第43卷 第5期　　华 中 农 业 大 学 学 报　　Vol.43 No.5
2024年 9月　　Journal of Huazhong Agricultural University　　Sep. 2024,31～40

贺英豪,唐德钊,倪铭,等.基于改进YOLOv5对果园环境中李的识别[J].华中农业大学学报,2024,43(5):31-40.
DOI:10.13300/j.cnki.hnlkxb.2024.05.004

基于改进YOLOv5对果园环境中李的识别

贺英豪[1,2],唐德钊[1],倪铭[1,2],蔡起起[1]

1.四川农业大学信息工程学院,雅安 625014;2.雅安市数字农业工程技术研究中心,雅安 625014

摘要　为提高果园中高遮挡和密集的李(*Prunus salicina* Lindl.)的检测精度,提出一种改进YOLOv5s模型,在促进模型轻量化的同时提高模型对李的检测精度。首先,使用新的结构Focus-Maxpool模块替换主干网络中的下采样卷积,使改进模型在下采样时能够保留更多高遮挡目标和小目标的特征信息。其次,使用focal loss和交叉熵函数的加权损失作为改进模型的分类损失,提升改进模型对粘连目标的识别能力。最后,设计若干组检测试验来评价改进模型的性能。结果显示,改进YOLOv5s模型的平均精度优于YOLOv5s、YOLOv4、Faster-RCNN、SSD和Centernet;与YOLOv5s模型的检测结果相比,改进YOLOv5s模型的平均精度、召回率和精度分别提高2.84、9.53和1.66百分点,检测速度可达到91.37帧/s,能够满足实时检测需求。研究结果表明,改进的YOLOv5s模型在真实果园环境下具有较高的检测精度和鲁棒性。

关键词　李;机械采摘;果实识别;YOLOv5;图像处理;注意力机制;目标检测

中图分类号　TP391.4　　**文献标识码**　A　　**文章编号**　1000-2421(2024)05-0031-10

李(*Prunus salicina* Lindl.),蔷薇科李属木本植物[1],果实香甜脆爽,营养丰富,深受人们喜爱。目前,我国李的种植面积特别广,但大部分果园机械化采摘应用非常少,果实的采摘任务都是由人工完成,增加了种植成本。机械化采摘既提高生产效率,也降低种植成本,是目前果园收获的主要发展方向之一。采摘过程中的目标识别准确率是影响机械化采摘推广重要影响因素之一。

近年来,为了解决目标因遮挡带来的漏检的问题,许多学者进行了相关研究。李欣等[2]使用YCb-Cr颜色空间模型和OTSU阈值分割技术对有遮挡的柑橘果实进行识别,检出率达到了90.48%。周文静等[3]针对田间复杂环境下传统算法识别目标准确度低的问题,选取了与人类视觉相近的RGB颜色空间,实现了田间环境下红葡萄果穗图像的分割。杨晓珍[4]使用Gabor小波变换和支持向量机对西红柿进行智能分类,提高了果实人工分类的效率。虽然数字图像处理技术在对各种果实识别方面有着较为广泛的应用,但是数字图像处理技术对数据进行处理和分析的过程比较复杂,并且对较为复杂背景下目标的识别效果非常不稳定,实时性较差,难以满足果园采摘任务。因此,需要一种更快、更精确、更稳定的识别方法。

随着深度学习的应用与发展,利用卷积神经网络对水果果实检测获得了优异的效果[5-8]。卷积神经网络的高特征提取能力,使其对复杂背景下目标的检测成为可能。目标检测模型主要为两阶段检测模型和单阶段检测模型。两阶段检测模型中,如Fast-RCNN[9]和Faster-RCNN[10]都是先通过区域建议网络生成区域候选框,然后再对候选框进行进一步预测从而得到检测结果。单阶段检测模型以YOLO系列模型[11-13]和SSD系列模型[14]为代表。这些模型使用多尺度检测,融合主干网络中不同层的特征值,可以适应各种不同尺度目标的检测任务。另外还有一类anchor-free目标检测网络,如Centernet[15]和Cornernet[16],它们不依赖于先验框,而是通过预测关键点从而预测出目标的位置。要实现对果园中李的正确识别,就要克服果园环境存在的背景复杂、目标被各种遮挡物遮挡等问题。石展鲲等[17]提出一种改进的Faster-RCNN模型,通过使用Mosaic数据增强和引入Soft NMS算法来增强模型对小目标和密集目标的识别能力,结果表明,改进后的模

收稿日期:2023-04-25
基金项目:四川省自然科学基金项目(2022NSFSC0172);四川农业大学专业建设支持计划(040-2121997775)
贺英豪,E-mail:heyinghao@sicau.edu.cn
通信作者:倪铭,E-mail:nm@sicau.edu.cn

型的召回率可达91.44%，准确率可达93.35%。Wang等[18]对YOLOv4进行改进，使用深度可分离卷积网络将网络压缩至54.05 MB，并额外添加1个检测尺度来增强模型对小目标和遮挡目标的识别能力，改进后的模型平均精度均值（mPA）可以达到88.56%，检测速度可达42.55帧/s。杨福增等[19]基于Centernet提出了1种密集场景下多苹果目标的快速识别方法，通过设计Tiny Hourglass-24轻量级骨干网络，同时优化残差模块以提高目标识别速度，并且最终模型的平均精度可达98.90%。曾乾等[20]提出一种基于SSD的循环融合特征增强目标检测模型，使用K-means算法改进先验框的尺寸以及引入DIoU-NMS和ECA注意力机制，以此来解决复杂背景下青瓜目标识别率低、定位效果不佳等问题，改进后的SSD模型的平均精度相比于原模型提高了4.61%。但上述的改进方法存在一些不足，如在提高模型精度的同时增加了模型的计算量，使得模型的训练速度和预测速度变慢，或者在压缩模型的同时降低模型的精度，难以做到模型的规模和精度平衡。

