

DOI:10.3969/j.issn.1005-202X.2024.12.007

医学影像物理

基于全尺度通道特征聚合编解码网络的肺结节分割算法

谢绍鹏,王明泉,耿宇杰,黄心玥,商然
中北大学信息与通信工程学院,山西 太原 030051

【摘要】针对不同性质下肺结节难以精准检测的难题,提出全尺度通道特征聚合编解码网络(FCFA-Net)去辅助经验丰富的医师进行诊断。本网络由SMC、全尺度特征聚合器、自相关特征增强器、通道特征层级提取解码器以及双项限制损失函数构成,用于充分提取CT图像中的浅层与深层特征,以达到分割大小各异、形状各异的肺结节病灶。本文方法FCFA-Net网络相较于UNet、UNet++和TransUnet方法精准率分别提升9.66%、7.84%、3.75%,召回率提升5.50%、2.96%、1.37%,均交并比提升11.35%、7.16%、4.18%,F1分数提升8.07%、5.87%、3.10%,且消融实验表明各个结构均发挥作用,在参数接受范围内达到最佳效果。

【关键词】肺结节分割;全尺度跳跃连接;双项限制损失函数;空洞卷积

【中图分类号】R318;TP391.41

【文献标志码】A

【文章编号】1005-202X(2024)12-1501-08

Lung nodule segmentation algorithm based on full-scale channel feature aggregation coding and decoding network

XIE Shaopeng, WANG Mingquan, GENG Yujie, HUANG Xinyue, SHANG Ran

School of Information and Communication Engineering, North University of China, Taiyuan 030051, China

Abstract: To address the difficulty in accurately detecting pulmonary nodules of different properties, a full-scale channel feature aggregation encoding and decoding network (FCFA-Net) is employed to assist experienced physicians in diagnosis. The network which consists of SMC, full-scale feature aggregator, autocorrelation feature enhancer, channel feature hierarchy extraction decoder and binomial constraint loss function can fully extract shallow and deep features from CT images for realizing the segmentation of pulmonary nodules of different sizes and shapes. Compared with UNet, UNet++ and TransUnet, FCFA-Net increases the accuracy by 9.96%, 7.84% and 3.75%, recall rate by 5.50%, 2.96% and 1.37%, mean intersection over union by 11.35%, 7.16% and 4.18%, F1 score by 8.07%, 5.87% and 3.10%, respectively. Additionally, ablation experiment results demonstrate that each structure is effective and can achieve the best result within the acceptable parameter range.

Keywords: lung nodule segmentation; full-scale skip connection; binomial constraint loss function; dilated convolution

前言

肺部结节是一种影像学的概念,通常指胸片或者计算机断层扫描(CT)上直径小于3 cm的团块状

阴影,反映肺部的实质性病灶。但区域内除病灶外的其他有形物质,也会在CT上显示为白色形状,对医师的诊断、分析和治疗产生很大的影响。目前辅助诊断的方法主要分为两种:第1种方法是利用边缘检测等传统图像处理方法对影像进行简单的分割;第2种方法是通过深度学习的方式将肺结节区域分割出来^[1]。相较于第1种方法,第2种方法在辅助诊断方面明显更具优势,可以利用高维的特征更准确地划分。传统方法上,Savic等^[2]提出区域的快速行进法,将图像分割成特征相似的区域,然后结合k均值增长的区域进行合并分割实质性结节,分割情况较好,但是对于非实质性结节的分割结果较差。Wang等^[3]提出一种全方差增强金字塔和改进Grab切

【收稿日期】2024-07-27

【基金项目】国家自然科学基金(61171177);国家重大科学仪器设备开发项目(2013YQ240803);山西省重点研发计划(201803D121069);山西省高等学校科技创新项目(2020L0624);山西省信息探测与处理重点实验室基金(ISPT-2020-5)

