

DOI:10.19651/j.cnki.emt.2518444

改进 Salience-DETR 的多部位小目标病灶检测算法^{*}

张茜¹ 武雨露^{1,2} 郑冰洁³ 董洁⁴ 杨关¹(1. 中原工学院人工智能学院 郑州 450007; 2. 中原工学院计算机学院 郑州 450007;
3. 河南省肿瘤医院放射科 郑州 450008; 4. 华北水利水电大学信息工程学院 郑州 450046)

摘要: 多部位病灶检测在临床中具有重要意义,但不同部位的病灶在大小和形状上差异较大,且 CT 图像中病灶区域通常较小、病灶区域与周围背景相似,导致检测难度较大。针对上述问题,提出了一种基于 Salience-DETR 模型改进的多部位小目标病灶检测算法。首先,设计一种高效空间通道协同注意力机制 ESCA,用于对 Backbone 提取的多尺度特征进行重构,加强模型对病灶重要信息的关注;其次,结合 DenseASPP 和 AugFusion 模块对跨层 token 融合网络进行优化,增强不同层级特征的多尺度融合能力;最后,引入 Inner-GIoU 损失函数,加速模型收敛并提高小目标病灶的检测性能。实验结果显示,在每张图像假阳性数目为 0.5 至 4 的情况下,改进后的模型在公开数据集 DeepLesion 和外部验证集上的平均检测灵敏度分别达到了 83.26% 和 82.33%。可见,所提算法在多部位小目标病灶检测任务中表现出较高的检测精度和良好的泛化能力,具有一定的实际临床应用价值。

关键词: 病灶检测;小目标;多尺度特征融合;Salience-DETR;注意力机制;Inner-GIoU

中图分类号: TP399; TN98 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Improved Salience-DETR for multi-organ small target lesion detection

Zhang Qian¹ Wu Yulu^{1,2} Zheng Bingjie³ Dong Jie⁴ Yang Guan¹

(1. School of Artificial Intelligence, Zhongyuan University of Technology, Zhengzhou 450007, China;

2. School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007, China;

3. Department of Radiology, Henan Cancer Hospital, Zhengzhou 450008, China; 4. School of Information Engineering Institute, North China University of Water Resources and Electric Power, Zhengzhou 450046, China)

Abstract: Multi-organ lesion detection is of great clinical significance. However, lesions in different anatomical regions vary significantly in size and shape, and in CT images, lesion areas are typically small and similar to surrounding tissues, which increases the difficulty of detection. To address these challenges, this paper proposes an improved multi-organ small lesion detection algorithm based on the Salience-DETR model. Firstly, an Efficient Spatial-Channel Collaborative Attention (ESCA) mechanism is designed to reconstruct the multi-scale features extracted by the backbone, enhancing the model's focus on important lesion information. Secondly, the DenseASPP and AugFusion modules are incorporated to optimize the cross-layer token fusion network, improving multi-scale feature fusion across different levels. Finally, an Inner-GIoU loss function is introduced to accelerate model convergence and improve the detection performance for small lesions. Experimental results show that, under the condition of 0.5 to 4 false positives per image, the improved model achieves average detection sensitivities of 83.26% and 82.33% on the public DeepLesion dataset and an external validation set, respectively. These results demonstrate that the proposed algorithm achieves high detection accuracy and good generalization performance for multi-organ small lesion detection, with promising potential for real-world clinical applications.

Keywords: lesion detection; small target; multi-scale feature fusion; Salience-DETR; attention mechanism; Inner-GIoU

0 引言

目前,多数研究仍集中于肝脏、肺部和脑部等单一部位

的病灶识别^[1-3]。但在实际临床中,医生往往需对患者多个部位的病灶进行综合分析,尤其是在癌症晚期,病灶易发生转移,形成多部位病灶^[4]。因此,开发一种可同时识别多个

收稿日期:2025-03-26

* 基金项目:国家自然科学基金(82202270)、河南省自然科学基金青年科学基金(252300420995)、河南省高等学校重点科研项目(24B520048)资助

部位病灶的检测模型,对提高诊断的准确性和全面性具有重要意义。

随着 CT 扫描数据集 DeepLesion^[5] 的公开,多部位病灶检测逐渐受到关注^[6-14],其中大部分研究基于卷积神经网络架构展开。例如,Liu 等^[6]提出了一种 3D 上下文增强的区域卷积神经网络,将 3D 上下文融合到 2D CNN 网络中,以提取和聚合特征,最终使用 2D 检测网络进行预测。Liu 等^[7]利用特征注意力增强和特征互补性机制,对原始的 YOLOv3 模型进行了全面改进。许金亚等^[8]设计了 APFN 模块以及新的损失函数,提升了小目标病灶的平均检测精度以及模型对负样本的鉴别能力。Zhu 等^[9]提出了一种基于循环金字塔结构的检测网络,旨在解决小病灶在检测过程中容易被忽略的问题。Guan 等^[10]提出了一种基于层内和跨层特征交互的新颖 FCOS 模型,增强了病灶细节特征的表示能力。Jian 等^[11]引入时间相关性加权融合模块,通过空间偏移注意机制提升网络对病变微小空间偏移的敏感性。

上述基于 CNN 架构进行病灶检测的方法,虽然从不同角度解决了部分检测难题,但其仅关注病灶区域的局部上下文信息,缺乏病灶与周围相似背景的关系建模能力。当面对多部位复杂的病灶检测场景,尤其是小病灶和背景相似情况下,检测效果不佳。近年来,基于 Transformer 的端到端目标检测器 DETR^[15]被提出,其通过交叉注意力机制捕捉全局信息和长程依赖,展现出强大的目标检测能力。但 DETR 仍存在信息冗余和尺度偏差的问题,为解决这一问题,Salience-DETR^[16]通过引入与尺度无关的显著性监督机制和分层过滤查询细化,减少了编码器中的信息冗余并提升了检测性能。

