

改进的YOLOv4算法在肺结核检测中的应用研究

王京华,袁金丽,郭志涛,王佳浩

河北工业大学电子信息工程学院,天津 300401

【摘要】针对CT影像中肺结核病灶复杂且尺度变化大造成检测精度低的问题,提出了一种改进特征融合方法的YOLOv4网络用于肺结核的检测。首先,采用尺度均衡的金字塔卷积来捕获不同尺度特征层之间的相互作用,并在此基础上以自适应空间特征融合的方式过滤掉不同尺度上的冲突信息,以进行特征的有效融合。其次,在低层特征上引入了坐标注意力以进一步提高小目标的检测精度。根据北京胸科医院提供的300例病患信息,搭建了一套规范的肺结核CT数据集,并在所构建的数据集上进行了实验。输入图片分辨率设定为512×512,与原始YOLOv4相比本文模型mAP提升了4.96%,且该指标优于现有主流肺结核检测算法,如Faster R_CNN、SSD、RetinaNet等。结果表明改进的YOLOv4算法能够有效解决检测目标尺度变化和小目标检测问题,提高检测精度。

【关键词】肺结核;深度学习;特征融合;坐标注意力;YOLOv4

【中图分类号】R318;TP391

【文献标志码】A

【文章编号】1005-202X(2023)01-0113-07

Application of improved YOLOv4 algorithm in the detection of pulmonary tuberculosis

WANG Jinghua, YUAN Jinli, GUO Zhitao, WANG Jiahao

School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China

Abstract: Aiming at the problem of low detection accuracy of pulmonary tuberculosis caused by the complex and large scale changes of tuberculosis lesions in CT images, YOLOv4 with an improved feature fusion block is proposed for the detection of pulmonary tuberculosis. Scale-equalizing pyramid convolution is used to capture the interaction between feature layers of different scales, and on this basis, the conflict information at different scales is filtered out by scale-equalizing adaptive spatial pyramid convolution, so as to achieve feature fusion effectively. In addition, coordinate attention is introduced on the low-level features for further improving the detection accuracy of small targets. A standardized tuberculosis CT data set is built using the information of 300 cases provided by Beijing Chest Hospital, and the experiments are conducted on the constructed data set. The input image resolution is set to 512×512. The results show that the proposed network increases mAP by 4.96% as compared with the original YOLOv4, and that it is better than the existing mainstream tuberculosis detection algorithms, such as Faster R_CNN, SSD, RetinaNet, etc. The improved YOLOv4 algorithm can effectively solve the problems of detection target scale changes and small target detection, thereby improving detection accuracy.

Keywords: pulmonary tuberculosis; deep learning; feature fusion; coordinate attention; YOLOv4

前言

肺结核是由结核分枝杆菌感染肺部引起的一种呼吸系统疾病,具有高度传染性。结核病的早期诊断在帮助医生发现结核病,降低死亡率方面具有重

要的临床意义。目前肺结核的检测方法主要有痰结核菌检查、X线胸片和CT断层扫描等。痰结核菌检查较为准确,但耗时久。X线胸片不易检测到位于肋骨或心脏前后等易被遮挡区域的病灶。CT在诊断肺部疾病和检测肺结核患者的肺部疾病严重程度方面具有重要的参考价值^[1-2]。例如CT图像可以更好地显示活动性肺结核相关的一些典型影像特征,包括空洞和树芽征等^[3]。肺结核诊断常采用的方法是人工阅片,但人工阅片易受主观影响。基于计算机辅助诊断的CT影像具有高效率 and 稳定的准确性等优势,对于提升肺结核检测与诊断的效率具有重要意义。

【收稿日期】2022-07-15

【基金项目】国家自然科学基金(61801164)

【作者简介】王京华,硕士研究生,研究方向:智能信息处理、计算机视觉、医学图像处理,E-mail: wjh_hebuter0676@163.com

【通信作者】袁金丽,博士,副教授,研究方向:机器学习、智能信息处理、医学图像处理,E-mail: 2005051@hebut.edu.cn

目前关于计算机辅助肺结核诊断,国内外有大量的研究。Yang 等^[4]通过改进的 Faster R-CNN (Faster Region-CNN)网络来检测肺结核^[5],并通过修改损失函数来缓解正负样本不平衡问题。Xie 等^[6]将 NAS-FPN (Neural Architecture Search-Feature Pyramid Network)引入 Faster R-CNN^[7],提高了小病灶的检测性能,但是其训练过程采用X线胸片作为输入,不利于对易遮挡区域病灶的检测。Griffin 等^[8]比较了 Faster R-CNN、Mask R-CNN 和各自级联版本用于检测肺结核的性能^[9],并通过比较被检测到的病灶位置与该病灶真实框的热图来减少假阳性。Liu 等^[10]在 FPN 上添加了一个分类网络^[11],并针对难例挖掘修改了损失函数。Liu 等^[12]采用 SSD (Single Shot MultiBox Detector)^[13]、RetinaNet^[14]、FCOS (Fully Convolutional One-Stage Object Detection)^[15]等网络,并通过添加额外的分类分支来减少假阳性。上述研究虽然一定程度上提高了模型的性能,但是它们并未考虑到肺结核病灶尺度变化大的特点,因此检测效果受到限制。

