

基于多尺度级联注意网络的肺实质分割^{*}

许圳兴¹ 余耀² 赵东² 陈园³ 范圣旺³

(1. 南京信息工程大学电子与信息工程学院 南京 210044; 2. 无锡学院电子与信息工程学院 无锡 214105;
3. 杭州海康威视数字技术股份有限公司 上海 200023)

摘要:针对肺实质分割任务中不同尺度特征的全局上下文信息利用率低、分割精度低、分割细节模糊等问题,提出一种多尺度级联注意网络(multiscale cascaded attention networks, MCANet)。该网络主要由多尺度特征提取网络(multi-scale feature extraction network, MSFENet)、多尺度注意力引导模块(multi-scale attention guidance module, MSAG)、解码特征整合器(decoding feature integrator, DFI)组成。首先,设计 MSFENet 以提高特征信息在不同通道维度上的空间交互能力,在采样过程中最大限度地保留图像的关键特征,丰富全局上下文信息。然后,设计 MSAG 提高模型在解码过程中对多尺度特征信息的利用率,并最大限度地融合两种注意力机制的优势。最后设计 DFI,重新整合解码器生成的解码特征,以提高模型对边缘信息的分割性能。在 LUNA16 数据集上对模型性能进行实验验证,得到了 0.993 的 Dice 和 3.864 的 HD,实验结果证明了 MCANet 与其他主流医学分割模型相比有更优异的分割性能,能更准确地分割肺实质。

关键词:肺实质分割;多尺度级联注意网络;多尺度特征提取网络;多尺度注意力引导模块;解码特征整合器

中图分类号: TP391.41 **文献标识码:** A **国家标准学科分类代码:** 510.40

Lung parenchymal segmentation based on multiscale cascaded attention network

Xu Zhenxing¹ Yu Yao² Zhao Dong² Chen Yuan³ Fan Shengwang³

(1. School of Electronic and Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. School of Electronic and Information Engineering, Wuxi University, Wuxi 214105, China;
3. Hangzhou Hikvision Digital Technology Co., Ltd., Shanghai 200023, China)

Abstract: A multiscale cascaded attention network (MCANet) is proposed to solve the problems of low utilization rate of global context information, low segmentation accuracy and fuzzy segmentation details of different scale features in lung parenchymal segmentation task. The network is mainly composed of a multi-scale feature extraction network (MSFENet), a multi-scale attention guidance (MSAG) module, and a decoding feature integrator (DFI). Firstly, the MSFENet network is designed to improve the spatial interaction ability of feature information in different channel dimensions, retain the key features of the image to the greatest extent in the sampling process, and enrich the global context information. Then, MSAG is designed to improve the utilization of multi-scale feature information in the decoding process of the model, and maximize the integration of the advantages of the two attention mechanisms. Finally, DFI was designed to reintegrate the decoding features generated by the decoder to improve the segmentation performance of the model for edge information. In this paper, the performance of the model was verified by experiments on the LUNA16 dataset, and the Dice of 0.993 and HD of 3.864 were obtained, and the experimental results proved that MCANet has better segmentation performance and can segment the lung parenchyma more accurately than other mainstream medical segmentation models.

Keywords: lung parenchymal segmentation; MCANet; MSFENet; MSAG; DFI

收稿日期:2023-11-15

^{*} 基金项目:江苏省高等学校基础科学(自然科学)研究面上省资助经费项目(22KJB140017)、无锡学院人才启动基金(2019r015,2022r006)项目资助

0 引言

肺癌作为目前死亡率最高的癌症之一,其早期症状以肺结节形式表现出来^[1]。传统的计算机辅助检测技术(computer aided detection, CAD)通过检测肺结节来发现并治疗早期肺癌^[2]。肺结节检测主要包括图像预处理、肺实质分割、肺结节检测和假阳性剔除4个部分^[3],其中,肺实质分割作为肺结节检测的基础,它影响着检测结果的准确性。

肺实质分割主要是剔除肺部的胸壁、心包等不相干的区域,保留肺实质部分^[4]。目前,肺实质分割方法主要分为两大类,基于传统机器学习的肺实质分割和基于深度学习的肺实质分割^[5]。

基于传统机器学习的肺实质分割主要包括阈值分割法^[6]、区域生长分割法^[7]、基于活动轮廓的分割方法等。阈值分割法简单高效,计算速度快,但对噪声敏感,且局限于明显灰度差异,难以处理复杂场景。区域生长分割法适应性强,对噪声有一定抵抗力,适用于不规则目标,但计算开销大,对种子点选择敏感。基于活动轮廓的分割方法适应不规则轮廓,对初始化要求低,但计算复杂度高,对噪声敏感。因此,基于传统机器学习方法的肺实质分割难以对复杂度高、拥有大量病理特征的肺部图像进行精准分割。基于此,医学图像处理领域逐渐出现基于深度学习的肺实质分割方法。

基于深度学习的肺实质分割方法采用卷积神经网络(convolutional neural network, CNN)^[8]的结构,主要包括全卷积网络(fully convolutional network, FCN)^[9]和对称卷积网络(U-Net)^[10]。FCN改进了传统的CNN结构,将全连接层替换为卷积层,通过上采样操作实现对肺实质的准确分割。U-Net在FCN基础上引入了两次卷积操作,通过拼接融合实现下采样和上采样阶段的特征传递,以输出准确的肺实质掩膜。随着U-Net在医学影像分割领域的发展,加速了如SegNet^[11]等更多经典分割网络的出现。

随着肺实质分割任务在医学影像分割领域的不断发展,国内外研究者开始积极探索各种肺实质分割模型。Xie等^[12]利用双流金字塔模块捕获不同层的互补特征,强化模型对局部细节特征的学习能力,并使用编码器-解码器结构来抑制复杂的背景,最后在3个不同医学分割数据集上取得不错的结果。Chen等^[13]使用剩余空间注意和门控注意两种机制,对肺实质的空间信息进行加权,并自适应校准权重,预测潜在的关键特征,并使用轻量级密集块和定位转置块代替传统上下采样来恢复图像分辨率,最后在LIDC-IDRI和CO-Seg数据集上获得了良好的分割效果。刘莹芳等^[14]使用一种改进的最大类间方差法(OS-TU)降低肺组织对分割性能的干扰,并结合形态学运算对分割到的肺实质进行弥合修补,虽然最后在LIDC数据集上取得0.9842的Dice系数,但是对模型的鲁棒性有着严格限制,不同CT图像的分割效果存在差异。毕蓉蓉等^[15]

