基于改进型 YOLOv4 的果园障碍物实时检测方法

蔡舒平，孙仲鸣，刘 慧※，吴翊轩，庄珍珍

（江苏大学电气信息工程学院，镇江 212013）

摘 要：针对农业机器人在复杂的果园环境中作业时需要精确快速识别障碍物的问题，该研究提出了一种改进型的 YOLOv4 目标检测模型对果园障碍物进行分类和识别。为了减少改进后模型的参数数量并提升检测速度，该研究使用了深度可分离卷积代替模型中原有的标准卷积，并将主干网络 CSP-Darknet 中的残差组件（Residual Unit）改进为逆残差组件（Inverted Residual Unit）。此外，为了进一步增强模型对目标密集区域的检测能力，使用了软性非极大值抑制（Soft DIoU-Non-Maximum Suppression，Soft-DIoU-NMS）算法。为了验证该研究所提方法的有效性，选取果园中常见的 3 种障碍物作为检测对象制作图像数据集，在 Tensorflow 深度学习框架上训练模型。然后将测试图片输入训练好的模型中检测不同距离下的目标障碍物，并在同一评价指标下，将该模型的测试结果与改进前 YOLOv4 模型的测试结果进行评价对比。试验结果表明，改进后的 YOLOv4 果园障碍物检测模型的平均准确率和召回率分别为 96.92%和 91.43%，视频流检测速度为 58.5 帧/s，相比于原模型，改进后的模型在不损失精度的情况下，将模型大小压缩了 75%，检测速度提高了 29.4%。且改进后的模型具有鲁棒性强、实时性更好、轻量化的优点，能够更好地实现果园环境下障碍物的检测，为果园智能机器人的避障提供了有力的保障。

关键词：农业；机器人；目标检测；深度学习；深度可分离卷积；逆残差组件；Soft-DIoU-NMS

doi：10.11975/j.issn.1002-6819.2021.2.005

中图分类号：TP391;S24 **文献标志码**：A **文章编号**：1002-6819(2021)-2-0036-08

蔡舒平，孙仲鸣，刘慧，等. 基于改进型 YOLOv4 的果园障碍物实时检测方法[J]. 农业工程学报，2021，37(2)：36-43.
doi：10.11975/j.issn.1002-6819.2021.2.005 http://www.tcsae.org

Cai Shuping, Sun Zhongming, Liu Hui, et al. Real-time detection methodology for obstacles in orchards using improved YOLOv4[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(2): 36-43. (in Chinese with English abstract) doi：10.11975/j.issn.1002-6819.2021.2.005 http://www.tcsae.org

0 引 言

随着中国果园种植面积的不断扩大，果园农药喷洒和修剪采摘作业量日趋繁重，仅仅依靠人力和简单的工具已经无法满足需求。近几年大力推动的“智慧化农业”中将机器人技术应用于农业生产，这为农业的现代化升级提供了新的途径[1]。然而农业机器人在提高作业效率，减轻劳动强度的同时[2]，也存在着不小的安全隐患。当机器人在果园内作业时，可能会撞到各种障碍物从而引发事故。所以农业机器人还必须具有一定的障碍物检测和识别的能力。

实际中的果园场景环境复杂，存在着各种障碍物，障碍物的存在会对机器人的作业造成影响。在观察多个果园的环境后，发现无人驾驶的农业机器人会对果园内的果农、机器操作人员以及其他行人的安全造成威胁，也会碰撞到电线杆和灯柱导致电气事故的发生，而树木本身不仅是农业机器人的作业对象，同时也是行驶途中的障碍物。因此本研究中选取了行人、果树、电线杆这 3 种典型的障碍物作为检测识别的对象。目前，国内外对于农业机器人避障检测的研究中，主要采用激光传感器检测、雷达传感器检测、机器视觉传感器检测等[3]。在室外的场景下，机器视觉传感器具有独特的优势，它比激光雷达传感器更廉价，并且具有采样周期短、实时性强、抗干扰性强、易于部署等优点[4]。

果园障碍物检测在机器视觉领域属于目标检测这一类任务，在传统目标检测中需要人工设计算子来提取出图像中目标物体的颜色、纹理和边缘等特征[5]然后进行定位和分类。如 HOG（Histogram of Oriented Gradient）、SIFT（Scale-invariant feature transform）、SURF（Speeded Up Robust Features）和 Canny 等[6-9]算法。但较低的准确率、复杂庞大的数据量、不同物体及不同环境下特征的设计难度大、实时性差等缺点使传统目标检测方法不再适用于农业障碍物检测识别。随着近几年深度学习（Deep Learning，DL）理论的迅速发展，计算机硬件和图像采集设备的性能不断提升，基于深度卷积神经网络的目标检测算法已被广泛使用[10]。相比于传统目标检测算法，深度学习目标检测算法的参数权重都是通过输入大量的数据，经过反复的训练迭代学习得来的，检测结果更加精确，具有很强的自适应性和鲁棒性[11]。在典型的深度学习目标检测算法中，一类是基于区域推荐（Region Proposal）的目标检测，代表性的算法有：R-CNN[12]、Fast

