
Detection Method of Chest X-ray Lung Nodules Based on Improved YOLOV3 Model

Yan Sitao

School of Software Engineering, Beijing University of Posts and Telecommunications, Beijing, China

Email address:

826719572@qq.com

To cite this article:

Yan Sitao. Detection Method of Chest X-ray Lung Nodules Based on Improved YOLOV3 Model. *Science Discovery*.

Vol. 8, No. 2, 2020, pp. 18-23. doi: 10.11648/j.sd.20200801.15

Received: April 11, 2020; **Accepted:** May 12, 2020; **Published:** May 19, 2020

Abstract: Lung cancer is a kind of cancer with high lethality in the world, and the detection of lung nodules becomes particularly important in the early clinical manifestations of lung cancer. However, due to the characteristics of the small size of the lung nodules on the front of the chest and the obstruction of the ribs, it is more difficult to manually detect the lung nodules. At the same time, because of the explosive growth of lung X-ray images and diagnostic reports, the application of deep learning technology to the identification of lung nodules has become an inevitable. In order to be able to detect lung nodules in real time, this paper is based on the YOLOV3 algorithm. In view of the characteristics of lung nodules imaging, such as small size and rib occlusion, a method that can be used Lung nodule detection algorithm (nodule-YOLOV3), the pre-processed lung nodule X-ray image is input to the nodule-YOLOV3 network to obtain the prediction results of lung nodules. The experimental results show that: nodule -YOLOV3 detection accuracy is 61%, compared with YOLOV3 target detection algorithm, the accuracy of nodule-YOLOV3 algorithm is improved by 3%.

Keywords: Lung Nodule, X-ray Images, Deep Learning, YOLOV3

基于改进的YOLOV3模型的胸部X线肺结节检测方法

闫思桃

北京邮电大学软件学院, 北京, 中国

邮箱

826719572@qq.com

摘要: 肺癌是全球致死较高的一种癌症, 肺癌早期的临床表现中肺结节的检测变得尤为重要。但由于胸部正面X线的肺结节存在着尺寸较小、肋骨遮挡等特点, 导致人工对肺结节的检测难度较大, 同时也因为肺部X线影像和诊断报告呈爆炸式增长, 将深度学习技术应用到肺结节的识别成为一种必然, 为能够对肺结节进行实时检测, 本文基于YOLOV3算法, 针对肺结节成像存在尺寸小、肋骨遮挡等特点, 提出一种能够用于肺结节检测的算法 (nodule-YOLOV3), 将经过预处理的肺结节X线影像输入到nodule-YOLOV3的网络中得到肺结节的预测结果, 实验结果表明: nodule -YOLOV3检测准确率为61%, 相比于YOLOV3目标检测算法, nodule-YOLOV3算法准确率提高了3%。

关键词: 肺结节, X线影像, 深度学习, YOLOV3

1. 引言

原发性肺癌是癌症发病率和死亡率最高的一种疾病, 2018年《CA》杂志发表的全球癌症报告显示[1], 全球发病率和致死率第一的癌症依旧是肺癌。肺癌早期的临床表现为肺结节, 在肺部的表现为直径不超过3cm的圆形或者类圆形的结节影, 且难以与肺部的其他组织区分开来。目前临床对肺结节的筛查主要通过CT和X线成像, 随着医学放射技术的不断发展, 螺旋CT技术的出现使得肺结节的筛查检出率得到明显提高。但是CT成本较高, 放射剂量也较大, 而且在一些欠发达的医院没有相关的设备, X线成像在肺结节的筛查中仍占有很大的比重。医生通过人工阅片对肺结节进行诊断, 但是由于肺结节存在着尺寸小、肋骨遮挡等特点, 纵使阅片丰富的熟练医生也很难保证完全准确的判断。因此, 将计算机视觉技术应用到肺结节的检测中对早期的肺癌确诊具有重要的意义。

在过去几年中, 肺结节相关研究者已经提出了一些传统的肺结节检测算法, 大多数算法是将图像处理算法和机器学习相结合, 主要分为三个步骤[2, 3]: 提取肺部可视区域; 产生肺结节的候选结果; 减少结果中的假阳性率。但使用这种流程的模型进行肺结节的检测, 每一个过程的性能都受到上一个过程结果的影响, 而且将候选区域的提取与分类进行分离, 会丢失一部分上下文信息, 同时候选区域的选择策略需要大量的计算, 无法满足实时检测的需求。

