

DOI:10.3969/j.issn.1005-202X.2025.01.007

医学影像物理

改进YOLOv5的肺结节检测算法

田吉, 杨萍, 刘佳, 王金华

北京联合大学智慧城市学院, 北京 100101

【摘要】针对肺部CT图像中大量小结节难以检测、以及现有肺结节检测算法难以实现轻量化和高精度兼顾的问题,提出改进YOLOv5的高精度轻量化肺结节检测算法。主要改进以下4个方面:(1)使用空间-深度下采样操作替换YOLOv5主干网络中步长为2的下采样操作,使细微特征提取更完整以便于发现微小结节;(2)在YOLOv5颈部使用渐进融合特征策略,构建不同路径特征图之间的联系以增强各个层级之间信息的交互;(3)创造性地提出了感知全局上下文注意力并将其应用在YOLOv5颈部网络的末端,提高模型从全局视角对肺结节关键特征和语义信息的理解能力;(4)采用损失排序挖掘方法重点训练困难样本,以此来强化模型的鉴别能力。改进后的算法在LUNA16数据集上得到了96.0%的精确度,95.0%的召回率和97.3%的平均精度,相比原始YOLOv5模型,精确度提高了14%,召回率提高了10.2%,平均精度提高了12.1%,上述结果表明,改进后的算法可有效检测出肺结节。

【关键词】YOLOv5;肺结节检测;下采样算法;注意力机制;困难样本

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2025)01-0043-09

Lung nodule detection algorithm based on improved YOLOv5

TIAN Ji, YANG Ping, LIU Jia, WANG Jinhua

College of Smart City, Beijing Union University, Beijing 100101, China

Abstract: To address the challenges of detecting small nodules in lung CT images and achieving a balance between lightweight and high-precision with the existing lung nodule detection algorithms, a high-precision and lightweight lung nodule detection algorithm based on improved YOLOv5 is proposed. The main improvements are focused on 4 aspects. (1) Replacing the stride-2 downsampling operation in the YOLOv5 backbone with space-to-depth downsampling operations to enhance fine feature extraction for detecting small nodules more comprehensively. (2) Employing an asymptotic feature pyramid network in the YOLOv5 neck to establish connections among feature maps from different paths, thereby enhancing interaction among different hierarchical levels. (3) Introducing global context-aware attention to the end of YOLOv5 neck network for improving the model's ability to understand key features and semantic information of lung nodules from a global perspective. (4) Utilizing the loss rank mining approach to strategically train on hard samples, thereby strengthening the model's discrimination ability. The improved algorithm achieves 96.0% precision, 95.0% recall rate and 97.3% average precision on the LUNA16 dataset, which are 14.0%, 10.2% and 12.1% higher than the original YOLOv5 model, demonstrating its effectiveness for lung nodule detection.

Keywords: YOLOv5; lung nodule detection; downsampling algorithm; attention mechanism; hard sample

前言

根据世界卫生组织国际癌症研究机构

(International Agency for Research on Cancer, IARC)发布的2022年全球癌症统计最新数据显示^[1],2022年肺癌导致的死亡人数高达181万例,占有癌症死亡病例的18.7%,是威胁人类健康的第一癌症杀手。研究显示,早期肺癌术后的10年生存率可达92%^[2]。肺癌早期主要表现为肺结节,因此早期筛查肺结节可有效降低死亡率。胸部CT扫描能够提供高分辨率的肺部图像,一位患者的胸部CT扫描图片通常在300张左右,医生需要仔细阅读每一张CT图像以甄别出结节,分析出结节的性质^[3]。对医生而言,需要

【收稿日期】2024-09-10

【基金项目】国家自然科学基金(62172045,62272049)

【作者简介】田吉,硕士研究生,研究方向:医学图像处理,E-mail: tj896979@163.com

【通信作者】杨萍,博士,副教授,研究方向:信号与信息处理、医学图像处理,E-mail: xxyangping@buu.edu.cn

较强的专业能力,阅片强度高,工作量大,很容易误诊,因此利用人工智能辅助医生进行临床诊断具有重要意义。

在深度学习得到充分发展之前,研究人员手动提取结节特征,再进行分类预测。这种传统的肺结节检测方法比较依赖研究人员的主观判断,需要复杂的调参过程,在不同场景中的适应性差,自动化程度较低,难以取得良好的检测效果。深度学习方法通过训练大量的样本来自动学习特征,不需要手动设计特征,从而减少了人为因素的干扰,并且可以更好地适应不同的数据分布和场景变化,因此取代了传统的肺结节检测方法。基于深度学习的肺结节检测算法主要分为双阶段目标检测算法和单阶段目标检测算法。对于双阶段目标检测算法,Su等^[3]以R-CNN为基础,从参数和模型结构两个方面进行改进,然后对几种常用的目标检测方法进行比较,并通过实验验证了网络的有效性;Ding等^[4]将反卷积结构引入Faster R-CNN中用于候选结节检测,并设计了用分类器来降低假阳率的双阶段网络。Mai等^[5]在Faster R-CNN的基础上引入多头检测器和跳跃连接来捕获多尺度特征,并设计了一个空间注意力模块,最后设计了一个降低假阳率模块。Guo等^[6]受Faster R-CNN的启发,提出融合Cascade R-CNN^[7]和特征金字塔的多尺度肺结节检测模型,可以较为有效地检测出肺结节。双阶段目标检测算法首先经过候选区域网络找出可能包含目标的区域,然后在此基础上进行分类和边界框回归,这种方法模型复杂、参数量大,在实际应用中不利于落地部署。而单阶段目标检测算法不经过区域候选网络,相比双阶段算法速度更快并且更轻量化。YOLO(You Only Look Once)系列算法是单阶段实时目标检测算法的代表,广大学者基于YOLO系列进行了大量研究和改进。王乾梁等^[8]调整了YOLOv3中骨干网络和检测网络的结构单元数量、更换了激活函数、引入含有空洞卷积的感受野模块层、最后修改损失函数,在LIDC-IDRI数据集上取得了94.73%的敏感度。Wu等^[9]在YOLOv7的基础上添加了小目标检测层、多尺度感受野模块以及高效全方位卷积,提高了YOLOv7对肺结节的检测能力。这些方法对肺结节有较好的检测效果,但模型参数量大。

