



基于DYOLO神经网络的超声图像肾脏检测

刘奇¹, 赵丽霞², 郑曙光², 赵希梅^{1,3}

(1. 青岛大学 计算机科学技术学院, 山东 青岛 266071; 2. 青岛大学附属医院 腹部超声科, 山东 青岛 266003;

3. 山东省数字医学与计算机辅助手术重点实验室, 山东 青岛 266071)

摘要: 为便于慢性肾脏疾病的计算机辅助诊断, 提出一种基于DYOLO神经网络学习模型的自动超声图像肾脏检测方法。将YOLOv3和可变形卷积网络集成在一个端到端学习框架中, 使得DYOLO可根据肾脏的大小和形状自适应调节接收域, 以适应肾脏的各种纹理特征形变, 实现临床超声图像中肾脏的自动检测。在自制KidneyDetec超声图像肾脏检测数据集上的实验结果表明, 该方法在DYOLO网络模型的图像输入尺寸为416像素×416像素和608像素×608像素的情况下分别取得了89.6%和90.5%的平均精度均值, 相比基于深度学习的目标检测方法具有更高的检测速度和检测精度, 适用于慢性肾脏疾病的早期诊断。

关键词: 慢性肾脏疾病; 计算机辅助诊断; 深度神经网络; 超声图像; 目标检测

开放科学(资源服务)标志码(OSID):



中文引用格式: 刘奇, 赵丽霞, 郑曙光, 等. 基于DYOLO神经网络的超声图像肾脏检测[J]. 计算机工程, 2021, 47(7): 307-313.

英文引用格式: LIU Q, ZHAO L X, ZHENG S G, et al. Kidney detection using ultrasound image based on DYOLO neural network[J]. Computer Engineering, 2021, 47(7): 307-313.

Kidney Detection Using Ultrasound Image Based on DYOLO Neural Network

LIU Qi¹, ZHAO Lixia², ZHENG Shuguang², ZHAO Ximei^{1,3}

(1. College of Computer Science and Technology, Qingdao University, Qingdao, Shandong 266071, China;

2. Department of Abdominal Ultrasound, The Affiliated Hospital of Qingdao University, Qingdao, Shandong 266003, China;

3. Shandong Key Laboratory of Digital Medicine and Computer-assisted Surgery, Qingdao, Shandong 266071, China)

[Abstract] In order to facilitate the Computer Aided Diagnosis (CAD) of Chronic Kidney Disease (CKD), a kidney detection using ultrasound image based on DYOLO neural network model is proposed. The method integrates YOLOv3 and Deformable Convolutional Network (DCN) into an end-to-end learning framework, making DYOLO adaptively adjust the receiving area according to the size and shape of the kidney to adapt to various texture feature deformations, and realizes automatic kidney detection in clinical ultrasound images. Experimental results on a self-made ultrasound image-based kidney detection dataset, KidneyDetec, show that the mean Average Precision (mAP) of the proposed method reaches 89.6% when the image size of the DYOLO input is 416×416 pixels, and 90.5% when the image size of the DYOLO input is 608×608 pixels. Compared with deep learning-based target detection methods, the proposed method has higher detection speed and accuracy, displaying excellent applicability to the early diagnosis of chronic kidney diseases.

[Key words] Chronic Kidney Disease (CKD); Computer Aided Diagnosis (CAD); Deep Neural Network (DNN); ultrasound image; object detection

DOI: 10.19678/j.issn.1000-3428.0058565

0 概述

目前, 慢性肾脏疾病(Chronic Kidney Disease, CKD)已经成为全球性的公共卫生问题, 其特征是永

久性肾脏损害, 患病率和死亡率高^[1]。近十年来, 我国CKD总患病率为10.8%且逐年增高, 给我国经济和社会带来了沉重的负担。早期发现和早期干预可显著降低慢性肾病患者的并发症, 提高生存率。长

