

金字塔卷积融合YOLOv4的乳腺肿块检测算法

罗 菁, 凌现慧

天津工业大学, 控制科学与工程学院, 天津

收稿日期: 2021年9月27日; 录用日期: 2021年10月20日; 发布日期: 2021年10月27日

摘 要

针对乳腺肿块与腺体对比度低而导致检测精度低, 假阳性率较高的问题, 提出了一种基于改进的YOLOv4网络的乳腺肿块检测方法。首先引入金字塔卷积, 通过不同大小和深度的卷积核, 对输入的特征进行提取; 其次, 将原特征提取网络中的普通卷积替换为深度可分离卷积, 减少网络训练参数, 提升网络训练速度。实验结果表明, 改进的YOLOv4算法在测试集上的敏感性达到81.49%, 较原网络高了2.81%, AP值达到86.85%, 比原网络的AP值高了4.27%; 平均每幅图假阳性个数为0.418个, 相较原网络降低了0.028个, 该算法的检测性能较YOLOv4有明显提升。

关键词

乳腺x线图像, YOLOv4, 金字塔卷积, 深度学习

Breast Mass Detection Algorithm Based on Pyramid Convolution Fusion with YOLOv4

Jing Luo, Xianhui Ling

School of Control Science and Engineering, Tiangong University, Tianjin

Received: Sep. 27th, 2021; accepted: Oct. 20th, 2021; published: Oct. 27th, 2021

Abstract

Aiming at the problems of low detection accuracy and high false positive rate caused by low contrast between breast mass and gland, a breast mass detection method based on improved YOLOv4 is proposed. Firstly, Pyramid Convolution is introduced to extract the input features through convolution kernels of different sizes and depths; Secondly, the ordinary convolution in the original feature extraction network is replaced by Depthwise Separable Convolution to reduce the network training parameters and improve the network training speed. The experimental results show that the sensitivity of the improved YOLOv4 algorithm on the test set is 81.49%, which is

2.81% higher than that of the original network, and the AP value is 86.85%, which is 4.27% higher than that of the original network; The average number of false positives per image is 0.418, which is 0.028 lower than the original network. The detection performance of the algorithm is significantly improved compared with YOLOv4.

Keywords

Mammography, YOLOv4, Pyramid Convolution, Deep Learning

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

当前, 乳腺癌已经成为全球女性最常见的癌症, 患乳腺癌的女性占女性新诊断得癌症病例的 30% [1]。乳腺癌的早期发现, 早期治疗大幅度降低了乳腺癌的致死率。因此, 早期的乳腺癌筛查是乳腺癌治愈率提升的关键。乳腺 x 线图像是乳腺早期筛查最常见的无创检测方法。但由于乳腺肿块与乳腺腺体对比度较低, 乳腺肿块较难识别, 并且专业医生由于任务量繁重以及主观因素, 诊断中仍会出现漏检、错检等现象。

随着科技的发展, 计算机辅助诊断技术(Computer Aided Diagnosis, CAD)被广泛应用在医学图像的处理上, 降低了医生的工作量, 并且也大大减少了医生因大量阅片产生误诊的概率[2]。对于乳腺肿块的检测算法, 传统的图像处理方法是利用边缘检测, 区域生长, 阈值分割等方法, 在乳腺 x 线图像上提取感兴趣区域后对其进行医学分析。随着深度学习的兴起, 各个领域都得到了较好的发展, 基于深度学习的医学图像检测也表现出了优异的成绩。目前主流的乳腺肿块检测算法主要分为两大类, 第一类是双阶段检测算法, 该方法首先在图像上生成大量的候选框, 然后对这些候选框进行分类并回归目标候选框位置, 最终得到包含目标的检测框以及对应类别的置信度。此类算法代表有 R-CNN [3]、Fast-RCNN [4]、Faster-RCNN [5]等, 该算法在检测精度上有较好表现, 但其检测速度慢, 对易混淆的目标区分差。郑杰等[6]提出了一种基于 Faster RCNN 的乳腺肿块检测算法, 其敏感性取得了较好的结果, 但是单张假阳性个数较大。第二类为单阶段检测算法, 该算法将检测问题转化为回归问题, 是一种端到端的目标检测算法。此类算法代表有 SSD [7]、YOLOv1 [8]、YOLOv2 [9]、YOLOv3 [10]等, 该算法检测速度较快, 但其对小目标检测不敏感。蒋慧琴等[11]设计了一种基于 YOLOv3 的双视图信息融合的乳腺肿块检测算法, 在乳腺肿块的检测上取得较好的效果。

