doi:10.6041/j.issn.1000-1298.2024.11.018

基于 OrchardYOLOP 的火龙果园多任务视觉感知方法

赵文锋　黄袁爵　钟敏悦　李振源　罗梓涛　黄家俊
（华南农业大学电子工程学院（人工智能学院），广州 510642）

摘要：现代果园机器人面临复杂环境、光线多变和非结构化环境等问题，需要高效处理大量环境信息，而传统顺序执行多个单一任务的算法受到计算能力的限制，难以满足现代果园机器人的需求。本文针对火龙果园环境中自动驾驶机器人处理多任务时所面临的实时性和准确性要求，基于 YOLOP 模型引入了焦点融合高效卷积模块，并采用 C2F 和 SPPF 模块，同时优化了分割任务的损失函数，从而构建出 OrchardYOLOP 模型。实验结果表明，在目标检测任务上的精确度达到 84.1%；在可行驶区域分割任务上的 mIoU 达到 89.7%；在果树区域分割任务上的 mIoU 提高到 90.8%；推理速度达到 33.33 f/s，而参数量仅有 9.67×10^6。与 YOLOP 模型相比，不仅在速度上满足了实时性要求，而且准确性上也有显著提升。这解决了火龙果园多任务视觉感知中的关键问题，为非结构化环境下的多任务自动驾驶视觉感知提供了一种有效的解决方案。

关键词：火龙果园；多任务；视觉感知；语义分割；目标检测；YOLOP
中图分类号：TP242.6　　**文献标识码**：A　　**文章编号**：1000-1298(2024)11-0160-11　　**OSID**：

Multi-task Visual Perception Method in Dragon Orchards Based on OrchardYOLOP

ZHAO Wenfeng　HUANG Yuanjue　ZHONG Minyue　LI Zhenyuan　LUO Zitao　HUANG Jiajun
(*College of Electronic Engineering* (*College of Artificial Intelligence*),
South China Agricultural University, *Guangzhou* 510642, *China*)

Abstract: In the face of challenges such as complex terrains, fluctuating lighting, and unstructured environments, modern orchard robots require the efficient processing of a vast array of environmental information. Traditional algorithms that sequentially execute multiple single tasks are limited by computational power which are unable to meet these demands. Aiming to address the requirements for real-time performance and accuracy in multi-tasking autonomous driving robots within dragon fruit orchard environments. Building upon the YOLOP, focus attention convolution module was introduced, C2F and SPPF modules were employed, and the loss function for segmentation tasks was optimized, culminating in the OrchardYOLOP. Experiments demonstrated that OrchardYOLOP achieved a precision of 84.1% in target detection tasks, an mIoU of 89.7% in drivable area segmentation tasks, and an mIoU increased to 90.8% in fruit tree region segmentation tasks, with an inference speed of 33.33 frames per second and a parameter count of only 9.67×10^6. Compared with the YOLOP algorithm, not only did it meet the real-time requirements in terms of speed, but also it significantly improved accuracy, addressing key issues in multi-task visual perception in dragon fruit orchards and providing an effective solution for multi-task autonomous driving visual perception in unstructured environments.

Key words: dragon orchard; multi-task; visual perception; semantic segmentation; object detection; YOLOP

0　引言

2022 年，中国水果产量达到 3.12×10^8 t，同比增长 4.42%[1]。这不仅满足了国内市场的需求，还使中国成为世界上主要的水果出口国之一。然而，尽管产量巨大，但是随着我国农业人口的减少特别

收稿日期：2024-06-04　修回日期：2024-08-07
基金项目：国家重点研发计划项目(2023YFD1400700)
作者简介：赵文锋(1979—)，男，副教授，博士，主要从事农业机器人研究，E-mail：zwf555@scau.edu.cn

是年轻一代劳动力的流失[2]，以及农业智能化程度仍然较低[3]，这些都限制了产业的进一步发展和效率提升。在果园中引入自动化机器人，能显著提升工作效率、减少人力依赖及损耗，提高果品品质和一致性，从而增强经济效益[4-5]。然而，果园机器人通常受限于其搭载的计算资源，而同时执行多个单一任务的算法会对算力提出更高的要求[6]。这些任务彼此间存在内在联系。鉴于此开发能够在一个统一的框架下同时处理多种感知算法，从而降低对算力的需求并提高处理速度，这对于果园机器人在复杂环境中的自主操作非常有价值。

在机器视觉研究领域，目标检测与图像分割是 2 项核心任务，它们是机器人理解复杂视觉场景的基础[7]。针对目标检测的模型有 Faster R-CNN[8]、DETR[9]和 YOLO[10-14]系列等；对于图像分割，TransUNet[15]、UNet[16]和 PSPNet[17]等在语义分割任务表现出色，适用于可行驶区域分割和果树区域分割。尽管上述算法在其特定任务上展现了优异性能。但在资源有限的非结构化农业场景中，顺序执行多个任务会导致处理时间显著增加，尤其是在需要快速、准确获取多种信息的自动化采摘场景下。

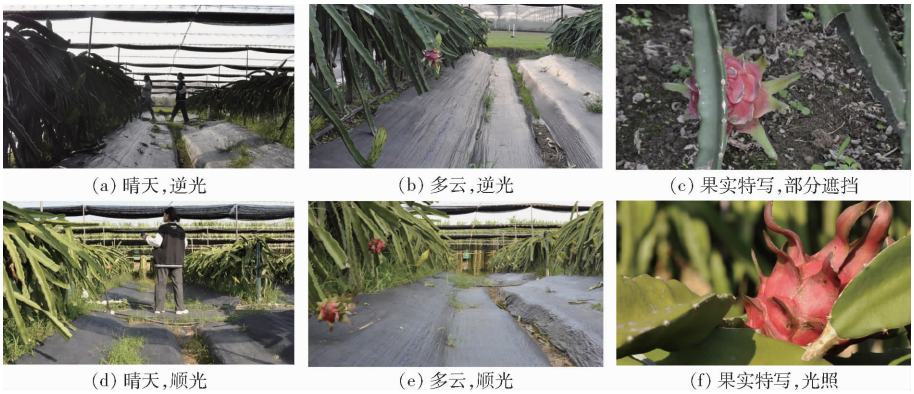
在结构化环境中，多任务自动驾驶视觉感知网络已经取得一定进展，例如，Multitask V3[18]、HybridNets[19]和 YOLOP[20]等，通过共享特征提取网络，能够在减少计算资源的同时，保持对各任务较高准确度。然而，在非结构化环境下，多任务视觉感知网络的研究尚处于起步阶段。这些环境的固有不确定性和复杂性对多任务感知系统提出了更高的要求。传统的视觉感知算法，通常基于预设的规则和模式，面对非结构化道路的不规则性和缺乏标准化标识，往往表现出适应性不足。例如，自动驾驶机器人在非结构化道路上必须能够应对不平整的路面和缺失的道路标识[21]，这些因素可能严重干扰其正常行驶。同时目前还缺乏一种能够综合感知作业环境及相关任务信息的视觉感知方法，这一方法对于提升自动驾驶机器人在复杂环境中的作业效率至关重要。