自卷积神经网络在ImageNet挑战中被引入, 研究人员对卷积神经网络(Convolutional neural network, CNN)在目标检测方面的探索得到了快速的发展。与传统的肺结节检测算法相比, 使用卷积神经网络对肺结节影像进行学习, 能够自动的提取肺结节的特征, 学习能力强而且识别速度快。目前基于深度学习的目标检测算法可以分为两类, 一类是先产生候选区域再进行卷积分类, 一类是对输入的图片通过模型直接得到对应的类别和相应的定位。其中先产生候选区域再进行目标检测包括Fast R-CNN[4]、Faster R-CNN[5]和FPN[6] (Feature Pyramid Networks)等算法。这些算法将产生候选区域和边界框与分类分为两步执行, 虽然提高了算法的精确度, 但是检测速度很慢, 无法达到实时检测的需求。单次检测包括YOLO[7] (You Only Look Once)、SSD[8](Single Shot MultiBox Detector)等算法, 这些算法通过卷积神经网络直接输出目标物体的边界框和分类结果, 而且能够满足实时检测的需求。上述这些深度学习算法在目标检测数据集PASCAL VOC 2007[9]上都取得了较好的成绩, 但在胸部X线中肺结节只占据很小的像素, 深度学习算法对肺结节的检测效果并不是很好。

为了能够对胸部X线中肺结节进行实时检测, 本文以YOLOV3模型为基础, 将胸部X线影像中的肺结节作为检测目标, 针对肺结节尺寸小、肋骨遮挡等特点的(删除)成像特点, 对YOLOV3模型进行改进, 提高了肺结节的检测效果。

2. 相关工作

近年来, 深度学习算法在图像处理领域取得了巨大的进步, 本节中将简要介绍现有的基于深度学习的目标检测算法和基于深度学习的肺结节检测算法。

目前基于深度学习的目标检测算法中, 特征提取阶段的方法分为基于区域和回归的目标检测算法, R-CNN[10]、Fast R-CNN、Faster R-CNN等目标检测算法主要是通过区域建议产生大量的边界框, 再用分类器对边界框中的对象进行判断。YOLO、SSD和RetinaNet[11]等模型将目标检测作为一个回归问题, 通过使用一个卷积神经网络对一张图像预测边界框的坐标。基于回归的目标检测算法很大程度上减少了参数的数量, 提高了检测速度, 并具有良好的检测性能。

上述算法在肺结节的检测中已被广泛使用。H Xie等提出了一种基于二维卷积神经网络的新型肺结节自动检测模型[12], 通过使用两个区域提议网络和一个反卷积层来优化Faster R-CNN的结构, 从而提高了对肺结节检测的敏感性。但是Faster R-CNN先产生候选区域再进行卷积分类不能对肺结节进行实时检测。Weikang Fan等采用了一种改进的快速的R-CNN模型[13]用来提高肺结节检测的性能。首先R-CNN模型先将图像通过反卷积输入到VGG-16[14]中提取共享的卷积特征。然后, 将得到的共享卷积特征输入到区域提议网络, 输出候选肺结节区域。之后将候选肺结节区域和先前共享的卷积特征同时输入到池化层中提取相应候选区域的特征。最后通过连接层, 多任务分类器用于定位候选区域的回归。然而整个过程将产生候选区域与分类进行分离, 会丢失一部分的上下文消息, 由于算法具有较高的复杂度, 不能够实时对肺结节进行检测。Sindhu Ramachandran S等使用基于YOLO模型的DetectNet架构[15]来检测肺部CT扫描的结节。使用YOLO模型训练和测试过程中可以看到整个图像, 能够隐式地学习有关类及其在图像中的上下文知识。这使得YOLO比基于滑动窗口和区域建议的检测方法更具优势。由于网络可以在更大的范围内看到对象, 因此与其他方法(如Fast R-CNN)相比, 将图像中的背景色块误认为是实际对象而引起的错误非常少。通过在卷积中纳入转移学习并在训练过程中保留邻域图像信息, 所提出的检测系统具有较高的准确性和灵敏度。