收稿日期：2020-12-10 修订日期：2021-01-05

基金项目：江苏省重点研发计划（BE2018372）；江苏省自然科学基金（BK20181443）；江苏高校青蓝工程资助；镇江市重点研发计划（NY2018001）

作者简介：蔡舒平，副教授，主要研究方向为人工智能算法在农业、电力方面的研究及应用。Email：spcai@ujs.edu.cn

※通信作者：刘慧，博士，教授，主要研究方向为农业电气化与自动化、智能控制与信号处理。Email：amity@ujs.edu.cn

R-CNN[13]、Faster-RCNN[14]、SPP-NET[15]等，另一类是基于回归的目标检测，利用端到端（End to End）的思想，将图像归一化到统一大小后直接放入一个卷积神经网络（Convolutional Neural Networks，CNN）中回归预测出目标物体的类别和位置信息。代表性的算法有：YOLO（You Only Look Once）[16]系列、SSD（Single Shot MultiBox Detector）[17]系列等。虽然基于区域推荐的目标检测算法在准确度上占有一定的优势，但候选区域的提取过程存在计算量大、过程复杂度高、检测速度较慢的缺点，使得这种算法无法满足农业机器人对实时性目标检测的需求。而 YOLO 系列的算法有着高准确率和高检测速度的优点。特征提取也更着眼于整体，因此训练后分类和识别的效果优秀，能够满足复杂的果园环境中农业机器人实时障碍检测的要求。

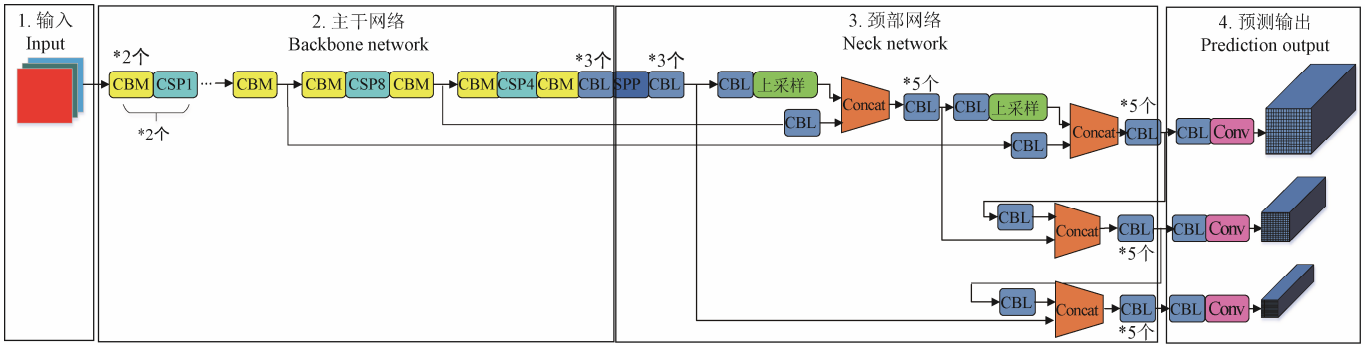
YOLOv1 目标检测网络于 2016 年由 Redmon 等[18]推出，随后它的 v2 和 v3 版本做出了不少改进。其中 YOLOv2 不仅将主干卷积的层数扩大到了 19 层，还借鉴了 Faster-RCNN 的锚框（anchor）方法来适应大小和长宽不同的检测目标，并将末尾的全连接结构替换成了 1×1 的卷积结构，使边框定位信息更准确。而 YOLOv3 借鉴了 He 等[19]提出的 ResNet 中的残差结构，有效解决了神经网络退化的问题，成功地将主干网络的卷积层数增加到了 53 层，同时使用多个尺度的检测头，在检测速度与检测效果上均达到了一个高峰。基于 YOLOv3 的改进与应用已经取得了不少研究成果，蔡逢煌等[20]加入了注意力机制，使卷积网络能更加专注于提取有用信息。刘洋等[21]在训练时加入 MSRCR 图像增强方法，提高了 YOLOv3 在雨雾天气下的检测精度。张健[22]运用了可变形卷积，使得 YOLOv3 的卷积网络在特征采样位置时能够随目标物的形状和大小自适应地改变。

Bochkovskiy 等[23]在 YOLOv3 的基础上提出了 YOLOv4 网络，准确率和检测速度再次得到提升。但是在应用于农业机器人障碍物检测时，存在以下缺点：对遮挡目标的检测不够理想、模型参数量过大、难以部署于低算力的嵌入式设备等。本研究在原有 YOLOv4 网络结构的基础上进行改进，并通过试验测试其运用在果园障碍物检测的上效果。

1　果园障碍物识别模型

1.1　YOLOv4 目标检测算法

YOLOv4 的网络结构简图如图 1 所示。主干网络 CSPDarknet 由 CSPX 模块和 CBM 模块交替叠加构成，CSPX 的结构如图 2a 所示，它的作用是将特征图一分为二。第一部分经过 CBM 和 X 个残差组件（Res unit）的卷积操作，第二部分直接与第一部分 Concat 结合。图 2b 中 CBM 模块由一个卷积层连接一个批量正则化（Batch Norm，BN）再连接一个 Mish 激活函数组成，而 CBL 模块与 CBM 的区别在于激活函数使用的是 Leaky Relu。图 2c 中的 Res unit 一部分经过 2 个 CBM 模块，与没有经过任何操作的另一部分进行 Add 连接操作。CSPDarknet 在特征图中集成了梯度的变化，有效地强化了神经网络的学习能力，且在减少计算量的同时保持了较高的准确度。空间金字塔池化（Spatial Pyramid Pooling，SPP）模块位于主干网络和颈部网络的结合处，它的作用如图 2d 所示，将输入特征图分别通过最大池化的方式变为不同尺度的特征图，然后将不同尺度的特征图与原特征图进行 Concat 操作结合起来输出。采用这种方式，相比于普通的最大池化操作，能够更好地增加卷积核的感受野。此外，YOLOv4 的颈部网络除了有特征金字塔网络（Feature Pyramid Networks，FPN）层外，还添加了路径增强网络（Path Aggregation Network，PAN）模块。如图 2e 所示，FPN 将顶层的特征图通过上采样的操作依次与下层的特征图连接起来，融合了丰富多样的特征信息。但是高层级与低层级之间的卷积层跨度大，需要耗费大量的时间，而 PAN 则解决了这个问题，它通过下采样连接底层特征与高层特征，缩短了各层特征之间融合的路径。YOLOv4 的输出预测部分除了损失函数和非极大值抑制（Non-Maximum Suppression，NMS）外都与 YOLOv3 保持一致。



