DOI: 10.12398/j.issn.2096-7217.2025.02.003

果园作业机器人自主导航多任务联合感知方法研究

张津国[1,2], 蔡建峰[1,2], 姜蓉蓉[3], 余山山[4], 王蓬勃[1,2*]

(1. 苏州大学机电工程学院,江苏苏州,215137; 2. 江苏省具身智能机器人技术重点实验室,江苏苏州,215137; 3. 苏州漕阳生态农业发展有限公司,江苏苏州,215143; 4. 农业农村部南京农业机械化研究所,江苏南京,210014)

摘要:针对果园复杂作业场景中植被异质性高、光照动态变化显著以及目标形态多样性等挑战,传统基于单任务学习的视觉感知模型因特征复用率低、计算冗余度高等问题,难以满足农业机器人实时环境感知需求。本研究提出一种面向果园场景的轻量化多任务联合感知框架AgriYOLOP,通过系统性地重构YOLOP网络架构,引入高效的主干网络,并添加增强的无锚检测技术以及增加了特征金字塔网络和路径聚合网络,并设计任务自适应的损失函数权重策略,实现树干检测、障碍物识别与可通行区域分割的并行化协同处理。在自主构建的果园数据集中进行验证,该数据集包含果园场景下4 765张分辨率为1 280像素×720像素的图像,涵盖不同季节、光照条件及植被生长阶段。试验结果表明,AgriYOLOP模型在目标检测任务中检测准确率、召回率以及*mAP*50分别达到92.7%、94.6%以及96.7%,在可行区域分割任务下召回率、*F*1分数以及*mIoU*达到98.3%、99.1%、98.1%,在NVIDIA RTX 4060平台下实现69f/s的实时推理速度,模型参数量仅14 M。对比结果表明,多任务协同学习架构能够显著提升特征共享效率,较单任务模型减少32.6%推理延迟,同时增强了对光照变化与季节特征的鲁棒性,有效地解决了农业机器人实时作业场景中目标检测精度与语义分割效率的协同优化难题。本研究为果园机器人自主导航提供了高精度、低延迟的实时感知解决方案。

关键词:多任务学习;果园环境感知;目标检测;语义分割;农业机器人

中图分类号:S24 **文献标识码**:A **文章编号**:2096-7217(2025)02-0035-09

张津国, 蔡建峰, 姜蓉蓉, 余山山, 王蓬勃. 果园作业机器人自主导航多任务联合感知方法研究[J]. 智能化农业装备学报(中英文), 2025, 6(2): 35-43

ZHANG Jinguo, CAI Jianfeng, JIANG Rongrong, YU Shanshan, WANG Pengbo. Multi-task joint perception framework for autonomous navigation in orchard robotics [J]. Journal of Intelligent Agricultural Mechanization, 2025, 6(2): 35-43

0 引言

随着我国农业产业结构升级与消费需求转型,果园种植规模持续扩张。截至2023年,我国果蔬作物栽培面积已突破22 867 khm^2,稳居全球首位[1-2]。然而,人口老龄化加剧导致的农业劳动力短缺已成为制约产业发展的核心矛盾。据统计,近十年农业从业人口年均减少约3.2%,且青壮年劳动力向第三产业转移趋势显著[3-4]。在此背景下,果园作业机器人因其在智能作业及自主导航等领域的技术优势,成为破解劳动力短缺、实现精细化管理的关键突破口[5]。果园作业机器人的视觉感知系统需应对植被异质性高、光照动态变化等复杂场景特性。通过多任务联合特征提取与参数共享机制,实现计算资源的优化配置与高效复用。这对于提升果园机器人在复杂非结构化环境下的自主作业能力具有重要的研究价值与应用意义。

在智能机器人环境感知技术领域,目标识别与语义分割构成视觉感知系统的2大核心研究领域,其算法演进为农业机器人实现自主环境交互提供了理论支撑[6]。目标识别领域呈现多元化架构发展,当前主流的识别模型Faster R-CNN[7]、DETR[8]和YOLO[9-11]系列等,图像分割领域也涌现一些优秀的模型,如DeepLab[12]、U-Net[13]和DFANet[14]等,这些模型在可行驶区域分割等特定任务中展现出优异的性能。

近年来,多任务学习(multi-task learning,MTL)在自动驾驶领域展现出显著优势。Uber公司的先进技术团队通过将物体检测、车道检测和场景分割等任务结合到一个模型中,研究了多任务学习[15]。由于利用

收稿日期:2025-04-10 修回日期:2025-05-10

基金项目:国家重点研发计划项目(2022YFB4702202);江苏省农业农村厅农机研发制造推广应用一体化试点项目(JSYTH07)

第一作者:张津国,男,2000年生,江苏连云港人,硕士研究生;研究方向为农业机器人导航技术。E-mail: 20225229128@stu.suda.edu.cn

*通信作者:王蓬勃,男,1984年生,安徽阜阳人,博士,副教授;研究方向为农业机器人、农业人工智能。E-mail: pbwang@suda.edu.cn

了共享表示,从而使得性能显著提升。类似地,SISTU研究团队[16]提出了一个多任务网络,可以同时检测和分割交通场景中的物体。该研究表明,多任务学习(MTL)可以通过允许不同任务之间共享卷积特征,从而减少处理需求并提高准确性。在报告中,MTL被视为一个多目标优化问题[17]。他们开发了一种基于帕累托前沿的方法平衡任务性能,证明通过仔细平衡任务贡献,自动驾驶感知系统的整体性能提升显著。同时,在制定有效的任务选择方法论和管理每个任务在训练过程中的影响方面,存在着诸多挑战[18-19]。