YOLOv5结合了较高精度、快速推理、轻量化和易部署等多方面的优势,在目标检测领域有广泛的应用,本文将YOLOv5用于肺结节检测,并在此基础上进行改进,使模型同时具备轻量化和高精度的优势。

1 YOLOv5目标检测算法

YOLOv5是YOLO系列的代表算法,主要由输入端、主干网络、颈部网络和头部网络构成。输入端采用了Mosaic方法对输入数据进行数据增广操作,同时加入自适应锚框计算操作以便自适应地计算出不同数据集中的最佳锚框值。主干网络通过一系列的卷积层和激活函数,将输入的图像逐步转化为由低到高不同层级的特征图,其中低层级特征图含有较多的细节信息,高层级特征图含有较多的语义信息。颈部网络采用“特征金字塔网络(Feature Pyramid Network, FPN)+路径聚合网络(Path Aggregation Network, PAN)”结构融合主干网络提取的不同层级特征^[10]。通过颈部网络将不同层级的特征进行融合后送到头部网络进行检测,得到目标的大小和位置。然而,针对肺结节检测问题,YOLOv5存在非对称下采样操作中容易丢失细微信息、颈部网络融合特征不充分等问题,导致小结节易漏检,对此,本文改进了YOLOv5的下采样操作,颈部网络采用了渐进特征融合策略,提出统领全局上下文的注意力模块,使用重点训练困难样本的策略,以提升YOLOv5对小肺结节的检测能力。

2 改进YOLOv5肺结节检测算法

本文提出的改进YOLOv5肺结节检测算法,主要分为以下几个方面:在主干网络部分,原始YOLOv5的下采样操作使用的是步长为2的卷积,这种操作是非对称采样,导致模型对细粒度信息提取不完整,用于肺结节检测时,容易漏检小结节,本文将其替换为空间-深度(Space-to-Depth, SPD)下采样操作^[11],实现对各像素点的均匀采样,从而保证了小结节信息的完整性。在颈部网络,YOLOv5采用FPN+PAN的多尺度特征融合策略,这种策略只是将相邻特征层简单拼接,导致非相邻特征层的融合不足,从而使不同尺度信息融合不充分,对此,本文引入渐进特征金字塔网络(Asymptotic Feature Pyramid Network, AFPN)^[12],以提升不同层级特征的融合,从而提高肺结节检测的精度;不止如此,在颈部网络,本文还创造性地提出了感知全局上下文注意力(Global Context Aware Attention, GCAA)模块,用于建立跨通道和空间的全局注意力,将各层级特征整合之后,模型能重点关注重点位置和关键信息,更好地适应肺结节特征的多样性和复杂性。在头部网络和颈部网络之间,使用损失排序挖掘(Loss Rank Mining, LRM)方法^[13],以实现在训练过程中挑选出

图像中置信度损失较高的像素点进行重点训练, 提高模型从复杂背景中鉴别出肺结节的能力。改进后的模型结构如图 1 所示。其中的 SPD、AFPN、GCAA、LRM 将在 2.1~2.4 节予以详细介绍。