基金项目: 国家自然科学基金(61303079)。

作者简介: 刘奇(1994—), 男, 硕士研究生, 主研方向为医学图像处理、深度学习; 赵丽霞、郑曙光, 主治医师; 赵希梅(通信作者), 副教授。

收稿日期: 2020-06-08 修回日期: 2020-07-14 E-mail: 907892691@qq.com

期以来,肾脏穿刺活检是CKD评估和诊断的金标准^[2],但是随着其临床应用的深入,这种有创性检查的缺陷愈发突出,主要表现为容易使患者引起出血等并发症^[2],诊断结果受医生的主观因素影响,并且需要经验丰富的医生操作,否则会造成无谓的创伤。

自20世纪90年代以来,计算机辅助诊断(Computer Aided Diagnosis, CAD)在国内外医学影像学领域获得较快发展并逐渐应用于临床^[3],在辅助提高影像科医生的诊断准确率方面获得广泛认同^[4]。目前,医生进行慢性肾病超声诊断主要采用手动测量和裸眼视觉诊断方式,通过超声手动测量获取肾脏长度、肾实质厚度和肾实质回声强度等定量指标以实现诊断^[5],然而以上指标测量分析过程繁琐,每例耗时约20 min~30 min,而且某些病症的细微特征在短时间内肉眼不易察觉。计算机辅助诊断在超声图像特征识别方面具有医师肉眼识别无法比拟的优势。随着深度学习理论的发展,基于深度学习技术的超声图像CKD肾功能预测模型相继出现。Kuo等^[6]应用ResNet残差网络模型对1 297个病例进行分类,但样本图像需要从原始图像中手动获取感兴趣区域(Region of Interest, ROI),具有一定的局限性。JACKSON等^[7-9]利用深度学习技术对肾脏进行分割,获取肾脏相关形态参数。ZHENG等^[10]将迁移学习技术引入肾脏检测数据集研究中。

本文提出一种DYOLO神经网络学习模型,通过将YOLOv3和可变形卷积网络集成在一个端到端学习框架中,实现对超声图像中肾脏的自动检测,利用对肾脏超声图像进行检测后得到的检测框来获取肾脏长度、肾实质厚度和肾实质回声强度等超声参数,以辅助医生进行慢性肾脏疾病的诊断。

1 相关理论

1.1 可变形卷积原理

可变形卷积网络(Deformable Convolutional Network, DCN)^[11]由微软亚洲研究院计算机视觉组的研究员于2017年提出,其在卷积神经网络中引入学习空间几何形变的能力,使得可变形卷积网络能够更好地解决具有空间形变的图像识别任务。可变形卷积将偏移量添加到标准卷积的常规网格采样位置,这样可以使采样网格自由变形。这些偏移量是通过附加的卷积层从前面的特征图中学习得到的,变形以密集、局部和自适应的输入特征为条件。因此,其感受野能够根据目标大小,适应各种不规则形变,提取更精准的物体特征^[12]。

图1给出了卷积核尺寸均为3×3的4种采样网格结构。图1(a)表示标准卷积的常规采样网格,该

卷积受限于固定形状,采样能力有限。图1(b)在可变形卷积中增加偏移量的变形采样位置,可产生不规则的采样位置。图1(c)和图1(d)可视为图1(b)的特殊情况^[13],表明可变形卷积能根据比例、旋转和尺度进行变换,其感受野依据目标的大小和形状自适应调节,适应各种不规则形变,提取更精准的物体特征。

