DOI:10.3969/j.issn.1671-024x.2024.01.010

基于 Transformer 和 CNN 交错混合的肺结节分割网络

吴 骏[1,2],侯宪哲[1,2],王 健[3],肖志涛[2],王 雯[2,4]

(1. 天津工业大学 电子与信息工程学院,天津 300387; 2. 天津工业大学 天津市光电检测技术与系统重点实验室,天津 300387; 3. 93756 部队教研部电子教研室,天津 300131; 4. 天津工业大学 生命科学学院,天津 300387)

摘 要: 针对肺结节尺寸多样、形状异质化高等问题,提出基于 Transformer 和卷积神经网络(CNN)交错混合(IMTC)的肺结节分割网络,该网络是一个对称的层次连接网络,具有很强的多尺度特征提取能力。该网络通过集成2种方案分别解决肺结节多尺寸与形状异质化问题:①采用感知注意力模块(inception attention module, IAM),通过并联多个不同大小的卷积核来增加浅层网络的感受野组合,以此捕获更为丰富的浅层特征;②为获取更具表示能力的高级语义特征,利用由 Transformer 和 CNN 组成的基本骨干网络交错提取结节特征,使得全局特征与局部特征充分融合,从而提高结节特征表示的泛化能力和鲁棒性。实验结果表明:本文模型可以准确分割直径较小以及边缘复杂的肺结节,在 LUNA16 公开数据集上分割性能良好,Dice 和 IOU 分别达到 86.15% 和 76.10%。

关键词: 肺结节;Transformer;卷积神经网络(CNN);感知注意力模块(IAM);交错混合

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1671-024X(2024)01-0074-08

Interlace mixed net of lung nodule segmentation based on Transformer and CNN

WU Jun[1,2], HOU Xianzhe[1,2], WANG Jian[3], XIAO Zhitao[2], WANG Wen[2,4]

(1. School of Electronics and Information Engineering, Tiangong University, Tianjin 300387, China; 2. Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tiangong University, Tianjin 300387, China; 3. Department of Electronic Teaching and Research, 93756 Units, Tianjin 300131, China; 4. School of Life Sciences, Tiangong University, Tianjin 300387, China))

Abstract: Aiming at the problems of multi-size and high heterogeneity of lung nodules, an interlace mixed network based Transformer and convolutional neural network interlace mixed (IMTC) is proposed. The network is a symmetrical hierarchical connection network with strong multi-scale feature extraction capabilities. It integrates two new schemes to solves the promblems of multi-size and shape heterogeneity. ① Inception attention module (IAM) is proposed to capture richer shallow features by paralleling multiple convolution kernels of different sizes to increase the combination of receptive fields. ② In order to extract deeper semantic features with more expressive ability, the basic backbone network composed of Transformer and CNN is used to extract nodule features alternately, so that the global features and local features are fully integrated, and then the generalization ability and robustness of nodule feature representation are improved. The experimental results show that the model in this paper can accurately segment nodules with small scale and complex margin, and has good segmentation performance on the LUNA16 public dataset, and the Dice and IOU reach 86.15% and 76.10%, respectively.

Key words: lung nodules; Transformer; convolutional neural networks(CNN); inception attention module(IAM); interlace mixed

肺癌是最常见的肺部原发性恶性肿瘤,2022 年中国预计约有 482 万新发癌症病例,其中最常见的癌症就是肺癌。肺癌在早期主要表现为肺结节,它们具有尺寸多样(3~30 mm)、对比度低、形状异质化高等特征,若在早期阶段发现这些结节可极大增加人类存活机率[1]。X 射线计算机断层扫描(CT)成像是检测和诊

收稿日期: 2022-09-13

基金项目: 天津市自然科学基金资助项目(21JCZXJC00170);京津冀基础研究合作专项(H2021202008)

通信作者: 吴 骏(1978—),男,博士,副教授,主要研究方向为图像处理与模式识别。E-mail:zhenkongwujun@163.com

断早期肺癌的重要手段,使用这种技术会产生大量的 CT 影像数据,在庞大的 CT 数据中检测肺结节对于放射科医生来说极其具有挑战性,因此开发一个稳健的自动肺结节分割模型对于避免繁琐的手工治疗和减少医生之间的诊断差异具有重要的临床意义[2]。

近些年,卷积神经网络(CNN)在医学图像分割等视觉任务中取得了长足的进步,并且随着 U-Net[3]网络在医学图像分割任务中的良好表现,一些基于编码-解码结构的变体网络也随之应用于肺结节分割。Tang 等[4]提出了一种新的端到端的多任务三维深度卷积神经网络(DCNN),该网络提出以多任务的方式联合解决结节检测、假阳性减少和结节分割问题。Singadkar 等[5]提出了一种基于深度反卷积残差网络的 CT 图像分割肺结节的方法,该方法以深度反卷积残差网络进行端到端训练,并在网络卷积部分到反卷积部分添加长跳跃连接从而保存了因池化操作过程丢失的空间信息,捕获全分辨率特征。但是上述方法均存在两方面问题:①结节具有复杂的形状和高度异质性纹理特征,仅使用基于局部偏置方法提取结节高级语义特征无法映射为高质量的分割特征图,以至于对具有非规则形状特征的肺结节造成欠分割;②虽然基于 CNN 方法具有很好的表示能力,但由于卷积局部归纳偏置的局限性使得网络很难建立一种显式的长距离依赖关系。卷积运算的局限性给学习全局语义信息增加了挑战,而全局语义信息对于肺结节分割任务至关重要。

