

引文格式:

朱雅琳, 陈宇骞, 常青玲, 等. 基于改进 U-Net 的 PET-CT 双模态头颈部肿瘤分割 [J]. 集成技术, 2023, 12(3): 94-104.
Zhu YL, Chen YQ, Chang QL, et al. ISA-DUNet: inception spatial-attention dense U-Net for head and neck tumor segmentation in PET-CT [J]. Journal of Integration Technology, 2023, 12(3): 94-104.

基于改进 U-Net 的 PET-CT 双模态头颈部肿瘤分割

朱雅琳 陈宇骞 常青玲 陈 涛*

(五邑大学智能制造学部 江门 529000)

摘要 放射肿瘤学领域中, 头颈部被认为是轮廓绘制最困难且最耗时的疾病部位之一。目前, 在临床中头颈部轮廓绘制通常由人工完成, 耗时且费力, 因此, 开发一种自动的医学图像分割方法十分必要, 可在节省人力和时间的同时, 避免由于不同医师的主观因素导致的诊断结果差异。该文使用正电子发射断层成像/计算机断层扫描双模态的数据对头颈部的肿瘤进行分割, 利用不同模态之间的信息互补, 实现了更加精确的分割。整个网络基于传统 U-Net 架构, 在编码器中增加 Inception 模块, 在解码器中增加 Dense 模块以及空间注意力来对网络进行改进。在头颈部肿瘤数据集上, 将该模型与不同的 U-Net 架构进行对比, 结果表明, 该文提出的方法多个指标均较优。该网络的 Dice 相似度系数为 0.782, 召回率为 0.846, Jaccard 系数为 0.675, 较原始 U-Net 分别提升 6.8%、13.4% 和 9.8%; 95% 豪斯多夫距离为 5.661, 较原始 U-Net 下降了 1.616。对比实验结果表明, 该文提出的 Inception Spatial-Attention Dense U-Net 分割模型在头颈部肿瘤数据集上有效改善了分割结果, 较标准的 U-Net 具有更高的性能, 提高了分割的准确率。

关键词 头颈部肿瘤; 多模态图像分割; PET-CT 图像; U-Net

中图分类号 TP 183 **文献标志码** A **doi:** 10.12146/j.issn.2095-3135.20230119001

ISA-DUNet: Inception Spatial-Attention Dense U-Net for Head and Neck Tumor Segmentation in PET-CT

ZHU Yalin CHEN Yuqian CHANG Qingling CHEN Tao*

(Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen 529000, China)

*Corresponding Author: chentao@wyu.edu.cn

Abstract In radiation oncology, it is usually difficult and time-consuming to manually profile the targets in the head and neck. Therefore, it is very necessary to develop an automatic medical image segmentation method, which not only saves time and energy, but also avoids the subjective variations among different

收稿日期: 2023-01-19 修回日期: 2023-02-19

基金项目: 广东省 2019 年省拨高建“冲补强”专项项目(5041700175); 教育部第二批新工科研究与实践项目(E-RGZN20201036)

作者简介: 朱雅琳, 硕士研究生, 研究方向为医学图像处理; 陈宇骞, 硕士研究生, 研究方向为人工智能、生物信号分析; 常青玲, 博士, 副教授, 研究方向为人工智能; 陈涛(通讯作者), 博士, 讲师, 研究方向为人工智能、知识推理, E-mail: chentao@wyu.edu.cn。

physicians. In this work, we used positron emission computed tomography and computed tomography image data to segment head and neck tumors, and realized more accurate segmentation by using the complementary information between them. The network was developed based on the U-Net architecture, and an inception module was added into the encoder module. In addition, dense modules and spatial attention are added to the decoder to improve the network performance. Experimental results show that our method outperforms the other U-Net networks. Quantitatively, the dice similarity coefficient, recall rate and Jaccard similarity coefficient are found to be 0.782, 0.846 and 0.675, respectively. Compared with the original U-Net, these results corresponds to an improvement by 6.8%, 13.4% and 9.8%, respectively. The 95% Hausdorff distance is found to be 5.661, which is 1.616 smaller than the original U-Net. In conclusion, this study demonstrates that the inception spatial-attention dense U-Net model can effectively improve the segmentation accuracy on the head and neck tumor PET-CT images.

Keywords head and neck tumor; multimodal image segmentation; PET-CT image; U-Net

Funding This work is supported by High-level University Construction Special Project of Guangdong Province in 2019, China (5041700175) and the Second Batch of New Engineering Research and Practice Project of Ministry of Education, China (E-RGZN20201036)

1 引言

放射治疗是现代癌症治疗的一种重要手段, 精确分割肿瘤区域对于放疗方案的设计和实施至关重要。目前, 临幊上主要采用医生手动标注肿瘤位置的方式, 极为耗时耗力, 且对医生技术要求较高。计算机辅助的放射治疗一直是国内外研究的热点, 其通过自动化的方式, 将医学图像中具有医学判断及应用价值的部分准确提取出来, 不仅能协助医生从多角度、多形式观察和诊断病变, 还能为放射治疗方案的制定提供更有价值的信息。基于深度学习的自动肿瘤分割对于指导临幊诊断治疗具有重要价值, 可以提高术前预判、术中监测和术后评价的准确率^[1]。本文提出的模型进一步提高了对头颈部肿瘤的识别精度, 有助于制定更完善的手术方案, 提高肿瘤手术的成功率, 因此, 研究精准自动化的分割技术对临幊工作具有重要意义。