注：Conv 表示卷积（Convolution），CBM 表示 Conv 加批量正则（Batch Norm，BN）加 Mish 激活函数的合成模块，CSP 表示跨阶段部分（Cross stage partial）的结构，CBL 表示 Conv 加 BN 加 Leaky relu 激活函数的合成模块，Concat 表示一种通道数相加的特征合成方式，图中的*i 个表示此处应有 i 个同样的模块构成。
Note: Conv represents convolution. CBM means the synthesis module includes convolution, batch norm and Mish activation function. CSP represents the structure of the Cross stage partial. CBL synthesis module includes convolution, Batch norm and Leaky relu activation function. Concat represents a feature fusion of channel numbers. The *i representation in the diagram should have i multiple of same module composition here.

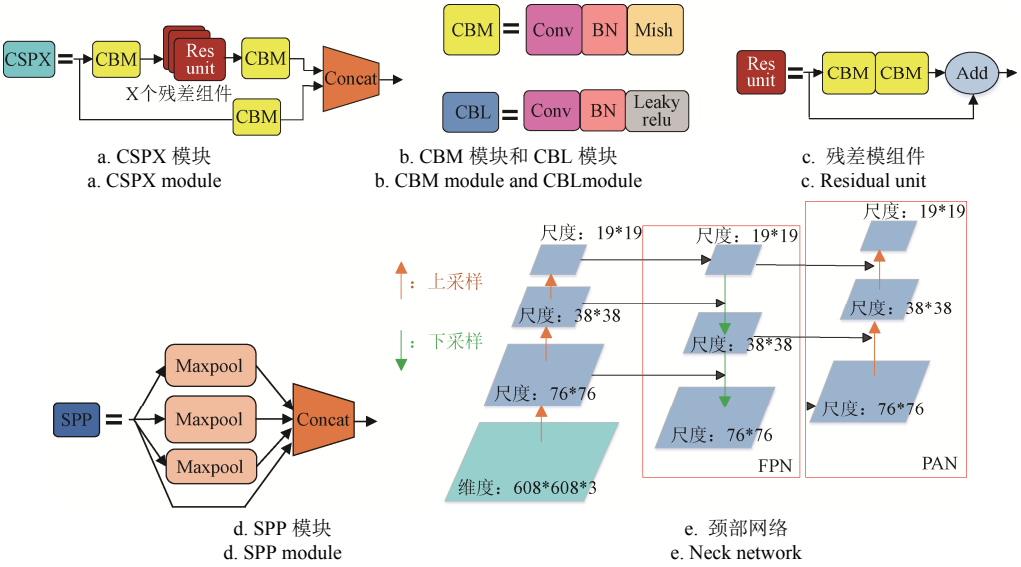
图 1　YOLOv4 网络整体框架

Fig.1　YOLOv4 network framework

1.2　改进的 YOLOv4 果园障碍物检测模型

YOLOv4 的作者公布了在 MS COCO 数据集上的训练结果[23]，准确率达到了 43.5%，比 YOLOv3 提高了 10%，并且速度也得到了提升，高达 45.2 帧/s，比 YOLOv3 快

了 12%。尽管 YOLOv4 检测模型性能优良，然而它使用的 CSPDarknet 主干网络参数量庞大，在特征提取的过程中计算参数量很大，需要耗费较长的时间。由于农业机器人在障碍物检测时应具有较高的实时性以便快速做出反应，所以改进模型来减少参数量势在必行。



a. CSPX 模块
a. CSPX module

b. CBM 模块和 CBL 模块
b. CBM module and CBLmodule

c. 残差模组件
c. Residual unit

d. SPP 模块
d. SPP module

e. 颈部网络
e. Neck network

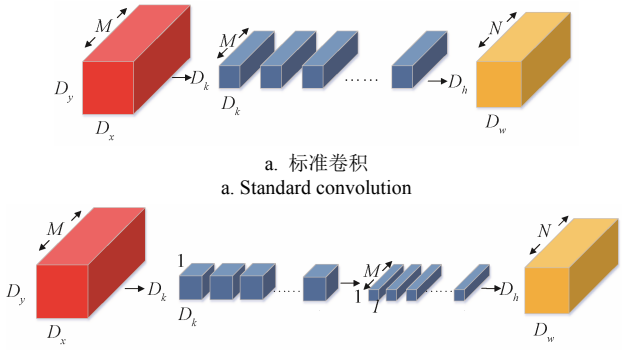
注：CSPX 表示内部带有 X 个残差组件的 CSP 模块，Maxpool 表示最大池化层。
Note: CSPX represents a CSP module with X residual components inside.Maxpool represents the maximum pool layer.

图 2 模块组成

Fig.2 Module components

1.2.1 深度可分离卷积

Sandler 等[24]在 2017 年提出了专用于嵌入式移动设备的轻量模型 Mobilenetv2，将标准卷积拆分为深度可分离卷积（Depthwise Separable Convolution）。如图 3 所示。



a. 标准卷积
a. Standard convolution

b. 深度可分离卷积
b. Depthwise separable convolution

注：M 和 N 分别为输入和输出数据的通道数；D_x 和 D_y 分别为输入数据的长度和宽度；D_k 为卷积核的尺寸；D_w 和 D_h 分别为输出数据的长度和宽度。
Note: M and N are the number of channels for input and output data; D_x and D_y are the length and width of input data; D_k is the size of convolution kernel; D_w and D_h are the length and width of output data.