针对农业机器人面临果园行间环境复杂、光线多变问题,在多任务串行执行机制导致计算复杂度增长的约束下,难以满足机器人实时作业的低延迟需求。因此,轻量化多任务联合学习框架成为解决该问题的关键技术路径。本研究提出的AgriYOLOP多任务联合优化框架。通过轻量化架构设计、动态任务平衡机制和农业场景适应性优化,为果园机器人自主导航提供了高实时性、强适应性的感知解决方案,并为农业多任务学习理论体系构建提供技术参考。

1 材料与方法

1.1 果园行间图像获取

果园行间数据集采集地点位于江苏省苏州市相城区倪家湾水之田枇杷果园种植基地,数据集采集设备由相机、相机固定三脚架、相机数据连接线以及笔记本电脑组成。利用编写的Python图像采集程序,每隔1 s保存1张图像,图像格式为JPG,图像分辨率为1 280像素×720像素。

考虑到室外果园在不同光照和气候条件下对于图像质量的影响,数据集采集的时间为2024年2月—5月、2024年9月—11月,覆盖了枇杷果园在秋季与春季中的果园行间图像,包含了不同光照条件、不同时间段(上午、中午、傍晚)下的行间图像,总计4 765幅原始RGB与深度图像,典型图像样本如图1所示。



(a) 春季枇杷果园行间顺光 (b) 春季枇杷果园行间逆光 (c) 秋季枇杷果园行间顺光 (d) 秋季枇杷果园行间逆光

图1 不同季节下果园行间采集数据集示例

Figure 1 Samples of datasets collected in orchard inter-row spaces across different seasons

1.2 果园数据集制作

模型的训练需要对原始图像中树干、障碍物以及进行标注,采用Labelme图像标注工具对前期采集的数据集进行标注,在软件中能够完成对树干与障碍物目标的标记,同时完成目标边界框的绘制及标注。

基于AgriYOLOP多任务视觉感知模型需要同时完成树干识别、障碍物检测以及分割3个任务,为了便于后续模型训练与后处理,标注过程需注意:

1) 树干标注:对作业路径双侧的枇杷果树主干进行标注,并在框内添加标签类别(trunk)。

2) 障碍物标注:当前作业行内出现的纸箱障碍物,添加标签(obstacle)。

3) 可导航区域标注:机器人行间为一个闭环,以单侧树行为起点,中间添加数条直线,最后连接到开端,为其添加标签(lane)。

数据集共计4 765张图像,按照训练集占80%,验证集占20%,最终3 812张标注图像用于感知模型训练,953张用于验证。在完成数据集标注工作之后,针对标注数据格式兼容性问题,开发基于Python语言的格式转换脚本,实现标注信息从JSON到BDD规范的结构化映射,为后续模型性能验证提供可靠数据基础。

2 果园行间多任务视觉联合感知算法

为辅助果园机器人自主导航任务,视觉感知模块应该能够理解果园场景,向运动规划模块提供导航数据,如障碍物的位置、树干与机器人的相对位置等。许多前沿的算法可分别执行这些任务。例如,Faster R-CNN和RetinaNet[20]处理目标检测,DeepLab和U-Net则用于语义分割。无论这些算法的效果如何,逐个处理单一任务比同时完成它们更耗时。在果园机器人作业平台上部署感知模型时,必须考虑延迟。

2.1 多任务学习模块

多任务学习(multi-task learning,MTL)是一种训练方法,其中机器学习(machine learning, ML)模型同时利用来自多个任务的数据进行训练[21]。为了避免为每个活动训练单独的模型,该方法利用共享表示来理解一组相关学习任务中的共享概念。这种策略通过识别任务之间共享的基本模式,提高了模型的泛化能力,从而增强了模型检测性能和鲁棒性。此外,共享表示能够加速相关或后续任务的学习过程,因为模型已经获得了可以再次利用的相关特征。图2展示了MTL网络的典型架构。

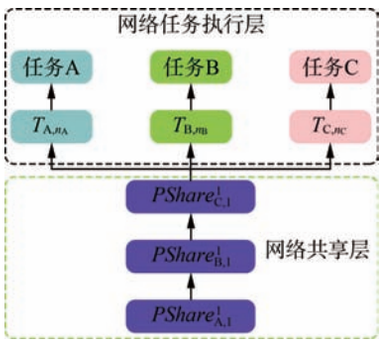


图 2　多任务学习神经网络参数共享示意图

Figure 2　Illustration of parameter sharing for MTL neural networks

在自动驾驶领域，针对感知的多任务学习进行了广泛的研究。这是因为感知任务，如检测交通物体、分割和检测车道，彼此密切相关，通常可以使用相同的输入进行预测。自动驾驶领域多任务学习的基础研究为整合多个感知任务奠定了基础。利用基于加权的方法用于多任务学习，依据其不确定性动态调整每个任务的损失贡献。该方法提升了语义分割、检测和深度估计等任务的性能，从而在训练过程中平衡了每个任务的重要性，以实现更精确和一致的结果。此外，注意力机制已被整合到自动驾驶的 MTL 框架中。在研究中引入了一种基于注意力驱动的 MTL 方法[22]。该方法使模型能够选择性地关注每个任务的相关特征，从而改善对象分割和车道检测等感知任务的性能。