针对多尺度物体检测的问题,现有研究多采用特征金字塔的方法。以 FPN 为代表,其通过自顶向下的方式和横向连接来融合高层特征与低层特征。在 FPN 的基础上,PAFNet(Path-Aggregation Network)添加了一个自底向上的融合路径来传递低层的位置等信息^[16]。Bi-FPN (Bi-directional Feature Pyramid Network)通过自顶向下和自底向上结构的重复利用来融合不同尺度的特征^[17]。DetectoRS 通过建立特征金字塔各层级与主干网络对应层级之间的反馈来

提高特征金字塔的特征表示能力^[18]。这些基于 FPN 的融合方式大多只是将不同尺度的特征图缩放到同一分辨率后进行相加或拼接,未考虑到各尺度特征层之间的相关性及其对整体特征表示的贡献程度。

考虑到肺结核病灶形态复杂且尺度变化明显的问题,本文提出了一种基于卷积神经网络的计算机辅助肺结核诊断方法。该方法采用具备高速度与高精度的单阶段检测算法 YOLOv4 作为检测网络^[19],并在此基础上做了以下改进:(1)提出了一种新的特征融合方法,该方法可以学习到各尺度特征之间的关系,以提升网络对复杂病灶的整体检测性能。(2)针对小目标分辨率低、信息少的特点,引入了坐标注意力机制 (Coordinate Attention, CA)^[20],以提高小区域病灶的检测精度。

由于肺结核空洞是肺结核在 CT 影像上的典型表现,本文基于医院提供的数据创建了一个空洞数据集,并在该数据集上进行实验。结果表明,该特征融合方法有效地提升了检测性能,且坐标注意力机制的引入进一步增强了小区域病灶的检测效果。

1 改进的 YOLOv4 算法

为了更好地融合不同尺度的特征,设计了新型特征融合结构尺度均衡的自适应空间金字塔卷积 (Scale-Equalizing Adaptive Spatial Pyramid Convolution, SEASPC)。其次,为了提高小目标的检测精度,在主干网络获取的两个特征层后添加了 CA。形成了改进后的 YOLOv4 算法,整体结构见图 1。

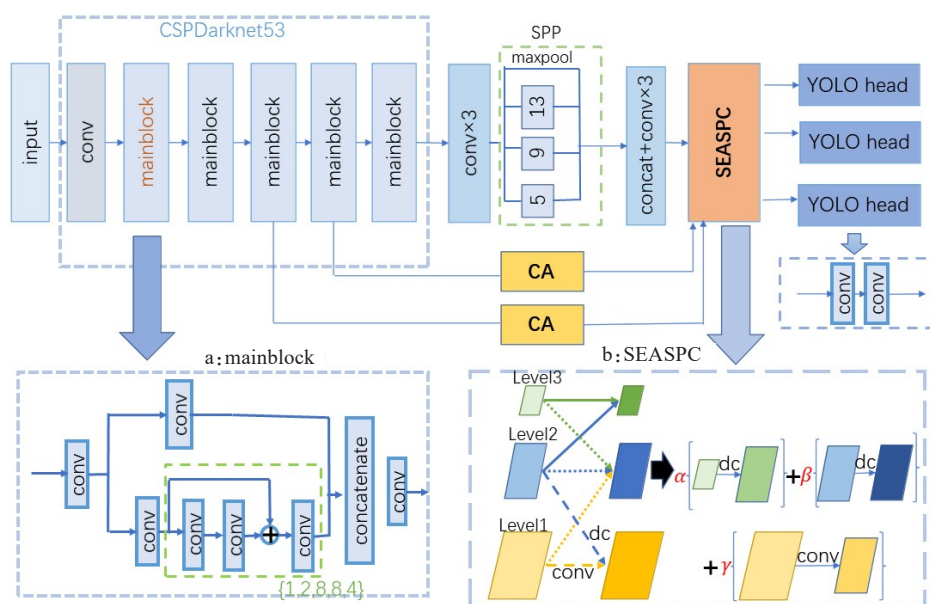


图1 改进后的YOLOv4结构图
Figure 1 Improved YOLOv4 structure

1.1 特征融合网络设计

1.1.1 尺度均衡的金字塔卷积 (Scale-Equalizing Pyramid Convolution, SEPC)