成像是放射治疗的必要步骤, 主要用于区

分肿瘤与正常的组织结构, 辅助医生更准确地确定治疗方案和剂量, 避免放射治疗对周围正常组织产生影响。不同的成像技术如核磁共振成像(magnetic resonance imaging, MRI)、计算机断层扫描(computed tomography, CT)和正电子发射断层成像(positron emission computed tomography, PET)都可用于癌症的诊断和分期^[2-3], 不同的成像技术可为肿瘤诊断提供互补的信息, 临幊上一般结合多种成像方式进行诊断。Ren 等^[4]在头颈部肿瘤分割方面, 结合分割技术, 对基于深度学习的不同模态图像进行了研究, 实验表明, 与其他模式相比, CT 与 PET 结合的分割方式表现较好。临幊上放射治疗的剂量计算主要基于 CT^[5], 通常利用 CT 确定正常组织位置和边界的解剖结构信息, 但其对于肿瘤等软组织区域的区分能力较弱; PET 能够通过分子成像的方式提供早期肿瘤的代谢信息^[6-7], 但 PET 图像分辨率较低, 不利于刻画肿瘤的边界信息。单独的 PET 和 CT 成像都难以提供肿瘤的

精准信息, PET-CT 双模态成像则结合了 PET 功能成像和 CT 结构成像的优势^[8], 通过 PET 的高灵敏度精准定位肿瘤, CT 的高分辨率刻画肿瘤边界, 提高了肿瘤分割的准确性。传统的自动化肿瘤分割方法主要是基于阈值、边界和区域的分割^[9], 基于阈值的方法通过灰度直方图将肿瘤与背景进行分割, 但对于边缘的分割效果较差; 基于边界的分割对于边缘灰度值差异较大的图像、噪声较小的图像分割效果较好, 但肿瘤区域与正常组织之间的表面过渡平滑, 容易导致分割失败; 基于区域的分割通常根据不同颜色和纹理进行分割, 但 CT 对于肿瘤等软组织区分能力较弱, PET 图像则对比度较低, 分割效果也较差。此后一系列基于概率统计的分割方法也缺乏来自相邻像素的上下文信息, 缺乏图像的空间信息。

深度学习的发展为上述问题提供了解决思路, 在医学图像分割领域, 将不同的神经网络应用于分割任务均取得了显著的成果。多个隐藏层组成的神经网络擅长学习具有多层次、抽象的数据表示, 能够发现高维数据中的复杂结构。2015 年, Ronneberger 等^[10]提出了 U-Net 结构, 因其分割效果较好、结构易扩展, 引起了广泛的研究和关注。Zhang 等^[11]在 U-Net 的基础上引入残差模块, 提出了 ResU-Net, ResU-Net 在 U-Net 的基础上简化了网络的训练, 减少 3/4 参数量的同时, 性能较 U-Net 提升了 1.5%。3D U-Net^[12]将 U-Net 原来的 2D 卷积、池化、上采样操作替换为对应的 3D 操作, 并加入 Batch Normalization 层, 实现了对三维医学图像的直接分割。V-Net^[13]是在残差基础上对 U-Net 的另一种三维实现, 上述对 U-Net 的改进进一步提升了神经网络的分割性能。

头颈癌是全球第六大常见癌症^[14], 其原发部位和病理类型居全球肿瘤之首。在基于深度学习的 PET-CT 头颈部肿瘤分割方面, Xie 等^[15]在微调的 3D nnU-Net 模型的基础上加入压缩与激

励(squeeze-and-excitation, SE)模块, 该方法以 0.778 的 Dice 相似度系数(Dice similarity coefficient, DSC)以及 3.088 的 95% 豪斯多夫距离(95% Hausdorff distance, HD95)获得了 MICCAI2021 头颈部肿瘤分割赛冠军。An 等^[16]提出一种由粗到细的训练框架, 用残差块训练了 3 个 3D U-Net 从粗到细进行预测, 第 1 个用于粗略分割肿瘤选择边界框, 第 2 个对较小的边界框执行更精细的分割, 第 3 个将 PET、CT 与先前分割的图像拼接后作为输入, 以细化预测, 该训练框架可逐步提高分割质量。Qayyum 等^[17]提出了一种 3D-Inception-ResNet 模型, 该模型在编码器端加入 3D Inception 及 SE 模块, 在解码器端加入 3D ResNet 模块, 以提升分割性能, Inception 模块使得网络结构更深更宽, 有能力学习更多的特征, SE 模块对输入特征图进行通道特征加强, ResNet 模块解决了梯度消失效应。

受上述不同网络结构的启发^[18], 本文提出一种基于 U-Net 的多模态头颈部肿瘤分割网络。与传统的 U-Net 结构相比, 本文提出的 ISA-DUNet 网络进行了如下改进: (1)为实现对特征的深层提取, 在编码器端加入 Inception 模块, 通过不同尺寸的卷积核, 实现不同深度的信息提取; (2)通过在解码器端加入空间注意力机制, 提高学习过程中肿瘤的分割权重, 实现对肿瘤区域更精确的分割; (3)在深层网络的训练中, 为避免梯度消失问题, 在解码器端加入 Dense 模块及深监督模块, 减少参数量的同时, 扩大对肿瘤区域的关注范围; (4)将原始的 2D 操作全部替换为 3D 操作, 充分提取肿瘤区域的空间结构信息, 以达到更好的三维医学图像分割效果。

