doi:10.6041/j.issn.1000-1298.2024.11.019

基于 PN－YOLO v8s－Pruned 的轻量化三七收获目标检测方法

王法安[1,2]　何忠平[1,2]　张兆国[1,2]　解开婷[2,3]　曾　悦[1,2]

（1.昆明理工大学现代农业工程学院，昆明 650500；2.云南省高校中药材机械化工程研究中心，昆明 650500；
3.昆明理工大学机电工程学院，昆明 650500）

摘要：为实现三七联合收获作业过程中的自适应分级输送和收获状态实时监测，本文针对三七根土复合体特征和复杂田间收获工况，提出一种基于 YOLO v8s 并适用于 Jetson Nano 端部署的三七目标检测方法。在 YOLO v8s 对三七准确识别的基础上，针对其新的模型结构特性，利用通道剪枝算法，制定相应剪枝策略，保证模型精度的同时提升实时检测性能。采用 TensorRT 推理加速框架将改进模型部署至 Jetson Nano，实现了三七目标检测模型的灵活部署。试验结果表明，改进后的 PN－YOLO v8s－Pruned 模型在主机端的平均精度均值为 93.71%，参数量、计算量、模型内存占用量分别为原始模型的 39.75%、57.69%、40.25%，检测速度提升 44.26%，与其他目标检测模型相比，本文改进模型在计算复杂度、检测精度和实时性方面具有更好的综合检测性能。在 Jetson Nano 端部署后，改进模型检测速度达 18.9 f/s，较加速前提升 2.7 倍，较原始模型提升 5.8 f/s。台架试验结果表明，4 种输送分离收获作业工况下三七目标检测的平均精度均值达 87% 以上，不同输送分离收获作业工况和不同流量等级下的目标三七计数平均正确率分别达 92.61%、91.76%。田间试验结果表明，三七目标检测平均精度均值达 84%，计数平均正确率达 88.11%，图像推理速度达 31.0 f/s。模型检测性能和计数效果能够满足复杂田间收获工况下目标三七的检测需求，可为基于边缘计算设备的三七联合收获作业自适应分级输送系统和收获作业质量监测系统提供技术支撑。

关键词：三七；复杂收获作业工况；目标检测；通道剪枝；Jetson Nano；YOLO v8s

中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1000-1298(2024)11-0171-13　　**OSID**：

Lightweight Object Detection Method for *Panax notoginseng* Based on PN－YOLO v8s－Pruned

WANG Faan[1,2]　HE Zhongping[1,2]　ZHANG Zhaoguo[1,2]　XIE Kaiting[2,3]　ZENG Yue[1,2]

(1. *Faculty of Modern Agricultural Engineering*, *Kunming University of Science and Technology*, *Kunming* 650500, *China*
2. *Research Center on Mechanization Engineering of Chinese Medicinal Materials in University of Yunnan Province*, *Kunming* 650500, *China*
3. *Faculty of Mechanical and Electrical Engineering*, *Kunming University of Science and Technology*, *Kunming* 650500, *China*)

Abstract: In order to realize the adaptive grading conveyance and real-time monitoring of harvesting status in the process of *Panax notoginseng* combined harvesting operation, focusing on the characteristics of *Panax notoginseng* root-soil complex and the complex field harvesting conditions, a *Panax notoginseng* object detection method based on YOLO v8s and suitable for deployment on the Jetson Nano was proposed. Based on the accurate recognition of *Panax notoginseng* by YOLO v8s, the channel pruning algorithm was utilized to formulate a corresponding pruning strategies for its new model structural characteristics, which ensured the accuracy and improved the real-time detection performance at the same time. The improved model was deployed to Jetson Nano by using the TensorRT inference acceleration framework, which realized the flexible deployment of the *Panax notoginseng* object detection model. The experimental results showed that the mean average precision of the improved PN－YOLO v8s－Pruned model on the host side was 93.71%, although it was decreased by 0.94 percentage points compared with

收稿日期：2024－05－30　修回日期：2024－08－06

基金项目：国家重点研发计划项目(2022YFD2002004)和云南省教育厅科学研究基础项目(2023J0151)

作者简介：王法安(1990—)，男，讲师，博士，主要从事智能农业装备设计制造研究，E-mail：wfa@kust.edu.cn

通信作者：张兆国(1966—)，男，教授，博士生导师，主要从事农业机械装备设计制造研究，E-mail：zzg@kust.edu.cn

that of the original model, the number of parameters, computational complexity, and model size were 39.75%, 57.69%, and 40.25% of the original model, respectively, and the detection speed was increased by 44.26%. Compared with other models, the improved model demonstrated superior comprehensive detection performance in terms of computational complexity, detection accuracy, and real-time performance. After deployment at the Jetson Nano, the improved model had a detection speed of 18.9 frames per second, which was 2.7 times higher than before acceleration and 5.8 frames per second higher than the original model, and the deployment detection effect was better than the original model. The results of the bench tests showed that the mean average precision of *Panax notoginseng* detection was more than 87% under four conveyor separation harvesting conditions. The average accuracy of the *Panax notoginseng* counting under different conveyor separation harvesting conditions and different flow levels reached 92.61% and 91.76%, respectively. The field test results showed that the mean average precision of *Panax notoginseng* detection was more than 84%, and the average accuracy of the *Panax notoginseng* counting reached 88.11%, which could meet the detection requirements of *Panax notoginseng* under complex field harvesting conditions, and could provide technical support for the monitoring system of harvesting quality and the adaptive grading transportation system of combined harvesting operation based on edge computing equipments.

Key words: *Panax notoginseng*; complex harvesting conditions; object detection; channel pruning; Jetson Nano; YOLO v8s

0 引言

三七作为云南省具有较高经济和药用价值的根茎类中药材,种植产业逐渐由小地块向规模化、集约化发展[1-2]。当前三七机械化收获环节存在的实时测产难、收获作业质量不明、输送分离效率低、智能化水平差等问题已经成为限制三七产业发展的关键因素[3]。如何在复杂田间收获工况下,实现三七的精准识别和快速检测,是解决这一问题的关键。

近年来,国内外学者利用深度学习算法对根茎类作物识别与分类开展了大量研究[4-7]。薯类、萝卜等传统根茎类作物目标检测主要集中在产品初加工阶段,具有特征明显、轮廓简单、容易识别等特点。然而,收获过程中的三七不但田间收获工况复杂,而且三七须根多且长、易缠绕、形态多样、易与土壤形成复杂的根土复合体,给识别与检测带来了困难。上述模型虽然在检测精度方面取得了优异效果,但同时模型复杂度过高导致检测实时性差,限制了模型的部署与应用。寻找精度高、检测速度快、适用于边缘计算设备部署的田间收获三七目标检测方法具有重要的实际工程价值。

为实现目标检测模型在边缘计算设备的部署,相关学者针对模型轻量化改进做了许多研究[8-11]。不同于其它领域目标检测中的静态、稳态识别场景,田间收获三七存在的根土复合体特征和收获工况复杂问题给三七目标检测的精度和速度均提出了更高要求,且目前尚无田间收获三七目标检测的实际工程应用,对三七收获作业质量监测和自适应分级输送的研究还处于起步阶段。如何平衡三七目标检测模型的精度和速度,并将其部署到边缘计算设备上高效工作,是亟需解决的关键问题。为此,本研究从三七机械化收获过程中的工程应用需求出发,提出一种基于 YOLO v8s 并适用于边缘计算设备部署的轻量化三七目标检测方法,采用通道剪枝算法对模型轻量化处理,保证精度的同时提升实时检测性能,并使用 TensorRT 推理加速框架完成三七目标检测模型的灵活部署与应用,拟为后续部署于边缘计算设备的三七联合收获智能化采收技术奠定基础。

1 材料与方法

1.1 总体技术路线

本文总体技术路线如图 1 所示。首先,根据田间复杂收获工况、三七根土复合体特征和输送分离作业参数建立田间收获三七样本数据集;其次,在 YOLO v8s 目标检测模型对田间收获三七准确、快速识别的基础上,利用通道剪枝算法压缩模型,去除冗余参数,保证模型精度的同时提升实时检测性能,从而获得精度高、速度快、体积小、鲁棒性强的三七目标检测模型。最后,采用 TensorRT 推理加速框架将改进模型部署至边缘计算设备,并开展相关台架试验,对田间收获三七进行实时检测,验证模型部署的有效性。