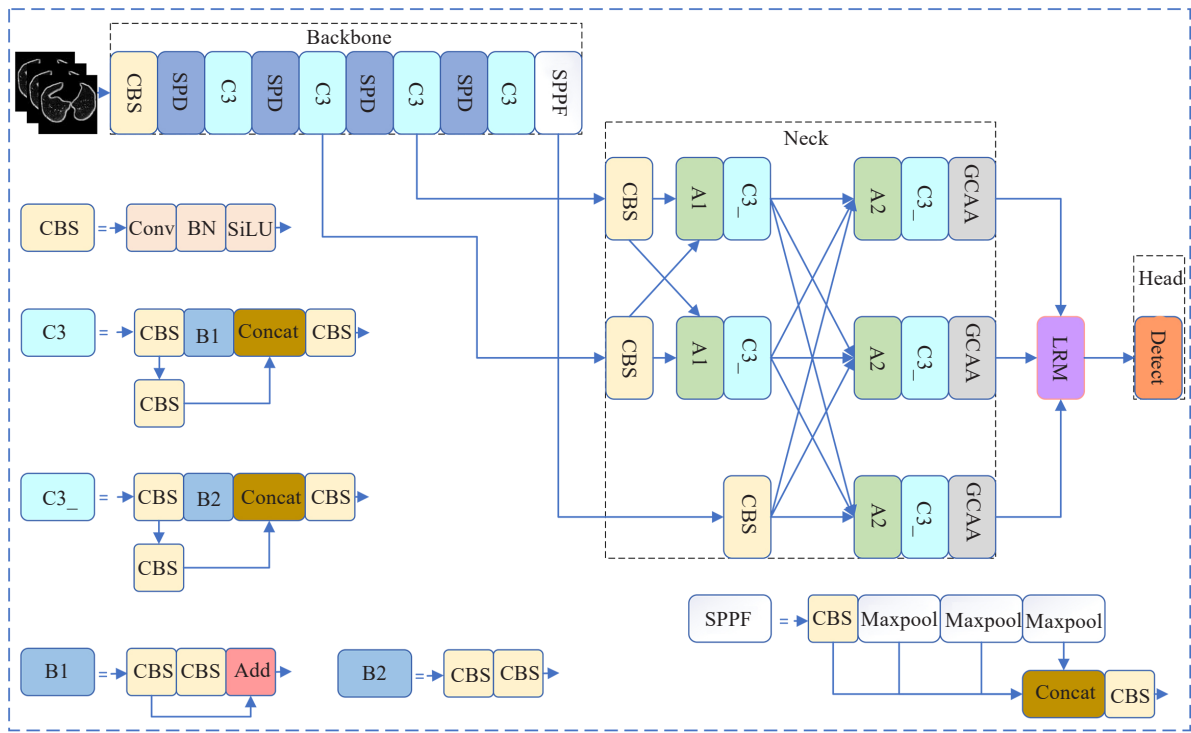


图1 改进后的YOLOv5结构
Figure 1 Improved YOLOv5 structure

2.1 SPD

原始YOLOv5中下采样由步长为2, 卷积核大小为3×3的卷积来完成, 这种操作是非对称采样, 即不同点的采样次数不同, 会导致细粒度信息丢失。如图2所示, 图中圆圈代表待下采样特征图中的像素点, 黑色框代表卷积核大小为3×3、步长为2的卷积每次卷积操作所覆盖的像素范围, 可见, 绿色像素点、黄色像素点、紫色像素点分别被采样了1次、2次、4次, 它们分别会对下一层级特征图的1个点、2个点、4个点产生影响。考虑肺结节检测中的小结节, 多为直径在3 mm(不到10个像素点)左右的类球体, 在网络中被多次下采样之后仅反应在一个像素点上, 若特征图中紫色点为背景信息, 绿色点为肺结节信息, 这种采样策略会使模型多次采样了无关背景而忽略了小结节的信息, 导致小结节漏检。此外, 当最大池化操作用作下采样时, 只保留了卷积核大小的框内最大像素的值, 忽略了其他像素信息, 导致信息丢失, 影响模型判断。

本文使用SPD操作将像素从空间重塑到深度, 输入张量X沿空间维度(宽度和高度)被分割成4个部分, 然后沿通道维度拼接起来。这个操作将输入

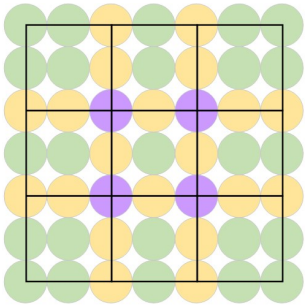


图2 原始YOLOv5下采样操作
Figure 2 Downsampling operation in original YOLOv5

张量的宽度和高度减小了一半, 同时将通道增加了4倍, 然后送入图1中的C3模块, 由C3中第一个卷积层将通道数进行复原。这种方法本质上是 将原始特征图中4个像素值加权求平均得到一个新的像素, 避免了非对称采样, 也解决了最大池化操作忽略非最大像素值的问题。SPD操作示意图如图3所示。

2.2 AFPN

原始YOLOv5采用自上而下和自下而上的特征金字塔网络, 如图4a所示。这种策略只是将相邻特征层简单拼接, 导致非相邻特征层特征的融合不足,

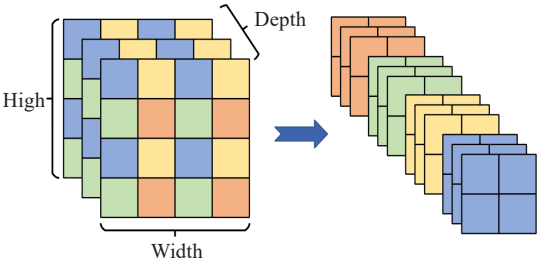


图 3 SPD 操作
Figure 3 SPD operation

从而使不同尺度信息融合不充分影响肺结节的检测精度。本文使用 AFPN 来解决上述问题,如图 4b 所示。由于非相邻层级特征图尺寸差异较大,直接融合会造成特征信息的丢失或退化,AFPN 首先通过加权求和的方式融合相邻两个层级的特征,之后在融合好相邻层级的基础上,进行非相邻层级的融合,这种融合策略减少了非相邻层级特征信息的差距,使不同尺度特征融合更充分。以图 4b 中特征图 A1 和 A2 为例,其融合策略的数学表示如式(1)、(2)所示:

$$A1 = \alpha1 \cdot X1 + \beta1 \cdot \text{Reshape}(X2) \tag{1}$$
$$A2 = \alpha2 \cdot \text{Reshape}(A1) + \beta2 \cdot (A1') + \gamma \cdot \text{Reshape}(C) \tag{2}$$

$\alpha1$ 、 $\beta1$ 、 $\alpha2$ 、 $\beta2$ 、 γ 是赋予不同路径的权重,通过学习得到,Reshape 表示采样操作。

