

基于 YOLO v5s 的作物叶片病害检测模型轻量化方法

杨佳昊¹ 左昊轩¹ 黄祺成² 孙 泉² 李思恩³ 李 莉¹

(1. 中国农业大学农业农村部农业信息获取技术重点实验室, 北京 100083;

2. 中国农业大学智慧农业系统集成研究教育部重点实验室, 北京 100083;

3. 中国农业大学水利与土木工程学院, 北京 100083)

摘要: 为保证识别性能前提下,对叶片病害检测模型进行有效轻量化,基于主干替换、模型剪枝以及知识蒸馏技术构建了一种模型轻量化方法,对以 YOLO v5s 为基础的叶片黄化曲叶病检测模型开展轻量化试验。首先,通过常见的性能优异的轻量级主干特征提取神经网络结构(Lightweight convolutional neural networks,LCNN)替换 YOLO v5s 主干对模型主体进行缩减;然后利用模型稀疏化训练和批归一化层(Batch normalization layer)的缩放因子分布状况,筛选并删减不重要的通道;最后,通过微调重新训练以及知识蒸馏,将模型精度调整到接近剪枝前的水平。试验结果表明,经轻量化处理的模型精确率、召回率和平均精度分别为 91.3%、87.4% 和 92.7%,模型内存占用量为 1.4 MB,台式机检测帧率 81.0 f/s,移动端检测帧率 1.2 f/s,相比原始 YOLO v5s 叶片病害检测模型,精确率、召回率和平均精度下降 3.7、4.6、2.7 个百分点,内存占用量仅为处理前的 10%,台式机和移动端检测的帧率分别提升近 27% 和 33%。本文所提出的方法在保持模型性能的前提下对模型有效轻量化,为移动端叶片病害检测部署提供了理论基础。

关键词: 病害检测; YOLO v5s; 轻量化模型; 网络剪枝; 知识蒸馏

中图分类号: TP18 **文献标识码:** A **文章编号:** 1000-1298(2023)S1-0222-08

Lightweight Method for Crop Leaf Disease Detection Model
Based on YOLO v5s

YANG Jiahao¹ ZUO Haoxuan¹ HUANG Qicheng² SUN Quan² LI Sien³ LI Li¹

(1. Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture and Rural Affairs,
China Agricultural University, Beijing 100083, China

2. Key Laboratory of Smart Agriculture System Integration, Ministry of Education, China Agricultural University, Beijing 100083, China

3. College of Water Resources and Civil Engineering, China Agricultural University, Beijing 100083, China)

Abstract: In order to effectively lightweight the leaf disease detection model under the premise of ensuring the recognition performance, a model lightweight method was constructed based on trunk replacement, model pruning and knowledge distillation technology, and a lightweight test was carried out on the leaf yellow leaf curl disease detection model based on YOLO v5s. Firstly, the main body of the model was reduced by replacing the YOLO v5s trunk with the common lightweight convolutional neural networks (LCNN) with excellent performance. Then, the unimportant channels were screened and deleted by using the sparse training of the model and the distribution of the scaling factors in the batch normalization layer. Finally, by fine-tuning retraining and knowledge distillation, the model accuracy was adjusted to a level close to that before pruning. The experimental results showed that the accuracy, recall and mean average accuracy of the lightweight model were 91.3%, 87.4% and 92.7%, respectively. The memory consumption of the model was 1.4 MB, and the detection frame rate of the desktop was 81.0 f/s. The detection frame rate of the mobile terminal was 1.2 f/s. Compared with the original YOLO v5s leaf disease detection model, the accuracy, recall and average accuracy were reduced by 3.7 percentage points, 4.6 percentage points and 2.7 percentage points, and the memory consumption was

收稿日期: 2023-06-28 修回日期: 2023-08-28

基金项目: 国家重点研发计划项目(2022YFD1900801)

作者简介: 杨佳昊(1999—),男,硕士生,主要从事智慧农业研究,E-mail: s20213081568@cau.edu.cn

通信作者: 李莉(1978—),女,副教授,博士生导师,主要从事农业信息获取与控制技术在设施园艺中的应用研究,E-mail: lily@cau.edu.cn

only 10% of that before processing. The frame rate of the desktop and mobile terminal detection was increased by nearly 27% and 33%, respectively. The proposed method can effectively reduce the weight of the model under the premise of keeping the performance, which provided a theoretical basis for the deployment of mobile leaf disease detection.

