

DOI:10.3969/j.issn.1005-202X.2022.12.004

医学影像物理

长尾分布下的PET-CT肺癌图像的深度学习网络检测与定位探索

梁智欣¹, 邓戈龙², 魏夏平³, 梁凤好¹, 惠贤娟¹, 梁怡君¹, 李谦¹, 黄小伟⁴, 罗荣城⁵

1. 广州中医药大学金沙洲医院核医学科, 广东 广州 510168; 2. 广东工业大学计算机学院, 广东 广州 510006; 3. 广州中医药大学金沙洲医院肿瘤放射治疗中心, 广东 广州 510168; 4. 东莞理工学院科学技术处, 广东 东莞 523808; 5. 广州中医药大学金沙洲医院肿瘤科, 广东 广州 510168

【摘要】目的:提出一种用于PET/CT图像在长尾分布下的肺癌检测方法,以提高PET/CT图像中的肺癌诊断效能。**方法:**以YOLOv5作为骨干网络(Backbone),通过将Backbone与自适应类损失函数(ACSLoss)相结合来构建一个基于自适应类损失函数的YOLO模型(ACS-YOLO),以此解决PET/CT肺癌图像真实数据集中的长尾分布问题并提高PET/CT图像中的肺癌诊断效能。**结果:**在Lung-PET/CT-Dx公开数据集上与现存的YOLOv5变体相比,本文提出的ACS-YOLO取得了最好的检测性能,Precision、Recall、mAP@0.5和mAP@0.5:0.95指标最好的值分别为0.9607、0.9489、0.9706和0.5583。与其他YOLOv5变体相比,ACS-YOLO的检测性能提升约1%~5%,而尾部类别检测性能提升约5%~11%。**结论:**提出的ACS-YOLO可有效地提高长尾分布下PET/CT图像中的肺癌检测效能,这表明本文提出的方法能够作为现实PET/CT肺癌诊断的辅助工具。

【关键词】肺癌;PET/CT;长尾分布;深度学习;目标检测

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2022)12-1473-12

Deep learning network for lung cancer detection and localization in PET-CT images under long-tailed distribution

LIANG Zhixin¹, DENG Gelong², WEI Xiaping³, LIANG Fenghao¹, HUI Xianjuan¹, LIANG Yijun¹, LI Qian¹, HUANG Xiaowei⁴, LUO Rongcheng⁵

1. Department of Nuclear Medicine, Jinshazhou Hospital, Guangzhou University of Chinese Medicine, Guangzhou 510168, China; 2. School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China; 3. Department of Radiation Oncology, Jinshazhou Hospital, Guangzhou University of Chinese Medicine, Guangzhou 510168, China; 4. Science and Technology Division, Dongguan University of Technology, Dongguan 523808, China; 5. Department of Oncology, Jinshazhou Hospital, Guangzhou University of Chinese Medicine, Guangzhou 510168, China

Abstract: Objective To propose a method for lung cancer detection in PET/CT images under long-tailed distribution for improving the efficiency of PET/CT images for lung cancer diagnosis. **Methods** A YOLO framework called adaptive class loss function-YOLO (ACS-YOLO) which mainly consists of a YOLOv5 as backbone network (Backbone) and an adaptive class loss function (ACSLoss) was constructed to solve the problem of long-tailed distribution in the real data set of PET/CT lung cancer images and to improve the efficiency of PET/CT images for lung cancer diagnosis. **Results** Among the existing YOLOv5 variants on the Lung-PET/CT-Dx public data set, the proposed ACS-YOLO achieved the best detection performances in the precision, recall, mAP@0.5 and mAP@0.5:0.95 which were 0.9607, 0.9489, 0.9706 and 0.5583, respectively. Compared with the other YOLOv5 variants, ACS-YOLO showed a 1%-5% improvement in detection performance and a 5%-11% improvement in tail category detection performance. **Conclusion** The proposed ACS-YOLO can effectively improve the lung cancer detection in PET/CT images under long-tailed distribution, which indicates that the proposed method can be used as an aid for lung cancer diagnosis using PET/CT images.

Keywords: lung cancer; PET/CT; long-tailed distribution; deep learning; target detection

【收稿日期】2022-07-24

【基金项目】国家自然科学基金青年基金(12004410)

【作者简介】梁智欣,主治医师,研究方向:肿瘤核医学,E-mail: paul8411@

163.com

前言

肺癌是全世界范围内发病率和死亡率均居于首位的恶性肿瘤,据2020年全球最新癌症数据显示,中

国新发肺癌人数约为81万人,占中国新发癌症总数的24.7%。早诊断、早治疗能够有效降低肺癌的发病率与死亡率,早期肺癌的有效诊断及治疗可提高患者5年生存期,降低死亡率,这对于早期肺癌患者有着极其重要的意义。如今,正电子发射计算机断层显像(PET/CT)是人工诊断肺癌患者最准确有效的方法之一,它既能像PET一样提供详细的功能影像和分子代谢信息,又能像普通CT那样提供病灶区域的精准定位和病灶结构信息。由于肺癌早期以肺结节的形式存在,形态各异、尺度变化较大,且容易与周围的血管组织等器官混淆,给PET/CT诊断带来了极大的困难。尽管PET/CT的先进放射技术和设备更新迭代,但肺癌病灶的人工诊断仍然需要多位放射专家对相关图像进行仔细检查才能完成诊断过程。因此,如何对PET/CT图像中的肺癌病灶图像实施快速且精准的自动检测和提高诊断效能成为现今需要解决的一个关键问题。

近年来,深度学习方法在医学图像检测领域发展迅猛。Girshick等^[1]提出双阶段目标检测算法R-CNN,该算法相较于传统的检测算法在准确率方面有着较大的提升,但模型检测速度缓慢,不适合用于大规模医学图像检测。而为提升检测效率,Ren等^[2]为改进R-CNN而提出了Faster R-CNN,该算法大幅度地提升了网络的检测速度。与此同时,Redmon等^[3]创造性地提出了单阶段的目标检测算法YOLOv1,它极大地降低了检测时间,但YOLOv1对于医学图像中病灶尺度大小不一的数据检测能力较差,甚至无法检测小尺度的病灶目标。为解决多尺度问题,Singh等^[4]采用多尺度训练策略对模型进行训练,以提高模型对多样性尺度的目标检测能力。Jocher等借鉴SPP^[5]和PAN^[6]等多尺度特征提取技术,创造性地提出YOLOv5以解决多尺度目标和小目标难检测的问题。Zhang等^[7]提出基于3D双路径网络和编码器-解码器结构的3D Faster R-CNN用于肺癌检测,在公开数据集LIDC-IDRI上的性能表现优秀。

受多尺度特征提取网络的启发,本研究创造性地将YOLOv5作为基础框架应用于PET/CT肺癌检测任务中,针对性地解决肺结节形态各异且尺度变化大的问题,并提高其对小目标的检测性能。同时,以上工作都是在数据分布平衡的数据集,如LIDC-IDRI数据集上表现良好。然而,在真实场景中肺癌PET/CT数据存在长尾分布及类别分布不平衡的问题,数据中的头部类别贡献了整个数据集中的绝大多数样本,导致出现频率较低的尾部类别数据检测性能下降。如果直接将上述工作应用在更贴近实际情况的长尾数据集上,由于

数据分布不均衡,性能表现可能会比较差。为了解决这个问题,本文将自适应类抑制损失函数(ACSLoss)和YOLOv5结合起来,设计一种可以有效处理肺癌PET/CT长尾分布数据集的自适应类抑制YOLO目标检测方法(ACS-YOLO)。