基于 CNN 在提取特征时的局部特性和权重分配的归纳偏差能力,许多视觉任务致力于扩大 CNN 的接受感受野来提高网络上下文建模的能力[6-7],然而卷积操作的局部性仍然将感受野限制在一个较小的区域。受 Transformer[8]在 NLP 成功的启发,多项计算机视觉研究尝试将 Transformer 直接应用于视觉任务,利用 Transformer 中自注意力机制先天性远程捕获的能力建模全局依赖关系。例如 Swin Transformer[9]通过引入卷积特征金字塔构建方式构建层次化 Transformer,并提出错位窗口移动方案,使得计算复杂度大大降低。

基于 Swin Transformer 在视觉任务的成功应用,本文提出一种将 Swin Transformer 与 CNN 相交错混合的肺结节分割网络(IMTC,interlace mixed net based on Transformer and CNN),主要工作有以下几个方面:

(1)为避免由于网络过深而导致小直径结节特征和边缘信息消失的问题,设计一种多个不同卷积核并联组合的感知注意力模块(inception attention module,IAM),通过增加浅层网络宽度来增加网络的感受野组合,以此捕获更为丰富的浅层特征。

(2)针对结节大尺寸与高度形状异质化等问题,设计了一种由 Transformer 和 CNN 组成的基本骨干网络交错提取结节特征以捕获更为丰富的深层信息。这么做的目的有 2 个:①在 Transformer 后面加上卷积层,可以将卷积操作的归纳偏置特性引入到 Transformer 提取的全局特征中,有益于特征聚合;②通过组成 Transformer 与 CNN 相结合的交错网络,可以使得全局特征与局部特征充分融合。

(3)将编码器生成的不同尺度的高分辨率特征通过跳跃连接与解码器中相对应位置特征进行融合,以生成更为精细的重建特征。

1　研究方法

本文提出的 IMTC 网络结构如图 1 所示。

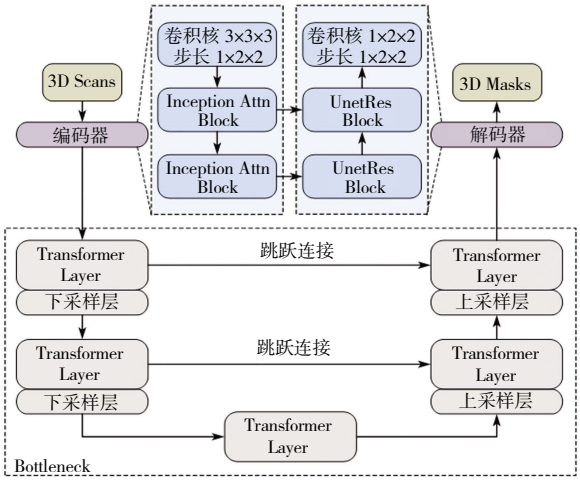


图 1　IMTC 网络结构示意

Fig.1　Schematic diagram of IMTC network structure

IMTC 是基于 U-Net 类型的双对称编码-解码结构,主要由编码器、Bottleneck 和解码器 3 个部分组成。编码器由 1 个卷积模块和 2 个 IAM 模块组成。为降低网络计算成本,首先经过卷积层将输入信息降采样,然后利用 2 个连续的 IAM 模块分别在不同感受野条件下充分提取结节浅层信息。此外,为了给解码器提供足够大的感受域,基于 Transformer Layer、下采样层和上采样层共同组成 Bottleneck。解码器与编码器结构相似,由 2 个 UnetRes Block 和扩展层组成,通过 2 个连续的 UnetRes Block 将提取的深层语义特征重建为更具识别能力的高分辨图像,并通过扩展层将重建的高分辨图像恢复至原分辨率进行像素级语义预测。此外,为了生成更为精细的重建特征,将分别包含有结节边缘信息的浅层特征以及下采样过程中生成的低分辨率高级语义特征通过跳跃连接方式传递至

上采样相同分辨率的位置，以进一步提高预测结果。下面将对模型各个结构进行更为详尽的叙述。

1.1　编码器

由于小结节的存在会在很大程度上限制网络的降采样能力，换句话说，小直径结节特征很容易在深层网络的一次次下采样和上采样过程丢失掉，这个时候可能需要感受野小的特征来帮助。因此若网络可以拥有不同大小的感受野，分割效果会更好。受 GoogLeNet[10] 启发，针对结节多尺度问题，本文提出感知注意力模块（IAM）。IAM 网络结构图分别由 Inception Block 和 Attention Module 组成，并以多残差方式连接，如图 2 所示。

