

基于改进YOLO V3的肺结节检测方法

王乾梁^{1,2}, 石宏理^{1,2}

1. 首都医科大学生物医学工程学院, 北京 100069; 2. 首都医科大学临床生物力学应用基础研究北京市重点实验室, 北京 100069

【摘要】针对肺结节占CT图像比例小、形状不规则及直接应用YOLO V3算法进行肺结节检测效果不佳的问题,提出基于改进YOLO V3的肺结节检测方法。首先进行重采样和肺实质分割等预处理操作。然后修改YOLO V3的基础网络结构,调整骨干网络和检测网络的结构单元数量;使用Mish激活函数替换Leaky ReLU激活函数,引入含有空洞卷积的感受野模块层;修改损失函数。最后使用改进的YOLO V3方法进行肺结节检测,完成对比实验。在LIDC-IDRI数据集上得到了88.89%的准确率和94.73%的高敏感度,实验结果表明该方法能够有效检测肺结节。

【关键词】深度学习;肺结节;YOLO V3;卷积神经网络

【中图分类号】R318;TP391.41

【文献标志码】A

【文章编号】1005-202X(2021)09-1179-06

Pulmonary nodule detection based on improved YOLO V3

WANG Qianliang^{1,2}, SHI Hongli^{1,2}

1. School of Biomedical Engineering, Capital Medical University, Beijing 100069, China; 2. Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing 100069, China

Abstract: In view of small proportion of pulmonary nodules in CT images, irregular shapes of pulmonary nodules and unsatisfactory results obtained by the direct application of YOLO V3 algorithm for pulmonary nodule detection, a pulmonary nodule detection method based on improved YOLOV3 is proposed in the study. Preprocessing such as resampling and parenchymal segmentation is carried out. Then the basic network structure of YOLO V3 is modified, and the number of structural units of the backbone network and the detection network is adjusted. Leaky ReLU activation function is replaced by Mish activation function, and receptive field block layers with dilated convolutions are added. Moreover, the loss function is modified. Finally, the improved YOLO V3 is used to detect pulmonary nodules, and the comparative experiment is completed. The proposed method is tested on LIDC-IDRI data set, and the results show that the improved YOLO V3 achieves an accuracy of 88.89% and a sensitivity of 94.73%, indicating that the proposed method can effectively detect pulmonary nodules.

Keywords: deep learning; pulmonary nodule; YOLO V3; convolutional neural network

前言

根据2019年国家癌症中心发布的全国癌症病情统计数据,肺癌的发病率和死亡率仍位居首位,肺癌及相关恶性肿瘤的治疗花费超过2 200亿元,防控形势十分严峻^[1]。早期诊断和早期治疗是降低肺癌死亡率的有效途径。肺结节作为早期肺癌的重要病征,准确检测肺结节具有极其重要的临床意义。临床上一般通过观察肺部CT图像进行肺结节的诊断

判别,但随着医疗需求的快速增长,CT图像数据量不断增加,医师的诊断难度也不断提高,凭借主观经验判断容易出现错误诊断。因此在CT图像诊断基础上,利用人工智能技术辅助医生做出更为客观、准确的判断,显得尤为必要。

传统的计算机辅助肺结节检测方法主要使用人工提取的特征进行训练,存在步骤繁琐、检测精度不高和系统整体性能较差等问题。随着机器学习领域的快速发展,以卷积神经网络为代表的深度学习方法广泛应用于目标检测领域:一种是基于候选区域的两阶段检测算法,如Faster RCNN^[2]和Mask RCNN^[3],另一种是基于回归的一阶段检测算法,如YOLO系列^[4-6]和SSD^[7],目前也有研究在尝试将两种方法结合起来使用^[8]。其中,YOLO V3作为YOLO系列的代表算法,在工程应用领域使用广泛,展现出