虽然使用带有残差单元的循环卷积层和特征映射的方式来降低网络参数,并使用特征累计的方式消除特征求和对网络性能的影响,让模型侧重于细小特征的提取,但是,模型的性能依赖于庞大的样本数量,并且没有有效利用编码与解码单元之间的特征信息,忽略了编码和解码单元之间的信息交互。侯英竹^[16]在U-Net的收缩路径和扩张路径利用空间注意力模块和通道注意力模块来改善网络对复杂环境的感知能力,降低环境对肺结节分割的干扰,但忽视了模型在采样过程中细节丢失的问题和特征在空间维度与通道维度之间的信息交互。张宇杰等^[17]在下采样编码部分使用残差连接降低网络参数,加速网络收敛,同时利用深度可分离卷积搭建Dep模块,以较少的运算量提取多维的信息,获取模型编码部分的尺度信息,模型更多关注浅层特征,因此,该模型可以很好的对小目标进行分割,在肺结节尺度上,在LUNA16数据集上实验得到了0.9672的Dice系数,但由于模型对深层特征的利用率低,因此,在肺实质尺度上仅实验得到了0.8333的Dice系数。

基于上述关联研究,为了同时兼顾高准确率和方便后续开展肺结节检测工作,本文依次提出多尺度特征提取网络(multi-scale feature extraction network, MSFENet)、多尺度注意力引导模块(multi-scale attention guidance, MSAG)和解码特征整合器(decoding feature integrator, DFI)结构来构建多尺度级联注意网络(multiscale cascaded attention networks, MCANet),并在编码器-解码器阶段以U-Net作基线,不仅提高模型在空间维度和通道维度上对不同尺度特征的全局上下文信息的利用率,还充分融合深层和浅层的特征信息,促进模型关注关键特征,避免因细节模糊导致分割不准确的问题,强化了对边界的分割效果,提高模型对肺实质分割的准确率。

1 网络结构

1.1 网络框架

本文提出的MCANet由多尺度特征提取网络、编码器-解码器和解码特征整合器3部分组成,如图1所示。为了解决模型在空间维度和通道维度上对不同尺度特征的全局上下文信息利用不足,导致分割精度低、分割细节模糊的问题,本文提出了如下改进策略。

1)MSFENet主要由多通道增强模块(multi-channel enhancement, MCE)和多级特征融合模块(multi-level feature fusion module, MLFF)组成。在模型进入编码阶段之前,MSFENet通过使用MCE和MLFF增强特征信息在不同通道维度上的空间交互能力,避免在采样过程中丢失细节信息,以获取不同尺度上具有代表性的全局上下文特征,提高编码阶段输入特征信息的丰富性。

2)为了提高模型在解码过程中对多尺度特征信息的利用率,增强模型在空间维度和通道维度上的信息交互能力,本文在编码-解码阶段引入MSAG模块。该模块首先通过多尺度融合模块(multi-scale fusion module, MSF)

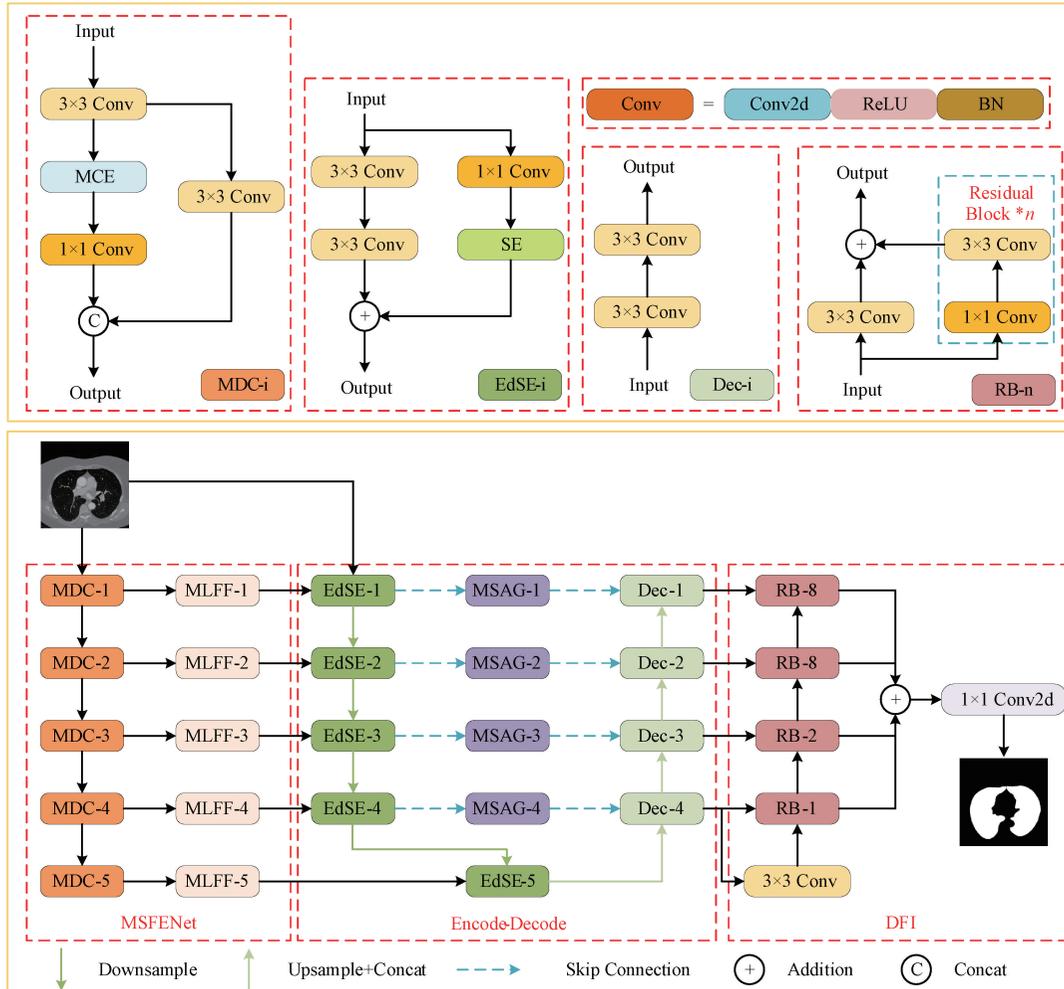


图1 多尺度级联注意力网络

Fig. 1 Multiscale cascaded attention networks

对不同尺度的特征信息进行信息整合,然后以串并行的方式同时使用空间和通道注意力模块,自适应的整合不同维度特征和全局特征的相关性,抑制冗余信息对模型解码过程的干扰,促进模型关注关键特征,提高分割性能。

3)为避免解码器在提高分辨率时造成关键特征模糊而丢失分割目标的边缘信息,并提高浅层网络的细节信息的利用率,本文利用 Darknet53^[18]基础结构构成解码特征整合器 DFI,对解码器生成的解码特征重新整合,并在最后以 1×1 卷积生成与原图像分辨率相同的肺实质掩模。

4)编码器-解码器阶段以 U-Net 为基线,为提高模型上下级网络的适应能力,在编码阶段引入通道注意力模块 SENet^[19],利用残差结构构建 EdSE 模块,提高编码器输出特征与上下级网络输入和输出特征的相关性。

