doi:10.6041/j.issn.1000-1298.2023.08.026

基于改进 YOLO v5s 的轻量化植物识别模型研究

马宏兴[1]　董凯兵[1]　王英菲[1]　魏淑花[2]　黄文广[3]　苟建平[4]

(1.北方民族大学电气信息工程学院,银川 750021;2.宁夏农林科学院植物保护研究所,银川 750002;
3.宁夏回族自治区草原工作站,银川 750002;4.西南大学计算机与信息科学学院,重庆 400715)

摘要:为方便调查宁夏全区荒漠草原植物种类及其分布,需对植物识别方法进行研究。针对 YOLO v5s 模型参数量大,对复杂背景下的植物不易识别等问题,提出一种复杂背景下植物目标识别轻量化模型 YOLO v5s-CBD。改进模型 YOLO v5s-CBD 在特征提取网络中引入带有 Transformer 模块的主干网络 BoTNet(Bottleneck transformer network),使卷积和自注意力相结合,提高模型的感受野;同时在特征提取网络融入坐标注意力(Coordinate attention,CA),有效捕获通道和位置的关系,提高模型的特征提取能力;引入 SIoU 函数计算回归损失,解决预测框与真实框不匹配问题;使用深度可分离卷积(Depthwise separable convolution,DSC)减小模型内存占用量。实验结果表明,YOLO v5s-CBD 模型在单块 Nvidia GTX A5000 GPU 单幅图像推理时间仅为 8 ms,模型内存占用量为 8.9 MB,精确率 P 为 95.1%,召回率 R 为 92.9%,综合评价指标 F1 值为 94.0%,平均精度均值(mAP)为 95.7%,在 VOC 数据集平均精度均值可达 80.09%。相比 YOLO v3-tiny、YOLO v4-tiny 和 YOLO v5s,改进模型内存占用量减小,平均精度均值提升。模型 YOLO v5s-CBD 在公开数据集和宁夏荒漠草原植物数据集都有良好的鲁棒性,推理速度更快,且易于部署,已应用在宁夏荒漠草原移动端植物图像识别 APP 和定点生态信息观测平台,可用来调查宁夏全区荒漠草原植物种类和分布,长期观测和跟踪宁夏盐池县大水坑、黄记场、麻黄山等地植物生态信息。

关键词:植物识别;YOLO v5s;BoTNet;坐标注意力;深度可分离卷积;轻量化

中图分类号:S4　　**文献标识码**:A　　**文章编号**:1000-1298(2023)08-0267-10　　**OSID**:

Lightweight Plant Recognition Model Based on Improved YOLO v5s

MA Hongxing[1]　DONG Kaibing[1]　WANG Yingfei[1]　WEI Shuhua[2]　HUANG Wenguang[3]　GOU Jianping[4]

(1. *College of Electrical and Information Engineering, North Minzu University, Yinchuan* 750021, *China*
2. *Institute of Plant Protection, Ningxia Academy of Agricultural and Forestry Sciences, Yinchuan* 750002, *China*
3. *Ningxia Grassland Workstation, Yinchuan* 750002, *China*
4. *College of Computer and Information Science, Southwest University, Chongqing* 400715, *China*)

Abstract: In ordered to facilitate the investigation of desert grassland plant species and their distribution in the whole Ningxia region, plant identification methods need to be studied. To address the problems of large number of parameters in YOLO v5s model, it is not easy to recognize plants in complex backgrounds, and a lightweight model of plant target recognition in complex backgrounds, YOLO v5s-CBD, was proposed. The improved model YOLO v5s-CBD introduced the BoTNet with Transformer module into the feature extraction network, to combine convolution and self-attention to improve the feeling field of the model. At the same time, coordinate attention was incorporated into the feature extraction network to effectively capture the relationship between channel and position and improve the feature extraction ability of the model. In terms of loss calculation, the SIoU function was introduced to calculate the regression loss to solve the problem of mismatch between the prediction box and the real box. Using depthwise separable convolution to reduce model volume. The experimental results showed that the model YOLO v5s-CBD infers a single image in only 8 ms, a model volume of 8.9 MB, a precision of 95.1%, a recall of 92.9%, a F1 value of 94.0%, and a mean average precision of 95.7% in a single Nvidia GTX A5000 GPU, and a mean average precision of 80.09% in the VOC dataset.

收稿日期:2023-03-29　修回日期:2023-05-25

基金项目:宁夏农业高质量发展和生态保护科技创新项目(NGSB-2021-14-05)、国家自然科学基金面上项目(61976107)和北方民族大学重点研究项目(2021JY005、YCX22134)

作者简介:马宏兴(1975—),男,副教授,博士,主要从事信息感知与智能处理研究,E-mail:mhx@nmu.edu.cn

Compared with YOLO v3 - tiny, YOLO v4 - tiny and YOLO v5s, the improved models reduced model volume and improved mean average precision. The model YOLO v5s - CBD had good robustness in both public dataset and Ningxia desert grassland plant dataset, faster inference speed and easy to deploy. It was applied in Ningxia desert grassland mobile plant image recognition APP and fixed ecological information observation platform, which can be used to investigate the species and distribution of desert grassland plants in the whole region of Ningxia, and long-term observation and tracking of Dashuikeng, Huangjichang, Mahuangshan and other places, Yanchi County, Ningxia.

Key words: plant recognition; YOLO v5s; BoTNet; coordinate attention; depthwise separable convolution; lightweight

0 引言

宁夏地处我国西北农牧交错带,在农牧业发展中占据重要的地位。宁夏荒漠草原是全区分布面积最大的天然草原,约占草地总面积的一半以上。近年来,在各种自然因素及人类行为干扰下,荒漠草原出现诸如土壤沙化、草地生态系统的生产力下降、土壤肥力流失等一系列生态问题[1-2]。面对这些问题,快速有效识别荒漠草原植物群落主要物种,是调查草地群落结构变化、划分草地类型、明确草地退化和修复现状,以及保护草地植物多样性的基础[3-4]。

