

张雨露, 张家琳. 基于 YOLOv5s 改进的肺结节检测[J]. 智能计算机与应用, 2025, 15(12): 55-61. DOI: 10.20169/j.issn.2095-2163.25051901

基于 YOLOv5s 改进的肺结节检测

张雨露, 张家琳

(河北地质大学 信息工程学院, 石家庄 052161)

摘要: 针对肺部 CT 图像中微小结节检测精度低的问题, 本文提出一种基于 YOLOv5s 的改进模型。首先, 在主干网络中融合坐标注意力 (Coordinate Attention, CA) 和高效通道注意力 (Efficient Channel Attention, ECA) 模块, 强化关键特征提取能力; 其次, 用 C3GC 模块替换原始 C3 模块, 通过全局上下文信息建模长程依赖关系; 最后, 在损失函数中 CIoU 结合归一化瓦瑟斯坦距离 (Normalized Wasserstein Distance, NWD), 优化小目标检测效果。在 LUNA16 数据集上的实验结果表明, 改进后的模型检测精确度达到 0.898, mAP@0.5 为 0.903, 相比 YOLOv5s 分别提升 6.2% 和 4.7%。

关键词: 肺结节检测; YOLOv5s; 归一化瓦瑟斯坦距离

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 2095-2163(2025)12-0055-07

Lung nodule detection based on improved YOLOv5s

ZHANG Yulu, ZHANG Jialin

(School of Information Engineering, Hebei GEO University, Shijiazhuang 052161, China)

Abstract: To address the issue of low detection accuracy for tiny nodules in lung CT images, an improved algorithm based on YOLOv5s is proposed. First, the Coordinate Attention (CA) and Efficient Channel Attention (ECA) modules are integrated into the backbone network to enhance the extraction of critical features. Second, the original C3 module is replaced with the C3GC module, which models long-range dependencies by capturing global contextual information. Finally, the loss function combines CIoU with Normalized Wasserstein Distance (NWD) to optimize the detection performance for small targets. Experimental results on the LUNA16 dataset demonstrate that the improved model achieves a detection precision of 0.898 and mAP@0.5 of 0.903, representing improvements of 6.2% and 4.7%, respectively, compared to the original YOLOv5s.

Key words: lung nodule detection; YOLOv5s; Normalized Wasserstein Distance

0 引言

根据国际癌症研究机构 (International Agency for Research on Cancer, IARC) 发布的 2022 年全球癌症统计数据, 肺癌也是发病率最高的癌种, 占全球癌症的八分之一^[1]。肺结节是肺癌早期的重要标志, 对 CT 图像中的肺结节检测有助于实现肺癌的早发现、早预防。

传统的肺结节检测主要通过手工设计特征和传统的机器学习算法进行肺结节检测。首先, 从 CT 图像中提取肺结节的形状和纹理等特征; 其次, 使用分类器对提取的特征进行分类, 判断是否存在肺结

节。这种方法虽然在一定程度上辅助诊断, 但主要依赖手工特征提取, 难以达到理想的检测效果。深度学习通过大量数据训练网络自动学习特征, 以适应数据差异性, 从而实现肺结节的精确检测。

区域卷积神经网络 (Regions with Convolutional Neural Network, R-CNN) 和快速卷积神经网络 (Faster Region-based Convolutional Neural Network, Faster R-CNN) 等都属于两阶段的目标检测算法模型。Su Y 等^[2]提出了一种基于 Faster R-CNN 框架的肺结节检测方法, 通过优化区域提议特征金字塔网络 (Feature Pyramid Network, FPN) 和分类模块, 提升了检测精度; Guo N 等^[3]提出了一种多尺度肺

基金项目: 河北地质大学 2025 年度研究生课程建设项目 (YALK2025006)。

作者简介: 张雨露 (2000—), 女, 硕士研究生, 主要研究方向: 目标检测。Email: 364485763@qq.com; 张家琳 (2000—), 男, 硕士研究生, 主要研究方向: 人工智能。