SEPC即在特征金字塔中进行3D卷积以学习特征层之间的关系^[21],并在3D卷积中对高层特征使用可变形卷积来实现尺度均衡。可变形卷积通过在采样点中引入偏移量使得采样区域集中于感兴趣区域。由于各特征层的尺度不同,选定一层作为基础特征层后,需要根据其他特征层的尺度来调整相应的步距决定进行上采样或下采样以满足融合要求。

SEPC的具体计算过程如下。假定 x 代表输入, w_1 、 w_0 、 w_{-1} 代表3种独立的二维卷积核。则第 l 个金字塔层的输出 y^l 可以表示为:

$$y^l = w_1 *_{s_{0.5}} x^{l+1} + w_0 * x^l + w_{-1} *_{s_2} x^{l-1} \quad (1)$$

$$= \text{Upsample}(w_1 * x^{l+1}) + w_0 * x^l + w_{-1} *_{s_2} x^{l-1}$$

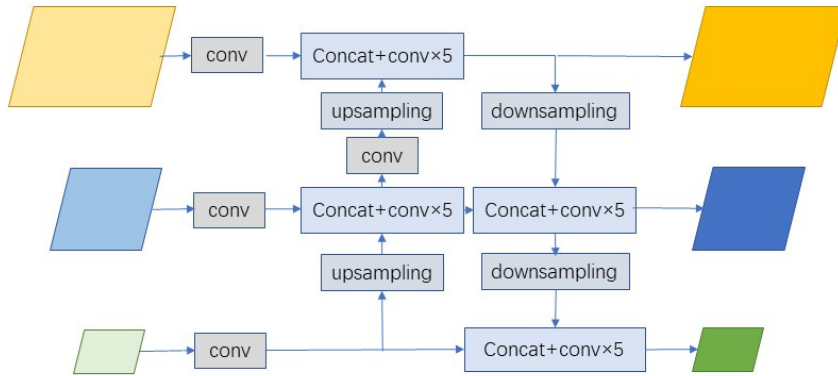


图2 PANet模块

Figure 2 PANet block

为此,本文使用SEPC来代替PANet,以捕获不同尺度特征间的相互作用。同时由于SEPC只进行了一次三维卷积,不仅同时提取了尺度及空间信息,而且减少了网络的参数量。考虑到不同特征层之间可能存在冲突信息,而SEPC中各尺度特征经过卷积之后直接相加,使得不同尺度的信息对最后输出特征图的贡献程度相等,不利于单一特征层上重要特征的保持。因此,受ASFF (Adaptively Spatial Feature Fusion)启发^[22],本文针对SEPC结构做了改进,在各尺度特征进行融合前首先赋予各特征层以初始权重,然后通过学习让包含不同尺度目标信息的特征在每个空间位置上进行融合,旨在通过自适应地调整融合时不同尺度特征的权重来过滤掉不同特征层上的冲突信息,以提高特征的尺度不变性。

改进后的特征融合网络如图1b所示,Level1、Level2、Level3是主干网络提取的3个尺度的特征层,“conv”和“dc”分别代表卷积和可变形卷积。

其中,*表示步距为1的卷积;*_{s₂}表示步距为2的卷积;*_{s_{0.5}}表示步距为0.5的卷积,通过一个步距为1的卷积和一个双线性上采样实现。

1.1.2 SEASPC 为了解决多尺度目标检测问题,YOLOv4中采用PANet结构来融合深层的语义信息和浅层的细节信息。然而,PANet结构存在以下缺点,首先,PANet通过多次的上采样与下采样操作以及横向连接将多个尺度的特征直接融合到一起,没有考虑到各尺度特征之间的相关性。具体来说,只有相邻的两个特征层之间是最相关的,而将一些相关性小甚至不相关的特征强行融合到某一特征层会损害该特征层的特征表达能力。其次,如图2所示,PANet中存在大量的卷积,大大增加网络层数的同时也增大了模型的学习难度。

SEASPC结构通过学习相邻尺度特征层之间的关系来更好地进行不同尺度特征的融合。由于各特征层的通道数和分辨率不同,采用不同步距的3×3卷积来同时控制通道数和分辨率以满足融合要求。

具体计算过程如下:

$$y_{ij}^l = \alpha_{ij}^l \cdot (w_1 *_{s_{0.5}} x^{l+1})_{ij} + \beta_{ij}^l \cdot (w_0 * x^l)_{ij} + \gamma_{ij}^l \cdot (w_{-1} *_{s_2} x^{l-1})_{ij} \quad (2)$$