在实际调查中,植物物种的识别主要按照传统方法进行,既耗时又复杂。随着计算机视觉技术的进步,在植物识别领域已有大量基于机器学习的植物识别方法,能够快速有效识别植物种类。目前对植物识别的研究中,有支持向量机[5]、K 最近邻[6]、深度学习和胶囊网络等[7]。在深度学习领域中,孙俊等[8]针对 CNN 训练收敛时间长、模型参数庞大的问题,对网络模型进行改进,提出一种批归一化与全局池化相结合的卷积神经网络识别模型,对 14 种不同植物共 26 类病害的平均测试识别准确率达到 99.56%,参数内存占用量为 2.6 MB,推理时间为 20.79 ms。GUANG 等[9]提出了一种基于高效通道注意力(ECAENet)的轻量级植物物种识别算法,使用神经架构搜索来获得基线网络,复合系数统一缩放深度、宽度和分辨率的所有维度,在 Swedish Leaf、Flavia Leaf 数据集的准确率分别达到 99.56% 和 99.75%。BARRADAS 等[10]基于 EfficientNet 模型的自然环境中药用植物物种的实时自动识别系统,识别精度可达到 78.5%。CHEN 等[11]基于改进 YOLO v5 对橡树病害进行识别,从橡胶树病害数据库中随机抽取样本图像形成训练集和测试集,改进网络平均精度达到 70%,相比原模型提高 5.4%。项和雨等[12]通过改进的深度卷积神经网络 RAN - 11,以残差注意力 Attention - 56 模块和 Attention - 92 为基础,将底层与顶层的通道特征融合,调整注意力模块和残差块个数压缩模型,对浮游植物识别准确率可达到 95.67%。在胶囊网络识别中,温长吉等[13]提出一种改进稠密胶囊网络模型,对 32 像素 × 32 像素的植物图像,平均识别准确率达到 77.2%,参数量仅为 1.8×10^6。

上述算法中,支持向量机针对高维度数据计算复杂;K 近邻需要大量数据集支撑训练,而且占用内存大; EfficientNet 模型体积较大,推理速度慢; YOLO v5 对于复杂植物图像识别精度较低;胶囊网络适合低分辨率图像,对分辨率较高图像不友好。

本文基于 YOLO v5s 网络,以宁夏荒漠草原植物为研究对象,拟对复杂背景下轻量级植物目标识别模型进行研究。在模型中拟引入 BoTNet[14]和坐标注意力机制 CA[15]来有效提升对植物特征的提取能力;引入深度可分离卷积 DSC 来减小模型体积[16],加快模型推理速度,实现模型的轻量化;引入回归损失函数 SIoU 来加快数据拟合[17];同时,对改进后的轻量化识别模型应用进行研究,以期应用于宁夏荒漠草原植物移动端识别和定点观测。

1 植物数据集

采集宁夏荒漠草原 17 类常见植物,对植物数据集进行清洗和增强处理,如图 1 所示。

1.1 植物数据

主要采集宁夏盐池县麻黄山、大水坑,同心县罗山等地区 17 类常见植物,分别为青海苜蓿(*Medicago archiducis-nicolaii*)、长芒草(*Stipa bungeana*)、甘草(*Glycyrrhiza uralensis*)、猪毛菜(*Salsola collina*)、兴安胡枝子(*Lespedeza davurica*)、二色补血草(*Limonium bicolor*)、蒲公英(*Taraxacum mongolicum*)、草木犀状黄芪(*Astragalus melilotoides*)、平车前(*Plantago depressa*)、小车前(*Plantago minuta*)、反枝苋(*Amaranthus blitum*)、乳苣(*Lactuca tatarica*)、刺齿枝子花(*Dracocephalum peregrinum*)、披针叶黄华(*Thermopsis lanceolata*)、地梢瓜(*Cynanchum thesioides*)、二裂委陵菜(*Potentilla bifurca*)、圆叶锦葵(*Malva pusilla*)。

由于宁夏属温带大陆性干旱、半干旱气候,降水

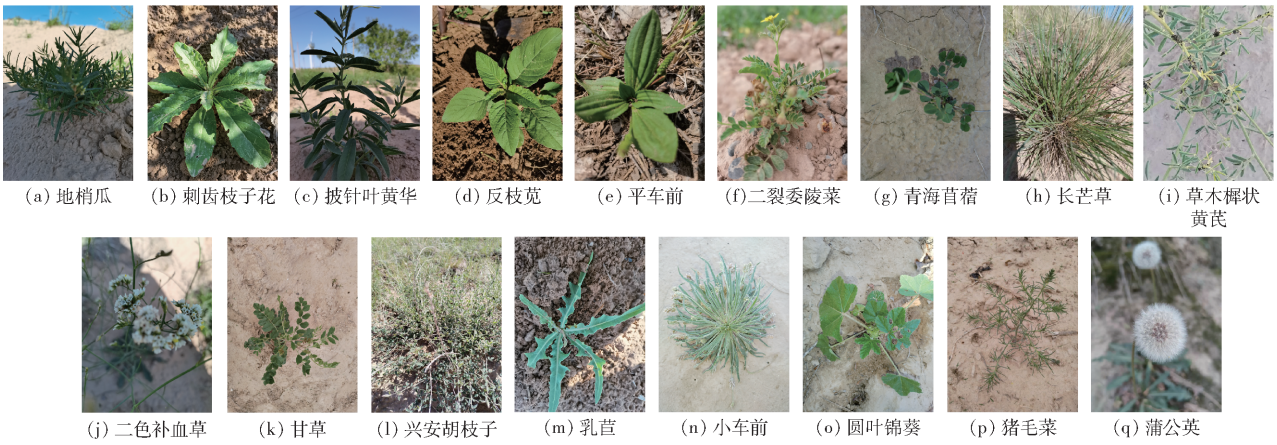


图 1　植物部分图像数据

Fig. 1　Images of plants

偏少且相对集中，光能丰富，蒸发强烈，植被覆盖以沙生植物为主，大部分植物叶片窄小。当对叶片进行采集时，难以获取有效信息，因此，植物数据集采用整株植物。由于荒漠草原植物在自然环境中与背景相似、特征不明显，不利于识别与分析，为丰富植物数据多样性，提高识别模型泛化能力，数据采集时选取不同角度、不同光照环境下的自然图像数据，分辨率为 3 000 像素 ×4 000 像素。

采集得到宁夏荒漠草原常见植物图像数据共 17 类，1 953 幅，7.2 GB。对部分拍摄模糊、枝叶遮挡等特征不明显的植物图像进行数据清洗后，得到植物图像数据共 17 类，1 892 幅，其中青海苜蓿 94 幅、长芒草 101 幅、甘草 66 幅、猪毛菜 103 幅、兴安胡枝子 88 幅、二色补血草 120 幅、蒲公英 156 幅、草木犀状黄芪 142 幅、平车前 136 幅、小车前 107 幅、反枝苋 101 幅、乳苣 132 幅、刺齿枝子花 132 幅、披针叶黄华 117 幅、地梢瓜 108 幅、二裂委陵菜 92 幅、圆叶锦葵 97 幅，7.0 GB。

