DOI:10.3969/j.issn.1005-202X.2025.02.012

医学信号处理与医学仪器

改进YOLOv5s的堆叠医疗器械检测算法

田昌锐[1],廖薇[1],徐震[2]
1.上海工程技术大学电子电气工程学院,上海 201620;2.上海工程技术大学机械与汽车工程学院,上海 201620

【摘要】针对医疗器械堆叠问题和提升医疗器械识别准确率,提出一种改进YOLOv5s的医疗器械检测方法。首先使用C2f模块优化YOLOv5s网络提升模型识别精度,其次在特征融合网络引入SENet,提升模型对有效信息的关注度,最后在DIOU损失函数的基础上引入Alpha交并比(α–IOU)构成α-DIOU,使边界框回归更加准确,精确定位图像中的医疗器械。实验结果表明,改进后的模型在验证集中对医疗器械的精确率、召回率、平均精度均值分别达到81.8%、93.7%、91.5%,相比于YOLOv5s模型分别提升3.2%、3.4%、4.6%。本研究方法简单有效,有望为医疗器械的检测方法提供新思路。

【关键词】医疗器械;YOLOv5s;注意力机制;α-DIOU;深度学习
【中图分类号】R318;TP391.41 【文献标志码】A 【文章编号】1005-202X(2025)02-0220-07

Detection of stacked medical devices using improved YOLOv5s

TIAN Changrui[1], LIAO Wei[1], XU Zhen[2]
1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China; 2. School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Abstract: A medical device detection method based on improved YOLOv5s was proposed to solve the problem of medical device stacking and further improve the detection accuracy of medical devices. The proposed method uses C2f module to optimize YOLOv5s network for improving the detection accuracy, introduces squeeze-and-excitation network into the feature fusion network for improving the model's attention to effective information, and constructs α-DIOU by introducing Alpha intersection union ratio (α−IOU) on the basis of the distance-intersection over union (DIOU) loss function, which makes the bounding box regression more accurate and enables the accurate detection of medical devices in the image. Experimental results show that the improved model has precision, recall rate and mean average precision of 81.8%, 93.7% and 91.5%, respectively, for medical device detection on the validation set, which are 3.2%, 3.4% and 4.6% higher than YOLOv5s model. The proposed method is simple and effective, and is expected to provide new ideas for the detection methods of medical devices.

Keywords: medical device; YOLOv5s; attention mechanism; α-DIOU; deep learning

前 言

实验室、医院、仓库、教学场地等地点医疗器械人工管理成本巨大且容易出现错误,并且对医疗器械进行自动识别有可能减少医疗事故的发生,降低风险。目前,目标检测算法发展已经相对成熟且涌现出多种算法,例如传统目标检测算法、深度学习目标检测算法。深度学习的发展摆脱人工提取特征的困境,实现医疗器械的自动检测[1]。目标检测按照深度学习分类可以分为单阶段和两阶段。两阶段目标检测算法主要基于候选区选取再使用卷积神经网络(Convolutional Neural Network, CNN)分类,主要代表有RCNN[2]、Fast R-CNN[3]、Faster R-CNN[4]、Mask R-CNN[5]、SPP-NET[6]等,RCNN系列首次将CNN用于目标检测,推动目标检测的进展[7]。单阶段目标检测算法不需要像两阶段那样提前生成候选框,通过把检测的问题变为单一的回归任务。通过直接回归目标的边界框位置和类别信息直接得到目标的位置和类别,而不需要额外的候选框生成步骤,代表作包括YOLO[8]、YOLOv2[9]、YOLOv3[10]、YOLOv4[11]系列

【收稿日期】2024-09-04
【基金项目】国家自然科学基金(62001282)
【作者简介】田昌锐,硕士研究生,研究方向:计算机视觉,E-mail: 1877132138@qq.com
【通信作者】廖薇,博士,副教授,研究方向:计算机视觉,E-mail: liaowei54@126.com;徐震,博士,副教授,研究方向:计算机视觉,E-mail: lcxuzhen@163.com