Key words: disease detection; YOLO v5s; lightweight model; network pruning; knowledge distillation

0 引言

YOLO 模型作为深度学习中单阶段的目标检测方法,由于其高效快捷的特性,在病害检测领域得到了广泛应用^[1-5]。在数据集采集方面,LIU 等^[6]收集制作了植物病害数据集 (Plant Disease Dataset 271, PDD271),其中包含 220 592 幅植物叶片图像,涵盖 271 种植物病害类别,SINGH 等^[7]以 13 种作物的 17 种病害建立了涵盖 2 598 个样本的 PlantDoc 数据库。

在模型改进方面,王卫星等^[8]在 YOLO v4 中使用 GhostNet、Ghost Module、新特征融合方法和注意力机制 CBAM 对 YOLO v4 - G 进行了改进,在尺寸上缩小 84%,检测速度提升 38%,平均精度提升 4.13%。何斌等^[9]通过改进 YOLO v5 实现对夜间环境下绿色、红色和综合番茄果实识别精度达 96.2%、97.6% 以及 96.8%,提高了被遮挡特征物与暗光下特征物的识别精度,改善了模型鲁棒性。王彦超等^[10]通过改进 YOLO v3 - tiny 的奶牛乳房炎自动检测方法,构建了自动检测奶牛关键部位模型,与传统模型相比,准确率提高 9.9 个百分点,平均检测精度提高 2.2 个百分点。

在应用方面,SHILL 等^[11]使用 YOLO v4 在 PlantDoc 数据集上对 13 种作物进行了检测,最优准确率达到 55.45% (IoU 为 50%)。WANG 等^[12]使用 MP - YOLO v3 在田间采集的番茄图像上进行检测,最优准确率达到 95.6%。ATILA 等^[13]在 PlantVillage 数据集上针对香蕉使用改进的 YOLO 进行了检测,最优准确率达到 98.4%。

但是深度学习的内存占用量较大,较难在移动设备、嵌入式设备上部署,模型的轻量化研究逐渐引起重视。黄小平等^[14]以 YOLO v5s 为基础,提出了一种改进的轻量级奶牛体况评分模型,改进的 YOLO v5s 模型的检测精度为 93.4%,计算量为 2.0×10^9 ,模型内存占用量仅为 2.28 MB。朱冬琴等^[15]基于剪枝对病害检测方法进行压缩,剪枝后的 Vgg16、ResNet164 和 DenseNet40 的平均准确率分别为 97.46%、99.12% 和 99.68%。李书琴等^[16]基于轻量级残差网络提出的植物叶片病害识别方法,在自建苹果叶片病害数据集上获得 1.52% 的低识别

错误率,参数量为 ResNet18 的 1/39,计算量约为其 1/10。方鹏等^[17]采用 Faster R - CNN 提出的模型经知识蒸馏后,性能达教师模型水平,参数量、计算量和模型推理时间减少 32%、33% 和 66%。

为在保证模型识别性能的前提下进行有效轻量化,减少计算资源占用,本文基于 YOLO v5s 提出一种模型轻量化方法,通过替换 YOLO v5s 的主干网络对模型进行大幅压缩,利用通道剪枝删除不重要的通道实现模型的进一步缩减,通过微调以及知识蒸馏将模型精度调整到接近剪枝前的水平。以期为后续作物叶片病害识别检测系统的移动端部署提供理论依据。

1 材料与方法

1.1 数据集建立

1.1.1 图像获取

数据集包括 200 幅 Plant Village 数据库的番茄黄化曲叶病图像和 445 幅在温室实地拍摄的番茄患病叶片图像。实地拍摄的图像于 2022 年 3—5 月在北京昌平区夏各庄镇谷新农业日光温室 E 区 4 号棚 (117°21'E, 40°14'N) 采集,番茄品种为普罗旺斯。使用小米 8 SE 智能手机后置相机作为图像采集设备,设置拍摄模式为自动,采用多个方位的俯视角度,距离叶片上方 20 ~ 30 cm 间,拍摄时尽量保证每个患病叶片的图像完整并且调整焦距使叶片清晰可见。数据集采集时还需考虑环境因素尤其是光线条件对图像的影响,分别在晴天、阴天、多云和雨天等 4 种不同天气状况,在同一天根据太阳入射角的变化情况,选择早晨、上午、中午和下午 4 个不同时间对不同番茄患黄化曲叶病毒病植株进行不同位置的拍摄采集,图像分辨率为 3 024 像素 × 4 032 像素,如图 1a 所示。

由于 Plant Village 数据集的样本图像均为裁剪的单叶图像,数据集较为简单,为了保证算法模型在实际复杂环境下的平均识别准确率,因此仅随机选取其中的 200 幅图像作为数据集的一部分。Plant Village 数据集图像如图 1b 所示。