1.2 图像采集与数据集构建

1.2.1 试验数据采集

由于田间收获三七根土复合体特征复杂,黏附土壤是影响三七识别的重要因素之一。为增加样本的丰富性,于昆明市石林县、昆明市西山区、文山州砚山县 3 个不同的三七主要种植区采集样本。图像

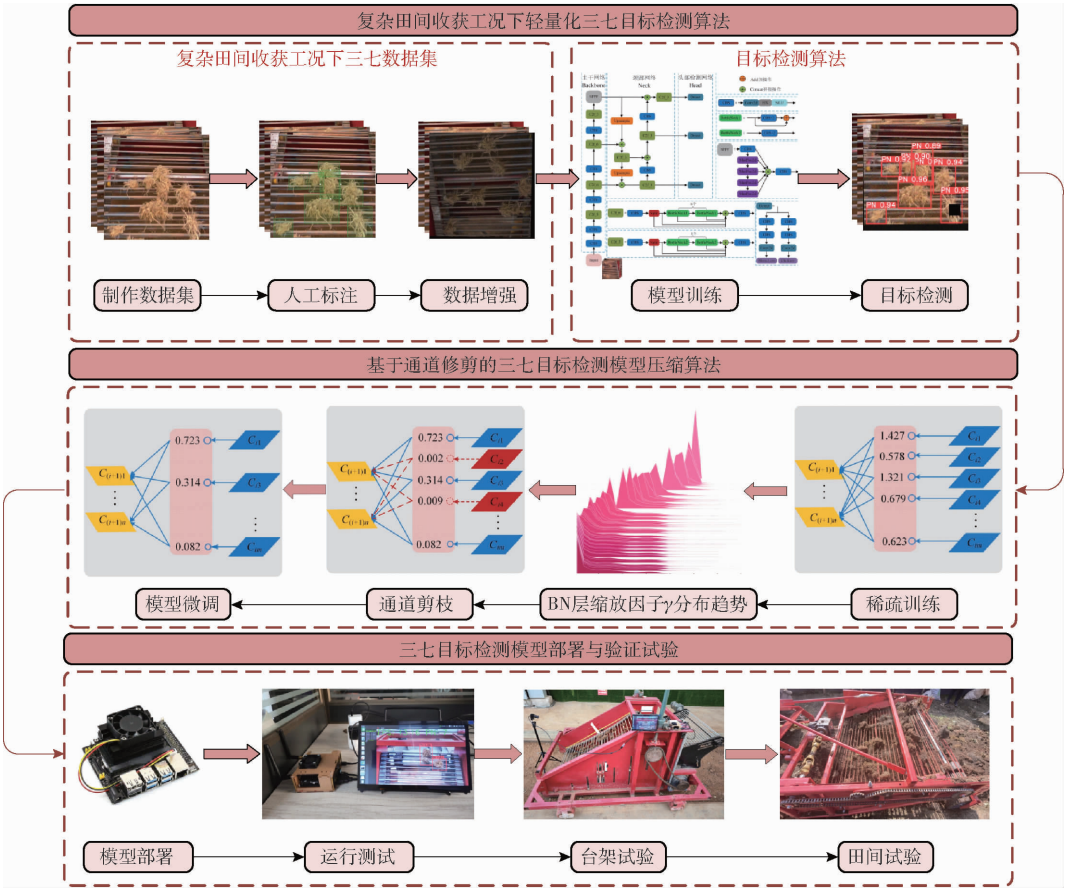

图 1　总体技术路线图

Fig. 1　Overall technology roadmap

采集地点为云南省高校中药材机械化工程研究中心(24°50′N,102°51′E)。采集日期为 2023 年 10 月 12—15 日、2024 年 1 月 26 日,拍摄时间为 09:00—12:00、15:00—18:00。采集设备为佳能 EOS M50 型相机和 IPHONE12 手机,分辨率分别为 1 920 像素 × 1 080 像素和 4 032 像素 × 3 024 像素。出土后的三七经一、二级输送分离装置后可去除大部分附着土壤,为模拟实际收获作业环境中存在的不同光照强度和不同输送分离状态等复杂场景,在输送分离试验台的一级升运链上于晴天和阴天不同时间段用相机拍摄不同视角、不同距离、不同须根缠绕程度、不同堆叠遮挡程度的三七输送分离视频,包含单个或多个目标,采样速率为 30 f/s,每次拍摄时长约40 s。利用 OpenCV 对采集到的三七输送分离视频分帧截图,并用手机拍摄静态环境下的三七图像,共采集 4 886 幅原始图像样本,图 2 为部分图像样本。



图 2　不同复杂场景下的三七图像

Fig. 2　*Panax notoginseng* images in different complex scenes

1.2.2　数据集制作

为保证田间收获三七数据集质量,去除误拍、相似度过高和高度模糊图像,通过尺寸调整、缩放等预处理方式最终获得 1 232 幅有效三七图像。采用 Make Sence 在线标注工具手动标出目标的边界框坐标,设定三七标签为 PN,共标注 4 004 个三七目标。为保证数据集图像的丰富性、提升模型泛化能力,对完成标注的图像采用改变亮度、添加噪声、镜像、平移、旋转 5 种带标签的图像增强方式进行数据增强,增强后的数据集图像达 7 392 幅。将数据集按照

80% 和 20% 的比例划分为训练集和验证集,其中训练集 5 914 幅,验证集 1 478 幅。

对训练集的标注文件进行可视化分析,如图 3 所示,标签的宽和高大多分布在样本尺寸的 1/10 至 1/4,小目标物体偏多,不易被模型检测出来。

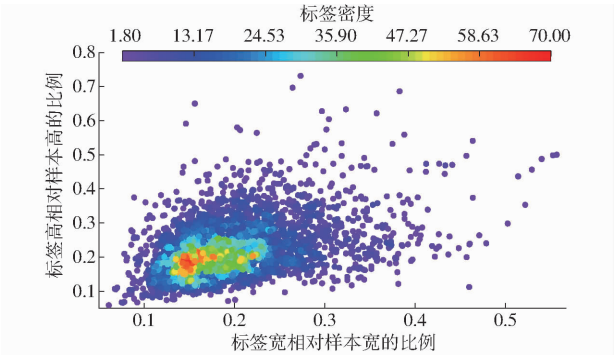

图 3 数据集标签分布情况

Fig. 3 Distribution of dataset labels

1.3 三七目标检测模型优化与部署

三七目标检测模型优化与部署技术路线如图 4 所示。首先,采用 YOLO v8s 目标检测模型对田间收获三七数据集进行训练;其次,对训练得到的网络进行稀疏训练、通道修剪和微调,以压缩模型,提高检测速度,实现实时、准确、高效的三七目标识别检测;最后,将优化后的模型部署至边缘计算设备,实现模型的部署与应用。

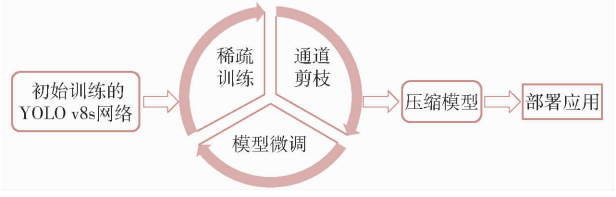

图 4 模型优化部署技术路线图

Fig. 4 Technical route for model optimization and deployment

1.3.1 基于 YOLO v8s 的三七目标检测模型

YOLO v8 模型继承了 YOLO 系列的优点,并引入新的功能和改进,进一步提升了模型性能。YOLO v8 模型结构与 YOLO v5[12] 大致相同,主要创新点在于新的骨干网络和头部网络。在主干网络和颈部网络中将 YOLO v5 的 C3 模块替换成 C2f 模块,在头部网络中将原本的耦合头改为解耦头结构,将分类和检测过程解耦[13],并引入了 Ancher-Free 无锚框检测头和分布焦点损失函数[14](Distribution focal loss,DFL)。在 YOLO v8 的 5 个版本中,YOLO v8s 在检测精度和速度方面有更好的综合性能,因此,为保证三七目标检测模型的综合检测性能和边缘计算设备的灵活部署,本文选择使用 YOLO v8s。