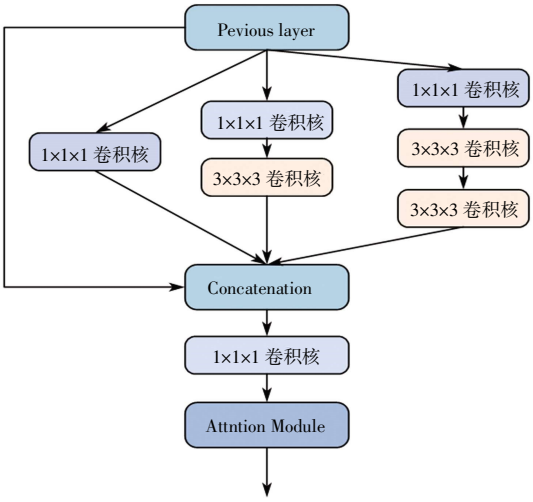


图 2　IAM 结构示意

Fig.2　Structure schematic of inception attention module

对于 Inception Block，为增加网络的感受野组合，分别结合 kernel 为 1×1×1、3×3×3 和 5×5×5 的卷积层提取结节特征并进行 Concat，从而有效提高对原始像素信息与网络内部资源的利用率。此外，为降低 Inception Block 计算成本，在 3×3×3 和 5×5×5 的卷积层前分别加 1×1×1 卷积对特征进行降维，然后再进行卷积运算，并且使用 2 个 3×3×3 卷积将 5×5×5 卷积进行替代，使得在减少参数量的同时实现相同大小的感受野。为增加网络非线性表达能力，该模块在每次卷积层后，均添加有 Layer Normalization[11] 和 GELU[12] 激活函数。

为了使融合特征具有更强劲的表征能力，本文设计了注意力模块（Attention Module），该模块是基于通道和空间注意力模块分别增强融合特征，其网络结构图如图 3 所示。

对于通道注意力模块，由于编码特征在前期训练阶段容易包含不相干区域，故基于规一化注意力模块

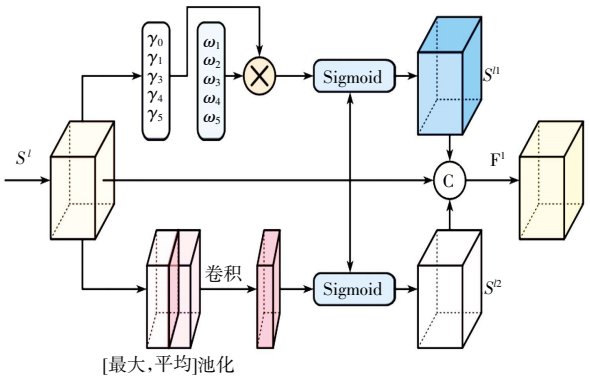


图 3　Attention Module 细节展示

Fig.3　Details show of Attention Module

NAM[13]（normalization-based attention module）中对归一化权重因素的重要性考虑，通过对权重做稀释惩罚进一步抑制不显著的通道或特征，进而增强重要特征信息。对于空间注意力模块，借助 CBAM [14]（Convolutional Block Attention Module）中空间滤波器池化与卷积等操作再次学习融合特征中的局部细节，以增强有用信息的传递。

最后将增强特征与初始融合特征进行 Concat 连接，作为最终的信息并传递至下一过程。通道注意力和空间注意力模块计算方式如式（1）所示，Attention Module 计算过程如式（2）所示：

$$\begin{cases} \hat{S}^{l1} = \sigma(W_\gamma(\mathrm{LN}(S^l))) \\ \hat{S}^{l2} = \sigma(\mathrm{f}^{7\times7\times7}([A(S^l);\mathrm{Maxpool}(S^l)])) \end{cases} \tag{1}$$

$$F^l = \mathrm{Concat}(S^l + \hat{S}^{l1} + \hat{S}^{l2}) \tag{2}$$

式中：l 为当前融合模块的索引值；W_γ 表示特征和的加权值；$\mathrm{f}^{7\times7\times7}$ 表示卷积核大小为 $7\times7\times7$ 的卷积操作；σ 表示 sigmoid 函数；LN 表示层归一化操作；F^l 为融合模块最终输出结果。

1.2　Bottleneck

若网络仅由连续的 IAM 模块组成，这样做虽可以很好地分割小直径结节，但是对于大直径和形状多变的肺结节会导致欠分割效果，因为这样的肺结节在对边缘进行分割的时候需要更深的网络结构。故基于对大尺度肺结节与形状高度异质化等问题的考虑，该方法利用由 Transformer 和 CNN 组成的基本骨干网络交错提取结节特征，从而为解码器提供足够大的感受域。在该方法中，Transformer Layer 利用其全局自注意机制增强特征全局上下文表征能力和建模远程依赖能力，卷积层依赖于卷积固有的归纳偏置特性可以将表征全局信息的特征进行局部增强，两者通过交错连接共同获取在多个尺度下特征的长期依赖

关系与局部空间信息,有助于提高特征表示的泛化能力和鲁棒性。

由于使用传统的多头自注意力（multi head selfattention, MHSA）方法具有极高的计算复杂度，参考 Swin Transformer 中提出的移位窗口机制计算窗口间自注意力方法,并将其应用在 3D 体积运算中。该方法利用非重叠窗口划分方法将 3D 体积数据划分为一个个局部体积块，并将自注意力限制在局部体积块内，降低计算复杂度为输入体积大小的线性关系。然后利用移位窗口方案增加体积块之间信息交互,从而建模全局上下文和捕获体素之间远程依赖关系。

