

基于SAM注意力机制的头颈部危及器官CT图像分割

林小惟, 张福全*

复旦大学附属中山医院放射治疗科, 上海

收稿日期: 2026年2月11日; 录用日期: 2026年3月4日; 发布日期: 2026年3月17日

摘要

目的: 本文提出了一种基于残差和注意力机制U-Net的头颈部医学影像危及器官分割新方法。方法: 在nnU-Net编码阶段, 提出将SAM注意力模块与残差方法相结合的残差注意力模块, 以增强特征表达能力; 在解码阶段引入残差注意力模块, 根据分割任务提高特征加权的相关性。结果: 该方法基于包含22个头颈部危及器官的真实医学影像数据集进行评估, 实验结果表明, 与现有方法相比, 所提方法的平均分割准确率提升了11.4%。对22种头颈部危及器官分割的平均骰子相似系数(Dice Similarity Coefficient, DSC)为87.2%。结论: 基于可分离卷积和注意力机制的U-Net卷积神经网络对鼻咽癌靶区达到了更好的分割精度, 表明该方法有望帮助临床医生提高放射治疗的准确性和效率。

关键词

深度学习, 卷积神经网络, 危及器官分割, 注意力机制

Automatic Segmentation of Organs-at-Risk in Head-and-Neck Carcinoma from Radiation Therapy Using Multi-Scale Fusion and Attention-Based Mechanisms

Xiaowei Lin, Fuquan Zhang*

Radiotherapy Department, Zhongshan Hospital, Fudan University, Shanghai

Received: February 11, 2026; accepted: March 4, 2026; published: March 17, 2026

*通讯作者。

文章引用: 林小惟, 张福全. 基于SAM注意力机制的头颈部危及器官CT图像分割[J]. 临床医学进展, 2026, 16(3): 2644-2650. DOI: 10.12677/acm.2026.1631064

Abstract

We proposed a new organs-at-risk segmentation method for medical images of heads and necks based on the U-Net with residuals and attention mechanism. A SAM block combined with a residual method is proposed as Residual-SAM Block in the nnU-Net encoding stage in order to enhance the feature expression ability. SAM are introduced in the decoding stage to increase the relevance of feature weighting in accord with segmentation tasks. The proposed method has been evaluated through a set of real world medical images including 22 organs at risk in the head and neck. Experimental results showed that our proposed method improved the average segmentation accuracy by 11.4% compared with current existing methods. The average Dice Similarity Coefficient (DSC) score on the segmentation of the 22 types of head-and-neck related organs at risk is 87.2%. It demonstrates that this approach is promising for clinical doctors to improve their accuracy and efficiency in radiotherapy.

Keywords

Deep Learning, Convolutional Neural Networks, Organs-At-Risk Image Segmentation, Attention Mechanisms

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

鼻咽癌(Nasopharyngeal Carcinoma, NPC)是全国常见的恶性肿瘤之一[1]。放射治疗是鼻咽癌最常用的治疗方式[2], 但部分健康的器官被照射野照射。放疗医生需要保证肿瘤能够接受剂量辐射的同时, 避免健康器官受损, 相关的健康器官被称为危及器官(Organs-at-Risk, OARs)。计算机断层扫描(Computed Tomography, CT)图像能提供肿瘤的信息和危及器官边界, 在放疗前的计划制定中起着关键作用。然而, 头颈部危及器官结构复杂, 其分割任务具有挑战性, 危及器官边界差异显著, 这些因素均增加了头颈部危及器官精准分割的难度。

传统上, CT 影像中危及器官边界的勾画主要依赖人工完成, 不仅耗时费力, 还存在操作者间和操作者的差异。目前, 基于图谱的方法[3][4]通过将带标签的影像配准到原始影像, 利用配准得到的形变场将模板标签转换为影像分割结果, 该方法对形状分布一致的分割结果较好。但鼻咽癌发展过程中常侵犯周围组织, 会降低配准精度, 影响分割准确性。统计形状模型[5][6]通过获取头颈部 OARs 的一致性特征, 基于图谱配准得到分割结果, 但该方法的融合模式和初始化原理往往需要人工干预, 且易陷入局部最优。周等[7]提出一种基于支持向量机的方法, 用于从磁共振成像中分割鼻咽癌病灶。