【作者简介】谢绍鹏,硕士研究生,研究方向:图像处理与人工智能,
E-mail: 2090117685@qq.com

【通信作者】王明泉,教授,博士,研究方向:智能信息处理、图像处理与信息反演等,E-mail: 3087378457@qq.com

割相结合的交互式分割方法,引入多尺度处理和优化能量项,克服不均匀性和轮廓模糊的缺点,但是由于医学图像周围灰色区域与背景的相似性较高,经常会出现假性检测。

针对深度学习的改进大多集中在两个部分:一种是改进网络基本块;另一种是通过更改网络基本结构实现改进。Li等^[4]通过将Resnet的变体ResNeSt作为网络主干,引入空洞空间金字塔池化等手段来提升网络性能。Wang等^[5]利用高精度肺结节分割模型SKV-Net在Luna16数据集中取得的Dice分数为79.6%。Wu等^[6]利用结合图像增强和双分支神经网络的分割方法,有效地从二维切片和相邻切片的关系之中发掘信息,在LIDC数据集以及其建立的非孤立结节数据集上拥有着较好的成效。Wang等^[7]针对UNet模型未能充分利用肺结节图像分割中编码层之间的权重关系和语义信息而导致分割精度不足的问题,利用并行注意模块完成肺结节图像的信息编码,并重置采样相关通道信息的权重。Zhang等^[8]提出一种将Transformer、UNet与局部二值能量函数模型相结合的两阶段模型Tunet-LBF,在用于视网膜血管的粗细分割上取得优异的效果。Chen等^[9]通过将Transformer与UNet结合提出一种TransUNet模型,实现腹部器官的精准分割。

以上的方法均是通过对传统方法的进一步探究或者针对分割网络的某一结构设计出发来提高网络的性能,但是传统方法的根本问题在于其鲁棒性较差,通常只能针对某一种特定状态下的肺结节进行分割。深度学习方法通常通过引入特征增强、更换主干或者注意力机制等,但是多数仍采用卷积层作为主要的特征提取方式,其较小的感受野所带来的网络性能下降仍是需要解决的问题。

1 问题与场景

通过CT影像对肺结节的分割任务是计算机辅助诊断癌症的一个关键组成部分,然而对于其图像内的相似成分过高,仍是一种极具挑战性的项目^[10]。针对上述问题以及难点,从分割网络整体的5个部分结构出发,通过分析特征的流向,逐一设计每个部分,以达到理想的效果。主干结构:首先将SwinTransformer引入主干网络,通过引入卷积层解决其4倍下采样带来的信息损失问题,将其作为一种特征集中的模块,添加到主干之中,为后面提供更为准确的语义信息。特征增强:针对主干传递过来的深层次、小尺寸、多通道、较大感受野的语义信息,其每一个样本以及样本之间的关系都十分重要,由此将具有全局性的Transformer引入,在特征增强的

同时,获取各个样本之间的相干关系,捕获长期的依赖。跳跃连接:通过前者的成功可以相对明显看出,在编解码器之间添加合理的特征通路是能够有效提升网络性能的。由此添加一个较为激进的全尺度跳跃连接,让越浅层的解码器拥有越丰富的信息去生成像素之间的关系。解码器:对于解码器来说,通常是为了前端的设计而服务的,仅由上采样和一部分卷积构成,采用叠加空洞卷积的方式,让其分别注意不同感受野下的信息。损失函数:损失函数的设计对于网络的性能是极为重要的,分割的损失函数通常有两种,分别基于分布的损失和区域的损失,通过选择一种较为合理的结合方式去作为当前网络的损失函数。

2 网络结构

2.1 整体结构

全尺度通道特征聚合编解码网络(FCFA-Net)整体结构如图1所示,采用编解码器的形式对肺结节病灶进行分割。其中输入图像尺寸为 $-1 \times 512 \times 512 \times 3$ 的肺结节影像,输出图像尺寸为 $-1 \times 512 \times 512 \times 2$ 的标签图像,通过设计的网络结构,增强对小目标的分割能力,实现对全尺度信息更高效的利用,提高分割效果。

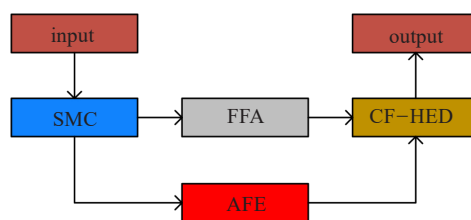


图1 FCFA-Net整体结构图

Figure 1 Overall structure of FCFA-Net

2.2 主要子结构

本文提出的FCFA-Net编解码器主要通过5个部分构成,分别为SMC(Swin Meet CNN)、全尺度特征聚合器、自相关特征增强器、通道特征层级提取解码器、双项限制损失函数。SMC是基于SwinTransformer的基础上改进而来,通过更改其内部的下采样方式将卷积融合进去,进而提升性能;全尺度特征聚合器主要是为了满足在网络全深度之间传输特征信息的同时,针对其各个通道的重要性进行特征聚合;自相关特征增强器选择Transformer作为主要部分,融合其他组件,来实现特征的增强;通道特征层级提取解码器选择融入空洞卷积来实现对于多尺度以及不同感受野下,特征的高效提取同时实现