1.2　数据增强

采用马赛克数据增强和传统图像增强相结合的方法，扩展宁夏荒漠草原植物图像样本的丰富度，提高模型的泛化能力和鲁棒性。在此选择 4 幅图像进行马赛克增强，马赛克增强实验结果如图 2 所示。



(a) 原始数据　　(b) 增强数据

图 2　数据增强

Fig. 2　Data augmentation

1.3　数据标注

为了提高目标检测模型输出的准确性，使得网络模型能够学习到植物的特征与位置信息，需要对图像中的植物进行标注。标注时使用 LabelImg 工具制作 VOC 格式数据集，文件中包含植物名称和位置信息。图像和标注文件均用于模型训练和验证。

2　轻量化植物识别模型设计

由于 YOLO v5s 对复杂目标识别精度低，参数量较大，所以在实际应用中需要对网络模型进行改进，提升识别精度的同时，降低模型的参数量，加快模型推理速度。

2.1　网络模型结构设计

改进的网络模型命名为 YOLO v5s－CBD，模型结构如图 3 所示。

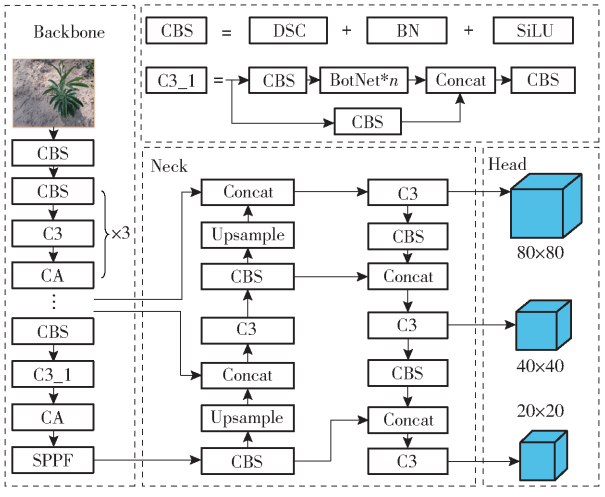


图 3　YOLO v5s－CBD 结构

Fig. 3　YOLO v5s－CBD structure

图 3 中，YOLO v5s－CBD 分为 3 部分，其中 Backbone 对输入植物图像进行特征提取，Neck 对获取的特征图进行特征融合，Head 进行回归预测。

YOLO v5s－CBD 首先在特征提取网络

Backbone 中引入带有 Transformer[18] 模块的轻量级主干网络 BoTNet,扩大模型的全局感受野;同时融入轻量级坐标注意力机制 CA 来有效捕获位置和通道的关系,提高模型对关键信息的特征提取能力;其次在输出端 Head 引入 SIoU 函数计算回归损失,提升模型的收敛能力;最后使用深度可分离卷积 DSC 减小模型的体积,实现改进模型的轻量化,提升模型识别速度。

2.2 BoTNet 网络

由于高分辨率植物图像中部分图像背景复杂,特征不明显,而 YOLO v5s 主干特征提取网络为 C3 网络,只具有平移不变性和局部性,不具有全局建模和长距离建模能力,为了提高模型的特征提取能力,在此引入 BoTNet。

BoTNet 是一种结合 CNN 和 Transformer 的骨干网络,同时使用了卷积和自注意力机制,其模型结构是在 ResNet[19] 的最后 3 个 bottleneck blocks 中使用全局多头自注意力(Multi-head self-attention, MHSA)替换 3×3 空间卷积。全局多头自注意力是 Transformer 的核心结构,可以汇总和处理特征图中包含的目标信息,提高目标的识别能力。全局多头自注意力的结构如图 4 所示。

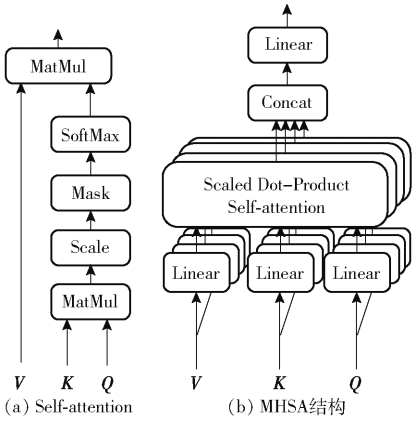


图 4 MHSA 的结构图

Fig. 4 Structural diagrams of MHSA

图 4a 为单头自注意力机制(Self-attention),计算公式为

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \quad (1)$$

式中 $\boldsymbol{V}$、$\boldsymbol{K}$、$\boldsymbol{Q}$——由输入矩阵通过线性变换得到的矩阵,维度与输入矩阵相同

$\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}$——经过矩阵乘法所得的矩阵,计算 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的相似度

$\sqrt{d_k}$——缩放因子,防止维度过高时 $\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}$ 的值过大导致 Softmax 函数反向传播时发生梯度消失

Softmax——归一化函数,与 $\boldsymbol{V}$ 进行映射,获得通道的相关性

图 4b 为全局多头自注意力,首先将数据输入 4 个 Self-attention,得到 4 个加权后的特征矩阵,然后将 4 个特征矩阵按列拼接成大矩阵,经过一层全连接得到最终的输出。

在 YOLO v5s 中引入 BoTNet,可以将卷积和自注意力机制相结合,弥补卷积信息丢失问题。BoTNet 可以利用卷积有效地从高分辨率图像中学习抽象的低分辨率特征图,同时也使用自注意力机制来处理和聚合卷积捕获的特征图中包含的信息,扩展了模型专注于不同位置的能力,而且相比 3×3 的矩阵,BoTNet 使用 1×1 的矩阵减小了模型参数量,有效提升目标识别性能的同时,兼顾模型的轻量化。

2.3 坐标注意力机制

注意力机制作为提高神经网络特征提取能力的一种有效方法,在自然语言处理、统计学习、语音和计算机视觉等领域有大量的应用[20]。

常见的注意力机制是通过加深网络的深度来提高模型的性能,但单纯加深网络深度会增加模型的计算量,一般轻量级网络所能承受的计算量有限,故常规的注意力机制在植物识别的轻量级网络中应用会受到影响。

由于采集的植物图像分辨率为 3 000 像素×4 000 像素,背景信息占比较高,在深度卷积神经网络中,背景信息卷积迭代会积累大量冗余的无效信息,导致目标信息淹没,识别的准确率降低。为提高改进模型 YOLO v5s-CBD 对荒漠草原植物关键信息的特征提取能力,解决无效冗余信息会降低植物识别准确率问题,在此采用轻量级坐标注意力机制(Coordinate attention, CA)。

CA 能充分利用通道注意力信息和空间注意力信息,有效提升网络模型的识别性能,结构如图 5 所示,其中 C 表示通道数,W 表示宽度,H 表示高度,r 表示维度[21]。