MTL 的实际优势通过其在自主交通中的应用得到了体现。研究探讨了多个视觉任务之间的知识转移[23]。试验表明，该模型通过利用任务关系和共享表示，能够在 MTL 框架下学习各种感知任务，从而实现最稳定的感知结果。

在自主驾驶的背景下，研究探讨了 MTL，并提出了一种多任务架构，该架构同时执行车道检测、物体检测和可行驶区域分割等任务[24]。这种方法强调了 MTL 在增强自主车辆感知系统的鲁棒性和准确性方面的优势。

尽管多任务学习在自动驾驶中具有潜在优势，但仍面临许多障碍。负迁移发生在一个任务的学习对另一个任务产生不利影响时，这凸显了开发减少任务干扰的方法的重要性。任务可扩展性是另一个障碍，因为管理和分配任务的复杂性随着工作量的增加而增加。

2.2　基于 AgriYOLOP 的枇杷果园行间多任务视觉联合感知方法

通过改进 YOLOP 模型，提出一种多任务联合优化框架 AgriYOLOP（agriculture you only look once for panoptic），以解决果园环境下多任务视觉感知问题。YOLOP 模型作为多任务联合学习框架的典型代表，其构建了基于特征共享机制的实时全景感知架构。该模型通过统一的编码器－解码器网络结构，并行实现动态目标检测（车辆、行人）、可通行区域语义分割以及结构化道路元素（车道线）识别 3 大核心任务。

本研究提出的 AgriYOLOP 多任务感知框架，基于 YOLOP 架构进行农业场景适应性改进，旨在解决果园环境多任务感知的技术难题，优化后的模型共享一个由主干（backbone）和颈部（neck）组成的编码器，并结合 2 个解码器完成目标检测和可行驶区域分割任务，同时该模型实现了低计算消耗、高精度和快速感知，本研究提出的 AgriYOLOP 模型的网络结构示意图如图 3 所示。

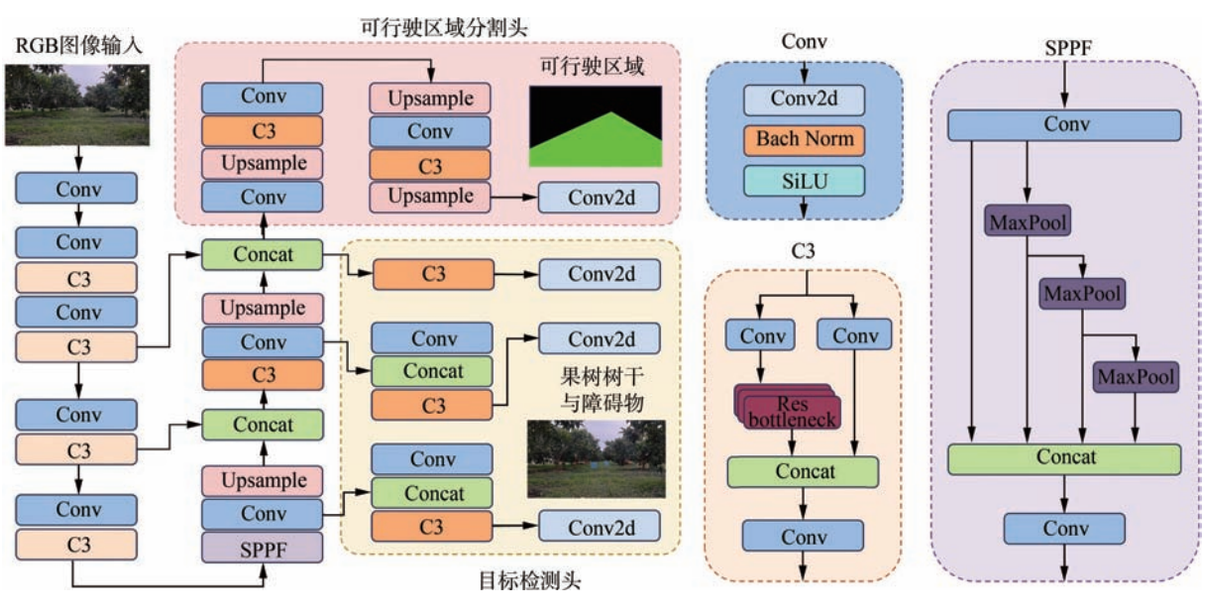


图 3　AgriYOLOP 网络模型结构示意图

Figure 3　Schematic diagram of the AgriYOLOP multi-task learning network architecture

首先对YOLOP模型负责目标检测的结构进行了改进,采用最新的YOLOv5[25]架构而不是原来的YOLOv4[26]网络。从YOLOv4升级到YOLOv5需要替换多个组件,包括更加高效的主干网络、增强的无锚检测技术,以及增加了特征金字塔网络[27]和路径聚合网络[28]利用率,这些特征的结合能够有效提高目标检测的精度。

其次,对YOLOP模型的检测头做了简化处理,以识别果园中特定的物体,农业环境中实际部署需要针对特定的任务进行优化处理。由于优化后的检测头减少了对无关对象的检测,模型可以通过专注于特定任务以实现更高的准确性、更快的识别速度。