【收稿日期】2020-11-14

【基金项目】北京市自然科学基金(7142022)。

【作者简介】王乾梁,硕士,研究方向:医学图像处理, E-mail: qianliang@ccmu.edu.cn

【通信作者】石宏理,副教授,研究方向:医学图像处理, E-mail: shl@ccmu.edu.cn

良好的效果。但在应用YOLO V3算法处理肺部CT图像时,存在目标检测精度不高等问题,难以达到理想效果。针对以上问题,本研究提出基于改进YOLO V3的肺结节检测方法,探究应用YOLO V3进行肺结节检测的有效性和可行性。

1 YOLO V3算法

YOLO V3的输入图像尺寸会被缩放或填充到416×416,再送入网络提取特征。YOLO V3的网络结构如图1所示,主要由CBL模块和残差单元(ResUnit)两个结构单元组成;CBL模块由卷积层、批归一化(Batch Normalization, BN)层^[9]和Leaky ReLU激活函数层构成,具体结构如图2a所示;残差单元主要由借鉴Resnet^[10]思想构建的残差模块(Residual block)组成,残差模块具体结构如图2b所示。1个CBL模块和多个残差模块组成1个残差单元,残差单元标注的数字代表包含的残差模块数量,每个残差单元进行一次下采样。

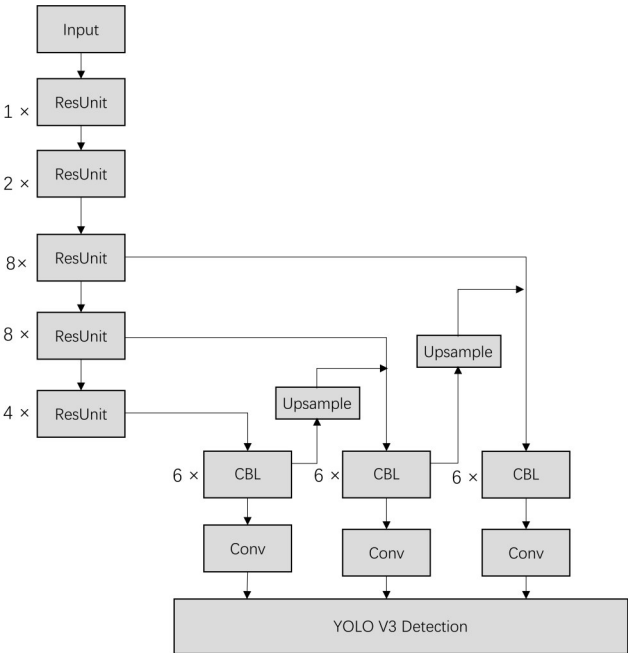


图1 YOLO V3网络结构
Fig.1 YOLO V3 network structure

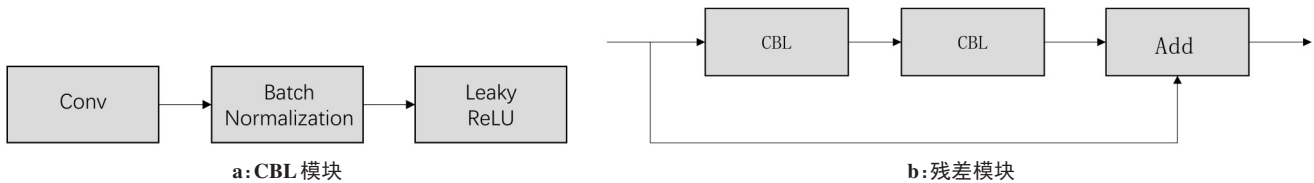


图2 YOLO V3的结构单元
Fig.2 Structural unit of YOLO V3

(Receptive Field Block, RFB)层,融合不同尺度特征并扩大特征图感受野,提升检测精度。二是修改

YOLO V3在检测网络上使用特征金字塔网络 (Feature Pyramid Network, FPN)^[11]结构将不同尺度的特征图融合,并在5次下采样中后3次输出的特征图上进行检测。YOLO V3将图像分别划分为13×13、26×26、52×52共3种尺度的网格单元,每个网格单元负责预测中心坐标落入其中的目标类别。同时,YOLO V3算法使用锚框(Anchor)机制,通过K均值聚类(Kmeans)算法在3种尺度上各预设3个先验框,每种尺度的每个网格单元使用先验框进行预测。生成的大量预测框使用非极大值抑制(Non-Maximum Suppression, NMS)^[12]算法进行筛选,从而得到目标检测框。因为YOLO V3在检测准确度和检测速度两方面都有很好表现,实现步骤简单,能够满足实时检测要求,所以YOLO V3在工程应用领域常作为首选算法。