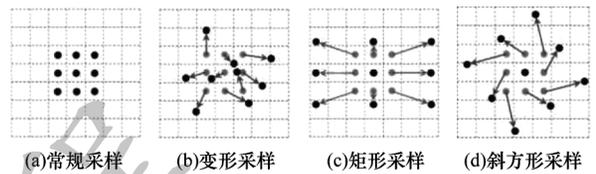


图1 4种采样网格结构

Fig.1 Four sampling grid structures

对于网格尺寸为1×1、卷积核大小为3×3的内核,常规网格G可形式化为:

$$G = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\} \quad (1)$$

将输出特征图y的每个位置 m_0 形式化为:

$$y(m_0) = \sum_{m_i \in G} w(m_i) \cdot x(m_0 + m_i) \quad (2)$$

其中: x 表示输入特征图; w 表示采样值的权重; m_i 表示G中的位置。

在可变形卷积中,常规网格G通过增强偏移量 Δm_i 得到:

$$y'(m_0) = \sum_{m_i \in G} w(m_i) \cdot x(m_0 + m_i + \Delta m_i) \quad (3)$$

由于偏移量 Δm_i 通常不是整数,因此应用双线性插值法确定偏移后的采样点值,并通过附加的卷积层学习偏移量 Δm_i 。

图2为可变形卷积结构,首先将输入的像素区域(input patch)通过一个小卷积层的输出(offset field)获得可变形卷积所需的偏移量(offsets)进而输出特征图(output feature map),然后将其作用在卷积核上达到可变形卷积效果。在得到偏移量的所有像素后,将得到的新图片作为输入数据传递到下一层。

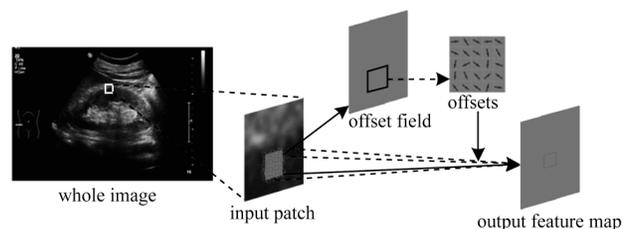


图2 可变形卷积结构

Fig.2 Structure of deformable convolution

综上所述,可变形卷积比标准卷积形式更灵活及广泛,对于具有空间几何形变的视觉任务有更好

的学习能力。与此同时,可变形卷积可能会产生一些计算开销,以更局部和自适应的方式进行卷积计算,而通过增加偏移量学习几何形变的思想还便于扩展到其他计算单元,由此可显著提升目标检测性能。

1.2 YOLOv3神经网络

随着卷积神经网络在计算机视觉领域的广泛应用,研究人员对卷积神经网络模型检测精度和速度的要求越来越高。REDMON等^[14-16]提出的YOLO系列神经网络将目标检测问题转化为回归问题,直接由图像像素优化得到物体边界位置和分类,相比Fast R-CNN^[17]、Faster R-CNN^[18]双阶段检测策略检测速度更快。YOLOv3^[14]是YOLO系列神经网络中速度和精度最均衡的目标检测网络,在业界得到广泛认可和应用。

图3给出了YOLOv3网络模型结构,YOLOv3具有以下特性:

1)YOLOv3的特征提取网络Darknet-53借鉴了Resnet^[19]的思想,引入残差模型(ResBlock),相比

YOLO9000^[15]使用的Darknet-19,网络更深且降低了梯度消失的风险,并且采用步长为2的卷积层代替池化层,避免了信息丢失问题。

2)YOLOv3采用多尺度融合方式进行预测,类似于特征金字塔网络(Feature Pyramid Network, FPN)^[20],利用非线性插值方法上采样(UpSampling)两次,获得3个不同尺寸(13像素×13像素、26像素×26像素、52像素×52像素)的特征图。由于深层且语义特征丰富的特征图负责预测大目标,浅层且几何特征丰富的特征图负责预测小目标,使得YOLOv3对于不同尺度的目标均具有较好的检测效果。

3)YOLOv3使用独立的逻辑分类器对目标进行分类识别,每个框利用多标签分类来预测边界框可能包含的类。在训练过程中使用二元交叉熵损失进行分类预测。

综上所述,YOLOv3算法不仅对于实物目标具有较好的预测效果,而且对于医学图像等目标同样具有较好的兼容性,可以取得较高的检测速度和检测准确率,并且简单易实现,实时性和鲁棒性更强。