2 材料与方法

2.1 数据预处理

本实验数据来自 MICCAI2021 头颈部肿瘤

分割挑战赛上的数据集^[19], 该数据集包括来自 5 个中心的 224 个病例, 每个病例都包含 CT、PET 图像以及由医生人工标记的肿瘤区域信息, 原始 CT 图像尺寸为 $512 \text{ px} \times 512 \text{ px} \times 91 \text{ px}$, 体素大小为 $0.997 \text{ mm} \times 0.997 \text{ mm} \times 3.27 \text{ mm}$, PET 图像尺寸为 $128 \text{ px} \times 128 \text{ px} \times 91 \text{ px}$, 体素大小为 $3.516 \text{ mm} \times 3.516 \text{ mm} \times 3.27 \text{ mm}$ 。将 224 个病例的图像数据进行随机划分, 80% 的数据(180 个病例)作为训练集, 20% 的数据作为测试集。

实验开始前, 对图像数据进行预处理使训练数据更适应网络的输入, 有助于加速网络收敛, 更好地进行特征提取。由于两种模态数据的图像尺寸及空间像素不一致, 需要对图像进行重采样和裁剪处理, 利用三重线性插值法, 将所有的 PET 和 CT 图像的空间像素大小重新采样至 $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$, 图像尺寸裁剪至 $144 \text{ px} \times 144 \text{ px} \times 144 \text{ px}$ 。对于 CT 图像, 将 HU 强度限制在 $[-1024, 1024]$ 的范围内, 并映射到 $[-1, 1]$ 之间, 由于 PET 图像没有标准的图像强度值, 本实验统一用 z-score 归一化所有的 PET 图像。

2.2 Inception 模块

2014 年, Inception 模块的概念在 GoogleNet^[20]

中被首次提出, 该模块的主要目的在于构建更深层、更宽的网络的同时, 保持计算成本的不变。在深度学习中, 更深层、更宽的网络会带来网络性能的提升, 但庞大的参数量易导致过拟合, 增加计算负担。Inception 模块通过稀疏连接解决了这两个问题, 其将不同的卷积块并联结合, 将不同分支处理后的结果进行拼接, 在形成更深层、更宽的网络的同时, 防止了过拟合现象的发生。GoogleNet^[20]的所有操作均基于二维图像, 本文将其扩展到三维图像。

本文 Inception 模块结构如图 1 所示, 其由 4 个不同的卷积分支并联而成, 从左到右的第 1 个分支采用一个 $1 \times 1 \times 1$ 的卷积核, 输入通道为 64, 输出通道为 16, 所有的归一化方法均采用 InstanceNorm3d, 所有的激活函数均使用 LeakyReLU; 第 2 个分支在 $1 \times 1 \times 1$ 的卷积块之后连接一个 $5 \times 5 \times 5$ 的卷积块, 然后池化操作, 为确保所有分支输出通道大小相同, 将池化卷积核大小设置为 2; 第 3 个分支将 $5 \times 5 \times 5$ 的卷积块替换成两个 $3 \times 3 \times 3$ 的卷积块, 卷积块尺寸不同将导致采用的感受野大小不同, 可以提取到不同的特征, 再经过拼接操作可将不同尺度的特征

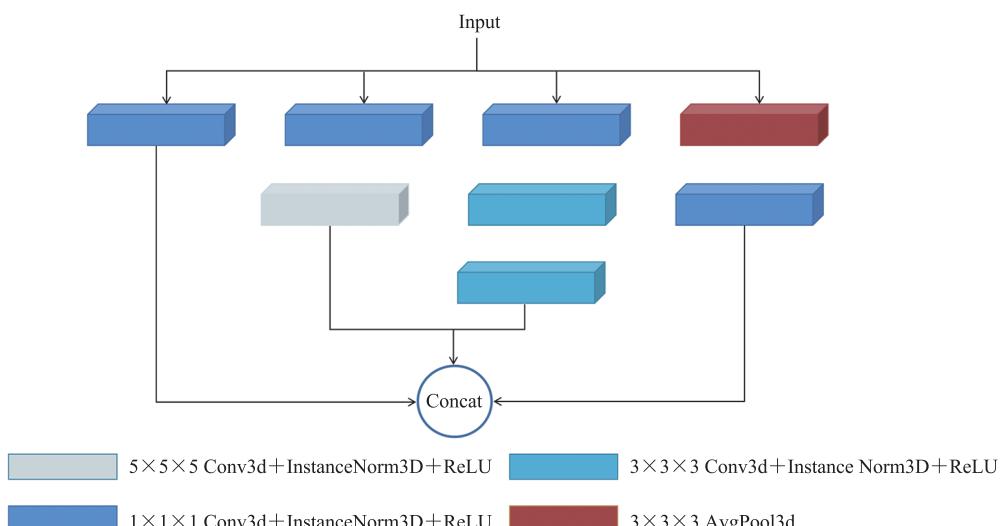


图 1 Inception 模块结构图

Fig. 1 Inception module structure

进行融合，每个 $3 \times 3 \times 3$ 的卷积块之后均进行池化操作，池化卷积核大小为 1；第 4 个分支进行 3D 的平均池化，池化卷积核为 3，之后连接一个 $1 \times 1 \times 1$ 的卷积块，所有分支均在通道的方向进行拼接后输入到下一层。