1 方法

为了提高长尾分布下的PET/CT肺癌图像的诊断效能,本研究使用YOLOv5作为基础框架。本文所提出方法的框架见图1,整个模型首先采用一个骨干网络对输入的图片数据进行特征提取输出特征矩阵。骨干网络中的Focus模块对图片进行切片使输入图片在极少信息丢失的情况下扩充图片通道,此外使用CBL(Convolutional Layer-Batch Normalization-LeakyReLU)模块进行特征提取和参数优化,BottleneckCSP模块^[8]可以减少参数量和优化梯度传递。随后利用SPP模块对特征图进行多尺度特征提取与融合,以捕获更丰富的上下文信息。更进一步地,使用PAN模块对浅层的低级定位特征与深层的高级语义特征进行反复聚合,为模型提供强语义信息^[9]和精准的定位信息,以此来提高模型的检测性能。最后通过自适应类损失函数优化尾部类别的检测性能。

1.1 基础结构

1.1.1 卷积层 ACS-YOLO中的CBL基础模块由3个部分组成,首先是卷积层(Convolutional Layer)^[10]提取特征,然后采用批归一化(Batch Normalization)^[11]对参数进行调整,最后采用LeakyReLU避免神经元死亡现象并且加速网络收敛。模块由一个卷积层开始,用来对上层传入的数据进行特征提取。卷积层中的卷积核数目为输入通道数 c_{in} 的两倍,卷积核尺寸为 3×3 ,步长为2,这样的卷积层设计可以使输出的特征矩阵通道数 c_{out} 提升至原来的两倍,特征矩阵的分辨率降低一半。

1.1.2 Batch Normalization 在神经网络中,训练过程中随着某层的前几层的参数发生了变化,该层的输入也会相对发生变化,并且会造成后面几层的协方差转移,使网络难以训练出饱和和非线性的模型。BN是优化神经网络的一种方法,主要目的是在参数进入激活函数前调整输入的一个批次(Batch)的特征图,使其满足均值为0、方差为1的分布规律。神经网络更容易在这种统一规格的数据中学习规律,使用BN可以加速网络的收敛并提升准确率,使网络可以使用更高的学习率,不需要担心初始化的问题,同时在某些情况下可以Dropout^[12]。其中计算均值和方差的公式为式(1)和式(2),标准化处理为式(3):

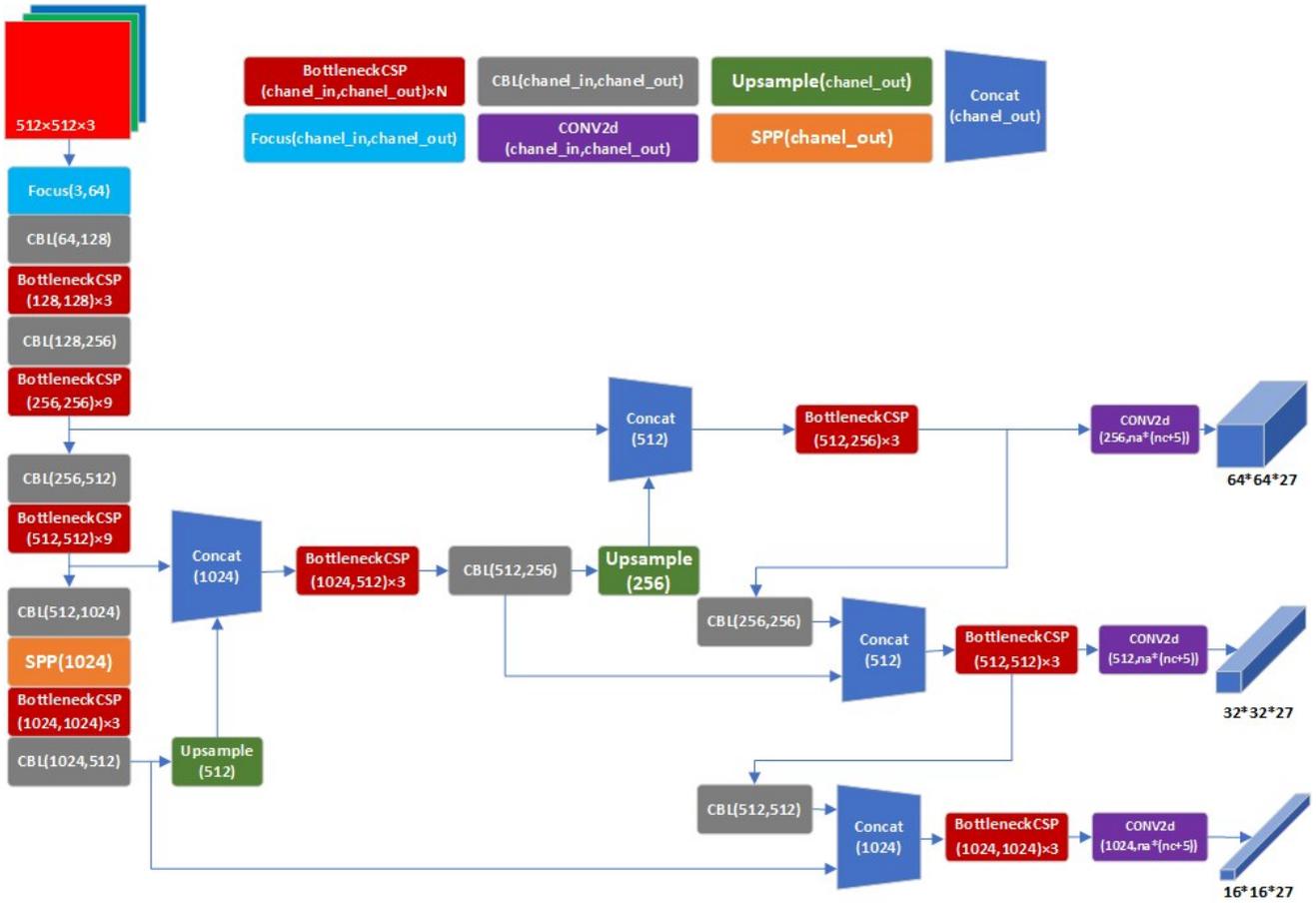


图1 ACS-YOLO的总体框架图

Figure 1 General framework of ACS-YOLO

$$\mu\beta \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad (1)$$

$$\sigma_\beta^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu\beta)^2 \quad (2)$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu\beta}{\sqrt{\sigma_\beta^2 + \varepsilon}} \quad (3)$$

其中,式(1)是计算同一批数据同一个通道参数的均值,计算结果为向量形式, $\mu\beta$ 中每个元素都代表一个通道参数的均值,其中 m 为每批数据中的图片数, x_i 为通道参数;式(2)是同批数据同一个通道的参数方差,计算结果为向量形式, σ_β^2 向量的每个元素代表一个通道的方差;而式(3)是对参数进行标准化处理,其中 ε 是一个超参数,为了避免公式分母为零。

1.1.3 LeakyReLU 激活函数为神经网络提供非线性映射学习能力,它的加入让神经网络有了非线性建模能力。修正线性单元(ReLU)是众多激活函数的一种,但ReLU也有缺点,随着网络训练的推进,当激活函数中的偏置项(bias)太小有可能导致激活函数的输出值总是负数,使得该处的反向传播梯度值为0,对应的权重参数无法进行更新,该神经元无法进行学习,称为神经元死亡(Dead Neuron)。由此提出新