YOLO v8s 网络结构主要由输入端(Input)、主干网络(Backbone)、颈部网络(Neck)和头部网络(Head)4 部分构成[15],如图 5 所示。输入端将经过 Mosaic 数据增强的收获三七图像输入网络。主干网络和颈部网络参考了 YOLO v7 的 ELAN 设计思想[16],将 YOLO v5 的 C3 结构换成了梯度流更丰富的 C2f 结构,提升了网络的特征提取能力,并去除了上采样之前的 1×1 卷积,同时针对不同尺度的模型调整不同的通道数,大幅提升了模型检测性能[17]。头部网络由原先的 Couple-Head 耦合头换为目前主流的 Decoupled-Head 解耦头结构,将分类和回归分支解耦,并舍弃 object 对象分支,模型也从 Anchor-Base 的预测方式变为 Ancher-Free 方法,用更加精简的方式来确定正负样本,提升了模型检测速度[18]。

1.3.2 模型压缩

本文建立的 YOLO v8s 模型虽能够实现田间收获三七的高效检测,但仍有大量的冗余参数,落地部署难,实时性有待提升。考虑到三七田间实际收获工况和边缘计算设备资源受限,本文在 YOLO v8s 检测模型的基础上,采用结构化剪枝中的通道剪枝方法[19],对模型进行压缩。通道剪枝原理如图 6 所示,主要包括稀疏训练、通道剪枝、模型微调 3 个过程[20],图中 $C_{(i+1)1}$、$C_{(i+1)n}$ 分别表示第 $i+1$ 个卷积层中的第 1 个卷积和第 n 个卷积。

为实现模型在稀疏化训练过程中的快速收敛,本文在 YOLO v8s 网络的批量归一化层(Batch normalization, BN)加入超参数 γ 和 β,采用缩放与平移两种方法对通道数据进行归一化处理[21],计算式为

$$Z_{out} = \gamma \hat{Z} + \beta \tag{1}$$

其中

$$\hat{Z} = \frac{Z_{in} - \mu_B}{\sqrt{\sigma_B^2 + \delta}} \tag{2}$$

$$\mu_B = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3}$$

$$\sigma_B = \frac{\sum_{i=1}^{n} (x_i - \mu_B)^2}{n} \tag{4}$$

式中 Z_{in}、Z_{out}——BN 层的输入和输出

$\hat{Z}$——BN 层输入的批归一化值

σ_B、μ_B——每个批样本数据的标准差和均值

n——批样本个数

x_i——批样本中的第 i 个输入值

δ——微小正数

取超参数 γ 为缩放因子,由式(1)可知,BN 层的输出与 γ 成正相关,γ 愈小,输出愈接近 0,说明位

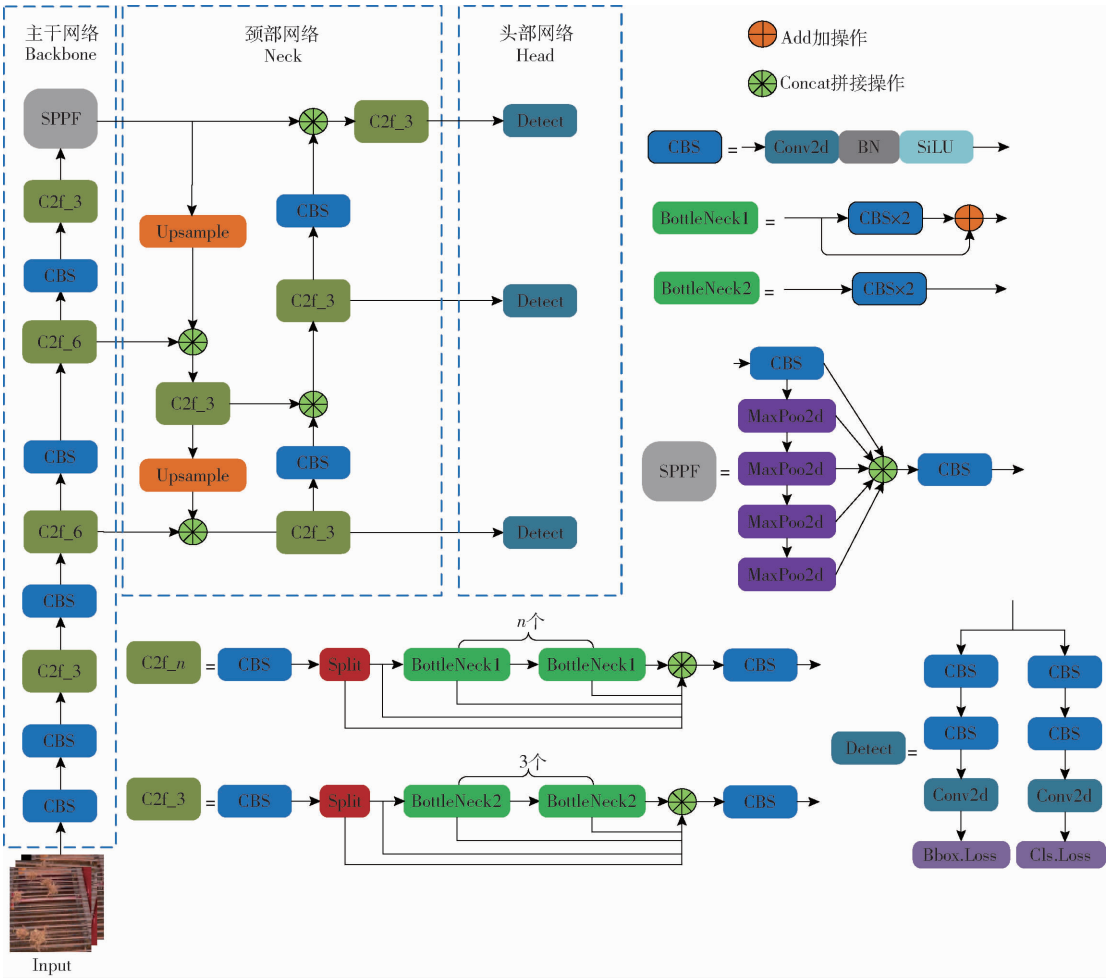

图 5　YOLO v8s 模型结构

Fig. 5　YOLO v8s model structure

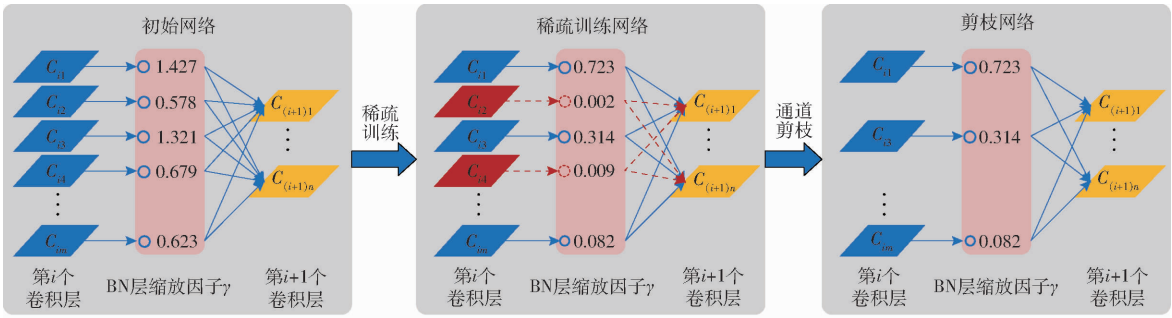

图 6　通道剪枝算法原理图

Fig. 6　Schematic of channel pruning algorithm

于 BN 层前的卷积层通道权重的影响愈低，可将其去除。因此，在稀疏训练过程中，选取缩放因子 γ 作为通道选择的代理[22]，并在正常训练的损失函数中引入 γ 的 L1 正则化约束，将其作为惩罚项，与网络权重联合训练，将缩放因子 γ 稀疏化[23]，稀疏训练损失函数计算式为

$$L=\sum_{(x,y)} l(f(x,W),y)+\lambda\sum_{\gamma\in\Gamma} g(\gamma) \tag{5}$$

式中　x——输入值　　W——权重

y——期望值　　λ——稀疏率

g——缩放因子的惩罚函数

f——原始网络的输出函数

l——原始网络的损失函数

从图 6 可以看出，经过稀疏训练后，BN 层缩放因子 γ 分布趋近于 0，表明所代表的通道对模型的贡献率低，可将这些通道及其对应的输入与输出进行修剪[24]，生成复杂度较低的压缩模型。最后对剪枝后的模型进行微调以恢复精度，在不影响模型检测精度的前提下压缩模型、优化结构、提高实时检测性能、使其更适于边缘计算设备的部署。