1.2 多尺度特征提取网络

在肺实质分割任务中,如果只是简单地对二维肺部图像进行上下采样或者其他多尺度操作,不仅会丢失图像中

大量的关键特征,还会提高模型对特征的提取难度,降低模型分割性能。为了解决上述问题,本文设计了一条自上而下的主干网络 MSFENet,该网络将提取到的关键特征以不同尺度形式输入到模型编码阶段,丰富模型在编码阶段生成的特征信息。

MSFENet 提取多尺度关键特征的过程如图 1 所示,MSFENet 主要进行下采样提取关键特征,由基础卷积块、MCE 和 MLFF 组成。对于 MCE,由于本文选用的 LUNA16 数据集,经过预处理后生成的二维肺部图片大多呈单色调灰色,因此难以捕捉图像在深度层面的特征信息。基于此,本文设计了一种多通道增强模块 MCE,加强对通道维度上特征信息的捕捉能力。如图 2 所示,MCE 通过对不同通道上的特征信息进行卷积、拼接和点积等操作来增强特征图不同通道之间的信息联系,提高特征图之间的空间交互能力。第 k 层 MDC 结构中 MCE 模块输出可以表示为:

$$F_{e/2}^k, F_{e/2}^k, F_e^k = Split(w_{3 \times 3} \cdot F_{3 \times 3}^k + b_{3 \times 3}) \quad (1)$$

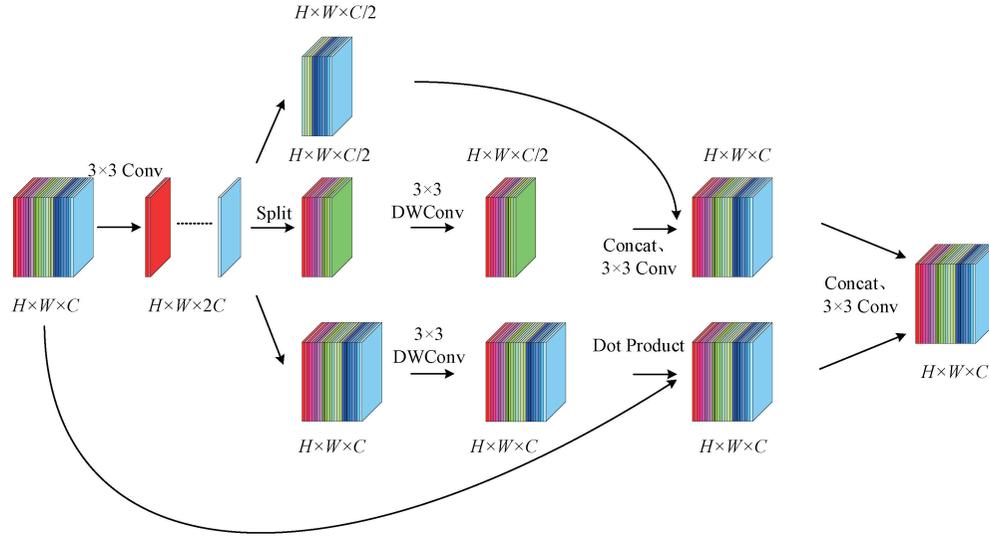


图2 多通道增强模块

Fig. 2 Multi-channel enhancement module

$$\mathbf{F}_{c/2}^{k''} = \mathbf{w}_{3 \times 3} \cdot \mathbf{F}_{c/2}^k + \mathbf{b}_{3 \times 3}, \mathbf{F}_c^{k'} = \mathbf{w}_{3 \times 3} \cdot \mathbf{F}_c^k + \mathbf{b}_{3 \times 3} \quad (2)$$

$$\mathbf{F}_c^{k''} = \mathbf{w}_{3 \times 3} \cdot (\{\mathbf{F}_{c/2}^{k''}, \mathbf{F}_c^{k'}\}) + \mathbf{b}_{3 \times 3} \quad (3)$$

$$\mathbf{F}_c^{k'''} = \mathbf{F}_{3 \times 3}^k \odot \mathbf{F}_c^{k''} \quad (4)$$

$$\mathbf{F}_{MCE}^k = \mathbf{w}_{3 \times 3} \cdot (\{\mathbf{F}_c^{k''}, \mathbf{F}_c^{k'''}\}) + \mathbf{b}_{3 \times 3} \quad (5)$$

式中: $k=1, 2, 3, 4, 5$; w 和 b 表示卷积层的权值和偏置; $\mathbf{F}_{c/2}^k, \mathbf{F}_c^k, \mathbf{F}_c^{k'}$ 表示经过 Split 切片得到的 3 个不同通道数的特征; $\mathbf{F}_{c/2}^{k''}, \mathbf{F}_c^{k'}$ 表示将 $c/2$ 和 c 通道数的特征分别通过 3×3 深度可分离卷积提取到的特征; $\mathbf{F}_c^{k''}$ 表示将 $\mathbf{F}_{c/2}^{k''}, \mathbf{F}_c^{k'}$ 按通道维度拼接后经过 3×3 卷积提取到的特征; $\mathbf{F}_c^{k'''}$ 表示将 $\mathbf{F}_{3 \times 3}^k, \mathbf{F}_c^{k''}$ 进行点积运算得到的特征; $\mathbf{F}_{3 \times 3}^k$ 表示第 k 层的 MDC 结构中第 1 个 3×3 卷积提取到的特征; $\{\}$ 表示拼接操作; \cdot 表示卷积运算; \odot 表示点积运算; \mathbf{F}_{MCE}^k 表示将 $\mathbf{F}_c^{k''}, \mathbf{F}_c^{k'''}$ 按通道维度拼接后经过 3×3 卷积提取到的最优特征。

另外,为了避免在下采样过程中丢失特征图中大量的位置信息,因此本文使用基础卷积块、MCE 和残差结构构成 MDC 模块,加强主干网络在不同尺度上对关键特征信息的捕捉能力,弥补下采样丢失的特征信息(图 1)。第 k 层 MDC 模块输出表示如下:

$$\mathbf{F}_{3 \times 3}^k = \mathbf{w}_{3 \times 3} \cdot \mathbf{F}_{MDC}^{k-1} + \mathbf{b}_{3 \times 3} \quad (6)$$

$$\mathbf{F}_{1 \times 1}^k = \mathbf{w}_{1 \times 1} \cdot \mathbf{F}_{MCE}^k + \mathbf{b}_{1 \times 1} \quad (7)$$

$$\mathbf{F}_{Res}^k = \mathbf{w}_{3 \times 3} \cdot \mathbf{F}_{3 \times 3}^k + \mathbf{b}_{3 \times 3} \quad (8)$$

$$\mathbf{F}_{MDC}^k = \{\mathbf{F}_{1 \times 1}^k, \mathbf{F}_{Res}^k\} \quad (9)$$