其中, y_{ij}^l 表示第 l 层输出上位置 (i,j) 处的向量。 w_1 、 w_0 、 w_{-1} 代表3种独立的二维卷积核。 $(\cdot)_{ij}$ 表示在 (i,j) 处的特征向量。 α_{ij}^l 、 β_{ij}^l 、 γ_{ij}^l 通过softmax函数求得:

$$\alpha_{ij}^l = \frac{e^{\lambda_{\alpha_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \quad (3)$$

β_{ij}^l 、 γ_{ij}^l 的计算类似。其中, λ_{α}^l 、 λ_{β}^l 、 λ_{γ}^l 通过1×1卷积求得,分别代表各尺度特征层的权重。

1.2 CA

注意力机制可以使模型重点关注输入的某一部

分,即感兴趣的区域,而不是整张图像。在注意力机制的研究中,通道注意力(Squeeze-and-Excitation, SE)^[23]对输入特征不同通道的重要程度进行建模,CBAM(Convolutional Block Attention Module)首先在通道维度上建模^[24],然后在空间维度上学习空间位置的重要性。虽然上述的注意力机制通过捕捉通道或空间的关系一定程度上提高了模型的性能,然而,无论是SE还是CBAM,均在二维空间上采用了全局平均池化或全局最大池化操作,容易导致目标具体位置信息的丢失。而CA通过在两个方向上分别使用一个一维的平均池化,保留了另一个方向的位置坐标信息,通过拼接操作使得两个方向上的位置信息都得以保留,然后将坐标信息整合到通道注意力中,再完成对通道关系的建模。

由于肺结核空洞具有尺度变化大的特点,而小型空洞本身包含的信息就少,且随着神经网络层数的加深,小区域病灶的特征信息,尤其是空间位置等信息,更是逐层减少^[25],但直接减少主干网络层数会导致大目标信息提取不足,故笔者在YOLOv4的主干网络CSPDarkNet53获取的两个较浅的特征层(64×64,32×32)后添加了CA。一方面保留了小目标的空间位置信息,有助于模型实现对小区域病灶的定位;另一方面也有助于减少背景信息的干扰,使得模型可以集中关注小目标,从而提高模型检测小目标的能力。

CA模块见图3。首先,沿着水平和竖直方向分别对输入 X 进行一个一维的全局池化。由此生成的特征图中,不仅保留了一个方向精确的位置信息,还建立了另一个方向的长距离依赖关系。设编码第 c 个通道的池化核分别为 $(H, 1)$ 和 $(1, W)$,则其在高度 h 上的输出 $z_c^h(h)$ 和其在宽度 w 上的输出 $z_c^w(w)$ 分别为:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (4)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \quad (5)$$

为了充分利用位置信息,将两个方向产生的特征图进行拼接。然后利用 1×1 卷积 F_1 来进行通道数的压缩,再经过激活函数 δ 。产生的特征图表示为:

$$f = \delta(F_1([\mathbf{z}^h, \mathbf{z}^w])) \quad (6)$$

其中, $[\cdot, \cdot]$ 表示拼接操作。 $f \in \mathbb{R}^{C/r \times (H \times W)}$, r 为控制通道压缩程度的因子。将 f 沿着竖直和水平方向分解,则产生的 $f^h \in \mathbb{R}^{C/r \times H}$, $f^w \in \mathbb{R}^{C/r \times W}$,然后分别利用 1×1 卷积 G_h 和 G_w 来扩充两个方向上的通道数,使其与输入的通道数保持一致:

$$g^h = \sigma(G_h(f^h)) \quad (7)$$

$$g^w = \sigma(G_w(f^w)) \quad (8)$$

然后经过Sigmoid函数,得到两个方向上的注意力 g^h, g^w ,将其作用到输入,则经过注意力加权后的输出为:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (9)$$

CA不仅对输入特征各通道的重要程度进行编码,而且通过一维平均池化保留了位置信息,对于模型实现精准定位有很大帮助。

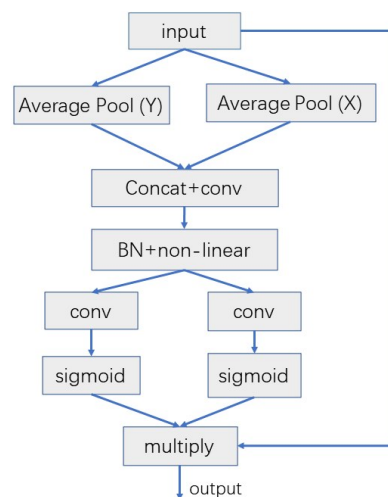


图3 坐标注意力模块

Figure 3 Coordinate attention block

1.3 算法实现

1.3.1 算法原理 首先通过K-means算法对数据集进行聚类,得到9种不同尺寸的先验框,这9种先验框按照尺寸大小被分配到特征金字塔3个尺度的特征层上,即每个特征层上有3种尺寸的先验框。3个特征层分别负责预测不同大小的目标,由于大目标的检测需要较大的感受野,所以尺度最小的特征层负责大目标的预测,而尺度最大的特征层则负责小目标的预测,中等目标由中间尺度的特征层来预测。对于每一个尺度的特征层来说,当物体的中心落在该尺度特征层的某一个网格点内时,则由该网格点负责该物体的预测。对3个特征层的预测结果进行解码后得到预测框的位置,通过得分排序与非极大值抑制得到最终的预测框^[26]。解码过程如下所示。