另一方面，需要提高模型本身的性能，提高模型对果实的定位精度。郜统哲等[21]使用YOLOv3对同发育期的杨梅进行识别，该模型对雌花开花、幼果和成熟果实的识别准确率分别可达90.91%、90.79%和90.83%。黄彤镔等[22]提出1种基于YOLOv5s的改进模型，该模型通过引入CBAM注意力模块来提高网络的特征提取能力，从而降低遮挡目标和小目标的漏检率，结果表明该模型的平均精度可达91.3%，单张图片的检测时间为16.7 ms。赵文清等[23]针对视觉模型识别小目标精度低的问题，基于CornerNet提出隔级融合特征金字塔的方法来提高小目标平均准确率，最终其精度相比于原模型有着较大的提升。

上文所述模型中，YOLOv3的参数量较大，不适用于实时检测，改进YOLOv5s模型则引入了额外的参数量，而CornerNet不适用于遮挡目标的检测。

目前，针对李的目标检测研究较少，并且由于李具有自然环境复杂、枝叶遮挡等特点，使得现有模型很容易出现漏检、误检等问题，不能满足实时精确的检测。针对以上问题，本研究基于YOLOv5s提出了一个能在复杂的果园环境下识别李的目标检测方法，旨在辅助果农检测李果实以及为果园采摘机器人等智能产品提供算法依据。

1 材料与方法

1.1 数据集采集

数据集采集于四川省成都市的李产业园区，采集设备使用佳能800D单反相机和荣耀V9智能手机。采集到的原始图像共621张，其中使用单反相机采集到的图像包括6 000像素×4 000像素和3 984像素×2 656像素2种分辨率，分别有429和89张，使用手机相机采集到103张图像，分辨率为3 968像素×2 976像素。另外，为增加数据丰富性，从网络上爬取各种分辨率的图像79张，共700张图像。

1.2 数据集构建

为了增强模型的鲁棒性以及数据的丰富性，同时为了防止由于数据过少发生过拟合，本研究对数据集进行数据增强处理，包括随机亮度调节、随机对比度调节、随机饱和度调节、加入高斯噪声、水平翻转、运动模糊和旋转15°[24-27]。每张图片生成7张新图片，最终得到5 600张图片。使用Labeling工具标注每张图片中的李果实目标区域（图1），遮挡目标只标注其未被遮挡部分，然后按照8∶1∶1的比例划分为训练集（4 480张）、验证集（560张）、测试集（560张），每个数据集中都包含每种分辨率的图像。

图1 图片标注示例

Fig.1 Example of image annotation

1.3 检测模型

本研究以YOLOv5s为基础进行了改进。

1）Focus-Maxpool（FM）模块。使用常规3×3卷积对特征图进行下采样，不仅反向传播的参数大，而且会损失特征图信息。在果园的大视场条件下，部分李目标就显得非常小，使用常规卷积对其进行下采样时，会损失这些小目标和高遮挡目标的特征信息，从而造成漏检。在YOLOv5中，将Focus放在网络的输入位置，用于将输入的特征图进行2倍下采样。其原理如图2所示，Focus在下采样的过程中不会丢失任何特征信息。

虽然Focus模块在下采样过程中不会丢失信息，但下采样之后特征图通道数会变为原来的4倍，因此，在Focus模块之后一般会连接1个卷积用于降

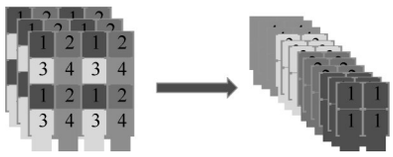


图2　Focus模块

Fig.2　Focus module

维。YOLOv5输入端的Focus模块之后连接的是1个3×3的卷积层，若是在模型中大量使用此模块，将会极大程度地增加模型的参数量和计算量。

为了弥补常规卷积在下采样时出现信息丢失的缺点，以及解决Focus模块参数量过大的问题，我们基于Focus模块提出了一种新的Focus-Maxpool结构（图3）。

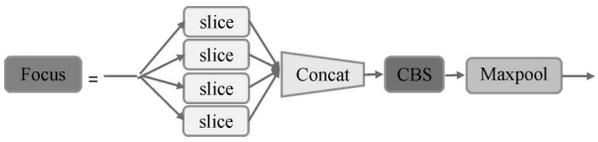


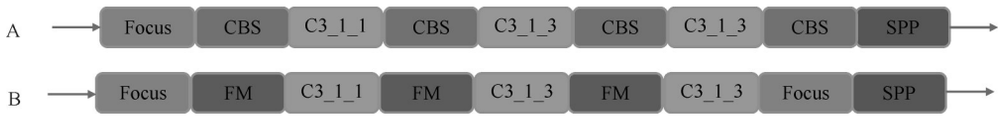
图3　Focus-Maxpool模块

Fig.3　Focus-Maxpool module

该模块由Focus模块、1×1的常规卷积和1个最大池化层组成，其中卷积层用于压缩特征图的通道数。YOLOv5模型中的下采样卷积的感受野是3×3，而Focus模块的感受野机制相当于一个全局感受野，它对目标具体特征的观察能力有限，因此在后面连接一个感受野为3×3的最大池化层用于优化Focus模块的感受野，该池化层不改变特征图大小。该模块在进行反向传播计算时，只会增加用于降维的卷积层的参数，而此卷积层的卷积核为1×1，因此，FM结构的参数量仅是使用3×3卷积下采样时的4/9。

2）改进YOLOv5s网络结构。图4是改进之后的YOLOv5s的主干网络结构，使用新的FM模块替换主干网络中的下采样卷积层。由于SPP模块为特征图提供了混合大小的感受野，因此，SPP模块前的下采样卷积层直接替换为Focus模块，该层的降维卷积层仍使用1×1的卷积核，以降低参数量。

图5是改进YOLOv5s的结构图。之所以只替换主干网络中的下采样卷积，是因为FM模块虽然参数量少，但在正向传播时的计算量较大，效率也比常规卷积稍慢，因此在网络中也不宜大量使用该模块。主干网络负责对输入的图像的主要特征的提取，并且主干网络的结构深度较深，更容易丢失信息。颈部网络中“自底向顶”的金字塔的网络结构较浅，使用常规卷积进行下采样也不会丢失太多信息。因此，只替换主干网络中的下采样卷积，以保证主干网络尽可能地减少信息丢失。



A：YOLOv5s主干网络 YOLOv5s backbone network；B：改进YOLOv5s主干网络 Improved YOLOv5s backbone network.