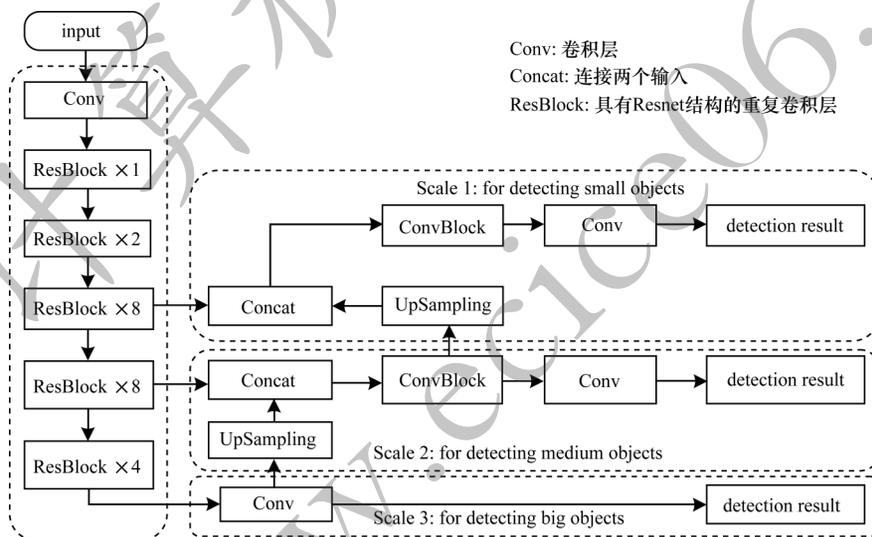


图3 YOLOv3网络模型结构

Fig.3 Structure of YOLOv3 network model

2 基于DYOLO的超声图像肾脏检测

2.1 Dec_Darknet-53特征提取网络

Darknet-53是YOLOv3的特征提取网络,在Imagenet中达到92.7%的Top-5测试准确率,领先于Darknet-19、Resnet-101和Resnet-152。Darknet-53网络包含53个接收域为3×3的卷积层,相比Darknet-19特征提取网络增加了残差模块,同时使用连续的3×3和1×1卷积层并且加入了shortcut连接,保证了在主干网络加深的同时不出现梯度消失现象,且网络训练效果更优。

可变形卷积(ConvOffset)可以为模型提供一个灵活的接收域,有利于检测超声图像中形态不规则的肾脏。本文将Darknet-53残差块中所有3×3卷积升级为可变形卷积,形成更强大的Dec_Darknet-53,作为DYOLO的特征提取网络。Dec_Darknet-53的网络模型结构如图4所示,其中,×n表示该模块重复n次,最后一列中的32、64、128、256、512、1024为通道数。Dec_Darknet-53强大而灵活的特征提取能力,为DYOLO检测网络提供了丰富的语义信息和几何信息,从而确保DYOLO的检测高效性和准确性。

	类型	核数量	核尺寸	步长	模型输出尺寸/像素
	Conv	32	3×3	1	416×416, 32
	Conv	64	3×3	2	208×208, 64
×1	Conv	32	1×1	1	208×208, 32
	ConvOffset	64	3×3	1	208×208, 64
	ResBlockoutput				208×208, 64
×2	Conv	128	3×3	2	104×104, 128
	Conv	64	1×1	1	104×104, 64
	ConvOffset	128	3×3	1	104×104, 128
×8	Conv	256	3×3	2	52×52, 256
	Conv	128	1×1	1	52×52, 128
	ConvOffset	256	3×3	1	52×52, 256
×8	Conv	512	3×3	2	26×26, 512
	Conv	256	1×1	1	26×26, 256
	ConvOffset	512	3×3	1	26×26, 512
×4	Conv	1 024	3×3	2	13×13, 1 024
	Conv	512	1×1	1	13×13, 512
	ConvOffset	1 024	3×3	1	13×13, 1 024