预测网络的输出为 $(t_x, t_y, t_w, t_h, \text{conf}, c)$ 。其中, t_x, t_y 表示预测框中心点相对于先验框中心点的偏移量, t_w, t_h 表示先验框的宽、高调整参数,conf表示目标置信度, c 表示种类置信度。则预测框的位置由以下公式求得:

$$b_x = \sigma(t_x) + c_x \quad (10)$$

$$b_y = \sigma(t_y) + c_y \quad (11)$$

$$b_w = p_w e^{t_w} \quad (12)$$

$$b_h = p_h e^{c_h} \quad (13)$$

其中, c_x 、 c_y 表示先验框的中心坐标, p_w 、 p_h 表示先验框的宽、高, b_x 、 b_y 表示预测框的中心坐标, b_w 、 b_h 表示预测框的宽、高。

1.3.2 网络结构 整体网络主要由主干网络 CSPDarknet53、融合部分 SEASPC 以及预测结构 YOLO head 3 个部分组成。首先,由检测器的主干网络对输入进行特征提取,当输入分辨率大小为 512×512 时,经过步距分别为 8、16、32 的下采样后,得到 3 个尺度分别为 64×64 、 32×32 、 16×16 的特征层。然后,来自较低网络层分辨率大小为 64×64 、 32×32 的特征层经过 CA,使网络集中关注中小目标的特征。分辨率为 16×16 的特征层先经过 3 次卷积,然后利用空间金字塔池化 (Spatial Pyramid Pooling, SPP) 结构来增大感受野^[27]。由此得到的 3 个特征层经过 SEASPC 结构来学习各尺度特征之间的关系,自适应地融合不同尺度的信息。最后对融合了不同尺度特征的 3 个特征层分别进行两次卷积来进行预测。

2 实验及结果分析

所有实验是在 Windows 系统上实现的,采用 pytorch1.7.1 深度学习框架,运算平台是 CUDA11.0。该计算机的处理器是 AMD Ryzen 7 4800H with Radeon Graphics 2.90GHz。

2.1 数据集与预处理

本研究使用的数据集由北京胸科医院提供,该数据集包括 300 个肺结核病例的 1 372 张 CT 切片。在这些切片中,肺结核空洞的直径范围为 10~126.4 mm。使用 LabelImg 软件将每张 CT 切片转换成格式为 PNG 的图片,分辨率大小为 512×512 ,同样使用该软件来标注空洞的位置,标注结果由 3 名来自北京胸科医院且经验超过 5 年的放射科专业医生进行校验,当校验结果不一致时,再由两名具有 10 年从医经验的资深放射科医师进行判断。

在实验中,将数据集按 8:2 的比例划分为训练集和测试集。训练集的 20% 用作验证集。迭代次数设为 400 次,初始学习率为 0.000 1,并使用余弦退火衰减策略来降低学习率,使用马赛克数据增强策略来提高模型对小物体的识别能力。此外,还使用了类标签平滑策略以提高模型的泛化能力。

2.2 评价指标

为了评估模型整体的性能,使用了以下评价指标:精确率、召回率、F1 分数、平均精度均值 (mAP)、对数平均漏检率 (Lamr)。AP 代表某一类物体的 P-R (Precision-Recall) 曲线下面积,计算如下:

$$AP = \sum_{i=1}^N P_i (R_i - R_{i-1}) \quad (14)$$

其中, i 表示取第 i 个置信度, N 为置信度的个数。 P_i 为第 i 个置信度下的精确率, R_i 为第 i 个置信度下的召回率。mAP 是 AP 的均值,用于衡量模型在所有类下的性能,计算公式如下:

$$mAP = \frac{\sum_{j=1}^M AP_j}{M} \quad (15)$$

其中, j 表示第 j 类, M 表示总的类别数, AP_j 表示第 j 类的平均精度。

对数平均漏检率的计算公式如下:

$$mr = \frac{FN}{TP + FN} \quad (16)$$

$$Lamr = \exp \left[\frac{1}{S} \sum_{i=1}^S \log (\max (mr, \varepsilon)) \right] \quad (17)$$