收稿日期: 2025-05-19

结节检测模型,通过融合级联区域卷积神经网络(Cascade Regions with Convolutional Neural Network, Cascade R-CNN)和特征金字塔网络(FPN),增强了模型对不同大小结节的识别能力;高璇等^[4]提出了一种改进的Faster R-CNN方法,通过融合挤压与激励网络(Squeeze-and-Excitation Network, SENet)通道注意力机制和二阶响应变换(Second-Order Response Transformation, SORT)模块,并优化锚框尺寸,显著提高了低剂量CT图像中肺结节的检测精度,平均精度达到89.83%。YOLO(You Only Look Once)模型采用单阶段检测方法^[5]。Facebook团队^[6]提出基于Transformer架构的目标检测器DETR(Detection Transformer);Zhao等^[7]在DETR的基础上提出RT-DETR(Real Time DetectionTransformer),采用单阶段检测方法。王乾梁等^[8]以YOLOv3作为基础网络结构,引入空洞卷积的感受野模块层修改损失函数,在LIDC-IDRI数据集上得到了88.89%的准确率;陈茂龙等^[9]在YOLOv4的基础上更换损失函数,引入卷积注意力模块,以提升算法的检测精度;张琼等^[10]提出了一种高效注意力特征金字塔网络,结合深度可分离卷积和注意力机制用于肺结节检

测;李益智等^[11]提出了一种基于快速傅里叶变换和3D卷积神经网络的肺结节检测模型,在Luna16数据集上显著提升模型性能;郭相均等^[12]改进了YOLOv8s算法,用于肺结节检测,通过引入空间深度转换卷积、坐标注意力机制和优化损失函数,在LUNA16数据集上实现了77.8%的检测精度。

本文基于YOLOv5s模型进行改进,通过引入坐标注意力机制和通道注意力机制,用C3GC模块替换原始的C3模块并优化损失函数,在保持实时性的同时提高检测精确度。

1 YOLOv5s 基本原理与结构

YOLOv5s是一种高效的单阶段目标检测模型,具有高精度和推理速度快的特点,其模型结构包含3部分:骨干网络(Backbone)通过卷积层、结合残差连接和分组卷积的C3模块和快速空间金字塔池化模块(SPPF)提取特征;多尺度特征融合网络(Neck)采用FPN+PAN结构实现多尺度特征融合;检测头(head)使用锚框机制在3个尺度上进行预测,并采用CIoU损失函数和非极大值抑制优化检测结果,模型结构紧凑,易于部署。YOLOv5s模型结构如图1所示。