图4 Dec_Darknet-53网络模型结构
Fig.4 Structure of Dec_Darknet-53 network model

2.2 DYOLO网络模型

鉴于YOLOv3在自然图像检测方面的出色表现, 本文以YOLOv3为基础, 充分考虑肾脏超声图像纹理信息的多样性, 对Darknet-53添加可变形卷积, 以进行更有针对性且更有效的特征提取, 改进后的Dec_Darknet-53更加适合超声图像的训练, 同时本文

将调整模型输入大小, 以适应不同尺寸目标的检测。

因此, 融合YOLOv3和可变形卷积的优势, 本文提出一种新的网络模型DYOLO。该模型结构如图5所示, 主要由Dec_Darknet-53网络提取特征, 经多尺度特征融合得到3种不同尺寸的特征图, 进而实现识别和检测任务。

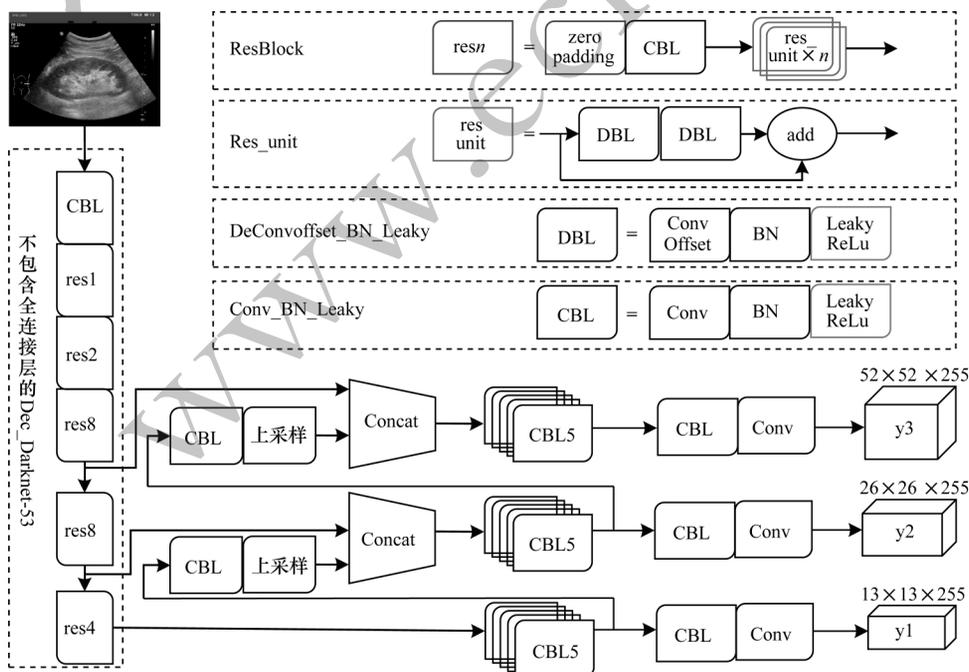


图5 DYOLO网络模型结构
Fig.5 Structure of DYOLO network model

图片在输入DYOLO网络后,首先经过多尺度调整至尺寸为416像素×416像素(或者为32倍数的其他尺寸)、通道数为3的网络模型输入,然后进入DYOLO的主干网络(backbone)Dec_Darknet-53进行特征提取:经过一层卷积操作后进入多个ResBlock层,其中卷积层为Conv+BN(批量归一化)+LeakyReLU(激活函数),ResBlock层中的resn包含 n 个残差单元(res unit),每个残差单元由ConvOffset(3×3)+BN+ReLU组成,每层ConvOffset(3×3)网络之前添加Conv(1×1)以减少特征图数量并提升网络计算效率和表达能力,形成Conv(1×1)+BN+ReLU+ConvOffset(3×3)+BN+ReLU的结构。整个网络的ResBlock部分采用串联方式,其输出直接作为后续网络的输入。