图 3 标准卷积与深度可分离卷积

Fig.3 Standard convolution and depthwise separable convolution

图 3a 中标准卷积的过程是将各通道的输入特征图与相应的卷积核做卷积操作后相加再输出特征。传统标准卷积操作的计算量 q_1 为

$$q_1 = D_k^2 \cdot M \cdot N \cdot D_w \cdot D_h \quad (1)$$

而图 3b 中深度可分离卷积把传统卷积中的一步卷积操作分离为一个 3×3 的深度卷积（Depthwise Convolution，DWC）和一个 1×1 的逐点卷积（Pointwise Convolution，PWC）两步操作，它的卷积操作计算量 q_2 为

$$q_2 = D_k \cdot D_k \cdot M \cdot D_w \cdot D_h + M \cdot N \cdot D_w \cdot D_h \quad (2)$$

最后可以推出深度可分离卷积与标准卷积之间的计算量比值为

$$\frac{q_2}{q_1} = \frac{(M \cdot D_w \cdot D_h) \cdot (D_k^2 + N)}{D_k^2 \cdot M \cdot N \cdot D_w \cdot D_h} = \frac{1}{N} + \frac{1}{D_k^2} \quad (3)$$

利用深度可分离卷积后，计算量和参数会下降到原来的 1/4 左右，这样的改进能够显著地提高检测速度。

1.2.2 逆残差组件

YOLOv4 原本的残差结构（Residual Unit）中使用了传统卷积。由深度可分离卷积改进的 Residual Unit，如果采用瓶颈结构（bottleneck）则先用 1×1 PWC 降维至 0.25 倍通道数后进行 3×3 DWC，再用 1×1 PWC 卷积升维。然而 Depthwise Convolutions 含有的参数较少，如果先进行降维压缩再进行 3×3 的 DWC 的话，会造成提取的特征过少，从而影响检测精度。因此本研究采用逆残差组件（Inverted Residual Unit，InvRes Unit），用 PWC 升维扩张至 6 倍通道数后进行 DWC 特征提取，再用 PWC 降维压缩。这样形似倒瓶颈的结构使得特征提取在高维度进行，有利于提取更多的信息，能够在减少参数量的同时保持高精度。

1.2.3 软性非极大值抑制 Soft-DIoU-NMS 算法

在深度学习目标算法中，对于同一检测对象会产生许多不同大小的预测框，而本研究只需要保留一个最佳的预测框。非极大值抑制（Non-Maximum Suppression，NMS）[25]算法的作用就是将某一类别的预测框按置信度排序，把最高分值的框设定为基准框，然后分别与剩余的框作交并比（Intersection over Union，IoU）计算，大于设定阈值的删除，小于阈值的保留，并依次循环，从而达到消除冗余重复的窗口，找到物体最佳位置的目

的。在 YOLOv4 中使用的（Distance-IoU-NMS，DIoU-NMS）则是改进过的版本，DIoU-NMS 除了考虑预测框重叠区域的 IoU 外，还考虑了两个预测框中心点之间的距离，能够有效解决检测框的损失函数收敛慢的问题。DIoU-NMS 的计算式为

$$\begin{cases} S_i=\begin{cases} S_i, & \mathrm{IoU}-R_{\mathrm{DIoU}}(\mu,B_i)<\varepsilon \\ 0, & \mathrm{IoU}-R_{\mathrm{DIoU}}(\mu,B_i)\geqslant\varepsilon \end{cases} \\ R_{\mathrm{DIoU}}=\dfrac{\rho^2(b,b^{gt})}{c^2} \end{cases} \tag{4}$$

式中 S_i 为当前类别的置信度得分，R_{DIoU} 为 DIoU 损失函数的惩罚项，B_i 表示当前类别中所有被比较的预测框，μ 表示所有预测框中置信度最高的那一个框，b 和 b^{gt} 表示 2 个预测框的中心像素点坐标，c 指的是两个预测框的外接框的对角线像素长度，ρ 表示欧式距离，ε 表示人为设定的阈值，一般取 0.5。

但是传统的非极大值抑制方式在检测紧密靠近且相互遮挡的物体时会存在漏检的问题，例如，当前检测出的结果为 n 个不同置信度的检测框。如果按照原来的 DIoU-NMS 方法进行处理，首先选中置信度最高的预测框，那么其余检测框在后续的 DIoU 比较中就会因为与置信度最高的预测框的重叠面积过大而被删除，造成误检率和漏检率增大。

在果园场景中，常常会出现树木或人的密集区域造成互相遮挡的情况[26]。针对这个问题，通过衰减函数来对与置信度最高的框有重叠部分的相邻检测框进行调整是个比较有效的方法。因此本研究中使用软性的 DIoU-NMS 算法（Soft-DIoU-NMS）。在该算法中，不再直接删除大于阈值的框，而是降低它们的置信度，与得分最高预测框的重叠度越高，其置信度下降得越快，反之则下降的越慢。进行目标检测时，如果使用改进后的 Soft-DIoU-NMS，首先按照置信度排序，选择得分最高的检测框为基准，其余的检测框为待处理框，经过第一次衰减后，计算置信度得分，保留置信度最高的检测框，并将置信度次高作为基准。经第二次衰减后，获取置信度得分，依次类推，处理后置信度不变。最终通过综合删选取得理想效果。由 Soft-NMS 原公式[27]的线性表达式结合 DIoU 方法得到 Soft-DIoU-NMS 的公式如下：

$$S_i=\begin{cases} S_i, & \mathrm{IoU}-R_{\mathrm{DIoU}}(\mu,B_i)<\varepsilon \\ S_i(1-\mathrm{IoU}(\mu,Bi)), & \mathrm{IoU}-R_{\mathrm{DIoU}}(\mu,B_i)\geqslant\varepsilon \end{cases} \tag{5}$$