3. YOLOV3算法原理及优化

3.1. YOLOV3模型分析

YOLOV3整合了YOLOV2[16]、Darknet-19和FPN模型。YOLOV3网络由53个卷积层组成, 称为Darknet-53。

Darknet-53在Darknet-19中添加残差网络的混合方式, 使用连续的1×1卷积层和3×3的卷积层进行网络的搭建。其中, 1×1卷积层用于跨通道的信息整合。为了提高模型的收敛速度, 同时达到正则化的效果, YOLOV3继续使用YOLOV2中在卷积层的后面添加的Batch Normalization, 用于对每一层的输入进行归一化。由于不同的物体形状大小有所差异, 在训练过程中模型需要学习适应不同物体的

形状, YOLOV3借鉴了Faster R-CNN中用于产生候选框的策略, 使用卷积和先验框(anchor box)预测边界框。为了能够找到合适的先验框, YOLOV3采用k-means聚类方法对数据集中的边界框做聚类分析, 得到9个先验框。YOLOV3采用了用于多尺度检测的FPN算法, 使用了三个尺度的特征图, 并对三个尺度的特征图进行融合, 在多个尺度的融合特征图上分别做独立检测。从而提高对小目标检测的效果。

YOLO在进行目标检测中将输入的肺部X线影像分为 $S \times S$ 个网格, 如果肺部X线影像中一个物体的中心点落在了该网格中, 该网格将会负责这一物体的预测。Darknet-53网络结构如图1所示。YOLOV3在三个不同的尺度上进行预测, 网络首先将经过预处理的肺结节X片影像缩放到

416×416 , YOLOV3网络分别在尺度 13×13 、尺度 26×26 、尺度 52×52 上进行检测。在尺度为 13×13 的特征图上执行8倍降采样后进行检测, 在尺度为 26×26 的特征图上执行8倍降采样后进行2倍的上采样, 之后对输出的特征图进行检测, 同时与上一层的特征图进行拼接。在尺度为 52×52 的特征图上执行8倍降采样后进行2倍的上采样, 之后对得到的特征图进行检测, 并与上一层的特征图进行拼接。每个特征图使用三个锚点预测边界框。

在PASCAL VOC 2007和2012的数据集中, 人, 车等目标会占据一张图片的大部分空间, 因此, YOLOV3在PASCAL VOC 2007和2012上会取得较好的结果, 但是对于肺结节尺寸小、肋骨遮挡等特点, YOLOV3的网络结构并不能取得在PASCAL VOC 2007和2012上的效果。

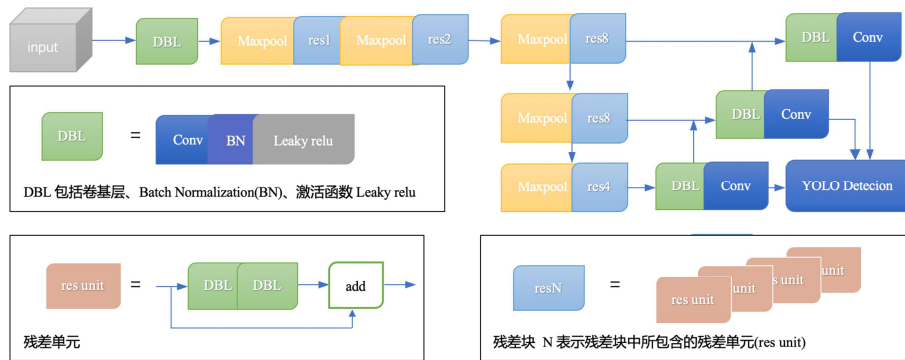


图1 Darknet-53网络结构, 在DBL块中, 所使用的卷积大小为 3×3 。在池化层(Maxpool)中, 使用最大池化(Max pooling), 过滤器大小为 3×3 , 步长为2。在残差单元(res unit)中, DBL块中卷积的大小分别是 3×3 、 1×1 。残差块(resN)中, N表示残差块中残差单元的个数。