后续网络为检测网络,主要任务是识别分类和预测目标检测框。DYOLO应用多尺度融合方式形成金字塔网络,提供3种尺寸不一的边界框。本文从backbone的后面多个图层中得到特征图并进行2次上采样,再从网络更早的图层中获得特征图,将高低2种分辨率的特征图相融合,从而找到早期特征映射中的上采样特征和细粒度特征,并获得更有意义的语义信息。之后,通过添加多个卷积层来处理该组合特征映射,最终的卷积层会预测出一个三维张量编码:边界box坐标,目标置信分数,各类物体的分类概率。

本文以尺寸为1024像素×768像素、通道数为3的肾脏超声图像样本作为模型输入,通过多尺度调整生成尺寸为416像素×416像素、通道数为3的模型输入进行特征提取、目标检测和分类识别,最终得到检测结果。

DYOLO在训练过程中使用锚框回归方式进行目标框预测,利用二元交叉熵损失进行类别预测,采用Adam优化器优化网络模型,最终通过逻辑分类器对训练样本的各类分类概率进行预测。二元交叉熵函数 $H(p, q)$ 表示真实值与预测值之间的差异,假设 x_i 为训练样本 x 的第 i 个样本, $y \in \{1, 2, \dots, Y\}$ 表示训练样本的各类分类概率,本文设定 y 为2,则交叉熵函数表示为:

$$H(p, q) = - \sum_{i=1}^m p(x_i) \ln q(x_i) \quad (4)$$

其中: $p(x_i)$ 是预测的概率值; $q(x_i)$ 是真实的概率值。

3 实验与结果分析

3.1 实验环境

实验开发环境为Windows10 64位操作系统、内存32.00 GB, Intel® Xeon™ W-2133处理器,显卡为NVIDIA GeForce GTX 1080Ti,显存为11 GB。在Anaconda3中的Spyder3.4平台下使用深度学习框架Pytorch1.2.0 GPU版本进行实验。实验结果的可视

化处理由tensorboard、pillow和matplotlib库实现。

3.2 肾脏检测数据集

实验使用LabelImg开源标签工具,根据Pascal VOC公共数据集自制肾脏检测数据集KidneyDetec。数据集图像来源于青岛大学附属医院,个人隐私信息均已从图像中剔除,共包含2911张超声图像和2911个标签文件。图像标注由一名临床超声医师手动完成,每张图像可标注为kidney_outside和kidney_pelvis 2类。KidneyDetec数据集图像收集自728名受试者,所有受试者均有双侧肾脏图像,保证了样本的代表性和有效性。所有图像尺寸为1024像素×768像素、位深度为24的肾脏长轴切面,超声仪器采用飞利浦超声扫描仪与腹部突阵探头。肾脏检测数据集中超声图像示例如图6所示。



图6 肾脏检测数据集超声图像示例
Fig.6 Examples of ultrasound images in Kidney Detection dataset

3.3 结果分析

实验在KidneyDetec肾脏检测数据集上进行,为证明本文提出的DYOLO网络模型的有效性,选取目标检测领域主流的双阶段检测模型(Fast R-CNN、Faster R-CNN)和单阶段检测模型(SSD300^[21]、YOLOv3)作为对比模型。

实验参数设置如下:epoch训练轮次为100,每次迭代输入样本的batch size为8,梯度优化使用Adam优化器,初始学习率为0.001,衰减系数为0.0005,动量为0.9,IoU阈值设置为0.5。实验数据集分为训练集和测试集,训练集和测试集分别随机分配2474张和474张图像。各主流目标检测模型对比结果如表1所示,其中各目标类别的平均精度均值(mean Average Precision, mAP)是目标检测任务中常用的

评价指标。可以看出,在相同数据集上本文提出的DYOLO检测模型的平均精度均值达到90.5%,高于Fast R-CNN和Faster R-CNN双阶段目标检测模型以及SSD和YOLOv3单阶段目标检测模型,能有效辅助医师进行CKD诊断。与此同时,本文还对比了网络模型输入尺寸为416像素×416像素和608像素×608像素时的mAP,结果表明本文网络模型对于大尺寸图像(1 024像素×768像素)具有更好的检测效果。