图4　YOLOv5s改进前后的主干网络结构

Fig.4　Backbone network structure before and after YOLOv5s improvements

3）损失函数。YOLOv5s的损失函数主要包含定位损失、分类损失和置信度损失。定位损失表示模型预测框的位置与真实框之间的差异，分类损失表示预测类别的概率与真实类别的概率的差异，置信度损失则是表示预测框位置有目标的概率。YOLOv5s的损失组成为：

$$l = a \times l_{\mathrm{loc}} + b \times l_{\mathrm{cls}} + c \times l_{\mathrm{conf}} \tag{1}$$

$$l_{\mathrm{loc}} = 1 - G_{\mathrm{iou}} \tag{2}$$

式（1）中，l表示总损失，l_{loc}表示位置损失，l_{cls}表示分类损失，l_{conf}表示置信度损失，a、b和c则是权重常数。l_{loc}的计算方式如式（2）所示，本研究为了降低模型计算的复杂度，使用GIOU损失（公式中用G_{iou}表示）。GIOU的计算公式如式（3）所示：

$$G_{\mathrm{iou}} = \frac{(A \cap B)}{(A \cup B)} - \frac{A^c - (A \cup B)}{A^c} \tag{3}$$

式（3）中，A和B分别代表2个不同框的面积，A^c表示A和B的最小外接矩形面积。

果园环境中粘连目标和高遮挡目标比较多，若是使用常规的交叉熵计算损失，模型会将权重集中在易分类的目标上，而容易忽略图像中不明显的遮挡目标和小目标。为了解决这种正、负样本不平衡的问题，引入焦点损失（focal loss）[28]替换原本的分类损失。focal loss的计算公式为：

$$F_{\mathrm{L}} = \begin{cases} -\alpha(1-p_{\mathrm{t}})^{\gamma}\ln(p_{\mathrm{t}}), \text{ if}, y=1 \\ -(1-\alpha)p_{\mathrm{t}}^{\gamma}\ln(1-p_{\mathrm{t}}), \text{ otherwise} \end{cases} \tag{4}$$

其中，F_{L}表示focal loss，y表示样本标签的数量。p_{t}表示目标类别的概率，α是平衡正、负样本权重的系数，α的取值范围一般是$0<\alpha<1$。γ为难分类样本的调制参数，并且$\gamma>0$。

p_{t}作为目标类别的概率，可直接反映样本分类的难易程度，p_{t}越大，样本越易分类，反之，p_{t}越小，样本

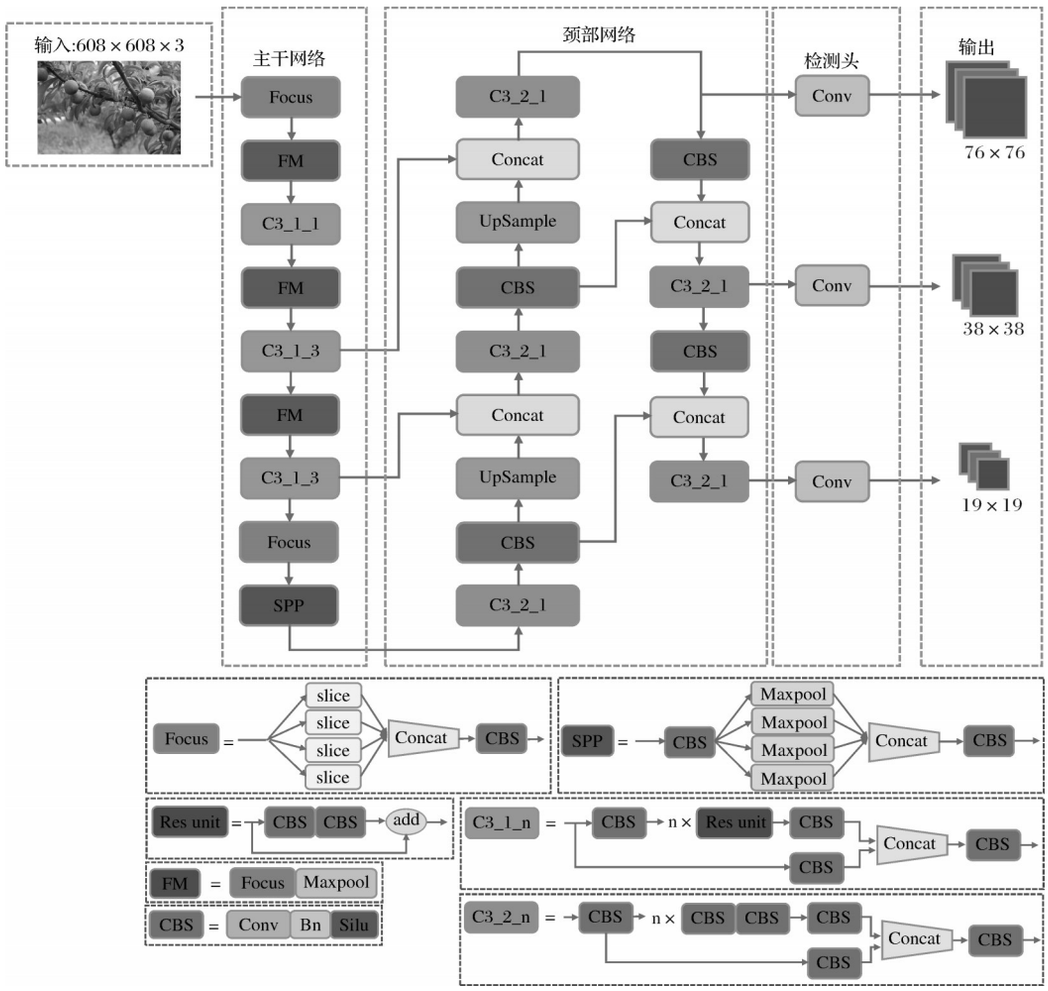


图5 改进YOLOv5s网络结构

Fig.5 Improved YOLOv5s network architecture

越难分类。而γ则是根据p_t的值来调整样本在总损失中所占的比例,γ越大,易分类样本的损失就越低,难分类样本的权重就增大了,因此,模型在训练时更容易将权重集中在难分类的样本上。