近年来, 卷积神经网络[8][9]已广泛应用于医学影像分割。卷积神经网络通过卷积操作提取影像特征, 能更有效地挖掘医学影像中的信息。Long 等[10]提出全卷积网络, 用卷积层替代所有全连接层, 大幅减少了模型参数并提升了性能; Ibragimov 和 Xing [11]基于全卷积网络提取特征进行分类, 取得了较高的准确率; Ronneberger 等[12]提出 U-Net 用于生物影像分割; 注意力 U-Net (Attention U-Net) [13]在 U-Net 中引入注意力, 无需额外的特征提取网络, 实现更优的医学影像分割效果; Gao 等[14]提出 FocusNet, 添加子网络实现小器官分割。

2. 方法

2.1. SAM 注意力模块

随着模型层数加深,图像分割结果容易因信息丢失产生误差。为解决这一问题,残差网络[15]被提出。残差连接模块构建的残差网络包含两条路径:一条路径将输入图像直接传递至下层的模块,另一条路径通过特征提取形成残差短连接,该模块为残差映射。残差网络中包含两个及以上卷积层,卷积层后均进行归一化(Batch Normalization)。这种设计可以使得模型结构更深,不会出现性能退化问题。

本文采用的 SAM 注意力模块是一种用于分割任务的注意力模块。多项研究[16][17]表明,注意力模型能帮助网络在全局影像上下文中聚焦于目标区域的学习,更好地关注危及器官区域的特征信息, SAM 由图像编码器、提示编码器和掩码解码器三部分组成。图像编码器提取与任务无关的通用图像特征,基于视觉 Transformer (ViT),将输入图像通过卷积划分为 16×16 的图像块(Patch),经嵌入层(Patch Embedding)转换为 768 维特征向量,并添加位置编码(Positional Embedding)以保留空间信息。该编码器对每张图像仅处理一次,输出全局图像嵌入(Image Embedding)。提示编码器将用户意图(提示)编码为任务相关的特征,通过卷积层嵌入,与图像嵌入逐元素相加。掩码解码器核心为双向 Transformer 交叉注意力模块,融合图像嵌入与提示嵌入,掩码解码器通过交叉注意力,让“提示特征”在“通用图像特征”中定位对应的区域,最终生成掩码。

这种设计让 SAM 无需针对特定任务微调,仅通过不同提示即可适配任意分割场景。SAM 注意力的网络能有效抑制影像中与目标任务无关区域的原始特征,同时突出与任务相关的区域特征表达,无需进行感兴趣区域提取操作。在残差模块(Residual Block)的基础上,引入 SAM 模块,结构如图 1 所示,其核心目的是获取通道信息和空间信息。每个残差-SAM 模块包含两条路径:第一条路径采用卷积层(由卷积、批量归一化、ReLU 和 SAM 模块组成的基本单元)对输入特征图进行一系列特征学习;第二条路径为输入特征图本身;两条路径相加后作为输出。该模块能学习不同通道间投影操作的相互依赖关系,结合空间和通道上下文进行重新校准,且在深层网络中仍可实现参数更新。

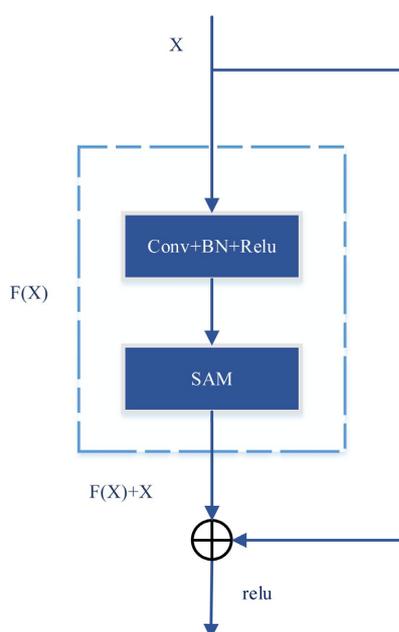


Figure 1. Residual SAM attention block
图 1. 残差 SAM 注意力模块

2.2. nnSAM-UNet 模型

图 2 展示了所提网络的整体架构。nnUNet 不固定网络层数, 而是根据数据特性选择“编码器-解码器”的组合:

编码器(下采样路径): 基础单元以“ $3 \times 3 \times 3(3D)/3 \times 3(2D)$ 卷积 + Batch Normalization + ReLU”为基础卷积块, 若图像维度深(如 3D 体积), 自动集成残差连接(Residual Block), 解决深层网络梯度消失问题。下采样用“最大池化 Max Pooling 实现, 自动调整池化次数, 用来平衡感受野大小与计算量(如小器官分割减少池化次数, 保留细节)。

解码器(上采样路径): 上采样单元用转置卷积(Transposed Convolution)恢复特征图尺寸, 可自动增加“卷积 + 上采样”的迭代次数来提升细节精度。

特征融合: 严格遵循 U-Net 的“拼接(Concat)”策略, 上采样生成的特征图与编码路径对应层级的特征图融合, 融合后的特征图既包含复杂抽象特征, 又提取了各层级低维与高维特征的细节信息, 能获取多尺度特征信息, 提升网络性能, 避免逐点相加导致的细节丢失。

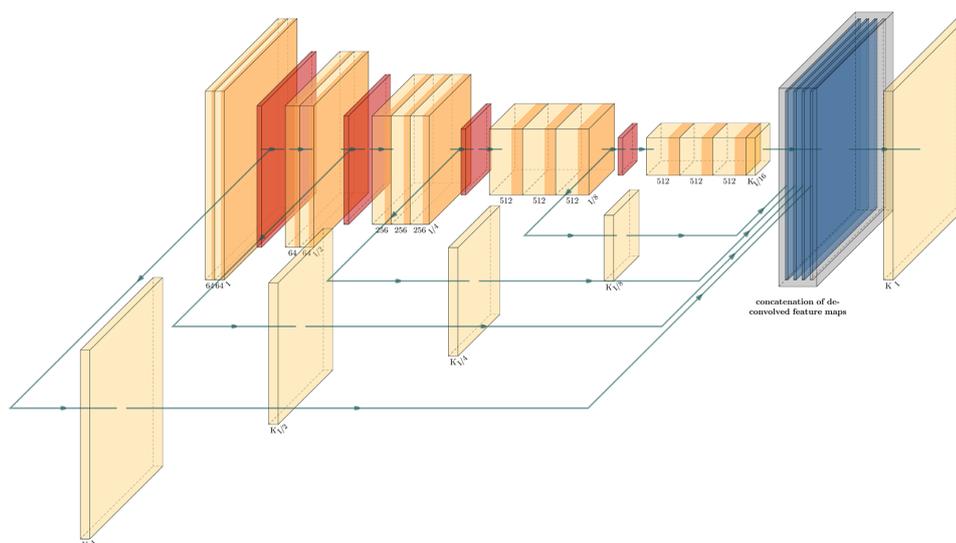


Figure 2. nnSAM-UNet neural network
图 2. nnSAM-UNet 网络模型

3. 数据集

3.1. 实验数据集

本文实验所采用的数据集由 MICCAI StructSeg 2019 比赛提供, 共包含 50 例头部至胸部区域的鼻咽癌患者的 CT 图像。CT 图像平面尺寸为 512×512 像素, 空间分辨率分布在 $0.97 \sim 1.2$ mm 之间, 数量分布在 100~152 层, 平均层厚为 3 mm。数据集中 22 个危及器官标注均由浙江省肿瘤医院的放疗专家进行手工标注, 作为金标准用于本文的模型训练和模型测试过程。

3.2. 单位

不同设备和参数会导致 CT 影像的空间分辨率存在差异, 直接将不同空间分辨率的影像作为网络输入会造成模型训练波动, 难以实现优化。为此, 将影像重采样至统一分辨率 $1 \times 1 \times 3$, 使输入影像包含相似的空间信息, 加快收敛速度。为增强头颈部器官在 CT 影像中的对比度, 对影像灰度值进行标准化处

理。nnUNet 模型对 50 位病人数据集随机划分五折交叉训练和 10 例测试集。所提网络在 PyTorch 框架下实现, 采用 Quadro GV100 32G GPU 进行训练, batchsize 为 13, nnSAM-UNet 采用 SGD 优化器, 初始学习率为 0.01, 指数衰减。医学影像分割的定量评估基于骰子相似系数(DSC)。