图 5 中,CA 首先利用 2 个一维全局池化操作分别将水平方向和垂直方向的输入特征聚合为 2 个不相关的方向感知特征图;接着 2 个嵌入垂直和水平方向信息的感知特征图分别被编码为 2 个注意力图,每个注意力图都捕获了输入特征图沿着一个空间方向的长程依赖,意味着位置信息被保存在生成的注意力图中;最后 2 个注意力图被乘到输入特征图上来增强特征图的表示能力。

在改进模型 YOLO v5s-CBD 中,CA 可以融入到主干网络 C3 模块之后或者颈部融合网络 C3 模

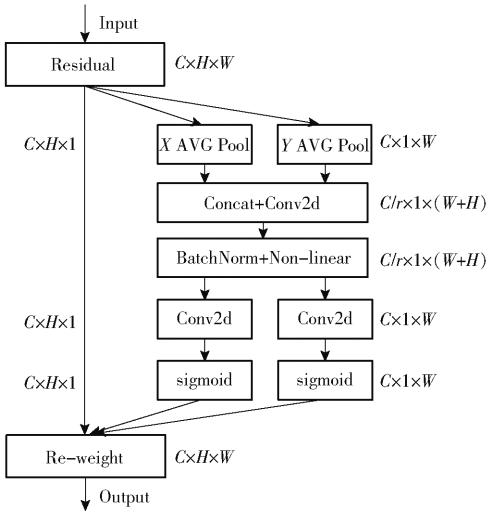


图 5　坐标注意力

Fig. 5　Coordinate attention

块之后。但在实际实验中，融入到主干网络 C3 模块之后性能更好，在提升识别精度的同时，每秒浮点运算次数（FLOPS）仅仅增加 1×10^8，几乎可以忽略不计。

2.4　SIoU

YOLO v5s 采用 CIoU 损失函数[22]，这种传统的目标检测损失函数依赖于边界框回归指标的聚合，没有考虑到所需真实框与预测框之间方向不匹配，导致收敛速度较慢，效率较低。为解决这一问题，改进模型中引入损失函数 SIoU，不仅考虑重叠区域、距离和长宽，还考虑预测框和真实框之间的角度问题。

总损失函数为

$$L = W_{box}L_{box} + W_{cks}L_{cls} \tag{2}$$

其中

$$L_{box} = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{3}$$

式中　IoU——传统回归损失

Δ——距离损失

Ω——形状损失

W_{box}——框损失模型体积

L_{box}——回归损失

W_{cks}——分类损失模型体积

L_{cls}——焦点损失

在改进模型引入了真实框与预测框之间的角度计算，可加速网络的收敛。

2.5　深度可分离卷积

模型压缩和加速不仅仅可以提升移动端模型性能，在服务端也可以加快推理响应速度，并减少服务器资源消耗，降低成本。

因此，为了进一步压缩网络参数，降低模型参数量，提升模型推理速度，减小资源的消耗。在改进模型 YOLO v5s－CBD 中引入轻量级卷积块深度可分离卷积来减小模型的参数量，在改进模型中原始卷积全部改进为深度可分离卷积。深度可分离卷积实现过程如图 6 所示。

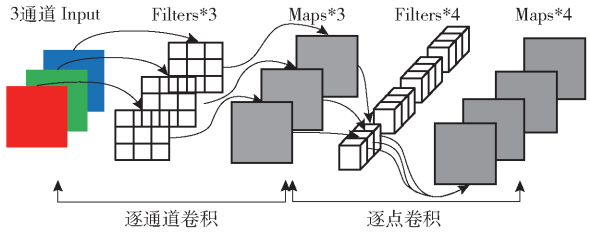


图 6　深度可分离卷积

Fig. 6　Depthwise separable convolution

由图 6 可知，深度可分离卷积分为逐通道卷积和逐点卷积 2 个过程，逐通道卷积是对输入的每个通道分别进行卷积运算，可在减少参数量的同时，保证模型的特征提取能力不变。但是这种卷积缺乏通道间的信息交流，而逐点卷积能够学习不同通道间在相同空间位置上的特征信息，重新组合生成新的特征图，从而增加了特征的学习利用率。

若常规卷积和深度可分离卷积输入相同，在得到 4 张 Feature map 的情况下，深度可分离卷积的参数个数是常规卷积的约 1/3。因此，选择深度可分离卷积可以有效降低模型体积，提升模型推理速度。

3　改进模型识别结果与分析

3.1　实验环境

所有模型在服务器上完成训练，服务器配置为 CPU：Intel（R）Xeon（R）Gold 6330，GPU：RTXA5000。模型训练环境为 PyTorch 1.11.0，Python 3.8（Ubuntu 20.04），Cuda 11.3。训练参数为 200 轮次，Batch-size（批处理）为 16，图像输入分辨率为 640 像素 ×640 像素，其他使用原始默认参数。总植物数据图像 1 892 幅，按 7∶2∶1 比例随机划分数据集，其中训练集 1 322 幅，验证集 380 幅，测试集 190 幅。

3.2　识别结果

采用改进模型 YOLO v5s－CBD 对测试集的 190 幅宁夏荒漠草原植物图像，共 17 类植物进行识别，部分植物识别结果如图 7 所示。

由图 7 可以看出，对多目标的植物图像、局部遮挡植物图像和光线变化的植物图像进行测试时，改进模型 YOLO v5s－CBD 可以准确予以识别。

3.3　植物数据集测试结果

为深入测试改进模型 YOLO v5s－CBD 的性能，选取精确率 P（Precision）、召回率 R（Recall）、平均精度均值（Mean average precision，mAP）和综合评价指标 F1 值作为模型的识别精度评价指标，选取模型