2 改进的YOLOV3模型

由于CT图像中肺结节所占面积较小、样本数据不平衡且形状不规则,直接使用YOLO V3算法应用于肺结节检测时,会产生小目标定位不精确和检测准确率低等问题。在进行网络模型训练时,较为复杂的网络结构和3个尺度特征图生成的大量预测框,都会产生庞大的计算量,影响网络模型的训练过程。在使用深度学习方法解决肺结节这样的小目标检测问题时,需要使用高分辨率的特征图进行预测,对于肺结节的不规则形状则需要提取不同尺度的特征信息和扩大浅层特征图感受野,以丰富提取的特征信息。同时,提升肺结节目标预测框的定位精度也尤为关键。基于以上几点的改进思路,本研究针对YOLO V3提出的改进方法主要从两个方面进行:一是针对肺结节的目标检测任务修改YOLO V3的基础网络结构,在骨干网络上增减残差单元中残差模块数量,提取更多的小目标特征信息;修改YOLO V3的检测网络结构,只使用8倍下采样的输出特征图作为检测尺度,降低训练时产生的计算量;使用Mish激活函数^[13]替换Leaky ReLU激活函数,提高网络模型的准确性和泛化性;加入含有空洞卷积的感受野模块

YOLO V3的损失函数,使用CIoU损失函数^[14]提升边框回归的收敛速度,提高预测框的定位精度。

2.1 改进的基础网络结构

对于YOLO V3的骨干网络,为了降低网络深度减少参数计算量、防止训练过程中产生过拟合问题,并提取更多的小目标特征信息,本研究增加了第2个残差单元的残差模块数量,减少了第3、4、5个残差单元的残差模块数量。同时本研究考虑到训练数据集的图像尺寸,将网络模型的图像输入尺寸设置为512×512,不再对图像尺寸进行缩放或者填充。对于检测网络,YOLO V3使用32倍和16倍下采样输出特征图对肺结节进行检测时,语义信息丢失严重,不能准确定位肺结节。考虑到使用不同分辨率特征图,生成预测框时所产生的参数计算量和内存使用量,本研究选择使用8倍下采样的输出特征图作为检测尺度,并使用Kmeans算法在训练数据集上计算得到3个预设先验框。

2.2 使用Mish激活函数

神经网络模型中的激活函数负责将神经元的输入映射到输出端,增加非线性变化,提高网络模型的表达能力,从而能够实现深度网络结构。本研究选择使用的激活函数为Mish激活函数,其表达式如式(1)所示。结合图3可知,Mish函数是一种非单调的激活函数,不会产生梯度消失,较Leaky ReLU激活函数更为平滑,具有更好的准确度和泛化能力。

$$f(x) = x \times \tanh(\log(1 + e^x))$$

(1)