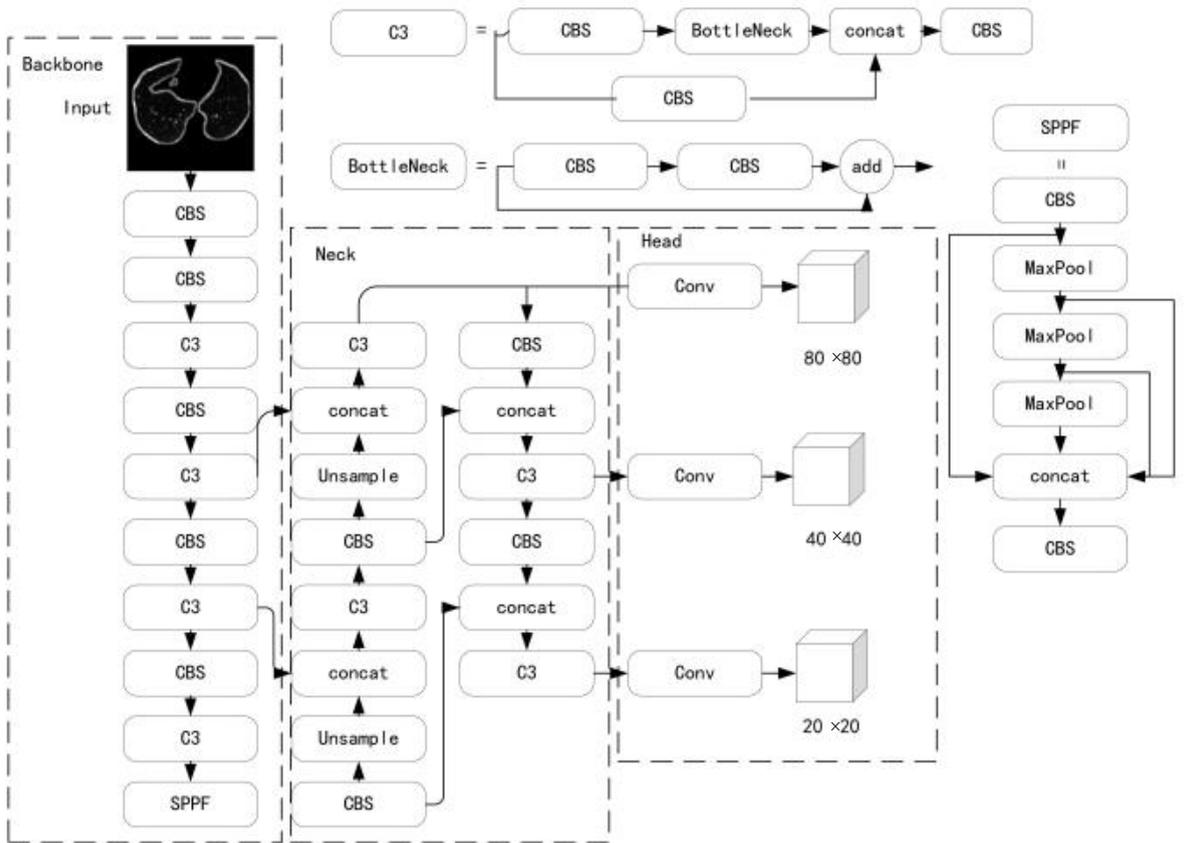


图 1 YOLOv5s 模型结构

Fig. 1 YOLOv5s model structure

2 改进的 YOLOv5s 肺结节检测算法

本文提出了一种基于 YOLOv5s 改进算法, 显著提升肺结节检测性能。首先, 在主干网络中引入 CA (Coordinate Attention) 和 ECA (Efficient Channel Attention) 模块, 通过将空间位置信息编码为注意力权重, 使网络能够更准确地定位关键区域, 提升特征选择的精度; 其次, 在特征提取方面, 将原始的 C3

模块替换为 C3GC 模块, 通过整合全局上下文信息, 增强模型对复杂背景和小尺寸目标的特征提取能力; 最后, 在损失函数设计上, 保留 CIoU 损失函数对边界框位置和尺寸回归优势的基础引入归一化瓦瑟斯坦距离 (NWD) 度量, 平衡边界框预测的敏感度和准确性, 提升小目标的检测性能。改进后的 YOLOv5s_EAGN 模型结构如图 2 所示。