2.3 空间注意力机制

本文引入空间注意力模块^[21]作为解码模块的一部分，空间注意力认为图像中所有区域并不是同等重要的，只有对当前任务有用的区域才是需要关心的，如分割任务中的肿瘤区域。空间注意力模块的任务就是加强网络对关心区域的寻找能力，从整个图像出发，关注面上每个局部的权重。

本工作采用的空间注意力模块结构如图 2 所示，首先，对输入特征 $\mathbf{F} \in R^{H \times W \times C}$ 进行平均池化和最大池化操作，平均池化关注特征图中整体特征的提取，最大池化关注特征图中细节特征的提取。其中，对特征图按照通道进行压缩，将每个通道对应位置的值相加后取平均，得到一个 $\mathbf{F}_{ap}^s \in R^{H \times W \times 1}$ 的特征图，即平均池化的结果；最大池化是将平均池化中的平均操作变成取最大值操作，同样可得到一个 $\mathbf{F}_{mp}^s \in R^{H \times W \times 1}$ 的特征图。池化后的两张特征图在通道维度进行堆叠，经过一个 7×7 的卷积层融合通道信息，并通过 Sigmoid 函数将特征图的空间权重缩放至 $[0,1]$ ，得到权重系数 $M^s(\mathbf{F}) \in R^{H \times W \times 1}$ 。最后，将权重系数与原始的输入特征相乘得出输出特征

$\mathbf{F}^s \in R^{H \times W \times C}$ ，公式描述如公式(1)所示。

$$\mathbf{F}^s = \mathbf{F} \cdot M^s(\mathbf{F})$$

$$= \mathbf{F} \cdot \sigma(f^{7 \times 7}([\text{MaxPool}(\mathbf{F}); \text{AvgPool}(\mathbf{F})])) \quad (1)$$

$$= \mathbf{F} \cdot \sigma(f^{7 \times 7}[\mathbf{F}_{mp}^s; \mathbf{F}_{ap}^s])$$

其中， $f^{7 \times 7}(\cdot)$ 为 kernel 是 7 的卷积运算； $\sigma(\cdot)$ 为 Sigmoid 函数。

2.4 网络结构

本文模型以 U-Net 网络^[11]作为基本框架，U-Net 网络包括一个编码器和一个解码器。编码器中使用 Inception 模块，在解码器中增加空间注意力机制和 Dense 模块^[22]。Inception 模块用于增加网络的宽度与深度，注意力机制可使网络聚焦于肿瘤等局部区域，Dense 模块通过密集连接获取每层的输出特征，使得较早层提取的特征仍可被较深层使用，增强了特征传递的同时，减轻了梯度消失现象。

本文模型中的归一化方式统一采用 InstanceNorm3d，激活函数为 LeakyReLU，Dense 模块使用 $3 \times 3 \times 3$ 的卷积块，并引入深监督模块^[23]，有助于注意力关注到更大范围的肿瘤区域。该网络输入一个 $2 \times 144 \times 144 \times 144$ 的张量，由 PET、CT 图像拼接而成；解码器的输出通道是一个 $1 \times 144 \times 44 \times 144$ 的张量，用于生成分割的结果。卷积操作每一层的输出特征图数量分别为 32、64、128、256、320、256、128、64、32、1，整体网络结构如图 3 所示。