2.3 GCAA

颈部网络将主干提取的不同尺度特征和语义信息进行融合,生成同时具有浅层细粒度信息和高层语义信息的特征图,在颈部网络之后,加入注意力机

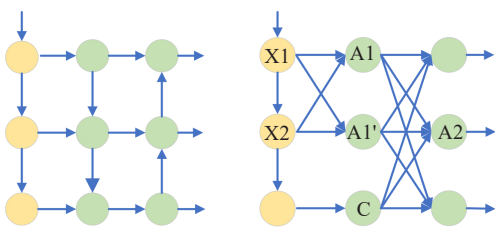


图 4 PAN 和 AFPN 结构
Figure 4 PAN and AFPN structures

制提取重要位置和关键特征可有效提高模型的性能。聚合上下文模块 (Context Aggregation Block, CABlock)^[14] 从空间和通道两个维度聚合全局上下文,然而 CABlock 只用一个卷积层加 Sigmoid 函数来学习空间维度的特征,缺少全局的关联,为此本文提出了 GCAA 模块,从通道和空间两个维度出发,根据跨通道跨空间不同位置之间的全局联系发现图像中的关键位置和特征,增强模型对语义信息的提取,更好地捕捉多维度的特征信息,形成更丰富、更全面的特征表示,提高模型对复杂结构的理解能力。就肺结节检测而言,增强了模型以全局视角从周围相似组织中甄别出肺结节的能力。CABlock 和 GCAA 结构如图 5 所示,图中 \otimes 表示矩阵乘法, \odot 表示广播哈达玛乘积, \oplus 表示矩阵相加, LR 表示 LayerNorm+Relu, 橙色虚线表示像素重新排列。

GCAA 由空间注意力和通道注意力组成。对于空间注意力,首先,输入 P 经过支路 1 中用作线性变

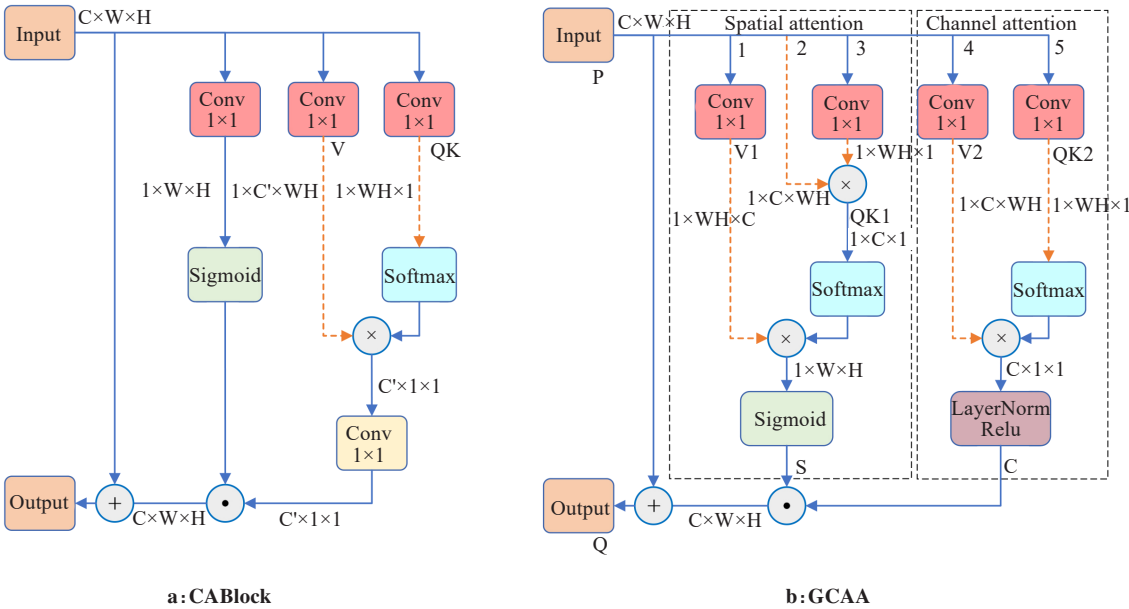


图 5 CABlock 和 GCAA 结构
Figure 5 CABlock and GCAA structures

换的卷积得到 $V1$;其次, P 经过支路2和支路3相乘完成查询和键的乘法运算后得到 $QK1$;最后将 $QK1$ 使用 Softmax 函数归一化后与 $V1$ 相乘再经过 Sigmoid 函数得到空间权重 S ,如式(3)所示:

$$S_i^j = \text{Sigmoid} \left(\sum_{j=1}^C \left(\frac{\exp(W_{QK1} P_i^j)}{\sum_{m=1}^C \exp(W_{QK1} P_m^j)} W_{V1} P_i^j \right) \right) \quad (3)$$

其中, P_i^j 和 S_i^j 分别表示第 i 通道中第 j 个像素的输入和空间权重, C 表示输入图像通道数。 W_{QK1} 、 W_{V1} 是线性变换矩阵, W_{QK1} 由输入经过支路3中的卷积后得到, W_{V1} 由支路1中卷积核大小为 1×1 的卷积完成。

对于通道注意力, P 首先经过支路4中用作线性变换的卷积得到 $V2$;其次 P 经过支路5中的卷积简化查询和键的乘法运算后得到 $QK2$,最后将 $QK2$ 使用 Softmax 函数归一化后与 $V2$ 相乘,再经过 LayerNorm 和 Relu 函数得到通道权重 C ,如公式(4)所示:

$$C_i^j = \text{Relu} \left(\text{LayerNorm} \left(\sum_{i=1}^N \left(\frac{\exp(W_{QK2} P_i^j)}{\sum_{n=1}^N \exp(W_{QK2} P_i^n)} W_{V2} P_i^j \right) \right) \right) \quad (4)$$