1.1.2 数据预处理

在构建番茄黄化曲叶病毒病数据集的过程中,先通过数据增强手段对图像进行预处理,然后使用

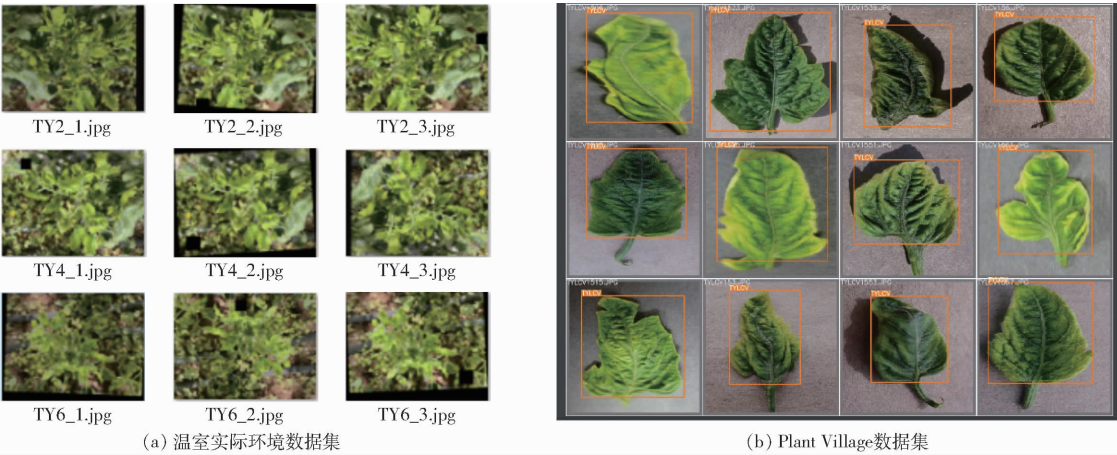


图 1 数据集图像
Fig. 1 Image of dataset

Labelimg 软件标注,生成可用于 YOLO v5s 训练的数据集。为保证模型的泛化性能和鲁棒性,所采用的数据增强包括随机添加噪声、裁剪、旋转等操作,扩充至 2 580 幅,以 VOC 格式制作数据集,按照 8:1:1 的比例划分为训练集、验证集和测试集。

1.2 轻量化网络设计

模型的轻量化必然导致精度降低,在保证模型的检测性能前提下进行轻量化是关键。与在每一步的轻量化处理中都追求较高的精度相比,在轻量化之后再行性能恢复更易于实现。因此首先对 YOLO v5s 进行网络轻量化设计来大幅压缩 YOLO v5s 模型的参数量、计算量、模型内存占用量,加快模型推理速度,后续再进行模型精度恢复。

YOLO v5s 网络结构包括输入端、主干特征提取神经网络 (Backbone)、颈部网络 (Neck) 和预测头 (Head),通常一个训练完毕的 YOLO v5s 目标检测模型的 Backbone 层所占用的计算资源和参数量最大,为有效缩减模型内存占用量,须对该部分进行改进。使用轻量化网络^[18]的方法能在保持网络模型性能的前提下减少参数量和计算量。

利用常见的性能优异的轻量级主干特征提取神经网络结构 (Lightweight convolutional neural networks, LCNN)^[19]对原始 YOLO v5s 模型的 Backbone 主干特征提取网络进行替换,共选取

MobileNetV2^[20]、MobileNetV3^[21]、ShuffleNetV2^[22]、GhostNet^[23]和 MobileNeXt^[24]5 种网络。

然后对各替换模型进行训练并根据训练结果选择合适的替换模型。训练过程中的参数设置为:输入图像尺寸 640 像素×640 像素,批处理大小 (batch-size)32,初始学习率 0.01,动量初始值 0.937,权重衰减系数 0.000 5, IoU 阈值 0.5,训练轮次 (epochs) 300。此外,还开启自适应锚框功能和 Mosaic 数据增强功能。

1.3 网络剪枝

经过轻量化网络设计后的模型,参数量、计算量和内存占用量得到大幅缩减,但是仍存在冗余,为进一步对模型轻量化,加快推理速度,继续采用结构化剪枝中的通道剪枝方法^[25]对模型继续压缩。

本文使用 YOLO v5s 中 BN 层的缩放因子作为剪枝方法中的通道缩放因子。在网络训练过程中,首先对 BN 层的缩放因子采用 L1 正则化进行稀疏训练,来自动识别 BN 层缩放因子分布接近于 0 的通道,也就是对于模型学习能力贡献较小的通道,然后根据稀疏训练后的结果,剪掉具有较小缩放因子的通道,获得剪枝后的紧凑模型,最后对其进行微调,优化其性能,通道剪枝过程如图 2 所示。

1.3.1 稀疏化训练

在通道剪枝的方法中,需要先对网络模型中的

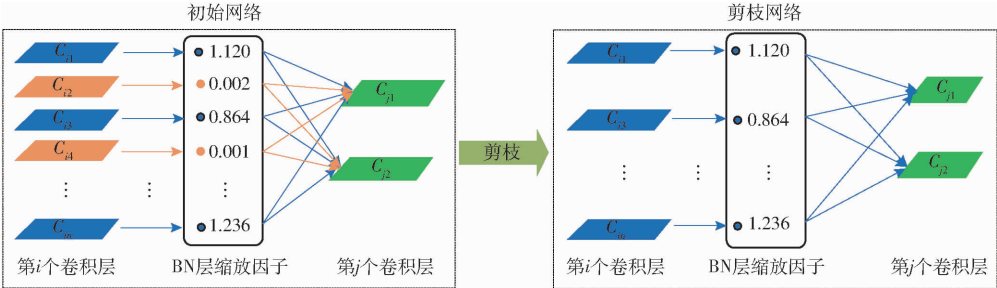


图 2 通道剪枝示意图
Fig. 2 Chanel pruning schematic

BN 层进行稀疏训练。在只进行正常训练的网络模型中,随着训练轮次增加,BN 层缩放因子基本满足均值为 1 的正态分布。在稀疏训练过程中,通过对 BN 层的缩放因子添加一个正则化 L1 约束改进正常训练的损失函数,将参数稀疏化,使得 BN 层的缩放因子的分布趋向于零,损失函数计算过程为

$$L = \sum_{(x,y)} l(f(x,W),y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma)$$