focal loss容易受噪声影响,当数据集中存在噪声或者错误标签的时候,focal loss会将权重集中在这些错误的样本上,并且会抑制较易分类样本的损失,从而提高噪音在总损失中的比例。因此,直接用focal loss替换原分类损失可能会影响训练效果。为了降低噪音对focal loss的干扰,使用focal loss和交叉熵损失的加权损失作为模型的分类损失,其计算公式为:

$$l_{cls}=j\times F_L+k\times C_E \tag{5}$$

$$C_E=t\ln a+(1-t)\ln(1-a) \tag{6}$$

式(5)~(6),j和k分别是focal loss和交叉熵损失的权重系数。C_E表示交叉熵损失,t是真实值,a是预测值。在本试验中,j和k的值皆取0.5。交叉熵函数可以让模型更容易识别目标是否为噪声,从而降低噪声对focal loss的干扰。

1.4 模型训练与测试

1)试验平台。本研究采用Windows10操作系统,基于Python 3.7.4和Pytorch 1.9.1搭建深度学习框架来完成网络模型的训练与测试。使用的计算机配置为Intel Core i9-10900K处理器,主频3.70 GHz,运行内存为128 GB,GPU使用NVIDIA Quadro RTX 5000,显存为16 GB,使用的CUDA版本是10.2,cudnn版本为7.6.5。

2)模型训练。本研究通过对所有图像中的目标的真实框进行K-means聚类得到先验框。本研究所使用的先验框为[8,10]、[12,17]、[18,23]、[22,35]、[36,30]、[31,48],[42,60]、[57,83]、[121,137]。模型训练时,将所有训练集图像的分辨率统一转换为608像素×608像素。优化器使用SGD优化器。为了更有效利用显存,训练时使用混合精

度训练模型，并且将批量(batch-size)设置为32。权值的衰减速率(weight-decay)设为0.000 5，动量(momentum)设为0.937。模型的总epoch数设为100，并自动保存模型识别精度最高的权重。训练时初始学习率为0.01，并使用余弦退火法来调整每个epoch的学习率。为了加快模型的收敛速度，先让模型在PASCAL VOC(http://host.robots.ox.ac.uk/pascal/VOC/index.html)数据集上训练并得到预训练模型，再将预训练模型用于本研究数据集的训练。

3)评价指标。采用推理速度作为模型检测速度的评判指标，推理速度代表每秒可识别的图像数量，帧/s。采用平均精度(average precision，AP；公式中用P_A表示)作为模型识别精度的评判指标。平均精度与模型的准确率(precision，P)和召回率有关(recall，R)有关：

$$P=\frac{T_P}{T_P+F_P}\times 100\% \tag{7}$$

$$R=\frac{T_P}{T_P+F_N}\times 100\% \tag{8}$$

$$P_A=\int_0^1 P(R)\mathrm{d}R\times 100\% \tag{9}$$

式(8)～(9)中，T_P为正确检测为李的数量，F_P表示被错误划分为李的数量，即误检数量，F_N则为图像中被漏检的目标的数量。

2　结果与分析

2.1　YOLOv5s模型改进前后的性能对比

1)整体效果对比。分别将YOLOv5s模型和改进YOLOv5s模型在同一测试集上进行测试(表1)。表1中小目标对应的是测试集中分辨率小于原图像分辨率0.5%的李果实目标，小目标在整个测试集中占比为65.63%。YOLOv5s模型在测试集上所有目标的平均精度、召回率和准确率分别为94.72%、84.56%和96.29%，其中小目标的平均精度、召回率和准确率分别85.22%、74.48%和91.78%。改进的YOLOv5s模型在测试集上所有目标的平均精度、召回率和准确率相比原模型分别提高了2.84、9.53和1.66百分点；对于小目标，则分别提高了6.79、14.33和3.75百分点。由此可以看出，改进YOLOv5s模型的主干网络由于保留了更多的细节信息，使得模型能够识别出更多的小目标，从而使得模型的性能在整体上有了提高。并且新的分类损失函数可以让模型更准确地预测出目标的位置，对模型的精度做出了一定的贡献。主干网络中添加的FM结构的参数量比常规3×3卷积少，因此，改进后的模型的大小比原模型降低3.3 MB。由于FM结构中需要对特征图进行切片、降维和池化操作，因此，FM结构的工作效率比常规卷积稍低，最终推理速度下降2.79帧/s，而这对模型的实时检测基本上没有影响。

图6是YOLOv5s模型改进前后对测试集中李的检测效果。由图6可知，改进后的模型对李的识别置信度普遍比原YOLOv5s模型高。相比图6B，图6A中YOLOv5s模型对高遮挡的李目标的检测均有漏检，而图6B中改进YOLOv5s模型可以将被树叶遮挡和被果实遮挡的目标全部检出。

YOLOv5s模型改进前后的损失图像如图7所示。由此也能看出改进后的模型在替换了下采样结构和分类损失后，训练效果要比原模型好，并且收敛过程稳定。

表1　YOLOv5s模型改进前后的结果对比

Table 1　Comparison of results before and after improvement of the YOLOv5s model

模型 Model	目标种类 Target type	P_A/%	R/%	P/%	推理速度/(帧/s) Inference speed	模型大小/MB Model size
YOLOv5s	所有目标 All target	94.72	84.56	96.29	94.16	27.1
	小目标 Small target	85.22	74.48	91.78		
改进 YOLOv5s Improved YOLOv5s	所有目标 All target	97.56	94.09	97.95	91.37	23.8
	小目标 Small target	92.01	88.81	95.53		

2)不同密度情况下的检测结果。为了进一步验证改进YOLOv5s模型对不同密度情况下李的检测都有着较高精度，本研究对测试集中的图片根据目标密度进行分类。其中，单位距离间像素大于40的图片被归类于高密度。由于数据集中有些图片是大视场下的场景，这些图片中的目标数较多，但目标成簇分布，每个簇中目标较为密集，但簇之间的距离较远，直接计算密度会导致密度的值偏低，因此，本研究直接将图片中目标个数大于20的也归类于高密度。而单位距离间像素小于等于20且大于等于10的图片被归类于中密度，单位距离间像素小于10的图片被归类于低密度。最终高密度的图片有205张，

A:YOLOv5s;B:改进YOLOv5s Improved YOLOv5s.