尽管 Salience-DETR 在一定程度上提升了目标检测性能,但其在医学图像检测中的表现仍有提升空间。医学图像中的检测任务相比自然图像更加困难,尤其在多部位小目标病灶检测任务中,不同部位的病灶位置分散、大小形态差异大、且部分病灶与背景组织高度相似,容易出现误检。为此,本文提出了一种基于 Salience-DETR 模型改进的多部位小目标病灶检测算法,旨在解决多部位小目标病灶难检测、病灶与周围背景相似易混淆等问题。主要贡献如下:

1)在 Backbone 之后加入提出的高效空间通道协同注意力机制 (efficient spatial-channel collaborative attention, ESCA),使得模型能够关注图像中的重要特征,从空间维度和通道维度抑制相似背景的干扰,加强网络特征提取能力。

2)构建 AugFusion 特征融合模块并结合密集空间金字塔池化模块改进跨层 token 融合网络,优化多尺度融合策略,增强模型对于不同尺寸病灶的检测能力。

3)将 Inner-IoU 与 GIoU 结合,引入一个由尺度因子 ratio 控制的辅助边界框,得到 Inner-GIoU 损失函数,加速模型收敛并提高小目标病灶的检测精度。

4)所提算法在公开数据集 DeepLesion 上表现良好,并使用外部验证集验证了模型的泛化能力。

1 方法改进

1.1 基准模型

Salience-DETR 是一种基于 Transformer 的端到端显著性引导的目标检测模型,并有 R50、ConvNeXt-L、Swin-L、FocalNet-L 四个版本。为了平衡检测精度和训练速度,选择了 Salience-DETR-Swin-L 作为基准模型。其由主干网络,显著性引导的分层过滤机制、编码器、查询细化模块以及解码器组成。其中,主干网络为 Swin-Transformer,负责对输入的图像进行特征提取。显著性引导的分层过滤机制通过额外的 MLP 预测 query 的显著性分数,并在特征层和编码器层进行监督过滤,保留最显著的 query 进行自注意力编码,增强前景信息的表达,同时降低计算成本。然而,该机制可能会导致查询之间的语义不一致,为此,Salience-DETR 在编码器之后又设计了 3 个查询细化模块,即背景嵌入 (Background embedding)、跨层 token 融合网络 (Cross-level token fusion) 和冗余去除 (Redundancy removal) 模块,从前景-背景信息的差异、多尺度特征融合和选择策略的角度来弥补差距和提升性能。最终,经过查询细化处理的特征输入解码器,生成最终的目标检测结果。

本文在基准模型上共有 3 个改进点,分别是设计 ESCA 注意力机制,对主干网络提取的多尺度特征进行注意力增强;构建 AugFusion 特征融合模块并结合密集空间金字塔池化模块改进跨层 token 融合网络;损失函数的优化。改进算法的整体流程如图 1 所示。

1.2 ESCA 注意力机制

CT 图像中的病灶区域较小,且病灶与背景区域相似,导致特征提取效果不佳,存在误检的风险。为解决这一问题,设计了高效空间通道协同注意力机制 ESCA,在空间维度和通道维度上对主干网络提取的多尺度特征进行注意力增强,提升网络对重要信息的关注,抑制背景信息干扰。

ESCA 基于 EMA (efficient multi-scale attention) 注意力机制^[17]进行构建,仍采用特征分组、并行子网络和跨空间学习等策略。其结构如图 2 所示,采用 3 条并行路径来提取分组特征图的注意力权重描述,其中 A、B 分支仍通过水平和垂直方向池化建模通道间长程依赖。而在 EMA 中,C 分支所使用的 3×3 卷积主要用于特征增强和跨空间信息的交互,弥补 A、B 分支因池化操作而导致的细节丢失,但标准的卷积会对病灶附近的背景区域进行过多卷积操作,导致关键信息和冗余信息混杂一起,导致无效特征的传播,影响注意力计算的有效性。为此,在 ESCA 中,对 C 分支进行优化,将原始 3×3 卷积替换为空间和通道重建卷积 (spatial and channel reconstruction convolution, SCConv)^[18],SCConv 由空间重建单元 (spatial reconstruction unit, SRU) 和