3.2. YOLOV3模型的改进

本文基于YOLOV3搭建了改进的YOLOV3网络结构 (nodule-YOLOV3), nodule-YOLOV3的网络结构如图2所示, nodule-YOLOV3包括50个卷积层, 5个最大池化层。

在YOLOV3网络中, 对于要检测的图像都会缩小到 416×416 进行检测, 但将图像缩小会直接影响目标检测的识别效果。本文采用经过预处理的 512×512 的肺结节X线影像数据集, 如果将肺结节X线影像缩放到 416×416 进行检测, 原图像中肺结节的分辨率会变低, 从而影响网络对肺结节的检测。因此, nodule-YOLOV3模型将图像的检测大小改为 $512 \text{ pixel} \times 512 \text{ pixel}$, 从而提高对目标检测的识别效果。

在YOLOV3网络中, 第二个残差块通过两个残差单元对小目标进行检测, 不能获得更多的小目标位置信息。为了能够获取更多的肺结节的目标位置信息, nodule-YOLOV3基于YOLOV3网络在第二个残差块中添加了两个残差单元, 通过4倍的降采样学习肺结节的特征。如图2所示, nodule-YOLOV3利用YOLOV3原有网络中在尺度 16×16 、尺度 32×32 、尺度 64×64 上进行降采样和上采样, 同时为了能够获取更多低层的小目标的位置信息, 在第二层添加了2个残差块。对第二层执行4倍的降采样后进行检测, 在尺度为 16×16 、 32×32 、 64×64 的特征图上执行8倍降采样后进行2倍的上采样, 将得到的特征图进行检测, 并与上一层的特征图进行拼接。

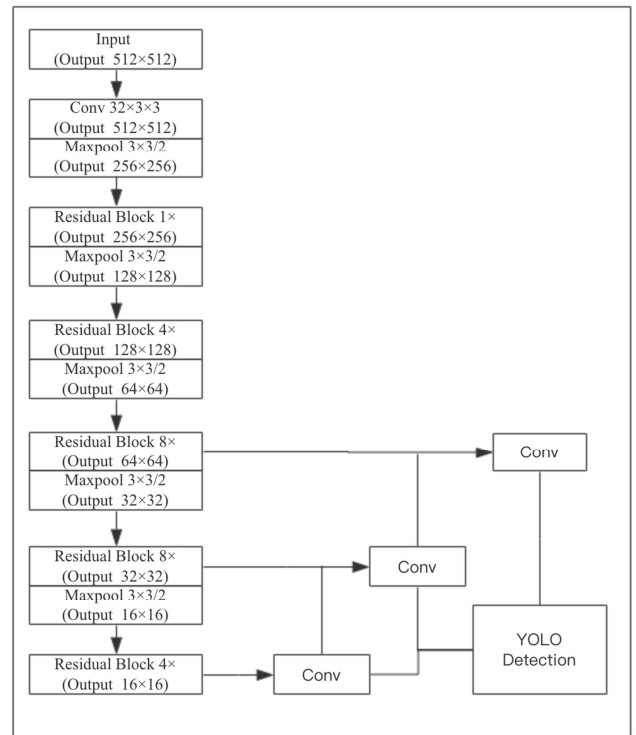


图2 nodule-YOLOV3网络结构, 图中Residual Block为残差块, N表示残差块中残差单元的个数, 图中Maxpool为池化层, 使用最大池化(Max pooling), 过滤器大小为 3×3 , 步长为2。

3.3. 残差网络模块

在卷积神经网络中,为了提高网络的性能可以通过增加网络的宽度和深度对网络进行优化。层数多的网络要比层数少的网络效果好,但随着层数的不断增多,卷积神经网络容易出现梯度爆炸等问题,从而阻碍了网络参数的收敛。随着反向传播梯度的消失整个网络的参数都无法再更新和优化,使得网络的性能不再发生变化,这是由于网络的层数增加引起的。