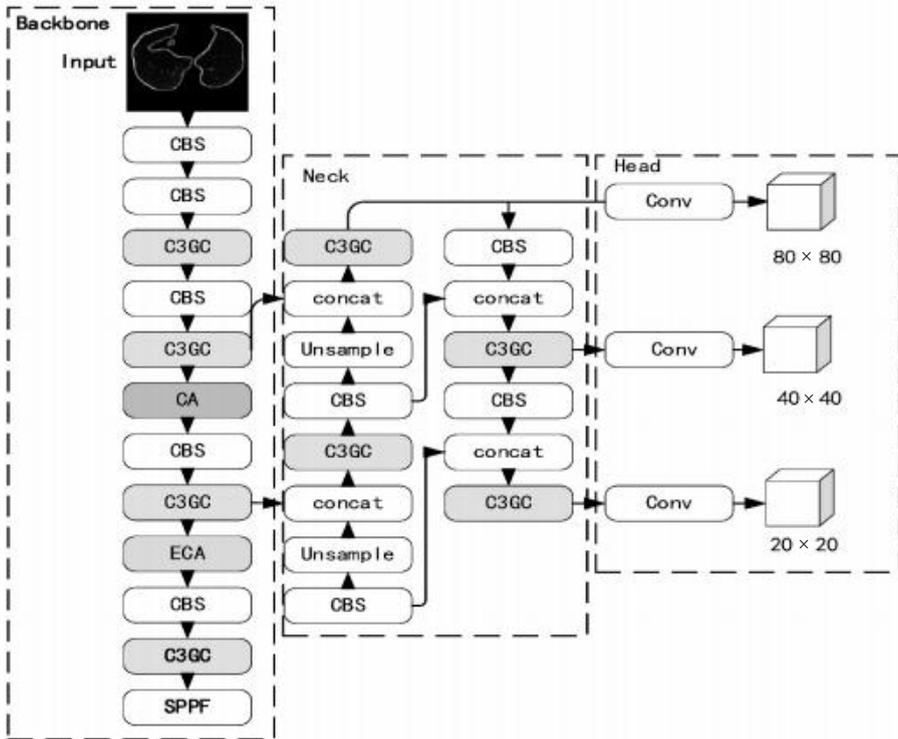


图 2 YOLOv5s_EAGN 模型结构

Fig. 2 YOLOv5s_EAGN model structure

2.1 坐标注意力模块

通道注意力会给模型带来比较显著的性能提升, SE (Squeeze - and - Excitation)、BAM (Bottleneck Attention Module) 和 CBAM (Convolutional Block Attention Module) 在提升模型性能方面取得了显著成果, 但是也存在一些局限性^[13-15]。SE 注意力只考虑了通道编码的信息, 忽略了位置信息的重要性,

BAM 和 CBAM 通过减少通道维度并使用卷积计算空间注意力来利用位置信息, 无法对视觉任务所必需的长期依赖关系建模。CA 通过将位置信息嵌入通道注意力中, 使模型能够在较大范围内参与并增强特征表示, 同时保留精确的位置信息, 使模型更加精准地定位和识别感兴趣的目标^[16]。CA 模块结构如图 3 所示。

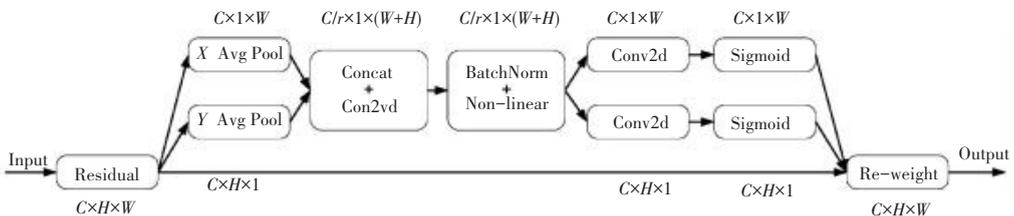


图 3 CA 模块结构图

Fig. 3 Structure of CA module

X Avg Pool 表示在水平方向使用 $1 \times W$ 的核进行全局平均池化, Y Avg Pool 表示在垂直方向使用 $H \times 1$ 的核进行全局平均池化。

2.2 高效通道注意力模块

通道注意力在提升深度卷积神经网络性能方面展现了巨大潜力,但大多数研究都侧重于设计结构更复杂的注意力模块以提升性能,不可避免地增加了模型的复杂度。通过对 SENet 通道注意力模块的深入剖析,发现避免降维并保持适当的跨通道交互,对降低模型复杂度的同时维持性能至关重要,这一观点在 ECA 模块中得到了很好的体现,该模块采用一维卷积实现了一种无需降维的局部跨信道交互策略^[17]。ECA 模块结构如图 4 所示。

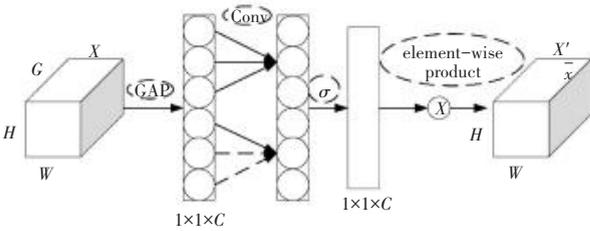


图 4 ECA 模块结构图

Fig. 4 Structure of ECA module

首先,对输入的三维特征图 x 进行全局平均池化(Global Average Pooling, GAP),将每个通道上的空间信息压缩成一个标量值,得到一个大小是 $1 \times 1 \times C$ 的向量;其次,对得到的向量进行 1×1 卷积操作,学习通道之间的相互依赖关系,使用 Sigmoid 函数作为激活函数生成一个范围在 $[0, 1]$ 之间的权重向量;最后,得到的权重向量和原始输入的 x 进行元素级乘法,输出与 x 具有相同尺寸的 x' 。

