

# U-Net 变体在鼻咽癌肿瘤分割中的应用回顾

丁字欣

浙江中医药大学 浙江省杭州 310053

**摘要:** 鼻咽癌 (NPC) 是头颈部最常见的恶性肿瘤之一。U-Net 因其灵活性、优化的模块化设计以及在所有医学图像模式中的成功而成为最广泛使用的图像分割架构。本文首先介绍了当前鼻咽癌现状和 U-Net 的工作原理, 然后, 根据时间线回顾了 U-Net 及其变体在鼻咽癌肿瘤分割中的应用进展。最后, 本文讨论了当前研究的不足以及 U-net 应用于鼻咽癌的未来发展方向, 并提出了一些建议。

**关键词:** U-Net; 鼻咽癌; 图像分割; 深度学习

## 引言

鼻咽癌是一种起源与鼻咽部上皮细胞的恶性肿瘤, 在东南亚地区发病率高<sup>[1]</sup>。鼻咽癌与种族、遗传、EB 病毒感染及环境因素有关, 鼻咽癌好发于中年人, 男性多见。鼻咽癌患者有鼻塞、涕血或鼻出血、耳鸣、听力减退、头痛、麻木、复视七大主诉症状。鼻咽癌对放疗和化疗敏感性高, 因此放射治疗是鼻咽癌的首选治疗方法, 鼻咽癌患者在及时的治疗下, 5 年生存率可达 80%<sup>[2]</sup>。此外, 不同成像方法对鼻咽癌的分期和放射治疗计划设计有着较大影响, 有 X 射线、计算机断层成像 (CT)、磁共振成像 (MRI) 和鼻咽镜等检查方法。其中, MRI 因其具有软组织分辨率高、对人体完全无害、实用信息等优点, 已被临床广泛用于评估鼻咽癌局部、区域和颅内浸润的最佳成像方式<sup>[3]</sup>。最近的研究表明, 在放疗中使用 MRI 可以降低对危险器官的毒性, 增加肿瘤部位的辐射剂量, 并获得更高的生存获益<sup>[4]</sup>。

临床放射治疗流程包括体位固定、计划设计、计划确认和计划执行阶段, 其中影响放疗效果的主要因素是靶区勾画的准确性<sup>[5]</sup>。大多数临床治疗通常采用手工勾画, 但是由于鼻咽癌可以浸润邻近的颅底和神经结构, 与周围组织之间的界限不清且模糊, 鼻咽癌的靶区勾画比其他类型的肿瘤更复杂。而深度学习方法的一个主要优势是可以直接从原始数据中学习复杂的特征<sup>[5]</sup>。此外, 精确头颈放射治疗的肿瘤靶区轮廓是一项劳动密集型工作, 医生勾画靶区的经验不一会导致结果的准确性也会有所不同<sup>[6]</sup>。因此, 借助人工智能来完成并提高靶区勾画的准确性和减少靶区勾画时间是有必要的, 例如机器学习 (ML)、支持向量机 (SVM)、

深度学习 (DL) 等。其中, U-Net 因其灵活性、优化的模块化设计以及在所有医学图像模式中的成功而成为最广泛使用的图像分割架构<sup>[7]</sup>。

随着人工智能和医学技术的不断发展, 医学成像速度加快的同时图像质量也得到飞升, 医学图像成本的下降和成像数量、信息量的增加导致医生的平均工作量翻倍增加。凭借医生的肉眼和经验, 可能会对一些信息出现遗漏<sup>[8]</sup>。但是, 深度学习不仅可以通过计算密集型学习过程实现卓越的性能, 还可以使用图形处理单元 (GPGPU) 上的通用计算从而高度加速深度学习过程<sup>[9]</sup>。因此, 应用深度学习于图像分割能够更好地利用图像信息并发挥其价值, 减少医生工作量, 实现“智能化”诊疗。

在这项工作中, 我们通过回顾相关文献, 梳理了先前研究中 U-net 应用于鼻咽癌的框架和模型, 并对多种算法进行比较, 从中发现新的关联、趋势和理论观点, 为后来者提供一个清晰的概览。本文不仅分析了现阶段分割方法的不足, 还提出了领域内的研究空白和未解决的问题。