Soft-DIoU-NMS 算法公式与普通的 NMS 相比复杂度几乎没有改变，且实现同样简单。

改进后的 YOLOv4 网络整体外观没有变化，内部模块中用深度可分离卷积代替标准卷积。将残差组件替换为逆残差组件。并将 NMS 算法更替为 Soft-DIoU-NMS 算法。

2　果园障碍物识别试验

2.1　试验平台

计算机视觉设备使用 ZED 高清相机，结合配套软件工具 SDK 和 OpenCV 库。深度学习硬件平台为一台拥有 Intel i9-10900K CPU、64 GB 内存、NVIDIA GTX 2080TI 型号 GPU 的计算机，安装有 CUDA10.0 版本的并行计算框架和 CUDNN7.6 版本的深度学习加速库。在 Tensorflow 深度学习框架上实现研究中 YOLOv4 目标检测模型的训练。试验平台如图 4 所示。

底座为一个由电机驱动的四轮差速转向小车，搭载本试验所用的计算机和摄像头，来模拟农业机器人行进过程中对障碍物的检测。

图 4　试验平台

Fig.4　Experiment platform

2.2　试验数据集的采集和标注

本试验中果园障碍物数据集于 7—9 月期间拍摄并制作而成，拍摄地点位于江苏大学校园内的一处果园，果园内有梨树、桃树、杏树等约 100 棵，高度 2～4 m，果树行间和两侧共有 15 座路灯和电线杆，采集图像时，让 3 名同学在果园内随意走动。用 800 万像素的摄像头共拍摄了 2 000 张原始图像，包含 3 种代表性障碍物：果树、行人、电线杆或灯杆。采集图片中所有目标障碍物按照距离分为近、中、远目标。对于包含行人的图像分别采集静止、移动中、站立、蹲、弯腰等姿态的人，以丰富数据集的多样性，从而提升目标检测模型的检测能力。

本研究训练模型采用 PASCAL VOC 的数据集格式，先用 Labelimg 标注工具对每张图片目标物体所在区域进行手工数据标注矩形框，得到真实框 ground truth 用于训练。本试验中设定果树的标签为 Tree，行人的标签为 Person，电线杆或灯杆的标签为 Pole。数据集按照 8∶1∶1 的比例划分为训练数据集、验证数据集与测试数据集。本试验中检测评价的指标包括准确率（Precision，P）、召回率（Recall，R）、调和均值 F_1、参数量（单位为 MB）。P、R 和 F_1 的计算式分别为

$$P=\frac{T_p}{T_p+F_p}\times 100\% \tag{6}$$

$$R=\frac{T_p}{T_p+F_N}\times 100\% \tag{7}$$

$$F_1=\frac{2PR}{P+R} \tag{8}$$

式中 T_p 表示正确检测到果树、行人或电线杆的数量，F_p 表示检测目标出现分类错误的数量，F_N 表示图片中的目

标漏检的数量，F_1 表示准确率 P 和召回率 R 的调和平均值。当 F_1 越逼近于 1 时说明模型优化得越好。

2.3 模型的训练

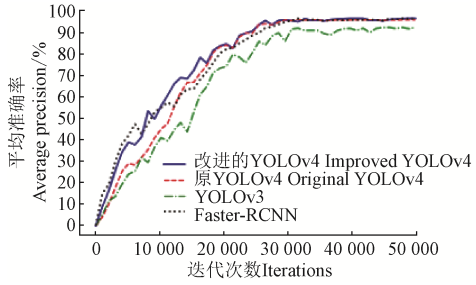
输入图像尺寸为 608×608 像素，为了增强模型的抗干扰能力，在训练时使用了多种数据增强的方法，包括随机裁剪、随机翻转、随机拉伸、随机失真、加入马赛克干扰。训练参数为：批量 16，动量 0.97，初始学习率 0.001，衰减系数为 0.9。

数据集内的图片通过图像增广将数量由原始的 2 000 张扩增到了 4 000。同时为了缩短训练时间加快迭代收敛，在试验中下载了 object365 数据集上的公开预训练模型用于迁移学习。将预训练模型的参数值除分类预测层之外都赋给 YOLOv4 模型。然后使用上述训练参数对 YOLOv4 的预训练模型进行训练调整。

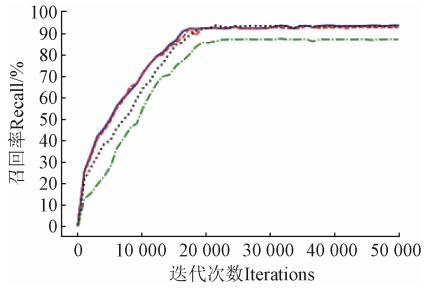
3 结果与分析

3.1 模型训练结果

为了验证改进后的模型与常用的典型模型的效果。分别对改进前后的 YOLOv4 模型、YOLOv3、Faster-RCNN 用同样的训练参数和数据集进行训练，4 种模型的训练都进行 50 000 次迭代，在训练集上每隔 2 000 次迭代就在验证集上测试一轮平均准确率和召回率，并保存一次模型。根据记录下的训练日志生成变化曲线图，如图 5 所示。



a. 平均准确率变化
a. Change of average precision



b. 召回率变化曲线
b. Change of recall

图 5 平均准确率和召回率变化

Fig.5 Change of average precision and recall

训练完成后，取果园障碍物数据集中的测试集图片用于测试各模型的指标得分并列表统计，统计结果如表 1 所示。由表中结果可知，改进后 YOLOv4 模型进行果园障碍物检测时在准确率方面比原 YOLOv4、YOLOv3、Faster-RCNN 分别提高了 0.61、4.18、0.04 个百分点。在召回率方面，分别提高了 0.68、6.37、0.11 个百分点。同时，改进后的 YOLOv4 模型参数量压缩了原 YOLOv4 的 75%，比 YOLOv3 小 68.7%，比 Faster-RCNN 缩小了 81%。在检测速度方面，改进后的 YOLOv4 比原 YOLOv4 快 29.4%，比 YOLOv3 快 22.1%，比 Faster-RCNN 快 346%。表明改进后的 YOLOv4 具有更优秀的性能。