最后,YOLOP模型中的车道线分割头在AgriYOLOP模型中不再需要,因为果园中的车道信息可通过准确识别树干获得。通过去除车道线分割头,AgriYOLOP模型不仅提升了处理速度,还增强了模型在果园环境中的适应性。在果园环境中,树干位置可以作为导航任务固定的参考点,帮助引导果园机器人在行间移动,这不仅降低了计算的复杂性,还满足了果园环境的独特需求。

2.3 损失函数

AgriYOLOP模型的损失函数定义为树干检测的损失与分割损失的加权和,如式(1)所示。

$$L=\gamma_1 L_{\text{det}}+\gamma_2 L_{\text{seg}}=\gamma_1\left(\alpha_1 L_{\text{class}}+\alpha_2 L_{\text{obj}}+\alpha_3 L_{\text{box}}\right)+\gamma_2 L_{\text{seg}} \tag{1}$$

式中:γ_1、γ_2、α_1、α_2——可调整超参数,用以平衡多任务感知模型中不同的子任务。

目标边界框回归包含所需预测的4个值分别为:目标框的中心坐标(x,y)以及目标框的尺寸(*height*,*width*)。L_{box}的回归损失通常由两部分组成:

1)预测中心坐标与实际中心坐标之间的差值所组成的中心坐标损失;

2)目标框高度宽度损失通常利用这些值的均方差(*MSE*)计算,同时也考虑了对数刻度以更好地处理不同对象尺寸。

对象损失L_{obj}表明预测边界框中包含目标对象的置信度,该数值范围从0到1。对象损失L_{box}通常采用BCE损失,将预测对象的分数与真值进行对比(如果目标框内包含一个对象则为1,反之则为0)。

分类损失L_{class}表明所预测类别与真实类别的符合程度,其通常利用交叉熵损失(cross-entropy loss)进行多类别分类工作,每个网格单元都可以预测所有可能类别的概率分布。

分割头是基于DeepLabv3语义分割模块,故可行驶区域分割损失与DeepLabv3模型语义分割损失相同。

3 试验方法与结果

根据前述提出的基于视觉的多任务感知模块。分别将AgriYOLOP模型的性能与前沿的深度神经网络在不同任务维度上进行了对比。

3.1 试验平台与评价指标

本试验计算环境为:硬件主机平台搭载了AMD锐龙R9-7945HX处理器(主频5.40 GHz),运行内存为16 GB、固态存储空间为1 TB。操作系统为Windows 11(64位),同时配备了CUDA12.6库,安装了Anaconda 3.5环境。GPU为配有8 GB显存的NVIDIA GeForce RTX 4060,使用Python作为编程语言。

算法性能评估体系采用6项核心指标进行综合度量:精确度(*Precision*)、召回率(*Recall*)、平均精确度均值(*mAP*50)、平均交并比(mean intersection over union,*mIoU*)、*F*1分数(*F*1 score)以及实时帧率(frames per second,*FPS*)。针对目标检测任务,主要选取精确度、召回率以及*mAP*50作为评价准则,而图像分割任务则采用*mIoU*作为核心评估指标。为全面评估模型效率特性,引入实时帧率与参数量大小两项量化指标。

3.2 模型训练策略

多任务视觉感知模型建立在YOLOv5紧凑版之上,并包含一个额外的可遍历区域分割头。为了改善训练过程,采用了原始YOLOv5实现中指定的几种网络模块。这些策略对于最大化模型效率以及确保模型在真实枇杷果园中的鲁棒性至关重要。

马赛克增强方法是视觉感知模型训练中使用的一种关键方法论。马赛克增强涉及将4张训练图像合并为一张图像,从而使模型能够在每个训练周期中从更广泛的多样化样本中获取知识。这种方法通过让模型接触到不同的物体尺度、纵横比和空间配置,增强了模型的泛化能力。由于其高效性和自适应调整学习率的能力,采用了Adam优化器。Adam算法通过根据梯度的一阶和二阶矩调整每个参数的学习率,结合了AdaGrad和RMSProp优化方法的优点。

此外,在初始训练期间实施了一种热身方法[29]。这种热身方法在几个周期内逐步将学习率从一个低值提高到初始学习率。在训练初始阶段,权重尚未得到精细调整,这一策略有助于减轻模型在训练初期的发散。

利用生成对抗网络(GANs)中应用的训练方法,

采用了一种不同的训练方法来训练多任务感知模型的两个输出头。该策略采用了一种在目标检测头和分割头之间切换训练重点的方法。通过交替的方式独立优化每个输出头，模型可以有效地学习这两个任务。该方法确保模型训练过程中，充分利用在目标识别和分割这两个关键任务中的性能。

3.3　单任务与多任务模型试验对比

遵循YOLOv5模型训练时的图像缩放策略，训练数据集的图像被缩放到384像素×640像素，然后基于交替更新策略对模型进行训练。选择了几种专注于目标检测或语义分割的模型，与AgriYOLOP模型进行比较。SSD[30]和RetinaNet是计算效率高的一阶段目标识别方法，而Faster-RCNN则是二阶段目标检测网络的典范。