YOLOv4 [12]相对于之前的系列网络来说, 其结合了之前的研究成果, 在检测精度和检测速度上有较好的表现。但对于乳腺肿块图像来说, 肿块较难识别, 导致检测精度低和假阳性率高的问题, YOLOv4 仍需要进一步的改进从而满足乳腺肿块的检测。本文对 YOLOv4 算法进行改进, 引入金字塔卷积[13] (Pyramid Convolution, PyConv), 提高特征提取能力; 使用深度可分离卷积在不降低网络检测精度的前提下, 大幅度降低网络训练参数。实验结果表明, 改进后的 YOLOv4 网络的检测精度提升, 平均每幅图像假阳性个数下降。

2. YOLOv4 网络结构

YOLO 系列网络不同于两阶段网络, 直接采用回归的方法对坐标框进行检测和分类, 是一个端到端的检测网络。该系列网络将输入的图像分割成若干网格, 每个网格负责预测落在此网格上的物体, 若待

测目标落在某网格, 那么该网格就负责检测目标, 这种方法大幅度提升了检测速度。YOLOv1 借鉴 GoogLeNet 设计主干网络, 将整张图片作为输入, 直接在输出层对 BBox 的位置和类别进行回归, 虽然 YOLOv1 检测速度快, 但其对小目标检测效果不好。YOLOv2 引入了 anchor 机制, 引入了 Darknet-19 作为主干网络。YOLOv3 引入了 Darknet-53 作为主干网络, 并且引入了 FPN 进行多尺度预测, 对不同尺度的特征图进行融合预测, 提高了小目标的检测精度。

YOLOv4 网络相比 YOLOv3 来说, 将原来的 Darknet-53 网络与 CSPNet 网络融合, 提出了 CSPdarknet53。特征提取网络由原来的特征金字塔网络转换为含有空间金字塔池化层(Spatial Pyramid Pooling, SPP)的 PANet。除此之外, YOLOv4 添加了 Mosaic 数据增强, 使用 Mish 激活函数, 并且引入了余弦退火学习率, 改进了 SOTA 的方法。YOLOv4 的结构图如图 1 所示。

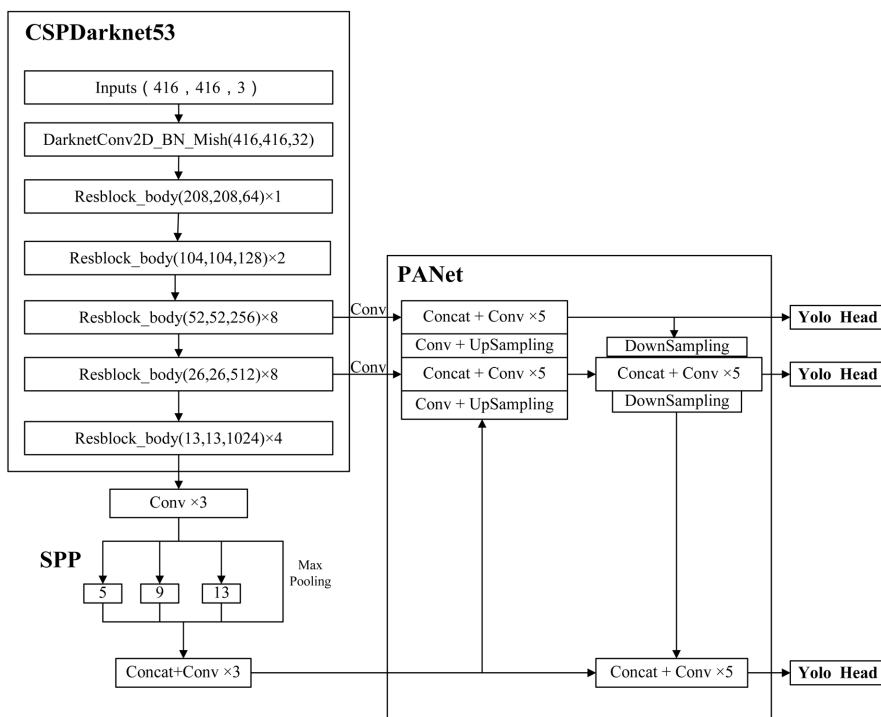


Figure 1. YOLOv4 network structure
图 1. YOLOv4 网络结构图

由图可知, YOLOv4 的主干网络主要由残差块堆叠而成, 这些残差块在 YOLOv3 中 Darknet53 的基础上进行了修改, 结合了 CSPNet [14]结构的思想。CSPNet 就是将原来的残差块进行拆分, 主干部分继续进行原来的残差块的堆叠, 另一部分则像一个残差边一样, 经过简单的卷积处理连接到最后, CPSdarknet53 结构如图 2 所示。