模型和SSD[12]网络。一阶段算法不需要候选框，直接完成检测任务，因此算法的检测速度快，二阶段算法相比于一阶段算法，虽然精度高，但是检测速度慢。很多学者对医疗器械检测进行研究。Abdulbaki等[13]提出使用CNN训练模型，再使用两个LSTM模型分别剪辑时间信息，对手术视频的时间依赖性进行建模来检测腹腔视频中的医疗器械。Hasan等[14]利用深度学习和几何3D视觉将CNN与代数几何相结合来对医疗器械进行检测。Jin等[15]将视频帧输入Faster R-CNN，输出7种医疗器械中任何一个手术器械的空间坐标。Zhang等[16]提出一种基于Faster R-CNN的网络用于检测腹腔镜手术中的医疗器械。

针对检测速度的问题，本文选用高实时性的YOLOv5s作为基础算法，YOLOv5s在检测重叠目标时表现较差，本文使用的数据集包括较多的重叠目标，对算法来说是一种考验。本文决定对YOLOv5s模型进行改进，以获得更好的重叠目标检测结果。

1 YOLOv5s结构

YOLOv5算法是一种检测速度较快的目标检测算法，根据深度和宽度划分可以分为4种不同的版本，分别是s、m、l、x，它们的深度宽度越来越大。该算法将图像划分成单元格，预测所有单元格的边框、置信度等信息，使用非极大抑制处理多个边框，删除重复和低置信度的边框。本文选取结构最小的YOLOv5s作为基础算法进行改进。其网络结构由骨干网络（Backbone）、颈部网络（Neck）和预测部分（Prediction）构成，YOLOv5s网络结构如图1所示。Backbone包括Conv、C3和SPPF模块，使用CSPNet来提取特征。Conv是网络中的卷积操作，负责提取特征图中的空间与通道信息。C3由Conv和残差结构（Bottleneck）组成，参考跨阶段局部网络，实现不同层次的特征融合。SPPF主要作用是特征提取，这一步骤通过池化层进行，它的池化核大小不同。Neck由PANet和FPN组成，它的作用是将低高层特征进行融合，使其更好地提取特征。融合后的特征传入预测层，预测部分使用Detect组成，它的损失函数是CIOU，使用NMS选出预测框，实现3种不同尺度的目标预测。

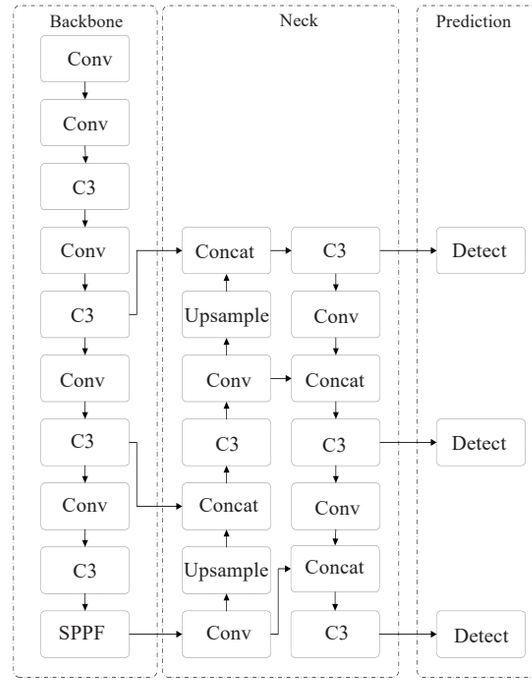


图1 YOLOv5s网络结构图

Figure 1 Network structure of YOLOv5s

2 YOLOv5s目标检测算法改进

2.1 C2f模块

在C2f模块中，输入的特征图经过第一个卷积层后会被划分为两个部分，两个部分都会使用不同的卷积进行处理，处理过后再将这两个部分拼接，通过这一步骤能够使模型得到更加丰富的上下文信息，从而让模型对目标的识别更加准确，提高精确度。拼接后的特征图要经过第二个卷积再次进行卷积操作，然后才能最终输出。C2f模块参考Bottleneck的设计思想进行设计，将输入的特征图分成两个部分，这种设计方法可以提高模型的非线性表示能力，因此能够对复杂的图像特征进行处理。C2f模块中有多个Bottleneck，每个Bottleneck模块都由两个卷积层构成，它们的作用是对图片的特征图进行卷积，通过卷积操作提取出高级特征。C2f模块还采用C3模块和ELAN的设计方法，和C3模块相比，C2f模块少了一次卷积层，并且使用Split对特征划分成两个部分，使用更多的跳层连接，这样设计的好处是在特征提取时可以获得更多的梯度流。C3和C2f模块结构图如图2所示。