图6 YOLOv5s改进前后的检测结果

Fig.6 Detection results before and after the improvement of YOLOv5s

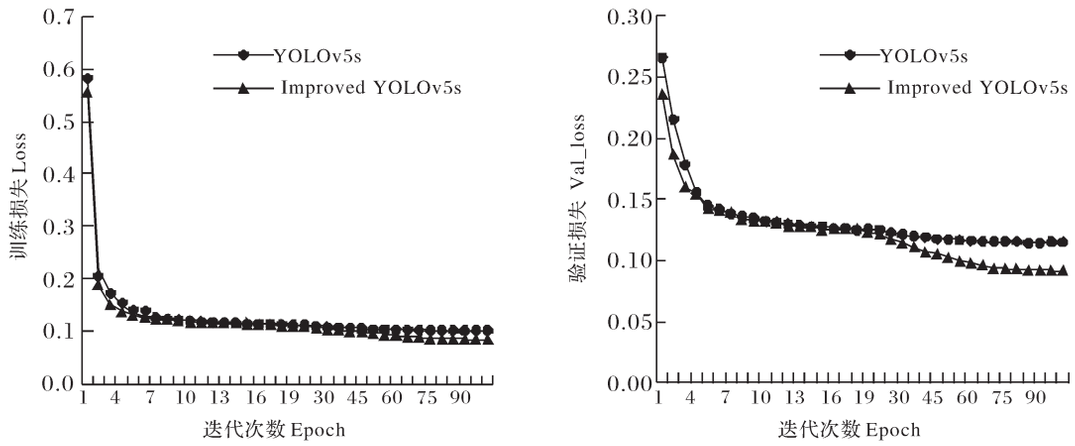


图7 YOLOv5s改进前后的损失曲线

Fig.7 Loss curve before and after YOLOv5s improvements

中密度的图片有188张,低密度的图片有167张。从表2可知,改进后模型对不同密度下的检测结果皆优于原模型,改进后模型的性能优势主要表现在召回率上,在高密度、中密度和低密度情况下,改进后模型的召回率比原模型分别提高10.46、8.36和7.57百分点。

对比图8的检测结果可以看到,改进后的模型计算出的目标置信度普遍要高于原模型,并且原模型漏检一些遮挡度较高的目标,而改进模型则识别出了所有目标。改进后的模型可以在果实比较密集且树叶比较密集的情况下有效识别出目标,且有较高的置信度,说明改进后的模型有着较强的对遮挡目标的识别能力。

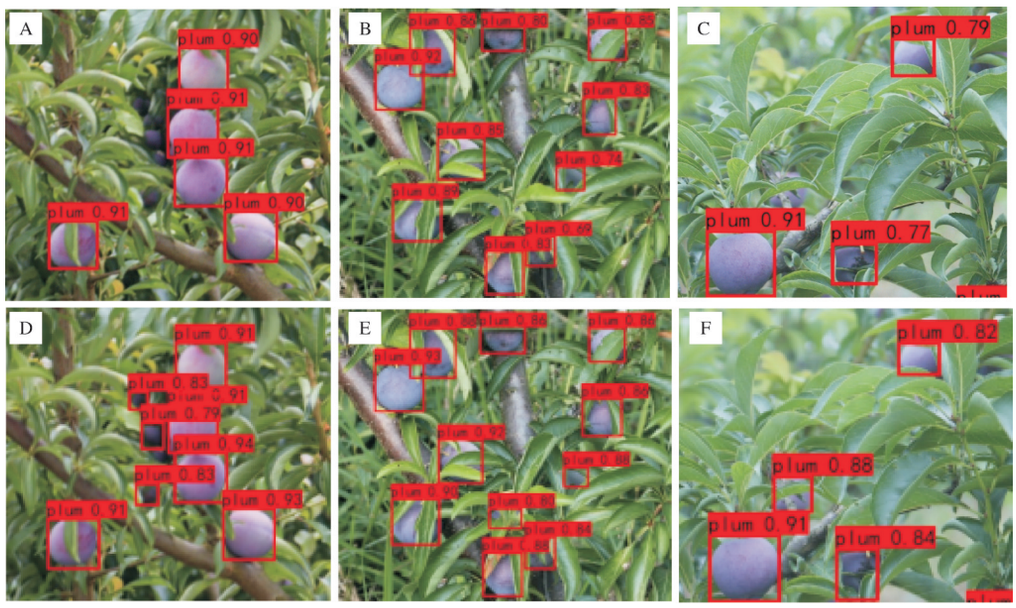
2.2 消融试验

表3中试验1是将YOLOv5s模型的主干网络中的下采样卷积替换为FM结构的测试结果,对比表1中原模型的测试结果可知,试验1的平均精度、召回率和准确率分别提升了2.71、9.34和1.53百分点。使用FM结构作下采样可以保留更多的细节信息,避免特征图在多次下采样后丢失小目标和遮挡目标的特征信息,从而降低了整体的漏检率。试验2是在改进模型的基础上将原分类损失替换为focal loss,从表3可知,试验2的召回率相对试验1降低3.02百分点,这是因为focal loss容易受噪音影响,而果园中环境

表2 不同密度情况下的检测结果

Table 2 Test results at different densities %

模型 Model	类型 Type	P_A	R	P
YOLOv5s	高密度 High density	92.90	81.25	95.47
	中密度 Medium density	96.74	88.98	97.17
	低密度 Low density	98.32	91.09	98.08
改进YOLOv5s Improved YOLOv5s	高密度 High density	96.46	91.71	97.40
	中密度 Medium density	98.77	97.34	98.65
	低密度 Low density	99.68	98.66	98.96



A-C：YOLOv5s；D-F：改进 YOLOv5s Improved YOLOv5s；A，D：高密度 High density；B，E：中密度 Medium density；C，F：低密度 Low density.