2.3 全局上下文注意力模块

GC(Global Context)模块是 GCNet 的一个核心组件^[18]。在传统卷积神经网络中,卷积操作通常只能在局部领域中建立像素关系,难以建模长距离依赖关系,尽管可以通过深度堆叠卷积层或者使用池化层来扩大感受野,但计算效率低下并且难以优化。GC 模块针对在捕捉长距离依赖关系方面的局限性,通过简化非局部块和结合 SE(Squeeze-Excitation)在特征转换步骤中的优点,更为有效地捕获通道间的依赖性,以较低的计算代价提升模型的性能。GC 模块与 C3GC 模块结构如图 5 所示。

C3GC 模块在 C3 模块的基础上引入 GC 机制,增强 YOLOv5 的特征提取能力。C3GC 采用双分支结构,一个分支经过 Conv 层和 Bottleneck 模块,用于对局部特征提取;另一个分支通过 GC 模块和

Conv 层,用于全局上下文特征提取。两个分支的输出特征进行拼接,随后通过 1×1 卷积进行特征融合,最后输出强化后的特征图。

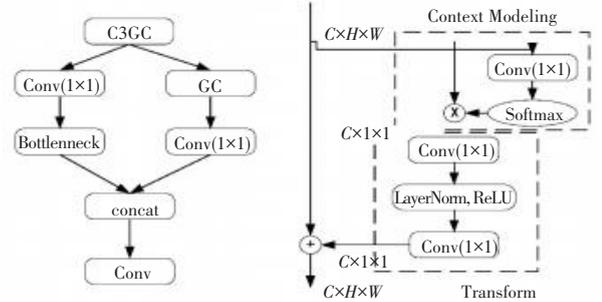


图 5 GC 模块与 C3GC 模块结构

Fig. 5 Structure diagrams of the GC module and the C3GC module

2.4 归一化瓦瑟斯坦距离

传统的 IoU 及其扩展在微小物体检测中存在局限性,尤其是对位置偏差过于敏感。微小物体的边界框通常只有几个像素大小,轻微的偏差就会导致 IoU 大幅下降,从而严重影响模型训练效果和最终性能。归一化瓦瑟斯坦距离(Normalized Wasserstein Distance, NWD)将边界框建模为二维高斯分布,并采用 Wasserstein 距离衡量其相似性,提供了一种对尺度不敏感、对位置偏差更鲁棒的相似性度量方法^[19]。利用最优传输理论中 Wasserstein 距离来计算分布距离。对于两个二维高斯分布 $\mu_1 = N(m_1, \Sigma_1)$ 和 $\mu_2 = N(m_2, \Sigma_2)$, μ_1 和 μ_2 之间的二阶 Wasserstein 距离定义如下:

$$W_2^2(\mu_1, \mu_2) = \|m_1 - m_2\|_2^2 + \|\Sigma_1^{1/2} - \Sigma_2^{1/2}\|_F^2 \quad (1)$$

其中, $\|\cdot\|_F$ 是 Frobenius 范数。

此外,分别根据边界框 a 和 b 建模得到 N_a 和 N_b , 公式(1)可以进一步简化为:

$$W_2^2(N_a, N_b) = \left\| \begin{bmatrix} c_{xa} & c_{ya} \\ c_{xb} & c_{yb} \end{bmatrix} - \begin{bmatrix} w_a^2 h_a^2 & w_b^2 h_b^2 \\ w_a^2 h_a^2 & w_b^2 h_b^2 \end{bmatrix} \right\|_2^2 \quad (2)$$

由于 $W_2^2(N_a, N_b)$ 是一个距离度量,不能直接作为相似度量,因此对其进行了指数形式的归一化处理,得到了新的度量 NWD:

$$NWD(N_a, N_b) = \exp\left(-\sqrt{\frac{W_2^2(N_a, N_b)}{C}}\right) \quad (3)$$

其中, C 是一个与数据集密切相关的常数。

基于 NWD 定义边界框回归的损失函数:

$$L_{NWD} = 1 - NWD(N_p, N_g) \quad (4)$$

其中, N_p 是预测框的高斯分布模型, N_g 是真实框的高斯分布模型。

3 实验与结果分析

3.1 数据集

本文实验中采用 LUNA16 数据集。LUNA16 数据集是对业界广泛使用的公共肺结节数据集 LIDC-IDRI 予以清洗和标注后得到的^[20]。LUNA16 数据集共包含 888 位患者肺部 CT 数据, 以及由 4 名影像学专家手工标记的 1 186 个结节标签。采用 K-means 聚类算法结合形态学处理方法对肺部 CT 图像进行分割, 提取肺实质区域, 去除图像中的噪声和无关信息。基于每个结节的注释, 提取结节所在的中心切片及其上下各一个切片。数据集共 3 558 张图片, 每张图片的大小为 640×640, 按 7:2:1 的比例分为训练集、验证集和测试集。