3.3. 实验结果

为了验证本文提出的多尺度融合模块和结合 scSE 空间和通道注意力模块的有效性, 本文进行了消融实验, 将 U-Net、Attention U-Net 模型的 SE U-Net 和我们提出的 nnU-Net 模型在 structseg2019 上进行实验测试, 测试结果如表 1 所示, 结果显示, 我们提出的 SAM 注意力模块集成到 nnU-Net 模型中分割效果有显著的提升, DSC 和提高了 11.4 个百分点, 证明了我们提出的深度学习模型对鼻咽危及器官分割的有效性。最后注意力模块集成的 nnSAM-UNet 取得最优的效果, 对补充的特征信息分配权重信息, 对重要的特征信息分配高权重, 不重要的特征信息分配低权重, 从而提高了模型的分割精确性。

Table 1. Segmentation results of head and neck organs at risk by different network models

表 1. 不同网络模型对头颈部危及器官分割结果

危及器官	戴斯相似性系数(%) ± 标准差			
	U-Net	Attention U-Net	SE U-Net	nnSAM-UNet
脑干	76.2 ± 6.1	79.9 ± 4.3	81.7 ± 3.9	88.7 ± 1.5
左眼	87.5 ± 2.4	86.0 ± 2.9	86.9 ± 4.0	90.4 ± 4.9
右眼	87.2 ± 2.9	86.9 ± 4.1	85.2 ± 3.9	89.6 ± 2.3
左晶状体	72.4 ± 11.0	65.3 ± 9.3	61.1 ± 12.8	74.7 ± 10.2
右晶状体	67.5 ± 14.3	61.0 ± 16.4	64.5 ± 15.51	73.4 ± 11.5
左视神经	65.1 ± 11.8	47.8 ± 13.9	58.1 ± 14.1	69.2 ± 6.6
右视神经	65.3 ± 7.7	42.5 ± 23.3	60.8 ± 7.0	74.5 ± 7.7
视神经交叉	46.5 ± 14.1	39.3 ± 16.6	34.8 ± 12.7	62.4 ± 14.0
左侧颞叶	83.6 ± 5.1	80.9 ± 6.5	82.0 ± 5.3	87.2 ± 5.6
右侧颞叶	84.0 ± 5.5	82.5 ± 5.9	82.3 ± 4.7	88.7 ± 4.4
脑垂体	51.5 ± 20.3	43.1 ± 13.7	51.3 ± 18.9	68.3 ± 14.9
左侧腮腺	77.5 ± 6.1	78.0 ± 4.9	75.6 ± 8.9	80.7 ± 5.2
右侧腮腺	81.1 ± 4.9	80.0 ± 5.4	67.3 ± 10.8	81.9 ± 5.5
左侧内耳	80.5 ± 5.9	82.5 ± 4.0	78.8 ± 5.7	76.3 ± 3.0
右侧内耳	83.5 ± 3.6	83.4 ± 3.4	83.7 ± 2.7	82.1 ± 2.4
左侧中耳	73.2 ± 16.4	70.8 ± 16.2	67.1 ± 24.9	83.1 ± 7.7
右侧中耳	79.8 ± 6.8	79.9 ± 4.8	80.6 ± 5.4	84.3 ± 4.4
左颞下颌	73.0 ± 5.1	73.6 ± 7.2	71.6 ± 9.4	75.6 ± 9.1
右颞下颌	73.4 ± 5.1	69.4 ± 5.5	70.0 ± 9.3	74.9 ± 8.0

续表

脊髓	81.3 ± 2.9	80.9 ± 2.4	75.8 ± 9.0	84.4 ± 5.6
左下颌骨	89.6 ± 2.8	88.1 ± 4.1	89.5 ± 2.4	89.3 ± 1.0
右下颌骨	90.0 ± 1.7	89.0 ± 2.6	88.3 ± 2.5	90.6 ± 1.6
平均值	75.8 ± 7.4	72.3 ± 8.1	72.6 ± 8.8	87.2 ± 6.2

4. 讨论

本文提出一种基于 nnU-Net 融合残差和注意力机制的医学影像危及器官分割新方法, 在 MICCAI 竞赛的真实数据集上进行了验证。实验结果表明, 与现有方法相比, 该模型的分割准确率提升了 11.4%; 对 22 种头颈部危及器官分割的平均骰子相似系数(DSC)为 87.2%。综上, 所提方法能有效提高分割准确率, 有望应用于临床放射治疗实践。但由于本文模型架构相比于对比方法模型更为复杂, 导致模型参数计算量较大, 进一步使网络模型更加简洁是接下来的研究重点。