## 1. U-net 的工作原理

医学图像分割方法从早期的边缘检测、模板匹配技术、区域生长、机器学习等其他数学方法发展到近年来基于 CNN 的方法在医学图像分析领域得到了普及。有文献证实, 人工智能轮廓工具提高了鼻咽癌肿瘤靶标轮廓的准确性<sup>[10]</sup>。U-Net 的架构呈 U 形, 由一个包含卷积层和池化层的编码器 (收缩路径) 和一个解码器 (扩展路径) 组成。输入的灰度图像根据每个像素的灰度值转换为单通道矩阵, 而彩色图像则转换为三通道矩阵。编码器的目的是捕获上下文信息<sup>[7]</sup>。其

中,卷积层中卷积核的数目决定着输出特征图的通道数,每个卷积核后有一个修正线性单元(ReLU)激活来增加神经网络不同层之间的非线性关系。此外,每通过一个卷积核输出特征图的尺寸减少2,若对输入图像边缘进行填充即可保持卷积后的特征图像与输入图像的尺寸大小相同。池化层又称下采样,每下采样一次,输出图像的尺寸减少为输入图像的1/2,可以增加网络的感受野。

解码器(扩展路径)包括卷积、上采样、跳级结构和由softmax函数激活的全连接层。输入解码器的特征图像经过卷积、上采样后与收缩路径中相对应裁剪后的特征图像进行融合,再经过多次重复后,经过softmax函数和全连接层得到输出图像。编码器和解码器路径之间的连接路径(也称为瓶颈)包括两个连续的 $3 \times 3$ 卷积,后跟ReLU激活。在这个过程中,网络逐渐学习高级语义特征而失去提取特征的定位。其中,跳级结构沿着网络传播重要的高分辨率上下文信息,有助于在分割任务中保留空间信息<sup>[7]</sup>。此外,全连接层将学习到的分布式特征表示映射到样本标签空间,并通过激活函数将定量值转换为非线性值,在模型中起到分类器的作用即产生所需的类预测(对象与背景)<sup>[11]</sup>。

从技术的角度来看,医学图像中的深度学习分割任务对于大多数不了解U-net概念的临床医生来说,这些概念是抽象的。因此,我们从需要解决的具体临床问题的角度总结了该领域已发表的研究,帮助对人工智能概念不了解的医生更好地应用深度学习于临床,使人工智能更好地为医生和病人服务<sup>[1]</sup>。

## 2. 临床应用及最新进展

计算机辅助诊断系统(CADx)能够通过算法和模型来检测图像中的异常或病变,并辅助医生进行初步诊断。这种技术有助于提高诊断的准确性和速度,同时也可以医学影像学研究和临床实践中发挥重要作用<sup>[12]</sup>。U-Net应用于鼻咽癌肿瘤分割旨在帮助临床医生和放射科医生提高诊断和治疗过程的效率。多年来,已经提出了该网络的几种扩展,以解决医疗任务造成的规模和复杂性并提高了肿瘤分割的准确性和精度。

Xue等人在U-net基础上提出了一种变体SI-Net,通过输入一张CT图像、该图像上的CTVp1轮廓以及下一张CT图像,得到下一张CT图像的预测CTVp1轮廓。在编码器中使用3D卷积核,并且在编码器和解码器的连接中增加了重

塑层,前向SI-Net模型的DSC和JI值分别比U-Net模型高5%和6%。SI-Net和U-Net训练200个epoch所需的时间分别为12小时和8小时,大大减少了模型训练时间。与传统的U-Net模型相比,SI-Net模型保留了相邻图像之间的连续性,从而提高了分割精度。然而,缺点是CT图像缺乏软组织对比度,数据集不够多,增加数据量可以提高准确性和鲁棒性<sup>[13]</sup>。