本文旨在克服火龙果园非结构化环境带来的问题，通过改进 YOLOP 模型，提出 OrchardYOLOP（Orchard you look only once for panoptic）模型。针对火龙果园环境下多任务视觉感知问题[22]（针对行人和成熟果实的目标检测、可行驶区域分割和果树区域分割）进行优化，以期提供一种快速、可靠和有效的技术方案。

1　材料与方法

1.1　图像数据获取

果园图像数据于 2023 年 10 月采集自广东省佛山市火龙果种植园。拍摄设备为尼康 D90 型单反相机，图像格式为 JPG，分辨率为 2 144 像素 ×1 424 像素。在采集数据时，相机距离地面约 1 m，每秒拍摄 1 幅图像，确保拍摄方向始终指向道路正前方。为了获取更多质量较高且细节丰富的火龙果图像，图像数据也采集了各种火龙果果实特写图像。考虑到不同光照和环境条件对图像的影响，图像数据采集在不同时间段（上午、傍晚）和不同光照条件（晴天、阴天、顺光、逆光）进行，共计得到 688 幅原始图像。图 1 展示了部分采集到的图像样本。



(a) 晴天，逆光　(b) 多云，逆光　(c) 果实特写，部分遮挡

(d) 晴天，顺光　(e) 多云，顺光　(f) 果实特写，光照

图 1　火龙果园图像示例

Fig. 1　Fire dragon orchard image examples

1.2　果园多任务数据集制作

通过相机直接获取的原始图像分辨率较高，在训练时会占用过多显存，降低训练速度，同时图像设有语义标签，需要预先进行标注，才能传入神经网络进行训练。为了减少训练时间，将原始图像分辨率等比例缩放至 1 280 像素 ×720 像素。

为了提高原始图像数据的标注效率，本研究采用了 label-studio-SAM 工具进行半自动化标注。label-studio-SAM 是一个基于 label-studio 开发的工具，它利用 SAM（Segment anything model）模型[23]进

行后端推理,自动识别并分割图像中的物体。针对图像中的可行驶区域与果树区域,主要采用掩码(mask)的形式进行精确标注。针对实际应用场景,为了确保在作业过程中能够有效规避人员并获取果园内成熟果实,采用标注框(bounding box)对行人和成熟果实进行标注。在标注完成后,可行驶区域和果树区域的标注信息以 PNG 格式的掩码文件存储,而目标检测任务的标注数据则按照 COCO 数据集格式导出,以便于后续的训练和验证工作。

针对现场采集的数据样本量不足问题,使用 Albumentations 库对数据集进行增强。Albumentations 库是一个专为图像数据增强设计的 Python 库[24],支持多种计算机视觉任务,如分类、分割和检测。每次增强会选择一种随机几何变换(平移、翻转、粗糙遮挡、网络畸变、像素遮挡等)和一种随机像素变换(高斯模糊、模糊、通道丢失、通道洗牌、失焦等)组合来对数据进行增强。增强后的图像共 6 880 幅,按照 8∶2 比例划分训练集和验证集。训练集用于训练模型,验证集用于训练过程中对模型参数进行调优。

在完成了对原始图像的缩放、标注以及数据增强后,面临的一个关键步骤是将处理后的数据集转换为适合 OrchardYOLOP 模型训练的数据格式。由于原模型采用 BDD(Berkeley DeepDrive)[25]数据集格式,为了确保数据的一致性和训练的顺利进行,编写了一个专门的 Python 脚本,用于将处理后的数据集转换为符合 BDD 数据集规范的数据集格式。将数据集划分为 4 个主要类别:图像(image)、目标检测标注(det_annotations)、可行驶区域分割标注(da_seg_annotations)以及果树区域分割标注(tree_seg_annotations)。每个类别都遵循 BDD 数据集的特定命名和存储规则。数据集可视化如图 2 所示。



(a) 原图 (b) 可行驶区域分割 (c) 果树区域分割 (d) 目标检测

图 2 3 种任务数据集可视化展示

Fig. 2 Visualization of datasets for three tasks

1.3 基于 OrchardYOLOP 视觉感知方法

OrchardYOLOP 模型是在 YOLOP 模型的基础上进行深度改进与优化,专为火龙果园环境下的多任务视觉感知任务(包括针对行人和成熟果实的目标检测、可行驶区域分割和果树区域分割)设计。该设计旨在减少计算消耗,并允许网络进行端到端的训练。骨干网络是高效层聚合网络(Efficient layer aggregation network, ELAN),针对快速高效的特征提取进行了优化。模型结构示意图如图 3 所示。

在骨干网络的第 1 层中使用焦点融合高效卷积(Focus attention convolution, FAC)模块替换原本的 Focus 模块作为骨干网络的第 1 层,新模块结合了深度可分离卷积模块和 CBAM[26]注意力机制,增强了模型对输入图像特征的捕获和筛选能力,可以通过自适应调整通道权重和利用注意力机制来优化特征表达,特别是在复杂背景下的目标识别和分割。

同时在骨干网络中使用 C2F 模块替换 BottleneckCSP 模块,旨在简化模型结构、降低计算复杂度、加速信息流动与融合过程,同时保持 CSP 结构的性能优势,在不牺牲或轻微牺牲精度的情况下提升效率和实用性。

颈部网络引入了快速空间金字塔池化(Spatial pyramid pooling-fast, SPPF)模块,这是空间金字塔池化(Spatial pyramid pooling, SPP)[27]的改进版本,保留了处理任意尺寸输入的灵活性,同时优化了计算效率,通过更高效的池化策略减少冗余计算,加快推理速度。

最后参考 HybridNets 的损失函数来增强模型中关于可行驶区域和果树区域分割相关功能的性能。

1.3.1 CBAM 模块

注意力机制作为一种有效的手段,已被广泛应用于计算机视觉任务中,用以引导模型聚焦于输入数据的关键区域,提升特征表达的针对性和有效性。在现有的工作里,虽然大部分注意力模块侧重于在高层次特征图上施加注意力,但对低层次特征图进行精细的注意力调控同样对提升模型性能具有重要意义。为此,本文采用了 CBAM(Convolutional block attention module)模块[26],该模块不仅在通道维度上对特征进行加权,还进一步引入了空间注意力机制,实现了对特征图的双重视觉注意力建模。

CBAM 模块的设计初衷在于同时捕获通道间和空间位置上的特征相关性,以增强模型对输入特征的全局理解和局部细节感知。具体而言,CBAM 模块包含通道注意力和空间注意力 2 个相互独立又互补的操作流程。图 4 展示了 CBAM 注意力结构图。