的激活函数LeakyReLU,它是在Relu激活函数的基础上,改变负半区的图像,在反向传播过程中,对于LeakyReLU激活函数输入小于零的部分,也可以计算得到梯度,允许模型继续基于梯度进行学习,解决Relu函数的输入在一定情况下进入负区间后,导致神经元死亡的问题。与此同时,使用LeakyReLU作为激活函数不会导致网络过拟合现象,相比Sigmod/tanh的计算速度和网络收敛速度都更快。ReLU和LeakyReLU的计算公式如下所示:

$$f(x) = \max(0, x) \quad (4)$$

$$f(x) = \max(\alpha x, x) \quad (5)$$

其中, α 为超参数,根据不同模型进行适当调整,两个函数如图2所示。

1.1.4 残差结构 在VGG^[13]中发现更深的网络会给神经网络带来显著的性能提升,随着神经网络中的网络层数不断增加一般会伴随着几个问题:计算资源需求增多,模型容易过拟合、梯度消失/爆炸问题的出现。这些问题可以通过多GPU训练、数据增强配合Dropout以及BN来解决。而事实是随着网络层数的继续增加,可以发现网络开始变得难以优化,无法学习到更多信息。更深的网络学习能力不如浅层网

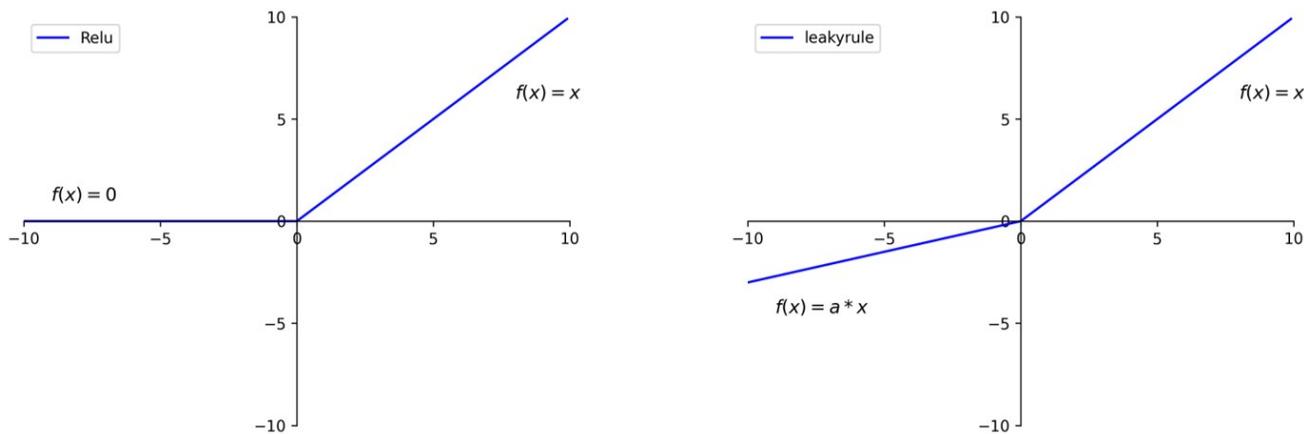


图2 ReLU和LeakyReLU的函数图像
Figure 2 Function images of ReLU and LeakyReLU

络,这种现象称为网络退化。He等^[14]在ResNet中提出残差网络实现了网络层间的恒等映射,使得深层的网络可以学习到浅层网络的特征信息,使得数据流可跨层流动,具体结构如图3所示。神经网络中的一层通常可以看作是 $y = H(x)$,通过学习去优化 $H(x)$ 这个未知形式的函数进行学习。残差结构(Residual Structure)让这个问题转化为了 $F(x)=H(x)-x$,使 $F(x)$ 去逼近0值,减少真实值 x 与预测值 $H(x)$ 的残差从而达到学习残差结构中的 $F(x)$ 这个未知形式的函数, $F(x)$ 可以学习到的上层网络输出和当前残差模块的卷积计算的融合信息,让网络蜕化问题得以解决。使用残差结构还可以让梯度信息跨层传输,加速网络学习需要的计算资源和时间,同时网络结构也可以进一步扩展到成百上千层,这使深层神经网络能学习到更多特征信息。

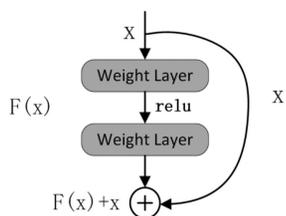


图3 残差结构图
Figure 3 Residual structure

1.2 BottleneckCSP

模型使用 BottleneckCSP 结构,在原 CSP Network^[15]上进行了一些优化,BottleneckCSP主要是通过基本层的特征图分为两个部分,其中一部分通过残差结构模块进行计算,另一部分通过一个Shotcut进行串联,具体结构如图4所示。这种架构可以实现更丰富的梯度组合,同时减少计算量。增加CSP结构后可以加强网络的学习能力,减少整个模型

的计算瓶颈,计算量可以减少10%~20%,使网络可以部署在CPU和移动GPU上而不会损失性能。同时可以减少网络参数,减少内存消耗,跨阶段的局部网络也使梯度传递更方便,整体上加速了模型的训练速度和准确率。

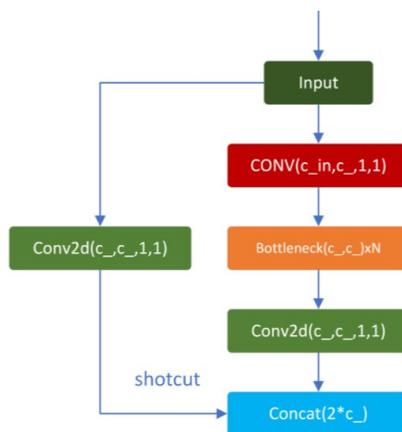


图4 BottleneckCSP网络架构图
Figure 4 BottleneckCSP architecture

1.3 SPP

SPP模块是一个“即插即用”的多尺度特征提取网络结构,其主要借鉴文献[16]中的空间金字塔结构思想。空间金字塔结构最初是为了解决神经网络需要输入固定尺寸图像的问题而被提出的,它有效地避免了由图像剪裁、缩放等操作而导致的图像失真问题。本文将SPP的思想应用到YOLOv5网络中,对图像进行多尺度特征提取与融合,以此来丰富特征图的表达能力,从而提高模型对病灶的检测性能。SPP模块位于BottleneckCSP主干特征提取网络之后,主要由池化核大小分别为 1×1 、 5×5 、 9×9 和 13×13 的最大池化层组成,采用不同尺寸的最大池化层对图