为了解决上述问题,Li等人提出了U-net的修改版模型来实现高精度勾画CT鼻咽部肿瘤靶区,对数据进行裁剪和归一化后在每个卷积层设置了Dropout以防过拟合并提高了模型的鲁棒性。然后使用测试集中不同T、N分期肿瘤的GTV<sub>nx</sub>、GTV<sub>nd</sub>的DSC值和HD值进行评估,证明归一化提高了U-net的性能,改进的U-Net模型可以比手动描绘具有更高的一致性和效率来描绘鼻咽癌肿瘤靶区,每位鼻咽癌患者的图像自动分割仅需18s而医生手动勾画则需要3小时。通过实验证实,早期勾画精度优于晚期勾画精度,原发肿瘤勾画精度优于淋巴结<sup>[4]</sup>。

Bai等人使用位置分割框架,分两步实现鼻咽癌GTV的精确分割。在定位步骤中,通过召回保留损失对训练的ResNeXt-50 U-Net进行微调从而对鼻咽癌的GTV(大体肿瘤体积)进行粗分割,在分割步骤中,通过骰子(Dice)损失对该结果进行微调,实现逐个分割定位步骤中检测到的边界框区域。该算法在StructSeg-NPC 2019的50个训练数据上实现了 $62.88 \pm 8.12\%$ 的Dice相似系数(DSC),优于广泛使用的2D和3D U-Net在StructSeg 2019 Challenge排行榜上排名第九<sup>[14]</sup>。

为了解决GTV和小GTV在3D图像空间中的分布不平衡很容易导致类别选择偏差的问题,Liu等人将对抗分类学习嵌入到U-Net中,创建了一个基于焦点损失的U-Net分割模型,从而平衡样本中小目标和大目标类别之间的分布差异<sup>[15]</sup>。此外,该文献对鼻咽癌图像进行直方图均衡化的预处理,以确保相同或近似CT值代表相同的组织。与Focal loss-based U-Net、VoxResNet、U-Net、U-Net++模型比较,该模型可以显著提高NPC分割的检测精度和灵敏度。

他们还提出如何利用迁移学习来提取鼻咽癌的抽象特征是非常有必要的。

面对靶区较小、临床CT图像分辨率各向异性、靶区与周围软组织对比度低等挑战时,Mei等人提出了2.5D卷积

神经网络来处理不同面分辨率,并在其中插入一个空间注意力模块,使用通道注意力提高分割性能。其次,采用多尺度采样方法对模型进行训练学习不同尺度的特征,并与多模型集成方法相结合,提高分割结果的鲁棒性。此外,通过模型集合估计了分割结果的不确定性。最终,实验结果表明:①对比发现我们提出的 2.5D 网络在各向异性分辨率的图像上比常用 3D 网络具有更好的性能。②注意力机制的结合可以更好地关注小的 GTV 区域并提高分割精度。③多尺度模型集成获得了更稳健的结果,并且发现 GTV 边界或难以分割的低对比度区域的不确定性通常较高,错误率越高,可以在提高临床决策的准确性和效率方面发挥重要作用<sup>[16]</sup>。

Zhang 等人基于经典的 U-Net,融合了空间注意力、残差连接、循环卷积、归一化等多种先进的计算机视觉方法,提出了一种全自动端到端分割 AttR2U-Net 模型,具有极高的 MRI 图像鼻咽癌分割性能。其中,循环卷积和残差连接结构使该模型提取细节的能力显著提高。此外,跳跃连接将空间注意力融合到网络中,可以更准确地查明癌症区域。计算结果表明,该模型的 DSC 值为 81.6%,并且在七种模型的比较中获得最佳性能,对不规则形状鼻咽癌的分割效果更稳健<sup>[17]</sup>。