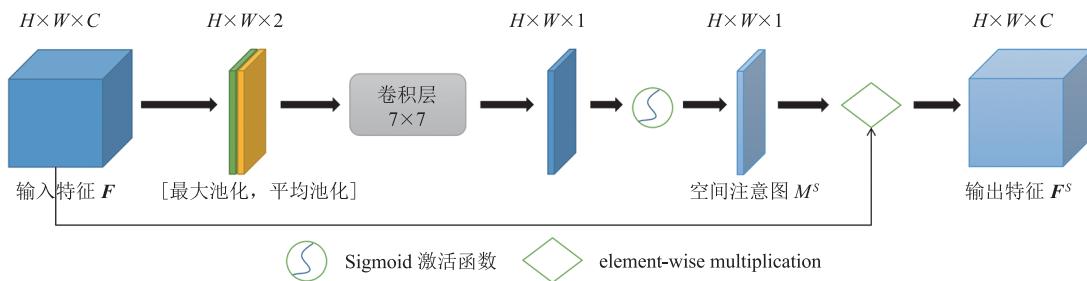


图 2 空间注意力机制结构图

Fig. 2 Structure of spatial attention mechanism

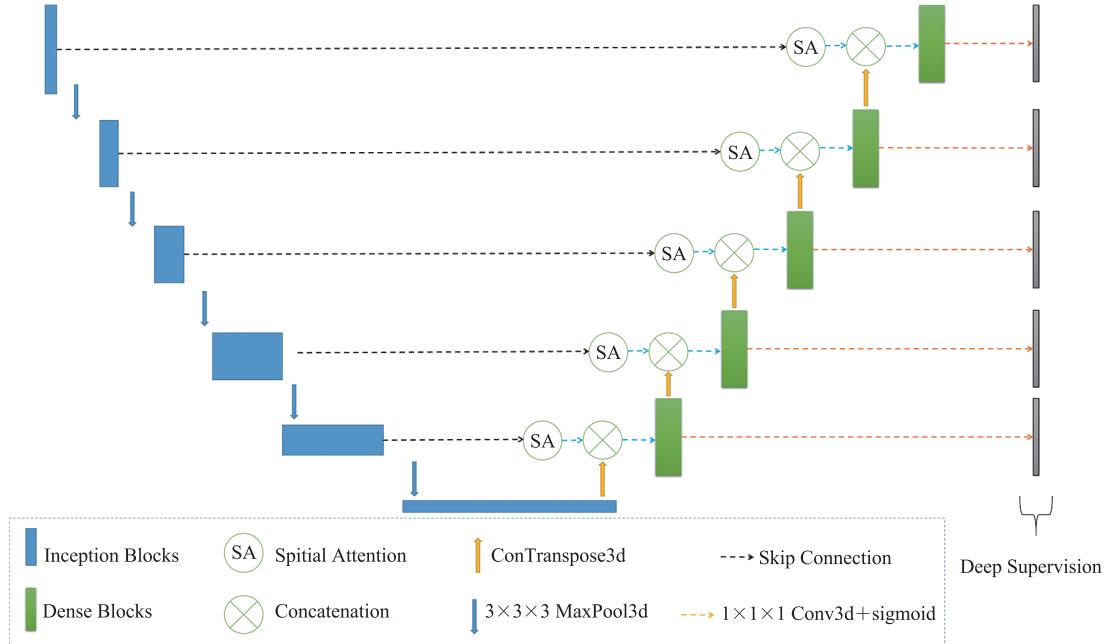


图 3 整体网络结构

Fig. 3 Overall network structure

2.5 损失函数

在医学图像分割任务中, Dice 损失函数是常用的损失函数之一, 在模型训练过程中, 通常被用于指导模型下一步训练方向, 其计算公式如公式(2)所示。

$$Loss_{Dice}(y, \hat{y}) = 1 - \frac{2 \sum_i^N y_i \hat{y}_i}{\sum_i^N y_i + \sum_i^N \hat{y}_i + \epsilon} \quad (2)$$

其中, y_i 是第 i 个样本的标签; \hat{y}_i 是第 i 个样本的预测结果; N 是样本的总数; ϵ 是一个很小的常数, 以防止分母为 0。

此外, 本实验还增加了 Focal 损失功能^[24], 用以最大化背景和分割目标之间的差异, 计算公式如公式(3)所示。

$$\begin{aligned} Loss_{Focal}(y, \hat{y}) = & \\ & -(1-\alpha)(1-y)\hat{y}^\gamma \log(1-\hat{y}) - \alpha y(1-\hat{y})^\gamma \log \hat{y} \end{aligned} \quad (3)$$

其中, α 为平衡权重因子, 其值设置为 0.5; γ 为重量减少的速率, 其值设置为 2。

本文将 Dice 损失与 Focal 损失组合, 以对模型进行训练, 训练公式如公式(4)所示。

$$Loss(y, \hat{y}) = Loss_{Dice}(y, \hat{y}) + Loss_{Focal}(y, \hat{y}) \quad (4)$$

3 实验结果与讨论

3.1 评价指标

本文尝试使用 DSC、HD95、召回率(Recall)、精度(Precision)以及 Jaccard 系数(Jaccard similarity coefficient)5 个评价指标验证网络的分割性能。

DSC: 范围为 0~1, 用以度量两个集合的相似度, 值越大, 代表两个集合越相似, 主要关注图像目标区域内部的分割效果, 其定义如公式(5)所示。

$$DSC(P, G) = \frac{2 \times |P \cap G|}{|P| + |G|} \quad (5)$$

其中, P 为模型预测的结果; G 为真实的标签。

HD95 是一种形状相似度的度量指标, 用以

量化轮廓精度, 表示两组集合之间最短距离的最大值。95% 指对两个集合中的边界点距离排序后取 95% 位数值, 以消除异常值子集的影响。其定义如公式(6)所示。

$$d_H(P, G) = \max \{d_{PG}, d_{GP}\} = \max \left\{ \max_{p \in P} \min_{g \in G} (p, g), \max_{g \in G} \min_{p \in P} (p, g) \right\} \quad (6)$$

召回率(Recall)和精度(Precision)的定义如公式(7)和公式(8)所示。

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

其中, TP (True Positive)为正类样本正确分类的样本数(真阳性); FP (False Positive)为正类样本错误分类的样本数(假阳性); FN (False Negative)为负类样本错误分类的样本数(假阴性)。

Jaccard 系数(Jaccard similarity coefficient)用于比较有限样本集之间的相似性与差异性。给定两个集合 A 和 B , Jaccard 系数定义为 A 与 B 交集的大小和 A 与 B 并集的大小的比值, 定义如公式(9)所示。