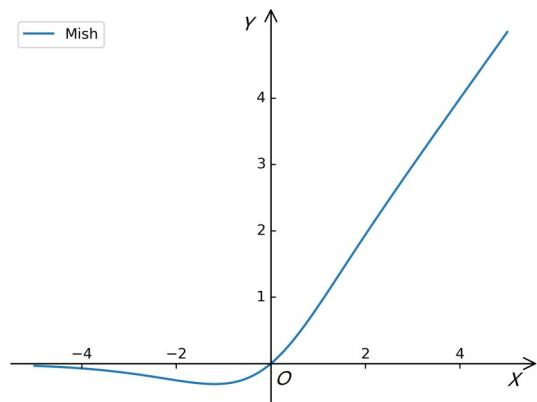


图3 Mish激活函数曲线
Fig.3 Mish activation function

使用Mish激活函数后组成新的网络结构单元CBM模块,由卷积层、批归一化层和Mish激活函数层构成,如图4所示。

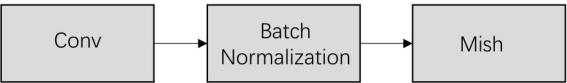


图4 CBM模块
Fig.4 CBM block

2.3 改进的RFB层

为了增强特征信息的提取,本研究在网络结构中引入RFBNet算法^[15]提出的RFB层。RFB在空间金字塔池化(Spatial Pyramid Pooling, SPP)^[16]模块的实现思路,模仿人类视觉系统进行设计,能够融合不同尺度的特征图并扩大输出特征图的感受野,获取更为全面的特征信息,提升判断结节位置的准确性。RFB结构如图5所示,基于Inception模块^[17]的网络结构,加入了不同膨胀率的空洞卷积^[18]。在分支中使用的卷积核越大其空洞卷积的膨胀率越大,其中膨胀率代表卷积核中的间隔距离。RFB首先在各个分支上使用1×1卷积层降低特征图的通道数,3个分支的空洞卷积的膨胀率依次为1、3、5。然后通过Concat拼接操作将不同尺度特征图融合,最后使用1×1卷积层和Add拼接操作实现跨通道的特征融合。

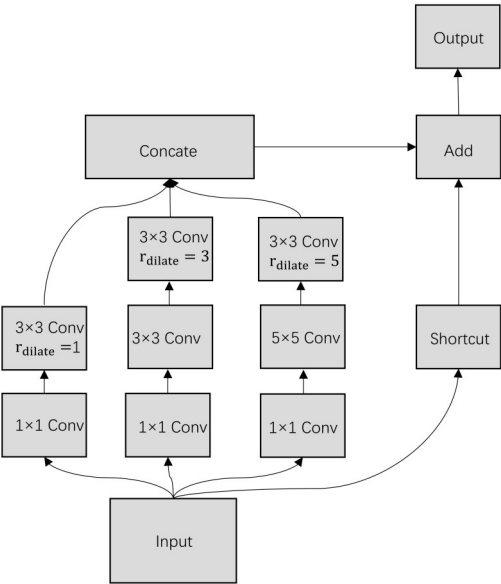


图5 RFB层结构
Fig.5 Receptive field block structure

本研究在应用时将第2个分支中1×1卷积层后的3×3卷积层替换为1×3和3×1卷积层,第3个分支中1×1卷积层后的5×5卷积层替换为两个3×3卷积层,使其在不改变感受野的情况下,减少参数量并增加非线性映射能力,有效提升网络的推理能力。

综合以上改进思路,本研究所提出方法使用的网络模型结构如图6所示。

2.4 改进的损失函数

损失函数在深度学习方法中用来评估预测的输出值与实际值之间的差距,以提升算法的运行情况。YOLO V3的损失函数对于预测框回归损失值、目标置信度损失值和目标类别损失值进行计算。