其中, mr 代表单张图片的漏检率, ε 为防止对数函数无意义的自设参数, S 为图片总数。

2.3 实验结果

2.3.1 消融实验 为了验证改进的结构对网络性能的影响,在本文创建的空洞数据集上进行了消融实验。首先,使用原始的 YOLOv4 作为基准。其次,分别对改进了特征融合结构和添加了 CA 的 YOLOv4 网络进行了实验。最后,对综合这两项改进后网络的整体性能进行验证。表 1 中, YOLOv4+CA 表示在分辨率分别为 64×64 、 32×32 的两个特征层后添加 CA。 YOLOv4+SEPC 表示将 YOLOv4 的特征融合部分换成 SEPC 结构后的网络。 YOLOv4+SEASPC 为采用本文设计的融合结构网络。 YOLOv4+CA+SEASPC 则是同时使用 SEASPC 结构及 CA 的网络。

由表 1 可以看出, SEPC 有效提高了模型的检测性能, mAP 提升了 2.22%。基于 SEPC 改进的特征融合模块 (SEASPC) 也使网络的 mAP 大幅提升,说明该模块不仅学习到了不同尺度特征层的相互关系,而且自适应融合的方式有效滤除了冲突信息。且 CA 的引入也提高了模型的精确率与召回率。另外,针对 YOLOv4 的每一项改进均有效降低了原始模型的漏检率。

除此之外,为验证本文设计的特征融合结构 SEASPC 的性能,将其与其他融合结构进行了比较,实验结果如表 2 所示。可以看出,本文设计的融合结构在精确率、召回率以及 mAP 等重要指标上均优于其他融合结构。

2.3.2 其他对比实验 在文献 [6]、[12] 中采用的 Faster R_CNN、SSD 等算法上进行了实验,结果如表 3 所示。可以看到, Faster R_CNN 的 mAP 值仅有 80.41%。 EfficientDet^[17]、RetinaNet^[14] 以及 SSD^[13] 等算法相比 Faster R_CNN 而言, mAP 值均有大幅提高,但仍有较大的提升空间。

虽然 YOLOv4 的表现优于 Faster R_CNN、SSD

表 1 不同结构的检测性能对比

Table 1 Comparison of detection performance among different networks

| 模型 | 精确率/% | 召回率/% | F1 分数 | 平均精度均值/% | 对数平均漏检率 | 参数量 |
|------------------|-------|-------|-------|----------|---------|------------|
| YOLOv4 | 92.86 | 90.43 | 0.92 | 92.09 | 0.122 6 | 64 040 001 |
| YOLOv4+CA | 92.44 | 92.17 | 0.92 | 93.90 | 0.096 6 | 65 026 881 |
| YOLOv4+SEPC | 92.17 | 92.17 | 0.92 | 94.31 | 0.088 3 | 50 787 429 |
| YOLOv4+SEASPC | 93.02 | 92.75 | 0.93 | 96.13 | 0.072 4 | 50 816 604 |
| YOLOv4+CA+SEASPC | 95.65 | 95.14 | 0.95 | 97.05 | 0.052 1 | 54 954 332 |

表 2 不同融合结构的对比(%)

Table 2 Comparison among different fusion networks (%)

| 融合结构 | 精确率 | 召回率 | 平均精度均值 |
|-------------|-------|-------|--------|
| HR-Net | 94.33 | 91.59 | 93.35 |
| Libra R-CNN | 94.72 | 93.62 | 95.34 |
| PANet | 92.44 | 92.17 | 93.90 |
| 本文融合结构 | 95.65 | 95.14 | 97.05 |

表 3 不同算法的检测性能对比(%)

Table 3 Comparison of detection performance among different algorithms (%)

| 算法 | 精确率 | 召回率 | 平均精度均值 |
|------------------------------|-------|-------|--------|
| Faster R_CNN ^[7] | 55.19 | 86.38 | 80.41 |
| EfficientDet ^[17] | 96.54 | 88.99 | 91.70 |
| SSD ^[13] | 91.25 | 84.64 | 90.54 |
| RetinaNet ^[14] | 89.81 | 81.74 | 88.42 |
| 本文算法 | 95.65 | 95.14 | 97.05 |

等网络,但是其 mAP 也只能达到 92.09%。原因可能在于 YOLOv4 的特征融合部分 PANet 为了利用上下文信息,不加分辨地将所有尺度的特征层信息融合到一起,从而导致网络无法做出准确的判断。本文设计的新型特征融合结构可以对主干网络获得的特征信息进行选择性融合,然后将融合后的特征信息送入预测网络进行预测。实验结果表明,有用的特征信息的充分利用提高了模型的性能,且 CA 保留的位置信息也有助于模型实现准确定位。