其中, C_i^j 是第 i 通道中第 j 个像素的通道权重, N 表示一个通道内所有像素总数。 W_{QK2} 、 W_{V2} 是线性变换矩阵,分别由支路4、支路5中卷积核大小为 1×1 的卷积完成。

最后将空间权重和通道进行哈达玛元素乘积操作,使空间与通道进一步关联之后与输入相加得到全局上下文权重 Q ,如公式(5)所示:

$$Q_i^j = S_i^j C_i^j + P_i^j \quad (5)$$

其中, Q_i^j 表示第 i 通道中第 j 个像素位置的输出。

2.4 重点训练困难样本

在肺结节检测问题中,背景数量远高于前景数量,并且背景和前景相似度高,特别容易忽略或误判。重点训练 CT 图像中难以分辨的区域可以帮助模型更准确地检测出较隐蔽以及容易混淆的肺结节,提高模型的检测精度。

本文使用 LRM 重点训练困难样本方法提高模型的鉴别能力。该方法在训练中首先将每个 Batchsize 中即将送入头部网络的特征图中所有像素按照置信度损失进行排序,然后选取一定比例的置信度损失较大点(即困难样本)进行训练,其余的点丢弃。特征图中置信度损失较大的点对应模型难以分辨的区域,重点训练置信度较大的点可使模型重点学习难以分辨的区域,从而加强模型的理解能力。

3 实验设计与结果分析

3.1 数据集介绍

实验中使用 LUNA16 数据集^[15],其内含 888 个患者的肺部 CT 扫描数据,涉及 1 186 个结节,共计 227 225 张 CT 图片和与其相关的肺结节注释信息。原始数据集包括 .mhd 和 .raw 格式的文件,其中 .mhd 文件存储了 CT 图像大小、CT 值范围以及结节的注释信息等;.raw 文件保存了实际的 CT 图像。在训练之前将 CT 图像转换成 .png 格式的普通图像,同时根据注释信息生成 txt 格式的标签文件。

3.2 数据预处理

本文使用传统方法对切片后的 LUNA16 数据集进行肺实质分割,可以排除 CT 图像中脂肪、空气以及非肺部区域等背景的影响。首先对 CT 图像进行二值化处理以及形态学处理消除边缘噪声,然后保留两块最大的连通区域,再进行开闭运算的形态学处理填充孔洞以及优化边缘得到掩膜,最后使用掩膜割取肺实质得到分割后的肺部图像。

为了提高模型的泛化能力,在训练之前使用 MedAugment^[16] 数据增广方法对预处理后的图像进行扩充,然后使用十折交叉验证方法划分数据集并训练验证。MedAugment 逐个筛选排除了自然图像数据扩充方法中不适用于医学图像分析的操作,例如反转、均衡化等操作,然后将数据增广操作分为像素级和空间级操作,并构建了两个增强空间,分别为 A_p 和 A_s ,其中 A_p 操作空间包括调整亮度、对比度、色调分离、锐度,以及添加高斯模糊和高斯噪声; A_s 操作空间包括旋转、水平翻转、垂直翻转、缩放、水平平移和垂直平移。

由于医学图像对亮度等属性非常敏感,并且 A_p 中连续的操作可能会破坏医学图像信息^[17],因此采用了一种新颖的操作采样策略,用于从 A_p 和 A_s 中进行采样操作:对于每次增强,首先从 A_p 随机选择 1 种或者 0 种方法,然后从 A_s 中随机选择 2 种或 3 种方法,最后将这些方法组合在一起完成对图像的一次增广。

3.3 实验配置及初始参数

采用 Python3.9、PyTorch1.12 框架进行编程,在 32 GB 内存、NVIDIA GeForce RTX 3090Ti、Intel Core i5-12600KF 3.70 GHz CPU 处理器和 Window 10 系统上进行网络训练与测试,其他详细参数设置如表 1 所示。

本文采用目标检测通用的评价指标:精确度(Precision)、召回率(Recall)、平均精度(Average

表1 训练参数设置
Table 1 Training parameter settings

训练参数名称	参数值
初始学习率	0.01
余弦退火超参数	0.01
权重衰减系数	0.000 5
批处理大小	16
训练批次	300
图像输入尺寸	640×640

Precision, AP)。精确度是正确检测出的正样本数占全部检测出正样本数的百分比,召回率是指正确检出的正样本占实际正样本的百分比,AP是在不同置信度下,以精确度为纵轴、召回率为横轴绘制成的曲线与坐标轴所包围的面积。精确度、召回率和AP的数学表示如式(6)、(7)、(8)所示:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{AP} = \int_0^1 P(R) dR \tag{8}$$

其中,TP为实际为正样本且被判定为正样本的数量,FP为实际为负样本但被判定为正样本的数量,FN为实际为正样本但被判定为负样本的数量,本文中肺结节为正样本,背景为负样本, R 表示召回率,作为自变量, $P(R)$ 表示当前置信度下召回率对应的精确度。

3.4 各个改进方法的对比实验

为了研究不同改进措施对算法性能的影响,本文进行了各部分改进的消融实验,实验结果如表2所

示。表2比较了在主干网络中将SPD操作作用作下采样,继而在颈部网络使用AFPN策略和在颈部网络末端添加GCAA后的效果。

表2 SPD、AFPN、GCAA对模型的影响(%)
Table 2 Effects of SPD, AFPN and GCAA on the model (%)

模型	精确度	召回率	AP
YOLOv5s	82.0	84.8	85.2
YOLOv5s+SPD	84.1	85.6	87.9
YOLOv5s+SPD+AFPN	88.2	89.2	90.4
YOLOv5s+SPD+AFPN+GCAA	92.3	91.8	93.1