在进行预测框回归运算时,使用交并比

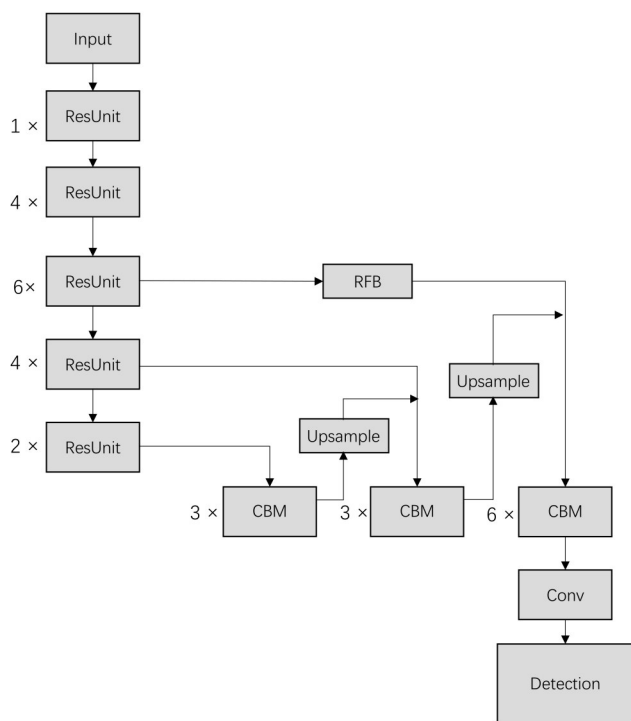


图6 改进的YOLO V3网络结构
Fig.6 Improved YOLO V3 network structure

(Intersection over Union, IOU)判别预测框与标注框的位置差距,计算式如式(2)所示, B 和 B^{gt} 分别代表预测框和标注框。常用IOU损失函数对于边框回归损失值进行计算,但分析式(2)中的IOU损失函数可以发现,面对预测框与标注框无重叠或者重叠面积大小一致等情况时,IOU不能给出优化方向,存在局限性。

$$IOU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}, \text{LOSS}_{IOU} = 1 - IOU \quad (2)$$

本研究使用CIOU损失函数计算预测框边框回归损失值,CIOU损失函数计算式如式(3)所示:

$$\text{Loss}_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (3)$$

式中, b 和 b^{gt} 代表预测框和标注框的中心位置坐标, ρ 代表计算两框中心位置的欧几里得距离, c 代表计算包含预测框和标注框的最小矩形的对角线长度, αv 是考虑预测框和标注框宽高情况的惩罚项。其中:

$$\alpha = \frac{v}{1 - IOU + v}, v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (4)$$

式中, α 为权重参数, v 代表考虑预测框和标注框宽高情况的计算项。

使用CIOU损失函数后,进行边框回归时面对预测框和标注框两框重叠面积相等、不相交及相包含等情况时,仍可以给出优化方向,使得预测框不断向标注框中心点靠近,加快网络训练的收敛速度。

3 实验

3.1 数据预处理

本研究使用的图像数据来自LIDC-IDRI数据集^[19],数据集共收录1 018个研究实例,每个研究实例由胸部医学图像文件和4位医师标注的肺结节病变信息组成。在读取图像文件头信息进行重新排序和重采样等预处理操作后,从中筛选出有至少3位医师标注且标注结节信息一致的CT图像共6 712张,组成实验使用的图像数据集。

肺结节所占CT图像比例较小,存在样本不平衡问题。为减少背景区域的干扰以提升检测性能,对得到的图像数据集进行肺实质分割处理。本研究采用阈值分割和形态学处理相结合的方法对图像数据集进行肺实质分割:首先选取合理阈值对于CT图像进行二值化处理,形态学处理消除边缘噪声,然后对比和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,最后通过得到的掩膜割取肺实质图像,组成训练数据集。过程可视化如图7所示。