式中, (x,y) 表示训练的输入和目标, W 表示进行正常训练的权重,第 1 个求和项表示正常训练的一个损失项,第 2 个求和项表示对缩放因子进行 L1 正则化的稀疏处理, $g(\lambda) = |\lambda|$ 表示稀疏训练的一个惩罚项,其中 γ 为 BN 层的缩放因子, λ 表示正常训练与稀疏训练的平衡因子,也称作稀疏正则项系数。

稀疏过程需要考虑精度与稀疏度的平衡,主要通过调整稀疏正则化系数来进行调整。在训练过程中,稀疏正则项系数设置的越大会使 BN 层缩放因子的分布趋近于 0 的速度越快,但是模型平均识别准确率会变得很低;稀疏正则项系数设置越小,趋近于 0 的速度就会越慢,平均识别准确率也会随之收敛的更稳定。

1.3.2 通道剪枝

在稀疏训练后,按照 BN 层缩放因子分布的均值进行排序,通过删除卷积层中不重要的通道来达到模型剪枝的目的。在深度网络中,各层都是相乘相加的关系,如果某通道网络模型缩放因子趋近于 0,则该通道对网络模型的贡献程度也越小,因此通过删除分布趋近于 0 的通道可以实现网络模型的压缩。

但删除的通道越多,越有可能导致准确率下降,因此需要综合考虑模型准确度和内存占用量缩小的程度来进行剪枝。

1.3.3 模型微调

剪枝后的模型不可避免地出现了精度损失,需要重新训练微调,让模型更好地适应新的网络结构,恢复网络模型的性能,通过重新导入数据集文件以及相应的配置文件,训练轮次设置为 300 轮,加载剪枝后的权重文件进行微调重新训练。

1.4 知识蒸馏

为恢复在轻量化处理中模型检测精度的下降,通过知识蒸馏从一个识别性能较好但参数量和计算量较大的模型中提取知识,将其传递给一个较小的模型来实现模型压缩^[26]。知识蒸馏的原理如图 3 所示。

选择前述数据集作为知识蒸馏试验的检测输入,选择替换主干后的模型作为知识蒸馏试验的大

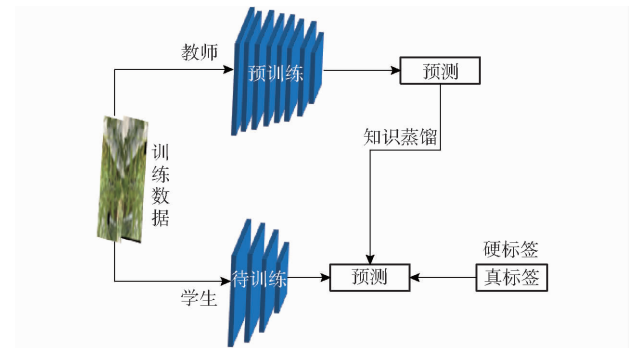


图 3 知识蒸馏原理图
Fig.3 Knowledge distillation

模型教师网络,微调后的剪枝模型作为小模型学生网络,通过知识蒸馏将大模型教师网络学习到的中间层软目标信息,传递给学生网络,使之具有更好的识别精度。在训练过程中采用与之前试验相同的训练策略,加载预训练权重文件,训练轮次设置为 500 轮。

2 试验与结果分析

2.1 试验环境及评价指标

2.1.1 试验环境配置

本文所有轻量化处理及训练均在台式计算机上进行,所用计算机配置为:Windows 10 操作系统, Intel Core i5 - 10400F CPU, Nvidia Geforce RTX 3070 GPU, 32 GB 内存。所用环境为 Pytorch 1.10.0, Cuda 11.3, cuDNN 8.2.0, Python 3.8。

2.1.2 试验评价指标

为综合表示轻量化处理对模型的影响,选取了 3 个模型识别性能指标、2 个计算性能指标以及模型内存占用量来评估模型。其中模型识别性能指标包括精确率 (Precision, P)、召回率 (Recall, R)、IoU 阈值为 0.5 时的平均精度均值 (Mean average precision, mAP); 计算性能指标包括参数量、浮点数计算量。

在知识蒸馏后,为进一步比较本文中轻量化处理对模型的影响,分别增加在与之之前配置相同的台式机和移动端(小米 12x 手机,骁龙 870 处理器)上检测的帧率作为计算性能指标。

2.2 轻量化设计对模型性能的影响

2.2.1 各替换模型训练过程

为选择适合检测任务的替换模型,评估各轻量化替换模型的性能,选取替换了主干的 YOLO v5s 模型进行比较。其中,将替换为 GhostNet 主干的 YOLO v5s 网络模型简称为 YOLO v5s - GhostNet,其他模型同理,包括 YOLO v5s - GhostNet、YOLO v5s - MobileNeXt、YOLO v5s - MobileNetV2、YOLO v5s -