SPP 结构是对主干网络最后一个特征层经过三次池化后进行最大池化。池化中使用四种不同尺度的池化层, 池化层大小分别为 1×1 , 5×5 , 9×9 , 13×13 。SPP 结构可以显著地增加感受野, 分离出重要的上下文特征, 并且对检测速度几乎没影响。PANet 结构由卷积, 上采样, 特征层融合, 下采样操作构成的金字塔结构, 相较于 FPN 结构, 在此基础上又增加了一个底层到高层的通道, 将底层信息直接向上传递, 提高底层信息的利用率。经过 PANet 结构, 不同层之间的特征层充分融合, 进一步提高特征提取的能力。YOLOv4 Head 结构与 YOLOv3 的原理相同, 对输出的特征进行 3×3 和 1×1 两次卷积, 对每个特征层的三个先验框进行判别, 并判断是否含有目标及其种类。

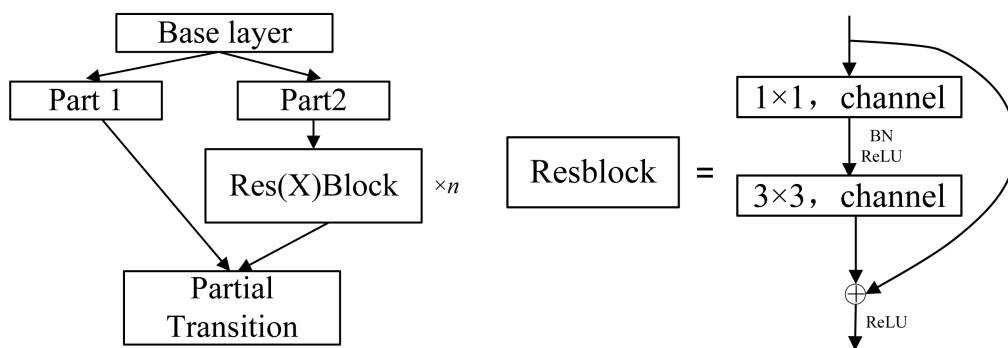


Figure 2. CSPDarknet53 structure
图 2. CSPDarknet53 结构图

3. YOLOv4 的改进

3.1. 融合 Pyconv 模块的 YOLOv4 主干网络

对于乳腺肿块图像来说, 乳腺肿块的尺度大小不一, 原 YOLOv4 主干网络采用的传统卷积的感受野是固定大小, 对于本文数据集的特征提取能力有限, 可能会导致有效信息丢失。针对该问题, 引入金字塔卷积 PyConv。PyConv 能够在多尺度的卷积核上处理输入的图像, 其中包含了不同大小和深度的卷积核, 能够捕捉不同的信息, 同时 PyConv 的效率较高, 它与传统卷积拥有相近的计算成本。因此本文引入金字塔卷积提高网络的特征提取能力, 使得网络可以提取到更加有效的信息。

对于传统卷积, 结构图如图 3 所示, 输入特征图维度为 FM_i , 经过 FM_o 个大小为 $K \times K$ 卷积核后输出维度为 FM_o 的特征图, 输出特征图大小为 $H \times W$ 。因此, 经过传统卷积的参数量和计算量公式如式(1)~(2)所示:

$$parameters_1 = K^2 \times FM_i \times FM_o \tag{1}$$

$$FLOPs_1 = parameter_1 \times (W \times H) \tag{2}$$

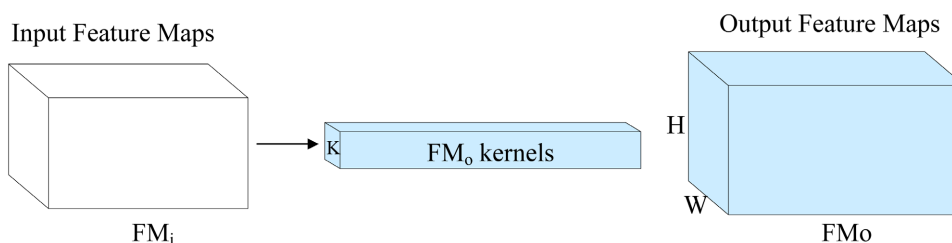


Figure 3. Standard convolution structure
图 3. 传统卷积结构图

对于金字塔卷积, 如图 4 所示, 包含 n 级不同类型卷积核的金字塔, 随着级别的加深, 卷积核大小不断减小, 每一层卷积核大小为 K_1, K_2, \dots, K_n , 其维度分别为 $FM_{o_1}, FM_{o_2}, \dots, FM_{o_n}$, 如公式(3)所示:

$$FM_{o_1} = FM_i, FM_{o_2} = \frac{FM_i}{\left(\frac{K_2^2}{K_1^2}\right)}, \dots, FM_{o_n} = \frac{FM_i}{\left(\frac{K_n^2}{K_1^2}\right)} \tag{3}$$