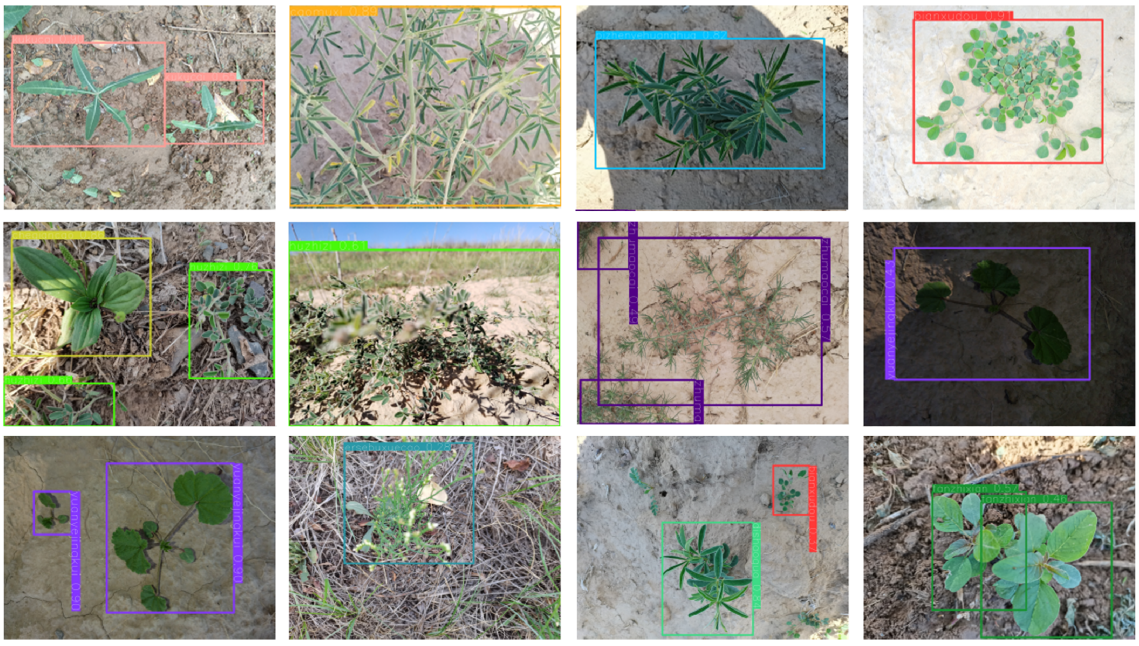


图 7 草原植物图像识别结果

Fig. 7 Image recognition results of grassland plant

体积和检测时间作为模型性能的评价指标[23-24]。同时考虑实际部署原因,分别使用 Nvidia GTX A5000 GPU 检测时间 T_1 和 Xeon(R)E-2136 CPU 检测时间 T_2 进行比较。

为有效评估改进模型 YOLO v5s-CBD 的性能,选取目标检测模型 YOLO v3-tiny[25]、YOLO v4-tiny[26]、YOLO v5s、YOLO v7-tiny[27]、YOLO v8n 与 YOLO v5s-CBD 进行比较,其中改进模型与其它模型 mAP 训练曲线变化如图 8 所示。

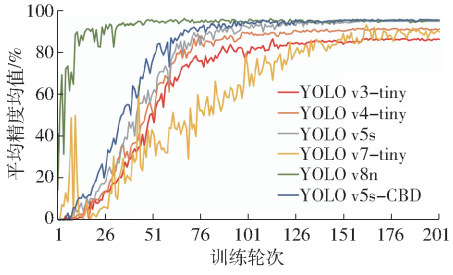


图 8 不同模型 mAP 变化曲线

Fig. 8 The mAP change curves of different models

在图 8 中,改进模型相比 YOLO v3-tiny、YOLO v4-tiny、YOLO v5s、YOLO v7-tiny,明显看出改进模型 YOLO v5s-CBD 平均精度均值最优,尽管相比于 YOLO v8n 模型,改进模型 YOLO v5s-CBD 收敛速度略差,但经过 300 轮次的充分训练之后,改进模型的收敛效果更好,且平均精度均值更优。

改进模型 YOLO v5s-CBD 与其它模型识别精度和性能实验结果如表 1 所示。

由表 1 可见,对植物识别时,YOLO v5s-CBD 相比 YOLO v5s,精确率提高 1.1 个百分点,召回率提高1.3个百分点,平均精度均值提高0.6个百分点,综合评价指标提升,模型内存占用量下降 38%,推理速度在 GPU 端提升 2 ms。在模型内存占用量下降的情况下,YOLO v5s-CBD 精度和性能仍然优于 YOLO v5s;与 YOLO v3-tiny 相比,YOLO v5s-CBD 精确率提高 10.3 个百分点,召回率提高 12.4 个百分点,平均精度均值提高 9.1 个百分点,综合评价指标提升,模型内存占用量下降 75%,推理速度在 GPU 端提升 7 ms;相比于 YOLO v4-tiny,YOLO v5s-CBD 精确率提高 5.9 个百分点,召回率提高 6.7 个百分点,平均精度均值提高 5.6 个百分点,综合评价指标提升,模型内存占用量下降 61%,推理速度在 GPU 端提升 5 ms;相比于 YOLO v7-tiny,YOLO v5s-CBD 精确率提高 4.9 个百分点,召回率提高 7.1 个百分点,平均精度均值提高 4.9 个百分点,综合评价指标提升,模型内存占用量下降 26%,推理速度提升;而相比于 YOLO v8n,尽管 YOLO v5s-CBD召回率下降 0.3 个百分点,模型内存占用

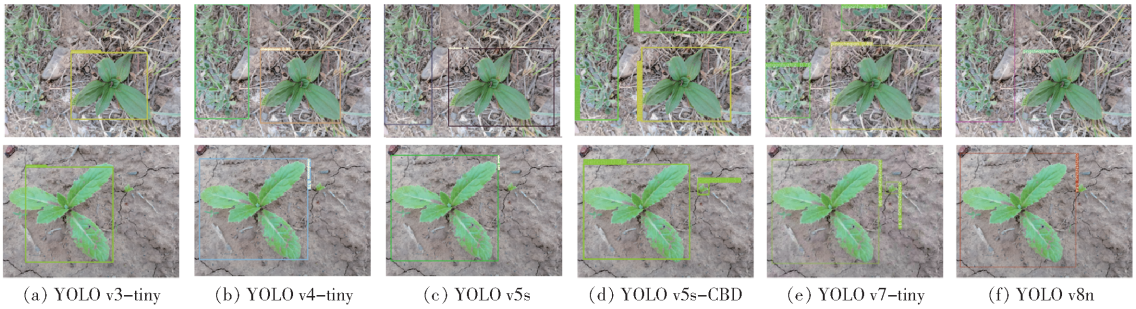
表 1 网络模型识别精度和性能比较

Tab. 1 Network model identification accuracy and performance comparison

网络模型	P/%	R/%	F1 值/%	mAP/%	模型内存占用量/MB	T_1/ms	T_2/ms
YOLO v3-tiny	84.8	80.5	82.5	86.6	36.6	15	316
YOLO v4-tiny	89.2	86.2	87.6	90.1	23.1	13	215
YOLO v5s	94.0	91.6	92.7	95.1	14.4	10	141
YOLO v7-tiny	90.2	85.8	87.0	90.8	12.1	12	138
YOLO v8n	92.6	93.2	93.0	95.6	6.1	14	133
YOLO v5s-CBD	95.1	92.9	94.0	95.7	8.9	8	125