由表2可见,将YOLOv5步长为2的卷积下采样操作替换为SPD操作时,由于步长为2的卷积丢失的部分信息得到保留,各项指标都有所提升;在YOLOv5颈部网络中使用AFPN策略后,各个层级的特征融合更充分,因此,精确度、召回率以及AP都得到一定的提升;最后加入GCAA,精确度、召回率以及AP分别提高了4.1%、2.6%、2.7%,表明本文提出的GCAA能够构建空间和通道之间的联系,从而找出关键位置和关键通道,有效整合信息,对算法的改进起到了比较关键的作用。

在前三项改进的基础上,重点训练困难样本,探究重点训练不同比例困难样本对最终结果的影响,实验结果如表3所示。对置信度损失从大到小进行排列,LRM($\alpha=x\%$)表示保留输入图像中前 $x\%$ 置信度损失较大的点进行重点训练,其余的点丢弃。

表3 重点训练不同比例困难样本对模型的影响(%)
Table 3 Effects of training on different ratios of hard samples on the model (%)

模型	精确度	召回率	AP
YOLOv5s+SPD+AFPN+GACC+LRM($\alpha=45\%$)	93.6	93.3	95.1
YOLOv5s+SPD+AFPN+GACC+LRM($\alpha=40\%$)	93.3	94.2	95.7
YOLOv5s+SPD+AFPN+GACC+LRM($\alpha=35\%$)	94.5	93.8	96.3
YOLOv5s+SPD+AFPN+GACC+LRM($\alpha=30\%$)	93.6	94.8	96.2
YOLOv5s+SPD+AFPN+GACC+LRM($\alpha=25\%$)	96.0	95.0	97.3
YOLOv5s+SPD+AFPN+GACC+LRM($\alpha=20\%$)	96.1	95.0	96.6
YOLOv5s+SPD+AFPN+GACC+LRM($\alpha=15\%$)	95.5	94.4	96.4

由表2、表3比较可知,重点训练困难样本的策略对整个模型都有改善;由表3可知,当 $\alpha=20\%$ 时精确度与召回率最高, $\alpha=25\%$ 时召回率和AP最

高,这与肺结节检测问题中肺结节和背景的比例有关。由于 $\alpha=25\%$ 时AP远高于其他情况,因此本文在前三项改进的基础上,将重点训练输入图像

中置信度损失较大点的比例确定为 25%, 此时的模型即为最终的模型。相比原始 YOLOv5, 改进后的算法将精确度提高了 14.0%, 召回率提高了 10.2%, AP 提高了 12.1%, 实验结果表明, 各项改进均取得了较好的效果。

3.5 不同注意力的对比实验

为了对比本文提出的 GCAA 与其他现行注意力机制在肺结节检测问题上的表现, 本文以 YOLOv5s 为基础, 在 YOLOv5 颈部网络的末端分别添加了卷积块注意力模块 (Convolutional Block Attention Module, CBAM)^[17]、挤压-激励注意力 (Squeeze-and-Excitation, SE) 模块^[18]、简易注意力 (Simple Attention Module, SimAM) 模块^[19]、坐标注意力 (Coordinate Attention, CA) 模块^[20]、高效通道注意力 (Efficient Channel Attention, ECA)^[21]、CABlock 和 GCAA 模块, 实验结果如表 4 所示。实验表明, 在 YOLOv5s 基础上添加不同注意力机制, 对检测指标均有不同程度的提高, 在加入 GCAA 之后, 精确度相比 YOLOv5s 提高了 7.2%, 召回率提高了 4.0%, AP 提高了 6.7%, 也大幅领先于其他注意力模块, 这是由于其他注意力机制通过最大池化或平均池化获取该通道或空间的重要性, 最大池化忽略其他非最大值信息、平均池化将所有信息同等看待不够灵活, 而 GCAA 中每一个权重都是根据全局上下文关系获取该空间或通道的重要程度, 这种方式更加灵活, 并且

兼顾了通道、空间和全局, 更能提高模型基于全局对关键位置、关键特征的理解能力。

表 4 GCAA 与不同注意力模块对比(%)
Table 4 Comparison of GCAA with different attention modules (%)

模型	精确度	召回率	AP
YOLOv5s	82.0	84.8	85.2
YOLOv5s+CBAM	86.4	84.7	87.2
YOLOv5s+SE	83.7	86.8	86.8
YOLOv5s+SimAM	84.6	84.0	87.0
YOLOv5s+CA	85.6	83.8	87.5
YOLOv5s+ECA	85.5	85.8	88.4
YOLOv5s+CABlock	85.9	83.9	87.1
YOLOv5s+GCAA	89.2	88.8	91.9

3.6 主流模型的对比实验

表 5 从检测效果和模型复杂度两个方面, 对比了本文算法和现有主流模型在肺结节检测问题上的表现。由表中数据可以看出, 本文算法在轻量化的基础上得到了不错的性能。其中精确度、召回率和 AP 分别为 96.0%、95.0% 和 97.3%, 相较于其他检测算法均有较大提高, 而参数量仅为其他算法的 1/10~1/5, 权重文件仅为 15.9 MB, 相较于其他算法, 部署和实际应用价值更高。

表 5 本文算法与其他主流检测模型对比
Table 5 Comparison of the proposed algorithm with other mainstream detection models