参考文献

- [1] Wei, W.I. and Sham, J.S. (2005) Nasopharyngeal Carcinoma. *The Lancet*, **365**, 2041-2054. [https://doi.org/10.1016/s0140-6736\(05\)66698-6](https://doi.org/10.1016/s0140-6736(05)66698-6)
- [2] Chang, E.T. and Adami, H. (2006) The Enigmatic Epidemiology of Nasopharyngeal Carcinoma. *Cancer Epidemiology, Biomarkers & Prevention*, **15**, 1765-1777. <https://doi.org/10.1158/1055-9965.epi-06-0353>
- [3] Han, X., Hoogeman, M.S., Levendag, P.C., Hibbard, L.S., Teguh, D.N., Voet, P., et al. (2008) Atlas-Based Auto-Segmentation of Head and Neck CT Images. In: Metaxas, D., Axel, L., Fichtinger, G. and Székely, G., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2008*, Springer, 434-441. https://doi.org/10.1007/978-3-540-85990-1_52
- [4] Daisne, J. and Blumhofer, A. (2013) Atlas-Based Automatic Segmentation of Head and Neck Organs at Risk and Nodal Target Volumes: A Clinical Validation. *Radiation Oncology*, **8**, Article No. 154. <https://doi.org/10.1186/1748-717x-8-154>
- [5] Gollmer, S.T. and M. Buzug, T. (2013) Relaxed Statistical Shape Models for 3D Image Segmentation—Application to Mandible Bone in Cone-Beam CT Data. *Current Medical Imaging Reviews*, **9**, 129-137. <https://doi.org/10.2174/1573405611309020008>
- [6] Fitton, I., Cornelissen, S.A.P., Duppen, J.C., Steenbakkers, R.J.H.M., Peeters, S.T.H., Hoebbers, F.J.P., et al. (2011) Semi-automatic Delineation Using Weighted CT-MRI Registered Images for Radiotherapy of Nasopharyngeal Cancer. *Medical Physics*, **38**, 4662-4666. <https://doi.org/10.1118/1.3611045>
- [7] Zhou, J., Chan, K.L., Xu, P. and Chong, V.F.H. (2006) Nasopharyngeal Carcinoma Lesion Segmentation from MR Images by Support Vector Machine. *3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano*, 2006, Arlington, 6-9 April 2006, 1364-1367. <https://doi.org/10.1109/isbi.2006.1625180>
- [8] Lei, W., Wang, H., Gu, R., Zhang, S., Zhang, S. and Wang, G. (2019) DeepGeoS-V2: Deep Interactive Segmentation of Multiple Organs from Head and Neck Images with Lightweight CNNs. In: Zhou, L., et al., Eds., *Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention*, Springer, 61-69. https://doi.org/10.1007/978-3-030-33642-4_7
- [9] Shin, H., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016) Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, **35**, 1285-1298. <https://doi.org/10.1109/tmi.2016.2528162>
- [10] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/cvpr.2015.7298965>
- [11] Ibragimov, B. and Xing, L. (2017) Segmentation of Organs-At-Risks in Head and Neck CT Images Using Convolutional Neural Networks. *Medical Physics*, **44**, 547-557. <https://doi.org/10.1002/mp.12045>
- [12] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

- [13] Kirillov, A., Heitz, G., Girshick, R., Rother, C. and Dollár, P. (2023) Segment Anything. arXiv: 2304.02643. <https://arxiv.org/abs/2304.02643>
- [14] Gao, Y., Huang, R., Chen, M., Wang, Z., Deng, J., Chen, Y., *et al.* (2019) FocusNet: Imbalanced Large and Small Organ Segmentation with an End-To-End Deep Neural Network for Head and Neck CT Images. In: Shen, D., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*, Springer, 829-838. https://doi.org/10.1007/978-3-030-32248-9_92
- [15] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/cvpr.2016.90>
- [16] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV 2018*, Springer, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [17] Roy, A.G., Navab, N. and Wachinger, C. (2018) Concurrent Spatial and Channel ‘Squeeze & Excitation’ in Fully Convolutional Networks. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C. and Fichtinger, G., Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*, Springer, 421-429. https://doi.org/10.1007/978-3-030-00928-1_48