Transformer Layer 网络结构如图 4 所示。

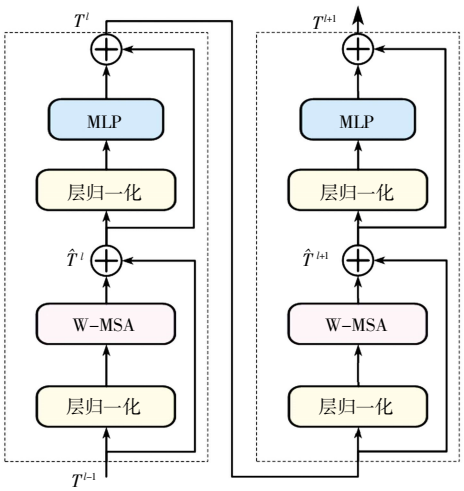


图 4　Transformer Layer 结构示意

Fig.4　Structure diagram of Transformer Layer

由图 4 可见,Transformer Layer 由 2 个连续Transformer 模块组成,分别表示为窗口多头自注意力(Window Multi-Head Self-Attention, W-MSA)模块和移位窗口多头自注意力（Shift Window Multi-Head Self-Attention,SW-MSA)模块,其中第 2 层表示为第 1 层的移位窗口版本。假设 $T \in R^{D\times H\times W\times C}$ 表示输入数据,W-MSA 首先利用非重叠窗口划分方法将数据重新表示为 $T \in R^{\frac{D\times H\times W}{M^3}\times M^3\times C}$。其中:$M$ 表示局部体积块长度;M^3 表示为非重叠体积块中体数量;$\frac{D\times H\times W}{M^3}$表示体积块的数量。然后对每个局部体积块进行自注意力计算,自注意力计算方式如公式(3)所示:

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{SoftMax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}} + B\right)V \quad (3)$$

式中:$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V} \in R^{M^3\times d}$ 分别表示 query、key 和 value 矩阵;$B \in R^{M^3}$ 表示相对位置编码。

为了促进非重叠体积块之间的信息交互,Transformer Layer 引入 SW-MSA,该模块是在 W-MSA 基础上对局部体积块移位，$\left(\left[\frac{M}{2},\frac{M}{2},\frac{M}{2}\right]\right)$ Transformer Layer 计算过程如式(4):

$$\begin{cases} \hat{T}^l = \text{W-MSA}(\text{LN}(T^{l-1})) + T^{l-1} \\ T^l = \text{MLP}(\text{LN}(\hat{T}^l)) + \hat{T}^l \\ \hat{T}^{l+1} = \text{SW-MSA}(\text{LN}(T^l)) + T \\ T^{l+1} = \text{MLP}(\text{LN}(\hat{T}^{l+1})) + \hat{T}^{l+1} \end{cases} \quad (4)$$

式中:l 表示当前 Transformer layer 索引值;LN 表示层归一化操作;$\hat{T}^l$ 与 T^l 分别表示(S)W-MSA 模块与多层感知机(MLP)输出结果。

在 Transformer Layer 后引入卷积层可以利用卷积操作固有的归纳偏置特性对获取的全局特征进行局部约束，从而提高特征表示的泛化能力和鲁棒性。IMTC 中模块细节如图 5 所示。Down-sampling 仅由 1 个 kernel 为 3 × 3 × 3 和 stride 为 2 × 2 × 2 的卷积层组成,且在卷积操作后添加 Layer Normalization 和 GELU 激活函数。为了验证不同的结果,该方法尝试使用不同下采样方式对全局特征进行约束,实际上,由于数据切片数量有限,过度下采样使得最终分割效果并不理想，所以综合考虑只在 Down-sampling 使用 1 个 kernel 为 3 × 3 × 3 的卷积进行特征下采样。

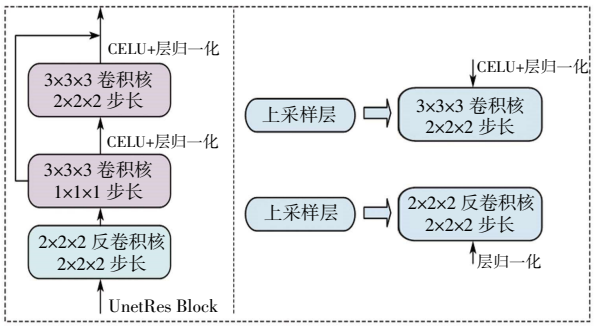


图 5　IMTC 中模块细节展示

Fig.5　Details of modules in IMTC

1.3　解码器

为了生成与输入医学图像相同分辨率的分割结果，解码器基于 2 个 UnetRes Block 和扩展层组成,用于对低分辨高级语义特征解码并生成最后的像素级分割结果。此外,为了弥合在下采样过程丢失的空间信息以及捕获更多语义信息，在编码器-解码器相同尺度内添加跳跃连接以重建更为精细的特征信息。同时,通过向具有不同尺度的解码器输出添加辅助损失来使用深监督策略,损失函数为 Dice 损失。