表 1 不同模型检测结果比较

Table 1 Comparison of detection results among different models

模型 Model	P/%	R/%	F_1/%	检测速度 Detection speed (帧·s^{-1})	参数量 Parameters quantity /MB
改进的 YOLOv4 Improved YOLOv4	96.92	91.43	94.09	58.5	35
原 YOLOv4 Original YOLOv4	96.31	90.75	93.44	45.2	140
YOLOv3	92.74	85.06	88.73	47.9	112
Faster-RCNN	96.88	91.25	93.98	13.1	186

注：P 为准确率；R 为召回率；F_1 为 P 和 R 的调和均值。

Note: P is precision; R is recall; F_1 is harmonic mean value of P and R.

3.2 不同距离下改进前后模型检测效果

果园中的障碍物按照与农业机器人的距离可分为：近距离目标、中距离目标、远距离目标。其中近距离目标定义为和摄像头距离 1～5 m，中距离目标定义为距离摄像头 5～10 m，远距离目标定义为距离摄像头 10～20 m。鉴于农业机器人的实际工作需要，这里 20 m 以上的目标不在考虑范围内。

为了详细地对比不同距离下改进后的模型与其他模型检测能力指标，额外准备 100 张不同于数据集的图片，每张图片包含不同距离下 3 种的果园障碍物。100 张图片内的不同类别果园障碍物的数量如表 2 所示。

表 2 不同距离下不同类别测试目标数量

Table 2 Number of test targets in different categories at different distances

距离 Distance	果树 Tree	行人 Person	电线杆 Pole
近 Close	87	283	30
中 Middle	46	39	15
远 Far	65	107	28

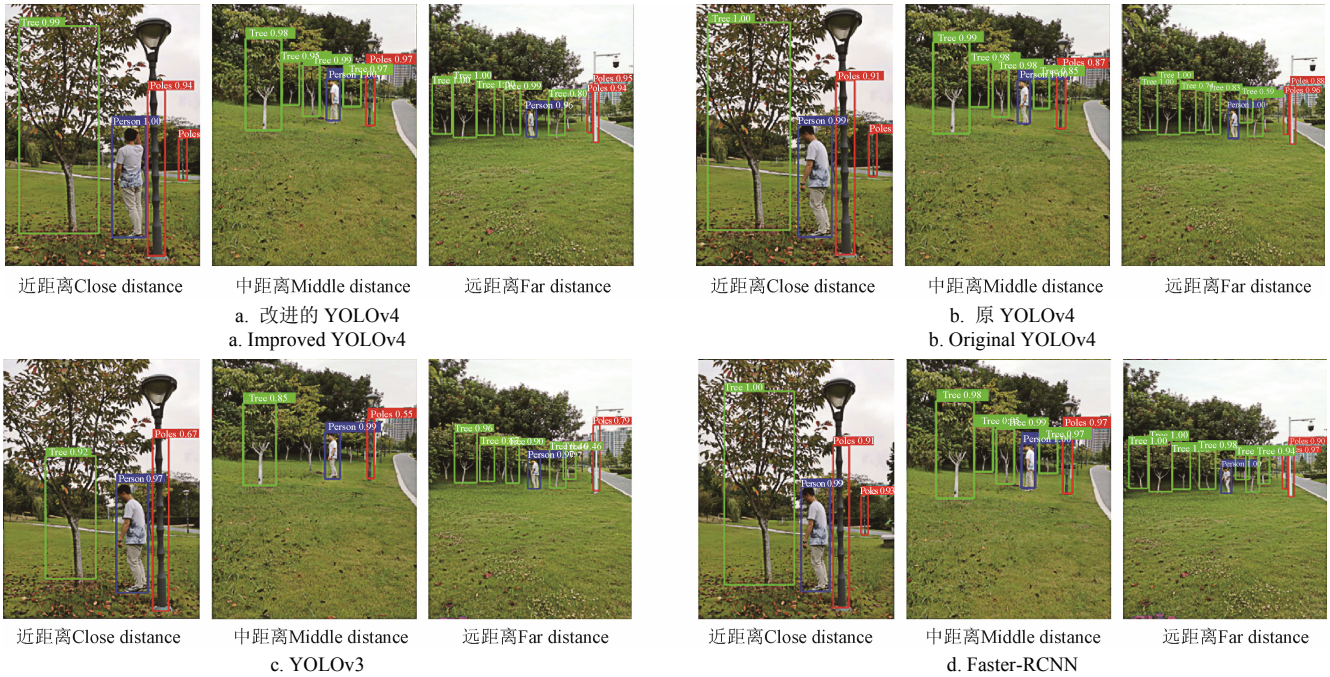
不同距离下不同模型的 P、R、F_1 指标如表 3 所示。

表 3 不同距离下不同模型的障碍物检测结果

Table 3 Obstacle detection results of different models at different distances

距离 Distance	模型 Model	P/%	R/%	F_1/%
近 Close	改进的 YOLOv4	97.13	90.76	93.84
	原 YOLOv4	96.89	90.53	93.60
	YOLOv3	94.37	88.04	91.09
	Faster-RCNN	97.05	91.12	93.99
中 Middle	改进的 YOLOv4	92.25	88.78	90.48
	原 YOLOv4	92.23	88.44	90.29
	YOLOv3	88.67	83.61	86.06
	Faster-RCNN	91.32	87.97	89.61
远 Far	改进的 YOLOv4	81.10	74.62	77.73
	原 YOLOv4	83.75	75.90	79.63
	YOLOv3	75.87	69.45	72.52
	Faster-RCNN	80.44	74.28	77.24

分别用 4 种模型对不同距离下的目标检验其检测效果，检测效果如图 6 所示。



注：图中标签为 Tree 的检测框代表检测到的果树，标签为 Person 的检测框代表检测到的行人，标签为 Pole 的检测框代表检测到的电线杆。
Note: The detection boxes labeled “Tree” represents the detected trees, the detection boxes labeled “Person” represents the detected pedestrians, and the detection boxes labeled “Pole” represents the detected poles.