为了解决深层网络梯度下降和梯度爆炸等问题,Kaiming He等人所提出了残差网络[17] (ResNet)。在残差网络中通过短路机制添加了残差单元用于训练更深的神经网络,残差单元如图3所示。一个残差块可以由公式(1)表示。

$$x_{l+1} = x_l + F(x_l, W_l) \quad (1)$$

由公式可以看出残差单元由两部分组成, x_l 为直接映射部分,如图3右边部分, $F(x_l, W_l)$ 为残差部分,如图3左边部分,由卷积层组成。

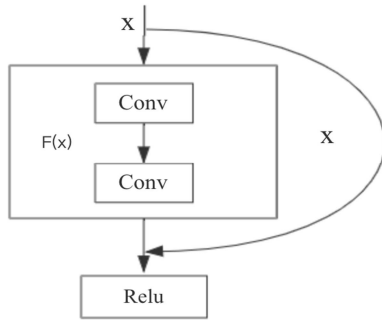


图3 残差单元。

在本文中,nodule-YOLOV3采用了基于二层的残差单元的网络结构,通过添加残差单元加深网络不会引起梯度消失和梯度爆炸的问题。残差网络能够训练更深层次的网络并因为网络深度的增加取得更高的精确度。

4. 数据集

4.1. 数据集介绍

本文采用了公开数据集ChestX-Ray8数据集[18],是迄今为止最大的公开可用的胸部X射线数据集。该数据集包含了从1992-2015年的32717幅胸片中的108948幅正面胸片。胸片大小为1024×1024,包含十四种常见的胸部疾病包括肺不张,巩固,浸润,气胸,水肿,肺气肿,纤维化,积液,肺炎,胸膜增厚,心脏扩大,结节,肿块和疝。本文引用了Kaggle中由zhangjin重新组织后的ChestX-Ray8的数据集进行肺结节的检测。该数据集仅对肺结节进行了标注,可以用于肺结节的检测。数据集一共有2016张肺部X线影像,将肺部X线影像分为训练集和验证集。其中训练集包括1500张,验证集包括516张。训练集和验证集都包括肺部X线影像和对应的标注。肺部X线影像如图4所示。

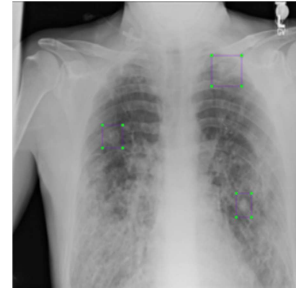


图4 肺部X线影像样本图像,图中紫色框标注的为肺结节。

4.2. 数据集预处理

X线图像是标准X线束穿过人体不同密度、厚度的组织结构的投影总和,将三维立体变为二维图像,因而X线图像与人体组织结构相比,产生形态失真、放大及相互重叠后的复合影像。根据X线影像这一成像特点,通过调整对比度对数据集进行数据为了提高肺结节的准确率,通过图像融合的方法进行对比度的调整,新建一张全为0的数组图片作为图2,将图1与X线影像进行融合,所用到的图像线性混合的数学原理公式如公式(2), $f_0(x)$ 和 $f_1(x)$ 表示参与混合的两张图像, $g(x)$ 表示输出图像。通过将两张图片的像素值进行先行加权计算得到最终的图像, α 表示图像1在叠加过程中所占的比例。在进行叠加需要保证两张图片的大小和类型必须完全一致,并具有相同的维度。

$$g(x) = (1 - \alpha)f_0(x) + \alpha f_1(x) \quad (2)$$

将调整对比度后的肺结节X线影像的大小缩小到512×512,用于适应nodule-YOLOV3搭建的网络。实验对比图如图5所示,可以看到,通过对比度的修改,肺部的影像效果更加的明显,有利于神经网络学习到肺结节的特征。

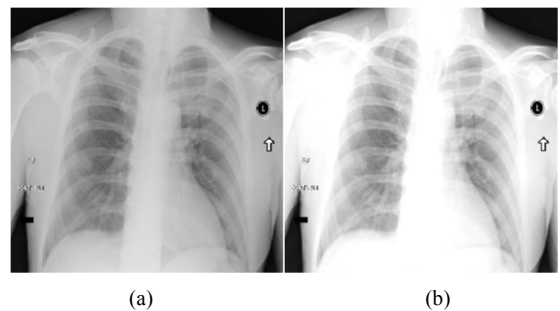


图5 图5(a)表示未调整对比度的肺部X线影像,图5(b)表示调整对比度后的肺部X线影像,相较于图5(a),图5(b)中前景和背景区分更加明显。

5. 实验及结果

5.1. 评估标准

实验使用平均检测准确率(mean aVerage precision, mAP)作为肺结节检测结果的评价标准。mAP值的计算由召回率和准确率计算得到,召回率的计算公式如公式(3),准确率的计算公式如公式(4)。其中 X_{tp} 表示预测与真实值