C2f模块可以交互浅层与深层的医疗器械特征信息，再将信息进行特征增强可有效提升医疗器械检测精度并降低漏检率，提升医疗器械目标检测精度。C2f模块可以增强网络的特征融合能力，它通过融合低级特征图和高级特征图，实现在不同尺度上利用语义信息和细节信息提高对医疗器械目标检测的准确性和鲁棒性，提升检测能力。王志新等[17]把C2f模块替换C3模块对行人进行检测，实现性能提升。通过引入C2f模块，让模型更有效地对复杂特征进行捕获，从而在目标检测任务中取得更显著的表现。

a：C3模块结构图　　b：C2f模块结构图

图2 C3与C2f模块结构图

Figure 2 C3 and C2f module diagrams

2.2 压缩和激励网络（Squeeze and Excitation Networks, SENet）

SENet亦称SE注意力机制，它是一种通过加权每个通道的特征图来增强神经网络性能的方法，它能够自适应地学习通道之间的重要性权重，并根据这些学习到的权重来调整特征图。这种注意力机制可以使算法更加关注对小目标更敏感的通道，通过分配更多的权重来提高对这些关键特征的提取能力。SENet网络结构如图3所示。SENet的工作如下：首先输入的是$H\times W\times C$的特征图，通过全局平均池化，将张量转化为一个$1\times1\times C$张量，然后经过两个全连接层，其中一个全连接层对张量进行降维，另一个进行升维从而增加非线性处理过程，通过这种操作可以把通道相关性更好地进行拟合，经过Sigmod激活函数层后得到$1\times1\times C$的特征图，最后将原始的$H\times W\times C$和$1\times1\times C$的特征图全乘得到不同通道的特征图。谭义镇等[18]在YOLOv4中添加SE注意力机制对行人进行检测，实现性能提升。

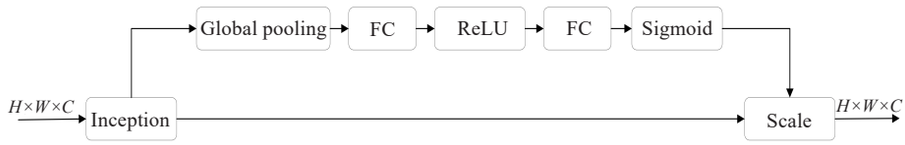


图3 SENet网络结构图

Figure 3 SENet structure

2.3 α-DIOU损失函数

医疗器械部分被遮挡，模型对遮挡识别具有一定的困难，并且为了让预测框在回归时与目标更贴近，提高模型的预测效果，考虑将CIOU损失函数替换为α-DIOU损失函数。边界框（Bounding box, Bbox）定位损失直接决定模型的目标定位能力。α-IOU可以超过其它IOU损失函数，本文在α-IOU的基础上将α-IOU改为α-DIOU，可以通过α灵活性调节Bbox回归精度，本文将α设置为3。

IOU是计算预测框与真实框的交并比。α-DIOU计算如式（1）所示，IOU^{α}计算如式（2）所示：

$$L_{\alpha\text{-}DIOU}=1-IOU^{\alpha}+\frac{\rho^{2\alpha}\left(b,b^{gt}\right)}{c^{2\alpha}} \tag{1}$$

$$L_{\alpha\text{-}IOU}=\frac{1-IOU^{\alpha}}{\alpha},\alpha>0 \tag{2}$$

α-DIOU对锚框和目标框之间的归一化距离进行建模，如图4所示。b代表锚框（图中黄色虚线）的中心点，b^{gt}代表目标框（图中红色虚线）的中心点，ρ代表计算b与b^{gt}之间的欧式距离。存在一个最小矩形来覆盖锚框与目标框（图中绿色虚线框），c代表这个最小矩形的对角线长度。

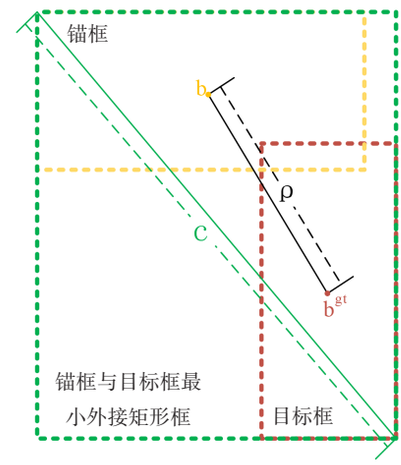


图4 α-DIOU锚框与目标框距离表示

Figure 4 Distance between anchor box and target box of α-DIOU

2.4 改进后的网络结构

主要改进如下：首先，通过引入C2f模块来改善模型的感受野和多尺度学习能力。其次，进行多种注意力机制的对比分析，并选择使用SENet，并将其整合到网络的第14层中学习通道的重要性权重来调