3.2 实验环境

实验在 Windows11 操作系统上进行, 硬件配置包含 Intel(R) Core(TM) i9-12900H 2.50 GHz CPU 和 NVIDIA GeForce RTX 3060 GPU。网络模型基于 PyTorch 深度学习框架构建, 开发环境为 PyTorch 2.5.1、CUDA 12.6 和 Python 3.9。初始学习率设为 0.01, 批量大小设为 16, 训练轮次为 200。

3.3 评价指标

本文采用精确度 (Precision, P), 召回率 (Recall, R) 和平均精度均值 (mean Average Precision, mAP) 作为评估算法检测性能的指标。

精确度是在识别出的物体中正确的正向预测 (True Positive, TP) 所占的比率, 计算公式如下式:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

其中, TP 为正确检测的目标数量, FP 为误检的背景或错误目标数量。

平均精度均值是对所有类别的单类平均准确率 (Average-Precision, AP) 取平均值, 得到全局性能指标, 计算公式如下式:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (6)$$

其中, N 表示类别总数, AP_i 表示第 i 个类别的平均精度。

召回率是正确识别出的物体占总物体的比率, 计算公式如下式:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

其中, FN 为被模型漏检的真实目标数量。

3.4 消融实验

为了验证改进点对模型性能的影响, 通过在 YOLOv5s 基础上逐一添加模块进行消融实验, 实验结果见表 1。基于 YOLOv5s 训练 200 个周期后精确率、召回率和 mAP@0.5 分别取得 83.6%、81.2% 和 85.6%; 引入 CA 模块后, 模型在精确率方面表现出明显提升, 但召回率出现一定程度下降; 引入 C3GC 模块则有效提升了模型的综合检测性能, 其 mAP@0.5 值较基准模型提高 3%; 最终, 通过整合 4 项改进, YOLOv5s_EAGN 相较于 YOLOv5s 在精确率、召回率和 mAP@0.5 分别提升了 6.2%、1.3% 和 4.6%。上述结果表明, 各模块有效提升了模型的检测精度。

表 1 消融实验结果

Table 1 Results of ablation experiments

模型	Epoch	$P/\%$	$R/\%$	mAP@0.5/%
YOLOv5s	200	83.6	81.2	85.6
YOLOv5s+CA	200	86.1	79.4	85.6
YOLOv5s+ECA	200	85.9	79.3	85.4
YOLOv5s+C3GC	200	84.3	82.1	88.6
YOLOv5s+NWD	200	85.5	82.4	86.5
YOLOv5s+NWD+ECA	200	84.7	81.7	86.4
YOLOv5s+C3GC+CA	200	85.8	81.2	87.2
YOLOv5s_EAGN	200	89.8	82.5	90.3

3.5 与其他主流模型对比实验

为了验证 YOLOv5s_EAGN 模型的改进效果, 选取 SSD (Single Shot MultiBox Detector)、Faster-RCNN、YOLO 系列在内的主流检测模型进行对比实验, 实验结果见表 2。从表 2 的实验数据可以看出, 与基线模型 YOLOv5s 相比 YOLOv5s_EAGN 的精确率和 mAP@0.5 分别提升了 6.2% 和 4.6%; 与 YOLOv8s 相比, YOLOv5s_EAGN 的 mAP@0.5 提高了 3.1%, 精确率和召回率分别提升了 2.3% 和 3.2%。