为更好地压缩模型，平衡模型的精度与速度，

本文针对 YOLO v8s 新的模型结构特性,制定了相应剪枝策略,便于模型在边缘计算设备部署。本文的剪枝策略基本方法为通过修剪 BN 层的通道来调整对应卷积层的输出通道以及下一个卷积层的输入通道,采用分模块剪枝的设计思想,具体方法为:

(1)避免维度计算和浅层网络剪枝[25]。从图 5 的 YOLO v8s 模型结构可以看出,存在的模块结构链接方式包括:CBS - CBS、CBS - C2f、C2f - CBS、C2f - C2f、C2f - SPPF、C2f - Detect、C2f - Concat、Concat - C2f。剪枝时需要对前一个剪枝层的输出通道和后一个剪枝层的输入通道进行剪枝操作调整,若后一个模块结构中有 Split 或 Concat 等操作,在对前一个模块结构的输出通道进行剪枝后会引起 Split 或 Concat 层的输入通道改变,与其进行 Split 或 Concat 等操作的层的输入通道也应随之改变,容易导致维度不一致问题。而对于有残差连接的模块的 Add 操作,通道剪枝后也难以保证残差连接的 2 个输入维度一致,因此,为避免维度计算,剪枝过程中不考虑 Split、Concat 和 Add 相关层。此外,由图 5 可知,模型初始的卷积层能够提取更加重要的特征信息,为保证模型精度,应避免浅层网络剪枝。

(2)不同模块设计不同剪枝方法[26]。从图 5 可以看出,C2f 作为 YOLO v8 中新增的网络模块,在主干网络和颈部网络中结构不一,应采用不同的剪枝方法。对于主干网络中含有残差连接的 18 个 C2f 模块,Split、Concat、Add 相关层不进行剪枝,对于颈部网络中的 12 个 C2f 模块,均可对其内部 BottleNeck2 中的 2 个卷积层进行剪枝。若 C2f 或 SPPF 为剪枝模块链接结构中的前一个剪枝层,则取其 Concat 操作后的下一个卷积层进行通道修剪,若为后一个剪枝层,则取其内部结构的第 1 个卷积层进行通道修剪。同样地,对 SPPF 内部的 Concat 相关层也不进行剪枝。对于 Detect 层,均可对内部的卷积层进行剪枝。

(3)为实现模型在边缘计算设备的加速,各卷积层保留通道数应大于等于 8 且能被 8 整除。

2 结果与分析

2.1 训练与部署平台

在 PC 端搭建模型训练环境,硬件配置主要包括 13th gen Intel(R)Core(TM)i5 - 13600K 处理器,主频 3.5 GHz,运行内存 32 GB,NVIDIA Quadro P4000 GPU,显存为 8 GB,软件环境为 64 位 Windows 10 操作系统、Pytorch 1.10 深度学习框架、Python 3.8 和 CUDA 11.3。

训练过程中保持模型输入图像尺寸为 640 像素 × 640 像素,线程数为 8,批量大小为 16,初始学习率为 0.01,动量为 0.937,权重衰减系数为 0.000 5,早停机制训练轮数为 100,迭代次数为 200,采用随机梯度下降法(Stochastic gradient descent, SGD)优化网络参数并开启 Mosaic 数据增强功能。

秉持降本增效原则,选取 NVIDIA Jetson Nano 为田间收获三七目标检测模型部署与应用平台,主要软硬件环境如表 1 所示。

表 1 边缘计算设备配置信息

Tab.1 Edge computing device configuration information

平台	名称	版本配置
Jetson Nano	操作系统	Ubuntu 18.04
	AI 算力	472 GFLOPS
	GPU	128 - core Maxwell @ 921 MHz
	CPU	4 - core ARM A57 @ 1.43 GHz
	内存	4 GB, 64 位 LPDDR4, 25.6 GB/s
	编程语言	Python 3.6
	Jetpack	4.6.1
	TensorRT	8.0.1.6

2.2 模型评价指标

采用精确率(Precision)、召回率(Recall)、平均精度均值(Mean average precision)来衡量模型的准确性;采用模型内存占用量、参数量、计算量来评估模型复杂度;采用检测帧率来评估边缘计算设备实时检测性能。

2.3 模型压缩试验分析

2.3.1 稀疏训练试验结果分析

为确保三七目标检测模型在稀疏化过程中保持较好的检测性能,采用不同的稀疏率 λ 进行试验,稀疏率不同,模型 BN 层权重分布和平均精度均值均会发生改变。为确定最佳稀疏率,将 λ 值分别设置为 0.002 5、0.005、0.01、0.02 后进行稀疏训练,不同稀疏率下的 BN 层缩放因子分布变化情况如图 7 所示,模型检测性能如表 2 所示。

图 7 表明,BN 层缩放因子 γ 随着稀疏训练进行逐渐趋近 0,稀疏率越大,γ 收敛于 0 的速度越快。原始模型的 γ 分布如图 7a 所示,总体上近似正态分布,分布中心为 1.0,此时区分不出通道重要性,无法进行通道筛选。结合表 2 分析,当 λ 为 0.002 5,经过 136 次迭代后 γ 分布开始收敛于 0.2 并趋于稳定,模型的平均精度均值较原始模型变化不大,精确率和召回率分别较原始模型下降 0.36、0.51 个百分点,γ 趋近于 0 的速度过慢,表明稀疏率过小;当 λ 为 0.005,稀疏训练经过 88 次迭代后 γ 分布开始收敛于 0 并趋于稳定,模型的平均精度均值较原始模型基本相同,精确率和召回率分别较原始模型下降

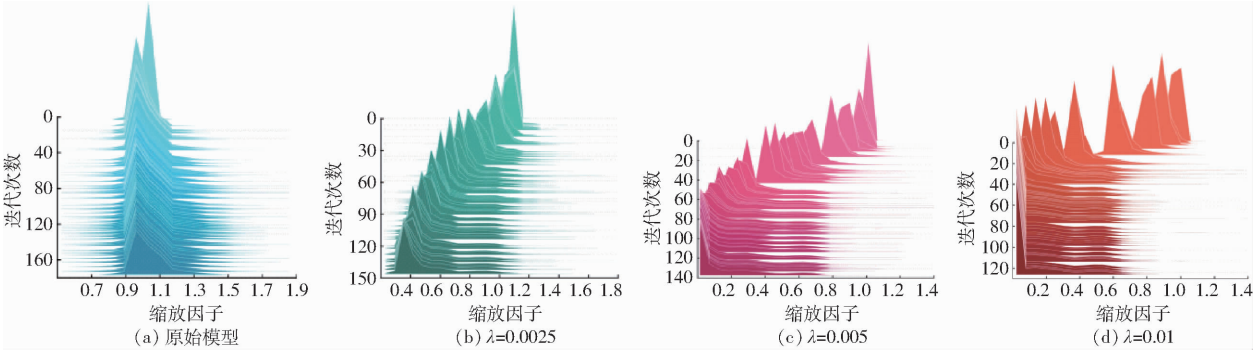

图 7　BN 层缩放因子 γ 分布情况

Fig. 7　Distribution of BN layer scaling factor γ

表 2　不同稀疏率下模型检测性能

Tab. 2　Model performance at different sparsity rates

%

稀疏率	平均精度均值	精确率	召回率
0	94.65	92.14	89.88
0.002 5	94.69	91.78	89.37
0.005	94.64	91.93	88.39
0.01	94.32	91.29	89.42
0.02	94.06	91.14	86.91

0.21、1.49 个百分点，精度损失较小；当 λ 值为 0.01，稀疏训练经过 52 次迭代后 γ 分布开始收敛于 0 并趋于稳定，模型的平均精度均值、精确率和召回率分别较原始模型下降 0.33、0.85、0.46 个百分点，虽然模型 γ 趋于 0 的速度快，但精度损失较大；当 λ 值为 0.02，模型的平均精度均值、精确率和召回率分别较原始模型下降 0.59、1.00、2.97 个百分点，精度损失最大。综上，本研究稀疏率 γ 取 0.005。

2.3.2　剪枝试验结果分析

通过稀疏训练试验结果得出当稀疏率为 0.005，迭代 140 次时，模型 BN 层缩放因子 γ 的值分布在零值附近的数量为 4.98×10^3 个，为平衡三七目标检测模型在压缩过程中的精度和模型复杂度，开展剪枝试验，不同剪枝率下的模型性能如图 8 所示。