整特征图。最后，采用α-DIOU损失函数代替CIOU损失函数，以优化边界框的定位精度，进一步提升模型在目标检测任务中的性能。改进后的网络框架如图5所示。

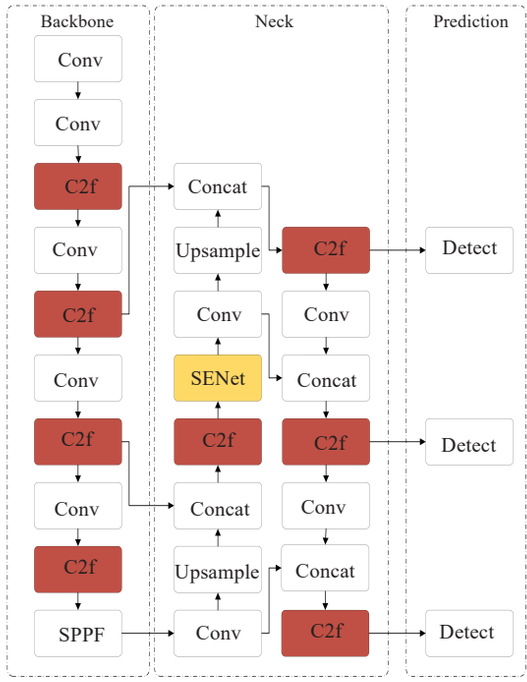


图5 改进后的YOLOv5s结构图

Figure 5 Improved YOLOv5s structure

3 实验与结果分析

3.1 实验环境

本实验使用的CPU为Intel(R) Core(TM) i7-8750H，GPU为NVIDIA GeForce GTX-1060 8 GB，8 GB运行内存，Windows10操作系统，深度学习框架PyTorch2.0.1，语言Python3.8.10，CUDA11.8。本文设置的参数：训练100轮，使用SGD优化器，学习率设置为0.01，使用GPU，batch-size设置为12，图像分辨率设置为640。

3.2 数据集

本文以8种医疗器械作为研究对象，其中5种剪刀类型的医疗器械具有部分相似性，8种医疗器械及部分数据集如图6所示。采集的数据集包含单目标、多目标、多种距离、堆叠等情况，使用数据增强随机遮挡的方式对部分数据进行数据增强，以增强数据的多样性。采用手机相机进行拍摄，设计单目标、双目标、四目标组合，从多种距离、多种角度、不同位置拍摄，让目标出现在图片的不同位置以及使目标出现多种角度，其中四目标占比37.5%，双目标占比30.0%，使每一种医疗器械出现次数均等。采集图片2 200张，采集完成后使用Labelimg标注工具对所有的医疗器械进行标注，标注时使用最小矩形框将目标全部包括。最后按照7:3:1分为训练集、验证集、测试集，其中训练集1 400张，验证集600张，测试集200张。

a：本实验涉及的8种医疗器械

b：部分4目标堆叠及遮挡数据集

图6 8种医疗器械及部分数据集

Figure 6 Eight types of medical devices and some of datasets

3.3 评价指标

本文选取平均精确度均值（mean Average Precision，mAP）、精确率（Precision，P）、召回率（Recal，R）[19]作为主要的评价指标，公式分别如下所示：

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{TP}{TP + FN} \tag{4}$$

$$AP = \int_0^1 P(R)\,dR \tag{5}$$

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{6}$$

其中，AP表示平均精度；TP（真正例）表示正样本被正确识别；FP（假正例）表示负样本被误报；TN（真反例）表示负样本被识别；FN（假反例）表示漏报正样本。

3.4 C2f模块对模型的影响

为验证C2f模块对模型的作用，本文对替换前后的模型效果进行对比，结果如表1所示。从表1可知，使用C2f模块后模型大小增加4.3 MB，参数量从7.0 M增加到9.2 M，增加2.2 M，但是改进后的模型在精确率、召回率、平均精度均值相比YOLOv5s都得

到提升，分别提升0.4%、0.6%、2.4%。虽然YOLOv5s-C2f模型的参数量、模型大小相比于YOLOv5s模型都增加，但是YOLOv5s-C2f模型提升精确率、召回率、平均精度均值这3个重要指标。