图 5 中,UnetRes Block 由 1 个 kernel 为 2×2×2 和 stride 为 2×2×2 的反卷积层和 2 个连续的 kernel 为 3×3×3和 stride 为 1×1×1 的卷积层依次连接组成。该 Block 首先对输入特征进行 2 倍上采样, 然后将采样特征送入连续卷积层学习特征之间的相关性。此外,为避免在训练过程中梯度消失,在连续卷积层之间使用残差连接。

2　实验结果分析

2.1　实验数据

本次实验使用 LUNA16[15]公开数据集来评估 IMTC 的性能,LUNA16 来自于一个更大的公共数据集 LIDC-LDRI[16],该数据集包含从多个胸部部位收集的具有不同厚片切度的 1 018 组 CT 影像,且直径等于或大于 3 mm 的结节由多达 4 位专家分别勾勒其轮廓。为保证数据集样本规范化,LUNA16 去除切片厚度大于 3 mm 和肺结节小于 3 mm 的 CT 影像,并将中心距离小于半径之和的 2 个相邻结节合并,所以共得到 2 290、1 602、1 186、777 个至少由 1、2、3、4 个专家标注的结节,LUNA16 选取至少由 3 位专家标注的 1 186 个结节作为最后要检测的区域。在本次实验中随机抽取 815 个肺结节作为训练集,验证集和测试集分别为 128 和 243 个肺结节, 并将数据集裁剪成大小为(64, 128, 128)的格式后输入网络。

2.2　评价指标

本次实验使用 Dice 和 IOU 评估指标来衡量模型分割精度。Dice 是医学图像分割比赛中使用频率最高的衡量指标,是一种集合相似度度量指标,用于衡量 2 个样本之间的相似度, 主要对分割内部填充比较敏感,该评价指标越高表示分割的效果越好。IOU 同样是描述 2 组点集之间相似程度的一种度量,但该指标重点集中于衡量两组点集之间的重叠度,Dice 和 IOU 计算公式分别如式(5)和式(6)所示:

$$\mathrm{Dice}(A,B)=\frac{2|A\cap B|}{|A|+|B|} \tag{5}$$

$$\mathrm{IOU}(A,B)=\frac{|A\cap B|}{|A\cup B|} \tag{6}$$

式中:A 和 B 分别表示预测值与真实值的集合;∩表示预测值与真实值之间的交集;∪表示预测值与真实值之间的并集。

2.3　实验环境及超参数设置

硬件配置为 GPU:2 个 NVIDIA 2080Ti 显卡;深度学习框架为 Pytorch1.6.0 + python3.6,网络在训练时采用 Adam 优化器训练模型,权重衰减系数为 $1e^{-4}$,初始学习率为 0.000 1,batch size 为 4,共训练 300 个epoch,并使用'ploy'学习率衰减方法。为减轻有限数据的过拟合问题,还采用了中心点随机裁剪、旋转、随机缩放、高斯白噪声、高斯模糊、调整亮度和对比度数据增强策略,以增加数据集多样化。

2.4　实验结果与分析

2.4.1　实验结果对比

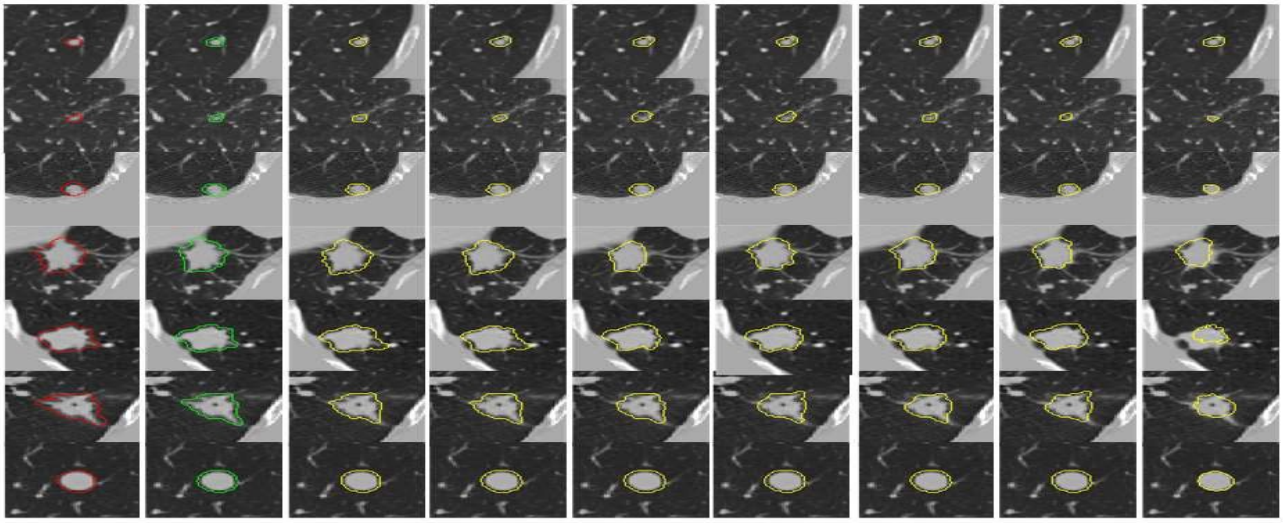
将所提出方法与不同肺结节分割网络模型进行了比较,并根据定量结果与定性结果进行了评估。在定量分析方面,为了得到更为全面的分析结果,使用 Dice 和 IOU 作为主要评估指标。对比的肺结节分割模型包括 SegNet[17]、VNet[18]、3D ResU-Net[19]、Attention-U-Net[20]、3D U-Net[21]、3D U-Net++[22]和 nnFormer。本文方法与不同分割网络模型的分割评价指标如表 1 所示。由表 1 可见本文提出的分割模型取得了最好的分割性能。