特征的上采样;双项限制损失函数则是通过分析结合图像分割任务中损失函数的两种不同的优化方向提出的。

2.2.1 SMC 在深度学习图像处理领域,无论是目标检测还是目标分割,能够有效地提取特征图是重要的一环。由于肺结节目标区域小,形状多变,提取特征是非常困难的。近些年来各种网络层出不穷,为了能够在避免卷积层过多而造成的梯度爆炸,同时提取抽象的语义信息,Resnet提出的残差机制在不断叠加网络层的同时避免问题的发生,卷积层堆叠后仍能极为有效地提取特征^[11]。Densenet则是以一种更为激进的方式添加残差结构,不仅在结构块内部同时还在各个块之间添加残差,这种方式虽然有着较高的参数量,但是也有着较为不错的性能。此时已经可以证明,残差对于现阶段神经网络来说是极为重要的一部分。Transformer以其高效的注意力机制,从自然语言处理领域进入机器视觉领域,并取得

非常好的效果^[12]。ViT^[13]、DETR^[14]和SwinTransformer^[15]都拥有着较为不错的性能,甚至在许多领域都做到了SOTA(State of the Arts)的地位。

本文借此提出SMC主干,整体结构如图2a所示,结合CNN与Transformer的优势,实现对肺结节的特征高效提取。其中密集残差块结构如图2b所示,SwinTransformer结构如图2c所示。在特征图输入最开始的部分采用密集残差块的目的是为了满足Transformer运算的条件,并且最开始进行大尺度下采样会对特征信息造成严重破坏,这对于小目标检测是极为有害的,因此在浅层采取卷积是更优的选择。对于深层信息,利用SwinTransformer在减少参数数量的情况下引入自注意力机制可以有效集中深层较为抽象的语义信息,引入相对位置编码,捕获长距离依赖。在窗口内计算自注意力的同时,滑动窗口机制也使得运算没有完全隔断各个窗口之间的信息,而是在窗口之间建立关系,保留一定的全局特征。

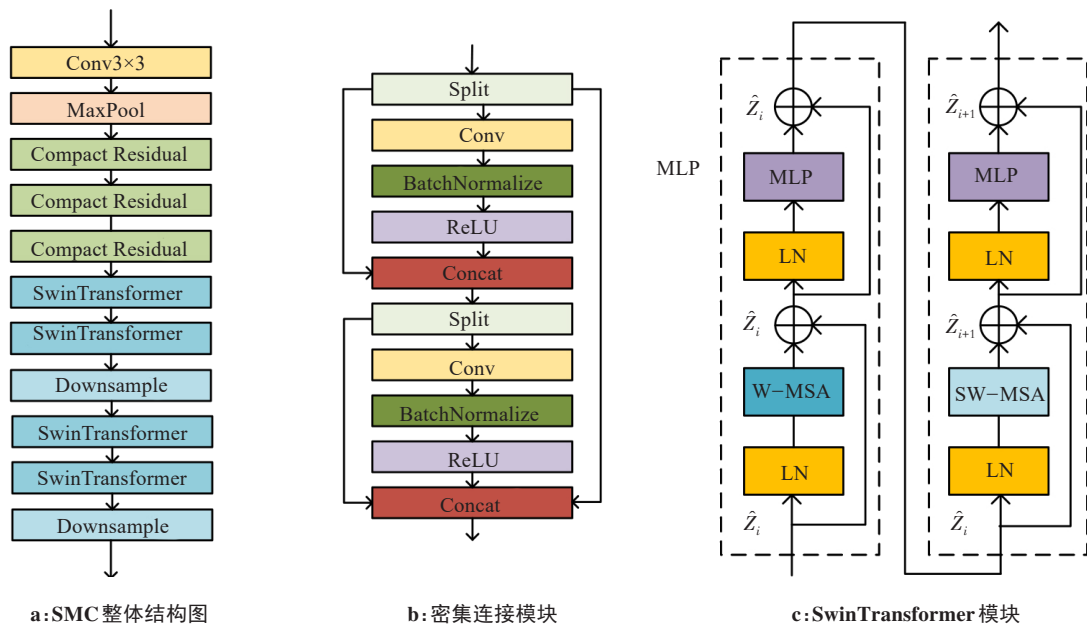


图2 SMC及其内部结构设计
Figure 2 SMC and its internal structure

2.2.2 全尺度特征聚合器 对于医学图像分割任务,提取的浅层与深层特征均有其重要性,在编码器与解码器之间构建跳跃连接,去沟通特征信息是一种必要的结构,因此本文设计一种全尺度跳跃连接方式,在结构间构成完整的特征通路。传统方式通常可以分成两种方法。一种方法是类似Unet将对应尺度的特征图直接进行连接,通过直接传输信息的方式去减轻下采样带来的信息损失^[16]。这种连接方式虽然简单有效,但是由于没有进行非线性操作,传递过