表1 使用C2f模块后的比较结果（验证集）

Table 1 Results after using C2f (Validation set)

模型	精确率/%	召回率/%	平均精确度均值/%	参数量/M	模型大小/MB
YOLOv5s	78.6	90.3	86.9	7.0	13.7
YOLOv5s-C2f	79.0	90.9	89.3	9.2	18.0

3.5 不同注意力机制对比分析

为了选择比较适合本实验数据集的注意力机制，本文在C2f的基础上（YOLOv5s-C2f）进行添加注意力机制进行实验。首先使用不同的注意力机制，本文选取比较常见的几种注意力机制进行注意力机制的对比实验，如CA[20]、ECA[21]、CBAM[22]、SE[23]，将这4种注意力机制放在改进模型的不同层来观察实验结果。同时为了找出这4种注意力机制放在YOLOv5s-C2f模型的某层较为合适，本实验将每一种注意力机制分别放在模型的第3、5、9、10、14、18层，使每一种注意力机制均等出现在模型的每个位置。

SE注意力机制在不同层表现结果如表2所示。从表2可知，在YOLOv5s-C2f中的不同位置添加注意力机制后，虽然召回率全部得到提升，但是大多数其他评价指标都有所下降。当SE被添加到第3、5、9、10、18层时，精确率相比YOLOv5s-C2f分别降低5.3%、5.4%、1.1%、1.2%、2.3%，说明在这几层添加SE注意力机制抑制了医疗器械数据集的通道特征，使精确率和平均精确度均值都出现不同程度的下降，也说明使用具有相似目标的数据集增加了数据集的识别复杂度。当SE注意力机制添加到第14层时，平均精确度均值相比YOLOv5s-C2f降低0.2%，但是精确率提升1.6%，召回率提升0.1%，说明第14层的SE注意力机制更加注重通道特征，增强特征之间的关系，提升精确率与召回率的数值。综上所述，在YOLOv5s-C2f模型第14层添加SE注意力机制适合本实验所使用的数据集。

ECA注意力机制在不同层表现结果如表3所示。由表3可知，将ECA加入第14、18层时，精确率分别提升0.8%、1.3%，但是召回率与平均精度均值都出现不同程度降低。将ECA加入到其它层，只有第5层的召回率与YOLOv5s-C2f保持持平。ECA增加少量参数可以获得性能提升，在处理上下文依赖和通道关系上效果不佳。ECA注意力机制在本数据集的表现不如SE注意力机制。

表2 SE注意力机制在不同层的比较结果（验证集）（%）

Table 2 Results of SE attention mechanism at different layers (Validation set) (%)

注意力添加位置	精确率	召回率	平均精确度均值
Layer=3	73.7	92.1	85.5
Layer=5	73.6	91.4	84.1
Layer=9	77.9	92.6	87.3
Layer=10	77.8	92.3	87.9
Layer=14	80.6	91.0	89.1
Layer=18	76.7	91.3	87.0

表3 ECA注意力机制在不同层的比较结果（验证集）（%）

Table 3 ECA attention mechanism results at different layers (Validation set) (%)

注意力添加位置	精确率	召回率	平均精确度均值
Layer=3	75.8	89.4	85.6
Layer=5	74.2	90.9	85.5
Layer=9	78.3	89.9	87.6
Layer=10	75.7	92.8	86.9
Layer=14	79.8	90.3	89.2
Layer=18	80.3	90.5	88.7

CBAM虽然可以从空间和通道上对目标进行关注，但是需要更多的计算资源，计算复杂度高。将CBAM插入到网络的不同位置出现的结果如表4所示。由表4可知，将CBAM插入到第3、5层时的精确率与平均精确度均值最低，由此可知，将CBAM加入到此位置不仅增加复杂度，而且3项指标都出现下降的趋势，不能对医疗器械进行精确识别，而且可能出现误判的情况。将CBAM加入到其他层也没有提升重要指标，说明CBAM注意力机制不适合本文使用的数据集。

CA注意力机制具有高效、新颖的特性，使用CA

表4 CBAM注意力机制在不同层的比较结果（验证集）（%）
Table 4 CBAM attention mechanism results at different layers (Validation set) (%)