通道注意力(Channel attention, AC):首先,

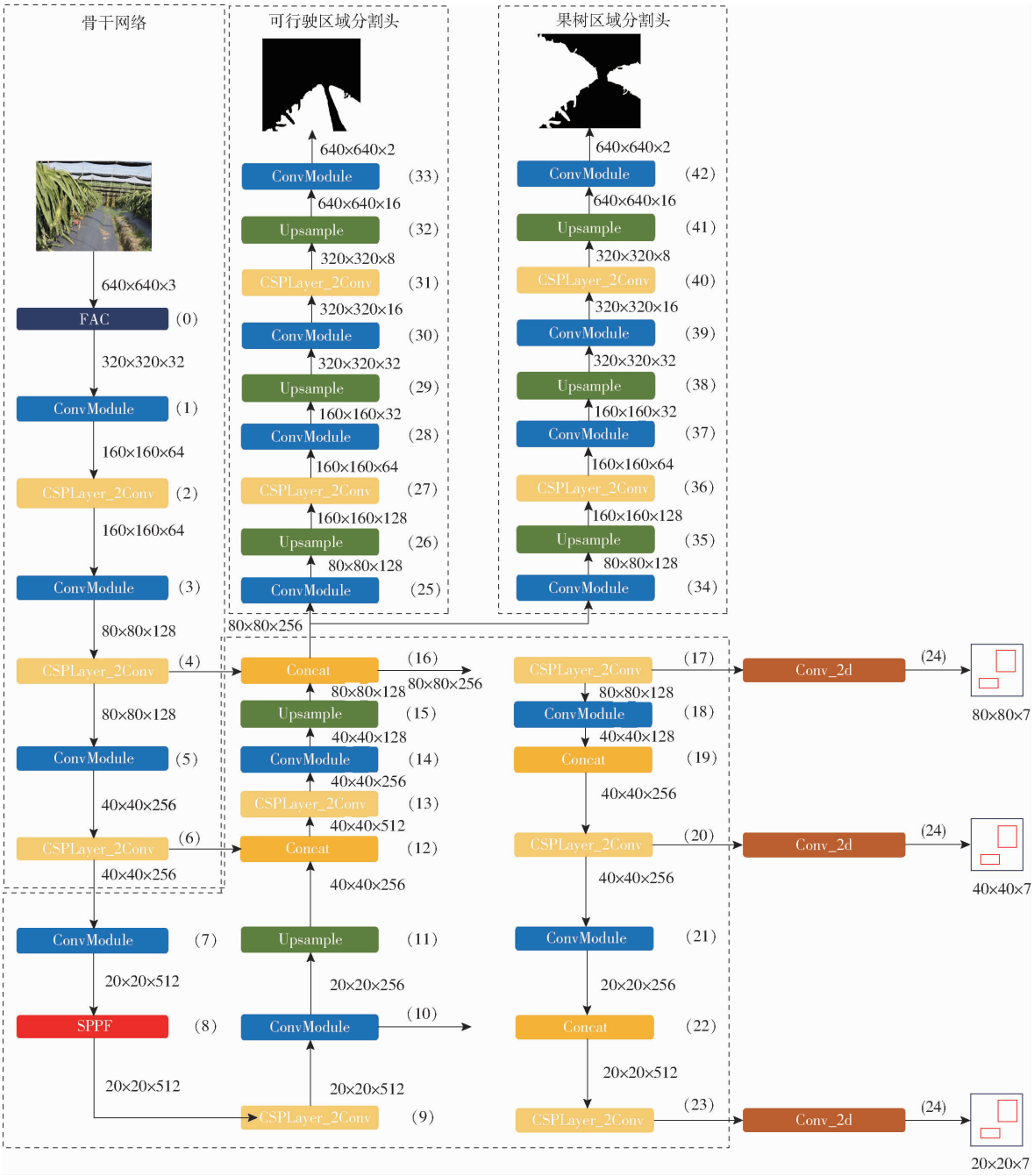


图 3　OrchardYOLOP 模型结构示意图

Fig. 3　Structure schematic of OrchardYOLOP

CBAM 通过计算全局平均池化（Global average pooling，GAP）和全局最大池化（Global max pooling，GMP）的结果，分别提取特征图的全局统计信息和显著特征。这些汇总信息随后被送入 2 个连续的全连接层（FC layers）进行非线性变换，以学习出反映各通道重要性的权重向量。最后，将这些权重向量逐元素地与原特征图进行按通道的乘法操作，实现对特征图在通道维度上的注意力加权[28]。这一过程有助于模型动态地强调或抑制某些特征通道，根据其对当前任务的贡献度进行自适应调整。

空间注意力（Spatial attention，SP）：在完成通道注意力之后，CBAM 进一步实施空间注意力机制。它首先沿通道方向对特征图进行平均池化和最大池化操作，分别得到反映每个像素位置上平均特征和显著特征的空间上下文信息。然后，将这 2 组空间上下文信息分别通过一维卷积层（1D convolutional layers）进行特征学习，生成空间注意力图。空间注意力图以像素为单位，表征了每个位置在空间维度上的关注度。最后，将空间注意力图与经过通道注意力处理后的特征图逐元素相乘，实现对特征图在空间维度上的注意力加权。这一过程使得模型能够对不同空间位置赋予不同的关注强度，凸显重要区域，抑制无关背景。

通过整合通道注意力和空间注意力，CBAM 模

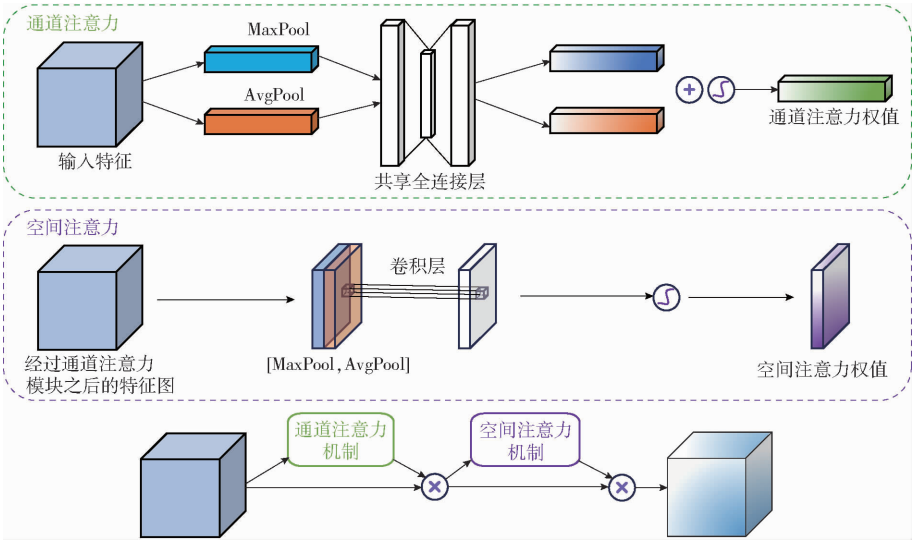


图 4 CBAM 结构图

Fig. 4 Structure diagram of convolutional block attention module

块能够在不显著增加模型复杂度的情况下,对特征图进行深度且细致的注意力调控。其双重视觉注意力机制不仅增强了模型对输入特征的整体把握,还提升对局部细节的敏感性,从而有助于在火龙果园多任务视觉感知任务中,尤其是目标检测、可行驶区域分割和果树区域分割任务上,达到更为精准和稳定的性能。