量增加，但是改进模型在精确率、平均精度均值、综合评价指标和推理时间更优。综合来看，改进的模型 YOLO v5s－CBD 的性能均优于 YOLO v3－tiny、YOLO v4－tiny、YOLO v5s、YOLO v7－tiny、YOLO v8n。

为验证模型改进的有效性，选择测试集中两幅背景复杂的多目标图像，使用 YOLO v5s－CBD 与以上网络进行测试，测试结果如图 9 所示。



(a) YOLO v3-tiny　(b) YOLO v4-tiny　(c) YOLO v5s　(d) YOLO v5s-CBD　(e) YOLO v7-tiny　(f) YOLO v8n

图 9　不同模型识别对比

Fig. 9　Identification and comparison between different models

由图 9 可知，对于背景复杂的植物图像，YOLO v3－tiny、YOLO v4－tiny、YOLO v5s 和 YOLO v7－tiny 均存在漏检现象，YOLO v8n 虽能识别图像中的植物类别，但定位不准确。而改进模型 YOLO v5s－CBD 可以有效识别植物种类和定位目标。

为进一步验证改进模型 YOLO v5s－CBD 的性能，选择精确率、平均精度均值、推理时间、模型内存占用量，比较 YOLO v7 和改进模型，结果如表 2 所示。

表 2　网络模型性能

Tab. 2　Network model performance

模型	P/%	mAP/%	模型内存占用量/MB	T_1/ms
YOLO v5s－CBD	95.1	95.7	8.9	8
YOLO v7	93.0	94.4	74.9	7

由表 2 可知，与 YOLO v7 相比，YOLO v5s－CBD 精确率提高 2.1 个百分点，mAP 提高 1.3 个百分点，模型内存占用量更小，由于 YOLO v7 设计高效聚合网络和引入 RepVGG 结构，所以速度更快，但是 YOLO v5s－CBD 推理时间仅仅比 YOLO v7 慢 1 ms，产生的影响可以忽略不计，因此，相比于 YOLO v7，改进模型 YOLO v5s－CBD 拥有更好的检测性能。

3.4　公开数据集测试结果

为进一步客观评估本文所提模型的性能，选取公共数据集 PASCAL VOC07＋12（共包含 20 个目标）对 YOLO v5s－CBD 进行测试，并与 Faster R－CNN、SDD（Mobilenet）[28]、YOLO v3－tiny、YOLO v4－tiny[29]和 YOLO v7－tiny 进行比较，比较结果见表 3。

由表 3 可知，相比于双阶段目标检测模型 Faster R－CNN，YOLO v5s－CBD 模型平均精度均值提升 7.7 个百分点，推理速度提升明显。相比于 SSD 模型，YOLO v5s－CBD 模型 mAP 提升 5.6 个百分点，模型内存占用量降低 91%，推理速度提升 212.5%；相比于 YOLO v3－tiny，YOLO v5s－CBD 模型 mAP 提升 6.6 个百分点，尽管浮点运算次数增加，但是模型内存占用量降低 71%，推理速度提升 87.5%；相比于 YOLO v4－tiny，YOLO v5s－CBD 模型 mAP 提升 3.3 个百分点，模型内存占用量降低 62%，推理速度提升 62.5%；相比于 YOLO v7－tiny，YOLO v5s－CBD 模型 mAP 提升 2.5 个百分点，模型内存占用量降低 26%，推理速度提升 50%。这说明本文改进模型 YOLO v5s－CBD 在公共数据集 PASCAL VOC07＋12 上相比其他一些轻量级网络，有着更好的检测性能。

表 3　网络模型 VOC 数据集测试性能

Tab. 3　Network model VOC dataset performance

网络模型	模型内存占用量/MB	mAP/%	T_1/ms
Faster R－CNN		73.2	200
SSD	101.0	75.3	25
YOLO v3－tiny	36.6	74.3	15
YOLO v4－tiny	23.1	77.6	13
YOLO v7－tiny	12.1	78.4	12
YOLO v5s－CBD	8.9	80.9	8

3.5　消融实验

为验证改进模型 YOLO v5s－CBD 的有效性，需要通过消融实验对模型进行对比分析，在此使用精确率、召回率、综合评价指标和平均精度均值验证。

消融实验验证结果如表 4 所示。由表 4 可知，①相比 YOLO v5s，模型 2 的精确率提高 0.9 个百分点，召回率提高 0.5 个百分点，F1 值提高 0.5 个百分点，平均精度均值提高 0.4 个百分点，说明采用 BoTNet 网络，能够较好地提取目标特征，有效提高

表 4 网络模型消融实验

Tab. 4 Network model ablation experiment

网络模型	基线模型	BoTNet	CA	SIoU	DSC	P/%	R/%	F1 值/%	mAP/%
模型 1	YOLO v5s	×	×	×	×	94.0	91.6	92.7	95.1
模型 2	YOLO v5s	√	×	×	×	94.9	92.1	93.2	95.5
模型 3	YOLO v5s	×	×	√	×	94.2	93.5	93.8	95.3
模型 4	YOLO v5s	√	×	√	×	95.0	93.5	94.2	95.7
模型 5	YOLO v5s	√	√	√	×	96.4	90.5	93.3	95.9
模型 6	YOLO v5s	√	√	√	√	95.1	92.9	94.0	95.7

注:“√”表示包含此模块,“×”表示不包含此模块。

网络对植物的检测性能,降低模型参数量。②相比 YOLO v5s,模型 3 的精确率提高 0.2 个百分点,召回率提高 1.9 个百分点,F1 值提高 1.1 个百分点,mAP 提高 0.2 个百分点,说明使用 SIoU,可以提高模型的拟合,提高模型识别精度。③相比 YOLO v5s,模型 4 的精确率提高 1 个百分点,召回率提高 1.9 个百分点,F1 值提高 1.5 个百分点,mAP 提高 0.6 个百分点,说明 BoTNet 和 SIoU 有效提升网络模型性能。④与模型 YOLO v5s 相比,模型 5 各方面均有提升,性能达到最优,说明改进模型的有效性。⑤模型 6 经过模型压缩之后,与模型 5 相比,精确率下降 1.3 个百分点,召回率提高 2.4 个百分点,F1 值提高 0.7 个百分点,而 mAP 仅下降 0.2 个百分点,因此,模型压缩检测精度基本没有变化,而与 YOLO v5s 模型相比,模型精确率、召回率、F1 值和 mAP 均有提升,说明本文改进 YOLO v5s - CBD 在压缩模型之后,尽管模型参数量下降,但检测性能仍然优于原始模型。