输出特征维度为 $FM_{o_1}, FM_{o_2}, \dots, FM_{o_n}$, 即输出特征如维度为 $H \times W \times FM_o$ 。因此, PyConv 的参数量和计算量公式如式(10)、(11)所示:

$$parameters_2 = K_n^2 \times \frac{FM_n}{\left(\frac{K_n^2}{K_1^2}\right)} \times FM_{O_n} + \dots + K_2^2 \times \frac{FM_2}{\left(\frac{K_2^2}{K_1^2}\right)} \times FM_{O_2} + K_1^2 \times \frac{FM_1}{\left(\frac{K_1^2}{K_1^2}\right)} \times FM_{O_1} \quad (4)$$

$$= \sum_{i=1}^n K_i^2 \times FM_{O_i} \times \frac{FM_i}{\left(\frac{K_i^2}{K_1^2}\right)}$$

$$FLOPs_2 = parameters_2 \times (W \times H) \quad (5)$$

其中, $FM_{O_n} + \dots + FM_{O_2} + FM_{O_1} = FM_{O_0}$ 。

通过上述公式不难看出, 当每一层级的输出通道数相近, 那么每层参数量及计算复杂度与传统卷积相近。同时在金字塔卷积的基础上使用分组卷积, 将输入特征划分为不同的组进行卷积计算, 每组特征图通道数不同, 通过分组卷积的方法改变卷积核深度。同时, 当分组数增加时, 减少卷积核深度, FLOPs 与参数量也相应地减少。

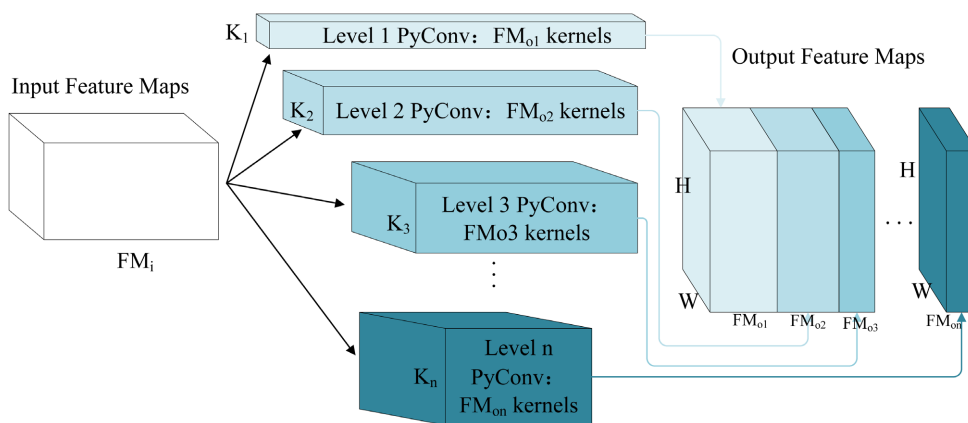
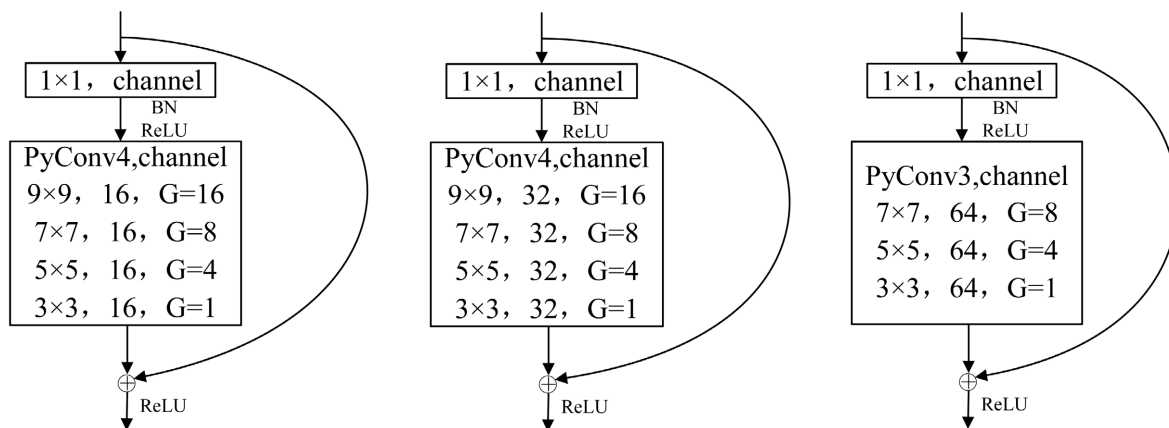