像进行多尺度特征提取,以增强模型的特征表达能力。在SPP模块中,不同尺度的特征图将在通道的维

度上进行堆叠并进行3次的 3×3 卷积操作,以实现多尺度特征融合。SPP模块的网络结构如图5所示。

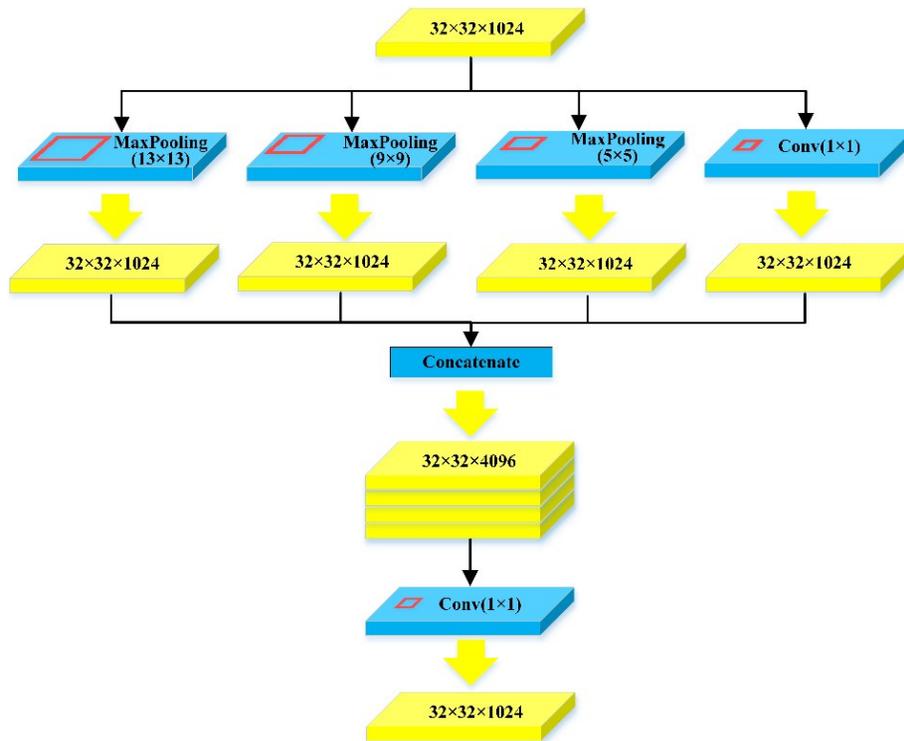


图5 SPP模块结构图

Figure 5 SPP module structure

SPP模块能够极大地增加模型的感受野^[17],显著地分离出最重要的上下文特征,感受野的具体计算公式如下所示:

$$RF_{l+1} = RF_l + (\text{kernal}_{\text{size}} - 1) \times \text{stride} \quad (6)$$

其中, RF_{l+1} 代表第 $l+1$ 个隐藏层的感受野大小, $\text{kernal}_{\text{size}}$ 代表当前最大池化核的大小, stride 代表最大池化的步长。由上述公式可知,SPP能通过使用多个不同大小的最大池化层来增大模型的感受野,以此来捕获更多的上下文信息,从而增强模型的检测性能。

1.4 PAN

PAN模块是一个能将低级定位特征与高级语义特征进行反复聚合的网络结构,其主要借鉴了特征提取路径扩张与聚合的思想。PAN能分别从自顶向下的扩张路径和自底向上的聚合路径对不同尺度的特征进行传递与融合,为模型提供强语义信息和精准的定位信息,以此来提高模型的检测性能。PAN模块位于主干特征提取网络BottleneckCSP和SPP模块之后,其输入的特征图分别来自于3个不同层次的有效特征提取层,特征图的大小分别为 76×76 、 38×38 和 19×19 。PAN模块是一个包含有上采样、下采样和特征融合的模块,它由大量的上采样块、下采样块、跳

跃连接与卷积核大小分别为 3×3 和 1×1 的卷积层构成。在PAN的特征扩张路径中,采用最近邻插值的方式对输入的特征图进行上采样操作,在特征图成倍扩大的同时,深层的强语义特征不断被自顶向下地传递。随后,在PAN的特征聚合路径中,采用卷积核大小为 3×3 、步长为2的卷积层对尺寸较大特征图进行下采样操作以充分提取图像的高级语义特征。与此同时,得益于特征聚合路径中的跳跃连接与特征融合结构,浅层的边缘等特征被较好地保留下来并与深层的强语义特征进行融合。该操作将为模型提供精确的定位信息和语义信息,从而提高模型对病灶的检测能力。PAN通过嵌入多个跳跃连接的方式,有效地避免了由于神经网络加深而导致的浅层特征丢失和网络难以训练的问题,其结构立体图如图6所示。

值得注意的是,相较于Shortcut特征融合方式,在PAN模块中采用了与其不一样的特征融合策略——Concat特征融合方式^[18]。在对两个特征图进行融合时,Shortcut是对两个特征图对应位置上的各个元素进行逐一相加,并不会增加特征图的通道数;而Concat则是在通道的维度上将两个特征图进行拼接,使得融合后的特征图通道数增加。相比于Shortcut

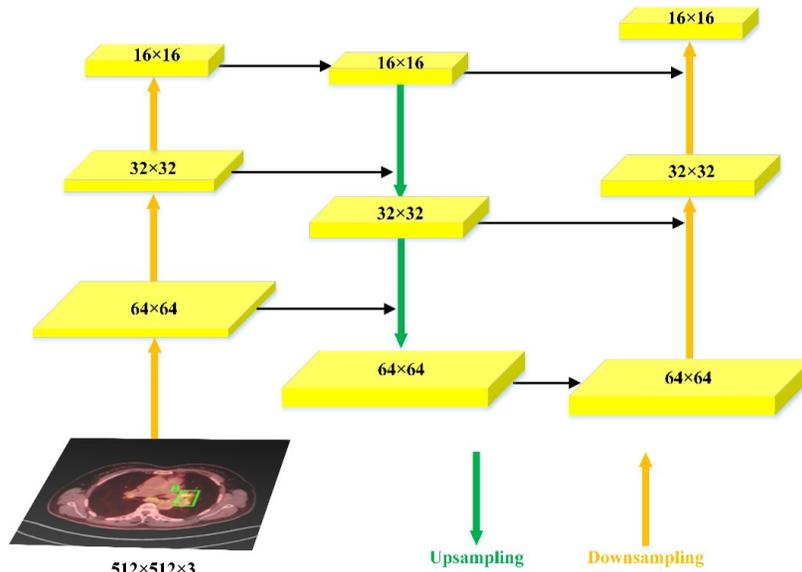


图 6 PAN 模块结构图
Figure 6 PAN module structure

特征融合,Concat 会使得特征图的深度加深,从而有利于模型对不同尺度特征图的语义信息加以利用。

Shortcut 与 Concat 特征融合方式的可视化插图如图 7 所示。

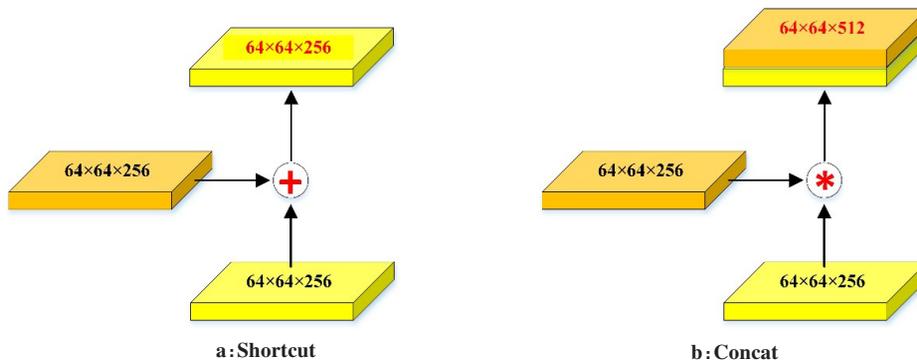


图 7 Shortcut 与 Concat 特征融合方式的可视化插图
Figure 7 Visual illustrations of Shortcut and Concat feature fusion approaches