注意力添加位置	精确率	召回率	平均精确度均值
Layer=3	71.9	90.0	82.7
Layer=5	74.9	90.6	84.6
Layer=9	77.9	91.9	89.0
Layer=10	78.7	91.9	87.6
Layer=14	77.8	90.1	86.8
Layer=18	76.2	89.9	86.9

可以让模型获取更大的范围信息，提升特征捕捉能力从而提升检测性能，CA注意力机制在网络不同层的表现如表5所示。由表5可知，将CA注意力机制加在不同网络层，模型对医疗器械的检测性能有所下降，所有添加的网络层模型都出现明显的精确率下降，同时模型的平均精度均值也出现下降，召回率有提升也有下降，召回率的下降幅度相比于精确率、平均精度均值要小，说明CA不适合本实验的模型对医疗器械的识别。

表5 CA注意力机制在不同层的比较结果（验证集）（%）
Table 5 CA attention mechanism results at different layers (Validation set) (%)

注意力添加位置	精确率	召回率	平均精确度均值
Layer=3	72.7	91.0	84.3
Layer=5	76.0	91.2	86.8
Layer=9	77.8	90.6	87.1
Layer=10	78.0	90.8	87.4
Layer=14	76.8	91.7	86.4
Layer=18	78.5	91.4	88.8

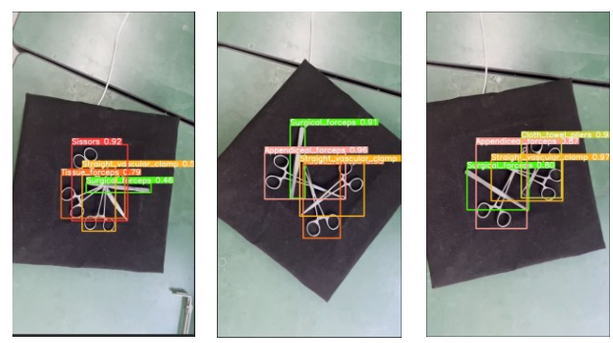
3.6 损失函数对比实验

通过分析不同注意力机制的对比实验可知SE注意力机制在其他注意力机制中效果最佳，而且通过实验可知将SE放在第14层效果优于其他层，因此本文在YOLOv5s-C2f模型的第14层添加注意力机制的情况下继续实验。为了找出适合本实验的损失函数，本文设置GIOU[24]、DIOU[25]、SIOU、EIOU、α-DIOU进行对比实验，将添加SE注意力机制模型的CIOU分别替换为这5种损失函数，不同损失函数的表现结果如表6所示。当损失函数为α-DIOU，3项指标全部提升，相比于添加SE注意力机制模型的CIOU，精确率、召回率、平均精度均值分别提升1.2%、2.7%、2.4%，相比于表1中改进前YOLOv5s分别提升3.2%、3.4%、4.6%，提升了检测性能，说明α-DIOU在本实验的有效性。α-DIOU相比于其他损失函数适合本文的数据集。

表6 不同损失函数对模型的影响（验证集）（%）
Table 6 Effects of different loss functions on the model (Validation set) (%)

损失函数	精确率	召回率	平均精确度均值
CIOU	80.6	91.0	89.1
GIOU	76.6	91.9	86.8
DIOU	81.7	91.8	91.4
SIOU	80.4	90.9	89.6
EIOU	72.9	87.8	84.6
α-DIOU	81.8	93.7	91.5

最终改进YOLOv5s模型的识别效果如图7所示。为验证模型在测试集中的效果，本文对数据集中的测试集使用原模型与改进模型进行测试。测试结果如表7所示，改进后的模型在精确率、召回率、平均精度均值达到81.2%、92.1%、88.9%，与YOLOv5s模型相比分别提升3.8%、3.4%、4.5%。



a：检测效果图1　b：检测效果图2　c：检测效果图3

图7 最终改进模型识别效果
Figure 7 Final improvement of model recognition performance

表7 模型在测试集中的表现结果（%）
Table 7 Performance of the model on the test set (%)