为了全面比较准确性和效率，选取具有不同骨干网络的目标模型，包括SSD-VGG16、RetinaNet-ResNet50和Faster-RCNNResNet-50，并在自定义枇杷果园数据集上进行了重训练。同样，出色的语义分割模型也采用了不同的编码器，如U-Net-MobileNet、LinkNet-VGG16和DeepLabv3-ResNet50。

3.3.1　目标检测任务试验对比

枇杷果园行间目标检测任务的评估指标如精确度、召回率和*mAP*50被用来量化模型在从枇杷果园收集的测试数据集上的表现。最终不同模型训练的结果参数比对如表1所示，从表中可以看出，AgriYOLOP模型在果园目标检测任务方面准确率为92.7%、召回率为94.6%、*mAP*50结果为96.7%，其综合性能优于SSD、RetinaNet和Faster-RCNN模型。

表1　枇杷果园行间目标检测任务对比结果

Table 1　Comparative results of object detection tasks in loquat orchard inter-row spaces

模型	*Precision*/%	*Recall*/%	*mAP*50/%	*FPS*/(f•s^{-1})	*Param*/M
AgriYOLOP	92.7	94.6	96.7	69	14
SSD	69.8	62.3	66.1	107	24
RetinaNet	89.3	70.4	85.4	25	32
Faster-RCNN	71.1	57.1	62.7	24	41

如图4所示，该模型在远离相机的小树干上的准确性更高，并且还能检测出行间可能存在的障碍物（纸箱）。此外，多任务学习策略并没有降低模型相对于YOLOv5的目标检测性能。相反，AgriYOLOP模型在某些场景中表现出更高的准确性，这得益于多任务学习的信息共享和交替训练策略。

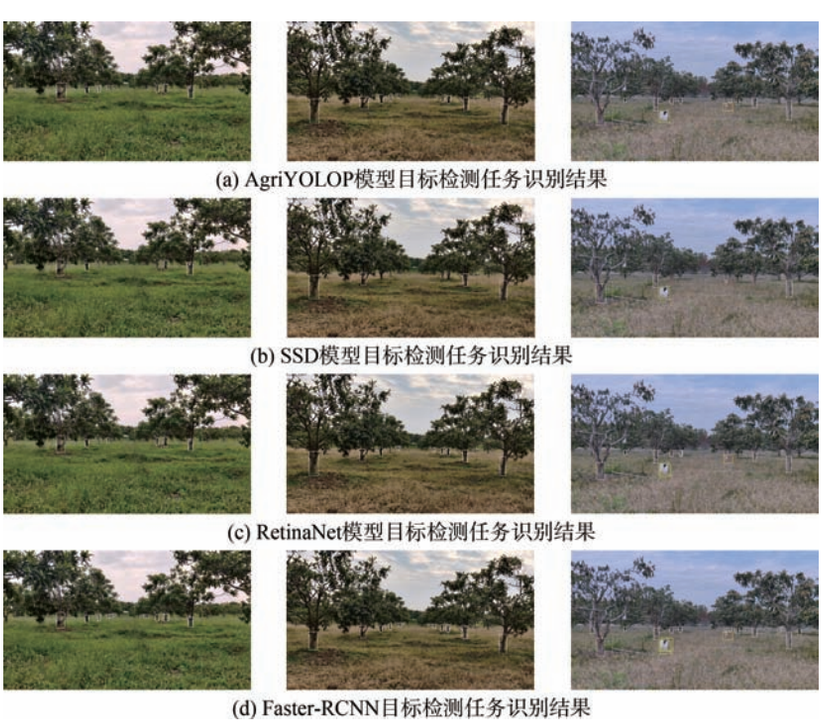


图4　果园行间不同模型目标检测任务试验对比图

Figure 4　Comparative experimental results of multi-model object detection in orchard inter-row spaces

3.3.2 分割任务试验对比

在枇杷果园果树内行场景中,由于被果树树干包围,多任务感知模型在分割任务的表现优于单任务模型。AgriYOLOP模型检测到的与其它语义分割模型对比,从图3的第1列和第2列分割的可视化结果看出,准确性远高于其他模型,这得益于树干检测任务提供的共享信息。分割任务对比结果如表2所示。

值得注意的是,树干特征检测为的边界界定提供了关键性判别依据,从而有效缓解了由光照条件及季节更替所导致的阴影与纹理特征变异对分割性能的干扰。同样,AgriYOLOP模型不会混淆存在障碍物的区域。如图5的最后1列所示,在包含非结构化障碍物的复杂场景中,模型能够准确区分障碍物(如纸箱),其分割输出严格遵循地形语义约束,确保分割结果的拓扑完整性。这种基于语义先验的判别机制显著提升了模型在农业非结构化环境下的场景适应性与分割鲁棒性。

表2 枇杷果园行间分割任务对比结果

Table 2 Comparative results of navigable region segmentation tasks in loquat orchard inter-row spaces

模型	*F*1 score	*Recall* /%	*mIoU* /%	*FPS* /(f•s^{-1})	*Param* /M
AgriYOLOP	98.3	99.1	98.1	69	14
U-Net	95.9	95.8	92.2	74	7
LinkNet	95.5	95.2	91.6	54	16
DeepLabv3	95.8	95.8	92.1	30	39