1.5 ACSLoss

为了追求尾部分类器的均衡训练,有几篇文章提出了基于类别频率先验将整个类别划分为若干组,并对不同的类别采用不同的训练策略。例如,均衡化损失忽略头部类别样本的负梯度,以防止尾部分类器被负信号所主导。平衡组 Softmax 首先将样本数相近的类别放入同一组,并在每组内应用 Softmax 交叉熵损失。但是对于公开数据集来说,这并不是一个好的类别分区,因为类别少同时每个类别样本偏多。为新数据集找到合适的组分区策略可能是费力和耗时的,这将限制基于组的方法在现实场景中的应用。

ACSLoss 不需要进行组划分,也不依赖于类别的先验频率,它可以无缝地与不同的数据集结合。自

适应类抑制损失有以下两个优点:(1)尾部分类器应避免被大量的头部样本过度抑制,同时对于容易混淆的类别,要保持负梯度以进行判别学习;(2)Loss 函数最好不依赖于标签分布,这样就可以无缝地应用于不同的数据集,而无需重新计算新的类别分布统计信息。

假设给定属于头部类别 m 的一个样本 x_i ,其标签为 one-hot 的 Y_i 。其中 y_m 值为 1, $y_t(t \neq m)$ 的值为 0。深度学习模型对于样本 x_i 的预测输出值为 Z_i ,其中 z_t 表示类别为 t 的 logit 值。将 z_t 输入式(7)的 Sigmoid 激活函数可以得到当前样本属于类别 t 的置信 p_t 。交叉熵损失函数^[19]如式(8)所示,其中 \hat{p}_t 的取值如式(9)所示:

$$p_t = \frac{1}{1 + e^{-z_t}} \quad (7)$$

$$L_{CE}(x_i) = -\sum_{t=1}^c \log(\hat{p}_t) \quad (8)$$

$$\hat{p}_t = \begin{cases} p_t, & t = m \\ 1 - p_t, & t \neq m \end{cases} \quad (9)$$

$$\frac{\partial L_{CE}}{\partial z_t} = \begin{cases} p_t - 1, & t = m \\ p_t, & t \neq m \end{cases} \quad (10)$$

因此,自适应类别抑制损失来保护尾部分类器的训练,并根据学习状态自适应地选择哪些类别应该被抑制。如式(11)所示,将一个二类权重项 w_t 与式(8)中的损失项 $-\log(\hat{p}_t)$ 相乘。对于类别 m ,由于当前类别样本属于类别 m ,所以 w_m 的值设为1。对于其他类别 $t(t \neq m)$, w_t 可以控制是否对类别 t 进行抑制。选择置信度 p_t 来判断是否抑制第 t 类。如果 p_t 大于预设的阈值 $\gamma(\gamma=0.7)$,那么这就表示模型混淆在类别 t 和 m 之间。因此,将 w_m 设置为1来帮助模型判别正确类别。然后将 w_t 设置为0,避免产生大量不必要的负面抑制。因此,ACSLoss 不依赖于类别分布统计数据,而只依赖于网络输出置信度,这节省了切换到新数据集时寻找最优类别统计相关超参数的努力。公式定义如下:

$$L_{ACSL}(x_i) = -\sum_{t=1}^c w_t \log(\hat{p}_t) \quad (11)$$

$$w_t = \begin{cases} 1, & t = m \\ 1, & t \neq m \text{ and } p_t \geq \gamma \\ 0, & t \neq m \text{ and } p_t < \gamma \end{cases} \quad (12)$$

ACSLoss 对 z_t 可以导出为式(13):

$$\frac{\partial L_{ACSL}}{\partial z_t} = \begin{cases} p_t - 1, & t = m \\ w_t p_t, & t \neq m \end{cases} \quad (13)$$

2 数据集

2.1 数据集来源

为了验证 ACS_YOLO 模型的实际性能,本研究使用 A Large-Scale CT and PET/CT Dataset for Lung Cancer Diagnosis(Lung-PET-CT-Dx)这个公开数据集进行全面的实验。该数据集共包含 355 位患者总共 1 295 张 CT 和 PET/CT 序列图像。根据组织病理学诊断将受试者分成 4 组,分为腺癌(A)、小细胞癌(B)、大细胞癌(E)和鳞状细胞癌(G)。这些患者接受了标准的肺活检和 PET/CT 检查,对纵隔(window width, 350 HU; level, 40 HU)和肺(window width, 1 400 HU; level, -700 HU)的图像进行分析。CT 层间间隔为 0.625~5.000 mm,扫描方式包括平扫、对比和三维重建。检查前空腹至少 6 h,每位患者血糖均小于 11 mmol/L。静脉注射 ^{18}F -FDG(4.44 MBq/kg, 0.12 mCi/kg) 60 min

后,取患者仰卧位在 PET/CT 扫描仪(西门子 mCT 64 排)上进行全身扫描。FDG 剂量和摄取次数分别为 168.72~468.79 MBq [(295.8±64.8) MBq] 和 27~171 min [(70.4±24.9) min]。 ^{18}F -FDG 放射化学纯度为 95%。在 PET 和 CT 采集过程中允许患者正常呼吸。利用 CT 数据采用混合分割方法对 PET 图像进行衰减校正。衰减校正使用 CT 协议(180 mAs, 120 kV, 1.0pitch)进行。每项研究包括一个 CT 容积,一个 PET 容积和融合的 PET 与 CT 图像:1 mm×1 mm 处的 CT 分辨率为 512×512 像素,4.07 mm×4.07 mm 处的 PET 分辨率为 200×200 像素,层厚 1 mm,层间距离 1 mm。两个体量用相同的切片数进行重建。从颅骨基部到股骨中部进行三维发射和透射扫描。采用 TrueX TOF 方法重建 PET 图像,切片厚度为 1 mm。

每个肿瘤的位置由 5 名具有肺癌专业知识的胸科放射科医师标注,保证数据集成为开发医疗诊断算法的有效工具和资源。其中两名放射科医生有超过 15 年的工作经验,其余的均有超过 5 年的工作经验。在其中 1 名放射科医生标记 1 位受试者后,其他 4 名放射科医生需要对该标注进行验证,保证所有 5 名放射科医生都检查了数据集中的每个标注结果。

2.2 数据集设置

根据官方提供的标注结果,先从原始的序列图像中按横截面切分提取出所有包含标注的二维切片,并转成 png 格式的图像。最终得到 30 324 张二维图像,图像尺寸为 512×512 的新数据。数据集中的图像和官方标注示例如图 8 所示。

在所有数据集中,腺癌总共有 19 885 张,占有所有样本的 65.57%,提供了绝大多数样本。鳞状细胞癌和小细胞癌分别有 7 216 张和 3 022 张,占比分别为 23.80% 和 9.97%。特殊的是大细胞癌,只有 201 张,占比仅为 0.67%。因此该数据集符合长尾分布情况,其中腺癌为头部类别,提供了数据集里绝大部分的样本。大细胞癌为尾部类别,出现频率很低。

使用 skleran 库中的 StratifiedKFold 方法将数据集按照 60%:20%:20% 的比例随机分为训练集、验证集和测试集。训练集用来训练模型参数权值,验证集用来验证模型效果和选择模型参数权值,测试集则用来测试模型的实际性能。StratifiedKFold 方法会根据原始数据集中的样本标签分布比例来切分样本,使得切分后的训练集、验证集和测试集中各样本的标签比例和原始数据集近似,均符合长尾分布。

2.3 Mosaic 数据增强

Mosaic 数据增强方法是由 Bochkovskiy 等^[20]在 2018 年提出的,它在理论上与 CutMix 数据增强^[21]具有一定的相似性,均是对多张训练集的图像进行拼