表1 平均精度均值对比

Table 1 Comparison of mAP

网络模型	主干网络	mAP/%
Fast R-CNN	VGG-16	73.6
Faster R-CNN	VGG-16	76.9
SSD300	VGG-16	77.8
YOLOv3-416	Darknet-53	86.8
DYOLO-416	Dec_Darknet-53	89.6
YOLOv3-608	Darknet-53	87.6
DYOLO-608	Dec_Darknet-53	90.5

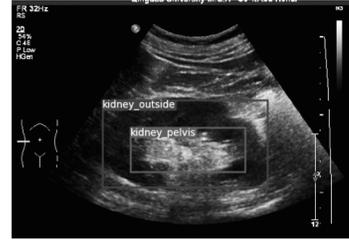
如表2所示,本文对YOLOv3和DYOLO网络模型的检测速度进行对比。可以看出,添加了可变形卷积后的DYOLO网络模型检测速度稍有降低,但是该检测速度依然可满足实时检测的应用需求,保障临床辅助诊断的高效性和实用性。

表2 目标检测速度对比

Table 2 Comparison of object detection speed

网络模型	检测速度/(frame·s ⁻¹)
YOLOv3-416	57
DYOLO-416	53

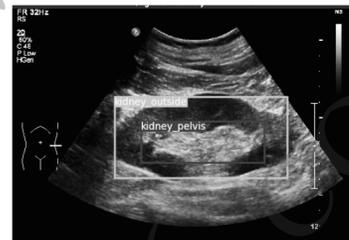
图7为原始YOLOv3网络模型与本文提出的DYOLO网络模型的检测效果对比图,其中肾脏超声图像选取自KidneyDetec肾脏检测数据集的测试集。图7(a)表示使用YOLOv3算法的肾脏右侧超声图像检测效果图,图7(b)表示使用DYOLO算法的肾脏右侧超声图像检测效果图,图7(c)表示使用YOLOv3算法的肾脏左侧超声图像检测效果图,图7(d)表示使用DYOLO算法的肾脏左侧超声图像检测效果图。可以看出,使用YOLOv3网络模型的检测结果存在检测框与目标贴合不精确的情况,而利用DYOLO网络模型检测的边界框更加贴合肾脏轮廓,边缘信息更加精细,从而证明DYOLO网络模型可以更全面完整高效地学习到不同形状的目标的特征信息,显著提升了检测效果。通过综合以上网络模型的检测效果和检测速度可知,本文提出的DYOLO网络模型既能取得较高的平均精度均值,又具有较好的实时性和鲁棒性,适用于临床辅助诊断^[22]。



(a)YOLOv3检测效果图1



(b)DYOLO检测效果图1



(c)YOLOv3检测效果图2



(d)DYOLO检测效果图2

图7 目标检测效果对比

Fig.7 Comparison of object detection effect

4 结束语

肾脏超声检查在肾脏疾病的诊治中具有重要作用,然而临床上超声检查受多种因素影响,容易导致超声图像质量欠佳,使得超声图像中的肾脏检测仍具有一定的挑战性。本文提出一种基于DYOLO网络模型的超声图像肾脏检测方法,通过将YOLOv3和可变形卷积网络集成在一个端到端学习框架中,实现临床超声图像中肾脏的检测。实验结果表明,该方法在DYOLO网络模型输入尺寸为608像素×608像素的情况下取得了90.5%的平均精度均值,并且相比对比方法具有更高的检测速度和检测精度,保证了CKD计算机辅助诊断的可靠性及临床应用的实用性。后续将拓展DYOLO网络模型在医学领域的应用范围,并利用多任务深度学习技术进一步提升计算机辅助诊断的应用效果。