式中: $k=1, 2, 3, 4, 5$; $\mathbf{F}_{1 \times 1}^k$ 表示将第 k 层 MDC 结构中 MCE 的输出经过 1×1 卷积提取到的特征; \mathbf{F}_{Res}^k 表示在残差结构中将 $\mathbf{F}_{3 \times 3}^k$ 经过 3×3 卷积提取到的特征; \mathbf{F}_{MDC}^k 表示将 $\mathbf{F}_{Res}^k, \mathbf{F}_{1 \times 1}^k$ 在通道维度进行拼接得到的特征; \mathbf{F}_{MDC}^0 表示原始输入图像。

为了使分割网络更多地关注特征图的肺实质部分,加

强网络对肺实质的分割能力,同时避免因过多上下采样而丢失特征信息的问题,本文在 MSFENet 中设计了一个多级特征融合模块 MLFF,用于将 MDC 提取到的上下文信息更易于融入到编码器中,在避免训练过程中丢失特征信息的同时丰富特征信息量,如图 3 所示。该模块分别使用 $1 \times 1, 3 \times 3$ 和 5×5 三个大小不同的空洞卷积,通过使用放大因子的步长对像素进行加权,可以保证在不改变分辨率的情况下增加感受野,有效捕捉不同尺度的上下文信息,并与输入图像拼接后进入一个 1×1 的降维卷积,提高模型的分割性能。第 k 层 MLFF 模块输出表示如下:

$$\mathbf{f}_{1 \times 1}^k = \sigma(\mathbf{w}_{1 \times 1} \cdot \mathbf{F}_{MDC}^k + \mathbf{b}_{1 \times 1}) \quad (10)$$

$$\mathbf{f}_{3 \times 3}^k = \sigma(\mathbf{w}_{3 \times 3} \cdot \mathbf{F}_{MDC}^k + \mathbf{b}_{3 \times 3}) \quad (11)$$

$$\mathbf{f}_{5 \times 5}^k = \sigma(\mathbf{w}_{5 \times 5} \cdot \mathbf{F}_{MDC}^k + \mathbf{b}_{5 \times 5}) \quad (12)$$

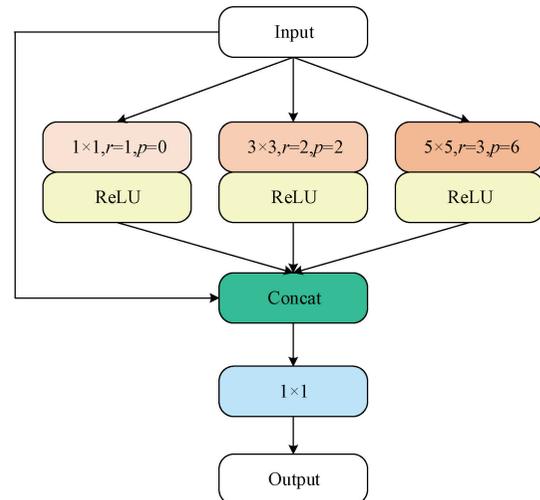


图3 多级特征融合模块

Fig. 3 Multi-level feature fusion module

$F_{MLFF}^k = w_{1 \times 1} \cdot (\{f_{1 \times 1}^k, f_{3 \times 3}^k, f_{5 \times 5}^k, F_{MDC}^k\}) + b_{1 \times 1}$ (13)
式中: $k=1,2,3,4,5$; σ 表示 ReLU 激活函数; $f_{1 \times 1}^k, f_{3 \times 3}^k, f_{5 \times 5}^k$ 表示将 F_{MDC}^k 分别经过 $1 \times 1, 3 \times 3, 5 \times 5$ 的空洞卷积提取到的特征; F_{MLFF}^k 表示将 $f_{1 \times 1}^k, f_{3 \times 3}^k, f_{5 \times 5}^k, F_{MDC}^k$ 在通道维度上进行拼接并经过一个 1×1 的卷积降维后得到的特征。

1.3 多尺度注意力引导模块

为改善 U-Net 网络对多尺度信息利用不充分以及多次下采样导致的语义信息丢失和失真的问题,本文设计了

MSAG 模块,如图 4 所示。MSAG 首先将每一层 EdSE 生成的 3 个不同尺度的特征信息进行融合,然后将融合后的特征信息分别输入进通道注意力(efficient channel attention, ECA)^[20]和空间注意力(spatial attention module, SAM)^[21]中以捕获关键区域特征,并在最后自适应融合提取到的关键特征。因此,MSAG 可以更有效地关注关键区域的特征,过滤掉特征图中不相关的信息,最终生成更准确的肺实质分割图片。

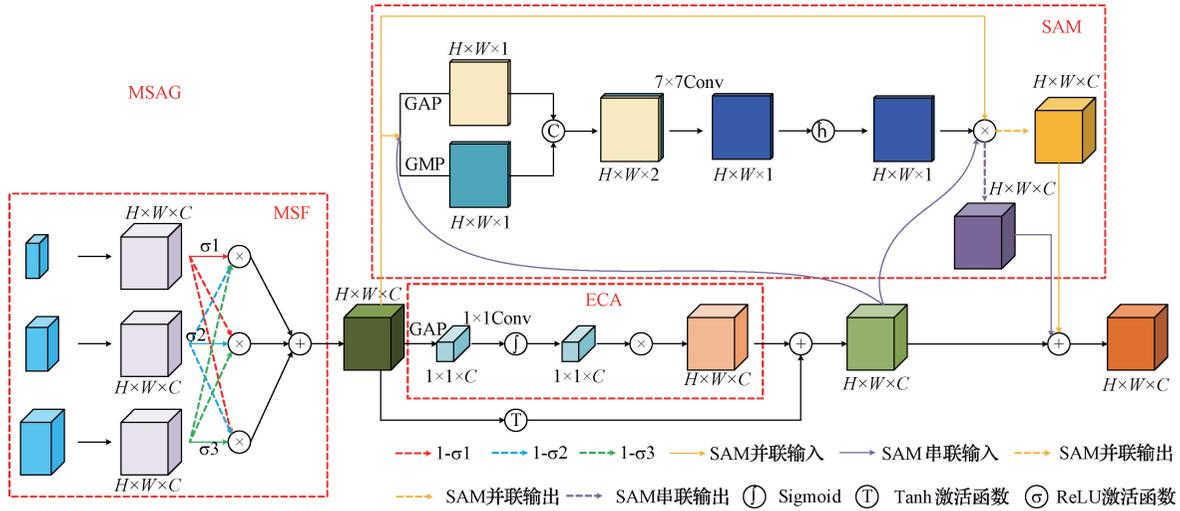


图 4 多尺度注意力引导模块

Fig. 4 Multi-scale attention guidance module

本文设计的 MCE 和 MSAG 两个模块都注重在通道维度和空间维度上进行信息交互,特别是在通道维度上都提高了模型的信息交互和提取能力。因此,在编码阶段,为了自适应的融合来自 MSFENet 的输入特征并生成最利于 MSAG 的输入特征,进一步加强编码阶段在通道维度上对关键特征的捕捉能力,本文在原 U-Net 编码器结构中引入通道注意力模块 SENet,构成 EdSE 模块(图 1)。利用残差结构与每一层最后的输出特征进行拼接,生成富含通道信息的关键特征,提高模型性能。第 i 层 EdSE 模块输出表示如下:

$$E_n^i = \begin{cases} w_{3 \times 3} \cdot (E_3^0) + b_{1 \times 1}, & n = 1, i = 1 \\ w_{3 \times 3} \cdot (Down(E_3^{i-1})) + b_{1 \times 1}, & n = 1, i \neq 1 \\ w_{3 \times 3} \cdot E_1^i + b_{3 \times 3}, & n = 2 \\ F_{MLFF}^i + S^i + E_2^i, & n = 3 \end{cases} \quad (14)$$

$$S^i = \begin{cases} SE(w_{1 \times 1} \cdot E_3^0 + b_{1 \times 1}), & i = 1 \\ SE(w_{1 \times 1} \cdot (Down(E_3^{i-1})) + b_{1 \times 1}), & i \neq 1 \end{cases} \quad (15)$$

式中: $i=1,2,3,4,5$; $n=1,2,3$; E_n^i 表示第 i 层 EdSE 生成的第 n 个特征; S^i 表示将 E_3^{i-1} 经过下采样、 1×1 卷积和 SE 注意力机制后提取到的特征; $Down$ 表示下采样操作; E_3^0 表示原始输入图像。

对于不同尺度的特征信息,本文设计了 MSF 模块。为避免使用过多的上下采样而丢失特征信息以及不同输出层语义信息的差异,MSF 首先通过 3 个 1×1 的卷积将 3 个不同尺度的特征信息缩放到同一尺度,在上下采样过程中更好地将对应尺度特征图中的关键信息保留下来,实现跨通道信息交互并增加特征信息的非线性特性;然后使用 ReLU 函数作为激活函数,对缩放后的特征实现非线性映射,并在最后利用乘法器和加法器将非线性映射的特征自适应融合成一个包含多尺度特征信息的关键特征。第 i 层 MSAG 模块中 MSF 的输出表示如下:

$$m_n^i = w_{1 \times 1} \cdot E_n^i + b_{1 \times 1} \quad (16)$$

$$M_n^i = \left(\sum_{n=1}^3 \sigma_n - 2 \sum_{n=1}^3 \sum_{m=1, m \neq n}^3 \sigma_n \sigma_m + 3 \prod_{n=1}^3 \sigma_n \right) \prod_{n=1}^3 m_n^i \quad (17)$$

式中: $i=1,2,3,4$; $n, m=1,2,3$; m_n^i 表示将 3 个不同尺度大小的 E_n^i 特征经过 1×1 卷积生成 3 个同尺度的特征; M_n^i 表示将 3 个 m_n^i 特征经过累加累乘生成的特征; \sum 表示累加符号; \prod 表示累乘符号。

本文通过串并行使用 ECA 和 SAM,引导各注意力关注关键区域特征,最后将捕获到的关键特征与全局特征自适应结合在一起,提高模型分割性能。其中,考虑到 ECA 和 SAM 的串并联方式对模型分割性能的影响,如果只使

用串联的方式,虽然可以更合理的分配各注意力的权重,但是会破坏特征信息在在空间维度和通道维度上的独立性,难以发挥 ECA 的通道交互能力,丢失通道维度上的空间信息;而如果只是用并联的方式,虽然更好地利用了特征图的空间信息和特征信息,但是 ECA 在通道维度上并没有充分的与 SAM 相互作用,相同的,SAM 在空间维度上也没有充分的与 ECA 相互作用,因此降低了通道维度和空间维度中各自的空间信息和通道信息。基于此,本文首先将 MSF 融合后的特征通过并联的方式分别进入 ECA 和 SAM,以弥补 ECA 在通道维度上丢失的空间信息;然后将 ECA 得到的特征与经过 Tanh 激活函数映射的特征相融合后再次串联进入 SAM,在通道维度上可以更好地与 SAM 进行交互,合理地分配注意力权重,捕捉关键特征区域;最后将串联得到的特征、并联得到的特征和经过 ECA 与 Tanh 融合后的特征相结合,生成最后的关键特征信息。第 i 层 MSAG 模块的输出表示如下:

$$\mathbf{F}_{ER}^i = \text{ECA}(\mathbf{M}_n^i) + \text{Tanh}(\mathbf{M}_n^i) \quad (18)$$

$$\mathbf{F}_{MSAG}^i = \text{SAM}(\mathbf{F}_{ER}^i) + \text{SAM}(\mathbf{M}_n^i) + \mathbf{F}_{ER}^i \quad (19)$$

式中: $i=1,2,3,4$; $n,m=1,2,3$; ECA、SAM、Tanh 分别表示通道注意力、空间注意力和 Tanh 激活函数; \mathbf{F}_{ER}^i 表示将 \mathbf{M}_n^i 分别经过 ECA 和 Tanh 并如何生成的特征; \mathbf{F}_{MSAG}^i 表示将 \mathbf{M}_n^i 经过 SAM 并联生成的特征和将 \mathbf{F}_{ER}^i 经过 SAM 串联生成的特征与 \mathbf{F}_{ER}^i 累加生成的特征。

1.4 解码特征整合器

在解码器阶段,U-Net 利用上采样和跳跃连接方式与浅层网络进行交互。对于第 i 层 Dec 模块的输出可以表示如下:

$$\mathbf{C}^i = \begin{cases} \{\text{Up}(\mathbf{E}_3^5), \mathbf{F}_{MSAG}^4\}, & i = 4 \\ \{\text{Up}(\mathbf{D}^{i+1}), \mathbf{F}_{MSAG}^i\}, & i \neq 4 \end{cases} \quad (20)$$

$$\mathbf{D}^i = \mathbf{w}_{3 \times 3} \cdot (\mathbf{w}_{3 \times 3} \cdot \mathbf{C}^i + \mathbf{b}_{3 \times 3}) + \mathbf{b}_{3 \times 3} \quad (21)$$

式中: $i=1,2,3,4$; \mathbf{C}^i 表示由 \mathbf{F}_{MSAG}^i 与上采样后的 \mathbf{E}_3^5 在通道维度拼接后生成的特征; \mathbf{C}^i 表示 \mathbf{F}_{MSAG}^i 与第 $i+1$ 层 Dec 模块的输出经过上采样后在通道维度拼接后生成的特征; \mathbf{D}^i 表示 \mathbf{C}^i 经过两次 3×3 卷积后生成的特征。

虽然本文在跳跃连接阶段设计了 MSAG,可以更好地融合浅层网络的细节特征,但是由于多次的上采样提高特征图分辨率造成细节特征模糊,丢失肺实质的边缘信息,影响模型分割性能。因此,本文设计了解码特征整合器 DFI,可以有效地整合修复解码器捕捉的特征,并将最后分割的特征恢复成与输入图像相同分辨率的掩模图像。