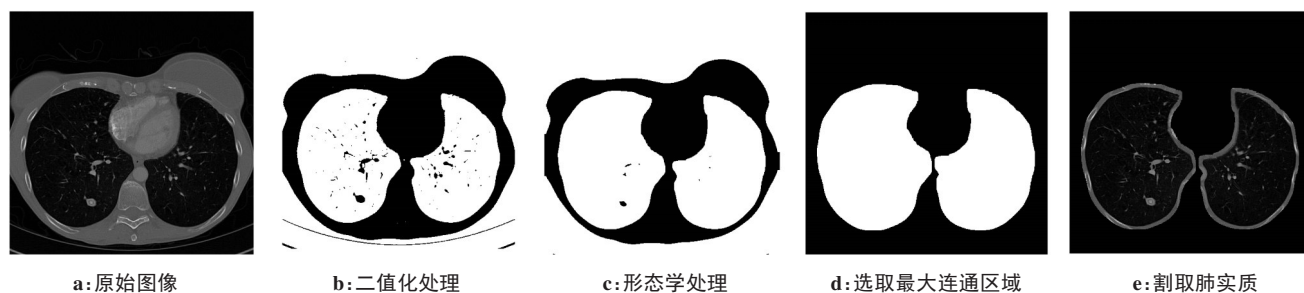


图7 肺实质分割可视化

Fig.7 Visualization of lung parenchyma segmentation

对于处理后的训练数据集进行角度翻转和添加高斯噪声等数据增强操作,将数据集扩充6倍。再将

数据增强后的训练数据集分成10份,1份作为测试集,其余图像数据送入网络进行训练。用于训练的

数据集再分为10份,1份作为验证集,其余9份作为训练集。

3.2 实验环境配置和实验评价标准

网络模型基于后端为TensorFlow的Keras神经网络库实现,使用的Python版本为3.6,TensorFlow-GPU版本为1.13.1,Keras版本为2.2.4。网络模型训练使用Adam优化算法,批次大小为32,初始学习率为0.001,在NVIDIA Quadro P6000显卡上完成训练。

在网络模型训练完成后,使用测试集验证网络模型的肺结节检测性能。根据测试集结果,计算目标检测常用评价指标准确率(Precision)和召回率(Recall),其中IOU阈值设定为0.5。由于通过准确率和召回率并不能完整评估模型性能,本研究进一步通过召回率-准确率曲线,计算得到精度检测指标AP(Average Precision)。为了更全面评价检测效果,本研究选择计算敏感度(Sensitivity)数据以评价肺结节的检出情况。以计算得到的AP值和敏感度数据两项指标综合评价实验方法的肺结节检测性能。

3.3 实验结果

为了充分验证改进方法的肺结节检测性能,进行两组对比实验:第1组实验对于不同组合的改进方法进行对比实验,验证本研究所提出方法的可行性,并直观展示不同组合的改进方法是如何提升网络模型的检测性能;第2组实验将本研究所提出的方法与其他深度学习检测方法进行对比,验证本研究方法的有效性。

3.3.1 使用不同改进方法的对比实验结果 在改进的YOLO V3基础网络结构上,搭建不同组合的改进方法的网络模型,测试这些改进方法对肺结节检测性能的提升效果。实验结果如表1所示。

在相同的实验环境下,从前4个实验的结果可以看出,激活函数的替换、加入RFB层和修改损失函数

表1 不同改进方法的实验结果对比
Tab.1 Experimental results comparison among different modified methods

检测方法	肺结节检测	
	AP/%	敏感度/%
改进基础结构的YOLO V3	77.81	87.56
改进的基础结构+Mish函数	79.16	88.77
改进的基础结构+RFB层	80.97	90.23
改进的基础结构+CIOU损失函数	82.81	91.17
改进的基础结构+Mish函数+RFB层	83.83	90.64
改进的基础结构+Mish函数+CIOU损失函数	84.42	91.45
改进的基础结构+RFB层+CIOU损失函数	86.59	93.66
本研究方法	88.89	94.73

这3种改进方法的实验结果均在两项检测指标上取得了一定提升,表明针对YOLO V3的改进思路是有效的,所提出的改进方法能够提升肺结节的检测效果。从后4个实验的结果可以看出,改进的损失函数对于肺结节检测性能的提升起到了较为重要的作用,在不同组合的改进方法中明显提升了肺结节检测的AP值和敏感度,明使用CIOU损失函数能够更好地定位肺结节。同时,应用RFB层加CIOU损失函数的改进方法相比单独应用CIOU损失函数的改进方法,提升了3.78%的AP值和2.49%的敏感度,与其他不同组合的改进方法相比表现更好,表明能够更好地提取肺结节特征信息。本研究所提出的改进方法结合了以上方法的优点,能够显著提升模型的检测性能,对于肺结节检测具有可行性。