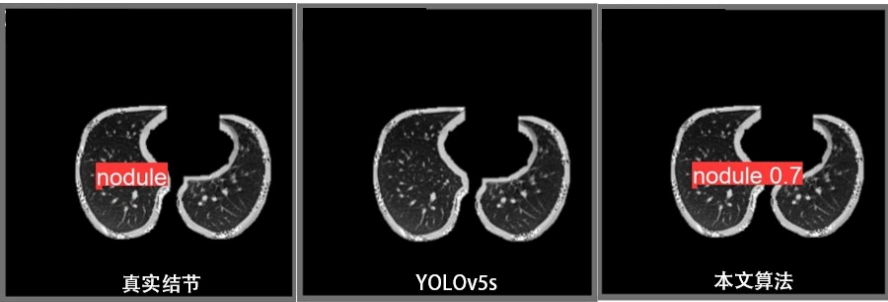
模型	精确度/%	召回率/%	AP/%	参数量	GFLOS	权重文件大小/MB
YOLOv3	93.6	93.0	94.9	6.2×10 ⁷	155.3	117.8
YOLOv4 ^[22]	68.5	54.9	57.3	5.2×10 ⁷	119.7	137.6
YOLOv7 ^[23]	71.2	69.4	68.2	3.7×10 ⁷	105.1	72.1
YOLOv8m	92.7	93.8	95.5	2.6×10 ⁷	78.9	49.6
FasterRCNN ^[24]	73.2	72.1	71.0	-	-	163.0
SSD ^[25]	67.3	64.8	65.2	-	-	102.0
YOLOv9m ^[26]	89.7	91.0	94.5	2.6×10 ⁷	76.3	38.8
YOLOv10m ^[27]	94.8	93.2	97.3	1.5×10 ⁷	59.1	63.8
本文算法	96.0	95.0	97.3	7.7×10 ⁶	27.7	15.9

GFLOS: 每秒 10 亿次浮点运算数

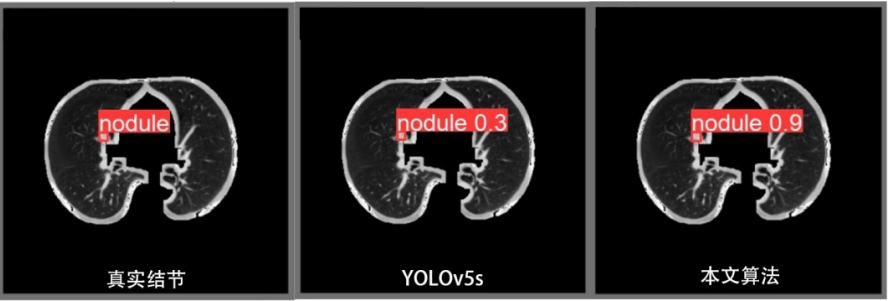
3.7 检测效果对比

图 6 展示了 YOLOv5s 和本文算法在不同大小结节检测中的表现。其中, 图 6a~c 是 3~5 mm 小结节检测结果, 图 6d、e 是 5~10 mm 和 10~30 mm 结节检测结

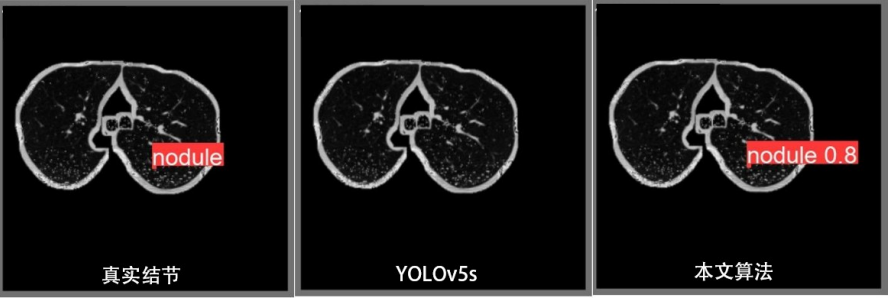
果。可以看出, 原始 YOLOv5s 网络检测结果存在置信度较小、检测错误率较高、尤其是小结节漏检的问题, 本文模型可以准确地检测出 CT 图片中不同大小的肺结节, 而且具有很高的置信度。