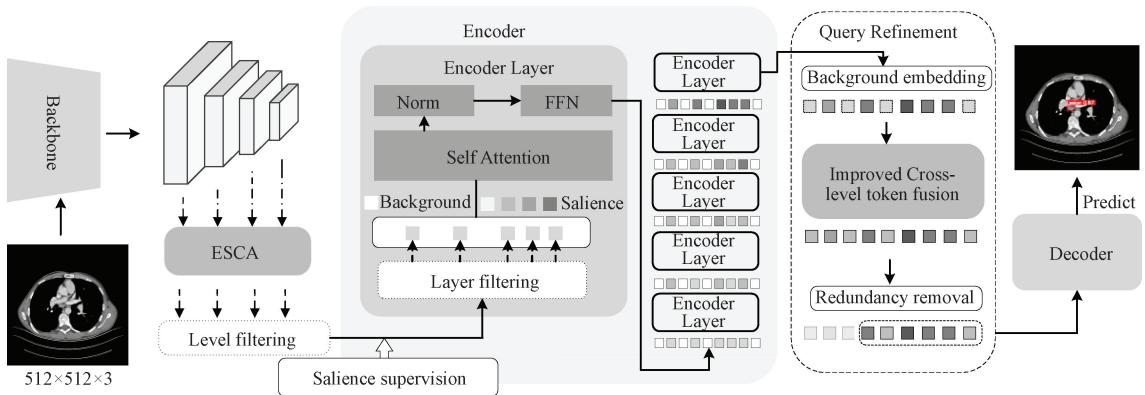


图 1 改进算法的结构

Fig. 1 The structure of the improved algorithm

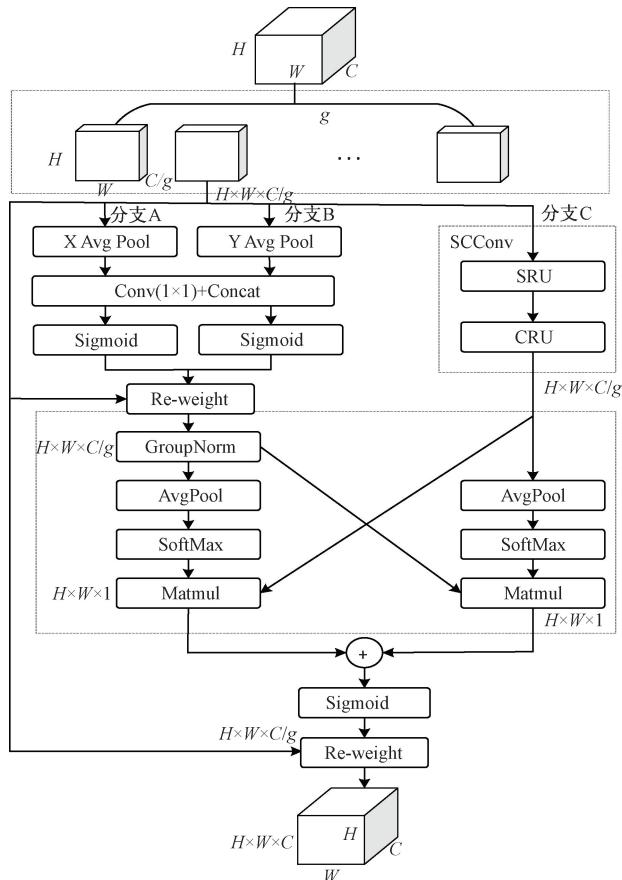


图 2 ESCA 注意力机制

Fig. 2 Efficient spatial-channel collaborative attention

通道重建单元(channel reconstruction unit, CRU)共同组成;其中 SRU 用于减少空间冗余,避免对无关区域的过度卷积;CRU 则用于降低通道冗余,并提升通道间的交互能力。如此,ESCA 注意力机制不仅能够捕获通道的长程依赖关系,还能在通道与空间信息之间建立更有效的交互,增强特征表达能力。

具体而言,给定输入特征图 $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$, 先按照通道维

度划分为 g 个子特征。即 $\mathbf{X} = [X_0, X_1, X_2, X_3, \dots, X_{g-1}]$, $\mathbf{X} \in \mathbb{R}^{(\frac{C}{g}) \times H \times W}$ 。其中, C 、 H 、 W 分别表示特征图的通道数、高度和宽度。

A 分支和 B 分支首先沿着对应方向进行池化操作、分别计算感受野,如式(1)和(2)所示。

$$z_h^A(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (1)$$

$$z_w^B(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (2)$$

其中, $z_h^A(h)$ 为水平方向的感受特征, $z_w^B(w)$ 为垂直方向的感受特征, x_c 为输入特征。随后, 将两者进行拼接和 1×1 卷积操作, 再经过 Sigmoid 激活函数与原始特征进行矩阵乘法运算生成第一个空间注意力图, 如式(3)所示。

$$f_A = f_B = \delta(\text{Conv}_{1 \times 1}(\text{Concat}(z_h^A(h), z_w^B(w)))) \times X \quad (3)$$

C 分支中, 输入特征先经过空间重建单元 SRU, 得到空间细化特征;再经过通道重建单元 CRU, 得到通道细化特征, 如式(4)所示。

$$f_c = \text{CRU}(\text{SRU}(x_c)) \quad (4)$$

其中, SRU 通过“权重分离-重构”操作处理空间冗余,如图 3 所示。其利用组归一化(GN)中的可学习参数 γ 评估特征信息量, 并根据 γ 设定阈值, 将特征划分为信息丰富(W1)和信息较少(W2)两类, 随后交叉重构整合信息, 增强空间细节表达, 同时抑制冗余信息。CRU 采用“拆分-变换-融合”策略, 如图 4 所示。先按比例 α 沿通道维度划分为 αC 和 $(1-\alpha)C$ 个通道, 并通过 1×1 卷积降低计算量。变换时, 上半部分采用分组卷积和逐点卷积提取高级特征, 下半部分通过 1×1 逐点卷积复用上半部分信息。最后, 采用选择性卷积自适应融合策略生成通道细化特征图, 在保留完整特征的同时, 大幅减少参数和计算量。

最后, A、B 分支生成的第一个注意力图与 C 分支进行跨空间交互, 将交互之后的注意力图进行拼接, 经过 Sigmoid 生成的最终权重与原始特征图进行矩阵乘法操