MobileNetV3、YOLO v5s - ShuffleNetV2 共 5 个模型。各替换模型训练过程中平均精度均值变化如图 4 所示。

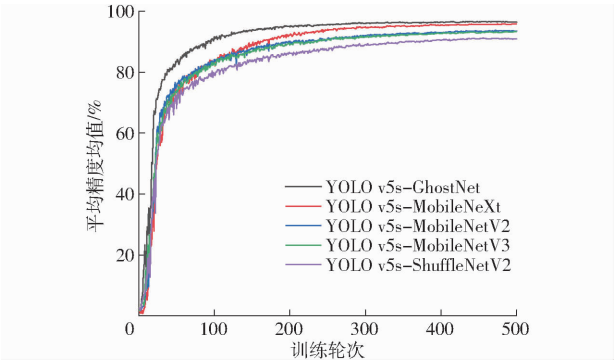


图 4 不同模型的 mAP 变化曲线
Fig.4 The mAP change curves of different models

表 1 各模型在测试集上的性能
Tab.1 Model performance on test set

模型名称	P/%	R/%	mAP/%	参数量	浮点数计算量	模型内存占用量/MB
YOLO v5s - GhostNet	96.3	93.0	96.5	3.90×10^6	8.90×10^9	7.80
YOLO v5s - MobileNetXt	95.3	92.4	95.8	4.48×10^6	8.40×10^9	9.20
YOLO v5s - ShuffleNetV2	89.5	84.9	91.0	8.50×10^5	1.90×10^9	2.00
YOLO v5s - MobileNetV3	92.3	88.1	93.2	1.38×10^6	2.30×10^9	3.10
YOLO v5s - MobileNetV2	93.7	86.7	93.6	7.60×10^5	2.20×10^9	1.90

YOLO v5s - GhostNet 和 YOLO v5s - MobileNetXt 模型尽管在识别性能指标上表现较好,但是在计算性能和模型内存占用量的缩减不足。

YOLO v5s - ShuffleNetV2 在浮点数计算量上缩减最多,但是所有指标都差于 YOLO v5s - MobileNetV2,其中精确率、召回率、平均精度均值分别降低 4.2、1.8、2.6 个百分点。

YOLO v5s - MobileNetV3 的识别性能指标接近 MobileNetV2,但是模型内存占用量为后者的 1.6 倍。本部分处理的目标是在牺牲一定精度的情况下对网络模型尽可能地轻量化,后续处理采用合理优化策略,也能够有效地恢复轻量化模型的识别精度。所以将在采用性能更好的 MobileNetV2 替换 YOLO v5s 主干网络基础上,继续探讨模型的轻量化以及综合性能的提升。

2.3 模型剪枝的效果

2.3.1 稀疏化训练对模型影响

为在模型进行稀疏化的同时保持较好的识别性能,本文采用不同的稀疏正则项系数进行了多次试验。设置不同稀疏正则项系数时,模型 BN 层的权重和平均精度均值均会有相应的变化,通过 Pytorch 框架中的可视化模块 tensorboard 进行数据查看,不同稀疏正则项系数选择下的 BN 层缩放因子分布变化情况如图 5 所示。

图 5a 显示了稀疏正则项系数为 0.000 1,训练

由图 4 可知,由于各替换模型在训练过程中均加载了预训练模型,试验中各模型的收敛速度基本一致,各模型训练平均精度均值也相差甚小,均能达到 90% 以上,其中 GhostNet 网络的平均精度最高,为 96.5%,ShuffleNetV2 的表现最差,为 91%。

2.2.2 轻量化模型在测试集上的性能对比

由表 1 可知,YOLO v5s - MobileNetV2 是轻量化程度最好的模型,计算量和模型内存占用量分别为 7.6×10^5 和 1.90 MB,浮点数计算量相比 MobileNetV3、MobileNetXt、GhostNet 分别缩减 1.0×10^8 、 6.20×10^9 、 6.70×10^9 ,仅比 ShuffleNet 少压缩 3.0×10^8 。精确率、召回率和平均精度均值略低于 GhostNet,分别降低 2.6、6.3、2.9 个百分点。

300 轮时 BN 层的权重变化。从图中可以看出,经过 300 轮的稀疏化训练后,BN 层缩放因子分布情况并无明显变化,并未找出缩放因子为 0 的通道层数,说明 0.000 1 过小。

从图 5b 中可以看出,BN 层缩放因子分布情况迅速趋向于 0,在稀疏化训练的初期就筛选出了大多数的 BN 层缩放因子分布情况为零的通道数,会影响模型精度,说明稀疏正则项系数 0.01 过大。

从图 5c 和图 5d 可以看出,经过 110 轮的稀疏化训练后,BN 层缩放因子分布情况出现逐渐趋近于 0 的趋势,但并未完全收敛,经过 300 轮稀疏化训练后,权重趋近于 0 变化已经达到了稳定,此时模型已基本满足剪枝试验需求。

图 6 显示了稀疏正则项系数为 0.01 与 0.005 时,稀疏训练时的平均精度变化情况,从图 6 中可以看出,相比稀疏正则项系数为 0.05 时,模型平均精度均值明显高于 0.01 时,最终结果相差 10.6 个百分点。

2.3.2 剪枝后模型的变化

通过试验结果得出在稀疏正则项系数 λ 为 0.005,训练轮次在 300 轮时,模型 BN 层缩放因子在零附近的大约有 2 500 个,通过不断地尝试剪枝率的设置,选择保留原模型 30% 通道的剪枝模型为最终模型,剪枝前后模型部分通道数如表 2 所示。