来的信息冗余较大,存在弊端。另一种方法是Unet++提出的紧密嵌套连接,由一系列层级间的短连接构成,它的优势是能够将相邻尺度的语义信息进行融合并传递给解码器,但是这种方式将信息集中在浅层解码,并不利于解码的进行^[17]。

本文提出的全尺度特征聚合器结构如图3所示,通过融合上下文的语义信息之后,将信息向更深层次传递,让浅层信息传递到解码器的入口,让解码器更好地理解抽象的语义信息。除此之外,利用线性

层进行通道判别,让其有选择地将信息传递给融合器,但是效果并不好。具体的原因可能是因为主干提取出来的语义特征往往已经进行过多次抽象,其各个通道的重要程度也早已被优化完成,添加多余的卷积结构反而会影响特征传递。

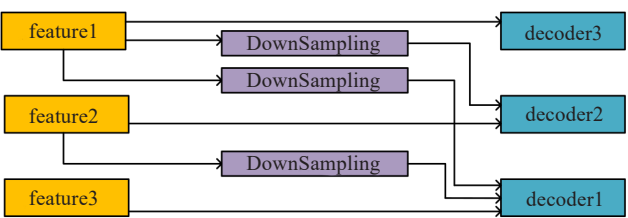


图3 全尺度特征聚合器结构图
Figure 3 Full-scale feature aggregator

2.2.3 Transformer 特征增强器 在编码器的设计上,随着主干的逐层降采样,特征信息逐渐浓缩,添加特征增强器用以获取上下文的多尺度信息,让传递给解码器的信息更为精准。对于高度浓缩过后的特征

图来说,利用卷积构建的CBMA^[18]、ASPP^[19]等方案,由于卷积存在的局部性的问题,存在着信息提取不全面的问题。由此,本文提出利用Transformer来实现这个目的,其整体及其内部Transformer块结构如图4所示。Transformer拥有着优秀的捕获长期依赖以及特征加权融合的能力以及全局信息处理能力,而且此时小尺寸多通道的特征图也可以满足计算需求。Transformer的主要结构是多头自注意力机制,首先根据输入的特征图运算得到QKV矩阵,然后将每个矩阵分割出数个头并逐个计算注意力得到注意力图,然后将注意力图与值相乘得出输出特征。计算注意力的过程是通过位置与键值矩阵计算各个位置之间的相关性,然后将相关性通过Softmax函数计算得到注意力图。它通过将特征通道划分区域,在各个区域内部计算自注意力并生成注意力图,然后将其与原图相乘算出结果。这种计算注意力的方式可以让其计算当前每个特征图内部各个样本之间的关系并赋予权值,也可通过调整区域划分的大小来调整通道之间的影响范围。