1.3.2 焦点融合高效卷积模块

焦点融合高效卷积模块是一个结合了深度可分离卷积和 CBAM 注意力机制的模块。相比于原本的 Focus 模块,它在保持计算效率的同时,增强了模型对空间信息的感知能力,并引入了自适应特征选择,从而可能提供更好的性能。该模块被置于骨干网络的第 1 层,对于输入图像进行初步的特征提取和处理从而提高网络精度。图 5 为焦点融合高效卷积模块。

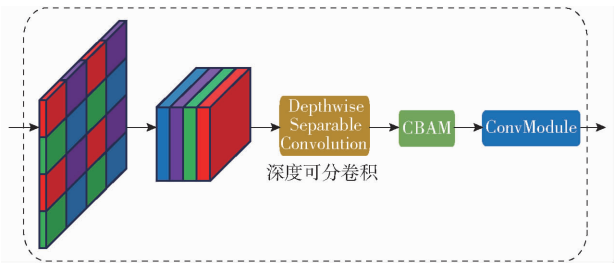


图 5 焦点融合高效卷积模块

Fig. 5 Structure of focus attention convolution module

该模块首先执行 Focus 操作,通过在输入的空间维度上进行切片操作来创建 4 个不同的部分,每个部分分别捕获原始输入的偶数行和偶数列、奇数行和偶数列、偶数行和奇数列、奇数行和奇数列的像素。这些部分随后沿着通道维度被连接起来,从而在不改变空间分辨率的情况下增加了通道的深度。这一策略有效地促进了网络对图像局部细节的捕捉能力。接下来,对经过 Focus 操作处理的图像应用深度可分离卷积,这是一种将传统卷积分解为深度卷积和逐点卷积的高效卷积方法。此操作不仅减少了计算成本和参数量,而且保持了良好的性能。随后,将深度可分离卷积的输出送入 CBAM 注意力机制,该机制能够使模型在空间和通道 2 个维度上识别出关键特征。这进一步增强了模型对输入数据中重要信息的关注。最终,通过标准卷积操作对经 CBAM 处理的特征进行进一步提炼,以提取有价值的信息并调整通道数量以满足后续网络层的需求。

1.3.3 C2F 模块

在骨干网络中使用 C2F 模块替换 BottleneckCSP 模块。C2F 模块通过简化设计和并行处理能力,显著提高了计算速度,这对于处理大规模图像数据至关重要。它通过减少内存使用和计算成本,提高了资源利用率,特别适合资源受限的环境。

首先,输入通过其首个卷积层,将输入特征图的通道数扩展为 2 倍的隐藏通道数,这一步骤为后续的特征处理奠定了基础。接着,模块通过一系列瓶颈单元,进一步对特征进行细致的提炼和优化。每个瓶颈单元都是一个独立的特征处理单元,它们串联起来,共同作用于输入特征图的一部分,以此来增强模型对特征的学习和表达能力。在这个过程中,C2F 模块的独特之处在于其使用 split 操作,这些操作将首次卷积的输出在通道维度上均匀分割,然后分别送入瓶颈单元。这种方法不仅提高了数据处理的并行性,还确保了特征图的每部分都能得到充分学习和利用。最后,经过一系列瓶颈单元处理后,所有特征图在通道维度上被合并,并通过第 2 个卷积层进行整合。这不仅调整了特征图的通道数以适应

后续网络层的需求，还进一步提炼了特征，确保了网络能够从输入数据中提取出最有价值的信息。图 6 所示为 C2F 模块示意图。图中，h、w 分别表示高度像素数、宽度像素数，c_{in} 为输入通道数，c_{out} 为输出通道数，n 为重复次数。

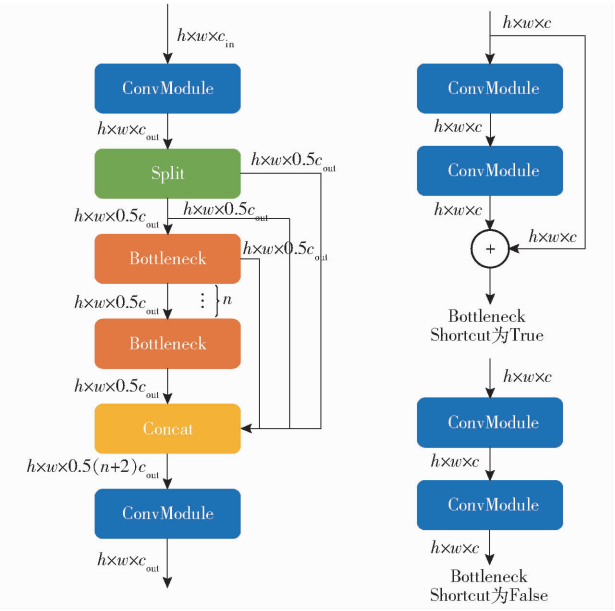


图 6　C2F 模块

Fig. 6　C2F module

1.3.4　SPPF 模块

颈部网络中使用了 SPPF 模块，它是 SPP[27] 的一种快速变体。SPPF 模块通过简化的池化操作，有效地减少了计算量，同时保持了特征融合的效果。图 7 所示为 SPPF 模块示意图。

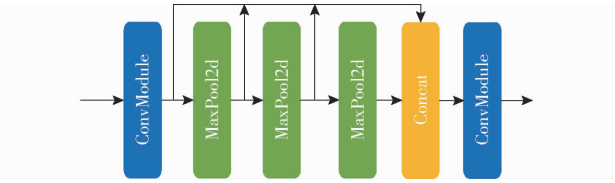


图 7　SPPF 模块

Fig. 7　SPPF module

首先，SPPF 模块通过其第 1 个卷积层，将输入特征图的通道数减半，这为后续的多尺度特征提取奠定了基础。然后，模块利用最大池化层，对特征图进行多次池化操作，以捕获不同尺度的上下文信息。这一系列池化操作能够在不同的空间分辨率上提取特征，从而增强模型对于尺度变化的适应性。在这个过程中，SPPF 模块的独特之处在于其高效的池化策略。通过对同一特征图进行连续池化操作，并将这些池化后的特征图与原始特征图进行拼接，SPPF 模块能够在极小的计算代价下模拟传统 SPP 中的多尺度池化效果。最后，经过池化操作的特征图被拼接在一起，并通过第 2 个卷积层进行整合。这一步骤不仅调整了特征图的通道数以适应后续网络层的需求，还进一步提炼了特征，确保了网络能够从输入数据中提取出最有价值的信息，同时保持了特征的空间分辨率。