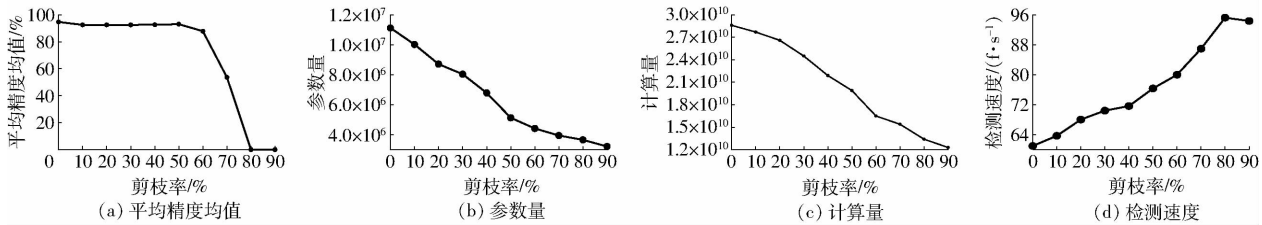

图 8　不同剪枝率下模型检测性能

Fig. 8　Model detection performance at different pruning rates

图 8 表明，采用不同的剪枝率，模型的检测性能均有不同程度变化。随着剪枝率增加，平均精度均值在初始阶段保持相对稳定，基本维持在 92.5%，当剪枝率大于 60%，平均精度均值开始急剧下降。原始模型的检测速度为 61 f/s，随着剪枝率增加，模型参数量和计算量不断下降，检测速度逐渐提高，当剪枝率大于 80%，检测速度略有下降。综上，在保证模型检测精度的前提下，采用 60% 的剪枝率，可有效压缩模型、减小计算复杂度、较大幅度提升检测速度，使模型更适用于边缘计算设备部署。

以 60% 的剪枝率进行剪枝后，模型各卷积层的通道数变化如图 9 所示。图 9 表明，对原始模型的 10 416 个通道进行剪枝后，部分卷积层通道数量大幅降低，最大修剪通道数为 504，平均每层被剪枝约 46 个通道，证明了该剪枝方法的有效性。

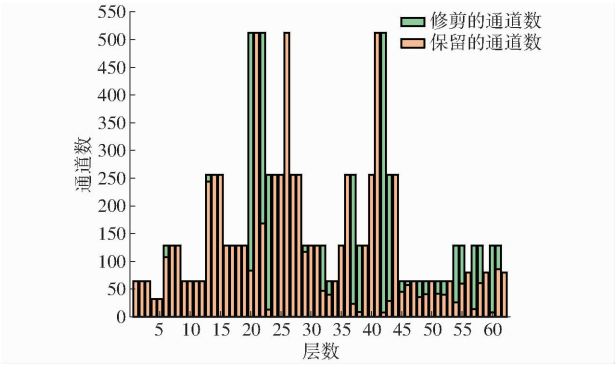

图 9　剪枝前后通道数量对比

Fig. 9　Channel changes before and after pruning

2.3.3　模型优化结果分析

田间收获三七目标检测模型在优化过程中的性能指标如表 3 所示。首先，采用 0.005 的稀疏率对原始 YOLO v8s 模型训练 200 次后，模型的精度和复杂度变化较小。其次，采用 60% 的剪枝率进行剪

表 3 优化模型性能

Tab. 3 Performance analysis of optimized model

指标	原始模型	稀疏训练	剪枝	微调
平均精度均值/%	94.65	94.64	87.83	93.71
精确率/%	92.14	91.93	87.95	92.19
召回率/%	89.88	88.39	83.12	88.08
参数量	1.112×10^7	1.112×10^7	4.42×10^6	4.42×10^6
计算量	2.86×10^{10}	2.86×10^{10}	1.65×10^{10}	1.65×10^{10}
模型内存占用量/MB	22.04	22.04	17.60	8.87
帧率/(f·s^{-1})	61	62	80	88

枝,模型的平均精度均值、精确率、召回率分别较原始模型下降 6.82、4.19、6.76 个百分点,模型的参数量、计算量、模型内存占用量分别为原始模型的 39.75%、57.69%、79.85%,检测速度提升至 80 f/s。最后,对模型微调以恢复精度,模型的平均精度均值、精确率、召回率分别为 93.71%、92.19%、88.08%,较微调前分别提升 5.88、4.24、4.96 个百分点,模型内存占用量为 8.87 MB,仅为剪枝模型的 1/2。将稀疏训练、剪枝、微调后的改进模型命名为 PN-YOLO v8s-Pruned(*Panax notoginseng* YOLO v8s-Pruned)。

图 10 为原始模型、稀疏训练模型和微调模型训练过程中在验证集上的平均精度均值变化情况。原始 YOLO v8s 模型经过 140 次迭代训练后,平均精度均值稳定在 93.5% 左右。在稀疏训练过程中,由于引入了 L1 正则化约束,模型的平均精度均值有小幅下降,经 120 次迭代训练后,稳定在 92.8% 左右。在微调模型的过程中,由于加载了稀疏训练后的预训练权重,模型初始平均精度均值较高,迭代 140 次后,基本稳定在 91.9% 左右。

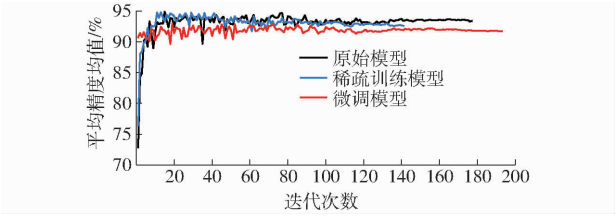

图 10 平均精度均值曲线

Fig. 10 Mean average precision curves

为验证改进后的 PN-YOLO v8s-Pruned 三七目标检测模型在复杂田间收获工况下的识别效果,采集三七在不同静态环境和动态输送分离阶段的图像进行测试,主机端检测效果如图 11 所示。图 11 表明,原始 YOLO v8s 模型和改进后模型均可识别光照较弱下的单个三七和光照充足下的多个三七,但改进模型预测框回归更准确,置信度更高;在识别须根缠绕下的多个三七时,原始模型存在 1 处误检(图中黄色三角形)、2 处漏检(图中黄色圆圈),改进模型存在 1 处漏检;原始模型和改进模型在识别晴天动态输送分离阶段须根缠绕的三七和阴天升运链上堆叠的三七时均存在 2 处漏检;原始模型和改进模型在识别阴天升运链上高速运动导致模糊的三七时均存在 1 处漏检,但改进模型置信度更高。综合上述分析并结合表 3 试验结果,表明所提出的 PN-YOLO v8s-Pruned 模型在保证检测精度的同时计算复杂度更小、检测实时性能更强、检测效果更好,更适用于模型在边缘计算设备的部署。

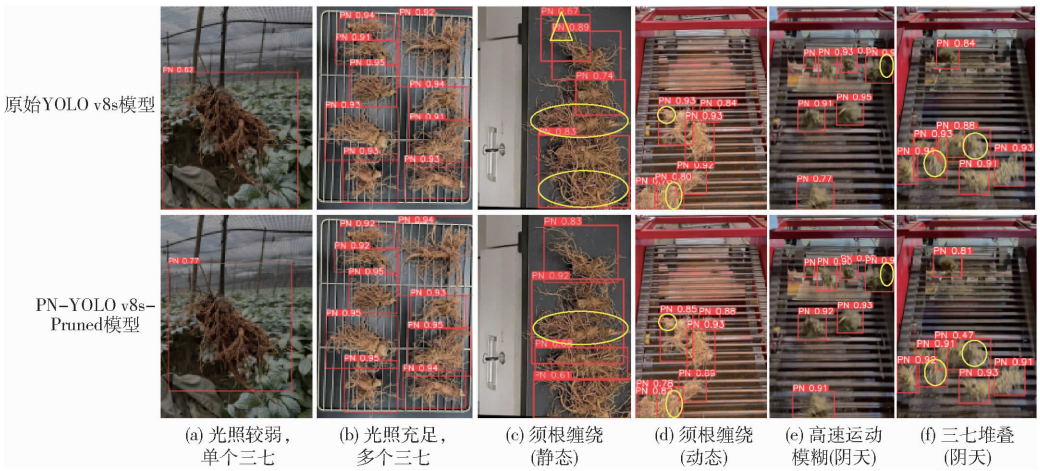

图 11 主机端改进前后网络检测效果对比

Fig. 11 Comparison of detection effect of network on host side before and after improvement