图 8　YOLOv5s模型改进前后对不同密度李的检测结果

Fig.8　Detection results of different densities of plums before and after improvement of YOLOv5s model

背景较为复杂，可能会出现较多的干扰。因此，试验3在试验2的基础上，将 focal loss 与交叉熵损失函数相结合，取二者的加权值作为模型的分类损失，而最终的平均精度、召回率和准确率相比试验1分别提升0.13、0.19和0.13百分点，加权损失函数可以使模型的性能在整体上有一个小幅度的提升。

表 3　消融试验结果

Table 3　Result of ablation test　　%

NO.	模型 Model	P_A	R	P
1	YOLOv5s＋FM	97.43	93.90	97.82
2	YOLOv5s＋FM＋focal loss	97.40	90.88	98.81
3	YOLOv5s＋FM＋加权分类损失 YOLOv5s＋FM＋ Weighted classification loss	97.56	94.09	97.95

2.3　改进 YOLOv5s 与其他模型的性能对比试验

为了进一步证明改进 YOLOv5s 模型的有效性，本研究将同样的数据集分别在 Faster-RCNN、YOLOv4、SSD 和 Centernet 上进行了训练。表4所示是各个模型的性能评估结果。对比表1中的原 YOLOv5s 模型的性能可以看出，虽然 YOLOv5s 的召回率不如 Faster-RCNN，准确率不如 SSD 和 Centernet，但其平均精度高于其他4个模型，且模型大小比其他模型小，推理速度也最快，说明其综合性能最好。这是因为 YOLOv5s 有着较优秀的结构设计和训练策略，使其在保持较好性能的同时，也能有较高的推理速度，这也表明了本研究选取 YOLOv5s 模型的合理性。

对于改进 YOLOv5s 模型，其平均精度比 Faster-RCNN、YOLOv4、SSD 和 Centernet 模型分别高13.69、8.77、6.77 和 6.02 百分点。其中，Faster-RCNN 在处理密集目标时会出现重叠区域较多的情况，这会使得检测结果不准确，从而导致准确率低。对于 YOLOv4，其颈部网络中采用的都是普通的卷积操作，对来自不同层的特征信息的融合能力有限，而 YOLOv5s 的颈部网络中，采用的 CSP 结构，拥有更强的网络特征融合的能力，能够提取更多浅层的关于小目标的特征信息。SSD 和 Centernet 虽然有着较高的准确率和平均精度，但它们的召回率较低，对小目标和遮挡目标的识别能力有限。Centernet 则是因为需要预测中心点来预测目标位置，在处理遮挡目标和小目标时会丢失细节信息，从而导致召回率较低。综上所述，改进 YOLOv5s 模型的综合性能最好，其模型大小仅为 SSD 的1/4，推理速度也比 Centernet 高35.84帧/s，可以快速有效地识别出李。

图9是上述5个模型对测试集中不同尺寸的目标的识别效果。从图9A，B可以看出，改进 YOLOv5s 模型和 Faster-RCNN 能够准确识别图片中的每个目标，但 Faster-RCNN 的模型较大，且速度较慢。如图9C，D所示，YOLOv4 模型和 SSD 模型都将紫色瓶子误检为了李。图9E图中 Centernet 模型

表4 改进YOLOv5s与其他模型的性能对比试验结果

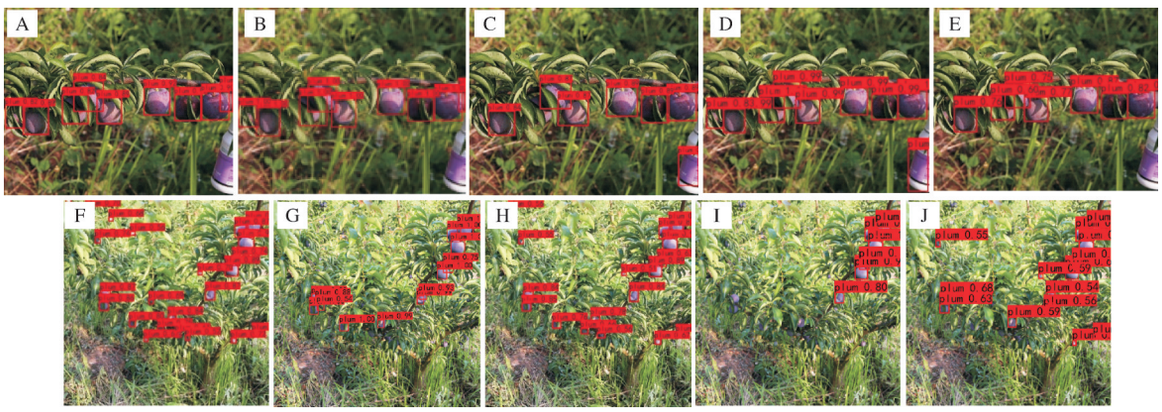
Table 4 Experimental results comparing performance of improved YOLOv5s with other models

模型 Model	P_A/%	R/%	P/%	模型大小/MB Model size	推理速度/(帧/s) Inference speed
Faster-RCNN[10]	83.87	85.64	74.70	521	17.07
YOLOv4[13]	88.79	78.44	89.88	244	32.02
SSD[14]	90.82	72.99	95.26	90.6	36.06
Centernet[15]	91.57	72.20	98.20	124	55.53
改进YOLOv5s Improved YOLOv5s	97.56	94.09	97.95	23.8	91.37