参考文献

- [1] WEBSTER A C, NAGLER E V, MORTON R L, et al. Chronic kidney disease[J]. *Lancet*, 2017, 389(10075): 1238-1252.
- [2] 朱丹丹,李世军,黄丽璇,等. 降低肾活检并发症发生率[J]. *中国卫生质量管理*, 2018, 25(2): 80-83.
ZHU D D, LI S J, HUANG L X, et al. Reducing the incidence of renal biopsy complications[J]. *Chinese Health Quality Management*, 2018, 25(2): 80-83. (in Chinese)
- [3] 宋波,宋同峰. 慢性肾病辅助诊断系统的设计与实现[J]. *计算机测量与控制*, 2019, 27(10): 50-54.
SONG B, SONG T F. Design and implementation of chronic kidney disease auxiliary diagnosis system [J]. *Computer Measurement and Control*, 2019, 27(10): 50-54. (in Chinese)
- [4] CASTELLINO R A. Computer Aided Detection (CAD): an overview[J]. *Cancer Imaging*, 2005, 5: 17-19.
- [5] 庞衍平,王培军,王秀艳. 超声造影定量分析技术诊断早期慢性肾病的临床价值[J]. *医学影像学杂志*, 2016, 26(8): 1460-1463.
PANG Y P, WANG P J, WANG X Y. The clinical value of ultrasound contrast quantitative analysis technology in the diagnosis of early chronic kidney disease [J]. *Journal of Medical Imaging*, 2016, 26(8): 1460-1463. (in Chinese)
- [6] KUO C C, CHANG C M, LIU K T, et al. Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning[J]. *Digital Medicine*, 2019, 2(1): 1-9.
- [7] JACKSON P, HARDCASTLE N, DAWE N, et al. Deep learning renal segmentation for fully automated radiation dose estimation in unsealed source therapy[J]. *Frontiers in Oncology*, 2018, 8: 215-224.
- [8] YIN S, PENG Q M, LI H M, et al. Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks[J]. *Medical Image Analysis*, 2020, 60: 101602.
- [9] NITHYA A, APPATHURAI A, VENKATADRIN, et al. Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images[J]. *Measurement*, 2020, 149: 106952.
- [10] ZHENG Q, TASTAN G, FAN Y. Transfer learning for diagnosis of congenital abnormalities of the kidney and urinary tract in children based on ultrasound imaging data [C]//*Proceedings of IEEE International Symposium on Biomedical Imaging*. Washington D. C., USA: IEEE Press, 2018: 97-104.
- [11] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks [C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Washington D. C., USA: IEEE Press, 2017: 127-132.
- [12] 郭正华,陈立福. 基于自适应感受野的DC-SSD目标检测算法[J]. *工业控制计算机*, 2019, 32(5): 90-91, 94.
GUO Z H, CHEN L F. DC-SSD target detection algorithm based on adaptive receptive field [J]. *Industrial Control Computer*, 2019, 32(5): 90-91, 94. (in Chinese)
- [13] 刘宗宝. 基于深度学习的水下视频目标检测与跟踪[D]. 哈尔滨: 哈尔滨工程大学, 2019.
LIU Z B. Underwater video target detection and tracking based on deep learning [D]. Harbin: Harbin Engineering University, 2019. (in Chinese)
- [14] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. [2020-05-12]. <https://arxiv.org/abs/1804.02767>.
- [15] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2017: 6517-6525.
- [16] REDMON J, DIVVALA S K, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2016: 779-788.
- [17] GIRSHICK R. Fast R-CNN [C]//*Proceedings of IEEE International Conference on Computer Vision*. Washington D. C., USA: IEEE Press, 2015: 1440-1448.
- [18] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [19] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//*Proceedings of IEEE Conference on Computer Vision & Pattern Recognition*. Washington D. C., USA: IEEE Press, 2016: 770-778.
- [20] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2017: 227-233.
- [21] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//*Proceedings of European Conference on Computer Vision*. Berlin, Germany: Springer, 2016: 127-131.
- [22] GAO F, YOON H, WU T, et al. A feature transfer enabled multi-task deep learning model on medical imaging [J]. *Expert Systems with Applications*, 2020, 143: 112957.