模型	精确率	召回率	平均精确度均值
YOLOv5s	77.4	88.7	84.4
本文方法	81.2	92.1	88.9

4 结 语

本文针对医疗器械堆叠和部分医疗器械相似，且为了进一步提升模型对医疗器械识别的性能，提出改进YOLOv5s医疗器械检测模型，可对重叠医疗器械进行实时检测。通过自建医疗器械数据集对改进后的模型进行评价。本文主要进行以下改进：(1)在YOLOv5s中引入C2f模块；(2)通过对比不同注意力机制在不同层的结果找到适合本文数据集的注意力机制；(3)在α-IOU的基础上引入DIOU作为YOLOv5s损失函数。在验证集中，改进后的模型相比于YOLOv5s模型在精确率、召回率、平均精度均值分别提升3.2%、3.4%、4.6%；测试集中，分别提升3.8%、3.4%、4.5%。相比于YOLOv5s模型，虽然本文使用的方法导致参数量增加，参数量增加导致推理速度有所降低，但是模型的精确率得到提升。本文主要在C2f模块实验中出现大量的参数，参数量的降低能够大大提高模型的推理速度，同时要考虑参数量降低的同时提升精确率、召回率等重要指标。在检测医疗器械的过程中，要对目标进行跟踪检测，使模型更加符合实际要求。

【参考文献】

[1] 米吾尔依提·海拉提, 热娜古丽·艾合麦提尼亚孜, 卡迪力亚·库尔班, 等. 基于改进YOLOv7的肝囊型包虫病超声图像小病灶检测[J]. 中国医学物理学杂志, 2024, 41(3): 299-308.
Hailati MW, Aihemaitiniyazi RN, Kuerban KD, et al. Small lesion detection in ultrasound images of hepatic cystic echinococcosis based on improved YOLOv7[J]. Chinese Journal of Medical Physics, 2024, 41(3): 299-308.

[2] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2014: 580-587.

[3] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2015: 1440-1448.

[4] Ren SQ, He KM, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. Cambridge, MA, USA: MIT Press, 2015: 91-99.

[5] He KM, Gkioxari G, Dollár P, et al. Mask R-CNN[J]. IEEE Trans Pattern Anal Mach Intell, 2020, 42(2): 386-397.

[6] He KM, Zhang XY, Ren SQ, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Trans Pattern Anal Mach Intell, 2015, 37(9): 1904-1916.

[7] Kim Y. Convolutional neural networks for sentence classification[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA, USA: ACL, 2014: 1746-1751.

[8] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 779-788.

[9] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6517-6525.

[10] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08). https://arxiv.org/abs/1804.02767.

[11] Wang CY, Bochkovskiy A, Liao HY. Scaled-YOLOv4: scaling cross stage partial network[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 13024-13033.

[12] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [C]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[13] Abdulbaki AT, Jalal NA, Möller K. A convolutional neural network with a two-stage LSTM model for tool presence detection in laparoscopic videos[J]. Curr Dir Biomed Eng, 2020, 6(1): 20200002.

[14] Hasan MK, Calvet L, Rabbani N, et al. Detection, segmentation, and 3D pose estimation of surgical tools using convolutional neural networks and algebraic geometry[J]. Med Image Anal, 2021, 70: 101994.

[15] Jin A, Yeung S, Jopling J, et al. Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2018: 691-699.

[16] Zhang BB, Wang SS, Dong LY, et al. Surgical tools detection based on modulated anchoring network in laparoscopic videos[J]. IEEE Access, 2020, 8: 23748-23758.

[17] 王志新, 王如刚, 王媛媛, 等. 基于混合注意力机制与C2f的行人检测算法研究[J]. 软件导刊, 2024, 23(1): 135-142.
Wang ZX, Wang RG, Wang YY, et al. Research on pedestrian detection algorithm based on mixed attention mechanism and C2f[J]. Software Guide, 2024, 23(1): 135-142.

[18] 谭义镇, 王飞, 李楠鑫. 基于改进YOLOv4的行人检测算法[C]//《人文与科技》第十辑会议论文集. 贵阳: 贵州民族大学人文科技学院, 2023: 261-273.
Tan YZ, Wang F, Li NX. Pedestrian detection algorithm based on improved YOLOv4 [C]//Humanities and Technology, Vol. 10. Guiyang: College of Humanities and Technology, Guizhou Minzu University, 2023: 261-273.

[19] Jiang BR, Luo RX, Mao JY, et al. Acquisition of localization confidence for accurate object detection[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 816-832.

[20] Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 13708-13717.

[21] Wang QL, Wu BG, Zhu PF, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 11531-11539.

[22] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.

[23] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Trans Pattern Anal Mach Intell, 2020, 42(8): 2011-2023.

[24] Rezatofighi H, Tsoi N, Gwak JY, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 658-666.

[25] Zheng ZH, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2020: 12993-13000.

（编辑：陈丽霞）