图 6　不同模型对不同距离目标检测效果

Fig.6　Detection results of the different models for different distance targets

由于使用了 Soft-DIoU-NMS 算法，改进后的模型在检测目标密集或重叠区域时漏检率大幅减少，所以改进后的模型在中近距离上检测目标的能力均超过了原模型和其他模型。在远距离上，改进后的 YOLOv4 的检测能力，强于 YOLOv3 和 Faster-RCNN 模型，但比原 YOLOv4 模型在准确率上低了 2.65%。这是由于逆残差结构的存在使得对小目标的特征敏感度变低造成的。然而在实际农业机器人的应用中，远处的障碍物随着机器人的运动由远及近，因此障碍物的避让决策会优先处理中近距离的目标，同时农业机器人的运动速度一般较慢，所以远目标的检测能力稍弱对农业机器人的避障几乎没有影响。因此从数据结果分析可知：改进后的 YOLOv4 模型更适合完成农业机器人对果园障碍物的检测任务。

4　结　论

1）本研究提出了一种改进型的 YOLOv4 目标检测模型用于果园多种障碍物的检测。利用深度可分离卷积代替原有的标准卷积，并在主干网络中用逆残差组件代替了原有的残差组件。使得模型的参数量和计算量仅为原来的 1/4 左右，更轻量化，利于农业嵌入式移动设备的模型部署。同时，改进了非极大值抑制的方式，采用 Soft-DIoU-NMS 来减少冗余框，对重叠目标的检测精度更高。使得果园农业机器人能更安全地行驶作业。

2）根据果园内的主要障碍物的类别，制作了包括果树、行人、电线杆这 3 类障碍物的图像数据集分别用于改进前后 YOLOv4 目标检测模型的训练和测试，并分别在近、中、远目标上对改进前后的模型和 YOLOv3、Faster-RCNN 模型进行对比试验。结果表明，改进后的模型具有较高的准确度和实时性，准确率和召回率分别达到了 96.92%和 91.43%，视频流检测速度达到了 58.5 帧/s，模型参数量仅有 35 MB。本研究中改进的 YOLOv4 在提升了精准度的情况下，大幅度减少了模型的参数量，增强了实时性。本研究所提方法特别适用于中近距离目标的检测，更好地满足了农业机器人的实际应用场景。

[参　考　文　献]

[1] 赵献立，王志明. 机器学习算法在农业机器视觉系统中的应用[J]. 江苏农业科学，2020，48(12)：226-231.
Zhao Xianli, Wang Zhiming. Application of machine learning algorithm in agricultural machine vision system[J]. Jiangsu Agricultural Sciences, 2020, 48(12): 226-231. (in Chinese with English abstract)

[2] 王辉. 机器视觉技术在果园自动化中的应用研究[D]. 北京：中国农业机械化科学研究院，2011.
Wang Hui. Applied Research of Machine Vision In Orchard Automation[D]. Beijing: Chinese Academy of Agricultural Mechanization Sciences, 2011. (in Chinese with English abstract)

[3] 张舜，郝泳涛. 基于深度学习的障碍物检测研究[J]. 电脑知识与技术，2019，15(34)：185-187.
Zhang Shun, Hao Yongtao. Research on obstacle Detection Based on Deep Learning[J]. Computer Knowledge and Technology, 2019, 15(34): 185-187. (in Chinese with English abstract)

[4] 邓淇天. 基于激光雷达和视觉传感器融合的障碍物识别技术研究[D]. 南京：东南大学，2019.
Deng Qitian. Reasearch on Object Recognition Based on

Lidar and Camera Fusion[D]. Nanjing: Southeast University, 2019. (in Chinese with English abstract)

[5] 向海涛，郑加强，周宏平. 基于机器视觉的树木图像实时采集与识别系统[J]. 林业科学，2004，40(3)：144-148.
Xiang Haitao, Zhen Jiaqiang, Zhou Hongping. Real-time tree image acquisition and recognition system based on machine vision[J]. Scientia Silvae Sinicae, 2004, 40(3): 144-148. (in Chinese with English abstract)

[6] Dalal N , Triggs B . Histograms of oriented gradients for human detection[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE, 2005.

[7] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[8] Bay H, Ess A, Tuytelaars T. Speeded-up robust features (SURF)[J]. Computer Vision & Image Understanding, 2008, 110(3): 346-359.

[9] Deriche R. Using Canny's criteria to derive a recursively implemented optimal edge detector[J]. International Journal of Computer Vision, 1987, 1(2): 167-187.

[10] 刘慧，张礼帅，沈跃，等. 基于改进 SSD 的果园行人实时检测方法[J]. 农业机械学报，2019，50(4)：29-35，101.
Liu Hui, Zhang Lishuai, Shen Yue. Real-time pedestrian detection in orchard based on improved SSD[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(4): 29-35, 101. (in Chinese with English abstract)

[11] 景亮，王瑞，刘慧.基于双目相机与改进 YOLOv3 算法的果园行人检测与定位[J]. 农业机械学报，2020，51(9)：34-39，25.
Jing Liang, Wang Rui, Liu Hui. Orchard pedestrian detection and location based on binocular camera and improved YOLOv3 algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery,2020, 51(9):34-39, 25. (in Chinese with English abstract)

[12] 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.

[13] Girshick R. Fast R-CNN[C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[14] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.