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (9)$$

其中, $J(A, B) \in [0, 1]$, Jaccard 系数值越大, 样本相似度越高。

3.2 训练细节

本文使用一个 NVIDIA RTX 3090 GPU 训

练网络, 网络中使用的超参数均保持一致。使用 PyTorch 深度学习框架进行模型训练, 其中 batch_size 设置为 1, 使用 Adam 作为优化器^[25]。网络训练 epoch 设置为 100, 初始学习率为 2×10^{-4} , 为避免陷入局部最优, 本实验使用余弦退火算法^[26]将学习率从 2×10^{-4} 降至 1×10^{-6} 。数据增强的方法是以 0.5 的概率按[0.7,1.2]的比例进行随机扩大和缩小, 以 $[-30^\circ, 30^\circ]$ 之间的随机角度旋转来增加数据, 提高模型的泛化能力。

3.3 实验结果

本文提出的模型与一系列经典模型在 HECKTOR 数据集上的实验结果如表 1 所示。由表 1 可知, 本文方法的 DSC 为 0.782, 召回率为 0.846, Jaccard 系数为 0.675, 与原始的 U-Net 相比, 分别提升 6.8%、13.4% 和 9.8%; HD95 为 5.661, 较原始的 U-Net 下降了 1.616。此外, 与一些改进的 U-Net 结构相比, 本文网络的性能也有一定提升。就 DSC 而言, 本文网络较 3D-UNet 提升了 3.4%, 较 Residual-UNet 提升了 4.0%, 较 Dense-UNet 提升了 4.3%。但本文网络的网络准确率仅为 0.783, 较原始的 U-Net 降低了 2.6%, 本文模型在分割精确度方面有所欠缺, 需进一步改进。

表 2 为本文模型与其他头颈部分割相关模型的对比结果, Iantsen 等^[27]在传统 U-Net 的基础上加入了残差层以及 SE 模块, 解决了梯度消失问题的同时, DSC 达到了 0.772; Xie 等^[28]使用一个微调的 3D nnU-Net 结合 SE 模块对肿瘤进行分

表 1 基于 HECKTOR 数据集的不同方法的比较结果

Table 1 Comparison results of different methods based on HECKTOR dataset

方法	DSC	HD95	Recall	Precision	Jaccard
2D-UNet	0.732	7.277	0.746	0.804	0.615
3D-UNet	0.756	5.927	0.816	0.763	0.647
Residual-UNet	0.752	6.213	0.825	0.755	0.642
Dense-UNet	0.750	9.745	0.843	0.738	0.638
Spatial-Attention Dense U-Net	0.760	7.551	0.834	0.772	0.649
ISA-DUNet(本工作)	0.782	5.661	0.846	0.783	0.675

割, 通过空间和通道的双重注意力机制, DSC 达到了 0.766。实验表明, 与 Xie^[28]和 Iantsen^[27]等的方法相比, 本文方法的 DSC 和 HD95 均有较大提升, 此外, HD95 的结果也证明了本文方法在肿瘤边缘处理方面的优越性。因此, 本文提出的 ISA-DUNet 模型不仅可以较为准确地找到肿瘤位置, 还可以提取肿瘤的边缘信息, 以取得较好的分割准确性。

表 2 与其他头颈部分割方法的比较结果

Table 2 Comparison results with other head and neck tumor segmentation methods

数据集	方法	DSC	HD95
HECKTOR	Iantsen 等 ^[27]	0.772	6.428
	Xie 等 ^[28]	0.766	7.218
	ISA-DUNet (本工作)	0.782	5.661

为探究多个模态是否对分割性能有所提升, 本文还进行了单模态实验, 使用本文提出的模型, 将 CT 和 PET 单个模态的图像作为输入, 与多个模态的结果进行对比, 结果表明, 多模态的图像分割有助于提高分割性能。

不同模态对比实验的结果如表 3 所示, 选用 DSC 和 HD95 作为评价指标, 从肿瘤内部分割效果和边缘分割效果两方面进行评估。由表 3 可知, 单独使用 CT 图像的分割效果最差, PET 图像的分割效果略优, PET-CT 双模态图像的 DSC 得分最高。PET 作为一种分子成像技术, 通过放射性示踪剂进行成像, 与 CT 相比, PET 更容易定位肿瘤位置, 弥补了 CT 软组织成像方面的不足。不同模态之间图像信息的互补则更有助于肿

表 3 基于本文方法不同输入模态的结果

Table 3 Results of different input modes based on the method presented in this paper

数据集	模态	DSC	HD95
HECKTOR	CT	0.628	6.752
	PET	0.745	5.456
	PET-CT	0.782	5.661

瘤的分割。

在本文模型中, 采用 PET-CT 双模态作为输入信息, 其中, CT 图像可以提供准确的解剖和肿瘤轮廓信息, PET 图像可以精确定位肿瘤的位置, 两种成像技术优势互补, 可实现更加精确的分割效果。因此, 本文采取 PET-CT 双模态输入, 将来自两种不同模式的数据进行拼接作为输入, 以提升分割性能。

为了更直观地反映本文模型的分割性能, 本实验选取部分病人的肿瘤切片进行可视化, 结果如图 4 所示。其中, 红色区域为肿瘤区域, 白色区域为不同模型预测的肿瘤区域, 蓝色箭头标注部分为本文模型预测的肿瘤区域与其他模型相比明显不同之处, 所有图像均为不同病人矢状面的切片。观察图 4 中病人 1 的图像可知, 对于左上角的无肿瘤区域, 本文模型对其进行了精确的预测, 而其他模型均有错误预测。观察图 4 中病人 3 的图像可知, 就边缘细微区域的分割效果而言, 本文模型明显优于对比实验中的其他模型, 如蓝色箭头所指的地方, 对于上端的区域, 2D-UNet 预测过多, 3D-UNet 和其他模型的预测又过少, 本文提出的模型是最贴近标记肿瘤区域的模型。图 4 进一步反映了本文模型对于肿瘤区域的识别明显优于其他模型, 特别是对边缘一些细节部位的识别, 本文模型均能更好地捕获。