漏检了图像中被遮挡的目标。由此可见,改进后模型对遮挡目标有较高的检出率和识别精度。

对比图9F～J,可以看到,相比于其他模型,改进YOLOv5s模型可以检测出更多的目标。Faster-RCNN、SSD和Centernet模型漏检了许多尺寸较小,遮挡度较高的目标,而YOLOv4可以检测出较多的目标,但还是漏检了一些被枝叶遮挡的目标。在大视场情况下,果实目标就会显得较小,一些被遮挡的目标则变得更加不明显,而改进YOLOv5s中的FM模块可以让模型在训练时更多地保留小目标信息和被遮挡目标的信息,从而提升模型对这些较难识别的目标的检测能力。

综上所述,改进YOLOv5s模型不仅能够对较明显的目标有着高精度和高检出率,并且也能精确识别出尺寸较小和遮挡度较高的目标的位置。本研究提出的改进YOLOv5s模型在能保证比较高的精度的情况下,又不会使模型在推理过程中引入太多的参数而影响推理速度,改进后YOLOv5s模型规模较原模型有所降低,所需的算力资源较少,适合部署于嵌入式设备。



A-E:大目标Bigger target;F-J:小目标Small target;A,F:Improved YOLOv5s;B,G:Faster-RCNN;C,H:YOLOv4;D,I:SSD;E,J:Centernet.

图9 不同模型对不同大小目标的检测结果

Fig.9 Detection results of different models for targets of different sizes

3 讨 论

本研究基于YOLOv5s模型提出一种果园自然环境下李果实目标检测模型,制作了自然场景下的李果实数据集用于模型的训练和测试。试验结果表明,改进YOLOv5s模型识别李果实的平均精度提高到97.56%,识别小目标的平均精度达到92.01%,其模型大小只有23.8 MB,在GPU上的推理速度为91.37帧/s,可以满足于嵌入式设备的部署条件。

本研究使用Focus-Maxpool模块替换主干网络中的下采样卷积,使模型在下采样时能够保留更多高遮挡目标和小目标的特征信息,使用focal loss和交叉熵函数的加权损失作为模型的分类损失,提升模型对密集目标的识别能力。本研究还设计不同的试验来对比模型改进前后的性能,并且将改进后的模型与YOLOv4、Faster-RCNN、SSD和Centernet的作对比,说明改进后模型的优越性。在以后的研究中,可将该模型进一步改良,丰富数据集,实现模型对李子成熟度和病虫害的识别。

参考文献 References

[1] 李东,李静雅,雷雨,等.李子的品质特性及其精深加工产品研究进展[J].北方园艺,2022(19):128-135.LI D,LI J Y,LEI Y,et al.Research progress on quality characteristics and intensive processing products of plum (*Prunus salicina* Lindl.)[J].Northern horticulture,2022(19):128-135(in Chinese with English abstract).

[2] 李欣，王玉德．基于颜色模型和阈值分割的有遮挡的柑橘果实识别算法[J]．计算技术与自动化，2022，41(2)：136-140. LI X，WANG Y D.Occluded *Citrus* fruit recognition algorithm based on color model and threshold segmentation[J].Computing technology and automation，2022，41(2)：136-140(in Chinese with English abstract).

[3] 周文静，马晓晓，张洚宇，等．基于颜色因子逻辑与运算的田间红葡萄果穗识别[J]．南方农机，2022，53(7)：16-18.ZHOU W J，MA X X，ZHANG J Y，et al.Ear recognition of field red grape based on color factor logic and operation[J].China southern agricultural machinery，2022，53(7)：16-18(in Chinese).

[4] 杨晓珍．基于数字图像技术的西红柿智能分类[J]．单片机与嵌入式系统应用，2022，22(1)：79-83.YANG X Z.Intelligent classification of tomatoes based on digital image technology [J].Microcontrollers & embedded systems，2022，22(1)：79-83(in Chinese with English abstract).

[5] 吕佳，李帅军，曾梦瑶，等．基于半监督 SPM-YOLOv5 的套袋柑橘检测算法[J]．农业工程学报，2022，38(18)：204-211. LYU J，LI S J，ZENG M Y，et al.Detecting bagged citrus using a semi-supervised SPM-YOLOv5[J].Transactions of the CSAE，2022，38(18)：204-211(in Chinese with English abstract).

[6] 姚松鹏，朱隆睿，乔波．基于目标检测的苹果探测与定位研究[J]．电脑与信息技术，2022，30(5)：22-24.YAO S P，ZHU L R，QIAO B.Research on fruit detection and location based on target detection [J]. Computer and information technology，2022，30(5)：22-24(in Chinese with English abstract).

[7] 周焜．基于深度学习的杨梅成熟度检测仪的研发[D]．杭州：浙江科技学院，2022.ZHOU K.Development of myrica rubra maturity detector based on deep learning[D].Hangzhou：Zhejiang University of Science & Technology，2022 (in Chinese with English abstract).

[8] 李竹，牟昌红，嵇康轩，等．基于深度学习的蓝莓成熟度预测[J]．安徽农业科学，2023，51(5)：232-236.LI Z，MOU C H，JI K X，et al. Prediction of blueberry maturity based on deep learning [J]. Journal of Anhui agricultural sciences，2023，51(5)：232-236(in Chinese with English abstract).

[9] GIRSHICK R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV). December 7-13，2015.Santiago，Chile.Santiago：IEEE，2015：1440-1448.

[10] REN S Q，HE K M，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.

[11] 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).June 27-30，2016.Las Vegas，NV，USA.Las Vegas：IEEE，2016：779-788.

[12] REDMON J，FARHADI A. YOLOv3：an incremental improvement[DB/OL].arXiv，2018：1804.02767[2023-04-25]. https://doi.org/10.48550/arXiv.1804.02767.

[13] BOCHKOVSKIY A，WANG C Y，LIAO H Y M.Yolov4：Optimal speed and accuracy of object detection[DB/OL].arXiv，2020：2004.10934 [2023-04-25]. https://doi. org/10.48550/arXiv.2004.10934.

[14] LIU W，ANGUELOV D，ERHAN D，et al.SSD：single shot MultiBox detector [M]//Computer Vision - ECCV 2016. Cham：Springer International Publishing，2016：21-37.