Figure 4. Pyramidal convolution structure
图 4. 金字塔卷积结构

本文将 PyConv 镶嵌到 CSPdarknet53 主干网络的残差结构中, 由于原主干网络是由 5 个不同的 CPS 残差模块构成, 因此将不同级别的特征金字塔引入到各个残差模块中, 本文主要将原来残差结构中 3×3 卷积替换成 5 种不同层级的金字塔卷积, 使得网络可以有效的提取完整的特征, 以及避免提取过多无效信息。改进后的残差部分结构如图 5 所示, 其中从左到右依次是 PyConv_Resblock1 至 PyConv_Resblock5。



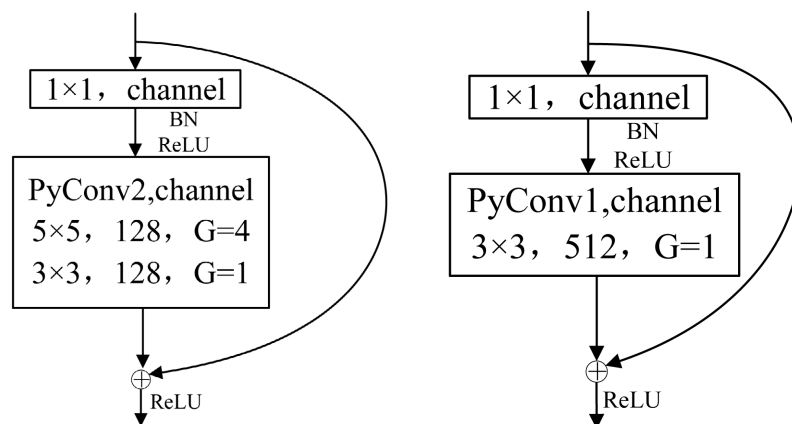


Figure 5. Improved Resblock structure
图 5. 改进后 Resblock 结构

3.2. 传统卷积的改进

为了减少参数量, 本文借鉴 Mobilenet [15]的轻量化思想, 用深度可分离卷积(Depthwise Separable Convolutions)替代 PANet 特征提取网络中所用到的普通卷积。深度可分离卷积由逐深度卷积(Depthwise Convolution)和逐点卷积(Pointwise Convolution)组成, 逐深度卷积对应输入每一个输入特征图通道; 逐点卷积与常规卷积相差无几, 其作用是负责将逐深度卷积提取的特征进行融合。通过逐深度卷积和逐点卷积两个步骤可以有效的降低计算复杂度和模型容量。

对于传统卷积来说, 结构图如图 6 所示, 假设输入特征图的大小为 $D_F \times D_F \times M$, 卷积核大小为 $D_K \times D_K \times M \times N$, 输出特征图的大小为 $D_F \times D_F \times N$, 则标准卷积对应的计算量如式(6):

$$D_K \times D_K \times M \times N \times D_F \times D_F \tag{6}$$

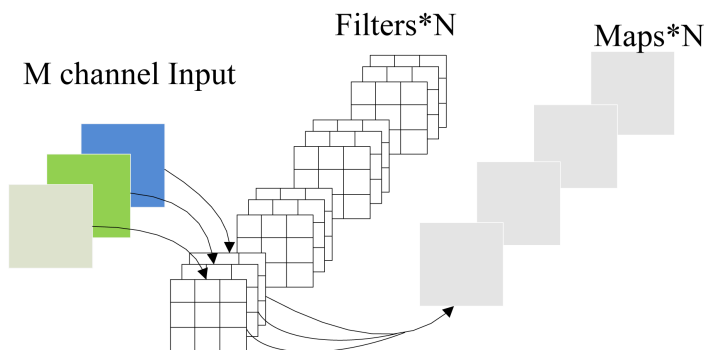


Figure 6. Traditional convolution structure
图 6. 传统卷积结构

对于深度可分离卷积来说, 结构图如图 7 所示, 首先经过逐深度卷积操作, 将单个滤波器应用到每个输入通道, 深度可分离卷积的第一步是深度卷积, 对于输入的每一个通道分别用 1 个 $D_K \times D_K \times 1$ 的卷积核进行卷积, 共计 M 个卷积核, 得到 M 个的特征图, 这 M 个特征图相互独立。第二步就是逐点卷积, 将第一步得到的 M 个特征图作为输入, 用 N 个 $1 \times 1 \times M$ 的卷积核进行标准卷积操作, 得到 $D_F \times D_F \times N$ 的输出特征图。则深度可分离卷积对应的计算量如式(7):