1.3.5　检测头

本文采用基于锚点的多尺度目标检测策略。具体来说，利用了路径聚合网络（Path aggregation network，PAN）[29]，这是一种从底层到高层的特征金字塔（FPN）[30] 结构。与 FPN 不同，PAN 不仅自顶向下传递语义信息，还自底向上传递精确的定位信息。这种双向特征流动实现了特征的有效融合，从而在 PAN 生成的多尺度特征图上进行目标检测。

在多尺度特征图上为每个单元格分配 3 种不同纵横比的预设锚点（Anchor boxes）[31]。检测头负责预测锚点位置的偏移量、目标的尺寸缩放比例，以及各个类别的概率和置信度。这种方法不仅提高了检测的准确性，也增强了模型对不同尺度目标的适应能力。

1.3.6　可行驶区域分割头和果树区域分割头

可行驶区域分割头和果树区域分割头采用相同的网络结构，首先将 FPN 的底层输入传递给分割分支，其大小为（$w/8$，$h/8$，256）。经过 3 次上采样过程，将输出特征图恢复到（w，h，2）的大小，表示输入图像中每个像素对可行驶区域（或果树区域）与背景的概率。此外，上采样层中使用最近插值法代替反卷积方法来减少计算成本。

1.3.7　损失函数

由于网络中有 3 个解码器，所以多任务损失包含 3 部分。分别是目标检测的损失 L_{det}、可行驶区域的损失 L_{seg_da}、果树区域的损失 L_{seg_tree}。

对于目标检测的损失 L_{det} 由 3 部分组成，即

$$L_{det} = \alpha_1 L_{class} + \alpha_2 L_{obj} + \alpha_3 L_{box} \tag{1}$$

对于 L_{class} 和 L_{obj} 都使用焦点损失[32]，用于减少分类良好的样本的损失，从而使得网络更专注于难分类的样本。分类损失 L_{class} 负责惩罚分类错误，而对象损失 L_{obj} 用于预测对象置信度。这 2 个损失都是通过焦点损失来实现的。L_{box} 是预测的置信度，通过考虑重叠度来表示预测结果和真实值之间的相似性，考虑了预测框与真实值之间的距离、重叠度、尺度相似性和横纵比。系数 α_1、α_2 和 α_3 都是用来平衡损失函数的超参数。

本研究改进了分割任务损失函数，具体如下，果树区域分割由 3 部分组成，即

$$L_{seg_tree} = \beta_1 L_{Tversky} + \beta_2 L_{Focal} + \beta_3 L_{Jaccard} \tag{2}$$

$L_{Tversky}$[33] 用于解决数据不平衡问题，并在精度和召回率之间实现更好的权衡，L_{Focal}[32] 旨在最小化像素之间的分类误差，重点关注硬标签。$L_{Jaccard}$ 用于

度量预测和真实值分割掩码之间相似性。系数 β_1、β_2 和 β_3 都是用来平衡损失的超参数。

可行驶区域分割任务的损失函数由 2 部分组成,即

$$L_{\text{seg_da}} = \gamma_1 L_{\text{Tversky}} + \gamma_2 L_{\text{Focal}} \tag{3}$$

式中系数 γ_1 和 γ_2 是用于平衡损失的超参数。

最终损失函数 L_{all} 是 3 部分的加权和,即

$$L_{\text{all}} = \delta_1 L_{\text{det}} + \delta_2 L_{\text{seg_da}} + \delta_3 L_{\text{seg_tree}} \tag{4}$$

式中 δ_1、δ_2、δ_3——用于平衡检测损失和分割损失的超参数

2 实验

2.1 实验运行平台

本实验计算环境为:操作系统 Windows 11(64 位)、配备 Anaconda 3.5 环境和 CUDA 12.4 库。实验平台搭载了 Intel Core i5 - 11400F 处理器(主频 2.60 GHz),配备 32 GB 运行内存和 1 TB 的固态存储。图形处理单元为配有 24 GB 显存的 NVIDIA GeForce RTX 3090 GPU。实验采用了开源的深度学习框架 PyTorch 1.7.1,并使用 Python 3.8 作为编程语言。

2.2 评价指标

采用 4 个主要指标来评估算法的性能:精确度(Precision, P)、召回率(Recall, R)、平均精确度(mean Average Precision, mAP50)和平均交并比(mean Intersection over Union, mIoU)。对于检测任务使用精确度、召回率和 mAP50 作为评价指标;而对于分割任务,则采用 mIoU 作为评价指标。

mAP50 的计算较为复杂,具体步骤如下:首先,对于每个类别,根据预测的置信度将预测结果进行排序。接着,计算在不同置信度阈值下的精确度和召回率。然后,计算精确度-召回率曲线下的面积,即为该类别的 AP50。最终,计算所有类别的 AP50 的平均值,得到 mAP50。

除此之外,模型性能还通过实时帧率(FPS)和检测运行时间来评估。FPS 衡量的是摄像头将图像送入网络模型进行实时处理的速度,而检测运行时间则是基于连续识别 20 幅图像所需的平均时间来计算的。

2.3 训练过程

以 OrchardYOLOP 作为检测模型框架。在训练过程中,为每个 GPU 配置了 32 个样本的批量大小并设定动量为 0.937,以确保权重更新的连贯性和稳定性。初始学习率设置为 0.001,以促进模型在 700 个训练周期内的稳健学习。此外,选择 Adam 优化器[34]并设定了 0.000 5 的权重衰减系数,这有助于模型在面对未知数据时展现出更强的泛化能力。

在训练策略方面,采用了预热期学习率递增的方法,持续 3 个周期,以平滑地引导模型进入训练状态。为了更精确地调整模型,实施了交替优化策略,允许模型在单一任务上进行深入训练。具体来说,模型将在可行驶区域分割、果树区域分割和目标检测这 3 个单一任务上各自进行 200 个训练周期的专注训练。在这 600 个周期的基础上,模型将进入最后 100 个周期的综合训练阶段,此时 3 个任务将被同时训练,以实现任务间的协同和性能的最优化。每完成一个周期,模型都会在验证集上进行评估,确保每一步训练都是有效的,从而保证了整个训练过程的质量和模型的可靠性。

2.4 消融实验

为验证所提出的改进措施对于提升 OrchardYOLOP 模型在复杂场景下小目标检测、可行驶区域分割和果树区域分割性能的有效性,进行了一系列消融实验。基于原始 YOLOP 模型,分别构建了 3 种变体,每一种变体均添加了一项改进措施,以量化该措施对模型整体性能的贡献。具体而言,这些变体分别为:

(1)添加焦点融合高效卷积模块:第 1 层添加了焦点融合高效卷积模块,来代替原始 Focus 模块。

(2)替换 C2F 模块和 SPPF 模块:将 C2F 模块替换原始的 BottleneckCSP 模块,颈部网络中的 SPP 模块替换为 SPPF 模块。

(3)替换分割任务损失函数:采用改进分割损失函数来代替原本的损失函数。

表 1 为消融实验性能指标对比,从消融实验来看,添加焦点融合高效卷积模块虽然对分割任务的效果有着轻微下降,但是仅增加极小参数量和计算量,对于目标检测任务精确度提升非常明显。使用 C2F 模块和 SPPF 模块虽然对于各个任务的性能提升不算很大,但是可以有效地减少模型的参数量和计算量。替换分割任务损失函数对于提升分割任务的性能较为明显。

2.5 单分割头与双分割头对比实验

为了验证使用 2 个分割头分别处理对应任务的优势大于使用 1 个分割头来同时处理 2 个任务,本文设计了单分割头与双分割头的对比实验。具体为:首先写脚本处理分割任务数据集,将可行驶区域灰度值设为 0,果树区域设为 1,背景设为 255。原本为 2 个任务的 2 幅掩码图像,合并为 1 幅掩码图像。接着更改分割任务类别数量的相关配置,修改读取数据集和验证分割任务的相关代码、删除多余

表 1　消融实验性能指标对比

Tab. 1　Comparison of performance of ablation experiment

模块	目标检测精确度/%	可行驶区域分割 mIoU/%	果树区域分割 mIoU/%	参数量	计算量
YOLOP(原始模型)	68.2	86.0	88.7	9.80×10^{6}	1.861×10^{10}
FAC	86.3	85.7	88.1	9.81×10^{6}	1.898×10^{10}
FAC + C2F + SSPF	84.1	86.4	88.9	9.67×10^{6}	1.840×10^{10}
FAC + C2F + SPPF + Loss	84.1	89.7	90.8	9.67×10^{6}	1.840×10^{10}

分支相关代码、修改损失函数等操作，实现一个单分割头的火龙果园多任务视觉感知网络。

表 2 展示了单分割头模型与双分割头模型的可行驶区域和果树区域 mIoU 对比结果。结果表明，双分割头模型在可行驶区域分割任务 mIoU 提升 2.3 个百分点，果树区域分割任务中 mIoU 提升 4.2 个百分点。

表 2　单分割头与双分割头模型性能对比

Tab. 2　Comparison of single segmentation head and double segmentation head　%

模型	可行驶区域 mIoU	果树区域 mIoU
双分割头模型	89.7	90.8
单分割头模型	87.4	86.6

图 8 可视化对比了双分割头与单分割头的性能。具体而言，单分割头在实际应用中显现出了将部分背景错误地识别为果树区域、将部分果树误检为背景，以及可行驶区域分割不完整等问题。这些问题可归因于双分割头相较于单分割头所具备的优势——即其能独立地从图像的不同层次中抽取与各任务相关的特征，这一机制有效减少了任务间特征的相互干扰，从而提升了识别的准确性。此外，本研究根据每个任务的特点，设计了各任务各自的损失函数，进一步强化了模型在特定任务上的表现。正是这些策略和双分割头的结构性优势，共同促成了其在性能上对单分割头的优势。

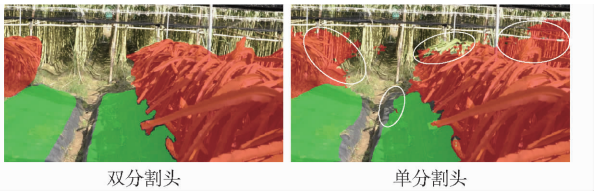


图 8　单分割头与双分割头可视化对比

Fig. 8　Visual comparison between single segmentation head and double segmentation head

2.6　计算成本与性能

表 3 展示了 OrchardYOLOP 模型与其他多任务网络的性能对比。OrchardYOLOP 模型在参数量上实现了优化，从 YOLOP 模型的 9.80×10^{6} 减少到 9.67×10^{6}，降低了 1.3×10^{5}。这一改进助于有减少模型的存储和内存需求。在计算量上，

表 3　各种多任务网络的计算成本与性能

Tab. 3　Computing cost and performance of various multitask networks

模型	参数量	计算量	实时帧率/(f·s^{-1})
OrchardYOLOP	9.67×10^{6}	1.84×10^{10}	33.33
YOLOP	9.80×10^{6}	1.86×10^{10}	31.63
HybridNets	1.34×10^{7}	1.56×10^{10}	42.95

OrchardYOLOP 模型的 1.84×10^{10} 实际上略低于 YOLOP 模型的 1.86×10^{10}，而在 FPS 上，OrchardYOLOP 模型的 33.33 f/s 则优于 YOLOP 模型的 31.63 f/s。

OrchardYOLOP 模型参数量与 HybridNets 相比显著减少，从 1.34×10^{7} 降至 9.67×10^{6}，减少了 3.73×10^{6}。尽管在计算量上 OrchardYOLOP 模型与 HybridNets 模型存在一定差距，但在目标检测、可行驶区域检测和果树区域检测这 3 个任务的精度、召回率以及平均交并比(mIoU)等多个关键性能指标上，都实现了显著提升。这些结果展示了 OrchardYOLOP 模型在执行多样化任务时的出色能力。

2.7　多任务性能

目标检测、可行驶区域分割和果树区域分割的实验结果如表 4 所示。表 4 给出了目标检测结果，并将其与其他 4 个模型在精确度、召回率和 mAP50 指标进行对比。

表 4　目标检测任务对比结果

Tab. 4　Comparison of object detection　%

模型	精确度	召回率	mAP50
OrchardYOLOP	84.1	86.5	82.9
YOLOP	68.2	73.2	68.0
HybridNets	63.9	78.1	67.3
YOLO v8	93.7	80.8	88.4
RT-DETR	81.3	67.3	77.7

表 4 展示了 OrchardYOLOP 模型与其他 4 个模型在目标检测任务上的对比结果。这些模型包括 2 个单任务目标检测模型(YOLO v8 和 RT-DETR)以及 2 个多任务模型(YOLOP 和 HybridNets)。OrchardYOLOP 精确度、召回率和 mAP50 均表现出色，与主流的单任务目标检测模型相当。虽然与

YOLO v8 模型相比,OrchardYOLOP 模型精确度和 mAP50 仍有一定差距,但是 OrchardYOLOP 模型召回率对比 YOLO v8 模型存在优势,并且 OrchardYOLOP 模型在多任务处理上具有优势。

针对非结构化环境优化的 OrchardYOLOP 模型,在目标检测任务上远超 YOLOP 模型和 HybridNets 模型。OrchardYOLOP 模型精确度、召回率和 mAP50 分别领先 YOLOP 模型 15.9、13.3、14.9 个百分点。得益于在模型底层结构中融入的深度可分卷积和 CBAM 注意力机制,这些技术显著增强了模型对目标特征的捕捉能力。