为进一步分析改进前后模型的性能,训练 200 个迭代周期后,得到 YOLO v5s - CBD 和 YOLO v5s 的位置损失值如图 10 所示。

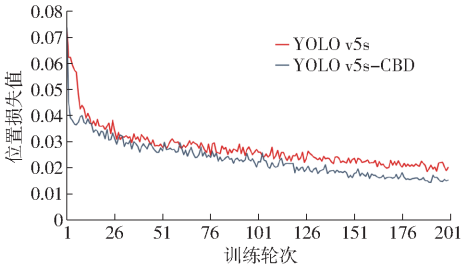


图 10 位置损失值曲线

Fig. 10 Curves of location loss

由图 10 可知,由于改进模型 YOLO v5s - CBD 引入了 BoTNet 和 CA 注意力机制结构,收敛更快且收敛性能优于 YOLO v5s。

4 应用

为方便调查宁夏全区荒漠草原植物种类及其分布,跟踪植物生长形态,基于改进模型 YOLO v5s - CBD,设计并开发了移动端植物图像识别 APP,同时将植物识别模型与定点生态信息观测平台结合[30-32],实现远程跟踪。

植物图像识别 APP 基于 uni - app 框架开发,使用 Android Studio 工具打包部署,为方便后期模型的升级,在此使用 Flask 前端框架将植物识别模型部署到阿里云服务器。

植物图像识别 APP 识别流程如图 11 所示。

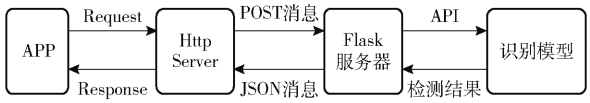


图 11 植物图像识别实现流程图

Fig. 11 Plant image recognition implementation flowchart

APP 主要实现功能为:①用户可通过 APP 拍照或者图库选取植物图像,上传至云服务器,云服务器调用 YOLO v5s - CBD 对图像进行自动识别后将结果反馈给客户端,实测在正常 4G 网络中只需 1 s 便可返回识别结果。返回结果之后,可以查询植物的详细信息,包括植物的草原类型、密度、氮磷钾、植物的干湿重、分布的经纬度等。②用户可以通过 APP 上传图像到数据库,扩充原始植物数据集,增加模型识别种类,方便科研工作者使用。

植物图像识别 APP 各模块的 UI 页面如图 12 所示。



(a) APP主页 (b) 识别结果 (c) 详细信息

图 12 植物图像识别 APP 界面

Fig. 12 Plant image recognition APP interface

定点生态信息观测平台以树莓派作为控制核

心，可完成光照强度、降雨量、空气温湿度、土壤氮磷钾主要元素等 10 余类生态环境数据的实时监测和跟踪，同时可对观测点植物图像数据进行实时采集和跟踪。定点生态信息观测平台实物如图 13 所示。



图 13　定点生态信息观测平台实物

Fig. 13　Fixed-point plant ecological information observation platform

为了对观测点的图像进行实时采集并查看，使用 UIkit 轻量级框架构建 web 界面，通过定点生态信息观测平台数据采集模块的高清图像采集设备，使用 RTMP 推流的方式，与萤石云开放平台结合，实时获取宁夏草原地区植物图像，同时利用部署在阿里云服务器的 YOLO v5 - CBD 模型，使用图像采集设备的 RSTP 地址，可完成观测点植物图像的跟踪，图像跟踪界面如图 14 所示。在图 14 中，利用 YOLO v5 - CBD 模型，实现了长芒草植物图像的远程实时跟踪识别。

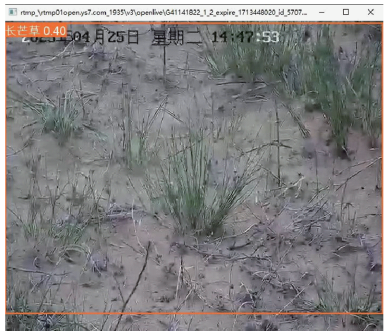


图 14　图像跟踪界面

Fig. 14　Image tracking interface

5　结论

（1）针对宁夏荒漠草原植物样本不一，背景复杂，采用 YOLO 系列检测算法参数量较大、识别精度不高等问题，提出了一种复杂背景下轻量化目标检测识别模型 YOLO v5s - CBD，该模型使用 BoTNet 和 CA 改进特征提取网络，引入 SIoU 提高模型收敛能力，使用深度可分离卷积压缩模型来进行轻量化。

（2）在对宁夏草原植物检测识别时，改进模型 YOLO v5s - CBD 识别精确率 P 为 95.1%，召回率 R 为 92.9%，F1 值为 94.0%，平均精度均值（mAP）为 95.7%，模型内存占用量为 8.9 MB。与 YOLO v3 - tiny、YOLO v4 - tiny、YOLO v5s 模型相比，改进模型 YOLO v5s - CBD 对宁夏荒漠草原常见植物具有更好的检测性能。

（3）结合实际应用，设计并开发了移动端植物识别 APP 和定点植物生态信息观测平台，可完成宁夏荒漠草原植物图像的移动识别和观测点的生态信息实时检测和长期跟踪。

参　考　文　献

[1] 卞莹莹，陈林，王建明，等. 平茬对荒漠草原区人工柠条林地土壤理化性质的影响[J]. 草地学报，2019，39(19)：7249 - 7257.
BIAN Yingying, CHEN Lin, WANG Jianming, et al. Effects of flatstubble on soil physicochemical properties in artificial lemon woodlands in desert grassland areas[J]. Journal of Grassland Science, 2019, 39(19): 7249 - 7257. (in Chinese)

[2] 马子元，钱志豪，马红彬，等. 宁夏荒漠草原 5 种乡土植物适应性评价[J]. 草业科学，2022，39(5)：1006 - 1014.
MA Ziyuan, QIAN Zhihao, MA Hongbin, et al. Adaptability evaluation of five native plants in Ningxia desert grassland[J]. Grassland Science, 2022, 39(5): 1006 - 1014. (in Chinese)

[3] YANG Z, HE W, FAN X, et al. PlantNet: transfer learning-based fine-grained network for high-through put plants recognition [J]. Soft Comput, 2022, 26: 10581 - 10590.

[4] WANG Z, LI H, ZHU Y, et al. Review of plant identification based on image processing[J]. Archives of Computational Methods in Engineering, 2017, 24: 637 - 654.

[5] GOYAL N, GUPTAK, KUMAR N. Multiclass twin support vector nachine for plant species identification[J]. Multimed Tools Applications, 2019, 78: 27785 - 27808.

[6] PATIL S, SASILKALA M. Segmentation and identification of medicinal plant through weighted KNN[J]. Multimed Tools Applications, 2023, 82: 2805 - 2819.