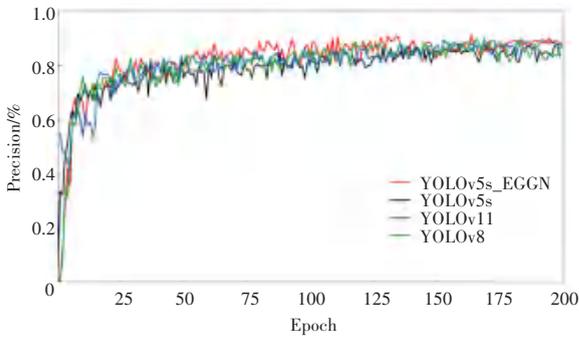
表 2 与主流模型对比实验结果

Table 2 Comparative experimental results with mainstream models

模型	Epoch	$P/\%$	$R/\%$	mAP@0.5/%
SSD	200	83.6	79.8	85.4
Faster-RCNN	200	84.7	82.2	80.5
YOLOv5s	200	83.6	81.2	85.6
YOLOv8s	200	87.5	79.3	87.2
YOLOv9t	200	85.4	81.6	88.4
YOLOv9s	200	84.8	83.9	87.7
YOLOv10s	200	87.9	82.4	87.8
YOLO11n	200	82.9	81.5	83.7
YOLOv5s_EAGN	200	89.8	82.5	90.3

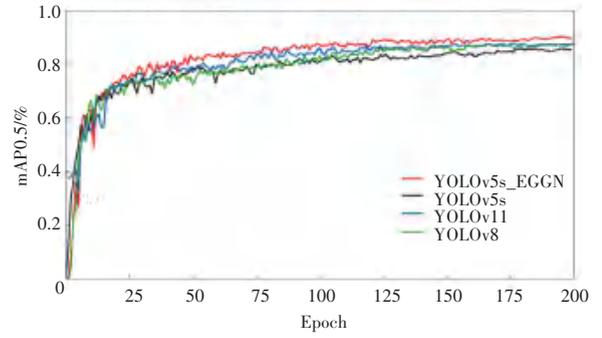
3.6 可视化对比

YOLOv5s_EAGN 和其他主流检测模型之间的精确度和 $mAP@0.5$ 比较如图 6 所示,可以看出



(a) 精确度性能对比

YOLOv5s_EAGN 在检测精度方面相比基线模型有明显的提升。



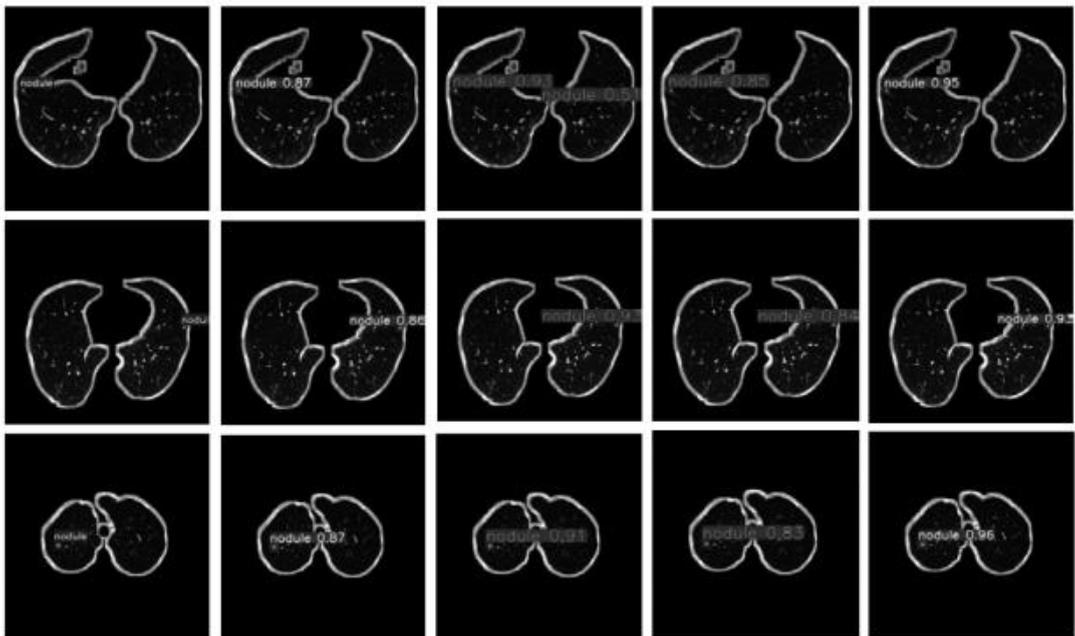
(b) mAP0.5 性能对比

图 6 模型性能对比曲线

Fig. 6 Comparative analysis of model performance metrics

为了验证 YOLOv5s_EAGN 对肺结节的检测性能,将其与 YOLOv5s、YOLOv10s 和 YOLO11n 在验证集上进行测试,得到的结果如图 7 所示。Nodule