OrchardYOLOP、YOLOP、HybridNets、TransUNet、UNet 模型在可行驶区域分割任务上的 mIoU 分别为 89.7%、86.0%、63.0%、82.3%、83.0%。这些模型包括 2 个单任务分割模型(UNet 和 TransUNet)以及 2 个多任务模型(YOLOP 和 HybridNets)。OrchardYOLOP 模型的 mIoU 最佳为 89.7%。

OrchardYOLOP、YOLOP、HybridNets、TransUNet、UNet 模型在果树区域分割任务上的 mIoU 分别为 90.8%、88.7%、68.2%、84.5%、86.4%。这些模型包括 2 个单任务分割模型(UNet 和 TransUNet)以及 2 个多任务模型(YOLOP 和 HybridNets)。OrchardYOLOP 模型的 mIoU 最佳为 90.8%。

根据实验表明,OrchardYOLOP 模型在目标检测与分割任务中,展现出了显著的性能优势。对于目标检测任务,在准确识别与覆盖广泛目标方面表现优异。对于分割任务,无论是可行驶区域还是果树区域分割,OrchardYOLOP 模型都展现出了较高的分割精度。总体而言,OrchardYOLOP 模型在各项任务中的表现不仅与当前流行的单任务网络相媲美,更在各项任务的性能上显著超越了其他多任务网络。

2.8 可视化比较

图 9 和图 10 提供了 HybridNets、YOLOP 和 OrchardYOLOP 3 个多任务模型在目标检测和分割任务上的直观比较。为清晰展示,目标检测和分割结果分开呈现。图 9 聚焦于目标检测性能,其中 HybridNets 在 3 种不同场景下的表现如下:第 1 场景中未能检测到小尺寸的火龙果目标;第 2 场景中出现了大量火龙果小目标的误检;第 3 场景在复杂的人物背景下产生了人物的误检。YOLOP 模型的表现在中间列展示,第 2 场景中小尺寸火龙果目标频繁缺失,而第 3 场景中将多个人物错误地识别为单一目标。OrchardYOLOP 模型在各种测试场景下均展现出优越的性能。

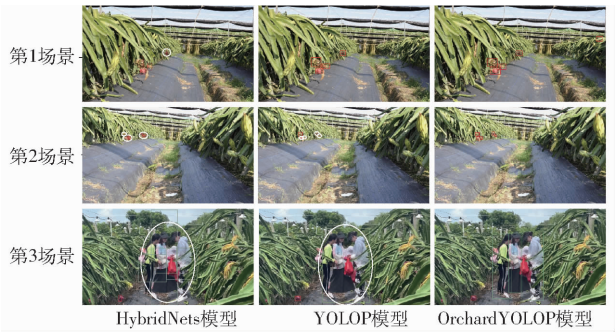


图 9 不同模型在不同条件下目标检测效果

Fig. 9 Object detection effects of different models under different conditions

在目标检测方面,HybridNets 模型虽然检测目标较全面,但准确度不高;YOLOP 模型则在准确性方面表现良好,但容易出现漏检;相比之下,OrchardYOLOP 模型在保持较高准确度的同时,大幅减少了漏检情况。

图 10 则关注于分割任务的表现。HybridNets 模型在 3 种场景下的分割结果显示,第 1 场景中可

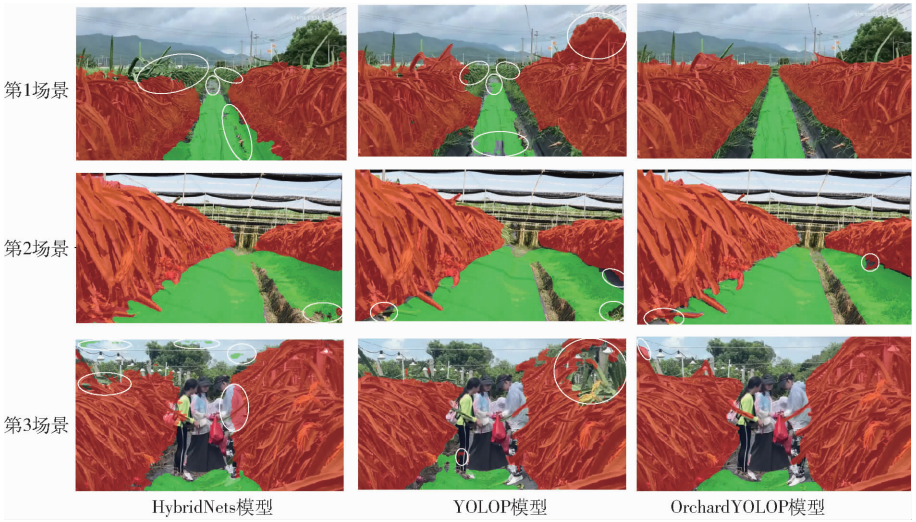


图 10 不同模型在不同条件下的分割效果

Fig. 10 Segmentation effect of different models under different conditions

行驶区域和果树区域的分割存在缺失；第 2 场景中可行驶区域缺失；第 3 场景中部分天空错误地被标记为可行驶区域，同时部分果树区域的分割也存在缺失。YOLOP 模型的结果在中间列呈现，其中第 1 场景中可行驶区域分割缺失，果树区域分割出现误报，将后方的树木错误地标记为果树；第 2 场景中部分可行驶区域未被检测到；第 3 场景中部分果树区域出现漏检和误报。OrchardYOLOP 模型的结果在右列展示，该模型在分割任务上仅在部分果树区域边缘出现轻微漏检，但这些漏检并不影响整体分割任务的性能，显示出模型的高效性能。

语义分割方面，OrchardYOLOP 模型不仅区别果树区域和背景区域区分上表现良好，还对远处场景分割能力表现优异。相比之下，YOLOP 和 HybridNets 模型在远处场景的分割上存在不足，而 OrchardYOLOP 模型则有效地克服了这一问题。

总体而言，OrchardYOLOP 模型在目标检测和分割任务上均展现出了卓越的性能，特别是在处理复杂场景和小目标检测方面，相较于 HybridNets 和 YOLOP 模型，提供了更为精确和可靠的结果。这些结果强调了 OrchardYOLOP 模型在实际应用中的潜力和优势。

3　结论

（1）针对火龙果园复杂环境下采摘机器人多任务视觉感知问题，本文提出了 OrchardYOLOP 模型。该模型可同时处理针对行人和成熟果实的目标检测、可行驶区域分割和果树区域分割 3 个视觉感知任务，且速度块、精度高。

（2）OrchardYOLOP 模型在 YOLOP 模型基础上进行改进，通过在骨干网络第一层引入焦点融合高效卷积模块，结合深度可分离卷积和 CBAM 注意力机制，增强了对复杂背景下目标检测能力；使用 C2F 模块替代了 BottleneckCSP 模块，简化了网络结构，降低了计算复杂度，提高了效率；在颈部网络中引入了一种高效的空间金字塔池化方法 SPPF 模块，加快了推理速度；最后改进分割任务损失函数，进一步提升分割任务相关性能。