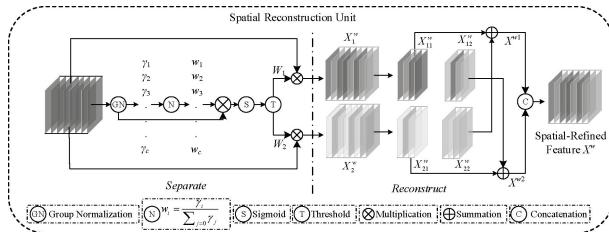


图 3 空间重建单元

Fig. 3 The architecture of spatial reconstruction unit

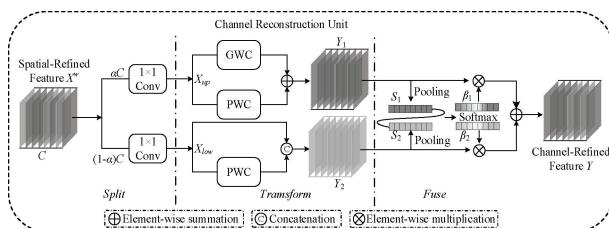


图 4 通道重建单元

Fig. 4 The architecture of channel reconstruction unit

作,形成空间和通道协同增强的注意力映射,具体如式(5)~(7)所示。

$$f_1 = f_c \times S(\text{Avg}(\text{GN}(f_A))) \quad (5)$$

$$f_2 = S(\text{Avg}(f_c)) \times \text{GN}(f_A) \quad (6)$$

$$O_c = X \times \delta(f_1 + f_2) \quad (7)$$

其中, δ 表示 Sigmoid 激活函数, S 表示 SoftMax 激活函数, Avg 表示全局平均池化, GN 表示组归一化, O_c 最终的注意力增强特征图。

1.3 改进的跨层 token 融合

原算法中的跨层 token 融合网络利用路径聚合网络结构(path aggregation network, PAN),结合 Fusion 特征融合模块实现跨层信息交互和多尺度特征融合。其有两点不足之处:1)在自顶向下的融合过程中,由于最高特征图分辨率低,直接对其进行上采样会导致部分语义信息和位置信息丢失,影响小目标病灶的检测精度;2)Fusion 模块基于 CSPNet 进行通道划分,如图 5 所示,仅对部分通道进行 RepVGGPlusBlock 处理,另一部分未经处理直接传递,导致特征利用率不足。

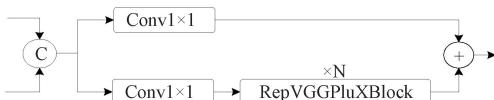


图 5 Salience-DETR 中的 Fusion 模块

Fig. 5 The Fusion module in Salience-DETR

为此,本文对跨层 token 融合网络进行优化,主要有以下两个改进:1)引入密集空间金字塔池化模块(densely connected atrous spatial pyramid pooling, DenseASPP)^[19],在 PAN 结构进行自顶向下融合之前,先通过 DenseASPP 模块对最高特征图进行增强。DenseASPP 利用多尺度空洞卷积结合密集连接扩大感受野,在不显著增加计算量的

情况下,提高小目标的表达能力。2)在原 Fusion 模块基础上,设计 AugFusion 特征融合模块进行更有效的融合操作。如图 6 所示,先对两个特征图进行通道维度拼接,再通过通道注意力机制计算重要性权重,进一步利用 3×3 卷积增强局部信息交互,最后,生成的特征与原始输入拼接,并送入 Fusion 模块进一步融合。

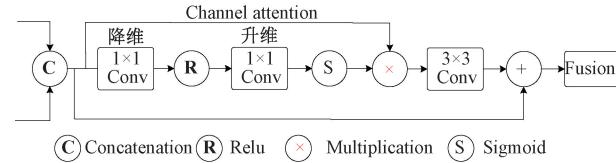


图 6 AugFusion 模块

Fig. 6 The AugFusion module

改进的跨层 token 融合网络(improved cross-level token fusion)结构如图 7 所示,最高特征图经过 DenseASPP 模块增强,然后送入结合 AugFusion 模块的 PAN 结构进行自顶向下和自底向上的多尺度特征融合,进一步提高小目标和相似背景下的病灶检测效果。

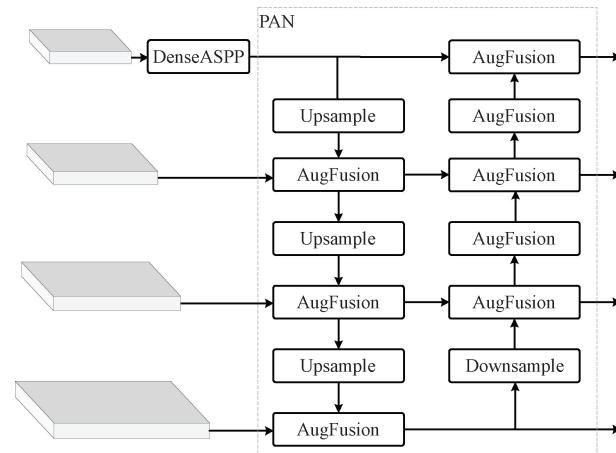


图 7 改进的跨层 token 融合网络

Fig. 7 Improved cross-level token fusion

1.4 损失函数优化

GIoU 损失函数在理论上能够描述边界框的回归状态,但由于不同部位的病灶在大小、形状和位置上差异显著,GIoU 难自适应调整,导致定位不准确。此外,当预测框与真实框区域完全重合时,GIoU 的计算等同于 IoU,此时损失值无法进一步优化,尤其在小目标病灶检测中更明显,进而减缓模型收敛速度。为解决上述问题,本研究借助 Inner-IoU^[20]辅助边界框的思想,引入 Inner-GIoU,对原算法使用的 GIoU 损失函数进行优化。该方法使用一个尺度因子 ratio 来控制不同尺度辅助边界框的生成,从而加速收敛并提高小目标病灶的检测精度。

Inner-GIoU 的计算如式(8)所示。

$$L_{\text{Inner-GIoU}} = L_{\text{GIoU}} + \text{IoU} - \text{IoU}^{\text{inner}} \quad (8)$$