保持一致的目标个数， X_{fp} 表示预测与真实值不符合的目标个数， X_{fn} 表示没有被检测到的真实框的值。

$$recall = \frac{X_{tp}}{X_{tp}+X_{fn}} \tag{3}$$

$$precision = \frac{X_{tp}}{X_{tp}+X_{fp}} \tag{4}$$

5.2. nodule-YOLOV3网络训练及实验环境

本文方法基于胸部X线影像数据集进行训练，类别为1，训练过程中将学习率设置为0.001，权重衰减正则项设置为为0.005。YOLOV3中引入了Faster R-CNN中anchor boxes的思想，anchor boxes是一组高宽固定的候选框，针对YOLOV3中三个不同尺度的预测，通过k-means算法得到候选框的高宽，尺度1对应的高宽(36, 37) (26, 23) (24, 23)，尺度2对应的高宽(24, 20) (22, 27) (21, 23)，尺度3对应的高宽(21, 25) (17, 14) (16, 16)。

本文通过对比实验来验证增强后的数据集以及改进后的模型。实验的测试环境为Linux，GPU为Tesla P100，将数据集集中的516张肺部X线影像数据输入到训练好的模型中进行训练可得到检测图像及对应的置信度。

5.3. 实验结果与分析

通过对训练集进行训练得到权重，将测试集中肺部X线影像输入至训练得到的两个网络中进行测试，选择其中一例的样本，标注如图6所示。

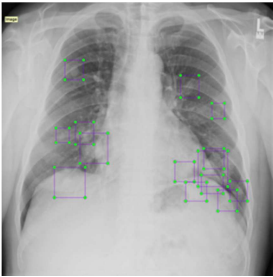


图6 训练样本标注,图中紫色框标注的为肺结节。

图7为两个模型在同一实验环境下的检测结果对比。由图中可以看出，使用nodule-YOLOV3模型能够检测到更多被肋骨遮挡部分肺结节，降低了漏诊率。

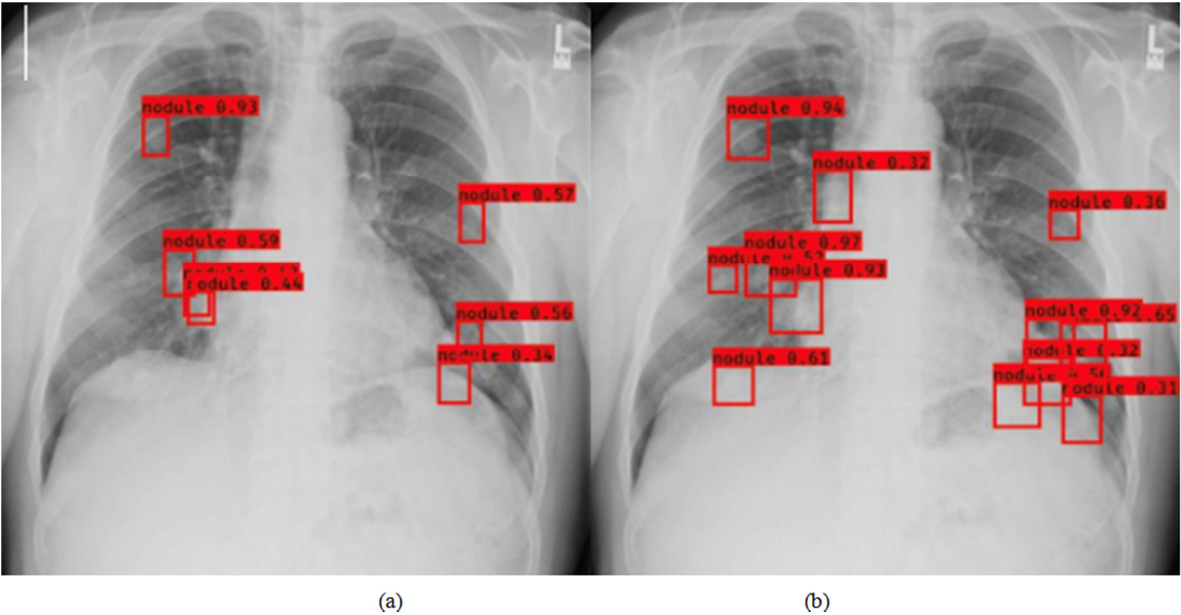


图7 检测结果对比。

图7(a)表示通过YOLOV3模型进行检测得到的实验结果，

图7(b)表示通过nodule-YOLOV3模型进行检测得到的实验结果，相较于图7(a)，图7(b)更接近于样本(如图6)中肺结节的标注。

通过计算目标检测中衡量识别精度的mAP值验证修改后的模型在精确度上得到提高。实验结果如表1，与YOLO V3 相比，nodule-YOLOV3算法对小目标的检测准确率由65%提高到70%，召回率由60%提高到65%。结果表明nodule-YOLOV3得到的mAP值为61%。与 YOLOV3 算法相比，其mAP值提高了 3%。

表1 实验结果。

Detection algorithm	recall	precision	mAP
YOLOV3	60%	65%	58%
nodule-YOLOV3	65%	70%	61%

6. 结束语

本文针对肺结节的成像特点提出了一种基于改进的YOLOV3的检测模型。所提出的nodule-YOLOV3目标检测算法取得了较好的效果。首先通过修改对比度进行数据增强处理，其次为了对小目标的肺结节进行更好的检测，通过在YOLOV3的模型中增加残差块的方式进行改进。实验

结果表明,改进后的YOLOV3算法对肺结节的准确率和召回率均有提高。将改进后的模型应用到肺结节的检测能够更加有效的帮助医生对肺结节进行诊断。为了进一步判断肺结节是否是引起肺癌的临床表现,在使用nodule-YOLOV3对肺结节进行检测后,对肺结节的良恶性进行检测将是本文下一阶段的研究目标。

致谢

本文为国家重点研发计划项目(2017YFC1307705)的阶段成果之一。

参考文献

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA: a cancer journal for clinicians, 2018, 68 (6): 394-424.
- [2] Wang C, Elazab A, Wu J, et al. Lung nodule classification using deep feature fusion in chest radiography [J]. Computerized Medical Imaging & Graphics, 2017, 57 (4): 10-18.