如图 1 所示,DFI 级联了来自 Darknet53 的前 4 个 Residual Block 结构,构成了一条自下而上的路径,不仅可以利用残差结构避免模型在训练过程中出现梯度发散现象,还可以充分利用不同尺度的解码特征,而不过分依赖某一个尺度特征信息。为了更好地利用 DFI 输出的解码特征,本文利用解码器将最后 4 个解码特征整合在一起,并输入进一个 1×1 的卷积,生成最后的分割结果。第 i

层 RB 模块的输出可以表示如下:

$$\mathbf{F}_m^i = \begin{cases} \mathbf{w}_{3 \times 3} \cdot \mathbf{D}^i + \mathbf{b}_{3 \times 3}, & i = 4, m = 0 \\ \mathbf{R}_m^i + \mathbf{B}_m^i + \mathbf{D}^i, & m \neq 0 \end{cases} \quad (22)$$

$$\mathbf{B}_m^i = \mathbf{w}_{3 \times 3} \cdot \mathbf{F}_{n-1}^i + \mathbf{b}_{3 \times 3} \quad (23)$$

$$\mathbf{R}_m^i = \mathbf{w}_{3 \times 3} \cdot (\mathbf{w}_{1 \times 1} \cdot \mathbf{F}_{n-1}^i + \mathbf{b}_{1 \times 1}) + \mathbf{b}_{3 \times 3} \quad (24)$$

式中: $i=1,2,3,4$; $n=1,2,8$; $m=0,1,2,\dots,8$; \mathbf{F}_m^i 表示第 i 层第 m 个 Residual Block 提取到的特征; \mathbf{B}_m^i 表示将 \mathbf{F}_{n-1}^i 经过 3×3 卷积提取到的特征; \mathbf{R}_m^i 表示将 \mathbf{F}_{n-1}^i 经过 RB 模块中残差块提取到的特征。

因此,最后的输出表示为:

$$\mathbf{F}_{out} = \mathbf{w}_{1 \times 1} \cdot (\mathbf{F}_1^4 + \mathbf{F}_2^3 + \mathbf{F}_8^2 + \mathbf{F}_8^1) + \mathbf{b}_{1 \times 1} \quad (25)$$

2 实验

2.1 实验准备

1) 数据集

为提高模型分割性能的公正性,本文采用 LUNA16 数据集进行实验。该数据集作为 LIDC-IDRI 数据集的一个子集,其 CT 图像以 MHD 和 RAW 的格式存储。本文首先需要对 MHD 和 RAW 格式的三维图像做切片处理,得到供网络训练的 PNG 格式,切片后的像素大小为 512×512 。本文使用含有肺结节的 1 186 张图片作为原始数据集,由于原始数据集较少,因此,本文采用旋转、缩放、翻转等数据增强技术对每一个样本随机做一次变换,扩充样本数量至 2 372 张,缓解数据集稀缺的问题。

2) 评价指标

本文使用 Dice 系数、豪斯多夫距离 (Hausdorff distance, HD)、准确率 (ACC) 和敏感度 (Sen) 评价指标对模型分割性能进行评价。其中,Dice 和 HD 为主要评价指标。Dice 系数主要用来度量两个集合的相似度,取值范围为 $[0,1]$,值越大代表两个集合越相似,模型分割性能越好;HD 主要用来度量边界的分割准确度,值越小模型的分割性能越好。各评价指标的表现形式如下:

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \quad (26)$$

$$\text{HD} = \max \left\{ \max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y) \right\} \quad (27)$$

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \quad (28)$$

$$\text{Sen} = \frac{TP}{TP + FN} \quad (29)$$

式中: TP 、 TN 、 FP 、 FN 分别是真阳性数、真阴性数、假阳性数、假阴性数; $d(x, y)$ 表示 x 和 y 点之间的欧氏距离; $k+1$ 表示以识别的类别总数。

3) 实验环境设置

本文实验硬件环境:CPU 为 AMD Ryzen 5 4600 H with Radeon Graphics 3.00 GHz, GPU 为 NVIDIA GeForce GTX1650;软件环境:Python3.8, Pytorch1.10.1;操

作系统为 Windows 10。模型使用 SGD 优化器训练,学习率为 0.01,动量 0.9,权重衰减为 1×10^{-4} ,Batch size 设为 8,训练集、验证集、测试集比例为 8 : 1 : 1。

2.2 实验结果

1) 对比试验

利用以上评价指标,本文将 MCANet 与其他主流的肺实质分割算法进行比较,以验证 MCANet 的性能,评价结果如表 1 所示。

表 1 不同模型在 LUNA16 数据集上的分割结果
Table 1 Segmentation results of different models on the LUNA16 dataset

序号	年份	网络模型	Dice	HD	ACC	Sen
1	2015	U-Net ^[10]	0.93	4.967	0.949	0.964
2	2015	SegNet ^[11]	0.937	4.695	0.954	0.956
3	2015	FCNet ^[9]	0.968	4.539	0.969	0.976
4	2023	CANet ^[12]	0.978	4.37	0.991	0.981
5	2023	LDANet ^[13]	0.983	4.366	0.995	0.978
6	2023	MCANet	0.993	3.864	0.99	0.99

从表 1 可以看出,MCANet 在 4 个指标上都有明显的优化,特别是 HD,已经下降到了 3.864,与各模型相比,最多降低了 1.103,最少降低了 0.502。值得注意的是,HD 可以准确描述边界分割的效果,因此可以说明 MCANet 可以很好的对边界进行分割。并且,Dice 和 Sen 也分别达到 0.993 和 0.99,相较于其他主流模型已经得到明显的提升;此外,ACC 得分提高到 0.99,相较于 CANet 和 LDA-

Net 仅降低了 0.001 和 0.005,而相较于 U-Net、SegNet 和 FCNet 分别提高了 0.041、0.036 和 0.021。因此,4 个指标的优化可以证明 MCANet 预测的分割图与标签极为相似,能够达到更精密的分割效果。

上述模型在 3 张切片上的分割效果如图 5 所示,其中图 5(a)~(h) 分别为原始 CT 切片、肺实质标签及上述 6 个网络模型的分割结果图。从图 5 可以看出,像切片 3 这种边界简单、轮廓清晰并且没有多余肺组织的肺切片,各模型都可以准确的分割出肺实质的基本形态,但是从细节上看,如果肺实质的边界复杂、轮廓模糊或者肺切片中有其他的组织信息时,各模型的分割效果都不理想。由切片 1 看出,U-Net、SegNet、FCNet 和 CANet 分割掉了肺实质中的肺组织,从而不能最大限度地保留肺实质中的信息;由切片 2 可以看出,SegNet、FCNet 和 CANet 在分割完肺实质后还保留了一部分的躯干结构,并没有达到只分割肺实质的效果;另外,在切片 2 中,除了保留了一部分躯干结构外,从 U-Net 到 LDANet,各模型都没有达到准确分割肺实质内边界的效果,其中,U-Net、SegNet 和 CANet 虽然分割出了内边界的大致形态,但与标签相比相差较大,而 FCNet 和 LDANet 则直接将边界外组织信息辨别为肺实质,从而没有分割出准确的肺实质边界。最后,本文所提 MCANet 不仅可以准确分割出肺实质,而且可以最大限度保留肺实质中的组织信息,准确的分割出肺实质的内边界,虽然与标签还有所差别,但是相较于其他模型已经有明显提升,达到理想的分割效果。