Inner-IoU 的计算过程如式(9)~(15)所示。

$$b_l^{\text{gt}} = x_c^{\text{gt}} - \frac{w^{\text{gt}} \times \text{ratio}}{2}, b_r^{\text{gt}} = x_c^{\text{gt}} + \frac{w^{\text{gt}} \times \text{ratio}}{2} \quad (9)$$

$$b_t^{\text{gt}} = y_c^{\text{gt}} - \frac{h^{\text{gt}} \times \text{ratio}}{2}, b_b^{\text{gt}} = y_c^{\text{gt}} + \frac{h^{\text{gt}} \times \text{ratio}}{2} \quad (10)$$

$$b_l = x_c - \frac{w \times \text{ratio}}{2}, b_r = x_c + \frac{w \times \text{ratio}}{2} \quad (11)$$

$$b_t = y_c - \frac{h \times \text{ratio}}{2}, b_b = y_c + \frac{h \times \text{ratio}}{2} \quad (12)$$

$$\begin{aligned} \text{inter} &= (\min(b_r^{\text{gt}}, b_r) - \max(b_l^{\text{gt}}, b_l)) \times \\ &(\min(b_b^{\text{gt}}, b_b) - \max(b_t^{\text{gt}}, b_t)) \end{aligned} \quad (13)$$

$$\begin{aligned} \text{union} &= (w^{\text{gt}} \times h^{\text{gt}}) \times \text{ratio}^2 + (w \times h) \times \text{ratio}^2 - \text{inter} \\ &\quad (14) \end{aligned}$$

$$\text{IoU}_{\text{inner}} = \frac{\text{inter}}{\text{union}} \quad (15)$$

其中, (x_c, y_c) 表示真实框的中心点; $(x_c^{\text{gt}}, y_c^{\text{gt}})$ 表示锚框的中心点; w^{gt} 和 h^{gt} 分别表示真实框的宽度和高度; w 和 h 分别表示锚框的宽度和高度; 从真实框中心点计算得到的左、右、上、下边界, 根据尺度因子缩放后分别用 b_l^{gt} 、 b_r^{gt} 、 b_t^{gt} 、 b_b^{gt} 表示; 类似, 从锚框中心点计算得到的左、右、上、下边界, 使用相同尺度因子缩放后分别用 b_l 、 b_r 、 b_t 、 b_b 表示; inter 为真实框和锚框之间的交集面积; union 为真实框和锚框之间的并集面积。

2 数据集和实验设置

2.1 数据集

实验采用了两个数据集: 一是 DeepLesion^[5] 公开数据集, 二是河南省肿瘤医院提供的包含 26 位患者全身 CT 图像的私有数据集, 以便于体现临床辅助诊断效果。所有数据均获得了各机构审查委员会的图标审查批准, 放弃了知情同意的必要性。

DeepLesion 是目前最大规模、具有不同部位病灶级别标注的临床医学 CT 图像数据集, 由 4 427 位患者的 10 594 次 CT 研究中截取的 32 120 张切片组成。每张切片包含 1~3 个病灶, 总共 32 735 个病灶。数据集涵盖了 8 种病灶实例, 包括骨骼、腹部、纵隔、肝脏、肺、肾脏、软组织和骨盆。按照官方的数据集划分, 70% 用作训练集, 15% 用作验证集, 15% 用作测试集。DeepLesion 中标注的病灶类型仅在验证集和测试集中给出, 以便有助于可视化不同病灶类型的检测结果, 这意味着检测实际上是一个二元检测任务。私有数据集包含 7 830 张 CT 图像, 经筛选与预处理后, 保留 933 张病灶清晰的图像, 作为外部验证集, 用于评估模型的泛化能力。

2.2 实验环境配置

本实验基于 Linux 操作系统完成, 在显卡 NVIDIA GeForce RTX 3090, 显存为 24 GB 的 GPU 上进行, 采用 Python3.8 作为编译环境, 并使用 Pytorch1.12.0 深度学习框架进行训练。训练阶段采用 AdmaW 作为优化器, 初

始学习率设置为 0.0001, 训练批次大小 batch_size 为 2, 训练轮次 epoch 为 12, 经过 10 轮之后衰减为原来的 1/10, 其他训练超参数均采用默认值。

2.3 评价指标

在医学图像检测任务中, 每张图像上不同假阳性 (False positives per image, FPs) 数目下的灵敏度 (Sensitivity) 是一个常用的评估指标。在 DeepLesion 数据集和外部验证集上展示了每张图像在 0.5、1、2 和 4 个假阳性情况下的检测灵敏度和平均检测灵敏度结果。此外, 为了进一步评估所提算法在不同尺寸病灶大小上的检测性能, 根据 RECIST(实体瘤最大径度标准)指南, 该指南由两条正交线表示病灶的长轴和短轴。本文按照该标准, 取长短轴之和的平均值作为病灶大小的划分依据, 将 DeepLesion 测试集划分为 <10 mm、10~30 mm、>30 mm 三组。

3 实验结果分析

3.1 消融实验

通过消融实验来验证本文所提算法模块的有效性。在基线模型 (Salience-DETR) 上加入 ESCA 注意力机制, 改进方法记为 A; 加入改进的跨层 token 融合网络, 改进方法记为 B; 引入 Inner-GIoU 损失函数, 改进方法记为 C。实验在 DeepLesion 测试集上进行, 结果如表 1 所示。

表 1 DeepLesion 测试集上的消融实验

Table 1 Ablation experiments on the DeepLesion test set

Method	Sensitivity(%)@FPs				Avg
	@0.5	@1	@2	@4	
Baseline	71.73	80.30	86.27	90.76	82.26
Baseline+A	72.73	80.47	87.06	91.17	82.73
Baseline+B	72.54	80.51	86.40	91.13	82.65
Baseline+A+B	73.65	81.08	86.92	90.55	83.05
Baseline+A+B+C	74.13	81.24	87.09	90.59	83.26