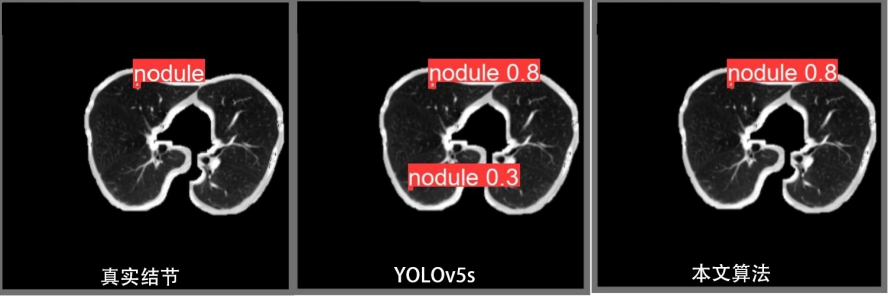
a:3~5 mm 小肺结节检测结果 1



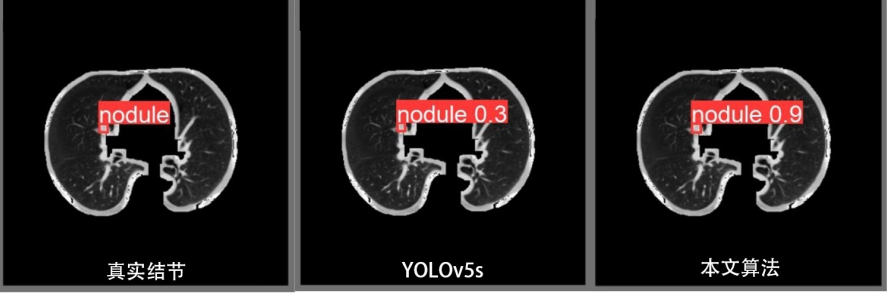
b:3~5 mm 小肺结节检测结果 2



c:3~5 mm 小肺结节检测结果 3



d:5~10 mm 肺结节检测结果



e:10~30 mm 肺结节检测结果

图6 肺结节检测结果

Figure 6 Lung nodule detection results

4 结 论

本文针对小肺结节难以检测,现有算法难以同时兼顾高精度和轻量化的问题,以轻量化的YOLOv5s模型为基础,首先改进了下采样操作,以减小细粒度信息的丢失;其次采用渐进融合特征的策略,有效融合了各个层级的特征;还提出适合肺结节检测的注意力机制,从全局视角关注关键通道和重要位置,提升模型的检测精度;最后探究了重点训练不同比例困难样本对肺结节检测的影响。实验结果显示,本文算法同时具备高精度和轻量化的优点,能够有效检测出肺结节并且更容易落地部署,具备实际应用价值。

【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 Cancer J Clin, 2024, 74(3): 229-263.
- [2] The International Early Lung Cancer Action Program Investigators. Survival of patients with stage I lung cancer detected on CT screening [J]. N Engl J Med, 2006, 355(17): 1763-1771.
- [3] Su Y, Li D, Chen XD. Lung nodule detection based on faster R-CNN framework [J]. Comput Methods Programs Biomed, 2021, 200: 105866.
- [4] Ding J, Li AX, Hu ZQ, et al. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2017. Cham: Springer International Publishing, 2017: 559-567.
- [5] Mai JY, Wang MH, Zheng JY, et al. MHSnet: multi-head and spatial attention network with false-positive reduction for lung nodule detection [C]//2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Piscataway, NJ, USA: IEEE, 2022: 1108-1114.
- [6] Guo N, Bai ZY. Multi-scale pulmonary nodule detection by fusion of cascade R-CNN and FPN [C]//2021 International Conference on Computer Communication and Artificial Intelligence (CCAI). Piscataway, NJ, USA: IEEE, 2021: 15-19.
- [7] Cai ZW, Vasconcelos N. Cascade R-CNN: delving into high quality object detection [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 6154-6162.
- [8] 王乾梁, 石宏理. 基于改进YOLO V3的肺结节检测方法[J]. 中国医学物理学杂志, 2021, 38(9): 1179-1184.
Wang QL, Shi HL. Pulmonary nodule detection based on improved YOLO V3 [J]. Chinese Journal of Medical Physics, 2021, 38(9): 1179-1184.
- [9] Wu XS, Zhang H, Sun JD, et al. YOLO-MSRF for lung nodule detection [J]. Biomed Signal Process Control, 2024, 94: 106318.
- [10] Liu S, Qi L, Qin HF, et al. Path aggregation network for instance segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 8759-8768.
- [11] Sunkara R, Luo T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects [C]//Joint European conference on machine learning and knowledge discovery in databases. Cham: Springer Nature Switzerland, 2023: 443-459.
- [12] Yang GY, Lei J, Zhu ZK, et al. AFPN: asymptotic feature pyramid network for object detection [C]//2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Piscataway, NJ, USA: IEEE, 2023: 2184-2189.
- [13] Köksal A, Tuzcuoglu Ö, Ince KG, et al. Improved hard example mining approach for single shot object detectors [C]//2022 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2022: 3536-3540.
- [14] Liu Y, Li HF, Hu C, et al. Learning to aggregate multi-scale context for instance segmentation in remote sensing images [J/OL]. IEEE Trans Neural Netw Learn Syst, 2024. (2024-01-23). <https://doi.org/10.1109/TNNLS.2023.3336563>.
- [15] 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.
- [16] Liu ZS, Lü QJ, Li YF, et al. MedAugment: universal automatic data augmentation plug-in for medical image analysis [EB/OL]. (2024-08-14). <https://arxiv.org/abs/2306.17466>.
- [17] 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.
- [18] Hu J, Shen L, Sun G. Squeeze-and-Excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.
- [19] Yang LX, Zhang RY, Li LD, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks [C]//Proceedings of the 38th International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2021: 11863-11874.
- [20] Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 13708-13717.
- [21] Wang QL, Wu BG, Zhu PF, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 11531-11539.
- [22] Bochkovskiy A, Wang CY, Liao HY. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. (2020-04-23). <https://arxiv.org/abs/2004.10934>.
- [23] Wang CY, Bochkovskiy A, Liao HY. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Chia Laguna Resort, Sardinia, Italy: PMLR, 2023: 7464-7475.
- [24] Girshick R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV). Chia Laguna Resort, Sardinia, Italy: PMLR, 2015: 1440-1448.
- [25] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [C]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [26] Wang CY, Yeh IH, Mark Liao HY. YOLOv9: learning what you want to learn using programmable gradient information [C]//Computer Vision-ECCV 2024. Cham: Springer Nature Switzerland, 2025: 1-21.
- [27] Wang A, Chen H, Liu LH, et al. YOLOv10: real-time end-to-end object detection [EB/OL]. (2024-10-30). <https://arxiv.org/abs/2405.14458>.

(编辑:薛泽玲)