$$D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F \tag{7}$$

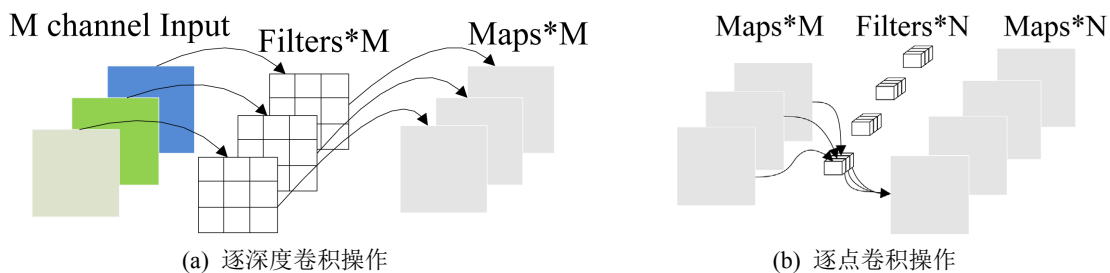


Figure 7. Depthwise separable convolution

图 7. 深度可分离卷积结构

由此, 深度可分离卷积与传统卷积相比计算量减少如式(8):

$$\frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (8)$$

4. 实验与分析

本实验在 Windows10 的环境下完成, 计算机内存为 16G, CPU 为 Inter Core i7 9700 3.0GHz, GPU 为 GTX2080, 显存为 8GB, 使用的开发框架为 Pytorch, 编程语言为 Python。同时安装了 CUDA10.2 核 cudnn7.6 以支持 GPU 使用。

4.1. 数据集

本实验的数据集采用的 CBIS-DDSM [16]乳腺数据集, 该数据集是南佛罗里达大学创建的公开乳腺数据集 DDSM [17] (Digital Database for Screening Mammography)更新的标准化版本, 实验将选取数据集中 1412 张含有肿块的图像进行处理, 通过反转, 旋转等操作扩充数据集, 最终得到含有肿块数据集 5648 张, 其中 4760 张作为训练集, 888 张作为测试集。本文图像中的肿块通过 labeling 软件进行标注, 数据标注图例如图 8 所示。

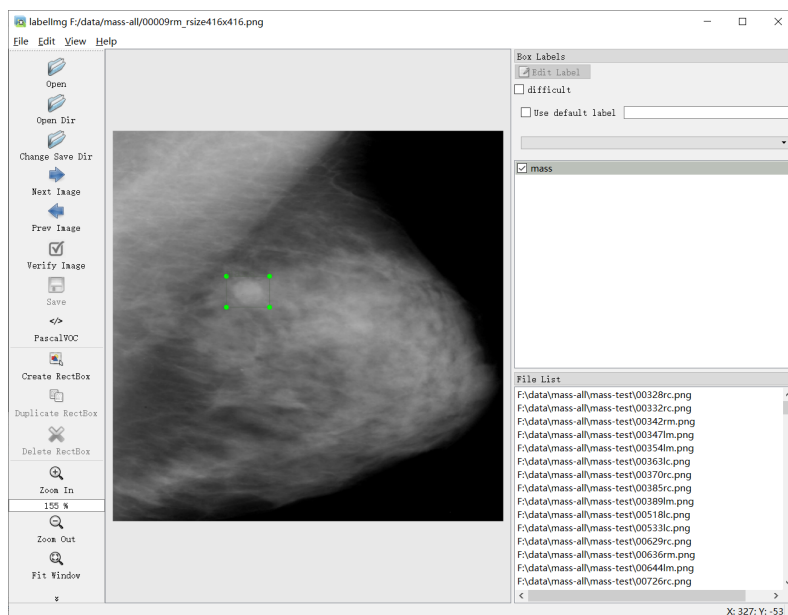


Figure 8. An example of one labeled image

图 8. 数据标注图例

乳腺钼靶图像中的肿块尺寸大多都属于小目标, 而原始的 YOLOv4 算法提供的锚框大小是通过 COCO 数据集聚类得来的, 难以适用本文数据集, 因此为了得到更符合本文检测目标的先验框, 在训练之前使用 K-means 算法对目标锚框进行聚类。参数设置方面, 考虑到网络训练效率方面, 输入图像尺寸设置为 416×416 ; 批次为一次训练需要用到的样本数, 它的大小影响到训练的速度, 考虑到 GPU 的内存, 批次设置为 8; Epoch 为对训练集的全部数据进行一次完整的训练, 根据数据集的样本量, 设置为 150。学习率是将输出的误差反向传播给网络参数, 用来拟合样本的输出, 本质上是一个优化的过程, 初始学习率设置为 0.001。