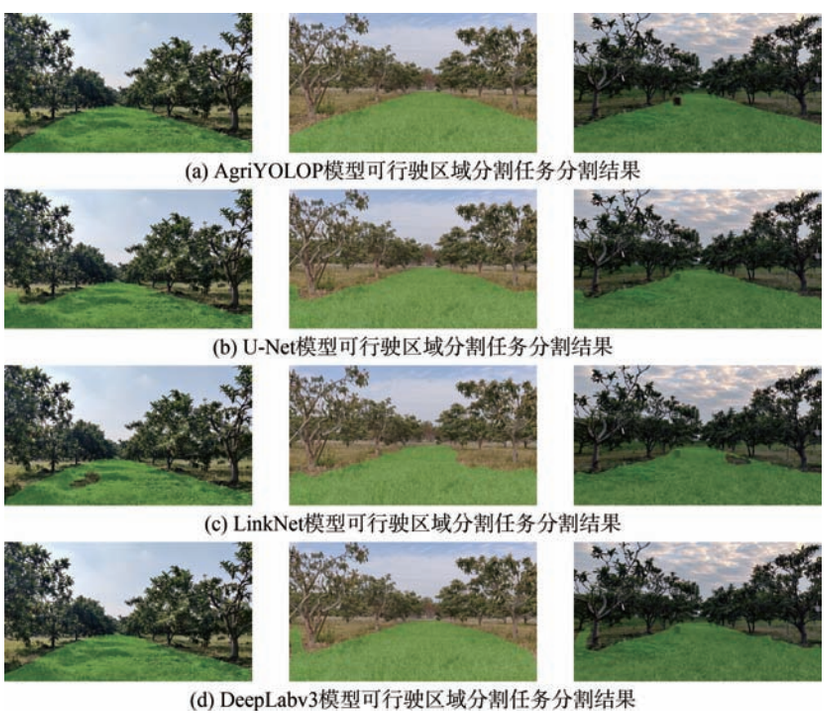


图5 果园行间不同模型目标检测任务试验对比图

Figure 5 Comparative experimental results of multi-model object detection in orchard inter-row spaces

3.4 多任务感知模型试验对比

为验证AgriYOLOP模型在多任务视觉感知性能,如表3所示,分别与Hybrid-Net和R-CNNP模型比较。HybridNets模型采用EfficientNet-B3作为提取图像特征的基础框架。通过利用Efficient-Net-B3的简化架构和增强的缩放技术,提高了效率。然而,这种效率的提升却伴随着准确性的降低,尤其是在对小物体进行识别的时候。相机视野中的小树干和其他障碍物识别,对于机器人路径规划和运动控制起到至关重要的作用[31]。

如图6所示,HybridNets模型在尝试检测较小物体时表现较差。R-CNNP是FasterRCNN的增强迭代版本,结合了YOLOP模型中相同分割头。尽管有所改进,R-CNNP在多个感知任务中与YOLOP相比仍显著不足。R-CNNP在目标检测方面的表现低于原始的Faster-RCNN。R-CNNP模型采用的基于区域的检测方法与基于语义分割方法有所不同。这种差异导致将两项视觉感知任务整合到单一模型中时效率降低。AgriYOLOP模型结合了多个解码器,以实现同时的目标检测和语义分割任务,模型能够更好地利用来自不同任务的信息,这与多任务学习模型的要求一致。其一致性通过使模型能够更有效地在任务之间共享学习到的特征,从而提高了模型的整体预测性能,还增强了模型在面对复杂背景和多目标时的鲁棒性。

表 3　枇杷果园行间多任务视觉感知对比结果

Table 3　Comparative results of multi-task visual perception in loquat orchard inter-row spaces

模型	目标检测任务			分割任务		
	Precision/%	*Recall*/%	*mAP*50/%	*F*1 score	*Recall*/%	*mIoU*/%
AgriYOLOP	92.7	94.6	96.7	98.3	99.1	98.1
HybridNets	88.6	87.1	89.5	97.3	97.0	96.6
R-CNNP	60.1	51.0	52.9	85.8	84.9	85.2

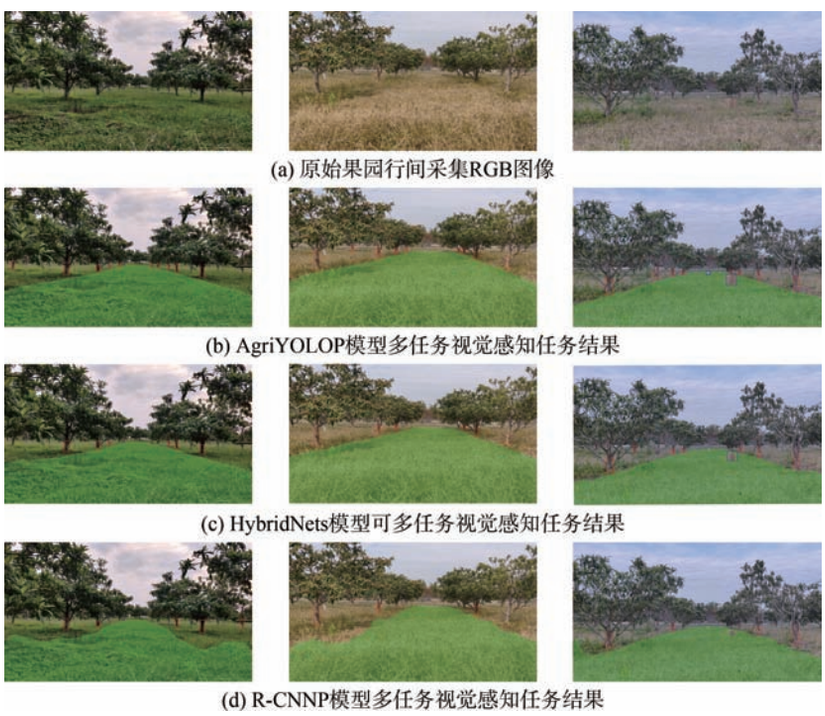


图 6　枇杷果园行间不同模型多任务视觉感知任务试验对比图

Figure 6　Comparative experimental results of multi-task visual perception in orchard inter-row spaces