实验结果表明, 在基线模型上单独加入 ESCA 注意力机制, 在不同假阳性数目下的检测灵敏度均有所提升, 且平均检测灵敏度提高了 0.47%, 表明加入 ESCA 注意力机制有助于增强模型对于病灶区域特征的关注度, 减少特征图中的冗余信息; 在此基础上, 采用改进的跨层 token 融合网络来优化多尺度特征融合的能力, 平均检测灵敏度进一步提升了 0.32%; 最后在前两种改进的基础之上引入 Inner_GIOU 损失函数, 通过优化边界框回归, 使检测更加准确, 此时, 模型的检测效果进一步达到了最佳, 平均检测灵敏度提升至 83.26%。从表中整体结果可见, 相较于基线模型, 改进方法在 FPs 为 0.5、1 和 2 时分别提升了 2.4%、0.94%、0.82%。尽管在 FPs 为 4 时, 灵敏度略微下降, 但平均检测灵敏度仍有所提高, 整体表现优秀。综

上,本文对基线模型的改进工作在多部位病灶检测中均有所贡献。

3.2 对比实验

将本文改进算法与目标检测领域中具有代表性的几个主流算法:Faster-RCNN、Cascade-RCNN、FCOS、DETR、DINO、Salience-DETR、Relation-DETR、MI-DETR 在 DeepLesion 测试集上做对比实验。如表 2 所示,相较于这些主流算法,本文算法在每张图像假阳性数目为 0.5、1、2 情况下均展现出了较优的检测效果。

表 2 DeepLesion 测试集上与主流算法的对比

Table 2 Comparison with mainstream algorithms on the DeepLesion test set

Method	Sensitivity(%)@FPs				Avg
	@0.5	@1	@2	@4	
Faster-RCNN ^[21]	58.47	69.04	77.94	82.50	71.99
Cascade-RCNN ^[22]	57.44	68.50	76.68	80.98	70.90
FCOS ^[23]	60.89	69.88	76.96	82.40	72.53
DETR	61.68	70.59	77.25	81.37	72.73
DINO ^[24]	70.33	77.35	82.02	86.62	79.08
Salience-DETR	71.73	80.30	86.27	90.76	82.26
Relation-DETR ^[25]	72.96	80.99	86.71	90.63	82.83
MI-DETR ^[26]	68.33	77.69	84.33	89.26	79.90
Ours	74.13	81.24	87.09	90.59	83.26

此外,为了进一步验证本文改进算法的有效性,与已有研究也进行了对比实验,如表 3 所示。其中,文献[7]通过优化像素级注意力提升了特征提取的能力,但未对目标框的尺寸进行自适应调整,导致目标框回归误差较大。文献[9]利用循环金字塔结构,并结合多尺度空洞卷积使特征提取更加充分,并提高了网路的感受野,但平均的检测灵敏度仅为 78.26%。文献[11]结合时序相关性加权融合策略和空间偏移注意力优化病灶检测,文献[12]引入了深度感知和跳跃分层训练机制。但由于多切片的输入,会给模型带来更多的噪声。尽管上述方法在一定程度上改善了病灶检测性能,但整体检测效果仍不理想,均低于本文算法。文献[13]设计了一种具有位置注意力引导的多路径特征金字塔网络,其平均检测灵敏度较高,但该注意力机制过度依赖于精确的位置编码。文献[14]基于 Mask-RCNN 模型进行了病灶的检测与分割,并引入 BiFPN 进行多尺度特征融合,但由于其加权机制可能导致小病灶在较深层特征图中的表征被稀释,影响检测精度。相比之下,文献[13]和文献[14]检测效果具有挑战性,本文模型在每张图像假阳性数目为 4 下的灵敏度略微较低,但平均检测灵敏度仍达到较优水平,整体表现更为优秀。

3.3 在不同病灶大小上的表现

如图 8 所示,在 DeepLesion 测试集上,分别报告了

表 3 DeepLesion 测试集上与已有研究的对比
Table 3 Comparison with existing studies on the DeepLesion test set

Method	Sensitivity(%)@FPs				Avg
	@0.5	@1	@2	@4	
文献[7]	61.27	70.43	78.67	85.07	73.86
文献[9]	66.10	75.40	83.10	88.30	78.26
文献[11]	68.02	79.8	86.71	91.42	81.49
文献[12]	64.86	74.39	81.41	86.04	76.68
文献[13]	73.19	81.12	86.91	91.37	83.15
文献[14]	73.09	81.33	87.03	91.16	83.15
Ours	74.13	81.24	87.09	90.59	83.26

3 组不同尺寸大小的病灶在每张图像假阳性数目为 [0.5, 1, 2, 4] 时的平均检测灵敏度。实验结果表明,相较于 Salience-DETR,本文算法在不同大小的病灶检测上均有所提升。其中,对于 <10 mm 的小目标病灶,平均检测灵敏度为 79.27%,提高了 1.17%;10~30 mm 的病灶,平均检测灵敏度为 85.67%,提高了 0.5%;对于 >30 mm 的病灶,平均检测灵敏度为 83.79%,提高了 2.12%。综上所述,本文算法在多部位不同病灶大小的检测上均展现出较优的性能。