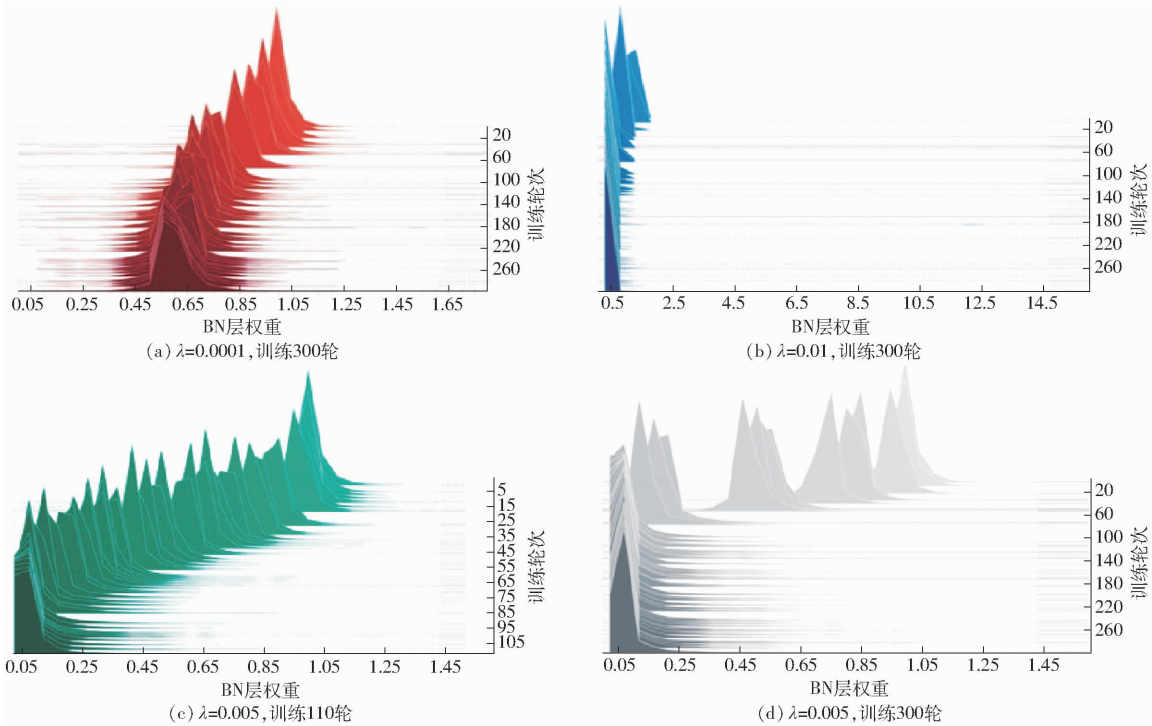


图 5 不同稀疏化训练后 BN 层缩放因子分布情况

Fig. 5 Distributions of BN layer scaling factors after different sparse trainings

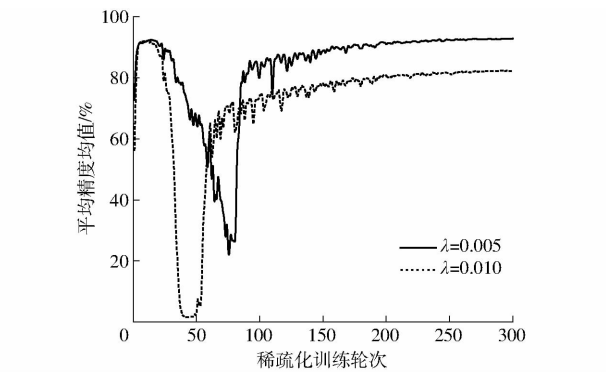


图 6 不同 λ 下 mAP 变化曲线

Fig. 6 The mAP change curves under different λ

表 2 模型部分通道数变化

Tab.2 Changes of partial channels in model		
网络层	剪枝前通道数	剪枝后通道数
model. 18. bn	80	3
model. 21. cv1. bn	40	25
model. 21. cv2. bn	40	4
model. 21. cv3. bn	80	21
model. 21. m. 0. cv1. bn	40	29
model. 21. m. 0. cv2. bn	40	27
model. 22. bn	40	16
model. 25. cv2. bn	20	6
model. 25. cv3. bn	40	28
model. 26. bn	40	22
model. 28. cv1. bn	40	14

2.3.3 微调后的模型精度变化

经过微调训练后,模型的精度相比剪枝前有轻微降低,将剪枝后的模型命名为 YOLO v5s -

MobileNetV2 - Pruned,剪枝前后的具体模型表现情况如图 7 所示。因为加载了预训练权重,故从图中可以看出,剪枝后的 YOLO v5s - MobileNetV2 - Pruned 在模型微调训练初期,收敛速度较剪枝前的 YOLO v5s - MobileNetV2 模型有一个微小的提高。经剪枝后的模型,最终在精度下降 4.7 个百分点的同时,模型内存占用量缩减 30%。