随着医学成像技术的发展,大量医学图像数据具有 3 维信息,包括 CT、MRI 等成像模式。为了更好地利用这些图像,我们可以使用多模态或多尺度(参数)模型对肿瘤进行分割。由于鼻咽癌的特殊性,如何在特征提取和选择过程中避免丢失小病灶的特征,或者如何扩展和增强特征,具有重要的研究价值。Hao 等人使用鼻咽癌 MRI 图像的三种模式作为网络输入,即 T1 加权(T1)成像、T2 加权(T2)成像和动态成像,提出了空间连续性信息提取没款(SCIEM)和多尺度语义特征提取模块(MSFFM),并使用 LCo-P ER 进行网络训练从而增强网络模型对不同大小病灶的识别和特征提取,提出了一种基于像素边缘区域级协作损失(LCo-P ER)的多尺度敏感 U-Net(MSU-Net)用于 MRI 鼻咽癌肿瘤分割。测试集的全局 Dice、Precision、Recall 和 IOU 分别为 84.50%、97.48%、84.33% 和 82.41%。该文献提出的特征融合模块对于具有大前部背景间隙或大吃寻间隙的医学图像网络分割具有启发意义<sup>[18]</sup>。

对于上述启发,Zeng 等人基于 U-Net 网络,结合了扩展卷积模块、迁移学习模块和残差模块,创建的新的深度学

习网络模型可以有效解决卷积感知领域受限的问题,实现全局和局部多尺度特征融合,显著提高鼻咽癌目标区域的分割精度。添加的三个模块可以进行多尺度提取,补充特征信息并增强感知野,进一步提高 NPC 的分割精度<sup>[19]</sup>。

计算机模型的更新日新月异,我们需要不断学习和应用新模型,不仅可以通过交叉结合多种模型来创建新模型,还可以将 U-Net 设计为级联深度学习的一部分,从而能够更加有效地捕捉数据中的特征,并且通过有效调整参数使得分割性能最好。

### 3. 讨论及结论

总体而言,U-Net 模型及其变体广泛应用于鼻咽癌肿瘤分割,不仅提高了肿瘤分割准确性和效率,也减少了医生工作量。但各种模型依旧存在一些不足和未来可研究方向。首先,深度学习架构需要大量的训练数据和计算能力。一般来说,浅层网络用于数据稀疏的情况,深层网络的架构性能更好。但计算能力的缺乏将导致需要更多时间来训练网络,这取决于所使用的训练数据的大小。在之前的研究中,Wang 等人比较了相同数据集训练下影像组学和深度学习的模型性能,缺乏数据是限制深度学习模型效果的直接因素之一<sup>[20]</sup>。因为在医学方面,样本收集比较困难,数据量难以达到那么多。但是在数据可用性有限的情况下,可以利用各种增强技术,包括随机裁剪、颜色抖动、图像翻转和随机旋转等。训练大型数据是深度学习无法避免的一个话题,而这个过程会产生的大量碳排放会在一定程度上污染环境。

除此之外,目前的临床数据无统一标准。因为患者检查时不同医生开出的影像检查不一,技师选择的采集参数不一,医学图像还存在固有的噪声和照明,即使经过滤波、强度离散化等预处理,这些信息还是存在一定的差别。为了解决上述问题,有必要统一图像的质量或扩大模型的包容性,以提高在完全标准化的临床环境中的性能。未来研究者需要继续寻找公共 NPC 数据集来评估网络模型的有效性,以评估网络对于不同数据的泛化能力。

其次,我们可以使用多种技术来使模型更好地学习和泛化。这可能包括 L1、L2 正则化器、dropout 和批量归一化等。结果表明,使用批量归一化不仅可以加快训练速度,而且在某些情况下,预形成正则化可以消除使用 dropout 层的需要。在大多数临床应用中,整个数据采集通常发生在一个中心,这限制了数据多样性并导致深度神经网络的通用性较差。相

比之下, 具有授权数据隐私的多中心数据集可以为临床目的提供更真实的数据集。

最重要的是, 深度学习有一个黑箱问题, 很多东西无法解释。然而, 模型解释不仅可以帮助放射科医生理解医学数据中隐藏的标记, 还提供了一种进一步检查网络架构的方法, 以便通过建模临床假设进行可能的改进。在这方面, U-Net 扩展的未来方向可能会考虑可解释特征在其网络设计中的重要性。

#### 参考文献

[1] 吴伯恒, 曹鸿斌, 马永康等. 基于多模态图像深度学习局部晚期鼻咽癌肿瘤靶体积自动勾画的研究 [J]. 肿瘤, 2024, 43:1-10.