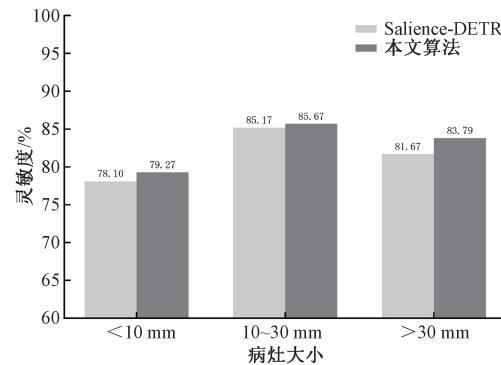


图 8 不同病灶大小的检测性能

Fig. 8 Detection performance for lesions of different sizes

3.4 可视化分析

为了评估本文算法在不同部位上的病灶检测效果,将其与 Salience-DETR 在 DeepLesion 测试集上的推理结果进行可视化对比,展示了部分对比结果,如图 9 所示。第 1 行为病灶的真实标注,第 2 行和第 3 行分别展示了基线模型和本文算法的检测结果。第 1 列至最后一列分别为纵隔、肾脏、软组织和肺部 4 种病灶实例。从第二列图像中可以看出,相较于基线模型,本文算法在小目标病灶的检测上置信度更高,能够更准确地识别出小目标病灶。从第 3 列图像中可以观察到,当病灶与周围背景相似时,Salience-DETR 出现了误检,而本文算法避免了这一情况,能够准确地将病灶检测出来。整体来看,本文方法对不同

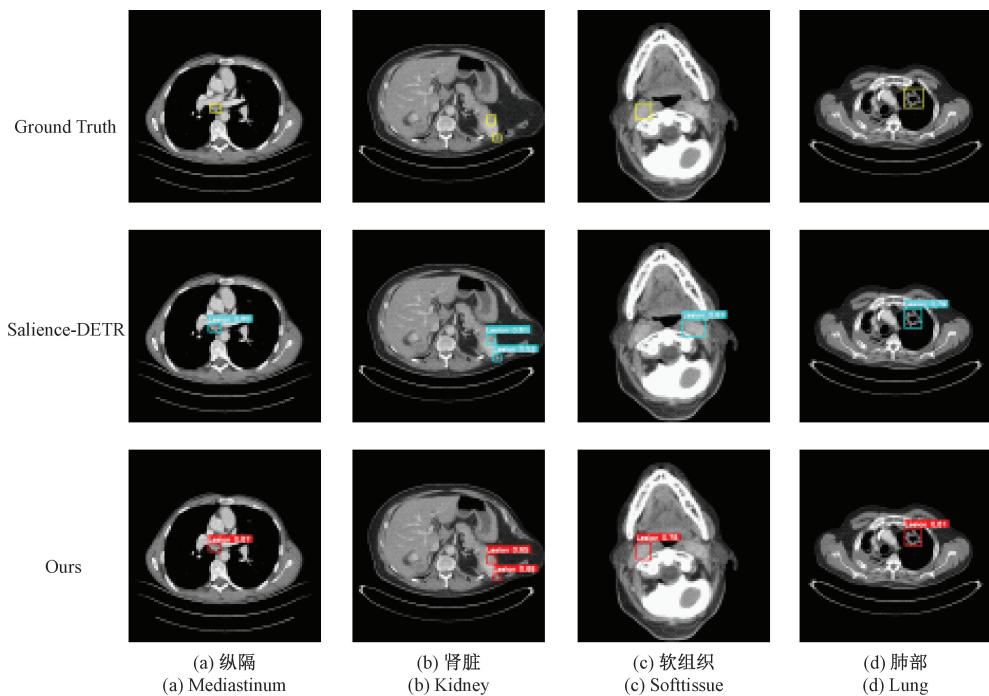


图 9 不同部位的病灶检测结果可视化

Fig. 9 Visualization of detection results for lesions in different organs

部位的病灶检测更精确。

3.5 模型泛化性验证

为了验证本文改进算法的泛化能力,在外部验证集上进行了实验,并与 Salience-DETR 进行对比。实验结果如表 4 所示。从表 4 中可以看出,相较于基线模型,改进后的算法在外部验证集上的检测效果有所提升,表明本文算法具有较好的泛化能力和实际临床应用价值。

表 4 不同算法在外部验证集上的表现

Table 4 The performance of different algorithms on the external validation set

Method	Sensitivity (%)@FPs				Avg
	@0.5	@1	@2	@4	
Salience-DETR	71.37	78.85	85.46	87.67	80.84
Ours	74.23	80.40	84.80	89.87	82.33

4 结论

为实现对患者多部位病灶的统一识别,并解决 CT 图像中小目标病灶难检测、病灶与背景相似等问题。本文提出了一种基于 Salience-DETR 模型改进的多部位小目标病灶检测算法。在 Backbone 之后加入 ESCA 注意力机制,强化特征图处理,提升网络对病灶关键特征的捕捉能力。同时,结合 DenseASPP 和 AugFusion 模块改进跨层 token 融合网络,优化多尺度特征融合策略,增强模型对不同尺寸病灶的检测能力。此外,使用 Inner-GIoU 损失函数进一

步优化小目标病灶的定位精度,并加快模型收敛速度。实验结果表明,改进算法在公开数据集 DeepLesion 和外部验证集上的平均检测灵敏度分别达到了 83.26% 和 82.33%。综上,本文算法有望满足临幊上多部位小目标病灶识别的需求,为医生提供更可靠的辅助诊断支持。未来,将持续优化模型架构,从不同的角度提升病灶检测性能。