[15] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904.

[16] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]// Computer Vision & Pattern Recognition. IEEE, 2016.

[17] Liu W, Anguelov D, Erhan D, et al.SSD: Single shot multiBox detector[C]// European Conference on Computer Vision. Springer International Publishing, 2016.

[18] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017: 6517-6525.

[19] He K, Zhang X, Ren S, et al. deep residual learning for image recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2016.

[20] 蔡逢煌，张岳鑫，黄捷. 基于 YOLOv3 与注意力机制的桥梁表面裂痕检测算法[J]. 模式识别与人工智能，2020，33(10)：926-933.
Cai Fenghuang, Zhang Yuexin, Huang Jie. Bridge surface crack detection algorithm based on YOLOv3 and attention mechanism[J]. Pattern Recognition and Artificial Intelligence, 2020, 33(10):926-933. (in Chinese with English abstract)

[21] 刘洋，姜涛，段学鹏. 基于 YOLOv3 的复杂天气条件下人车识别方法的研究[J]. 长春理工大学学报：自然科学版，2020，43(6)：57-65.
Liu Yang, Jiang Tao, Duan Xuepeng. Research on recognition method of pedestrian and vehicle under complex weather conditions based on YOLOv3[J]. Journal of Changchun University of Science and Technology: Natural Science Edition, 2020, 43(6): 57-65. (in Chinese with English abstract)

[22] 张健. 基于改进 YOLOv3 的果园行人检测方法研究[D]. 镇江：江苏大学，2020.
Zhang Jian. Research on Orchard Pedestrian Detection Method Based on Improved YOLOv3[D]. Zhenjiang: Jiangsu University, 2020. (in Chinese with English abstract)

[23] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934,2020.

[24] Sandler M, Howard A, Zhu M L, et al. MobileNetV2:Inverted residuals and linear bottlenecks[C] //Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.

[25] Neubeck A, Van Gool L. Efficient Non-Maximum Suppression[C]//International Conference on Pattern recognition, 2006:850-855.

[26] 汤庆闻. 互斥损失优化的密集行人检测算法[D]. 武汉：华中科技大学，2019.
Tang Qinwen. Intensive Pedestrian Detection Algorithm Optimized With Exclusion Loss[D]. Wuhan: Huazhong University of Science and Technology, 2019. (in Chinese with English abstract)

[27] Bodla N, Singh B, Chellappa R, et al. Soft-NMS improving object detection with one line of code[C]//2017 IEEE International Conference on Computer Vision. 2017:5561-5569.

Real-time detection methodology for obstacles in orchards using improved YOLOv4

Cai Shuping, Sun Zhongming, Liu Hui※, Wu Hongxuan, Zhuang Zhenzhen

(*School of Electrical and Information Engineering, Jiangsu University, Zhenjiang* 212013*, China*)

Abstract: China is one of the countries with the largest cultivation area of the orchards in the world. The traditional orchard planting is quite time-consuming and laborious. An orchard robot can be expected as an important artificial intelligence (AI) tool to replace the manual labor in orchard management. However, the robot can encounter various obstacles in the actual operation, due to the complex and changeable environment of an orchard. It is necessary for the agricultural robots to real-time detect obstacles during operation. In recent years, various target detection systems have been widely used for agricultural robot avoidance, such as YOLOv4, YOLOv3, and Faster-RCNN, particularly with the rise of intelligent deep learning. Generally, there were some problems, including the unsatisfactory detection accuracy, a large number of parameters required in the models, low real-time performance, and difficulty in detecting densely overlapping target areas. In this study, an improved YOLOv4-based model of target detection was proposed with the help of the latest vision sensor technology to realize that the agricultural robots can quickly and accurately identify and classify the obstacles in the orchard. A deep separable convolution was utilized to reduce the number of parameters, and further improve the detection speed. An Inverted Residual Unit was selected to replace the Residual Unit in the core network CSP-Darknet in the previous model. In addition, a Soft DIoU-Non-Maximum Suppression (Soft-DIoU-NMS) algorithm was employed to detect the dense areas. Three common obstacles, including pedestrians, fruit trees, and telegraph poles, in the orchards were selected as the detection objects to generate an image dataset. The improved model was trained on the Tensorflow deep learning framework, and then the test images were input into the trained model to detect target obstacles at different distances. Under the same evaluation index, an evaluation was made on the improved YOLOv4, the original YOLOv4, YOLOv3, and Faster-RCNN. The results showed that the improved YOLOv4-based detection model for orchard obstacles achieved an average accuracy rate of 96.92%, 0.61 percent point higher than that of the original YOLOv4 model, 4.18 percent point higher than that of the YOLOv3 model, and 0.04 percent point higher than that of Faster-RCNN model. The recall rate of the proposed model reached 96.31%, 0.68 percent point higher than that of the original YOLOv4, 6.37 percent point higher than that of YOLOv3, and 0.18 percent point higher than that of Faster-RCNN. The detection speed in the improved YOLOv4-based video stream was 58.5 frames/s, 29.4% faster than that in the original YOLOv4, 22.1% faster than that in YOLOv3, and 346% faster than that in Faster-RCNN. The number of parameters in the improved YOLOv4-based model was reduced by 75%, compared with the original YOLOv4 model, 68.7% less than that of the YOLOv3 model, and 81% less than that of the Fasters-RCNN model. In general, the proposed model can greatly reduce its size without losing accuracy, and thereby enhance the real-time performance and robustness in the actual orchard environment. The improved YOLOv4-based model achieved ideal effects in different distance tests, indicating better performance for the obstacle detection in the orchard environment. The findings can provide a strong guarantee for the obstacle avoidance of intelligent robots in orchards.

Keywords: agriculture; robots; object detection; deep learning; depthwise separable convolution; inverted residual unit; Soft-DIoU-NMS