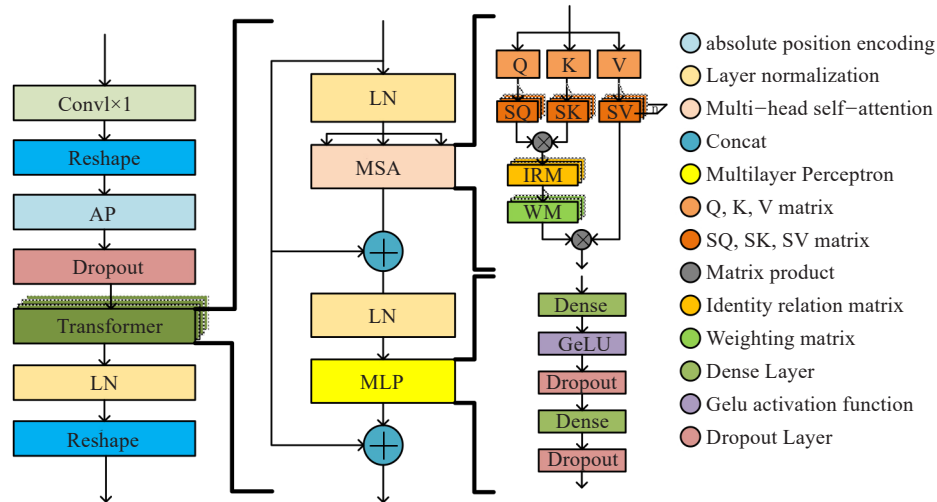


图4 Transformer 特征增强器设计
Figure 4 Transformer feature enhancer

2.2.4 多尺度通道特征解码器 本文根据传递过来的各种多尺度语义信息,设计一种通道特征解码器,其结构如图5a所示。对于每个解码块来说其结构如图5b所示。针对于每个解码器,虽然接受的信息不同,但大致上可以分为比例不同的解码器内部传递信息以及全尺度连接传递信息两部分,这些信息具有不同的抽象程度,因此需要不同的感受野去处理,然后再进行合并。空洞卷积使用扩张参数来实现对感受野的控制,用不同的感受野匹配不同通道传递过来

的语义信息。解码块之间采用短连接的方式作为信息补充让编码器传递的深层次语义信息可以向上影响浅层的解码器。分割头采用Sigmoid作为激活函数将特征图分割成前景和背景两部分,来实现最终的语义分割。
2.2.5 双项限制损失函数 在深度学习中,损失函数的设计是十分重要的环节,其与网络的收敛速度以及训练结果息息相关。在目标分割领域,与其他任务不同的是损失函数的构成更为复杂,可以将其大

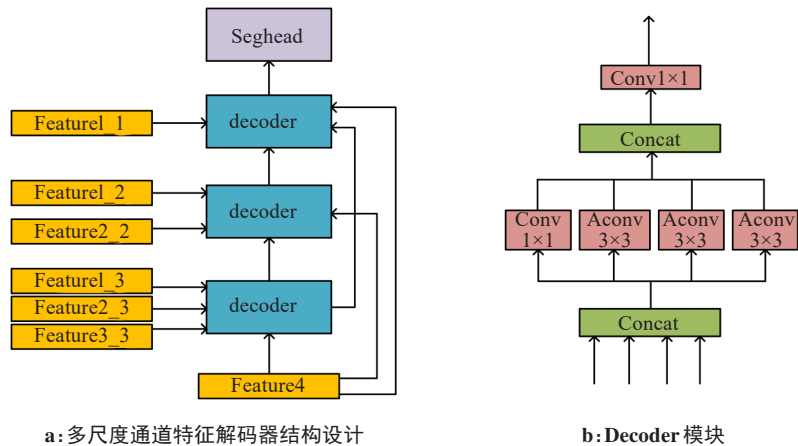


图5 多尺度通道特征解码器

Figure 5 Multi-scale channel feature decoder

致分为基于分布的损失和基于区域的损失。基于分布的损失主要的理论来源于概率分布,将样本间的相似性转换为随机事件出现的可能性,通过度量样本的真实分布与它估计的分布之间的距离,判断两者之间的相似性。基于区域的损失大多数都是在混淆矩阵的基础上发展而来,通过调整混淆矩阵中4种结果样本的比例,让网络在训练过程中不断优化。

基于分布的损失函数通常是将分割方向精确定位到需要分割的位置,并根据正负样本的分布进行优化。本文拟采用的第一项损失函数是 Focal Loss^[20],其公式如下:

$$\text{Focal_Loss}(p_i) = -(1 - p_i)^\gamma \log(p_i) \quad (1)$$

$$p_i = \begin{cases} p, & y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (2)$$

其中,对于常数 γ 是一个大于0的调节因子; p_i 是预测概率,反映模型对正确类别的预测置信度。当 p_i 越大,即模型对正确类别的预测置信度越高, $(1 - p_i)^\gamma$ 就越趋近于0,对于已经能正确分类的样本,其损失会被降低。相反,对于那些模型预测错误的样本,即 p_i 较小,那么 $(1 - p_i)^\gamma$ 就趋近于1,这样模型就会更加关注这些被错误分类的样本。在模型训练过程中,占有极大比例的样本会在训练过程中不断减少影响,而小样本逐渐占据主要位置。

基于区域的损失函数通常是根据混淆矩阵作为调整损失的偏向来实现分割任务。目前大多数都是通过 Tversky 系数异化而来,包括 Dice 系数^[21]、Jaccard 系数。Jaccard 系数也被称为 IoU 系数,它直接将位置信息作为一个整体进行训练,其计算公式如下:

$$\text{IoU} = 1 - \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (3)$$

Dice 系数计算公式如下:

$$\text{DiceLoss} = 1 - \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (4)$$