4 结 论

本文提出一种基于 ISA-DUNet 的 PET-CT 双模态头颈部肿瘤分割方法, 测试结果表明, ISADUNet 提高了肿瘤分割的准确度, 通过同时采用双模态的图像输入充分利用了 PET 对肿瘤的高灵敏度特征和 CT 在肿瘤边界结构中的高精度来提取图像特征信息。该方法不仅可以精确地对肿瘤内部区域进行分割, 而且可以准确地提取肿瘤的边缘信息, 从而获得更高的分割精度。

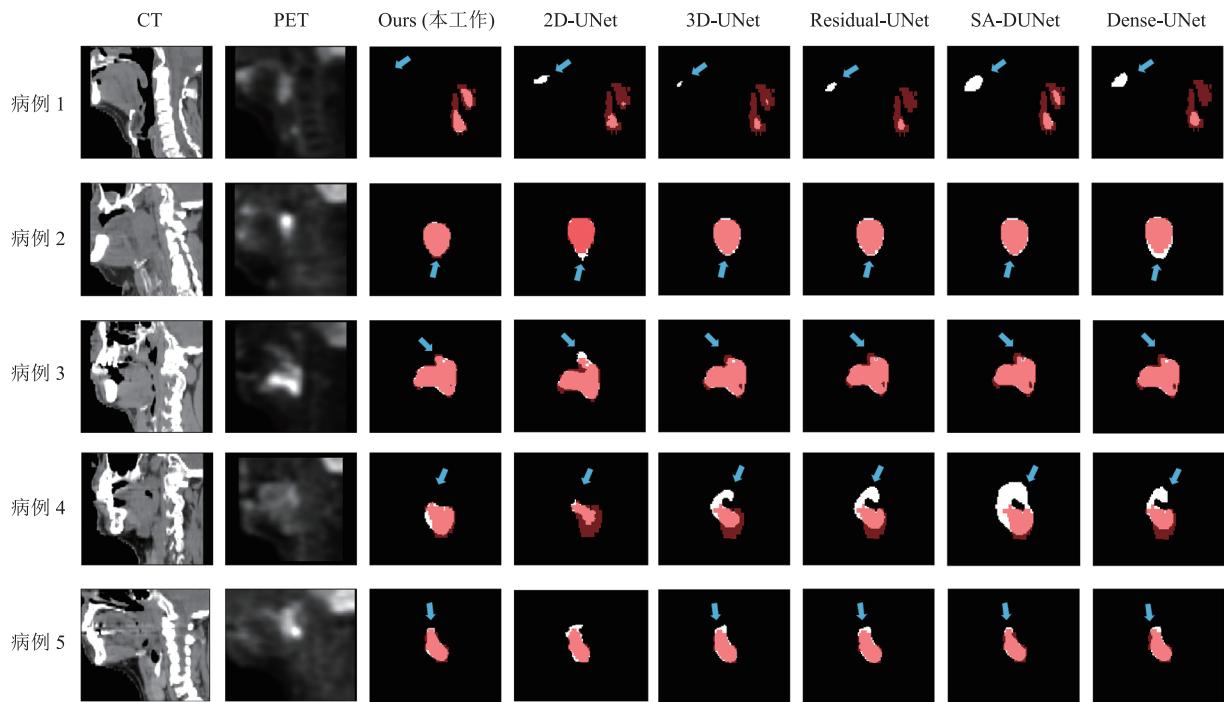


图 4 5 个不同病人肿瘤切片的分割结果

Fig. 4 Segmentation of tumor sections from five different patients

本团队在传统 U-Net 结构的基础上，将编码器端的卷积神经网络替换为 Inception 模块，实现了对特征更深层的提取，在解码器端引入空间注意力机制，提高肿瘤区域在学习过程中的权重，实现对肿瘤区域的精确分割，同时加入 Dense 模块以及深监督模块，减缓学习过程中的梯度消失效应。此外，利用 ISA-DUNet 对头颈部肿瘤数据集进行处理，并使用三折交叉验证评估其在训练集上的性能。与多个经典网络的对比实验结果表明，本文方法优于目前经典的头颈部肿瘤分割方法。

但是本文网络的分割精确度仍有待提升，原始 U-Net 的分割精确度指标高于本文其他测试模型。此外，本文缺少对于其他不同部位肿瘤分割效果的讨论，对于该网络的泛化能力讨论不足，后续研究可通过增加其他不同的数据集进行对比。对于不同网络结构的贡献，本文也未进行消融实验分析，后续研究可增加消融实验，以进