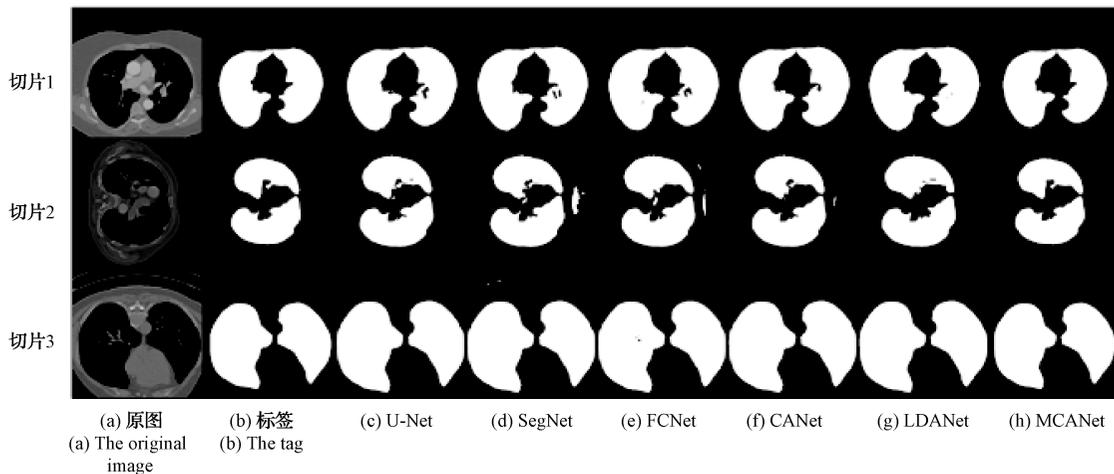


图 5 各模型分割效果对比

Fig. 5 Comparison of the segmentation effect of each model

2) 消融实验

为了评估 MCANet 不同模块的分割效果,本文在相同的实验环境下进行消融实验,各模块的分割效果如表 2 所示。

对提出的 MCANet 模型的 4 种变体进行消融实验,分别对应表 2 的方法 A~D,所有变体都使用改进后的

U-Net 作主干。如表 2 所示,方法 D 表示完整的 MCANet 模型,除了在 ACC 系数上与方法 B 和方法 D 分别相差 0.005 和 0.004,但在其他方面,其性能明显优于另外 3 个方法,这说明 MCANet 有着更好地分割效果。比较方法 A、方法 B 和方法 C 的 ACC 和 Sen 指标,可以看出,两项指标首先分别提高了 0.02 和 0.05,然后又分别降低了

表 2 具有不同变体的模型分割性能对比

Table 2 Comparison of model segmentation performance with different variants

	U-Net	MSFENet	MSAG	DFI	Dice	HD	ACC	Sen
方法 A	✓	✓			0.966	4.08	0.975	0.981
方法 B	✓		✓		0.98	3.939	0.995	0.986
方法 C	✓	✓	✓		0.987	3.926	0.994	0.982
方法 D	✓	✓	✓	✓	0.993	3.864	0.99	0.99

0.001 和 0.004, 这两项指标的波动间接说明了 MSFENet 可能会对 ACC 和 Sen 指标有一定的影响, 但是相比于提高的比例, 降低的比例几乎可以忽略不计。另外, 对方法 A 和方法 B 而言, 方法 B 的各项指标都优于方法 A, 从而证明 MSAG 可以更大程度的结合 ECA 和 SAM 机制的优点, 从而提升 MCANet 的性能。从方法 A 到方法 D, Dice 和 HD 系数都有所优化, 其中, Dice 系数共提高了 0.027, HD 系数共降低了 0.216, 这更加说明了 MSFENet、MSAG 和 DFI 可以使 MCANet 获得更优异的性能。

方法 A~D 在测试集上的分割效果比较如图 6 所示。从图 6 可以看出, 各方法都可以准确分割出肺实质的基本形态, 唯一的区别是各方法在细节上对肺实质边界的分割效果不同, 如图中箭头部分。在切片 3 中, 4 个方法分割效果大致相同, 但在切片 1 和切片 2 中, 方法 A、方法 B 和方法 C 对边界的细节把控不佳, 容易将肺实质外部组织辨认为肺实质内部的肺气管或者肺血管, 并将其保留下来。由此可以看出, 包含所有变体的 MCANet, 分割性能会更优异。

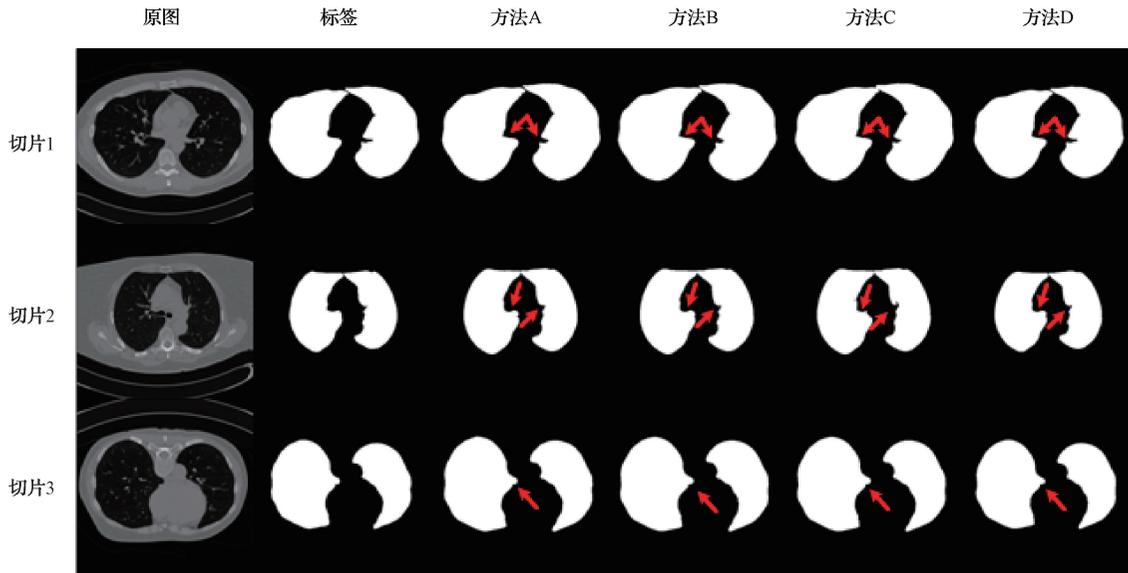


图 6 具有不同变体的模型分割效果对比

Fig. 6 Comparison of model segmentation effects with different variants

3 结论

本文提出基于 MSFENet 和 MSAG 的 MCANet 模型, 在 LUNA16 数据集上对模型性能进行实验验证, 得到了 0.993 的 Dice 和 3.864 的 HD, 并通过对比实验证明了 MCANet 模型有较高的分割性能。此外, 本文还通过消融实验证明了所设计的 MSFENet、MSAG 和 DFI 的有效性, 并且能够很好的利用在不同尺度上、不同维度上融合全局的上下文信息, 提高模型性能。