表 1　与不同分割网络模型比较的分割评价指标

Tab.1　Segmentation evaluation indexes compared with those of different segmentation network models

网络结构	Dice	IOU
SegNet	0.769 4 ± 0.101 0	0.672 9 ± 0.114 7
VNet	0.783 7 ± 0.087 2	0.682 4 ± 0.105 9
3D ResU-Net	0.820 6 ± 0.068 0	0.719 1 ± 0.089 0
Attention-U-Net	0.819 1 ± 0.075 5	0.716 3 ± 0.093 1
3D U-Net	0.825 8 ± 0.061 2	0.716 8 ± 0.086 3
3D U-Net++	0.832 9 ± 0.060 2	0.727 7 ± 0.081 9
nnFormer	0.845 1 ± 0.053 4	0.741 7 ± 0.079 2
IMTC	0.861 5 ± 0.048 2	0.761 0 ± 0.072 4

图 6 展现了不同种类的肺结节在上述几种网络中的分割结果与本文提出的 IMTC 方法的分割结果对比图。图 6 中从上到下可依次将肺结节分为小直径结节和大直径结节,其中小直径结节和大直径结节中又包括规则形状结节与非规则形状结节。本文以图 6(a)为标准对各分割方法进行分析,通过分析可见,对于各种类型的肺结节,图 6(b)得到的分割结果始终与图 6(a)最为相似。图 6(c)得益于 Transformer 与 CNN 相结合的结构,取得相对较好的分割结果。图 6(c)通过在多个级别上添加密集的跳跃连接, 改进了 U 型结构,取得不错的分割结果。图 6(f)—图 6(i)则表现出较差的分割结果,虽然这几种方法均可以初步定位肺结节的位置信息,但对于不规则形状肺结节的分割结果均出现不同程度的误分割与欠分割,这可能是由卷积操作的局部性造成的,表明这几种网络对特征分布复杂的肺结节目标在特征提取方面还有待加强。



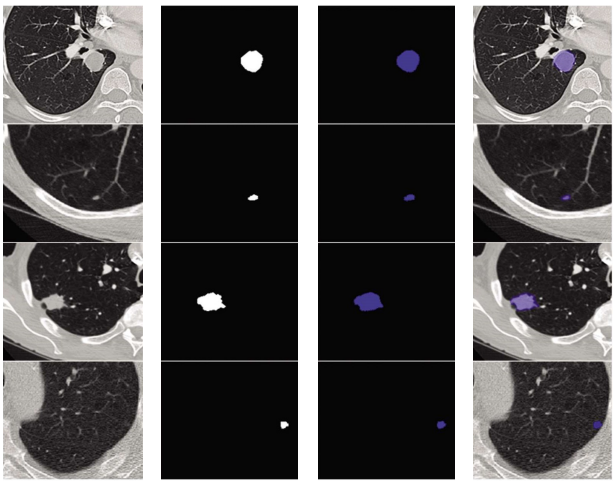
(a) 金标准标签　(b) IMTC 分割结果　(c) nnFormer 分割结果　(d) 3DU-Net++ 分割结果　(e) 3D U-Net 分割结果　(f) Attention-U-Net 分割结果　(g) 3D ResU-Net 分割结果　(h) VNet 分割结果　(i) SegNet 分割结果

注:其中金标准、本文提出方法的分割结果和其他方法的分割结果以红色、绿色和黄色分别标出。

图 6　肺结节分割可视化对比

Fig.6　Visual comparison of lung nodule segmentation

图 7 展示了提出方法在不同肺结节类别上的分割结果,从上到下依次为孤立性结节、血管黏连性结节、胸腔黏连性结节和磨玻璃结节。由图 7 可见,针对于不同种类的肺结节,图 7(b)与图 7(c)虽然在边缘位置有一些不同,但是在大致位置与结节形状还是非常相似的。证明该模型可以成功分割黏连性结节与磨玻璃结节,且分割结果与 Ground truth 极为相近。