参考文献

- [1] 马金林,欧阳轲,马自萍,等. 多尺度自适应融合的肝脏肿瘤检测[J]. 中国图象图形学报, 2023, 28(1): 260-276.
MA J L, OUYANG K, MA Z P, et al. Liver tumor detection based on multi-scale adaptive fusion [J]. Journal of Image and Graphics, 2023, 28 (1): 260-276.
- [2] 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. A lung nodule malignancy diagnosis method based on expert knowledge fusion in CT images[J]. Foreign Electronic Measurement Technology, 2023, 42(7): 181-187.
- [3] 廉小亲,王梓桐,高超,等. 基于 MSAPNet 的青少年精神分裂症脑电识别研究[J]. 电子测量技术, 2024, 47(16): 155-164.
LIAN X Q, WANG Z T, GAO CH, et al. Research on EEG recognition of adolescent schizophrenia based on MSAPNet[J]. Electronic Measurement Technology, 2024, 47(16): 155-164.
- [4] YU Q, WANG Y, YAN K, et al. Effective lymph nodes detection in CT scans using location debiased

- query selection and contrastive query representation in transformer [C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 180-198.
- [5] YAN K, WANG X S, LU L, et al. DeepLesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning [J]. Journal of Medical Imaging, 2018, 5(3): 036501.
- [6] LIU ZH, XIE X, SONG Y Q, et al. MLANet: Multi-layer anchor-free network for generic lesion detection [J]. Engineering Applications of Artificial Intelligence, 2021, 102: 104255.
- [7] LIU ZH, HAN K, XUE K F, et al. Improving CT-image universal lesion detection with comprehensive data and feature enhancements [J]. Multimedia Systems, 2022, 28(5): 1741-1752.
- [8] 许金亚, 裴颂文. 面向 CT 图像小目标病灶区域的智能化检测算法研究[J]. 小型微型计算机系统, 2023, 44(8): 1792-1797.
- XU J Y, PEI S W. Research on intelligent detection algorithm for small lesion regions in CT images [J]. Mini-micro Computer Systems, 2023, 44 (8): 1792-1797.
- [9] ZHU Y, LIU ZH, SONG Y Q, et al. CPSNet: A cyclic pyramid-based small lesion detection network [J]. Multimedia Tools and Applications, 2023, 83 (13): 39983-40001.
- [10] GUAN Q, PAN M J, CHEN F, et al. IAFI-FCOS: Intra- and across-layer feature interaction FCOS model for lesion detection of CT images [C]. 2024 International Joint Conference on Neural Networks (IJCNN). IEEE, 2024: 1-8.
- [11] JIAN M W, JIN Y, WANG R, et al. Inter-slice correlation weighted fusion for universal lesion detection [C]. 2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). IEEE, 2023: 636-643.
- ZHANG H, CHUNG S C A. Depth-aware networks for multi-organ lesion detection in chest CT scans [J]. Bioengineering, 2024, 11(10): 998-998.
- [13] XU ZH H, LI T R, LIU Y X, et al. PAC-Net: Multi-pathway FPN with position attention guided connections and vertex distance IoU for 3D medical image detection [J]. Frontiers in Bioengineering and Biotechnology, 2023, 11: 1049555.
- [14] 黎生丹, 柏正尧. 深度学习多部位病灶检测与分割[J]. 中国图象图形学报, 2021, 26(11): 2723-2731.
- LI SH D, BAI ZH Y. Multi-organ lesion detection and segmentation based on deep learning [J]. Journal of Image and Graphics, 2021, 26(11): 2723-2731.
- [15] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.
- [16] HOU X Q, LIU M Q, ZHANG S L, et al. Salience DETR: Enhancing detection transformer with hierarchical salience filtering refinement [C]. IEEE/ CVF Conference on Computer Vision and Pattern Recognition, 2024: 17574-17583.
- [17] OUYANG D L, SU H, ZHANG G ZH, et al. Efficient multi-scale attention module with cross-spatial learning [C]. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). IEEE, 2023: 1-5.
- [18] LI J, WEN Y, HE L. SCCConv: Spatial and channelreconstruction convolution for feature redundancy[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6153-6162.
- [19] YANG M K, YU K, ZHANG CH, et al. DenseASPP for semantic segmentation in street scenes[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 3684-3692.
- [20] ZHANG H, XU C, ZHANG SH J. Inner-IoU: More effective intersection over union loss with auxiliary bounding box [J]. ArXiv preprint arXiv: 2311. 02877, 2023.
- [21] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.
- [22] CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [23] TIAN ZH, SHEN CH H, CHEN H, et al. FCOS: A simple and strong anchor-free object detector [J]. IEEE Transactions on Pattern Analysisand Machine Intelligence, 2022, 44(4): 1922-1933.
- [24] ZHANG H, LI F, LIU SH L, et al. DINO: DETR with improved denoising anchor boxes for end-to-end object detection [J]. ArXiv preprint arXiv: 2203.03605, 2022.
- [25] HOU X Q, LIU M Q, ZHANG S L, et al. Relation DETR: Exploring explicit position relation prior for object detection [C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 89-105.
- [26] NAN Z X, LI X H, DAI J F, et al. MI-DETR: An object detection model with multi-time inquiries mechanism [J]. ArXiv preprint arXiv:2503.01463, 2025.

作者简介

张茜,博士,副教授,主要研究方向为医学影像处理。

E-mail:6382@zut.edu.cn

武雨露,硕士研究生,主要研究方向为医学影像处理。

E-mail:2023007305@zut.edu.cn

郑冰洁,博士,副主任医师,主要研究方向为医学影像诊断。

E-mail:zlyyzbj2100@zzu.edu.cn

董洁,博士,讲师,主要研究方向为医学影像处理和人工智能。

E-mail:dongjie@ncwu.edu.cn

杨关(通信作者),博士,副教授,主要研究方为机器学习、计算机视觉。

E-mail:yangguan@zut.edu.cn