4.2. 评价指标

本文所采用的评价指标包括: 精确率(Precision), 敏感性(Sensitivity), 平均精确率(AP)、以及平均每幅图片假阳性个数(False Positive Marks Per Image, FPI)。敏感性用来评价漏诊率, FPI 表示每幅图像过检的肿块个数, 使用 P-R (precision recall)曲线评价模型的综合性能。计算方法如表所示, 其中, TP 表示将肿块目标预测为肿块目标, FP 表示将背景预测为肿块目标, FN 表示没有将肿块检测出来。此外, 本文采用 Times 表示检测单张图像所需的时间。

$$\text{precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (10)$$

$$AP = \frac{\sum \text{precision}}{n} \quad (11)$$

$$FPI = \frac{FP}{\text{Total number of image}} \quad (12)$$

$$\text{Times} = \frac{\text{Total time}}{\text{Total number of image}} \quad (13)$$

4.3. 实验结果与分析

本文在不同的主流模型上进行实验对比, 得到的测试结果见表 1。从表 1 可以看出, YOLOv3 平均单张假阳性 0.629 个, 敏感性 77.31%。YOLOv4 算法的单张假阳性 0.446 个, 敏感性为 78.68%。本文算法平均单张假阳性为 0.418 个, 相较于原 YOLOv4 网络减少 0.028 个, 敏感性为 81.49%, 相较于原网络提高 2.81%。YOLOv3 和 YOLOv4 单张图像检测时间分别为 0.014 和 0.026。本文改进算法单张检测时间为 0.038, 虽然改进算法的单张检测时间略低于原网络, 但是 FPI 以及敏感性相较于原网络均有显著提升。因此本文改进算法在牺牲较少检测时间的前提下, 检测性能有较好的表现。

Table 1. Comparison of evaluation indexes of different models

表 1. 不同模型的评价指标对比

模型	单张假阳性/个	敏感性/%	Times/s
Faster RCNN	0.881	80.76	/
YOLOv3	0.629	77.31	0.014
YOLOv4	0.446	78.68	0.026
本文算法	0.418	81.49	0.038

通过敏感性和精确率, 绘制出 P-R 曲线, 对网络的性能进行综合评价。P-R 曲线所围成的面积表示乳腺肿块检测的平均精度。对于图 9 分析得到, YOLOv4 网络的 AP 值为 82.58%, 本文改进的 YOLOv4 网络 AP 值达到 86.85, 相较于改进前提高了 4.27%。由此可知, 网络在改进后整体性能有所提升。

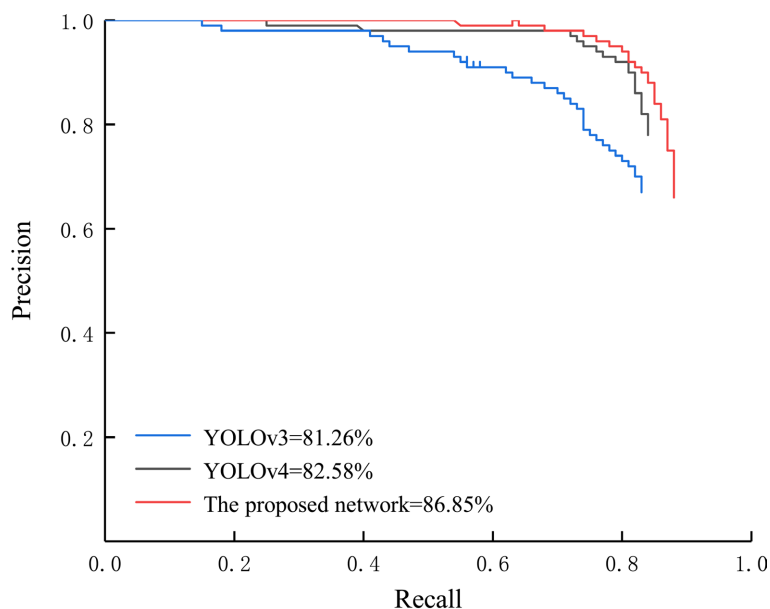


Figure 9. P-R curve of different network models
图 9. 不同网络模型的 P-R 曲线

乳腺肿块图像在 YOLOv4 和本文算法在数据集上的可视化检测结果如图 10 所示, 其中蓝色框表示标注框, 绿色框表示预测框与真实框交并比大于 0.5 的检测框, 红色框表示标注错误。观察可视化结果发现, YOLOv3 和 YOLOv4 网络都容易将与肿块相似的正常组织错判为肿块, 假阳性率较高。而本文方法更能准确的识别出正确的肿块, 避免将正常组织识别为肿块, 降低假阳性率。

