CT tumor segmentation[J]. *Comput Meth Prog Bio*, 2021, 203: 106043.
 - [23] Huang B, Chen Z, Wu PM, et al. Fully automated delineation of gross tumor volume for head and neck cancer on PET-CT using deep learning: a dual-center study[J]. *Contrast Media Mol Imaging*, 2018, 2018: 8923028.
 - [24] Jemaa S, Fredrickson J, Carano RA, et al. Tumor segmentation and feature extraction from whole-body FDG-PET/CT using cascaded 2D and 3D convolutional neural networks[J]. *J Digit Imaging*, 2020, 33(4): 888-894.
 - [25] Ren J, Eriksen JG, Nijkamp J, et al. Comparing different CT, PET and MRI multi-modality image combinations for deep learning-based head and neck tumor segmentation[J]. *Acta Oncol*, 2021, 60(11): 1399-1406.
 - [26] Zhao Y, Gafita A, Vollnberg B, et al. Deep neural network for automatic characterization of lesions on 68Ga-PSMA-11 PET/CT[J]. *Eur J Nucl Med Mol Imaging*, 2020, 47(3): 603-613.
 - [27] Li L, Zhao X, Lu W, et al. Deep learning for variational multimodality tumor segmentation in PET/CT[J]. *Neurocomputing*, 2020, 392: 277-295.
 - [28] Zhao X, Li L, Lu W, et al. Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network[J]. *Phys Med Biol*, 2018, 64(1): 015011.
 - [29] Zhong Z, Kim Y, Zhou L, et al. 3D fully convolutional networks for co-segmentation of tumors on PET-CT images[C]//2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018: 228-231.
 - [30] Song Q, Bai J, Han D, et al. Optimal co-segmentation of tumor in PET-CT images with context information[J]. *IEEE Trans Med Imaging*, 2013, 32(9): 1685-1697.
 - [31] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. *IEEE Trans Pattern Anal*, 2016, 39(4): 640-651.
 - [32] Ghafoorian M, Mehrtash A, Kapur T, et al. Transfer learning for domain adaptation in MRI: application in brain lesion segmentation [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2017: 20th International Conference. Springer International Publishing, 2017: 516-524.
 - [33] Fu X, Bi L, Kumar A, et al. Multimodal spatial attention module for targeting multimodal PET-CT lung tumor segmentation[J]. *IEEE J Biomed Health*, 2021, 25(9): 3507-3516.
 - [34] Kumar A, Fulham M, Feng D, et al. Co-learning feature fusion maps from PET-CT images of lung cancer[J]. *IEEE Trans Med Imaging*, 2019, 39(1): 204-217.
 - [35] Zhong Z, Kim Y, Plichta K, et al. Simultaneous cosegmentation of tumors in PET-CT images using deep fully convolutional networks[J]. *Med Phys*, 2019, 46(2): 619-633.
 - [36] Jin D, Guo D, Ho TY, et al. DeepTarget: gross tumor and clinical target volume segmentation in esophageal cancer radiotherapy[J]. *Med Image Anal*, 2021, 68: 101909.
 - [37] Cabrera DL, Gogin N, Morland D, et al. Segmentation of axillary and supraclavicular tumoral lymph nodes in PET/CT: a hybrid CNN/component-tree approach[C]//2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021: 6672-6679.
 - [38] Salembier P, Oliveras A. Antiextensive connected operators for image and sequence processing[J]. *IEEE Trans Image Process*, 1998, 7(4): 555.
 - [39] Wilkinson M, Westenberg MA. Shape preserving filament enhancement filtering [C]// Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer-Verlag, 2002.
 - [40] Dufour A, Tankyevych O, Naegel B, et al. Filtering and segmentation of 3D angiographic data: advances based on mathematical morphology [J]. *Med Image Anal*, 2013, 17(2): 147-164.
 - [41] Padilla F, Romaniuk B, Naegel B, et al. Matching filtering by region-based attributes on hierarchical structures for image co-segmentation [C]//2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.
 - [42] Grossiord L, Passat N, Talbot H, et al. Shaping for PET image analysis [J]. *Pattern Recogn Lett*, 2020, 131: 307-313.
 - [43] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
 - [44] Lin G, Liu F, Milan A, et al. Refinenet: multi-path refinement networks for dense prediction[J]. *IEEE Trans Pattern Anal*, 2019, 42(5): 1228-1242.

(编辑:陈丽霞)