[15] DUAN K W，BAI S，XIE L X，et al.CenterNet：keypoint triplets for object detection [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV).October 27-November 2，2019.Seoul，Korea (South).Seoul ：IEEE，2019：6569-6578.

[16] LAW H，DENG J.CornerNet：detecting objects as paired keypoints [J]. International journal of computer vision，2020，128(3)：642-656.

[17] 石展鲲，杨凤，韩建宁，等．基于 Faster-RCNN 的自然环境下苹果识别[J]．计算机与现代化，2023(2)：62-65.SHI Z K，YANG F，HAN J N，et al.Apples recognition in natural environment based on faster-RCNN[J].Computer and modernization，2023(2)：62-65(in Chinese with English abstract).

[18] WANG L L，ZHAO Y J，LIU S B，et al.Precision detection of dense plums in orchards using the improved YOLOv4 model [J/OL].Frontiers in plant science，2022，13：839269[2023-04-25].https://doi.org/10.3389/fpls.2022.839269.

[19] 杨福增，雷小燕，刘志杰，等．基于 CenterNet 的密集场景下多苹果目标快速识别方法[J]．农业机械学报，2022，53(2)：265-273.YANG F Z，LEI X Y，LIU Z J，et al.Fast recognition method for multiple apple targets in dense scenes based on CenterNet[J].Transactions of the CSAM，2022，53(2)：265-273(in Chinese with English abstract).

[20] 曾乾，李博．基于改进 SSD 的青瓜检测算法[J]．国外电子测量技术，2023，42(4)：158-165.ZENG Q，LI B.Cucumber detection algorithm based on improved SSD[J].Foreign electronic measurement technology，2023，42(4)：158-165(in Chinese with English abstract).

[21] 郜统哲，王亮，张晓，等．基于 YOLOv3 的杨梅关键发育期识别技术[J]．现代农业科技，2023(5)：79-82.GAO T Z，WANG L，ZHANG X，et al. Identification technology of key developmental stages of *Myrica rubra* based on YOLOv3[J]. Modern agricultural science and technology，2023(5)：79-82 (in Chinese).

[22] 黄彤镔，黄河清，李震，等．基于 YOLOv5 改进模型的柑橘果实识别方法[J]．华中农业大学学报，2022，41(4)：170-177. HUANG T B，HUANG H Q，LI Z，et al.*Citrus* fruit recognition method based on the improved model of YOLOv5[J]. Journal of Huazhong Agricultural University，2022，41(4)：170-177(in Chinese with English abstract).

[23] 赵文清，孔子旭，赵振兵．隔级融合特征金字塔与 CornerNet 相结合的小目标检测[J]．智能系统学报，2021，16(1)：108-

116.ZHAO W Q,KONG Z X,ZHAO Z B.Small target detection based on a combination of feature pyramid and CornerNet [J].CAAI transactions on intelligent systems,2021,16(1):108-116(in Chinese with English abstract).

[24] 李兆旭,蒋红海,杨肖,等.基于轻量化深度学习模型的豆角苗-杂草检测方法[J].农业装备与车辆工程,2022,60(9):98-102.LI Z X,JIANG H H,YANG X,et al.Bean seedling-weed detection method based on lightweight deep learning model[J].Agricultural equipment & vehicle engineering,2022,60(9):98-102(in Chinese with English abstract).

[25] 赵越,赵辉,姜永成,等.基于深度学习的马铃薯叶片病害检测方法[J].中国农机化学报,2022,43(10):183-189.ZHAO Y,ZHAO H,JIANG Y C,et al.Detection method of potato leaf diseases based on deep learning[J].Journal of Chinese agricultural mechanization,2022,43(10):183-189(in Chinese with English abstract).

[26] 祁金文.基于YOLOv5的苹果目标识别方法研究[J].电脑编程技巧与维护,2022(8):137-139.QI J W.Research on apple target recognition method based on YOLOv5[J].Computer programming skills & maintenance,2022(8):137-139(in Chinese).

[27] 邵良玉.基于迁移学习的花类图像分类方法研究[J].农业装备与车辆工程,2022,60(7):62-64.SHAO L Y.Flower image classification based on transfer learning[J].Agricultural equipment & vehicle engineering,2022,60(7):62-64(in Chinese with English abstract).

[28] LIN T Y,GOYAL P,GIRSHICK R,et al.Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV).October 22-29,2017.Venice.Venice:IEEE,2017:2980-2988.

Recognizing plums in orchard environment based on improved YOLOv5

HE Yinghao[1,2],TANG Dezhao[1],NI Ming[1,2],CAI Qiqi[1]

1.*College of Information Engineering Sichuan Agricultural University*, *Ya'an 625014*, *China*;
2.*Ya'an Digital Agricultural Engineering Technology Research Center*, *Ya'an 625014*, *China*

Abstract This article proposed an improved YOLOv5s model to improve the accuracy of detecting plums (*Prunus salicina* Lindl.) with high occlusion and density in orchards and the lightweight. Firstly, a new Focus-Maxpool module was used to replace the down-sampling convolution in the backbone network, enabling the model to retain more feature information of small and highly occluded targets during down-sampling. Secondly, the weighted loss of focal loss and cross-entropy function was used as the classification loss of the model to improve its recognition ability for adhesive targets. Finally, several sets of detection experiments were designed to evaluate the performance of the model. The results showed that the average accuracy of the improved YOLOv5s model was better than that of YOLOv5s, YOLOv4, Faster RCNN, SSD, and Centernet. Compared with the detection results of the YOLOv5s model, the average accuracy, recall rate, and accuracy of the improved model increased by 2.84, 9.53, and 1.66 percentages, respectively. The detection speed of the improved model reached 91.37 frames per second, meeting the requirements of real-time detection. It is indicated that the model improved has higher accuracy of detection and robustness in real orchard environments. It will provide data reference for studying fruit-picking robots and monitoring orchard environments.

Keywords *Prunus salicina* Lindl.; mechanical harvesting; fruit recognition; YOLOv5; image processing; attention mechanism; object detection

(责任编辑:陆文昌)