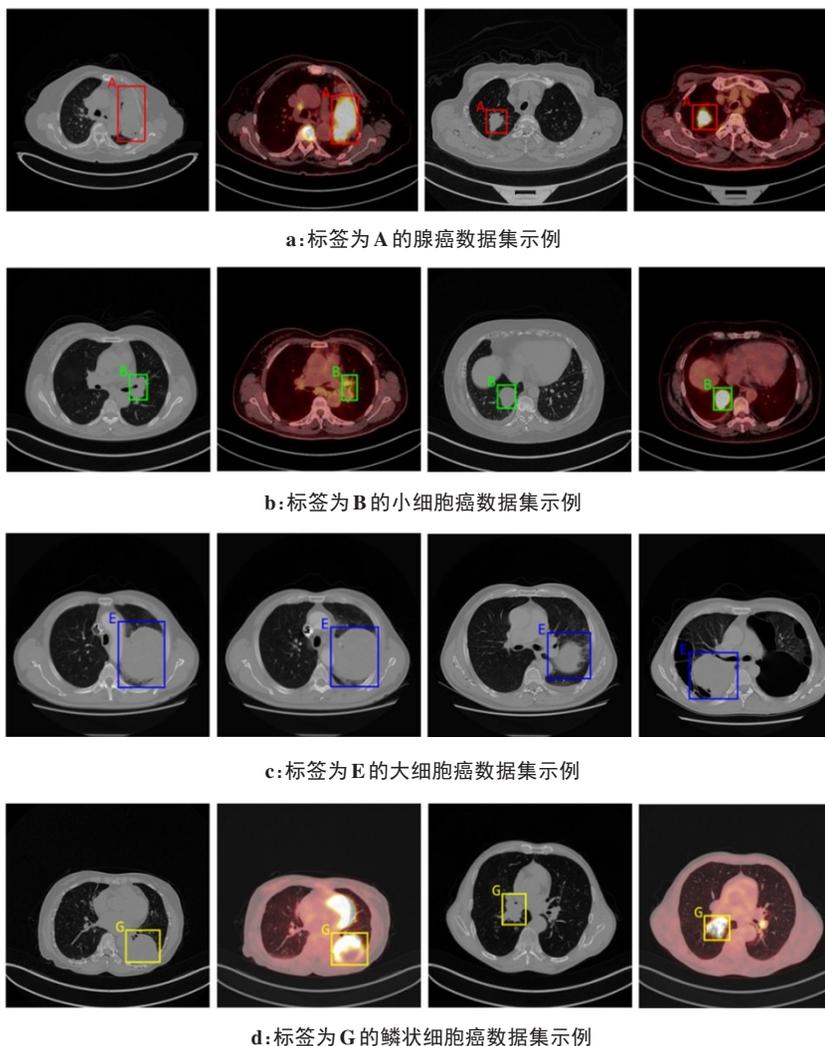


图 8 数据集图像示例

Figure 8 Examples from data sets

接组合。CutMix 数据增强是将两张训练集图像进行随机拼接,而本文采用的 Mosaic 数据增强则是一性将 4 张训练集图像按照预先设置好的流程进行拼接组合,两者的可视化结果对比如图 9 所示。其中,标签 0 对应腺癌,标签 1 对应鳞状细胞癌和小细胞癌,标签 2 对应大细胞癌。Mosaic 数据增强的实现步骤如下:(1)从训练集中随机选取 4 张不重复的图像;(2)对 4 张图像分别进行左右翻转、随机缩放和色域变换等处理;(3)将处理后的图像按照预先规定好的顺序依次拼接成一张训练图像。

与以往的数据增强方法相比,Mosaic 数据增强允许模型检测正常检测环境以外的目标,增加了单张训练图像的可检测目标个数,有效地增强了训练数据的多样性。此外,Mosaic 数据增强能使模型在仅接收一张训练图像的情况下,即可一次性计算 4 张图像的均值和方差,使得计算而来的均值和方差更加接近于整个训练集图像的均值和方差,在一定程

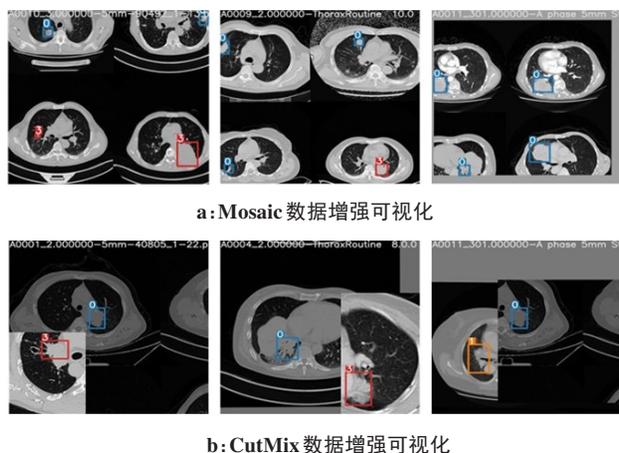


图 9 Mosaic 与 CutMix 数据增强可视化结果对比

Figure 9 Comparison of Mosaic and CutMix data augmentations

度上解决了由于设备资源限制而带来的样本均值和方差难以准确估计的问题。

3 结果

3.1 评估指标

在测试阶段,采用精确度(Precision)、召回率(Recall)、mAP@0.5、mAP@0.5:0.95对结果进行定量分析。精确度(Precision)表示被分为正例中实际为正例的比例,如式(14)所示。召回率(Recall)表示在正例中有被正确预测的比例,如式(15)所示。IoU为模型产生的目标窗口和原来标记窗口的交并比,即检测结果(最终生成的框)与真实值 Ground Truth 的交集除以它们的并集,如式(16)所示:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (14)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (15)$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \quad (16)$$

其中,TP(True Positives)表示正类判断为正类的数量,FP(False Positives)表示负类判断为正类的数量,FN(False Negatives)表示正类判定为负类的数量,TN(True Negatives)表示负类判定为负类的数量。

平均精度(Average Precision, AP)是计算单类别的模型平均准确度。对于目标检测任务,每一个类都可以计算出其 Precision 和 Recall,每个类都可以得到一条 P-R 曲线,曲线下的面积就是 AP 的值。mAP@0.5(mean of Average Precision, IoU=0.5)即将 IoU 设为 0.5 时,计算每一类的所有图片的 AP,然后所有类别求平均。mAP@0.5:0.95 则表示在不同 IoU 阈值(从 0.5 到 0.95,步长 0.05)上的平均 mAP 值。

3.2 实验设置

本文的实验环境为 Ubuntu18.04LTS 系统、Intel Core i9-10900X、NVIDIA RTX 2080Ti 显卡,显卡显存 11 GB,内存 32 GB。在 Python 版本 3.8、CUDA 版本 10.1 的环境下配置了 Pytorch1.8.0 的开源深度学习框架。在实验中采用多种数据增强技术,包括随机上下、左右翻转,随机缩放裁剪,随机旋转,随机水平和垂直平移,Mosaic 和 Mixup。本研究选择 YOLOv5 作为基准模型,分别使用 YOLOv5s、YOLOv5m、YOLOv5x 3 个不同深度、不同宽度的模型进行实验。YOLOv5s 网络在 YOLOv5 系列里深度最小,特征图的宽度也是最小的。YOLOv5m 相比 YOLOv5s 网络更深更宽,而 YOLOv5x 则是深度最大,特征图的宽度最大的网络,模型性能也是最好的。在训练时,使用 Adam 优化器^[22]优化损失函数。训练参数设置为:初始学习率为 0.000 5,学习率动量设置为 0.843,权重衰减系数为 0.000 36。预热学习的 epoch 轮数为 2 轮,预热学习率和预热学习率