其中,TP 表示预测为正类的正样本;TN 表示预测为负类的负样本;FP 预测为正类的负样本;FN 预测为负类的正样本。Dice 系数通过计算精确率与召回率的调和平均数,提高召回率在损失函数中的占比,有助于减少假阴性样本。

3 实验数据与评价指标

3.1 数据集来源及其预处理

Luna16 数据集^[22]作为最大公用肺结节 LIDC-IDRI 数据集^[23]的子集,删除其中肺结节切片厚度大于 3 mm 以及肺结节小于 3 mm 的 CT 影像。该数据集中共有 1 018 例研究病例,收录 4 位医师共同标注的 1 186 个肺结节金标准信息。在所有的病例中,提取全部共 8 121 张肺结节切片进行人工筛选。

肺结节影像的预处理以及筛选步骤如下:(1)根据数据集肺结节位置文件,在对应的 mhd 文件中提取出所有的肺结节文件,初筛标准为是否包含 z 轴坐标以及是否含有 z 轴重叠部分;(2)再次通过位置文件,在医生标注的金标准文件中提取边缘坐标,利用其生成对应肺结节文件的标注文件;(3)人工筛选删去其中完全不含有肺结节以及标注与肺结节位置出入较大的文件。筛选结果为 3 528 张含有标准信息的肺结节以及其对应标注信息的图片。

3.2 实验环境与实验设置

本实验选用 Window11 操作系统作为基础平台,CPU 为 Intern®Xeon®CPUE5-2630v4, GPU 为 NVIDIA-GeForce-RTX2080Ti,显存 11 GB,CUDA 版本为 10.1。软件环境为 Python3.8、Tensorflow2.4。本文按照 8:2 的比例分割实验数据集,共 2 820 个训练样本,708 个测试样本,实验采用相同的测试集以及验证集。在训练阶

段,epoch 均为 200 轮次,每轮次 Batch size 为 2。优化器选择 Adam,初始学习率为 1e-4。

3.3 评估指标

本文评估指标为精准率(Precision, PRE)、召回率(Recall, REC)、均交并比(mean Intersection Over Union, mIoU)^[14]、F1 分数,公式如下所示:

$$PRE = \frac{TP}{TP + FP}$$
(5)

$$REC = \frac{TP}{TP + FN}$$
(6)

$$mIoU = \frac{1}{K + 1} \sum_{i=0}^K \frac{TP}{TP + FP + FN}$$
(7)

$$F1 = \frac{2TP}{2TP + FN + FP}$$
(8)

4 结果与分析

4.1 主干网络对比

本文选取作为对比的经典主干网络分别为强调使用残差结构的 Resnet、Densenet;利用小核卷积替换大核卷积的 VGG 网络;使用全 Transformer 结构的 SwinTransformer 网络。这些网络对主干网络的发展有较大的影响,且在相关领域做到过 SOTA 的地位,对比结果如表 1 所示。从实验数据中可以看出传统的深度卷积网络性能随着深度的改变或者残差密集程度提升而数值有所提升,Transformer 主干由于其较大的计算量所采用的高倍下采样对于较小的肺结节样本会产生丢失。因此结合 Transformer 以及 CNN 的 SMC 在一定程度上对于小目标检测更有优势。

表 1 不同主干网络之间结果比较(%)
Table 1 Comparison among different backbones (%)

方法	精准率	召回率	均交并比	F1 分数
Resnet	95.87	81.58	79.87	88.15
Densenet	94.98	82.87	80.54	88.52
VGG	92.77	85.60	81.77	89.04
Swin	89.76	61.63	60.94	73.09
SMC	95.22	86.56	82.21	90.53

4.2 特征增强对比

分别选取不添加任何增强方式、隐藏层、CBAM 3 种方式作为对比,如表 2 所示。Transformer 拥有着较为优秀的性能,虽然 CBAM 也能在空间上进行优化,但是感受野较小,效果略差。由于 Transformer 在全局下的特征提取机制,虽然此时的特征已经经过多层稀释,其仍然能够顺利提取到各个样本之间的信息并进行特征增强,但是其有一定的不稳定性。

表 2 不同特征增强结构结果比较(%)
Table 2 Comparison among different feature enhancers (%)