（3）OrchardYOLOP 模型在目标检测任务上精确度为 84.1%，召回率为 86.5%，mAP50 为 82.9%，可行驶区域分割任务上 mIoU 为 89.7%，果树区域分割任务 mIoU 为 90.8%，参数量为 9.67×10^6，计算量为 1.84×10^{10}，与原始 YOLOP 模型相比，目标检测任务上精确度、召回率和 mAP50 分别提升 15.9、13.3、14.9 个百分点，可行驶区域分割任务、果树区域分割任务的 mIoU 也分别提升 3.7、2.1 个百分点，参数量减少 1.3×10^5，计算量减少 2.1×10^8，与其他模型相比也均有不同程度提升，表明该模型适用于火龙果园环境下多任务视觉感知。

参　考　文　献

[1] 国家统计局. 中国统计年鉴. 北京：中国统计出版社，2023.

[2] 杜鹏，李子洋. 乡村振兴的劳动力基础：现状与变迁[J]. 济南大学学报(社会科学版)，2023，33(5)：106－117.

[3] 严婉瑜，韩旭，尚永强，等. 果业数字化发展趋势及其特征分析[J]. 中国果树，2023(1)：116－121.
YAN Wanyu, HAN Xu, SHANG Yongqiang, et al. Analysis on the development trend and characteristics of digital fruit industry [J]. China Fruits, 2023(1):116－121. (in Chinese)

[4] 窦汉杰，陈震宇，翟长远，等. 果园智能化作业装备自主导航技术研究进展[J]. 农业机械学报，2024，55(4)：1－22.
DOU Hanjie, CHEN Zhenyu, ZHAI Changyuan, et al. Research progress on autonomous navigation technology for orchard intelligent equipment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(4):1－22. (in Chinese)

[5] 陈青，殷程凯，郭自良，等. 苹果采摘机器人关键技术研究现状与发展趋势[J]. 农业工程学报，2023，39(4)：1－15.
CHEN Qing, YIN Chengkai, GUO Ziliang, et al. Current status and future development of the key technologies for apple picking robots[J]. Transactions of the CSAE, 2023, 39(4):1－15. (in Chinese)

[6] 莫冬炎，杨尘宇，黄沛琛，等. 基于环境感知的果园机器人自主导航技术研究进展[J]. 机电工程技术，2021，50(9)：145－150.
MO Dongyan, YANG Chenyu, HUANG Peichen, et al. Research progress of autonomous navigation technology for orchard robots based on environment perception[J]. Mechanical & Electrical Engineering Technology, 2021, 50(9):145－150. (in Chinese)

[7] 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.
ZHU Yun, LING Zhigang, ZHANG Yuqiang. Research progress and prospect of machine vision technology[J]. Journal of Graphics, 2020, 41(6):871－890. (in Chinese)

[8] REN S Q, HE K M, GIRSHICK R, et al. Faster R－CNN: towards real-time object detection with region proposal network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137－1149.

[9] CARION N, MASSA F, SYNNAVE G, et al. End-to-end object detection with transformers[J]. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020, 12346:213－229.

[10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263－7271.

[11] REDMON J, FARHADI A. YOLO v3: an incremental improvement[J]. arXiv Preprint, arXiv:1804.02767, 2018.

[12] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-YOLO v4: scaling cross stage partial network[C]//Computer Vision and Pattern Recognition. IEEE, 2021.
[13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLO v7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464 - 7475.
[14] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. You only learn one representation: unified network for multiple tasks[J]. arXiv Preprint, arXiv:2105.04206,2021.
[15] CHEN J, LU Y, YU Q, et al. Transunet: transformers make strong encoders for medical image segmentation[J]. arXiv Preprint, arXiv:2102.04306, 2021.
[16] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5 - 9, 2015, Proceedings, Part Ⅲ 18. Springer International Publishing, 2015: 234 - 241.
[17] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881 - 2890.
[18] BACZMANSKI M, SYNOCZEK R, WASALA M, et al. Detection-segmentation convolutional neural network for autonomous vehicle perception[C]//2023 27th International Conference on Methods and Models in Automation and Robotics (MMAR). IEEE, 2023: 117 - 122.
[19] VU D, NGO B, PHAN H. Hybridnets: end-to-end perception network[J]. arXiv Preprint, arXiv:2203.09035, 2022.
[20] DONG W, LIAO M W, ZHANG W T, et al. YOLOP: you only look once for panoptic driving perception[J]. Machine Intelligence Research, 2022,19(6):550 - 562.
[21] 孟庆宽,杨晓霞,张漫,等. 基于语义分割的非结构化田间道路场景识别[J]. 农业工程学报,2021,37(22):152 - 160.
MENG Qingkuan, YANG Xiaoxia, ZHANG Man, et al. Recognition of unstructured field road scene based on semantic segmentation model[J]. Transactions of the CSAE, 2021,37(22):152 - 160. (in Chinese)
[22] 周学成,肖明玮,梁英凯,等. 基于改进 DeepLabv3 + 的火龙果园视觉导航路径识别方法[J]. 农业机械学报,2023,54(9):35 - 43.
ZHOU Xuecheng, XIAO Mingwei, LIANG Yingkai, et al. Navigation path recognition between dragon orchard using improved DeepLabv3 + network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(9): 35 - 43. (in Chinese)
[23] LI F, ZHANG H, SUN P Z, et al. Semantic-sam: segment and recognize anything at any granularity[J]. arXiv Preprint, arXiv:2307.04767, 2023.
[24] BUSLAEV A, IGLOVIKOV I V, KHVEDCHENYA E, et al. Albumentations: fast and flexible image augmentations[J]. Information (Switzerland), 2020,11(2):125.
[25] YU F, CHEN H F, WANG X, et al. Bdd100k: a diverse driving dataset for heterogeneous multitask learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2636 - 2645.
[26] WOO S, PARK J, LEE J Y. CBAM: convolutional block attention module[J]. Lecture Notes in Computer Science, 2018, 11211:3 - 19.
[27] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904 - 1916.
[28] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//Computer Vision-ECCV 2014: 13th European Conference, Proceedings, Part I 13. Springer International Publishing, 2014: 818 - 833.
[29] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018:8759 - 8768.
[30] LIN T Y, DOLLAR 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.
[31] REN S Q, HE K M, GIRDHICK R, et al. Faster R - CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137 - 1149.
[32] 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.
[33] SALEHI S S M, ERDOGMUS D, GHOLIPOUR A. Tversky loss function for image segmentation using 3D fully convolutional deep networks[J]. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017,10541:379 - 387.
[34] KINGMA D, BA J. Adam: a method for stochastic optimization[J]. arXiv Preprint, arXiv:1412.6980, 2014.