动量分别为 0.05 和 0.50。批大小为 16,最大迭代 100 次。其他超参数均使用 YOLOv5 的默认设置。在测试时,选择置信度为 0.8 来提取检测框,并选择 IoU 阈值为 0.5 进行非极大值抑制去除多余检测框。之后使用自适应类抑制损失函数 ACSL 替换 YOLOv5 原始的分类损失函数进行实验,并与原版 YOLOv5 进行比较。其中自适应类抑制损失函数 ACSL 的阈值 γ 设置为原论文里的 0.7,其他超参数与基准模型时一样,并在表 1 中展示实验结果。

3.3 结果

如表 1 所示,ACS-YOLO 在相同条件下的各项指标均比原始 YOLO 提高约 1%~5%。并且,当 ACSL 与更加深的网络模型结合时,带来了更加显著的提升。最大的模型 ACS-YOLO-x 取得了最佳效果,达到了精确度 =0.960 7,召回率 =0.948 9, mAP0.5=0.970 6, mAP0.5:0.95=0.558 3。相比原始 YOLOv5x, ACS-YOLO-x 的精确度、召回率、mAP@0.5、mAP@0.5:0.95 分别提升约 4.4%、5.2%、2.9%、4.2%。而最小的模型 ACS-YOLO-s 相比于初始 YOLOv5s 的提升约 1.1%、1.8%、1.6%、0.4%。以上结果表明 ACSL 能够解决长尾数据集中的类不平衡问题。

表 1 ACS-YOLO 与原始 YOLOv5 检测性能的比较
Table 1 Comparison of the detection performance of ACS-YOLO with the original YOLOv5

模型	精确度	召回率	mAP0.5	mAP0.5:0.95
YOLOv5s	0.863 1	0.850 3	0.899 4	0.473 4
YOLOv5m	0.907 5	0.866 2	0.924 5	0.493 0
YOLOv5x	0.916 6	0.896 5	0.941 8	0.516 7
ACS-YOLO-s	0.874 4	0.868 6	0.915 4	0.477 7
ACS-YOLO-m	0.935 6	0.913 3	0.953 6	0.522 1
ACS-YOLO-x	0.960 7	0.948 9	0.970 6	0.558 3

为了进一步验证 ACSL 的有效性,分别选择 YOLOv5x 和 ACS-YOLO-x 模型,统计两个模型对于不同类别的预测。结果如表 2 所示,其中 YOLOv5x 的头部类别腺癌的检测效果非常好,但是作为尾部类别的小细胞癌和大细胞癌精确度仅有 0.884、0.879,召回率为 0.864、0.822, map@0.5 只有 0.924 和 0.892,各项指标远低于头部类别腺癌。而 ACS-YOLO-x 模型对小细胞癌和大细胞癌这两个尾部类别的检测性能提升非常显著,提升幅度约为 5%~11%。头部类别腺癌和鳞状细胞癌的检测性能仅有轻微降低,但是所有类别的总体性能有较大幅度的提升。

训练过程如图 10 所示,分别对验证集的精确

表2 ACS-YOLO-x与YOLOv5x对各类类别检测结果的比较

Table 2 Comparison of ACS-YOLO-x and YOLOv5x detection results for various categories

模型	类别	精确度	召回率	mAP@0.5	mAP@0.5:0.95
YOLOv5x	所有类	0.917	0.897	0.942	0.517
	腺癌	0.930	0.931	0.965	0.554
	小细胞癌	0.884	0.864	0.924	0.499
	大细胞癌	0.879	0.822	0.892	0.466
	鳞状细胞癌	0.973	0.968	0.987	0.548
ACS-YOLO-x	所有类	0.961	0.949	0.971	0.558
	腺癌	0.927	0.901	0.954	0.542
	小细胞癌	0.970	0.958	0.973	0.578
	大细胞癌	0.989	0.983	0.981	0.547
	鳞状细胞癌	0.959	0.952	0.974	0.567

度(Precision)、召回率(Recall)、mAP@0.5。训练结果表明,ACS-YOLO模型在第50轮 epoch左右的各

个损失函数都已收敛,同时精确度、召回率、mAP@0.5、mAP@0.5:0.95达到最高。

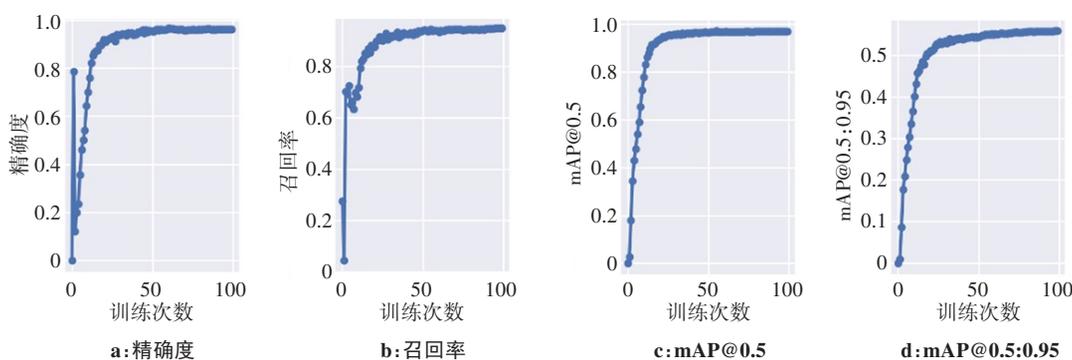


图10 ACS-YOLO模型训练曲线图

Figure 10 ACS-YOLO model training curves

图11展示了使用ACS-YOLO-x对测试集图像的部分检测结果。其中第1、3列为测试集图像的真实标签,第2、4列为ACS-YOLO-x的检测结果,检测框中的数字代表模型的预测置信度。从图中可以看出ACS-YOLO能够有效检测与定位病灶区域。图12展示了ACS-YOLO的错误检测结果,可以看出模型的错误主要是多检测,当病灶区域比较小的时候,模型虽然基本能够检测到病灶区域,不过往往会伴随着其他错误的检测结果。

4 结论

当前,虽然有很多工作将深度学习目标检测方法应用到了肺癌的检测研究中,但是其研究大多数是在数据分布平衡的数据集,如LIDC-IDRI数据集上进行的,其结果在这类数据平衡的数据集上表现良好。然而,在现实情况下,数据集的分布为非平衡状

态,如:Lung-PET/CT-Dx数据集等。Lung-PET/CT-Dx数据集中某一头部类别贡献了整个数据集中的大多数类别,而某些尾部类别可能只有几十个样本量。这会导致模型过度拟合头部类别,而降低了对尾部类别的检测性能。为了解决上述问题,本研究将ACSLoss和YOLOv5结合起来,设计一种可以有效处理长尾分布数据集的ACS-YOLO方法。本文提出的ACS-YOLO具有以下优点:(1)通过残差结构加速网络学习的速度,避免网络退化的现象同时使网络学习到更多肺癌病灶的特征,而CSP结构减少网络的参数,加速整个模型的训练速度同时提高网络识别病灶的准确率;(2)SPP模块通过采用多个尺寸不一的最大池化层来实现对特征图的多尺度特征提取与融合,对待检测的肺癌病灶具有平移和尺度不变性,能较好地捕获更多肺癌病灶的上下文信息并增强模型的特征表达能力;(3)PAN模块可以分别从自顶向