方法	精准率	召回率	均交并比	F1 分数
None	85.67	81.57	75.87	83.57
CNN	87.70	83.02	77.23	85.30
CBAM	88.64	84.23	78.09	86.38
Tranformer	96.24	82.39	80.59	89.89

4.3 跳跃连接对比

选取前文中提到的长连接与紧密嵌套连接两种方式以及添加额外通道注意力的全尺度跳跃连接作为对比方法,对比结果如表 3 所示。随着浅层的解码器能够获取更多深层的信息,网络性能也会随之有一定的提升,因此全尺度跳跃连接为更好的选择。

表 3 不同跳跃连接结果比较(%)
Table 3 Comparison among different skip connections (%)

方法	精准率	召回率	均交并比	F1 分数
长连接	95.87	81.58	79.87	88.15
紧密嵌套	94.46	82.30	79.88	87.96
通道尺度	95.95	82.01	80.28	88.43
全尺度	96.23	84.32	83.19	90.52

4.4 损失函数对比

针对于损失函数的比较,本文分别选取 CE_Dice_loss、Log_Dice_loss、Focal_IoU_loss、Focal_Fβ_loss、Focal_Dice_loss 5 种由分布以及区域相结合的损失函数作为对比,如表 4 所示。由于肺结节分割属于一种正负样本不均匀的任务,通过结果可以看出,3 种分布损失函数分别与 Dice_loss 相结合的结果中,Focal_loss 表现最好,但是训练过程中发现其收敛速度要低于 CE_loss 以及 Log_loss。将 Focal_loss 分别与 3 种区域损失函数相结合,可以发现 Dice_loss 的表现要优于 IoU_loss 以及 Fβ_loss。

表 4 不同损失函数结果比较(%)
Table 4 Comparison among different loss functions (%)

损失	精准率	召回率	均交并比	F1 分数
CE_Dice	95.87	81.58	79.87	88.15
Log_Dice	94.69	84.58	81.85	89.35
Focal_IoU	92.21	86.38	82.07	89.20
Focal_Fβ	93.00	85.18	81.55	88.92
Focal_Dice	93.33	87.24	83.41	90.48

4.5 网络对比

选取经典的医学图像分割网络 UNet 作为基准网络,同时选取近年来表现较优的网络 Unet++ 以及 TransUnet 作为对比网络,如表 5 所示。本文方法相较于 UNet、Unet++ 和 TransUnet 方法精准率分别提升 9.66%、7.84%、3.75%,召回率提升 5.50%、2.96%、1.37%,均交并比提升 11.35%、7.16%、4.18%,F1 分数提升 8.07%、5.87%、3.10%。实际的分割结果添加了与金标准的对比,结果如图 6 所示,最差的 UNet 边缘分割很差,甚至出现空洞,Unet++ 的分割结果毛刺较多,导致结果较差,TransUnet 分割表现不错,但是边缘较为锋利,与标签边缘贴合程度较差,本文方法无论是在边缘表现还是在位置准确性上均较为准确。

5 结 论

肺结节分割对于辅助临床肺结节诊断中拥有着较高的研究价值,本文设计的 FCFA-Net 网络可以用以精准分割肺部 CT 图像中的肺结节病灶,在 Luna16

表 5 本文方法与其他方法对比(%)
Table 5 Comparison of the proposed method with other methods (%)

方法	精准率	召回率	均交并比	F1 分数
UNet	87.70	83.02	77.23	85.29
Unet++	89.52	85.56	81.42	87.49
TransUnet	93.61	87.15	84.40	90.26
FCFA-Net	97.36	88.52	88.58	93.36

数据集中进行训练的结果表明,其整体上拥有着较为优秀的表现,同时本文的各项实验证明,本文在网络各部分结构上的设计均有效,表明本文提出的方法提高了现有计算机辅助诊断的自动化水平。但是由于肺结节的种类大小以及阶段不同,形状表现以及判断标准的复杂多变,因此在这个领域上还存在着更多进步的可能,计算机辅助系统也会在肺结节疾病检测中发挥更为重要的作用。