(a) 原始 CT　(b) 金标准　(c) 分割结果　(d) 最终分割结果

注:从上到下依次表示为:孤立性结节、血管黏连性结节、胸腔黏连性结节和磨玻璃结节。

图 7　肺结节分割结果

Fig.7　Lung nodules segmentation results

综合比较,本文提出网络模型性能优于上述几种方法,说明 IMTC 对于边缘复杂且具有不同尺度的肺结节可以得到较好的分割效果。因为提出的网络不仅是基于端到端训练,而且利用 IAM 模块可以捕获更为丰富的浅层特征,能很好地识别目标的轮廓等细节特征。此外,通过 Transformer 与卷积操作交错组合的下采样结构可以更好的学习结节特征在全局与局部之间的信息交互,从而产生更好的分割结果。

2.4.2　消融实验

为了探讨不同因素对模型性能的影响,本文分别对 IAM 模块组合、注意力模块组合、下采样方式和输入尺寸进行了消融实验。

本方法基于浅层特征对于肺结节分割重要性的考虑,将 IAM 模块分别放在第 2 次和第 3 次下采样过程中。在实际分割中,肺结节的边缘信息和小直径结节信息很容易在深层网络的一次次的下采样过程中丢失其原有特征信息,因此若在浅层网络提取结节特征时拥有不同大小的感受野,可以产生更好的分割效果。本文通过消融实验来验证 IAM 模块在网络不同位置处的有效性和实用性,其上角标表示 IAM 模块添加到第 n 次下采样的位置,如表 2 所示。由表 2 可见,本文所提出方法获得最好的分割精度。

表 2　IAM 模块不同组合设计的实验结果对比

Tab.2　Comparison of experimental results of different combination designs of IAM modules

网络名称	Dice	IOU
$IMTC^{2}$	0.851 2 ± 0.064 5	0.747 2 ± 0.091 0
$IMTC^{3}$	0.857 7 ± 0.054 0	0.756 1 ± 0.079 1
$IMTC^{2,3}$	0.861 5 ± 0.048 2	0.761 0 ± 0.072 4

表 3 所示为 IAM 模块中空间注意力与通道注意力在不同组合方式下对分割性能的影响。表 3 中 BackBone 表示提出方法将 IAM 完全剔除掉的网络，可以看出，在 BackBone 不变的情况下，在逐渐增加模块的同时肺结节评估指标也在稳步上升，而且全部模块组合的网络取得了最好的结果。

表 3　注意力模块不同组合方式的实验结果对比

Tab.3　Comparison of experimental results with different combinations of attention modules

网络结构	Dice	IOU
BackBone	0.837 9 ± 0.065 5	0.735 3 ± 0.091 9
+Inception Block	0.846 9 ± 0.062 3	0.740 6 ± 0.088 7
+Channel Attn	0.855 7 ± 0.054 7	0.753 0 ± 0.080 2
+Spatial Attn	0.861 5 ± 0.048 2	0.761 0 ± 0.072 4

为了探索不同大小的卷积核对于全局特征的归纳偏置能力，本文针对 Bottleneck 中的下采样方式进行消融实验，其中 kernel = 2 × 2 × 2 和 kernel = 3 × 3 × 3 分别表示下采样方式采用的不同卷积核的大小，patch merging 表示 Swin-Unet[23]所采用的全连接下采样方式。表 4 所示为使用卷积核为 3 × 3 × 3 的下采样方式可以获得最好的分割精度。

表 4　不同下采样方式的实验结果对比

Tab.4　Comparison of experimental resluts with different Down-sampling

下采样方式	Dice	IOU
kernel = 2 × 2 × 2	0.854 3 ± 0.058 0	0.751 3 ± 0.084 6
patch merging	0.854 9 ± 0.056 9	0.740 6 ± 0.088 7
kernel = 3 × 3 × 3	0.861 5 ± 0.048 2	0.761 0 ± 0.072 4

本文基于肺结节数据集中肺结节尺寸大小与胸腔背景之间的比例考量，决定将输入网络的数据尺寸裁剪为。因为若裁剪尺寸过小，则不能很好的学习肺结节边缘与胸腔背景之间的关系，可能会造成误分割。而若裁剪尺寸过大，则会导致肺结节目标与胸腔背景比例严重失衡，从而造成欠分割。当网络输入为 64 × 128 × 128 时取得最好的分割效果，如表 5 所示。

表 5　不同输入尺寸的实验结果对比

Tab.5　Comparison of experimental results with different input sizes

输入尺寸	Dice	IOU
64 × 64 × 64	0.843 6 ± 0.065 7	0.736 4 ± 0.092 6
128 × 128 × 128	0.835 1 ± 0.065 8	0.731 4 ± 0.092 4
64 × 128 × 128	0.861 5 ± 0.048 2	0.761 0 ± 0.072 4

本文通过大量实验定量分析与定性分析了所提出模型的性能。对于边缘复杂且尺度不一的肺结节，本文模型均可以准确地预测出结节位置，而且对于一些难分割结节比如黏连性结节和磨玻璃结节也表现出较好的性能，虽然预测结果可能稍有误差，但总的来看分割结果和金标准数据比较接近。此外，通过对比几种分割网络对于大尺度边缘复杂的肺结节和小直径结节 2 种情况的分割结果，均显示本文提出方法得到分割效果最好，进一步验证了本文方法能够对肺结节进行精确分割，充分说明说明本文所提出方法适用于三维 CT 图像的肺结节分割任务。