此外,绘制了交并比(Intersection over Union, IoU)阈值为 0.5 时几种模型的 P-R 曲线。AP 的值越大,模型的性能越好。如图 4 所示,YOLOv4 及综合两项改进后的 YOLOv4(本文模型)的性能均优于 Faster R_CNN。且本文模型的 P-R 曲线在 YOLOv4 之上,表明其性能优于 YOLOv4。与 SSD 和 RetinaNet 等算法相比,本文模型也表现出了良好的性能。

图 5 展示了本文模型以及其他算法在测试集上的

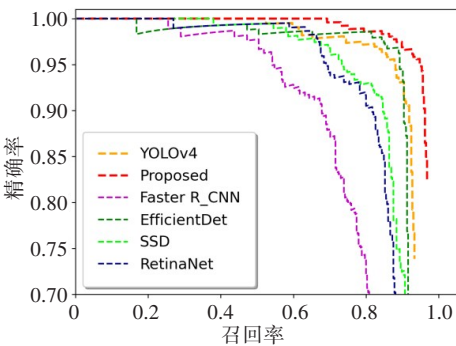


图 4 不同算法的 P-R 曲线图

Figure 4 P-R curves of different algorithms

结果,选取了其中 3 张进行比较。其中,图 5a 为原始 YOLOv4 的检测结果图,图 5b 为本文模型的检测结果图。如图 5b 所示,本文模型可以准确地检测出 CT 上不同尺寸的肺结核空洞,而且具有很高的置信度。而未修改融合网络部分的 YOLOv4 未能检测出所有病灶(图 5a),其没有检测出第二张 CT 图像中位于主气管左下角的小型空洞,由此可见,本文提出的特征融合结构是有效的,学习到了不同尺度病灶的特征,且 CA 的引入有效提高了小病灶的检测精度。图 5 中还展示了其他对比算法的检测结果图。可以看到,Faster R-CNN 在小型病灶的识别中出现了很多误检的情况(图 5c),在第一张与第三张 CT 图像中,Faster R-CNN 都将支气管误检为肺结核空洞,而且第二张图中也出现了漏检的情况。而 EfficientDet、SSD 以及 RetinaNet 也都出现了漏检的问题,且检测出来的病灶也具有很低的置信度。

3 结 论

本文提出一种基于卷积神经网络的方法用于肺结核典型病灶空洞的检测。该方法针对 YOLOv4 的融合结构存在的缺点,设计新型特征融合结构 SEASPC,有效提升了模型的性能。另外,考虑到小目标信息少且难识别的问题,引入 CA 来帮助模型定位,降低了模型的漏检率。在医院提供的数据集上进行消融实验,结果表明本文模型优于原始 YOLOv4 网络,mAP 提升了 4.96%,参数量也有所减少。将其