[2] Lee A W M, Ng W T, Chan L L K, et al. Evolution of treatment for nasopharyngeal cancer - success and setback in the intensity-modulated radiotherapy era[J]. Radiotherapy and Oncology, 2014, 110(3): 377-384.

[3] 苏晓红, 金观桥. 基于 MRI 的深度学习在鼻咽癌中的研究进展 [J]. 磁共振成像, 2023, 14(03):170-174+188.

[4] Tseng M, Ho F, Leong Y H, et al. Emerging radiotherapy technologies and trends in nasopharyngeal cancer[J]. Cancer Communications, 2020, 40(9): 395-405.

[5] Anwar S M, Majid M, Qayyum A, et al. Medical image analysis using convolutional neural networks: a review[J]. Journal of medical systems, 2018, 42: 1-13.

[6] 郗玉珍, 周敏, 丁忠祥. 影像人工智能在鼻咽癌诊疗中的应用 [J]. 临床放射学杂志, 2022, 41(11):2145-2148.

[7] Azad R, Aghdam E K, Rauland A, et al. Medical image segmentation review: The success of u-net[J]. arXiv preprint, 2022.

[8] Recht M P, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations[J]. European radiology, 2020, 30: 3576-3584.

[9] 李佳熾, 刘红英, 万亮. 基于深度学习的医学图像分析域自适应研究综述 [J/OL]. 计算机应用研究, 2024, 1-12.

[10] Lin L, Dou Q, Jin Y M, et al. Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma[J]. Radiology, 2019, 291(3): 677-686.

[11] Soffer S, Ben-Cohen A, Shimon O, et al. Convolutional neural networks for radiologic images: a radiologist's guide[J]. Radiology, 2019, 290(3): 590-606.

[12] Wang Y, Liu C, Zhang X, et al. Synthetic CT generation based on T2 weighted MRI of nasopharyngeal carcinoma (NPC) using a deep convolutional neural network (DCNN)[J]. Frontiers in oncology, 2019, 9: 1333.

[13] Xue X, Qin N, Hao X, et al. Sequential and iterative auto-segmentation of high-risk clinical target volume for radiotherapy of nasopharyngeal carcinoma in planning CT images[J]. Frontiers in oncology, 2020, 10: 1134.

[14] Bai X, Hu Y, Gong G, et al. A deep learning approach to segmentation of nasopharyngeal carcinoma using computed tomography[J]. Biomedical Signal Processing and Control, 2021, 64: 102246.

[15] Liu Y, Yuan X, Jiang X, et al. Dilated Adversarial U-Net Network for automatic gross tumor volume segmentation of nasopharyngeal carcinoma[J]. Applied Soft Computing, 2021, 111: 107722.

[16] Mei H, Lei W, Gu R, et al. Automatic segmentation of gross target volume of nasopharynx cancer using ensemble of multiscale deep neural networks with spatial attention[J]. Neurocomputing, 2021, 438: 211-222.

[17] Zhang J, Gu L, Han G, et al. AttR2U-Net: A fully automated model for MRI nasopharyngeal carcinoma segmentation based on spatial attention and residual recurrent convolution[J]. Frontiers in Oncology, 2022, 11: 816672.

[18] Hao Y, Jiang H, Diao Z, et al. MSU-Net: Multi-scale Sensitive U-Net based on pixel-edge-region level collaborative loss for nasopharyngeal MRI segmentation[J]. Computers in Biology and Medicine, 2023, 159: 106956.

[19] Zeng Y, Zeng P H, Shen S D, et al. DCTR U-Net: automatic segmentation algorithm for medical images of nasopharyngeal cancer in the context of deep learning[J]. Frontiers in Oncology, 2023, 13.

[20] 崔珂, 田启川, 廉露. 基于 U-Net 变体的医学图像分割算法综述 [J/OL]. 计算机工程与应用, 2024, 1-18.