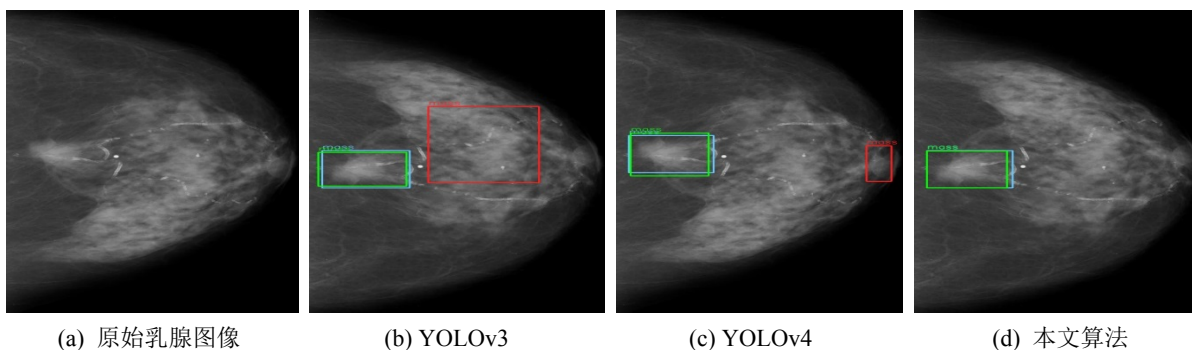


Figure 10. Visualization of mass detection
图 10. 肿块检测可视化

5. 总结

本文提出了一种基于改进 YOLOv4 算法的乳腺肿块检测方法。首先引入金字塔卷积, 使网络在通过多尺度卷积对输入图像进行特征提取; 其次将 YOLOv4 网络中的普通卷积替换成深度可分离卷积, 进一步提高网络识别速度; 最后在 YOLOv4 网络中增加一个尺度的输出特征图, 并且在 PANet 特征提取部分

引入注意力机制, 提升网络的识别能力。实验结果表明, 改进后的网络敏感性达到 81.49%, AP 值达到 86.85%, 单张图像假阳性个数为 0.418 个, 对于乳腺肿块的检测来说, 本文算法具有较好的表现。下一步研究将尝试设计肿块的多分类的网络, 使得研究更加符合临床需求。

参考文献

- [1] Siegel, R.L., Miller, K.D., Fuchs, H. and Jemal, A. (2021) Cancer Statistics. *CA: A Cancer Journal for Clinicians*, **71**, 7-33. <https://doi.org/10.3322/caac.21654>
- [2] Chung, J., Rajagopal, V., Nielsen, P., et al. (2008) A Biomechanical Model of Mammographic Compressions. *Biomechanics and Modeling in Mechanobiology*, **7**, 43-52. <https://doi.org/10.1007/s10237-006-0074-6>
- [3] Girshick, R., Donahue, J., Darrell, T., et al. (2013) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [4] Girshick, R. (2015) Fast R-CNN. *IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [5] Ren, S., He, K., Girshick, R., et al. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *International Conference on Neural Information Processing Systems*, Montreal, 7-12 December 2015, 91-99.
- [6] 郑杰, 郭朋. 一种改进 Faster R-CNN 检测乳腺肿块的方法[J]. 影像研究与医学应用, 2020, 4(14): 139-142.
- [7] Liu, W., Anguelov, D., Erhan, D., et al. (2016) SSD: Single Shot MultiBox Detector. *European Conference on Computer Vision*, Amsterdam, 11-14 October 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [8] Redmon, J., Divvala, S., Girshick, R., et al. (2016) You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [9] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
- [10] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. 2018 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 1-6.
- [11] 蒋慧琴, 王博霖, 马岭, 等. 一种双视图信息融合的乳腺肿块自动检测算法[J]. 郑州大学学报(理学版), 2020, 52(4): 28-36.
- [12] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection.
- [13] Duta, I.C., Liu, L., Zhu, F., et al. (2020) Pyramidal Convolution: Rethinking Convolutional Neural Networks for Visual Recognition.
- [14] Wang, C.Y., Liao, H., Yeh, I.H., et al. (2020) CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. *Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, 14-19 June 2020, 1571-1580. <https://doi.org/10.1109/CVPRW50498.2020.00203>
- [15] Howard, A.G., Zhu, M., Chen, B., et al. (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.
- [16] Lee, R.S., Gimenez, F., Hoogi, A., et al. (2017) A Curated Mammography Data Set for Use in Computer-Aided Detection and Diagnosis Research. *Scientific Data*, **4**, Article ID: 170177. <https://doi.org/10.1038/sdata.2017.177>
- [17] Heath, M., Bowyer, K., Kopans, D., et al. (2001) The Digital Database for Screening Mammography. Digital Mammography. Springer, Berlin.