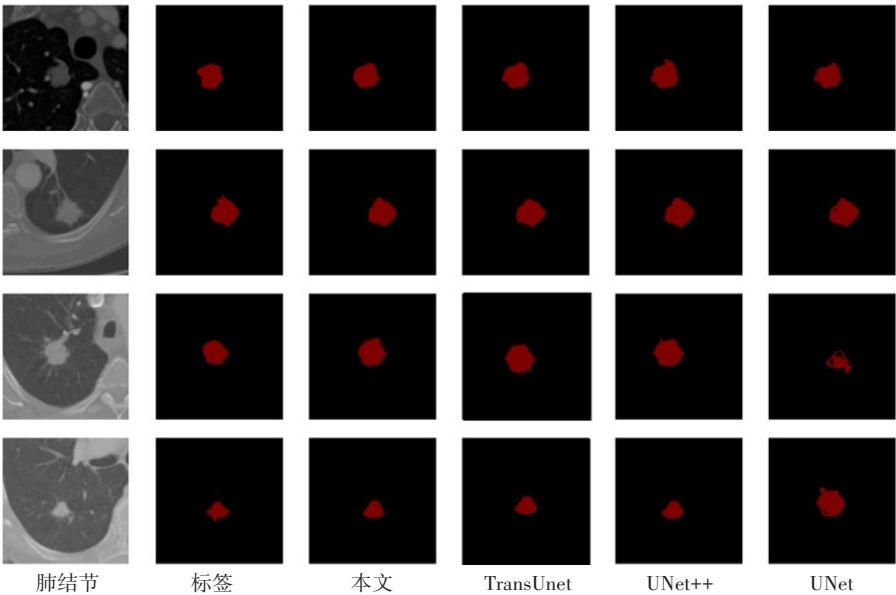


图 6 分割结果对比
Figure 6 Comparison of segmentation results

【参考文献】

[1] 董婷,魏琬,聂生东. CT 影像肺结节分割研究进展[J]. 中国图象图形学报, 2021, 26(4): 751-765.
Dong T, Wei L, Nie SD. Research progress of lung nodule segmentation based on CT images [J]. Journal of Image and Graphics, 2021, 26(4): 751-765.

[2] Savic M, Ma YH, Ramponi G, et al. Lung nodule segmentation with a region-based fast marching method[J]. Sensors (Basel), 2021, 21(5): 1908.

[3] Wang D, He K, Wang B, et al. Solitary pulmonary nodule segmentation based on pyramid and improved grab cut [J]. Comput Methods Programs Biomed, 2021, 199: 105910.

[4] Li DJ, Yuan SL, Yao G. Pulmonary nodule segmentation based on REMU-net[J]. Phys Eng Sci Med, 2022, 45(3): 995-1004.

[5] Wang ZR, Men JR, Zhang FC. Improved V-net lung nodule segmentation method based on selective kernel [J]. Signal Image Video Process, 2023, 17(5): 1763-1774.

[6] Wu ZZ, Li XS, Zuo JH. RAD-UNet: research on an improved lung nodule semantic segmentation algorithm based on deep learning[J]. Front Oncol, 2023, 13: 1084096.

[7] Wang X, Li MQ, Yang J. Segmentation of pulmonary nodules based on BBClstm Unet[J]. J Phys Conf Ser, 2021, 1966(1): 012037.

[8] Zhang HY, Ni WH, Luo Y, et al. TUnet-LBF: retinal fundus image fine segmentation model based on transformer Unet network and LBF[J]. Comput Biol Med, 2023, 159: 106937.

- [9] Chen JN, Lu YY, Yu QH, et al. TransUNet: transformers Make Strong encoders for medical image segmentation[J]. Comput Sci, 2021, 2102: 04306.
- [10] Wang YF, Zhou C, Chan HP, et al. Hybrid U-net-based deep learning model for volume segmentation of lung nodules in CT images[J]. Med Phys, 2022, 49(11): 7287-7302.
- [11] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017: 6000-6010.
- [13] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. (2021-06-03)[2023-09-20]. <https://arxiv.org/abs/2010.11929>.
- [14] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers [C]//Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 213-229.
- [15] Liu Z, Lin YT, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 9992-10002.
- [16] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [17] Zhou ZW, Rahman Siddiquee MM, Tajbakhsh N, et al. UNet++: a nested U-net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer International Publishing, 2018: 3-11.
- [18] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [19] Chen LC, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs [EB/OL]. (2016-06-07). <https://arxiv.org/abs/1412.7062>.
- [20] Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection [J]. IEEE Trans Pattern Anal Mach Intell, 2020, 42(2): 318-327.
- [21] Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C]//Proceedings of the 4th International Conference on 3D Vision. Stanford, CA, USA: IEEE, 2016.
- [22] Setio AA, 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]. Med Image Anal, 2017, 42: 1-13.
- [23] Armato SG, 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]. Med Phys, 2011, 38(2): 915-931.

(编辑:陈丽霞)