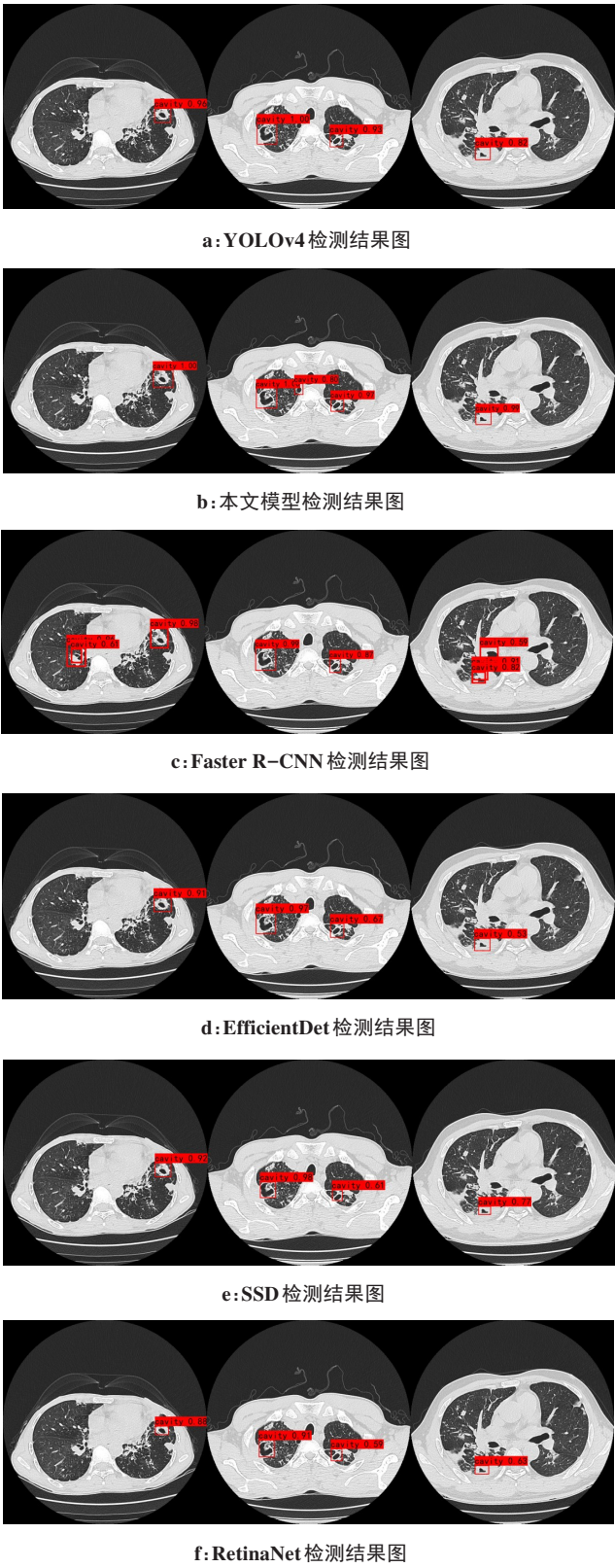


图5 不同算法检测结果

Figure 5 Detection results of different algorithms

从左至右分别为3个不同患者的CT切片

与当前经典的肺结核检测算法进行了比较,发现本文模型的mAP优于其他算法。实验表明,该方法对于肺结核空洞的检测是有效的,且在辅助医生进行肺结核诊断方面具有实用价值。

【参考文献】

[1] Leung CC. Reexamining the role of radiography in tuberculosis case finding[J]. Int J Tuberc Lung Dis, 2011, 15(10): 1279.

[2] Jaeger S, Karagyris A, Candemir S, et al. Automatic tuberculosis screening using chest Radiographs[J]. IEEE Trans Med Imaging, 2014, 33(2): 233-245.

[3] Schmidhuber J. Deep learning in neural networks: an overview[J]. Neural Netw, 2015, 61: 85-117.

[4] Yang A, Jin X, Li L. CT images recognition of pulmonary tuberculosis based on improved faster RCNN and U-Net [C]//2019 10th International Conference on Information Technology in Medicine and Education (ITME), 2019.

[5] Ren S, He KM, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(6): 1137-1149.

[6] Xie Y, Wu ZY, Han X, et al. Computer-aided system for the detection of multicategory pulmonary tuberculosis in radiographs[J]. J Healthc Eng, 2020(6): 1-12.

[7] Ghiasi G, Lin TY, Le QV. NAS-FPN: learning scalable feature pyramid architecture for object detection [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.

[8] Griffin T, Cao Y, Liu B, et al. Object detection and segmentation in chest X-rays for tuberculosis screening[C]//2020 Second International Conference on Transdisciplinary AI (TransAI), 2020.

[9] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017: 2961-2969.

[10] Liu J, Liu Y. A locating model for pulmonary tuberculosis diagnosis in radiographs[J]. arXiv preprint arXiv: 1910.09900, 2019.

[11] Lin TY, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.

[12] Liu Y, Wu YH, Ban Y, et al. Rethinking computer-aided tuberculosis diagnosis[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[13] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]. Springer, Cham, 2016.

[14] 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.

[15] Tian Z, Shen C, Chen H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2020.

[16] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[17] Tan M, Pang R, Le Q V. Efficientdet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 10781-10790.

[18] Qiao S, Chen LC, Yuille A. DetectoRS: detecting objects with recursive feature pyramid and switchable atrous convolution[J]. arXiv preprint arXiv: 2006.02334v1, 2020.

[19] Bochkovskiy A, Wang CY, Liao HY. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2004.10934, 2020.

[20] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 13713-13722.

[21] Wang X, Zhang S, Yu Z, et al. Scale-equalizing pyramid convolution for object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[22] Liu S, Huang D, Wang Y. Learning spatial fusion for single-shot object detection[J]. arXiv preprint arXiv: 1911.09516, 2019.

[23] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Trans Pattern Anal Mach Intell, 2019, 42(8): 2011-2023.

[24] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision, 2018.

[25] 赵鹏飞, 谢林柏, 彭力. 融合注意力机制的深层次小目标检测算法[J]. 计算机科学与探索, 2022, 16(4): 927-937.

[26] 卢迪, 马文强. 基于改进YOLOv4-tiny算法的手势识别[J]. 电子与信息学报, 2021, 43(11): 3257-3265.

[27] Lu D, Ma WQ. Gesture recognition based on improved YOLOv4-tiny algorithm[J]. Journal of Electronics and Information Technology, 2021, 43(11): 3257-3265.

[27] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Trans Pattern Anal Mach Intell, 2015, 37(9): 1904-1916.

(编辑:薛泽玲)