4　结论

针对果园复杂环境下多任务视觉感知需求，提出基于 AgriYOLOP 的轻量化联合感知框架，通过多尺度特征共享与动态任务平衡机制，实现树干检测、障碍物识别与可行区域分割的高效协同。主要结论如下。

1）高性能实时感知：AgriYOLOP 在目标检测任务中取得 96.7% 的 *mAP*50，可行区域分割任务中 *mIoU* 达 98.1%，推理速度达 69 f/s（NVIDIA RTX 4060），较单任务模型（Faster R-CNN＋DeepLabv3）延迟降低 32.6%，参数量压缩至 14 M，显著优于现有农业多任务模型（如 HybridNets *mAP*50：89.5%，*mIoU*：96.6%）。试验表明，多任务特征复用机制可减少 28.4% 的重复计算量，验证了模型在嵌入式平台的部署潜力。

2）动态场景强鲁棒性：通过融合跨季节数据集与无锚检测技术，模型在逆光、树冠阴影等复杂光照条件下的目标检测精度波动小于 2.3%，可通行区域分割边界误差控制在合理范围内。消融试验进一步证明，多尺度特征融合机制使远处树干等小目标识别率大幅提升，有效缓解了农业非结构化场景中的特征退化问题。

3）技术延展性：本研究提出的动态损失加权策略为农业多任务学习提供了可扩展框架，支持灵活接入新任务（如植保喷雾、果实计数、病害监测）。田间实测表明，融合深度相机数据后，导航路径规划误差显著降低，满足果园机器人作业精度要求。

未来工作将主要聚焦在果园场景动态变化、边缘计算优化和多模态协同方向，进一步解决果园环境感知局限，实现端侧的实时推理和长期作业稳定性。

参　考　文　献

[1] 张隆梅，刘岗微，齐彦栋，等. 农业机械无人驾驶系统关键技术研究进展（英文）[J]. 智能化农业装备学报（中英文），2022，3(1)：27－36.
ZHANG Longmei，LIU Gangwei，QI Yandong，et al. Research progress on key technologies of agricultural machinery unmanned driving system [J]. Journal of

Intelligent Agricultural Mechanization, 2022, 3(1): 27－36.

[2] 刘羊, 刘宇洋, 陈江春, 等. 标准化果园多功能作业平台关键技术研究现状与趋势[J]. 智能化农业装备学报(中英文), 2024, 5(1): 31－39.
LIU Yang, LIU Yuyang, CHEN Jiangchun, et al. Research status and trend of key technologies of standardized orchard multi-functional operation platform [J]. Journal of Intelligent Agricultural Mechanization, 2024, 5(1): 31－39.

[3] 李建平, 李绍波, 杨欣, 等. 苹果园生产管理智能机械化技术与装备研究进展与展望[J]. 智能化农业装备学报(中英文), 2024, 5(4): 66－83.
LI Jianping, LI Shaobo, YANG Xin, et al. Research progress on intelligent mechanization technology and equipment for apple orchard production management [J]. Journal of Intelligent Agricultural Mechanization, 2024, 5(4): 66－83.

[4] 张翼. 中国式现代化的推进、人口转型与弹性退休制度[J]. 北京大学学报(哲学社会科学版), 2024, 61(5): 128－136.
ZHANG Yi. Promotion of chinese-style modernization, population transformation and flexible retirement system [J]. Journal of Peking University (Philosophy and Social Sciences), 2024, 61(5): 128－136.

[5] 王文明, 王志强, 韩明明, 等. 我国智慧果园关键技术装备研究应用现状与展望[J]. 中国果树, 2024(7): 143－149.

[6] 朱云, 凌志刚, 张雨强. 机器视觉技术研究进展及展望[J]. 图学学报, 2020, 41(6): 871－890

[7] 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.

[8] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]// European conference on computer vision. Cham: Springer International Publishing, 2020.

[9] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [J]. IEEE, 2017: 6517－6525.

[10] BENJUMEA A, TEETI I, CUZZOLIN F, et al. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles [J]. 2021.

[11] HUANG Z, WANG J, FU X, et al. DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection [J]. Information Sciences, 2020, 522.

[12] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834－848.

[13] WENG W, ZHU X. INet: Convolutional networks for biomedical image segmentation [J]. IEEE Access, 2021, 9: 16591－16603.

[14] LI H, XIONG P, FAN H, et al. Dfanet: Deep feature aggregation for real-time semantic segmentation [C]// Proceedings of the IEE E/CVF conference on computer vision and pattern recognition, 2019: 9522－9531.