本文提出的 MCANet 有更加优异的分割性能, 能够为后续的医学检测研究或者临床诊断提供较好的协助。然而, 本文所提模型仍然具有较大的改进空间。本文设计

的 MSAG 只是利用 ECA 和 SAM 去计算不同维度特征信息内部之间的相关性, 导致模型会出现分割边缘缺失情况。另外, 本文分别在编码和解码阶段前后添加大量参数, 导致模型训练过程出现很大负担。因此, 在今后研究中, 将针对增强边界轮廓上每个像素的特征信息与全局特征信息之间的相似性, 降低模型参数量方面开展工作。

参考文献

- [1] LI R, XIAO C, HUANG Y, et al. Deep learning applications in computed tomography images for pulmonary nodule detection and diagnosis: A review[J]. Diagnostics, 2022, 12(2): 298.

- [2] 王桂棠, 林楨哲, 符秦沈, 等. 联合生成对抗网络的肺结节良恶性分类模型[J]. 仪器仪表学报, 2020, 41(11): 188-197.
WANG G T, LIN ZH ZH, FU Q SH, et al. Joint generative adversarial network model for classification of benign and malignant pulmonary nodules [J]. Chinese Journal of Scientific Instrument, 2020, 41(11): 188-197.
- [3] SUJI R J, GODFREY W W, DHAR J. Border to border distance based lung parenchyma segmentation including juxta-pleural nodules[J]. Multimedia Tools and Applications, 2023, 82(7): 10421-10443.
- [4] LIU X, SHEN H, GAO L, et al. Lung parenchyma segmentation based on semantic data augmentation and boundary attention consistency[J]. Biomedical Signal Processing and Control, 2023, 80: 104205.
- [5] 肖汉光, 冉智强, 黄金锋, 等. 基于电子计算机断层扫描图像的肺实质分割方法研究进展[J]. 生物医学工程学杂志, 2021, 38(2): 379-386.
XIAO H G, RAN ZH Q, HUANG J F, et al. Research progress in lung parenchyma segmentation based on computed tomography [J]. Journal of Biomedical Engineering, 2021, 38(2): 379-386.
- [6] 张杰, 施露露, 史振志, 等. 相差细胞图像的光照均衡与分割计数方法研究[J]. 电子测量与仪器学报, 2022, 36(5): 136-145.
ZHANG J, SHI L L, SHI ZH ZH, et al. Study of light equalization and segmentation counting method for cell phase contrast images [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(5): 136-145.
- [7] 王瑞丰, 朱铮涛, 冯端奇. 基于改进 LCCP 的堆叠物体分割算法[J]. 电子测量技术, 2022, 45(3): 118-124.
WANG R F, ZHU ZH T, FENG D Q. Improved LCCP-based stacked object segmentation algorithm [J]. Electronic Measurement Technology, 2022, 45(3): 118-124.
- [8] 禹文明, 刘伟, 张其超. CT 图像融合专家知识的肺结节良恶性诊断方法[J]. 国外电子测量技术, 2023, 42(7): 181-187.
YU W M, LIU W, ZHANG Q CH. CT image fusion expert knowledge-based method for benign and malignant diagnosis of pulmonary nodules [J]. Foreign Electronic Measurement Technology, 2023, 42(7): 181-187.
- [9] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference. Berlin: Springer International Publishing, 2015: 234-241.
- [11] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation [J]. CoRR, 2015, DOI: abs/1511.00561.
- [12] XIE X, ZHANG W, PAN X, et al. Canet: Context aware network with dual-stream pyramid for medical image segmentation[J]. Biomedical Signal Processing and Control, 2023, 81: 104437.
- [13] CHEN Y, FENG L, ZHENG C, et al. LDANet: Automatic lung parenchyma segmentation from CT images [J]. Computers in Biology and Medicine, 2023, 155: 106659.
- [14] 刘莹芳, 柏正尧, 李琼. 一种基于 CT 图像的肺实质分割方法[J]. 云南大学学报(自然科学版), 2019, 41(3): 456-463.
LIU Y F, BAI ZH Y, LI Q. A method of pulmonary parenchyma segmentation based on CT image [J]. Journal of Yunnan University (Natural Sciences Edition), 2019, 41(3): 456-463.
- [15] 毕蓉蓉, 王进科. CT 图像下结合 RCNN 与 U-Net 的肺实质自动分割方法[J]. 哈尔滨理工大学学报, 2021, 26(3): 74-82.
BI R R, WANG J K. Automatic lung segmentation using RCNN combing U-Net in CT [J]. Journal of Harbin University of Science and Technology, 2021, 26(3): 74-82.
- [16] 侯英竹. 融入双注意力模块的 U-Net 肺结节图像分割方法[J]. 计算机科学与应用, 2022, 12(7): 1756-1764.
HOU Y ZH. A U-Net lung nodule image segmentation method incorporating a dual attention module [J]. Computer Science and Application, 2022, 12(7): 1756-1764.
- [17] 张宇杰, 叶西宁. 基于 WU-Net 网络的肺结节图像分割算法[J]. 计算机应用研究, 2022, 39(3): 895-899,905.
ZHANG Y J, YE X N. Improved lung nodules segmentation algorithm based on WU-Net [J].

- Application Research of Computers, 2022, 39(3): 895-899,905.
- [18] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. arXiv preprint arXiv: 1804.02767, 2018.
- [19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [20] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [21] ZHU X, CHENG D, 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.

作者简介

许圳兴, 硕士研究生, 主要研究方向为医学图像处理、深度学习等。

E-mail: 1550269092@qq.com

余耀, 博士, 讲师, 主要研究方向为光电测试系统、图像处理等。

E-mail: yuyao0222@126.com

赵东, 博士, 讲师, 主要研究方向为人工智能图像处理等。

E-mail: dzhao@cw Xu.edu.cn

陈园, 工程师, 主要研究方向为数字图像处理与人脸识别等。

E-mail: chenYuan10@hikvision.com

范圣旺, 高级技术支持总监, 主要研究方向为智能数字图像处理等。

E-mail: fanshengwang@hikvision.com