3　结　论

本文基于对肺结节多尺度与多形状等问题设计网络模型，提出基于 Transformer 与 CNN 交错混合的肺结节分割方法。为了解决肺结节在不同网络深度下特征提取的有效性，分别设计 IAM 模块和基于 Transformer layer 与 CNN 交错提取特征的骨干网络。针对于小直径结节以及边缘复杂的结节，可以利用 IAM 模块通过拓宽网络宽度充分提取浅层细节特征。对于大直径和多形状变化的结节，该方法利用骨干网络交错提取结节特征，其中 Transformer layer 主要使特征建立全局依赖关系，从而增强全局上下文建模能力；而卷积层可以利用其固有的归纳偏置特性可以将表征全局信息的特征进行局部增强，两者相互补充，相辅相成。此外，该方法还通过跳跃连接将来自编码器的不同尺度的高分辨率特征与解码器上采样特征进行融合，以弥合由于下采样过程导致的空间信息丢失。实验表明，本文所提出的模型展现出良好的肺结节分割效果，Dice 和 IOU 分别达到 86.15%和 76.10%。

参考文献：

[1] DIEDERICH S, WORMANNS D, SEMIK M, et al. Screening for early lung cancer with low-dose spiral CT: Prevalence in 817 asymptomatic smokers[J]. Radiology, 2002, 222(3): 773-781.

[2] REEVES A P, CHAN A B, YANKELEVITZ D F, et al. On measuring the change in size of pulmonary nodules[J]. IEEE Transactions on Medical Imaging, 2006, 25(4): 435-450.

[3] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Lecture Notes in Computer Science, 2015: 234-241.

[4] TANG H, ZHANG C P, XIE X H. NoduleNet: Decoupled false positive reduction for pulmonary nodule detection and segmentation[C]//International Conference on Medical Image

Computing and Computer-Assisted Intervention. Shenzhen, China: Springer, 2019: 266-274.

[5] SINGADKAR G, MAHAJAN A, THAKUR M, et al. Deep deconvolutional residual network based automatic lung nodule segmentation[J]. Journal of Digital Imaging, 2020, 33(3): 678-684.

[6] ROCHA J, CUNHA A, MENDON?A A M. Conventional filtering versus U-net based models for pulmonary nodule segmentation in CT images[J]. Journal of Medical Systems, 2020, 44(4): 1-8.

[7] GU Z W, CHENG J, FU H Z, et al. CE-net: Context encoder network for 2D medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281-2292.

[8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30:5998-6008.

[9] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2022: 9992-10002.

[10] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 1-9.

[11] BA J L, KIROS J R, HINTON G E. Layer normalization[J]. 2016. DOI:10.48550/arXiv:1607.06450.

[12] HENDRYCKS D, GIMPEL K. Gaussian error linear units (gelus)[J]. 2016. DOI:10.48550/arXiv:1606.08415.

[13] LIU Y C, SHAO Z R, TENG Y Y, et al. NAM: normalization-based attention module[J]. 2021. DOI:10.48550/arXiv: 2111.12419.

[14] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision. Cham: Springer, 2018: 3-19.

[15] SETIO A A A, TRAVERSO A, DE BEL T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge [J]. Medical Image Analysis, 2017, 42: 1-13.

[16] ARMATO III S G, MCLENNAN G, BIDAUT L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans[J]. Medical Physics, 2011, 38(2): 915-931.

[17] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (12): 2481-2495.

[18] MILLETARI F, NAVAB N, AHMADI S A. V-net: fully convolutional neural networks for volumetric medical image segmentation [C]//2016 Fourth International Conference on 3D Vision (3DV). October 25-28, 2016, Stanford, CA, USA. IEEE, 2016: 565-571.

[19] YU L, YANG X, CHEN H, et al. Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images[C]//Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. New York: ACM, 2017: 66-72.

[20] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: Learning where to look for the pancreas[J]. 2018. DOI: 10.48550/arXiv.1804.03999.

[21] ÇIÇEK Ö, AABDULKADIR A, LIENKAMP S S, et al. 3D U-net: Learning dense volumetric segmentation from sparse annotation[C]//Medical Image Computing and Computer-Assisted Intervention –MICCAI 2016: 19th International Conference. Athens, Greece: Springer International Publishing, 2016: 424-432.

[22] ZHOU Z W, RAHMAN SIDDIQUEE M M, TAJBAKHSH N, et al. "Unet++: A nested u-net architecture for medical image segmentation." Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop. Granada, Spain, Cham: Springer, 2018: 3-11.

[23] CAO H, WANG Y, CHEN J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation[C]//European Conference on ComputerVision(ECCV). Israel. Cham: Springer Nature Switzerland, 2022: 205-218.

本文引文格式:

吴骏,侯宪哲,王健,等. 基于 Transformer 和 CNN 交错混合的肺结节分割网络[J]. 天津工业大学学报, 2024, 43(1): 74-81.

WU J, HOU X Z, WANG J, et al. Interlace mixed net of lung nodule segmentation based on Transformer and CNN[J]. Journal of Tiangong University, 2024, 43(1): 74-81(in Chinese).