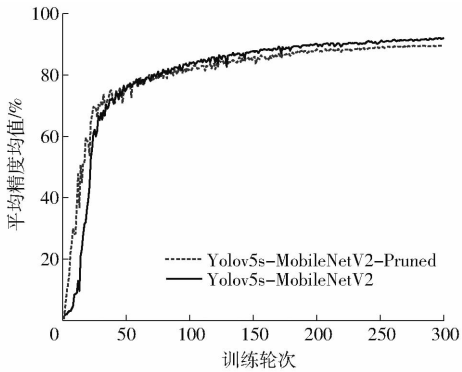


图 7 微调前后模型的平均精度均值对比曲线

Fig. 7 Comparison of model's mAP curves before and after finetuning

2.4 知识蒸馏效果

将进行知识蒸馏后模型命名为 YOLO v5s - MobileNetV2 - Pruned - Distill,试验结果如表 3 所示。

从表 3 可以看出,经知识蒸馏处理,模型在保持模型内存占用量和帧率不变的情况下,精确率、召回率和平均精度分别提高 3.6、4.7、4.7 个百分点。

表 3 模型性能
Tab.3 Performance of models

模型	<i>P</i> /%	<i>R</i> /%	mAP/%	模型内存占 用量/MB	台式机帧率/ (f·s ⁻¹)	移动端帧率/ (f·s ⁻¹)
YOLO v5s	95.0	92.0	95.4	14.0	64.0	0.9
YOLO v5s – MobileNetV2	93.7	86.7	93.6	1.9	78.0	1.1
YOLO v5s – MobileNetV2 – Pruned	87.7	82.7	88.0	1.4	81.0	1.2
YOLO v5s – MobileNetV2 – Pruned – Distill	91.3	87.4	92.7	1.4	81.0	1.2

YOLO v5s – MobileNetV2 – Pruned – Distill 相比 YOLO v5s – MobileNetV2,精确率和平均精度均值降低 2.4、0.9 个百分点,召回率提高 0.7 个百分点,模型内存占用量下降 0.5 MB,台式机帧率增加 3.0 f/s。综合来看,经本文轻量化处理后的用于叶片病害检测

的 YOLO v5s 模型内存占用量缩小 90%,台式机和移动端检测的帧率分别提升 27% 和 33%。

为验证轻量化处理后模型的有效性,选择数据集中背景复杂且形态方向不一的患病叶片图像进行测试,结果如图 8 所示。



图 8 模型验证结果

Fig.8 Results of model validation

由图 8 可知,在“正常-患病”叶片上的识别任务上,轻量化处理后的模型与原始模型的性能相近,均能将患病叶片全部识别出来(置信度在 50% 以上)。在带有卷曲特征(黄化曲叶病毒病的重要特征之一)正常叶片上,2 个模型均出现了轻微混淆(置信度在 10% 以下,在可接受的范围之内)也证明了与原模型的识别效果接近。

3 结束语

以番茄黄化曲叶病毒病为目标病害,研究了建立基于 YOLO v5s 的叶片病害识别轻量化模型

的方法。通过更换 YOLO v5s 的骨干网络,综合各项指标参数,选择了以 MobileNetV2 网络作为轻量化改进模型的主干,在此基础上,进行了网络模型通道剪枝,通过删除不重要的通道,进一步缩小了模型参数,在此过程中,模型精度有所下降,通过微调重新训练以及知识蒸馏的操作,将模型精度调整到接近剪枝前的水平,比原始 YOLO v5s 模型的内存占用量缩小 90%,台式机和移动端帧率提升 27% 和 33%,为后续作物叶片病害识别检测系统的移动端部署提供了理论依据。

参 考 文 献

[1] 张领先, 韩梦瑶, 丁俊琦, 等. 作物病害智能诊断与处方推荐技术研究进展[J]. 农业机械学报, 2023, 54(6): 1 – 18. ZHANG Lingxian, HAN Mengyao, DING Junqi, et al. Research progress in intelligent diagnosis and prescription recommendation of crop diseases[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(6): 1 – 18. (in Chinese)

[2] 宋玲, 曹勉, 胡小春, 等. 基于 YOLOX 的复杂背景下木薯叶病害检测方法[J]. 农业机械学报, 2023, 54(3): 301 – 307. SONG Ling, CAO Mian, HU Xiaochun, et al. Detection of cassava leaf diseases under complicated background based on YOLOX[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(3): 301 – 307. (in Chinese)

[3] XU Y, CHEN Q, KONG S, et al. Real-time object detection method of melon leaf diseases under complex background in greenhouse[J]. Journal of Real-Time Image Process, 2022, 19(5): 985 – 995.

[4] SHOAIB M, SHAH B, EL-SAPPAGH S, et al. An advanced deep learning models-based plant disease detection: a review of recent research[J]. Frontiers in Plant Science, 2023, 14: 1158933.

[5] 邵明月, 张建华, 冯全, 等. 深度学习在植物叶部病害检测与识别的研究进展[J]. 智慧农业(中英文), 2022, 4(1):

29 – 46.