[7] GAJJARV K, NAMBISAN A K, KOSBAR K L. Plant identification in a combined-imbalanced leaf dataset[J]. IEEE Access, 2022, 10: 37882 - 37891.

[8] 孙俊，谭文军，毛罕平，等. 基于改进卷积神经网络的多种植物叶片病害识别[J]. 农业工程学报，2017，33(19)：209 - 215.
SUN Jun, TAN Wenjun, MAO Hanping, et al. Recognition of multiple plant leaf diseases based on improved convolutional

neural network[J]. Transactions of the CSAE,2017, 33(19): 209 - 215. (in Chinese)
[9] GUANG J, XI Z. ECAENet: EfficientNet with efficient channel attention for plant species recognition[J]. Journal of Intelligent & Fuzzy Systems, 2022, 43(4): 4023 - 4035.
[10] BARRADAS A,SILVA J M D,MARIANO P, et al. Automated real-time identification of medicinal plants species in natural environment using deep learning models—a case study from Borneo region[J]. Plants, 2022, 11(15): 1952.
[11] CHEN Z, WU R, LIN Y, et al. Plant disease recognition model based on improved YOLOv5[J]. Agronomy, 2022, 12(2): 365.
[12] 项和雨,邹斌,唐亮,等. 基于残差注意力网络模型的浮游植物识别[J]. 生态学报,2021,41(17):6883 - 6892.
XIANG Heyu, ZOU Bin, TANG Liang, et al. Phytoplankton identification based on residual attention network model[J]. Journal of Ecology, 2021, 41(17): 6883 - 6892. (in Chinese)
[13] 温长吉,娄月,张笑然,等. 基于改进稠密胶囊网络模型的植物识别方法[J]. 农业工程学报,2020,36(8):143 - 155.
WEN Changji, LOU Yue, ZHANG Xiaoran, et al. Plant recognition method based on a improved dense CapsNet[J]. Transactions of the CSAE,2020, 36(8):143 - 155. (inChinese)
[14] SRINIVAS A, LIN T, PARMAR N, et al. Bottleneck transformers for visual recognition[J]. arXiv Preprint, arXiv: 2101. 11605, 2021.
[15] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[J]. arXiv Preprint,arXiv: 2103. 02907, 2021.
[16] ZHAO Z, YANG X, ZHOU Y, et al. Real-time detection of particleboard surface defects based on improved YOLOV5 target detection[J]. Scientific Reports,2021(11): 21777.
[17] GEVORGYANZ. SIoU loss: more powerful learning for bounding box regression[J]. arXiv Preprint, arXiv:2205. 12740, 2022.
[18] ASHIISH V, NOAM S, NIKI P, et al. Attention is all you need[J]. arXiv Preprint, arXiv: 1706. 03762, 2011.
[19] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. arXiv Preprint, arXiv: 1512. 03385, 2015.
[20] 李文举,张干,崔柳,等. 基于坐标注意力的轻量级交通标志识别模型[J]. 计算机应用,2023,43(2):608 - 614.
LI Wenju, ZHANG Gan, CUI Liu, et al. A lightweight traffic sign recognition model based on coordinate attention[J]. Computer Applications,2023,43(2):608 - 614. (in Chinese)
[21] 陈加敏,徐杨. 基于注意力拆分卷积残差网络的表情识别[J]. 激光与光电子学进展,2022,59(18):306 - 313.
CHEN Jiamin, XU Yang. Expression recognition based on attention-splitting convolutional residual networks[J]. Advances in Laser and Optoelectronics,2022,59(18):306 - 313. (in Chinese)
[22] ZHENG Z,WANG P, LIU W, et al. Distance - IoU loss: faster and better learning for bounding box regression[J]. arXiv Preprint, arXiv: 1911. 08287, 2019.
[23] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345 - 1359.
[24] 何斌,张亦博,龚健林,等. 基于改进 YOLO v5 的夜间温室番茄果实快速识别[J]. 农业机械学报,2022,53(5):201 - 208.
HE Bin,ZHANG Yibo,GONG Jianlin,et al. Fast recognition of tomato fruit in greenhouse at night based on improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(5):201 - 208. (in Chinese)
[25] REDMON J, FARHADR A. YOLO v3: an incremental improvement [J]. arXiv Preprint, arXiv: 1504. 08083, 2015.
[26] BOCHKOVSHIY A, WANG C Y, LIAO H Y M. YOLO v4: optimal speed and accuracy of object detection[J]. arXiv Preprint, arXiv: 1504. 08083, 2015.
[27] WANG C, BOCHKOVSKIY A, LIAO H. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv Preprint, arXiv:2207. 02696,2022.
[28] XU X,ZHENG H,GUO Z, et al. SDD - CNN: small data-driven convolution neural networks for subtle roller defect inspection[J]. Applied Sciences, 2019, 9(7): 1364.
[29] 王宏志,宋明轩,程超,等. 基于改进 YOLOv4 - tiny 算法的车距预警方法[J/OL]. 吉林大学学报(工学版). https://doi.org/10.13229/j.cnki.jdxbgxb20220473.
WANG Hongzhi, SONG Mingxuan, CHENG Chao, et al. A vehicle distance warning method based on improved YOLOv4 - tiny algorithm [J/OL]. Journal of Jilin University (Engineering Edition). https://doi.org/10.13229/j.cnki.jdxbgxb20220473. (in Chinese)
[30] 李文涛,张岩,莫锦秋,等. 基于改进 YOLO v3 - tiny 的田间行人与农机障碍物检测[J]. 农业机械学报,2020,51(增刊1):1 - 8,33.
LI Wentao, ZHANG Yan, MO Jinqiu, et al. Detection of pedestrian and agricultural vehicles in field based on improved YOLO v3 - tiny[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(Supp.1): 1 - 8, 33. (in Chinese)
[31] 马宏兴,张淼,董凯兵,等. 基于改进 YOLO v5 的宁夏草原蝗虫识别模型研究[J]. 农业机械学报, 2022,53(11):270 - 279.
MA Hongxin, ZHANG Miao, DONG Kaibing, et al. Research of locust recognition in Ningxia grassland based on improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(11):270 - 279. (in Chinese)
[32] 蔡舒平,潘文浩,刘慧, 等. 基于 D2 - YOLO 去模糊识别网络的果园障碍物检测[J]. 农业机械学报,2023,54(2):284 - 292.
CAI Shuping,PAN Wenhao,LIU Hui, et al. Orchard obstacle detection based on D2 - YOLO deblurring recognition network[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(2):284 - 292. (in Chinese)