2.3.4 不同目标检测模型对比试验分析

为验证 PN-YOLO v8s-Pruned 改进模型综合检测性能的优越性,与原始 YOLO v8s 模型、YOLO v5s、YOLO v7-tiny、YOLO v9s、YOLO v10s、Faster R-CNN、SSD 及采用 MobileNetv4[27]、FasterNet[28]、ShuffleNetv2[29] 主干网络改进后的 YOLO v8s 轻量化目标检测模型在相同条件下进行对比试验,结果如表 4 所示。

表 4　改进模型与不同模型检测性能对比

Tab. 4　Comparison of detection performance between improved model and different lightweight models

模型	参数量	计算量	模型内存占用量/MB	精确率/%	召回率/%	平均精度均值/%	帧率/(f·s^{-1})
Faster R－CNN	$1.366\,9\times10^{8}$	4.017×10^{11}	110.77	75.47	93.82	94.03	7
SSD	2.361×10^{7}	2.732×10^{11}	92.78	89.55	92.74	93.82	17
YOLO v5s	7.01×10^{6}	1.58×10^{10}	14.05	93.41	88.57	94.68	90
YOLO v7－tiny	6.00×10^{6}	1.30×10^{10}	11.98	93.27	88.72	92.18	49
MobileNetv4－YOLO v8s	5.70×10^{6}	2.25×10^{10}	11.47	90.92	89.74	94.47	91
FasterNet－YOLO v8s	4.32×10^{6}	1.14×10^{10}	49.49	92.76	87.48	93.22	88
ShuffleNetv2－YOLO v8s	6.38×10^{6}	1.64×10^{10}	12.71	90.81	88.64	94.07	87
YOLO v8s	1.112×10^{7}	2.86×10^{10}	22.04	92.14	89.88	94.65	61
YOLO v9s	9.74×10^{6}	3.87×10^{10}	19.83	92.03	92.13	94.37	49
YOLO v10s	8.07×10^{6}	2.48×10^{10}	16.14	88.82	87.94	93.50	54
PN－YOLO v8s－Pruned	4.42×10^{6}	1.65×10^{10}	8.87	92.19	88.08	93.71	88

由表 4 可知，改进的 PN－YOLO v8s－Pruned 模型参数量、计算量、模型内存占用量仅为原始 YOLO v8s 模型的 39.75%、57.69%、40.25%，检测速度提升了 27 f/s，虽然平均精度均值较原始模型下降 0.94 个百分点，但模型复杂度大幅降低，实时检测速度明显提升；改进模型较 Faster R－CNN、SSD 的平均精度均值相差不大，但后者检测实时性差、模型复杂度过大、难以在边缘计算设备部署；与 YOLO v5s 相比，改进模型在检测精度上略显劣势，检测速度相差不大，但前者具有更小的模型复杂度；与 YOLO v7－tiny 相比，改进模型在精确率和召回率上略显劣势，但平均精度均值提升 1.53 个百分点，检测速度提升了 39 f/s，且具有更低的模型复杂度；与 FasterNet－YOLO v8s 相比，改进模型在参数量和计算量方面略显劣势，检测精度和检测速度相差不大，但后者的模型过大，不适用于部署与应用；与 MobileNetv4－YOLO v8s、ShuffleNetv2－YOLO v8s 相比，改进模型精确率分别提升 1.27、1.38 个百分点，召回率、平均精度均值和检测速度相差不大，但前者具有更低的计算复杂度；与 YOLO v9s、YOLO v10s 相比，改进模型在检测精度相差不大的同时计算复杂度更小，检测速度分别提升 39、34 f/s。综上所述，改进模型在模型复杂度、检测精度和实时性方面具有更好的综合检测性能，更适用于复杂田间收获工况下模型的部署与应用。

2.4　模型部署试验结果分析

为实现 PN－YOLO v8s－Pruned 三七目标检测模型部署与应用，选择 TensorRT 推理部署框架，将改进模型部署到 Jetson Nano 边缘计算平台。首先，将改进模型权重文件导出为.onnx 格式的中间文件，并添加和优化模型节点，然后将.onnx 模型文件导出至 Jetson Nnao，使用 TensorRT 推理加速框架解析网络模型，利用 C＋＋接口层层搭建模型，生成 engine 推理引擎，实现模型灵活部署，部署流程如图 12 所示，Jetson Nano 平台模型运行情况如图 13 所示。

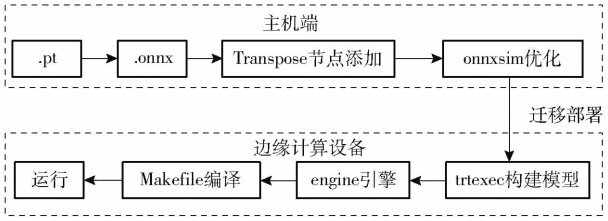

图 12　模型部署流程图

Fig. 12　Model deployment flowchart



图 13　Jetson Nano 平台模型运行图

Fig. 13　Diagram of model running on Jetson Nano

为验证模型部署后的性能，分别在主机端和 Jetson Nano 平台测试收获过程中的三七输送分离视频，视频持续 40 s，分辨率为 1 920 像素×1 080 像素，模型部署检测对比情况如表 5 所示。由表 5 可知，改进模型在 TensorRT 加速和未加速时检测速度分别较原始模型提升 5.8、1.8 f/s，加速前，改进模型由于边缘计算设备算力有限，检测速度较低，加速后，改进模型检测速度达 18.9 f/s，较加速前提升 2.7 倍。

为进一步验证改进模型部署于边缘计算设备后在复杂田间收获工况下目标三七的检测效果，选择晴天须根缠绕、三七堆叠和阴天高速运动模糊 3 种

表 5 模型部署检测速度对比

Tab. 5 Model deployment detection speed comparison

f/s

模型	主机端	Jetson Nano 平台	
		未加速	TensorRT 加速
YOLO v8s	61.0	5.2	13.1
PN－YOLO v8s－Pruned	88.0	7.0	18.9

较难识别的三七图像进行测试，检测效果如图 14 所示。

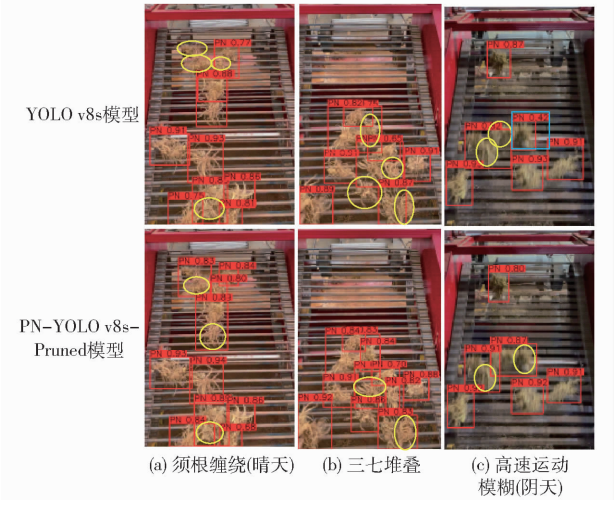

图 14 Jetson Nano 端改进前后模型检测效果对比

Fig. 14 Comparison of detection effect of model on Jetson Nano before and after improvement

由图 14 可知，在识别晴天升运链上须根缠绕的三七时，PN－YOLO v8s－Pruned 模型和原始模型分别存在 3 处、4 处漏检(图中黄色圆圈)；在识别晴天升运链上堆叠的三七时，改进模型和原始模型分别存在 2 处、4 处漏检，且改进模型置信度更高；在识别阴天升运链上因高速运动导致模糊的三七时，改进模型和原始模型均存在 2 处漏检，原始模型有 1 处目标置信度过低(图中蓝色方框)。3 种收获工况下改进模型的识别能力均优于原始模型。

3 验证试验

3.1 台架试验

为验证部署在边缘计算设备的 PN－YOLO v8s－Pruned 模型在不同流量等级和不同输送分离作业参数影响下的三七目标检测性能和收获作业实时监测效果，开展台架试验，如图 15 所示。在实际收获作业过程中，三七联合收获机振动频率和升运速度为可调变量，满足最佳输送分离效果时，输送分离试验台架升运速度为 0.86～0.92 m/s，振动频率为 1.4～1.7 Hz[3]。根据升运链上的三七数量，将三七流量等级分为Ⅰ级、Ⅱ级、Ⅲ级。在确定输送分离试验台架最佳升运速度和振动频率的条件下，三七流