表示结节,框上的数字表示类别标签的置信度,可以看出 YOLOv5s_EAGN 检测出的结节具有更高的置信度。



(a) 标签真实性

(b) YOLOv5s

(c) YOLOv10s

(d) YOLO11n

(e) YOLOv5s_EAGN

图 7 检测结果对比

Fig. 7 Comparison chart of test results

4 结束语

本文针对肺部 CT 图像中微小结节检测精度低的问题,提出一种基于 YOLOv5s 的改进模型。首先,结合肺结节检测的特点,引入了 CA 模块和 ECA 模块,以保留精确的位置信息,并提升网络对关键特征的专注能力;其次,用 C3GC 模块替换原有的 C3

模块,通过 GC 更有效地捕获通道间的依赖关系,从而提升特征提取能力;最后,针对小结节检测,引入了归一化瓦瑟斯坦距离作为损失函数的一部分,以增强模型在边界框预测中对位置敏感性与准确性之间的平衡。实验结果表明,本文提出的算法在肺结节检测中有较高的精度,能够准确识别肺结节。

参考文献

- [1] BRAY F, LAVERSANNE M, SUNG H, et al. Global cancer statistics 2022; GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2023, 73(6): 176-200.
- [2] SU Y, LI D, CHEN X. Lung nodule detection based on Faster R-CNN framework [J]. Computer Methods and Programs in Biomedicine, 2020, 200(1): 105866.
- [3] GUO N, BAI Z. Multi-scale pulmonary nodule detection by fusion of cascade R-CNN and FPN[C]//Proceedings of 2021 International Conference on Computer Communication and Artificial Intelligence (CCAI). 2021:15-19.
- [4] 高璇, 毕晓君. 基于改进 Faster R-CNN 的低剂量 CT 图像肺结节检测[J]. 计算机应用研究, 2020, 37(S2): 404-406.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016:779-788.
- [6] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]// Proceedings of European Conference on Computer Vision. Cham: Springer, 2020; 213-229.
- [7] ZHAO Y, LV W, XU S, et al. Detsr beat yolos on real-time object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2024; 16965-16974.
- [8] 王乾梁, 石宏理. 基于改进 YOLOv3 的肺结节检测方法[J]. 中国医学物理学杂志, 2021, 38(9): 1179-1184.
- [9] 陈茂龙, 李亚洲, 云家伟, 等. 基于 KCCS-YOLOv4 的肺结节检测算法[J]. 长春工业大学学报, 2023, 44(5): 424-433.
- [10] 张琼, 杭益柳, 邱建林, 等. 高效注意力金字塔网络在肺结节检测的应用[J]. 中国医学物理学杂志, 2024, 41(11): 1361-1369.
- [11] 李益智, 张佳, 林晓丹. 快速傅里叶变换引入小样本肺结节检测模型[J]. 癌症, 2025, 44(1): 47-56.
- [12] 郭相均, 蒋朝根. 基于改进 YOLOv8s 的肺结节检测算法[J]. 现代信息科技, 2025, 9(7): 87-92.
- [13] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [14] PARK J, WOO S, LEE J Y, et al. BAM: Bottleneck attention module[C]//Proceedings of British Machine Vision Conference. Cham: Springer, 2018:30-38.
- [15] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 2516-2529.
- [16] HOU Q, ZHANG L, CHENG M M, et al. Coordinate attention for efficient mobile network design [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021; 13708-13717.
- [17] 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. Piscataway, NJ: IEEE, 2020; 11531-11539.
- [18] CAO Y, XU J, LIN S, et al. GCNet: Non-local networks meet squeeze-excitation networks and beyond[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop. Piscataway, NJ: IEEE, 2019; 1971-1980.
- [19] WANG J, CHEN K, YANG S, et al. A normalized Gaussian Wasserstein distance for tiny object detection[EB/OL]. arXiv, 2110.13389, 2021.
- [20] SETIO 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.