步分析各个模块的贡献度。在今后的工作中将尝试优化上述问题，进一步提高模型的分割性能，并尝试在更多的数据集上推广应用。

参 考 文 献

- [1] 刘云鹏, 刘光品, 王仁芳, 等. 深度学习结合影像组学的肝脏肿瘤 CT 分割 [J]. 中国图象图形学报, 2020, 25(10): 2128-2141.
Liu YP, Liu GP, Wang RF, et al. Accurate segmentation method of liver tumor CT based on the combination of deep learning and radiomics [J]. Journal of Image and Graphics, 2020, 25(10): 2128-2141.
- [2] Wei L, El Naqa I. Artificial intelligence for response evaluation with PET/CT [J]. Seminars in Nuclear Medicine, 2021, 51(2): 157-169.
- [3] Kao YS, Yang J. Deep learning-based auto-segmentation of lung tumor PET/CT scans: a systematic review [J]. Clinical and Translational

- Imaging, 2022, 10(2): 217-223.
- [4] Ren JT, 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 Oncologica, 2021, 60(11): 1399-1406.
- [5] Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy [J]. Medical Physics, 2014, 41(5): 050902.
- [6] Savjani RR, Lauria M, Bose S, et al. Automated tumor segmentation in radiotherapy [J]. Seminars in Radiation Oncology, 2022, 32(4): 319-329.
- [7] Li LQ, Zhao XM, Lu W, et al. Deep learning for variational multimodality tumor segmentation in PET/CT [J]. Neurocomputing, 2020, 392: 277-295.
- [8] Lian CF, Ruan S, Denoeux T, et al. Joint tumor segmentation in PET-CT images using co-clustering and fusion based on belief functions [J]. IEEE Transactions on Image Processing, 2018, 28(2): 755-766.
- [9] Cardenas CE, Yang JZ, Anderson BM, et al. Advances in auto-segmentation [J]. Seminars in Radiation Oncology, 2019, 29(3): 185-197.
- [10] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C] // Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [11] Zhang ZX, Liu QJ, Wang YH. Road extraction by deep residual U-Net [J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753.
- [12] Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation [C] // Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016: 424-432.
- [13] Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation [C] // Proceedings of the 2016 Fourth International Conference on 3D Vision, 2016: 565-571.
- [14] Economopoulou P, Psyrris A. Head and neck cancers: essentials for clinicians, chap. 1 [M]. <https://oncologypro.esmo.org/content/download/113133/1971849/file/2017-ESMO-Essentials-for-Clinicians-Head-Neck-Cancers-Chapter-1.pdf>.
- [15] Xie JY, Peng Y. The head and neck tumor segmentation based on 3D U-Net [C] // Proceedings of the 3D Head and Neck Tumor Segmentation in PET/CT Challenge, 2022: 92-98.
- [16] An CY, Chen H, Wang LS. A coarse-to-fine framework for head and neck tumor segmentation in CT and PET images [C] // Proceedings of the 3D Head and Neck Tumor Segmentation in PET/CT Challenge, 2022: 50-57.
- [17] Qayyum A, Benzinou A, Mazher M, et al. Automatic segmentation of head and neck (H&N) primary tumors in PET and CT images using 3D-Inception-ResNet model [C] // Proceedings of the 3D Head and Neck Tumor Segmentation in PET/CT Challenge, 2022: 58-67.
- [18] 殷晓航, 王永才, 李德英. 基于 U-Net 结构改进的医学影像分割技术综述 [J]. 软件学报, 2021, 32(2): 519-550.
Yin XH, Wang YC, Li DY. Survey of medical image segmentation technology based on U-Net structure improvement [J]. Journal of Software, 2021, 32(2): 519-550.
- [19] Andrarczyk V, Oreiller V, Bougħadid S, et al. Overview of the HECKTOR challenge at MICCAI 2021: automatic head and neck tumor segmentation and outcome prediction in PET/CT images [C] // Proceedings of the 3D Head and Neck Tumor Segmentation in PET/CT Challenge, 2022: 1-37.
- [20] Szegedy C, Liu W, Jia YQ, et al. Going deeper with convolutions [C] // Proceedings of the IEEE

- Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [21] Zhu XZ, Cheng DZ, Zhang Z, et al. An empirical study of spatial attention mechanisms in deep networks [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6688-6697.
- [22] Guan S, Khan AA, Sikdar S, et al. Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal [J]. IEEE Journal of Biomedical and Health Informatics, 2019, 24(2): 568-576.
- [23] Lee CY, Xie SN, Gallagher P, et al. Deeply-supervised nets [C] // Proceedings of the 18th International Conference on Artificial Intelligence and Statistics, 2015: 562-570.
- [24] Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [25] Kingma DP, Ba J. Adam: a method for stochastic optimization [Z/OL]. arXiv Preprint, arXiv: 1412.6980, 2014.
- [26] Loshchilov I, Hutter F. SGDR: stochastic gradient descent with warm restarts [Z/OL]. arXiv Preprint, arXiv: 1608.03983, 2016.
- [27] Iantsen A, Visvikis D, Hatt M. Squeeze-and-excitation normalization for automated delineation of head and neck primary tumors in combined PET and CT images [C] // Proceedings of the 3D Head and Neck Tumor Segmentation in PET/CT Challenge, 2021: 37-43.
- [28] Xie JY, Peng Y. The head and neck tumor segmentation using nnU-Net with spatial and channel ‘squeeze & excitation’ blocks [C] // Proceedings of the 3D Head and Neck Tumor Segmentationin PET/CT Challenge, 2021: 28-36.