[15] GUO Y, ASHMAWY K, HUANG E, et al. Under the hood of uber atg′s machinelearning infrastructure and versioning control platform for self-driving vehicles [J]. Uber Engineering, March, 2020.

[16] SISTU G, LEANG I, YOGAMANI S. Real-time joint object detection and semantic segmentation network for automated driving [J]. 2019.

[17] SENER O, KOLTUN V. Multi-task learning as multi-objective optimization [J]. Advances in neural Information Processing Systems, 2018, 31.

[18] STANDLEY T, ZAMIR A, CHEN D, et al. Which tasks should be learned together in multi-task learning? [C]// International conference on machine learning. PMLR, 2020: 9120－9132.

[19] WU D, LIAO M W, ZHANG W T, et al. Correction to: YOLOP: You only look once for panoptic driving perception [J]. Machine Intelligence Research, 2023, 20(6): 952－952.

[20] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017(99): 2999－3007.

[21] CRAWSHAW M. Multi-task learning with deep neural networks: A survey [J]. arXiv preprint arXiv: 2009.09796, 2020.

[22] LIU S, JOHNS E, DAVISON A J. End-to-end multi-task learning with attention [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1871－1880.

[23] ZAMIR A R, SAX A, SHEN W, et al. Taskonomy: Disentangling task transfer learning [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3712－3722.

[24] VANDENHENDE S, GEORGOULIS S, VAN GANSBEKE W, et al. Multi-task learning for dense prediction tasks: A survey [J]. IEEE transactions on Pattern Analysis and Machine Intelligence, 2021, 44(7): 3614－3633.

[25] JOCHER G, CHAURASIA A, STOKEN A, et al. Ultralytics/yolov5: v7.0-YOLOv5 SOTA real-time instance segmentation [J]. Zenodo, 2022.

[26] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. arXiv preprint arXiv: 2004.10934, 2020.

[27] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature

pyramid networks for object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117－2125.

[28] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8759－8768.

[29] GOYAL P, DOLLÁR P, GIRSHICK R, et al. Accurate, large minibatch sgd: Training imagenet in 1 hour [J]. arXiv preprint arXiv:1706.02677, 2017.

[30] WEI L, DRAGOMIR A, DUMITRU E, et al. SSD: Single shot multibox detector [J]. Springer, Cham, 2016.

[31] VU D, NGO B, PHAN H. HybridNets: End-to-end perception network [J]. 2022.

Multi-task joint perception framework for autonomous navigation in orchard robotics

ZHANG Jinguo[1, 2], CAI Jianfeng[1, 2], JIANG Rongrong[3], YU Shanshan[4], WANG Pengbo[1, 2*]

(1. *College of Mechanical and Electrical Engineering, Soochow University, Suzhou* 215137, *China*; 2. *Jiangsu Key Laboratory of Embodied Intelligent Robot Technology, Suzhou* 215137, *China*; 3. *Suzhou Caoyang Ecological Agriculture Development Co., Ltd., Suzhou* 215143, *China*; 4. *Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing* 210014, *China*)

Abstract: Orchard operational scenarios present significant challenges for visual perception, including high vegetation heterogeneity, dynamic lighting variations, and diverse target morphologies. Traditional single-task visual perception models suffer from low feature reusability and high computational redundancy, thereby inadequately addressing real-time environmental perception demands for agricultural robots. This study proposes AgriYOLOP, a lightweight multi-task collaborative perception framework specifically designed for orchard environments. Through a systematic reconstruction of the YOLOP architecture, AgriYOLOP incorporates an efficient backbone network, enhanced anchor-free detection techniques, feature pyramid networks (FPN), path aggregation networks (PAN), and task-adaptive loss function weighting strategies. This framework faciliates parallel collaborative processing of three critical percetption tasks: trunk detection, obstacle recognition, and traversable region segmentation. The proposed framework was validated on a self-constructed orchard dataset comprising 4 765 images (1 280 pixels×720 pixels), captured across diverse seasons, lighting conditions, and vegetation growth stages. Experimental results demonstrate that AgriYOLOP achieves 92.7% *precision*, 94.6% *recall*, and 96.7% *mAP*50 in object detection tasks, along with 98.3% *recall*, 99.1 *F*1 score, and 98.1% *mIoU* in traversable region segmentation. Deployed on an NVIDIA RTX 4060 platform, the model attains 69 f/s real-time inference speed with only 14 M parameters. Comparative experiments reveal that the multi-task collaborative architecture significantly enhances feature-sharing efficiency, reducing inference latency by 32.6% compared to single-task models while improving robustness to illumination and seasonal variations. This approach effectively mitigates the conventional trade-off between target detection accuracy and semantic segmentation efficiency encountered in real-time agricultural robotic applications. The study provides a high-precision, low-latency real-time perception solution for autonomous orchard robot navigation.

Keywords: multi-task learning; orchard environment perception; target detection; semantic segmentation; agricultural robot

Foundation items: National Key Research and Development Program of China (2022YFB4702202); Integrated Pilot Project for Agricultural Machinery Research, Development, Manufacturing, Promotion and Application of Jiangsu Provincial Department of Agriculture and Rural Affairs (JSYTH07)

First Author: ZHANG Jinguo, E-mail:1432263736@qq.com

* **Correspondence Author:** WANG Pengbo, E-mail: zhulixue@zhku.edu.cn