- [3] Li X, Shen L, Luo S. A solitary feature-based lung nodule detection approach for chest X-ray radiographs [J]. IEEE J Biomed Health Inform, 2017, 22 (2): 516-524.
- [4] GIRSHICK R. Fast r-cnn [C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.
- [5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C]//Advances in Neural Information Processing Systems. 2015: 91-99.
- [6] LIN T Y, DOLL R P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//CVPR. 2017: 4.
- [7] 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. 2016: 779-788.
- [8] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, et al. SSD: Single Shot MultiBox Detector. ECCV 1 (2016), 21-37 (2016).
- [9] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge [J]. International journal of computer vision, 2010, 88 (2): 303-338.
- [10] 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.
- [11] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.
- [12] Xie H, Yang D, Sun N, et al. Automated pulmonary nodule detection in CT images using deep convolutional neural networks [J]. Pattern Recognition, 2019, 85: 109-119.
- [13] Fan W, Jiang H, Ma L, et al. A modified faster R-CNN method to improve the performance of the pulmonary nodule detection [C]//Tenth International Conference on Digital Image Processing (ICDIP 2018). International Society for Optics and Photonics, 2018, 10806: 108065A.
- [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv: 1409.1556, 2014.
- [15] George J, Skaria S, Varun V V. Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans [C]//Medical Imaging 2018: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2018, 10575: 105751I.
- [16] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
- [17] 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.
- [18] Wang X, Peng Y, Lu L, et al. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2097-2106.