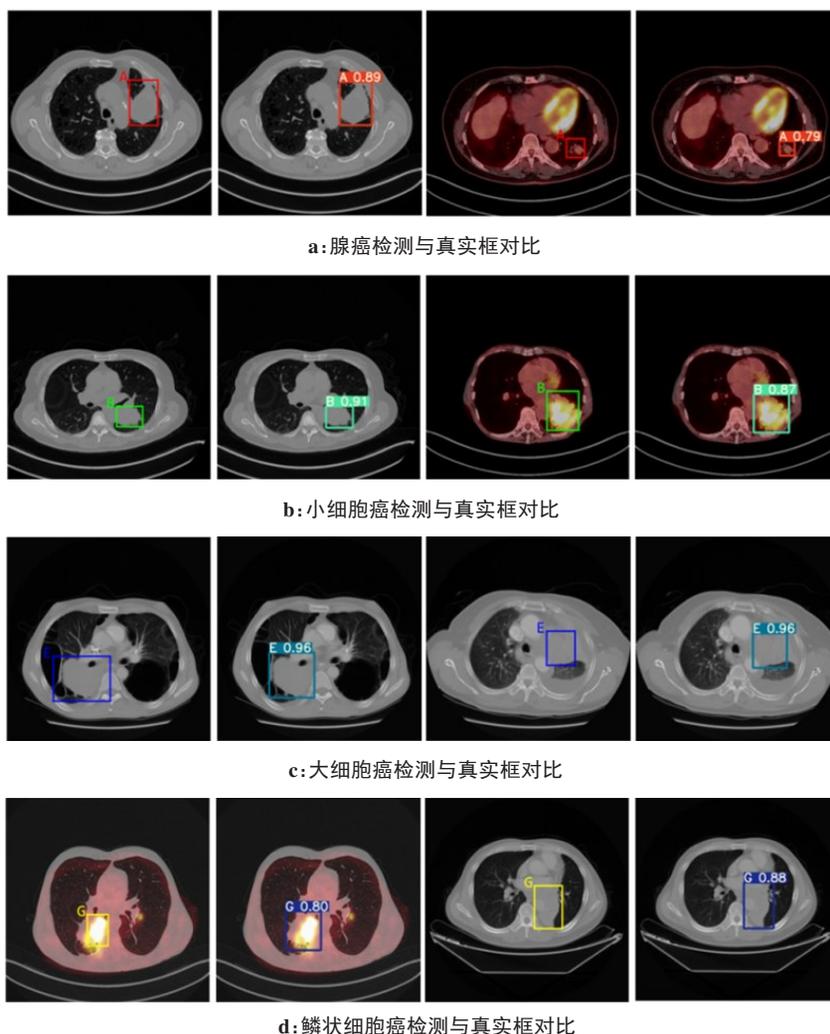


图11 ACS-YOLO检测结果与真实标签图像

Figure 11 ACS-YOLO detection results with true labeled images

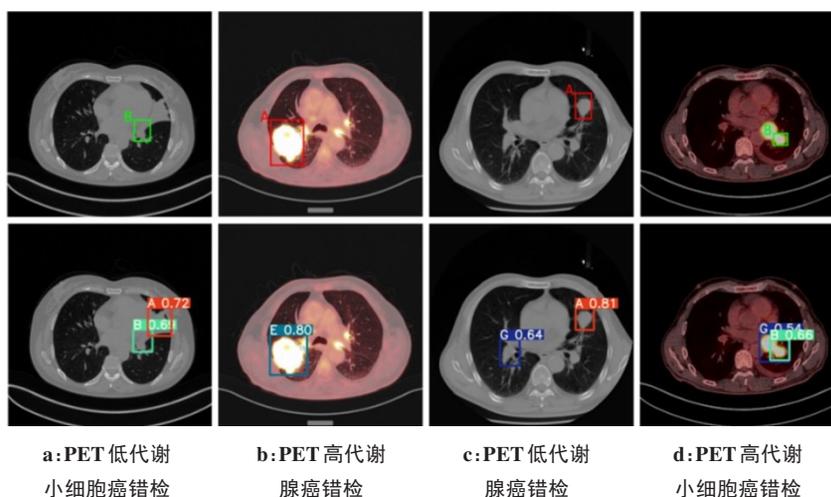


图12 ACS-YOLO错误检测结果

Figure 12 ACS-YOLO error detection results

下的扩张路径和自底向上的聚合路径对不同尺度的特征进行传递与融合,其优点在于能将浅层的低级定位特征与深层的高级语义特征进行反复聚合,为模型检测肺癌病灶提供精准的定位信息和语义信

息,从而提高模型对肺癌病灶的检测性能;(4)将ACS-Loss应用到YOLO上,使得模型能够防止ACS-YOLO模型过度拟合头部类别,提高模型对于尾部类别的检测精度,很好地解决了在实际场集中数据集

长尾分布的问题。

通过在 Lung-PET/CT-Dx 数据集验证了本文方法的有效性。相比于原始YOLO,本文提出的ACS-YOLO-X检测精确度为0.960 7,召回率为0.948 9,mAP0.5为0.970 6,mAP0.5:0.95为0.558 3,总体提升4%左右。同时,ACS-YOLO模型对于尾部类别的检测性能有极大的提升,普遍提升约5%~11%。

ACS-YOLO在长尾分布数据集上取得了比原始YOLO模型更加优秀的性能表现,对于尾部类别的检测性能也提升明显。但是该模型对于头部类别的检测性能却有轻微下降,在后续的工作里,将会研究如何在保证模型对于尾部类别的鉴别能力的情况下,不过度抑制模型对于头部类别的特征学习,以避免出现头部类别检测性能降低的问题。针对模型对于检测较小的目标时出现的多检测问题,计划引入双注意力机制来提升模型对于病灶区域位置的定位能力,并尝试引入transformer结构提高肺癌检测的效果。

【参考文献】

- [1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 580-587.
- [2] Ren SQ, 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.
- [3] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]. CVPR, 2016: 779-788.
- [4] Singh B, Davis LS. An analysis of scale invariance in object detection snip[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3578-3587.
- [5] 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.
- [6] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8759-8768.
- [7] Zhang J, Xia Y, Zeng H, et al. NODULe: combining constrained multi-scale LoG filters with densely dilated 3D deep convolutional neural network for pulmonary nodule detection[J]. Neurocomputing, 2018, 317: 159-167.
- [8] Redmon J, Farhadi A. Yolov3: an incremental improvement[J]. arXiv preprint arXiv: 1804. 02767, 2018.
- [9] Grover A, Leskovec J. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 855-864.
- [10] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems. 2012: 1097-1105.
- [11] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//International Conference on Machine Learning. PMLR, 2015: 448-456.
- [12] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. J Mach Learn Res, 2014, 15(1): 1929-1958.
- [13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv: 1409. 1556, 2014.
- [14] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [15] Wang CY, Liao HY, Wu YH, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 390-391.
- [16] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2117-2125.
- [17] Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2015, 38(1): 142-158.
- [18] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.
- [19] De Boer PT, Kroese DP, Mannor S, et al. A tutorial on the cross-entropy method[J]. Ann Oper Res, 2005, 134(1): 19-67.
- [20] Bochkovskiy A, Wang CY, Yolo HY. Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2004. 10934, 2020.
- [21] Yun S, Han D, Oh SJ, et al. Cutmix: regularization strategy to train strong classifiers with localizable features [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6023-6032.
- [22] Kingma DP, Ba J. Adam: a method for stochastic optimization[J]. arXiv preprint arXiv: 1412. 6980, 2014.

(编辑:陈丽霞)