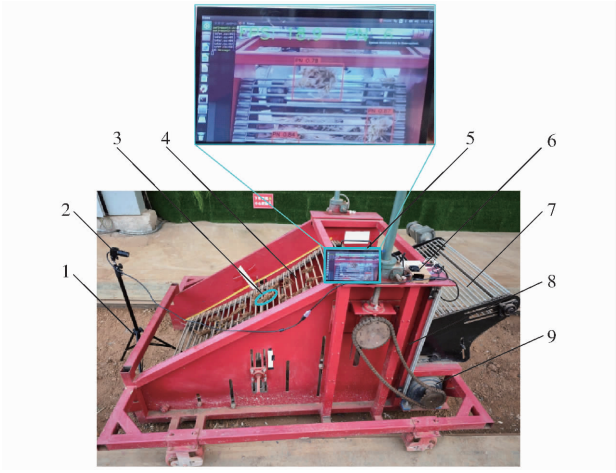

图 15 台架试验

Fig. 15 Bench tests

1. 三脚架 2. 工业相机 3. 收获出土的三七 4. 一级升运链 5. 显示屏 6. Jetson Nano 7. 二级升运链 8. 输送分离试验台架 9. 电动机

量等级评估如表 6 所示。试验过程中，不同流量等级和不同输送分离作业参数影响下的输送分离过程均持续 40 s，工业相机分辨率为 1 920 像素×1 080 像素。

表 6 流量等级评估

Tab. 6 Traffic level evaluation form

升运速度/(m·s^{-1})	振动频率/Hz	流量等级	数量
0.86～0.92	1.4～1.7	Ⅰ	(0,10]
		Ⅱ	(10,20]
		Ⅲ	(20,+∞)

收获状态监测系统可根据检测到的升运链上三七数量进行流量等级实时评估，当升运链上的收获三七喂入量过多时，调节一级升运链转速加快，避免三七之间堆叠堵塞，形成不必要的“滚动”与“回流”，大幅提升输送分离效率。当升运链上的收获三七喂入量减少时，调节一级升运链转速下降，减少三七与栅条之间的线性擦伤，并降低能耗。准确、实时预测三七流量等级，可为三七联合收获作业中的自适应分级调速提供技术支撑。不同流量等级下三七计数结果如表 7 所示。由表 7 可知，不同流量等级下三七计数正确率达 89% 以上，三七计数平均正确率达 91.76%，随着流量等级的提升，三七与三

表 7 不同流量等级下三七计数结果

Tab. 7 Counting results of *Panax notoginseng* at different flow grades

流量等级	人工观测值/个	模型预测值/个	正确率/%
Ⅰ	69	65	94.20
Ⅱ	165	151	91.52
Ⅲ	278	249	89.57

七、三七与土壤之间重合、遮挡程度过大，导致三七计数正确率有所下降。

不同输送分离作业参数影响下的三七目标检测性能和计数结果如表 8 所示。从表 8 可以看出，当振动频率为 1.4 Hz，升运速度为 0.86 m/s 时，部署模型的平均精度均值达 92.82%，目标三七计数正确率达 94.37%，模型检测性能和计数结果最佳。随着振动频率和升运速度增加，模型的召回率、平均精度均值、三七计数正确率呈下降趋势。原因在于，随着升运链振动频率和升运速度增加，输送分离过程中的无规律、非线性运动加剧，给模型检测带来了一定难度。整体而言，目标三七检测精确率达 90% 以上，召回率达 80% 以上，平均精度均值达 87% 以上，三七计数平均正确率达 92.61%，检测速度相差不大，模型检测性能和计数效果能够满足复杂田间收获工况下目标三七的检测需求。

表 8　台架试验模型部署检测性能和计数结果

Tab.8　Detection performance of deployed model and counting results in bench tests

试验编号	升运速度/($m\cdot s^{-1}$)	振动频率/Hz	精确率/%	召回率/%	平均精度均值/%	帧率/($f\cdot s^{-1}$)	人工观测值/个	模型预测值/个	计数正确率/%
1	0.86	1.4	91.02	87.44	92.82	31.5	71	67	94.37
2	0.88	1.5	90.91	84.61	91.32	32.1	71	66	92.96
3	0.90	1.6	91.15	82.24	90.37	33.1	71	66	92.96
4	0.92	1.7	90.04	80.91	87.25	31.3	71	64	90.14

3.2　田间试验

为测试部署于边缘计算设备的改进模型在实际收获工况下三七目标检测性能和收获作业监测效果，于云南省昆明市西山区三年生成熟三七种植地(24°24′N，102°27′E)开展田间试验。三七平均垄高 200 mm，平均垄宽 1 500 mm，土壤平均含水率 14.65%，收获前机械化杀秧，牵引式三七收获机作业速度为 0.6 m/s。实际收获过程中，一级输送分离装置升运速度和振动幅度是影响三七识别检测的重要因素，调节一级输送分离装置升运速度为 0.9～1.2 m/s，振动幅度为 20～50 mm，每次试验时长持续 40 s，田间试验如图 16 所示。



图 16　田间试验

Fig. 16　Field experiment

1. 东方红 500 型拖拉机　2. 三七轻简型收获机　3. 一级输送分离装置　4. 振动轮　5. 一级主动轮

田间试验三七目标检测性能和计数结果如表 9 所示。从表 9 可以看出，随着一级输送分离装置的升运速度和振动幅度增加，模型的精确率、召回率、平均精度均值、三七计数正确率均呈下降趋势。原因在于，实际田间收获作业工况下，随着升运速度和振动幅度增加，不但输送分离过程的无规律、非线性运动加剧，且自然环境复杂和七土分离不彻底给三七目标识别与检测带来了困难。整体而言，目标三七检测平均精度均值达 84% 以上，三七计数平均正确率达 88.11%，模型检测性能和计数效果能够满足实际田间收获工况下目标三七的检测需求。

表 9　田间试验模型部署检测性能和计数结果

Tab.9　Detection performance of deployed model and counting results in field experiments

试验编号	升运速度/($m\cdot s^{-1}$)	振动幅度/mm	精确率/%	召回率/%	平均精度均值/%	帧率/($f\cdot s^{-1}$)	人工观测值/个	模型预测值/个	计数正确率/%
1	0.9	20	89.04	86.01	88.61	33.0	153	139	90.85
2	1.0	30	88.65	83.79	87.32	31.5	167	148	88.62
3	1.1	40	87.92	81.34	86.17	32.1	142	124	87.32
4	1.2	50	87.14	79.53	84.96	31.0	181	155	85.64

4　结论

(1)为实现三七机械化收获过程中的目标三七准确、快速识别，本文融合 YOLO v8s 目标检测模型和通道剪枝算法，提出了一种可灵活部署于边缘计算设备的轻量化三七目标检测方法。模型压缩试验结果表明，当稀疏率 λ 为 0.005，剪枝率为 60% 时，剪枝模型微调后性能最佳，平均精度均值达 93.71%，尽管较原始模型下降 0.94 个百分点，但参

数量、计算量、模型内存占用量分别为原始模型的 39.75%、57.69%、40.25%，检测速度提升 27 f/s，改进模型较原始模型具有更低的计算复杂度，更适用于边缘计算设备部署在复杂田间收获工况具有更好的检测效果。

(2)改进后的 PN-YOLO v8s-Pruned 模型在计算复杂度、检测精度和实时性方面较其它目标检测模型相比具有更好的综合检测性能。模型部署试验结果表明，改进模型在 TensorRT 加速后检测速度达 18.9 f/s，较加速前提升 2.7 倍，较原始模型提升 5.8 f/s，在边缘计算设备的检测效果优于原始模型。

(3)台架试验结果表明，随着振动频率和升运速度增加，模型的召回率、平均精度均值呈下降趋势。4 种输送分离收获作业工况下三七目标检测的平均精度均值达 87% 以上，计数平均正确率达 92.61%，不同流量等级下三七计数平均正确率达 91.76%。田间试验结果表明，三七目标检测平均精度均值达 84% 以上，计数平均正确率达 88.11%，图像推理速度最低达 31.0 f/s。模型检测性能和计数效果能够满足复杂田间收获工况下目标三七的检测需求，可为后续三七联合收获机上集成自适应分级输送系统和收获作业质量监测系统奠定基础。

参 考 文 献

[1] LIN Y H, ZHANG F J, LI L X, et al. Identification of *Panax notoginseng* powder in different parts based on the electronic nose and time-domain feature extraction[J]. Journal of the Electrochemical Society, 2022, 169(4): 047510-047519.