3.3.2 其他深度学习方法的对比实验结果 本实验引入Faster RCNN和SSD深度学习方法进行对比实验,在进行网络训练时设置相同的随机数种子,保证图像训练数据的一致性和实验的可重复性。实验结果如表2所示。

表2 本研究方法与其他检测方法的实验结果比较
Tab.2 Experimental results comparison of the proposed method and other detection methods

检测方法	肺结节检测	
	AP/%	敏感度/%
SSD	74.56	82.53
Faster RCNN	82.78	90.46
本研究方法	88.89	94.73

从实验结果可知,在相同的实验环境下,本研究方法在测试集的AP值和敏感度两项评价标准上均取得更好的结果,优于Faster RCNN和SSD,能够更准确地定位结节。同时相对于Faster RCNN耗时较长的训练过程,本研究方法的处理过程简单直接,更适应于大量图像数据的检测需求。由此表明,本研究所提出的方法与其他深度学习检测方法相比,在肺结节检测中表现出更好效果。

4 讨论

本研究针对使用深度学习方法在CT图像中进行肺结节检测时遇到的问题,基于改进的YOLO V3提出一种肺结节检测方法:首先进行肺实质分割等预处理;然后修改基础网络结构,替换激活函数,引入含有空洞卷积的RFB层,提高肺结节特征提取能力;最后修改损失函数,提高对于肺结节的定位精度。该检测方法经过与其他深度学习检测方法的对

比实验,得到88.89%的准确率和94.73%的高敏感度,能够有效提升肺结节的检测效果,这表明所提出的方法在肺结节检测上具有可行性和有效性。但距离实际应用仍有一定差距,下一步将扩大研究范围,贴合临床工作实际要求去改进检测方法,进一步提升对肺结节的检测性能。

【参考文献】

- [1] 郑荣寿, 孙可欣, 张思维, 等. 2015年中国恶性肿瘤流行情况分析[J]. 中华肿瘤杂志, 2019, 41(1): 19-28.
ZHENG R S, SUN K X, ZHANG S W, et al. Report of cancer epidemiology in China, 2015[J]. Chinese Journal of Oncology, 2019, 41(1): 19-28.
- [2] REN S, HE K, 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] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2961-2969.
- [4] 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. Las Vegas: IEEE, 2016: 779-788.
- [5] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer vision and Pattern Recognition. Honolulu: IEEE, 2017: 7263-7271.
- [6] REDMON J, FARHADI A. Yolov3: an incremental improvement[EB/OL]. arXiv preprint, arXiv: 1804.02767.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision. Amsterdam: Springer, 2016: 21-37.
- [8] LU X, LI Q, LI B, et al. MimicDet: bridging the gap between one-stage and two-stage object detection [EB/OL]. arXiv preprint, arXiv: 2009.11528.
- [9] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[EB/OL]. arXiv preprint, arXiv: 1502.03167.
- [10] 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. Las Vegas: IEEE, 2016: 770-778.
- [11] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2117-2125.
- [12] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression [C]//18th International Conference on Pattern Recognition (ICPR'06). HongKong: IEEE, 2006: 850-855.
- [13] MISRA D. Mish: a self regularized non-monotonic neural activation function[EB/OL]. arXiv preprint, arXiv: 1908.08681.
- [14] ZHENG Z, WANG P, LIU W, et al. Distance-IoU Loss: faster and better learning for bounding box regression [C]//Thirty-Fourth AAAI Conference on Artificial Intelligence. New York: Elsevier, 2020: 12993-13000.
- [15] LIU S, HUANG D. Receptive field block net for accurate and fast object detection [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 385-400.
- [16] 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, 2014, 37(9): 1904-1916.
- [17] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [EB/OL]. arXiv preprint, arXiv: 1602.07261.
- [18] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[EB/OL]. arXiv preprint, arXiv: 1511.07122.
- [19] ARMATO III S G, MCLENNAN G, BIDAUT L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans[J]. Med Phys, 2011, 38(2): 915-931.

(编辑:黄开颜)