SHAO Mingyue, ZHANG Jianhua, FENG Quan, et al. Research progress of deep learning in detection and recognition of plant leaf diseases[J]. *Smart Agriculture*, 2022, 4(1): 29 – 46. (in Chinese)

[6] LIU X, MIN W, MEI S, et al. Plant disease recognition: a large-scale benchmark dataset and a visual region and loss reweighting approach[J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2021, 30: 2003 – 2015.

[7] SINGH D, JAIN N, JAIN P, et al. PlantDoc: a dataset for visual plant disease detection[C] // *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, 2020: 249 – 253.

[8] 王卫星, 刘泽乾, 高鹏, 等. 基于改进 YOLO v4 的荔枝病虫害检测模型[J]. *农业机械学报*, 2023, 54(5): 227 – 235.

WANG Weixing, LIU Zeqian, GAO Peng, et al. Detection of litchi diseases and insect pests based on improved YOLO v4 model[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(5): 227 – 235. (in Chinese)

[9] 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)

[10] 王彦超, 康熙, 李孟飞, 等. 基于改进 YOLO v3 – tiny 的奶牛乳房炎自动检测方法[J]. *农业机械学报*, 2021, 52(增刊): 276 – 283.

WANG Yanchao, KANG Xi, LI Mengfei, et al. Automatic detection method for dairy cow mastitis based on improved YOLO v3 tiny[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(Supp.): 276 – 283. (in Chinese)

[11] SHILL A, RAHMANMD A. Plant disease detection based on YOLO v3 and YOLO v4[C] // *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*, 2021: 1 – 6.

[12] WANG X, LIU J. Multiscale parallel algorithm for early detection of tomato gray mold in a complex natural environment[J]. *Frontiers in Plant Science*, 2021, 12: 620273.

[13] ATILA Ü, UÇAR M, AKYOL K, et al. Plant leaf disease classification using EfficientNet deep learning model[J]. *Ecological Informatics*, 2021, 61: 101182.

[14] 黄小平, 冯涛, 郭阳阳, 等. 基于改进 YOLO v5s 的轻量级奶牛体况评分方法[J]. *农业机械学报*, 2023, 54(6): 287 – 296.

HUANG Xiaoping, FENG Tao, GUO Yangyang, et al. Lightweight dairy cow body condition scoring method based on improved YOLO v5s[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(6): 287 – 296. (in Chinese)

[15] 朱冬琴, 冯全, 张建华. 基于剪枝的植物病害识别方法[J]. *浙江农业学报*, 2023, 35(6): 1462 – 1472.

ZHU Dongqin, FENG Quan, ZHANG Jianhua. Plant disease identification based on pruning[J]. *Acta Agriculturae Zhejiangensis*, 2023, 35(6): 1462 – 1472. (in Chinese)

[16] 李书琴, 陈聪, 朱彤, 等. 基于轻量级残差网络的植物叶片病害识别[J]. *农业机械学报*, 2022, 53(3): 243 – 250.

LI Shuqin, CHEN Cong, ZHU Tong, et al. Plant leaf disease identification based on lightweight residual network[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(3): 243 – 250. (in Chinese)

[17] 方鹏, 郝宏运, 王红英. 基于知识蒸馏的叠层笼养蛋鸡行为识别模型研究[J]. *农业机械学报*, 2021, 52(10): 300 – 306.

FANG Peng, HAO Hongyun, WANG Hongying. Behavior recognition model of stacked-cage layers based on knowledge distillation[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(10): 300 – 306. (in Chinese)

[18] 罗志聪, 李鹏博, 宋飞宇, 等. 嵌入式设备的轻量化百香果检测模型[J]. *农业机械学报*, 2022, 53(11): 262 – 269, 322.

LUO Zhicong, LI Pengbo, SONG Feiyu, et al. Lightweight passion fruit detection model based on embedded device[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(11): 262 – 269, 322. (in Chinese)

[19] 王志强, 于雪莹, 杨晓婧, 等. 基于 WGAN 和 MCA – MobileNet 的番茄叶片病害识别[J]. *农业机械学报*, 2023, 54(5): 244 – 252.

WANG Zhiqiang, YU Xueying, YANG Xiaojing, et al. Tomato leaf diseases recognition based on WGAN and MCA – MobileNet[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(5): 244 – 252. (in Chinese)

[20] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks[C] // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 4510 – 4520.

[21] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C] // *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019: 1314 – 1324.

[22] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C] // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 6848 – 6856.

[23] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations[C] // *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 1577 – 1586.

[24] ZHOU D, HOU Q, CHEN Y, et al. Rethinking bottleneck structure for efficient mobile network design[C] // *Computer Vision-ECCV 2020*. Cham: Springer International Publishing, 2020: 680 – 697.

[25] LIU Z, LI J, SHEN Z, et al. Learning efficient convolutional networks through network slimming[C] // *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, 2017: 2755 – 2763.

[26] 黄海松, 陈星燃, 韩正功, 等. 基于多尺度注意力机制和知识蒸馏的茶叶嫩芽分级方法[J]. *农业机械学报*, 2022, 53(9): 399 – 407, 458.

HUANG Haisong, CHEN Xingran, HAN Zhenggong, et al. Tea buds grading method based on multiscale attention mechanism and knowledge distillation[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(9): 399 – 407, 458. (in Chinese)