[2] 张兆国, 王一驰, 李汉青, 等. 液压控制履带自走式温室三七收获机设计与试验[J]. 农业机械学报, 2021, 52(6): 127-135.
ZHANG Zhaoguo, WANG Yichi, LI Hanqing, et al. Design and test of hydraulic control tracked self-propelled greenhouse *Panax notoginseng* harvester[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(6): 127-135. (in Chinese)

[3] 王法安, 温博, 谢晓红, 等. 三七收获机输送分离装置作业机理分析与参数优化[J]. 农业机械学报, 2023, 54(增刊 1): 201-211, 259.
WANG Faan, WEN Bo, XIE Xiaohong, et al. Operation mechanism analysis and parameter optimization of conveying and separating device of *Panax notoginseng* harvester[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(Supp. 1): 201-211, 259. (in Chinese)

[4] 刘毅君, 何亚凯, 吴晓媚, 等. 基于改进 Faster R-CNN 的马铃薯发芽与表面损伤检测方法[J]. 农业机械学报, 2024, 55(1): 371-378.
LIU Yijun, HE Yakai, WU Xiaomei, et al. Potato sprouting and surface damage detection method based on improved Faster R-CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1): 371-378. (in Chinese)

[5] 张兆国, 张振东, 李加念, 等. 采用改进 YoloV4 模型检测复杂环境下马铃薯[J]. 农业工程学报, 2021, 37(22): 170-178.
ZHANG Zhaoguo, ZHANG Zhendong, LI Jianian, et al. Potato detection in complex environment based on improved YoloV4 model[J]. Transactions of the CSAE, 2021, 37(22): 170-178. (in Chinese)

[6] LESTARI S, RAHMAN A Y. ISTIADI. Quality detection of export purple sweet potatoes using Yolov4-Tiny[C]// IEEE International Seminar on Intelligent Technology and Its Applications, Surabaya, Indonesia: IEEE, 2023:610-614.

[7] LIONG S T, WU Y L, LIONG G B, et al. Moving towards agriculture 4.0: an AI-AOI carrot inspection systemwith accurate geometric properties[J]. Journal of Food Engineering, 2023, 357: 111632.

[8] 黄家才, 唐安, 陈光明, 等. 基于 Compact-YOLO v4 的茶叶嫩芽移动端识别方法[J]. 农业机械学报, 2023, 54(3): 282-290.
HUANG Jiacai, TANG An, CHEN Guangming, et al. Mobile recognition solution of tea buds based on Compact-YOLO v4 algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(3): 282-290. (in Chinese)

[9] 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)

[10] LI Y, HE L, CHEN J, et al. High-efficiency tea shoot detection method via a compressed deep learning model[J]. International Journal of Agricultural and Biological Engineering, 2022, 15(3): 159-166.

[11] 刘莫尘, 高甜甜, 马宗旭, 等. 基于 MSRCR-YOLO v4-tiny 的田间玉米杂草检测模型[J]. 农业机械学报, 2022, 53(2): 246-255, 335.
LIU Mochen, GAO Tiantian, MA Zongxu, et al. Target detection model of corn weeds in field environment based on MSRCR algorithm and YOLO v4-tiny[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(2): 246-255, 335. (in Chinese)

[12] 黄华, 张昊, 胡晓林, 等. 基于改进 YOLO v5 的复杂环境下花椒簇识别与定位方法[J]. 农业机械学报, 2024, 55(3): 243-251.
HUANG Hua, ZHANG Hao, HU Xiaolin, et al. Recognition and localization method for pepper clusters in complex

environments based on improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(3): 243－251. (in Chinese)

[13] 李茂, 肖洋轶, 宗望远, 等. 基于改进 YOLO v8 模型的轻量化板栗果实识别方法[J]. 农业工程学报, 2024, 40(1): 201－209.
LI Mao, XIAO Yangyi, ZONG Wangyuan, et al. Detecting chestnuts using improved lightweight YOLO v8[J]. Transactions of the CSAE, 2024, 40(1): 201－209. (in Chinese)

[14] LI X, WANG W H, HU X L, et al. Generalized focal loss v2: learning reliable localization quality estimation for dense object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 11632－11641.

[15] LOU H T, DUAN X H, GUO J M, et al. DC－YOLO v8: small-size object detection algorithm based on camera sensor[J]. Electronics, 2023, 12(10): 2323.

[16] WANG C, BOCHKOVSKIY A, LIAO H. YOLO v7: trainable bag of freebies sets new state of the art for real-time object detectors[EB/OL]. (2022－07－06)[2024－04－06]. https://arxiv.org/abs/2207.02696.

[17] 马超伟, 张浩, 马新明, 等. 基于改进 YOLO v8 的轻量化小麦病害检测方法[J]. 农业工程学报, 2024, 40(5): 187－195.
MA Chaowei, ZHANG Hao, MA Xinming, et al. Method for the lightweight detection of wheat disease using improved YOLO v8[J]. Transactions of the CSAE, 2024, 40(5): 187－195. (in Chinese)

[18] 周宏平, 金寿祥, 周磊, 等. 基于迁移学习与 YOLO v8n 的田间油茶果分类识别[J]. 农业工程学报, 2023, 39(20): 159－166.
ZHOU Hongping, JIN Shouxiang, ZHOU Lei, et al. Classification and recognition of *Camellia oleifera* fruit in the field based on transfer learning and YOLO v8n[J]. Transactions of the CSAE, 2023, 39(20): 159－166. (in Chinese)

[19] LIU Z, LI J, SHEN Z, et al. Learning efficient convolutional networks through network slimming[C]// Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy: IEEE, 2017: 2736－2744.

[20] 蔡管鸿, 李国平, 王国中, 等. 基于改进 YOLO v5s 的轻量化交通灯检测算法[J]. 上海大学学报, 2024, 30(1): 94－105.
CAI Guanhong, LI Guoping, WANG Guozhong, et al. Lightweight traffic-light detection algorithm based on improved YOLO v5s[J]. Journal of Shanghai University, 2024, 30(1): 94－105. (in Chinese)

[21] 冀文欢. 多目标识别深度学习网络轻量化技术研究[D]. 西安: 西安工业大学, 2024.
JI Wenhuan. Research on lightweight technology of multi target recognition deep learning network[D]. Xi'an: Xi'an Technological University, 2024. (in Chinese)

[22] 王政, 许兴时, 华志新, 等. 融合 YOLO v5n 与通道剪枝算法的轻量化奶牛发情行为识别[J]. 农业工程学报, 2022, 38(23): 130－140.
WANG Zheng, XU Xingshi, HUA Zhixin, et al. Lightweight recognition for the oestrus behavior of dairy cows combining YOLO v5n and channel pruning[J]. Transactions of the CSAE, 2022, 38(23): 130－140. (in Chinese)

[23] WANG S, DU C, MENG Y, et al. Automatic inspection method for YOLO v5s railway freight vehicles based on channel pruning[C]// International Conference on Intelligent Communication and Networking, Changzhou, China: IEEE, 2023: 307－313.

[24] DU W S, JIA Z H, SUI S S, et al. Table grape inflorescence detection and clamping point localisation based on channel pruned YOLOV7－TP[J]. Biosystems Engineering, 2023, 235: 100－115.

[25] LIU Z J. Unmanned aerial vehicles general aerial person-vehicle recognition based on improved YOLO v8s algorithm[J]. Computers, Materials & Continua, 2024, 78(3):3787－3803.

[26] ZHANG J C, WANG P Y, ZHAO Z C, et al. Pruned-YOLO: learning efficient object detector using model pruning[C]// International Conference on Artificial Neural Networks, Springer, Cham, 2021:34－35.

[27] QIN D, LEICHNER C, DELAKIS M, et al. MobileNetV4－universal models for the mobile ecosystem[EB/OL]. (2024－04－16)[2024－05－01]. https://arxiv.org/abs/2404.10518.

[28] CHEN J, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[EB/OL]. (2023－03－07)[2024－05－06]. https://arxiv.org/abs/2303.03667.

[29] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: practical guidelines for efficient CNN architecture design[C]// Proceedings of the European Conference on Computervision, Munich, 2018: 116－131.